

ОДИННАДЦАТАЯ  
МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ  
ПО КОМПЬЮТЕРНОЙ ОБРАБОТКЕ  
ТЮРКСКИХ ЯЗЫКОВ  
«TURKLANG 2023»

---

Труды конференции

2023

УДК 004.8+81'32  
ББК 81.1

**Организаторы:**

**Министерство высшего образования, науки  
и инноваций Республики Узбекистан**

**Бухарский государственный университет**

**Академия наук Республики Татарстан  
Институт прикладной семиотики**

**Евразийский национальный университет имени Л. Н. Гумилёва  
Министерства образования и науки Республики Казахстан  
НИИ «Искусственный интеллект»**

**Стамбульский технический университет**

**Научные редакторы:**

Д.филол.н. Тоирова Г. И.  
Д.филол.н. Абдурахмонова Н. З.  
К.пед.н. Юлдашева Д. Н.,  
К.т.н. Гатиатуллин А. Р.

**Одиннадцатая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2023», – Труды конференции. – Казань: Издательство Академии наук Республики Татарстан, 2023. – 1030 с.**

ISBN 978-5-9690-1197-7

Сборник содержит материалы Одиннадцатой Международной конференции по компьютерной обработке тюркских языков «TurkLang-2023» (Бухара, Узбекистан, 20–22 октября 2023 г.)

Для научных работников, преподавателей, аспирантов и студентов, специализирующихся в области компьютерной лингвистики и ее приложений.

УДК 004.8+81'32  
ББК 81.1

ISBN 978-5-9690-1197-7

© Коллектив авторов, 2023

---

## ПРОГРАММНЫЙ КОМИТЕТ

1. Адалы Ешреф (Стамбул, Турция) – сопредседатель
2. Сулейманов Джавдет Шевкетович (Казань, Татарстан, РФ) – сопредседатель
3. Шарипбаев Алтынбек Амирович (Астана, Казахстан) – сопредседатель
4. Хамидов Обиджон Хафизович (Бухара, Узбекистан) – сопредседатель
5. Абдурахмонова Нилуфар Зайнобиддиновна (Ташкент, Узбекистан)
6. Алтынбек Гулила (Урумчи, Китай)
7. Арипов Мирсаид Мирсидикович (Ташкент, Узбекистан)
8. Бекманова Гульмира Тылеубердиевна (Астана, Казахстан)
9. Гатиатуллин Айрат Рафизович (Казань, Татарстан, РФ)
10. Гильмуллин Ринат Абрекович (Казань, Татарстан, РФ)
11. Дыбо Анна Владимировна (Москва, РФ)
12. Ергеш Бану Жантуганкызы (Астана, Казахстан)
13. Исраилова Нелла Амантаевна (Бишкек, Кыргызстан)
14. Кубединова Ленара Шакировна (Симферополь, Крым, РФ)
15. Мамедова Масума Гусейновна (Баку, Азербайджан)
16. Назирова Элмира Шодмоновна (Ташкент, Узбекистан)
17. Ооржак Байлак Чаш-ооловна. (Кызыл, Тыва, РФ)
18. Офлазер Кемаль (Доха, Катар)
19. Рахимова Диана Рамазановна (Алматы, Казахстан)
20. Садыков Ташполот (Бишкек, Кыргызстан)
21. Салчак Аэлига Яковлевна (Кызыл, Тыва, РФ)
22. Сиразитдинов Зиннур Амирович (Уфа, Башкортостан, РФ)
23. Сулайманов Мухаммад-али (Симферополь, Крым, РФ)
24. Татевосов Сергей Георгиевич (Москва, РФ)
25. Тоирова Гули Ибрагимовна (Бухара, Узбекистан)
26. Торотоев Гаврил Григорьевич (Якутск, Саха (Якутия), РФ)
27. Тукеев Уалишер Ануарбекович (Алматы, Казахстан)
28. Чумакаев Алексей Эдуардович (Горно-Алтайск, Алтай, РФ)

---

## ОРГАНИЗАЦИОННЫЙ КОМИТЕТ

Председатель – Хамидов Обиджон Хафизович (Бухара, Узбекистан)

Заместитель председателя – Расулов Тулкин Хусенбович – (Бухара, Узбекистан)

### ЧЛЕНЫ ОРГАНИЗАЦИОННОГО КОМИТЕТА:

Абдурахмонова Нилуфар (Ташкент, Узбекистан)

Тоирова Гули Ибрагимовна (Бухара, Узбекистан)

Гатиатуллин Айрат Рафизович (Казань, Республика Татарстан, РФ)

Жумаев Рустам Ганиевич – (Бухара, Узбекистан)

Жураев Аброр Туробович – (Бухара, Узбекистан)

Зарипов Гулмурод Тохирович – (Бухара, Узбекистан)

Ражабов Дилшод Зарипович – (Бухара, Узбекистан)

Ахмедов Атоуллох Рахматович – (Бухара, Узбекистан)

Эшонкулов Хусниддин Пиримович – (Бухара, Узбекистан)

Юлдашева Дилором Неъматовна – (Бухара, Узбекистан)

Қиличев Байрамали Эргашевич – (Бухара, Узбекистан)

Халилова Рухсора Рауповна – (Бухара, Узбекистан)

Хамроева Нафиса Низомиддиновна – (Бухара, Узбекистан)

Хайруллаева Гуласал Файзулло кизи – (Бухара, Узбекистан)

Зарипова Азиза Кахрамоновна – (Бухара, Узбекистан)

XI INTERNATIONAL CONFERENCE  
ON COMPUTER PROCESSING  
OF TURKIC LANGUAGES  
“TURKLANG 2023”

---

Proceedings of the conference

2023

UDC 004.8+81'32  
BBC 81.1

**Organizers:**

**Ministry of higher education, science and innovation  
of the Republic of Uzbekistan**

**Bukhara State University**

**Tatarstan Academy of Sciences  
Institute of Applied Semiotics**

**L.N. Gumilyov Eurasian National University  
“Artificial Intelligence” Scientific Research Institute**

**Technical University of Istanbul (ITU)**

**Scientific editors:**

Doctor of Philology Toirova G. I.  
Doctor of Philology Abdurakhmonova N. Z.  
Ph.D. Yuldasheva D. N.,  
Ph.D. Gatiatullin A. R.

**XI International Conference on Computer Processing of  
Turkic Languages “TurkLang 2023”, – Proceedings of the con-  
ference. – Kazan: Publishing House of the Academy of Sciences  
of the Republic of Tatarstan, 2023. – 1030 p.**

ISBN 978-5-9690-1197-7

The proceedings contains materials from the XI International Conference on Computer Processing of Turkic Languages “TurkLang-2023” (Bukhara, Uzbekistan, October 20–22, 2023) For researchers, teachers, graduate students and students specializing in the field of computational linguistics and its applications.

UDC 004.8+81'32  
BBC 81.1

ISBN 978-5-9690-1197-7

© Autors, 2023

---

## PROGRAMME COMMITTEE

1. Adaly Eshref (Istanbul, Turkey) – Co-chairman
2. Suleymanov Dzhavdet (Kazan, Tatarstan, Russia) – Co-chairman
3. Sharipbayev Altynbek (Astana, Kazakhstan) – Co-chairman
4. Khamidov Obidjon (Bukhara, Uzbekistan) – Co-chairman
5. Abdurakhmonova Nilufar (Tashkent, Uzbekistan)
6. Altenbek Gulila (Urumchi, China)
7. Aripov Mirsaid (Tashkent, Uzbekistan)
8. Bekmanova Gulmira (Astana, Kazakhstan)
9. Gatiatullin Ayrat (Kazan, Tatarstan, Russia)
10. Gilmullin Rinat (Kazan, Tatarstan, Russia)
11. Dybo Anna (Moscow, Russia)
12. Ergesh Banu (Astana, Kazakhstan)
13. Israilova Nella (Bishkek, Kyrgyzstan)
14. Kubedinova Lenara (Simferopol, Crimea, Russia)
15. Mamedova Masuma (Baku, Azerbaijan)
16. Nazirova Elmira (Tashkent, Uzbekistan)
17. Oorzhak Baylak (Kyzyl, Tuva, Russia)
18. Oflazer Kemal (Doha, Qatar)
19. Rakhimova Diana (Almaty, Kazakhstan)
20. Sadykov Tashpolot (Bishkek, Kyrgyzstan)
21. Salchak Aelita (Kyzyl, Tuva, Russia)
22. Sirazitdinov Zinnur (Ufa, Bashkortostan, Russia)
23. Sulaimanov Muhammad Ali (Simferopol, Crimea, Russia)
24. Tatevosov Sergei (Moscow, Russia)
25. Torotoev Gavril (Yakutsk, Saha, Russia)
26. Tukeev Ualisher (Almaty, Kazakhstan)
27. Toirova Guli (Bukhara, Uzbekistan)
28. Chumakaev Aleksey (Gorno-Altaysk, Altay, Russia)

## ORGANIZING COMMITTEE

Chairman – Khamidov Obidjon Khafizovich (Bukhara, Uzbekistan)  
Deputy Chairman – Rasulov Tulkin Husenovich (Bukhara, Uzbekistan)

## MEMBERS OF THE ORGANIZING COMMITTEE:

Abdurakhmonova Nilufar (Tashkent, Uzbekistan)  
Toirova Guli Ibragimovna (Bukhara, Uzbekistan)  
Gatiatullin Airat Rafizovich (Kazan, Republic of Tatarstan, Russian Federation)  
Zhumaev Rustam Ganievich (Bukhara, Uzbekistan)  
Zhuraev Abror Turobovich (Bukhara, Uzbekistan)  
Zaripov Gulmurod Tohirovich (Bukhara, Uzbekistan)  
Rajabov Dilshod Zaripovich (Bukhara, Uzbekistan)  
Akhmedov Atoullloh Rakhmatovich (Bukhara, Uzbekistan)  
Eshonkulov Khusniddin Pirimovich (Bukhara, Uzbekistan)  
Yuldasheva Dilorom Nematovna (Bukhara, Uzbekistan)  
Kilichev Bairamali Ergashevich (Bukhara, Uzbekistan)  
Khalilova Rukhsora Raupovna (Bukhara, Uzbekistan)  
Hamroeva Nafisa Nizomiddinovna (Bukhara, Uzbekistan)  
Xayrullaeva Gulasal Fayzullo kizi (Bukhara, Uzbekistan)  
Zaripova Aziza Kakhramonovna (Bukhara, Uzbekistan)



“TURKLANG 2023  
TURKIY TILLARNI KOMPYUTERDA QAYTA  
ISHLASH” NOMLI XI XALQARO  
KONFERENSIYA

---

Konferensiya materiallari

UOT 004.8+81'32  
KBT 81.1

**Tashkilotchilar:**

**O'zbekiston Respublikasi oliy ta'lim, fan va innovatsiyalar Vazirligi**

**Buxoro davlat universiteti**

**Tatariston Fanlar Akademiyasi**  
Amaliy semiotika instituti

**L.N. Gumilyov nomidagi Yevrosiyo milliy universiteti**  
“Sun'iy intellekt” ilmiy tadqiqot instituti

**Istanbul texnik universiteti (itu)**

**Mas'ul muharrir:**

Filologiya fanlari doktori Toirova G. I.  
Filologiya fanlari doktori Abduraxmonova N. Z.,  
Pedagogika fanlari nomzodi Yuldasheva D. N.,  
Texnika fanlari nomzodi Gatiatullin A. R.

**Turkiy tillarni kompyuterda qayta ishlash bo'yicha o'n birinchi xalqaro konferensiya “TurkLang 2023”** – Konferensiya materiallari. – Qozon: Tatariston Respublikasi Fanlar akademiyasining nashriyoti, 2023. – 1030 p.

ISBN 978-5-9690-1197-7

To'plamda turkiy tillarni kompyuterda qayta ishlash bo'yicha “TurkLang-2023” o'n birinchi xalqaro konferensiyasi (Buxoro, O'zbekiston, 20–22 oktyabr, 2023-yil) materiallari o'rin olgan. Mazkur konferensiya to'plami Kompyuter lingvistikasi va uning yo'nalishlariga doir sohalarda izlanish olib borayotgan tadqiqotchilar, professor-o'qituvchilar, tayanch doktorantlar va talabalar uchun mo'ljallangan.

UOT 004.8+81'32  
KBT 81.1

ISBN 978-5-9690-1197-7

© Mualliflar jamoasi, 2023

## DASTUR QO‘MITASI

1. Suleymanov Djavdet (Qozon, Tatariston, Rossiya) – hamrais
2. Sharipbayev Oltinbek (Nur-Sulton, Qozog‘iston) – hamrais
3. Adali Eshref (Istanbul, Turkiya) – hamrais
4. Xamidov Obidjon Xafizovich (Buxoro, O‘zbekiston) – hamrais
5. Abduraxmonova Nilufar (Toshkent, O‘zbekiston)
6. Altenbek Gulila (Urumchi, Xitoy)
7. Aripov Mersaid (Toshkent, O‘zbekiston)
8. Bekmanova Gulmira (Nur-Sulton, Qozog‘iston)
9. Gatiatullin Ayrat (Qozon, Tatariston, Rossiya)
10. Gilmullin Rinat (Qozon, Tatariston, Rossiya)
11. Dybo Anna (Moskva, Rossiya)
12. Ergesh Banu (Nur-Sulton, Qozog‘iston)
13. Isroilova Nella (Bishkek, Qirg‘iziston)
14. Kubedinova Lenara (Simferopol, Qrim, Rossiya)
15. Mamedova Masuma (Boku, Ozarbayjon)
16. Nazirova Elmira (Toshkent, O‘zbekiston)
17. Oorjak Baylak (Qizil, Tuva, Rossiya)
18. Oflazer Kamol (Doha, Qatar)
19. Rahimova Diana (Olmaota, Qozog‘iston)
20. Sodiqov Toshpolot (Bishkek, Qirg‘iziston)
21. Salchak Aelita (Qizil, Tuva, Rossiya)
22. Sirazitdinov Zinnur (Ufa, Boshqirdiston, Rossiya)
23. Sulaymonov Muhammad Ali (Simferopol, Qrim, Rossiya)
24. Tatevosov Sergey (Moskva, Rossiya)
25. Torotoyev Gavril (Yakutsk, Saha, Rossiya)
26. Tukeyev Ualisher (Olmaota, Qozog‘iston)
27. Toirova Guli Ibragimovna (Buxoro, O‘zbekiston)
28. Chumakaev Aleksey (Gorno-Oltaysk, Oltoy, Rossiya)

## TASHKILIY QO‘MITA

Rais – Xamidov Obidjon Xafizovich (Buxoro, O‘zbekiston)  
Rais o‘rinbosari – Rasulov To‘lqin Xusenbovich – (Buxoro, O‘zbekiston)

## TASHKILIY QO‘MITA A‘ZOLARI

Abduraxmonova Nilufar (Toshkent, O‘zbekiston)  
Toirova Guli Ibragimovna (Buxoro, O‘zbekiston)  
Gatiatullin Ayrat (Qozon, Tatariston, Rossiya)  
Amonov Azizbek Amonovich – (Buxoro, O‘zbekiston)  
Jumayev Rustam G‘aniyevich – (Buxoro, O‘zbekiston)  
Jo‘rayev Abror Turobovich – (Buxoro, O‘zbekiston)  
To‘xсанov Qahramon Rahimboyivich – (Buxoro, O‘zbekiston)  
Rajabov Dilshod Zaripovich – (Buxoro, O‘zbekiston)  
Eshonqulov Xusniddin Pirimovich – ( Buxoro, O‘zbekiston)  
O‘rinboyeva Dilbar Bazarovna – (Samarqand, O‘zbekiston)  
Hamroyeva Shahlo Mirjanovna – (Toshkent, O‘zbekiston)  
Yuldasheva Dilorom Ne‘matovna – (Buxoro, O‘zbekiston)  
Astanova Gulnora Aminovna – ( Buxoro, O‘zbekiston)  
Qilichev Bayramali Ergashevich – (Buxoro, O‘zbekiston)  
Ne‘matova Lola Hamidovna – (Buxoro, O‘zbekiston)  
Xalilova Ruxsora Raupovna – (Buxoro, O‘zbekiston)  
Hamroyeva Nafisa Nizomiddinovna – (Buxoro, O‘zbekiston)  
Xayrullayeva Gulasal Fayzullo kizi – ( Buxoro, O‘zbekiston)  
Ahmadova Umida Shavkatovna – ( Buxoro, O‘zbekiston)  
Hojiyeva Marifat – ( Buxoro, O‘zbekiston)  
Zaripova Aziza Kaxramonovna (Buxoro, O‘zbekiston)

## ПРЕДИСЛОВИЕ

В этом году на базе Бухарского государственного университета прошла уже одиннадцатая международная конференция по компьютерной обработке тюркских языков TurkLang-2023.

Предыдущие 10 конференций прошли в Астане (2013, 2022), Стамбуле (2014), Казани (2015, 2017), Бишкеке (2016), Ташкенте (2018), Симферополе (2019), Уфе (2020), Кызыле (2021). География проведения, количество представленных трудов и состав участников конференции подтверждают, что в настоящее время тематика конференции продолжает оставаться весьма актуальной. Целью международной конференции TurkLang является создание пространства совместных компьютерных лингвистических исследований для тюркских языков. На конференции представляются качественно новые результаты, связанные с разработкой компьютерных лингвистических моделей, приложений для тюркских языков. Тюркские языки обладают целым рядом структурно-функциональных особенностей, благодаря которым простой перенос решений, получаемых на материале других языков (в том числе английского и русского) порой практически невозможен. В последние годы направление компьютерной лингвистики бурно развивается в целом ряде университетов Республики Узбекистан, поэтому выбор места проведения конференции был не случаен. Организаторы надеются, что проведение этой конференции на базе Бухарского государственного университета послужит еще большему развитию исследований в области компьютерной обработки тюркских языков в Республике Узбекистан.

В сборник трудов включены статьи участников XI Международной конференции по компьютерной обработке тюркских языков «TurkLang-2023» (Бухара, Узбекистан, 20–22 октября 2023 г.). Участниками конференции, учеными и специалистами из Узбекистана, Казахстана, Кыргызстана, Турции, Азербайджана, России (Татарстан, Башкортостан, Москва, Санкт-Петербург, Саха (Якутия), Тува, Алтай и др.), были представлены доклады, посвященные актуальным проблемам компьютерной и прикладной лингвистики в плане разрешения их в контексте тюркских языков. В ходе работы конференции активно и плодотворно обсуждались вопросы разработки лингвистических моделей, электронных корпусов, а также исследования на базе корпусов, систем машинного перево-

да, речевых технологий, проблемы, связанные с функционированием тюркских языков в Интернет-пространстве.

Следует отметить, что тематика конференции находится в постоянном развитии в соответствии с развитием новых направлений исследований компьютерной лингвистики.

Организаторы и участники конференции выражают благодарность коллективу Бухарского университета за проведение конференции “TurkLang-2023” на высоком уровне.

---

## ЛИНГВИСТИЧЕСКИЕ МОДЕЛИ

УДК 81'322.2

### ЗАПРЕТЫ НА СОЧЕТАЕМОСТЬ МОРФЕМ В ХАКАССКОМ АВТОМАТИЧЕСКОМ МОРФОЛОГИЧЕСКОМ ПАРСЕРЕ

*А. В. Дыбо<sup>1</sup>, В. С. Мальцева<sup>1</sup>, Э. В. Султрекова<sup>2</sup>,  
А. В. Шеймович<sup>1</sup>, Ф. С. Крылов<sup>3</sup>*

<sup>1</sup>*Институт языкознания РАН, Москва, Россия*

<sup>2</sup>*независимый исследователь, Абакан, Россия*

<sup>3</sup>*независимый исследователь Кельн, Германия*

adybo@mail.ru, malt.wh@gmail.com, evsultrekova@gmail.com,  
asheimovich@yandex.ru, phil.krylov@gmail.com

Статья посвящена описанию морфологических правил, действующих в автоматическом парсере хакасского языка (<https://khakas.altai.ru/grammar/>). Любая автоматическая модель морфологического анализа, не включающая ограничений на сочетаемость морфем, при работе с большими объемами реального языкового материала выдаст большое количество некорректных результатов анализа, в то время как тюркские языки уже демонстрируют чрезвычайно большое количество правильных омонимичных вариантов анализа. Ограничения имеют разный статус. Некоторые из них обусловлены фундаментальным строением тюркской словоформы или несовместимостью семантических характеристик морфем. Другая часть является следствием того, что автомат нацелен на практическое применение для анализа текстов, входящих в корпус хакасского языка.

**Ключевые слова:** тюркские языки, хакасский язык, корпусная лингвистика, морфология, грамматика, автоматическая обработка языка.

### THE CONSTRAINTS ON THE COMPATIBILITY OF AFFIXES IN THE AUTOMATIC PARSER FOR THE KHAKAS LANGUAGE

*Anna Dybo<sup>1</sup>, Vera Maltseva<sup>1</sup>, Elvira Sultrekova<sup>2</sup>,  
Aleksandra Sheimovich<sup>1</sup>, Filipp Krylov<sup>3</sup>*

<sup>1</sup>*Institute of Linguistics of the Russian Academy of Sciences  
Moscow, Russia*

<sup>2</sup>*independent researcher, Abakan, Russia*

<sup>3</sup>*independent researcher, Cologne, Germany*

adybo@mail.ru, malt.wh@gmail.com, evsultrekova@gmail.com,  
asheimovich@yandex.ru, phil.krylov@gmail.com

The paper is devoted to the description of morphological rules operating in the automatic parser of the Khakas language (<https://khakas.altai.ru/grammar/>).

When working with large volumes of actual linguistic material, any automatic model of morphological analysis, which does not include constraints on the compatibility of morphemes, will come up with a large number of incorrect analyses, while the Turkic languages already show an extremely high amount of correct homonymous analyses. The constraints have a different status. Some of them are due to the fundamental structure of the Turkic word form or the incompatibility of the semantic characteristics of morphemes. The other part is a consequence of the fact that the automaton is aimed at practical application, the analysis of texts included into the Khakas Language Corpus.

**Key words:** Turkic languages, Khakas language, corpus linguistics, morphology, grammar, automatic language processing

The Electronic Corpus of the Khakas language (<https://khakas.altaica.ru>) with more than 500 thousand word forms has been successfully functioning for several years. This is a parallel Khakas-Russian corpus with morphological parsing of Khakas word forms, which is performed by an automatic parser. In this paper we want to present the types of rules used in this parser describing the mutual compatibility of morphological markers. While defining the rules of the parser, we consider not only the available grammatical descriptions of the Khakas language and its dialects, but also the material of the corpus itself. Thus, a cyclic process takes place: the more textual data we have analyzed, the more precisely we can formulate the parsing rules and the faster and better we can process the new data.

We have already written about the peculiarities of the ideology and structure of the automatic morphological analyzer working on the corpus website ([Maltseva 2004; Dybo, Sheimovich, Krylov 2016; Dybo et al. 2019; Dybo, Sheimovich 2014]). Both for the purposes of automatic analysis and for the purposes of describing the morphology of Turkic languages, we believe the most adequate is grammar of orders with cycles (that is, with the operation of repeating sequences of slots). About the grammar of orders cf. [Gleason 1959: 164] (original in [Gleason 1955: 112]).

The grammar of orders is a convenient tool for describing agglutinative languages whose morphology meets the following requirements: a) a fixed sequence of word-forming affixes; b) their grammatical unambiguity (absence / rarity / insignificance of cumulation); c) a single occurrence in a given word-form of a marker of a certain grammeme, cf. [Gleason 1959: 164] (original in [Gleason 1955: 112]): “Orders are mutually exclusive classes of morphemes which occupy a certain place in the sequence of morphemes forming a word”; d) low



degree of fusion, high degree of phonetic integrity of a morpheme: each grammatical meaning is expressed by an unbroken phonetic chain having a single phonetic prototype (see, e.g., [Plungian 2001] for more details on this). But for the purposes of describing the grammar of Turkic languages (as well as a number of other agglutinative languages – see, e.g., [Volodin 2004]) – this model should be supplemented. Our material clearly shows that the obligatory and one-time expression of grammatical categories within a word-form is not characteristic for Turkic languages. First, each grammatical slot can be filled in or empty, and morphological zeros in general case should not be introduced in the description [Guzev, Nasilov 1970; Solntsev et al. 1979:– 9-10]; cf. [Melchuk 1997: 247-250]). Practical consequence is that our search interface allows to search for word-forms with unfilled slots for all selected categories: - we can search for singular number, and we can search for word forms without distributive index. Secondly, grammatical categories with the same name can be expressed several times within a word-form, i.e. orders can be repeated (forming cycles) in the word-form structure. In Khakas language there is a possibility of cyclic use of affixes of possessive, case and differently expressed categories of number and negation. The cluster of number, possessive and case categories (in the indicated order) is observed in two places in Khakas word-forms: before and after transposition affixes.

An essential part of our model are rules for constraints on the compatibility of affixes within a word-form. For a detailed description of the constraints applied in the parser see [Dybo & al. 2023]. Here we give classification and examples of different types of these constraints.

In the classical ordinal model there are no constraints on compatibility (cf. the already mentioned work [Gleason 1955: 112]<sup>1</sup>; for more details on the history of the issue see [Volodin 2004]. [Volodin 2004]). Usually, however, a single model for all modifiable grammatical classes is not constructed either. It is important to note that before the development of automatic parsers, no grammatical models had been seriously tested in practice, and therefore the details of its design, in particular, constraints, could be neglected. When analyzing the actual language material, it turned out that the model without constraints produces a large number of incorrect parsings.

The number of correct homonymous parsings in Khakass is already extremely large.

– Lexical homonyms:

As an example, due to the fact that ProtoTurkic *\*s*, *\*š*, *\*d* and *\*č* be-

came *s/z* in modern literary Khakass, a large number of stems became homonymous:

therefore

*as* means ‘few’; ‘hungry’; ‘grain/food’; ‘weasel’, ‘ridge’,  
and *as-* means ‘to exceed’; ‘to stray’; ‘to open’; ‘to waste’

– Grammatical homonymy

There is also homonymy of affixes: for example, markers of the causative and the indicative TIR look the same. The causative marker was inherited from ProtoTurkic, and the indirective marker came from an analytical construction with the verb *tur-* ‘stand’

causative:

*as* ‘open!’ - *astyr* ‘make somebody open!’

indirective / mirative past:

*astyr* (< *asyp tur*) ‘he opened (as I know)’

Restrictions do not forbid homonymy of this kind, but only prohibit impossible parsings.

For example, the word *xarbir*, besides the correct parsing (*xarba-ar*: take\_in handfuls-Fut) ‘will take in handfuls’ also had an incorrect one (#*har-ba-ar*: grow old-Neg-Fut ‘will not grow old’). In fact, the negative form of the future tense in Turkic languages is cumulative, ‘will not grow old’ is expressed differently: *xarbas*. After the introduction of constraints, the markers of both simple future Fut Ar and negative future Neg.Fut PAs are forbidden to occur in the same word-form with the marker of negation PA (Restr 11), so the second parsing does not exist.

*xarbir*

<sup>OK</sup> *xarba-ar* ‘take\_in handfully-Fut’

\* *xar-ba-ar* ‘get\_old-Neg-Fut’

Restriction: Neg PA + Fut Ar is prohibited, a portmanteau morpheme Neg.Fut PAs can be only used for this form.

## Types of constraints

Restrictions on the compatibility of affixes can be classified as two different types.

**Principal constraints** are due to the principle structure of the Turkic word-form or incongruity of semantic characteristics of morphemes. Here are i.e., constraints for some types of stems (nominal / verbal) or

affixes (dialectal / coming from the auxiliary verbs), compatibility of the personal markers, etc.

**Occasional constraints** are a consequence of the fact that the practical purpose of the automatic analyzer is to analyze genuine texts that have been included in the Corpus of the Khakas language. It should be noted that Khakas literary grammar is poorly normed, so most texts also reflect the peculiarities of the dialects spoken by their authors. Dialectal norms differ quite significantly from each other and are not fully studied. In this regard, if we are not sure whether some parsing is possible, we do not forbid it. The introduction of new texts into the corpus can (and often has) led to a revision of the rules of the automatic analyzer, up to and including changing the order of slots. For the exploratory generative model, the construction of the entire set of principally possible forms is constrained by the actual presence of forms in the texts at the moment. Thus, there are rules that can be modified as the corpus grows further. But when some rarely used and semantically strange combination of the affixes is homonymic with some frequent combination, we are forced to impose a constraint prohibiting it until the validity of the combination is proven.

## Examples of principal constraints

### Semantic contradiction

Cunctative *-GALAK* can't be used in combination with the negative affixes (Neg *PA*, Neg.NF *Pin*, Neg.Conv *Pin*, Neg.Conv.Abl *PinAn*) or the variants of the perfective affix – because of its meaning.

*pīs-xalax* ‘is not cooked yet’

The constraint is due to the semantics of this form. Negative and antiperfective semas are already expressed in this form. Two negations in one form are not permitted. Cunctative *GALAK* is also not combined with durative forms, but we do not prescribe this prohibition in the constraints on combinability of morphemes for the parser – until we encounter any incorrect parsings that would require it.

### Dialectal combinations

Shor negative converb affix is *-PAA<sub>n</sub>* while in other dialects and in literary language it is *-Pin*

Negative present = (former) negative converb affix + present affix

Shor dialect *at-paan-ža* ‘doesn't shoot’

Saghay, Xaas, Lit. *at-pin-ža* ==

For the Shor negative converb affix *-PAAn* there is a rule that forbids its combining with the present forms of the other dialects (but not with the literary forms of these markers).

\* *at-paan-žadyr*

shoot-NegConv<sub>Shor</sub> -Pres<sub>Xaas</sub>

\* *at-paan-tur* / \**at-paan-dur*

shoot-NegConv<sub>Shor</sub> -Pres<sub>Kyzyl</sub>

### Combinations of variants of the affixes

Brief variants of the 2nd person markers *-ŋ* and *-ŋAr* can only combine with conditional *-SA*, recent past *-TI* and particles, while other non-imperative forms combine only with the full variants *-SIŋ* and *-SAr*

|                    |   |
|--------------------|---|
| <i>pas-ti-ŋ</i>    | (write-RPast-2sg <sub>Br</sub> ) ‘you wrote recently’         |
| <i>pas-xa-ŋ</i>    | * <i>pas-xa-ŋ</i> (write-Past-2sg <sub>Br</sub> ) ‘you wrote’ |
|                    | OK <i>pasxa-ŋ</i> (hammer-2pos.sg) ‘your hammer’              |
| <i>pas-xa-ziiŋ</i> | (write-Past-2sg <sub>Full</sub> ) ‘you wrote’                 |

This constraint is due to the following factors. There are three sets of personal verb markers in Khakas: brief, full and mixed ones [see Dybo et al. 2023 for details]. Their distribution in verb forms is historically unstable and varies across dialects. So, the distribution of 1sg forms in the literary language is close to morphophonological: after vowel-ending morphemes, the variant (*Im*) is more common, after consonant-ending morphemes *-PIŋ* (in the Sagai and Central Xaas dialects, *-SIŋ*). It is a result of an innovative morphophonological process. For names in predicative position this rule does not apply: *min xīspīn* ‘I am a girl’, *min mindabīn* ‘I am here’ – both after a vowel and after a consonant the archaic ending *PIŋ* is used. This situation resulted from the common Turkic process of gradual penetration of “short” endings from the paradigm of preterite on *\*-dI* into the paradigms of tenses formed from participle forms and therefore initially using nominal personal markers, namely postpositive personal pronouns. In each Turkic idiom we find different stages of this process. So far we have not met in the Khakas corpus any varieties only for the forms of the 2nd person. We cannot rule out, however, that this is not the case in some colloquialisms or in some speakers. If the variation is found, the constraint parameters will have to be changed.

### Lexical constraint

Durative / present affixes *-I(r)*, *-At*, *-It* can only be used with the stems *par-* ‘go’, *kil-* ‘come’, *apar-* ‘carry away’

*kil-i* ‘here he comes’, *par-at-syŋ* ‘you go (Kyzyl)’,  
*apar-it-se* ‘if he carry away (Saghay)’

*učuyim*            \* *učuy-i-m* (**fly-Dur-1sg**) ‘here I fly’  
                           OK *učuy-im* (**fly-Implsg**) ‘let me fly’  
*učuh parim* ‘here I fly (away)’

The verbs *bar-* and *kil-* are the only case of “irregular” verbs in Khakas. In general, verbs derived from ProtoTurkic *\*bar-* ‘to go’ and *\*gel-* ‘to come’ have special forms in other Turkic languages as well – cf., for example, the remark about the forms of the present tense in Uzbek dialects in [Shcherbak 1957: 20].

It should be noted that the usual durative marker *ČAT* can also combine with the bases *par-*, *apar-* and *kil-*. In [ГХЯ: 218] it is noted that there is a semantic difference between the forms *par-i-yan* (go-Dur1-Past) and *par-čat-xan* (go-Dur-Past) if *par-* is used in the function of auxiliary verb as a marker of aspect (cf. examples there).

### Selective compatibility of the affixes

Restrictions help not only to cut off the unnecessary, but also to allow the necessary. For example, the marker of Permissive *-TAK* can only be used in forms with the imperative meaning – i.e., after the bare stem, voice markers, distributive marker *-(G)LA* and imperative personal markers

*par-dax*            ‘well, go’  
*paradax*            \* *par-a-dax* (go-Conv<sub>A</sub> -Perm) ‘well, going’  
                           OK *par-ad-ax* (go-Dur<sub>Kyzyl</sub> -Impl+2)  
                           ‘let’s go (me and you)’

Durative / present affixes *-I(r)*, *-At*, *-It* can only combine with past *-GA(n)*, conditional *-SA*, converb *-PIn*, personal markers, or they can be used as final markers

*paradi*    \* *par-ad-i* (go-Dur<sub>Kyzyl</sub> -3pos) ‘his going’  
                           OK *par-adi* (go-GenerPres) ‘he goes usually’

The selective compatibility of the affixes is the most frequent type of constraint. Some of this cases can be explained by the semantics (cf.

cumulative) or by the origin of the markers (i.e., the affixes coming from the auxiliary verbs can't be used without affixes coming from the converb marker, despite of the fact, that the latter can be omitted in some phonological positions).

But the other cases are not clear yet – though we know about the lexical and the morphological constraints of the durative forms *-I(r)*, *-At*, *-It*, we can't explain them properly. The lexical constraint is well-known while the morphological constraint we discovered when we analyzed the corpus.

So we also have occasional constraints – they deal with the cases when we are not sure about the existence of the certain forms.

## Examples of occasional constraints

### Selective compatibility of the affixes

Allative *-SAR* can't be used immediately after the participle tense markers

*pol-ar-zar*      \* (be-Fut-All) 'toward your being'  
                   <sup>OK</sup> (be-Fut-2sg) 'you will be'

We met only substantivized participles in combination with the allative – this is not true for all the case markers, cf. using with locative:

*pol-ar-da* 'when it will be'

Inner additive particle *-TAA* can't be used with nominal stems

*tastaabīs*      \* *tas-taa-bīs* (stone-Add-1pl) 'we are even the stones'  
                   <sup>OK</sup> *tasta-ya-bīs* (throw-Past-1Pl) 'we threw'

Other particles are used rarely in this position:

*kizi-le-bin* (person-Emph-1sg) 'I'm only a human'  
*külüg-ök-pin* (hero-Ass-1sg) 'I'm also a hero'

Allowing such a possibility for the inner additive *TAA* would result in the generating a lot of extra parsing. Until we find a convincing example of *TAA* use in a nominative predicate, we have decided to rule out this possibility.

## Conclusion

In this paper, we have presented some examples of rules-constraints for an automatic parser of Khakas language and tried to demonstrate the linguistic data from which we derived them. The introduction of constraints on the co-occurrence of morphemes, of course, partly con-

tradicts what we declared at the beginning: obligatory expression is excluded from the notion of grammaticality of categories for agglutinative languages. It turns out that the obligatoriness of expression is present in some cases, but it is an implicative obligatoriness, not a classificatory one: “if X is expressed in a word-form then Y must (or must not) be expressed”. The single exception is the prohibition on the expression of proper verb categories in a name.

At the moment, applying the analyzer to the available corpus does not seem to yield incorrect parsing, although semantically and syntactically unnatural parsings occur. To exclude them, additional mechanisms are needed: a syntactic analyzer, semantic tags and filters on them (see [Dybo, Sheimovich, Krylov 2016] for some examples).

Now the parser successfully analyzes more than 95% of word forms. Most of the unanalyzed word forms are proper names, rare words, dialectal word-forms and non-adapted loanwords not included in the dictionary, productive derivatives (e.g., diminutives), words with misprints, numerals, and some punctuation marks. Nevertheless, there are still unrecorded grammatical phenomena that are rare and not described. Therefore, the list of constraints for the automatic parser can be changed in future according to the new data.

### Glosses and Abbreviations

The number after the gloss indicates another marker with the same meaning as the marker without the number (e.g. Pres1).

*Symbols:*

# – incorrect, but technically possible parsing;

\* – incorrect, but technically possible parsing;

OK – correct parsing;

. – separates the parts of cumulative grammemes (Neg.NF, Imp.1sg);

– – separates the inflectional markers in glossing: *ат-ты* horse-Acc;

= – separates derivational markers in glossing: *усті=л-бин-ібіс-кен* (hear=Pass-Neg.NF-Perf-Past) ‘стало неслышно’. has gone silent?

<sup>Sag, Хаас, Kyz, Shor</sup> at the end of the marker gloss indicates its belonging to the Sagai, Хаас, Kyzyl, Shor dialects of the Khakas language, respectively.

<sup>Br, Full</sup> at the end of the gloss of a person marker indicates whether it belongs to the short or full set of person markers

*Possessive markers:*

1pos.sg – 1st person singular of possessor;  
2pos.sg – 2nd person singular of possessor;  
1pos.pl – 1st person plural of possessor;  
2pos.pl – 2nd person plural of possessor;  
3pos – 3d person of possessor.

*Personal markers:*

1sg – 1st person singular;  
1pl – 1st person plural;  
2sg – 2nd person singular;  
2pl – 2nd person plural;  
3 – 3d person.

Abl – ablative case

Acc – accusative case

Add – additive particle

All – allative

Ass – assertive particle

Emph – emphatic particle

Cunc – cunctative

Conv<sub>A</sub> – деепричастие на -А

Dur, Dur1 – durative

Fut – future tense

Gener – common present tense

Imp – imperative

Neg – negation

NF – “not finished” | “non finitum” marker

Pl – Plural number

Pass – passive voice

Past – past tense

Perm – permissive

Person – person

Poss (pos) – possessiveness

Pres – present tense

RPast – recent past

Sg – Singular number



## REFERENCES

1. Dybo et al. 2019 – A. Dybo, V. Maltseva, A. Sheymovitch, E. Sul-trekova. The use of personal markers in the Beltyr dialect of the Khakas language from a comparative perspective. In: *Turkic Languages*. 2019. 23. P. 31–48.
2. Dybo et al. 2019 (1) – Дыбо А. В., Крылов Ф. С., Мальцева В. С., Шеймович А. В. Сегментные правила в автоматическом парсере Корпу-са хакасского языка // Урало-алтайские исследования. № 1 (32), 2019. С. 48–69. {A. V. Dybo, Ph. S. Krylov, V. S. Maltseva, A. V. Sheimovich. Segmental rules in the automatic parser for the Khakas corpus. In: *Ural-Al-taic Studies*. 2019, 1 (32). P. 48–69.}
3. Dybo et al. 2023 – А. В. Дыбо, В. С. Мальцева, А. В. Дыбо, В. С. Мальцева, Э. В. Султрекова, А. В. Шеймович, Ф. С. Крылов. Структу-ра хакасской словоформы и ограничения на сочетаемость аффиксов в автоматическом парсере хакасского языка // Урало-алтайские ис-следования. 2023, 2 (49). С. 42–75. {A. V. Dybo, V. S. Mal'tseva, A. V. Sheimovich, E. Sultrekova, F. S. Krylov. The structure of the Khakas word form and constraints on the compatibility of affixes in the automatic parser for the Khakas language. In: *Ural-Altaiic Studies*. 2023, 2 (49). P. 42–75}
4. Dybo, Sheimovich 2014 – Дыбо А.В., Шеймович А.В. Автома-тический морфологический анализ для корпусов хакасского и дре-внетюркского языков // “Научное обозрение Саяно-Алтая”. Вып. 1, 2014. С. 9–30. (<http://haknii.ru/?page=activities&subpage=magazine>) {A. V. Dybo, A. V. Sheimovich. Automatic morphological analysis for the cor-pora of the Khakas and Old Turkic languages. In: *Scientific Review of the Sayano-Altai*. 2014. No. 1. P. 9–30.}
5. Dybo, Sheimovich, Krylov 2016 – Дыбо А. В., Шеймович А. В., Крылов С.А. Расстановка семантических и деривационных тэгов в электронном хакасско-русском словаре // Российская тюркология, № 2(15), 2016, с. 28–39 {A. V. Dybo, A. V. Sheimovich, S. A. Krylov. Se-mantic and derivational tagging of the electronic Khakas-Russian dictionary. In: *Russian Turkology*. 2016, 2 (15). P. 28–39.}
6. Gleason 1955 – H. A. Gleason. Introduction to descriptive linguistics. New York: Holt, Rinehart and Winston, 1955.
7. Gleason 1959 – Г.Глисон. Введение в дескриптивную лингвисти-ку. М. 1959.
8. Guzev, Nasilov 1970 – Гузев В.Г., Насилов Д.М. О нулевых фор-мах тюркского имени существительного в различных грамматических категориях // Письменные памятники и проблемы истории культуры народов Востока. VI годичная научная сессия ЛО ИВ АН, посвящен-ная 100-летию со дня рождения В.И.Ленина. Апрель 1970 года. М.: ГРВЛ, 1970. С. 142–144. {V. G. Guzev, D. M. Nasilov. On zero forms of

the Turkic noun in various grammatical categories. In: Written monuments and problems of the history of culture of the peoples of the East. April 1970. Moscow: GRVL, 1970. P. 142–144.}

9. ГХЯ – Грамматика хакасского языка / Под ред. Н. А. Баскакова. М., 1975. {The grammar of Khakas. N. A. Baskakov (ed.). Moscow: Nauka. 1975.}

10. Maltseva 2004 – Мальцева В.С. Структура глагольной словоформы в сагайском диалекте хакасского языка (говор с. Казановка) / Дипломная работа. М., 2004. {V. S. Maltseva. The structure of the verbal wordform in the Sagai dialect of the Khakas language (sub-dialect of Kazanovka village). Master's thesis. Moscow, 2004.}

11. Melchuk 1997 – Мельчук И.А. Курс общей морфологии. Том I. Введение. Часть первая: Слово. М.: Языки русской культуры, 1997. {I. A. Melchuk. The course of general morphology. Vol. I. Introduction. Part One: The Word. Moscow: Yazyki russkoi kul'tury, 1997.} of Khakas. N. A. Baskakov (ed.). Moscow: Nauka. 1975.}

12. Plungian 2001 – V. Plungian. Agglutination and flexion. In: Language Typology and Language Universals / Sprachtypologie und Sprachliche Universalien / La typologie des Langues et les Universaux Linguistiques. Vol. 1. Berlin, New York: De Gruyter, Inc., 2001. P. 669–678.

13. Shcherbak 1957 – Щербак А.М. Способы выражения грамматических значений в тюркских языках. // ВЯ 1957, № 1. {A. M. Shcherbak. Ways of grammatical meanings expressing in Turkic languages. In: Voprosy jazykoznanija. 1957. № 1. P. 18–26.}

14. Solntsev et al. 1979 – Солнцев В.М. и др. О значении изучения восточных языков для общего языкознания // ВЯ 1979, №1. {V. M. Solntsev et al. On the Importance of Studying of Oriental Languages for General Linguistics. In: Voprosy jazykoznanija. 1979. № 1. P. 3–15.}

15. Volodin 2004 – Володин А.П. Проспект позиционной грамматики // 40 лет Санкт-Петербургской типологической школе. М.: Знак, 2004. С. 74–96 {A. P. Volodin. Prospect of Positional Grammar. In: 40th Anniversary of St. Petersburg Typological School. Moscow: Znak, 2004. P. 74–96.}

УДК 81'322.2

**ПРОБЛЕМА МОДЕЛИРОВАНИЯ ЛИНГВИСТИЧЕСКИХ  
СИНТАКСИЧЕСКИХ СЛОВСОЧЕТАНИЙ  
В ПРЕДЛОЖЕНИИ**

***О. Х. Абдуллаева***

*Ташкентский государственный университет имени  
Алишера Навои Узбекский язык и литература  
abdullayeva.oqila@navoiy-uni.uz*

Появление новых направлений, автоматический анализ лингвистических практик и анализ новейших исследований в узбекском языкознании развивались быстрыми темпами. Важной задачей является построение языковых корпусов, разработка морфоанализатора, а на следующем этапе - разработка лингвистического и программного обеспечения программы синтаксического анализатора. Именно для таких программ автоматического анализа необходимо определять лингвистические закономерности языковых единиц и разрабатывать модели. В данной статье определены общие группы лингво-синтаксических моделей словосочетаний в узбекском языке и предложены лингвистические модели.

**Ключевые слова:** словосочетание, лингвосинтаксический образец, моделирование, лингвистическая модель.

**GAPDA SO‘Z BIRIKMALARI LISONIY SINTAKTIK  
QOLIPLARINI MODELLASHTIRISH MASALASI**

***О. Х. Abdullayeva,***

*Alisher Navoiy nomidagi Toshkent davlat  
o‘zbek tili va adabiyoti universiteti  
abdullayeva.oqila@navoiy-uni.uz*

O‘zbek tilshunosligida amalga oshirilgan so‘nggi tadqiqotlarda yangi yo‘nalishlarning paydo bo‘lishi, til birliklarining lisoniy sintaktik qoliplarining aniqlanishi, lingvistik amallarni, tahlillarni avtomatik bajarish ishlari jadal rivojlandi. Til korpuslarining qurilishi, morfonalizadorning ishlab chiqilishi, keyingi qadam sifatida esa sintaktik analizator dasturining lingvistik va dasturiy ta‘minoti ishlab chiqilishi muhim vazifa hisoblanadi. Aynan mana shunday avtomatik tahlil dasturlari uchun til birliklarining lisoniy qoliplari aniqlanishi va modellari ishlab chiqilishi zarur. Mazkur maqolamizda o‘zbek tilidagi so‘z birikmalarining LSQlarining umumiy guruhlari aniqlandi va lingvistik modellar taklif qilindi.

**Калит so‘zlar:** so‘z birikmasi, lisoniy sintaktik qolip, LSQ, modellashtirish, lingvistik model.

## THE PROBLEM OF MODELING LINGUISTIC SYNTACTIC PATTERNS OF WORD COMBINATIONS IN A SENTENCE

*O. X. Abdullayeva*

*Alisher Navoi Tashkent State University of  
Uzbek Language and Literature.  
abdullayeva.oqila@navoiy-uni.uz*

The emergence of new directions, the automatic analysis of linguistic practices and analyzes in the latest researches in Uzbek linguistics have developed rapidly. The construction of language corpora, the development of a morphoanalyzer, and as the next step, the development of the linguistic and software of the syntactic analyzer program is an important task. It is for such automatic analysis programs that linguistic patterns of language units need to be determined and models developed. In this article, general groups of linguistic syntactic patterns of word combinations in Uzbek were determined and linguistic models were proposed.

**Keywords:** word combination, linguistic syntactic pattern, modeling, linguistic model.

Tilning o'ziga xos qonuniyatlari yillar mobaynida turli omillar ta'siri natijasida o'zgarishlarga uchraydi va bu o'zgarishlar shu til doirasida turli lingvistik qoliplarni shakllantiradi. Lingvistik qoliplar barcha jamiyat a'zolari lisoniy ongida bir-biriga o'xshash holda bir xil ko'rinishda shakllanib, lisoniy qolip sifatida yashaydi. Ushbu lisoniy qoliplar nutq jarayonida biri ikkinchisini takrorlamagan holda voqelalanadi. Demak, lisoniy qolip ma'lum bir tilning o'ziga xos xususiyatlari va imkoniyatlarini ochib beruvchi qonuniyatlarning inson miyasining til xotirasida to'planib, shakllanib borishidir. Bu qonuniyatlar barcha uchun umumiy bo'lsa, uning moddiylikda aks etishi nutqiy hosilani vujudga keltiradi va xususiylik sifatiga ega bo'ladi. Har qanday til o'zining qonuniyatlariga ega bo'ladi. Uning tarkibiy qismlari, ichki tuzilishi va ma'lum bir qoliplarni til xotirasida shakllantirish orqali o'rganadi va o'zlashtiradi. Tildagi turli qoliplar haqida Tatiyana Bushuy va Shahriyor Safarovlar - til tizimi turli darajadagi muayyan murakkablikka ega bo'lgan birliklar va ulardan foydalanish yig'indisidir. Til tizimiga turli qoliplar va sxemalar xos bo'lib, ular asosida turli murakkab birliklar, so'z birikmalari va jumlar yasaladi, - deya fikr bildirib o'tadilar [Bushuy, 2007: 42]. Tillar so'z yasalishi, gap qurilishi va boshqa jihatlari bilan bir-biridan farq qiladi. O'zbek tilida so'zlashuvchi kishi rus yoki ingliz tilini o'rganish jarayonida tilning barcha sathidagi lisoniy qoliplarni o'zlashtirib ongida shakllantiradi va mavjud qoliplar asosida shu tilda o'z fikrini ifodalaydi. Ko'rinadiki, lisoniy qoliplar til o'rganishda shu bilan birga tilshunoslikning zamonaviy yo'nalishlarida qo'llanilayotgan kompyuter dasturlarning lingvistik ta'minotini

yaratishda muhim rol o'ynaydi. Barcha tillar kabi o'zbek tili ham o'ziga xos va boshqalarga o'xshamaydigan alohida lisoniy qoliplarga ega. O'tgan asr mobaynida tilni lisoniy qoliplar asosida o'rganish muhim ahamiyatga ega bo'lmagan bo'lsa, bugungi kunga kelib ushbu yo'nalishdagi amaliy va nazariy ishlarga dolzarb masala sifatida qaralmoqda. Ma'lumki, tilning keng qamrovli va murakkab sathi sintaktik sath hisoblanadi. Ushbu sath gap va so'z birikmasi kabi til birliklarini o'z ichiga oladi. Ushbu birliklarni shakllantiruvchi lisoniy qoliplar – lisoniy sintaktik qoliplar (LSQ) deyiladi. Tadqiqotlarda ko'rsatilishicha, fonetika, leksika va morfologiyada bo'lgani kabi sintaksisda ham lisoniy va nutqiy jihat farqlanadi. Ma'lumki, lisoniy hodisa bevosita kuzatishda berilmaganlik (moddiylikdan holilik), miqdoran cheklilik, takrorlanuvchanlik, ijtimoiylik va majburiylik belgisiga ega, u bevosita kuzatishda berilganlik, miqdoriy cheklanmaganlik, betakrorlik, individuallik, ixtiyoriylik sifatiga ega bo'lgan nutqiy hodisaga qarama-qarshi turadi. Nutqiy sintaktik birlik sifatida nutqda qo'llaniladigan, sezgi a'zolariga ta'sir qiladigan, o'qish, yozish, aytish, eshitish mumkin bo'lgan so'z birikmasi va gap tushuniladi. **Lisoniy sintaktik birlik** esa, so'z birikmasi va gap hosil qilish qolipi. Lisoniy sathga tegishli bo'lganligi uchun ularni lisoniy sintaktik qolip (qisqacha LSQ) deb ataymiz. LSQ g'isht qolipiga o'xshaydi. Inson ongida ham so'zlash, nutqni shakllantirish maqsadida leksemalarni so'z birikmasi shakliga keltirish, gap hosil qilish qolipi mavjud. Ular LSQ, model, konstruktsiya, qurilma deb nomlansa-da, aslida bir tushunchani ifodalaydi. Masalan, *kitobni o'qimoq* kabi cheksiz birikmani chiqaradigan (ot tushum kelishigi+fel) so'z birikmasi qolipi qanday nomlanmasin, uning mohiyatiga ta'sir qilmaydi [Sayfullayeva, 2009: 115].

So'z birikmalari strukturasi morfologik kategoriyalar va til lug'at tarkibining leksik-grammatik tasnifiga asoslanib tahlil qilish mumkin bo'ladi. Shu o'rinda tipologik tahlilda yakka faktlarni emas, balki tilning alohida hodisalari o'zaro tizimlarni taqqoslash foydaliroq [Bushuy, 2007: 44]. Keltirilgan fikrlardan shuni aytish mumkinki, sintaktik jarayonlar va so'z birikmasini hosil qilishda morfologik birliklar hamda morfologik ko'rsatgichlarni tahlil qilish muhimdir. Grammatik munosabatlarda ishtirok etuvchi elementlar turli birikuv usullarini yuzaga keltiradi.

So'z birikmalarining formasi degan tushuncha (grammatik aspekt-da) mustaqil so'zlar o'rtasidagi sintaktik munosabatlarni ifodalash usullarini, sintaktik priyomlarni moslashuv, boshqaruv va bitishuv usullarini o'z ichiga oladi. So'zlar o'rtasidagi grammatik aloqani ta'minlovchi vositalar o'zbek tilida juda ko'p va xilma-xildir. Til vositalarining tabiatiga ko'ra ularni quyidagi asosiy turlarga bo'lish mumkin.

**1. Sintetik usul.** Bu usulning tub mohiyati shundan iboratki, soʻz birikmasini hosil qilgan komponentlardan biri biror qoʻshimchani oladi, shu affiks yordamida bir soʻz ikkinchi soʻzga grammatik jihatdan bogʻlanadi va fikr almashish uchun tayyor material vazifasini oʻtaydi. Bunday bogʻlanish soʻz oʻzgartiruvchi va shakl yasovchi deb atalgan qoʻshimchalar yordamida yuzaga keladi. Bunga kelishik qoʻshimchalari, egalik affikslari, sifatdosh va ravishdosh formalarini yasovchi koʻrsatkichlar kiradi. Baʼzi soʻz yasovchi affikslar ham, masalan: -li, -lik, -dagi, -day kabi yasovchilar ham soʻzlarning oʻzaro aloqaga kirishuvida ishtirok etadi. Bu va boshqa juda koʻp formal elementlar tilning butun morfologik strukturasi qamrab olgandir. Ayrim qoʻshimchalarning ish koʻrish doirasi bir qadar cheklangan, masalan, egalik va kelishik qoʻshimchalari faqat ot kategoriyasidagi soʻzlarga xos, umuman, soʻzlarning oʻzaro bogʻlanish koʻlami eʼtibori bilan yondashganda, -lar affiksi otga ham, feʼlga ham (feʼl formalariga ham) qoʻshilavradigan koʻrsatkichdir.

### **2. Analitik usul.**

Bu usulning asosiy xususiyati shundan iboratki, birikmani tashkil qilgan komponentlar hech qanday qoʻshimcha olmaydi yoki qoʻshimcha olganda ham, qoʻshimchanning soʻzlarning oʻzaro bogʻlanishidagi ahamiyati ikkinchi darajalidir. Bunday hollarda ikki mustaqil soʻzning oʻzaro grammatik munosabatga kirishuvini koʻmakchilar, umuman, yordamchi soʻzlar va har qanday bogʻlama vazifasida keluvchi soʻzlar (masalan, bor, yoʻq soʻzlari) taʼminlaydi. (injener boʻlib ishlaydi, bolasi bor xotin birikmalaridagi boʻlib, bor soʻzlarining funksiyasiga eʼtibor qilinsin).

### **3. Soʻz tartibi.**

Yuqorida keltirilgan ikkita asosiy usuldan tashqari, hozirgi oʻzbek tilida soʻz tartibi ham soʻz birikmasi komponentlarining sintaktik aloqasini taʼminlovchi muhim vositalardan sanaladi. Soʻz tartibi koʻp vaqt grammatik holatnigina emas, balki birikma ifoda etgan maʼnoni ham oʻzgartirib yuboradi:

A) ikkita olma-olma ikkita, farovon xalq-hayot farovon, pishgan olma-olma pishgan kabi birikmalarda komponentlarning oʻrin almashinuvi aniqlovchili birikmani predikativlik birikmaga, yaʼni gapga aylantirib yuborgan [Abdullayev, 1962: 10–11].

Yuqoridagi fikrlarning uzviy davomi sifatida quyidagi mulohazalarga eʼtiborimizni qaratamiz:

Soʻz birikmalarining turli-tuman konstruktiv koʻrinishlarida ishtirok qiluvchi komponentlarning bosh muchchasi (hokim soʻz) qa-

ysi so‘z turkumiga taalluqli ekanligiga qarab (asosan ot va fel), so‘z birikmalarini ikkita katta gruppaga bo‘lib o‘rganish ko‘zda tutiladi:

1. Otli so‘z birikmalari;
2. Fe‘lli so‘z birikmalari.

So‘z birikmalari sintaksisining asosini so‘zlarning o‘zaro bog‘lanish formalari-sintaktik priyomlar - bitishuv, boshqaruv va moslashuv tashkil qiladi. Grammatik bog‘lanishning eng keng tarqalgan turi bitishuv yo‘li bilan yuzaga kelgan birikuvdir. Shakl jihatdan boy va murakkabi esa boshqaruv yo‘li bilan aloqaga kirishgan bog‘lanishdir. Moslashuv yo‘li bilan yuzaga kelgan birikuv esa tarqalish e‘tibori bilan bir qadar cheklangandir [5;30].

Shuni aytish kerakki, yuqorida keltirilgan turli grammatik vositalar bog‘lanish usullari hamda komponentlar so‘z birikmalarining grammatik jihatdan mos ravishda birikuvini ta‘minlaydi, shu bilan birga turli LSQlarni shakllantiradi. So‘z birikmalarini hosil qilishda nafaqat grammatik jihatdan, balki leksik-semantik jihatdan so‘zlarning o‘zaro mos ravishda bog‘lanishi muhim ekanligi e‘tibordan chetda qolmasligi lozim.

Yuqorida keltirilgan fikrlar so‘z birikmasiga doir LSQni ilmiy nazariy jihatdan o‘rganish bo‘lsa, uni amaliy jihatdan o‘rganish o‘zbek tilidagi so‘z birikmasi doirasida mavjud LSQlarni to‘plash va nutq jarayonida uchraydigan boshqa LSQlarni aniqlashni ko‘zda tutadi.

### Otli birikma

| Bitishuvli birikma                            | Boshqaruvli birikma   | Moslashuvli birikma  |
|---|---|--|
| Ot + ot [yog‘och qoshiq, elektr chiroq]       | Ot +ga +ot [kitobga mehr, hayotga ishonch]                                      | Ot +ning +ot +e.q. [kitobning varog‘i, o‘qituvchining savoli]    |
| Olmosh + ot [barcha inson, butun olam]        | Ot +dan +ot [bobomdan yodgorlik, do‘stimdan esdalik]                            | Ot + e.q. [yaqt qadri, shahar ko‘shalari]                        |
| Son + ot [uchinchi sinf, beshta kitob]        | Olmosh +dan +ot [bizdan maktub, sizdan esdalik]                                 | Olmosh +ning +ot +e.q. [mening baxtim, barchaning orzusi]        |
| Sifat + ot [go‘zal shahar, mehribon murabbiy] | Ravish +dan +ot [yuqoridan topshiriq]   | Sifat +ning +ot+e.q. [yaxshining so‘zi, aqllining fikri]         |
| Sifatdosh +ot [kelgan mehmon, kulgan bola]    | Ot +da +ot [maktabda o‘qituvchi, zavodda ishchi]                                | Son +ning +ot+e.q. [o‘ntaning o‘rni, birinchining sovg‘asi]      |
|   | Ot +ko‘makchi +ot [savolga ko‘ra javob, yosh bilan suhbat]                      | Ravish +ning +ot +e.q. [kunning hikmati, ertaning ishonchi]      |
|   | Ot +dan+ko‘makchi+ot [navbatdan tashqari ushrashuv, mavzudan tashqaridagi fikr] | Sifatdosh +ning +ot+e.q. [so‘zlaganning so‘zi, yurganning yo‘li] |
|   | Olmosh +ko‘makchi +ot [biz bilan uchrashuv, hamma bilan tanishuv]               |  |

## Sifatli birikmalar

### Bitishuvli birikma

O'zbek tilida sifatli bitishuv mavjud emas. Avval aytganimizdek, bitishuv yo'li bilan birikadigan so'z birikmalarining hokim qismi sifat so'z turkumi bilan ifodalana olmaydi. Qachonki so'z birikmasi qismlari ana shu tarzda shakllansa ushbu grammatik munosabat tushuncha emas, balki fikr ifodalaydi.

### Boshqaruvli birikma

- 1.Ot +ga +sifat [*mehrga muhtoj, baxtga intizor*]
- 2.Ot +dan +sifat [*hayotdan mamnun, ilmdan xabardor*]
- 3.Ot +da +sifat [*maktabda a'lochi, ishda faol*]
- 4.Ot +ko'makchi sifat [*sham kabi yorug', quyosh kabi issiq*]
- 5.Olmosh +ga+ sifat [*hammaga manzur, menga ayon*]
- 6.Olmosh +dan +sifat hech kimdan so'roqsiz, bizdan bexabar
- 7.Olmosh +ko'makchi +sifat [*siz kabi sadoqatli, barcha kabi faol*]
- 8.Sifat +ga +sifat [*sovuqqa chidamli, issiqqa beparvo*]
9. Sifat + da +sifat [*g'oyatda go'zal, bag'oyatda mamnun*]
- 10.Ravish +ga +sifat [*ochlikka mahkum, ko'pga mehribon*]
- 11.Ravish +dan +sifat [*avvaldan mehribon, bugundan ma'lum*]
- 12.Ravish +da +sifat [*o'z vaqtida mashhur, paytida taniqli*]
- 13.Sifatdosh +ga +sifat [*kutganga aziz*]
- 14.Modal +da +sifat [*borligida beqadr, kerakligida muhim*]

### Moslashuvli birikma

- 1.Ot+ning +sifat +e.q. [*nonning issig'i, bolaning begonasi*]

## Ravishli birikmalar

### Boshqaruvli birikma

- 1.Ot +ga +ravish [*ishga loqayd, hayotga beparvo*]
- 2.Ot +dan +ravish [*uydan baland, devordan past*]
- 3.Ot +ko'makchi +ravish [*yil kabi uzun, osmon kabi yuksak*]
- 4.Olmosh +ga +ravish [*menga baribir, barchaga barobar*]
- 5.Olmosh +dan +ravish [*undan uzoq, hammadan keyin*]

### Moslashuvli birikma

- 1.Ot+ning +ravish +e.q. [*gapning ozi, uyning orqasi*]

## Sonli birikmalar

### Boshqaruvli birikma

- 1.Ot +lar +dan +son +e.q. [*kitoblardan biri, o'quvchilardan ikkita*]



2.Olmosh +lar +dan + son +e.q. [*ulardan biri*]

### **Moslashuvli birikma**

1.Ot + ning + son+e.q. [*savollarning biri*]

2.Ravish +lar+ning +son + e.q. + da [*kunlarning birida*]

3.Olmosh +lar +ning +son +e.q. [*sizlarning biringiz*]

### **Modalli birikma**

#### **Boshqaruvli birikma**

1.Ot +da +modal [*o'quvchilarda bor, rejada yo'q*]

2.Ot +ga +modal [*onamga zarur, do'stinga kerak*]

3.Ot +ko'makchi +modal [*hayot uchun zarur, inson uchun shart*]

3.Olmosh +da +modal [*bizda bor, hammada yo'q*]

4.Olmosh +ga+modal [*unga zarur, bizga kerak*]

5.Olmosh +ko'makchi + modal [*siz uchun kerak, barcha uchun zarur*]

### **Fe'li birikmalar**

#### **Bitishuvli birikma**

1.Ot +fe'1 [*quyoshdek porlamoq, guldek ochilmoq*]

2.Sifat +fe'1 [*chiroyli yozmoq, yolg'on gapirmoq*]

3.Ravish +fe'1 [*tez yurmoq, asta chaqirmoq*]

4.Son +fe'1 [*bitta ko'rmoq, birinchi kirmoq*]

5.Olmosh +fe'1 [*o'zi bilmoq*]

#### **Boshqaruvli birikma**

1.Ot +ni +fe'1 [*aqlni ishlatmoq, she'rni o'qimoq*]

2.Ot +ga +fe'1 [*maktabga bormoq, ko'chaga chiqmoq*]

3.Ot +da +fe'1 [*uyda o'tirmoq, bog'da yurmoq*]

4.Ot +dan +fe'1 [*osmondan tushmoq, nazardan qolmoq*]

5.Ot +ko'makchi +fe'1 [*baxt uchun kurashmoq, qalam bilan chizmoq*]

6.Sifat +ni +fe'1 [*shirinini yemoq, eskini tashlamoq*]

7.Sifat +ga +fe'1 [*sovuqqa chidamoq, yorug'ga chiqmoq*]

8.Sifat +da +fe'1 [*issiqda yurmoq, qorong'ida uxlamoq*]

9.Sifat +dan +fe'1 [*muzdagidan ichmoq, achchig'idan olmoq*]

10.Sifat +ko'makchi +fe'1 [*yaxshi bilan yurmoq, bechora kabi yashamoq*]

11.Olmosh +ni +fe'1 [*shuni aytmoq, hammani chaqirmoq*]

12.Olmosh +ga +fe'1 [*bizga ko'rsatmoq, barchaga bermoq*]

13.Olmosh + da +fe'1 [*o'zida saqlamoq*]

14.Olmosh +dan +fe'1 [*o'zdan kechmoq, barchadan nafratlanmoq*]

15.Olmosh+komakchi+fe'1 [*hamma uchun ishlamoq, shu haqida gapirmoq*]

16.Ravish +ni +fe'1 [*uzoqni ko'zlamoq, oldini to'smoq*]

17.Ravish +ga +fe'1 [*tashqariga qaramoq, pastga tushmoq*]

18.Ravish +da +fe'1 [*yuqorida o'tirmoq, ichida o'ylamoq*]

19.Ravish +dan +fe'1 [*yaqindan ko'rmoq, olisdan kuzatmoq*]

20.Ravish +ko'makchi +fe'1 [*orqa tamondan kelmoq, ichkari orqali o'tmoq*]

Ushbu LSQlar asosida keltirilgan misollardan tashqari boshqa ko'plab so'z birikmalarni hosil qilish mumkin. So'z birikmasining bog'lanish usullari yuzasidan keltirilgan tasniflarga asoslanib, tobe qism sifatdosh va ravishdosh bilan ifodalangan birikmalarni boshqaruvli birikmalar qatoriga kiritdik va quyida ularning LSQlarini berib o'tamiz [Sayfullayeva, 2009: 85].

#### **Sifatdoshli boshqaruv**

Fe'1 +gan (r, ar) +ot [*kutilgan mehmon, oqar suv*]

Fe'1+gan (gani)+sifat [*gapirgan aybdor, o'qigan foydali, topgani barakali, aytgani xayrli*]

Fe'1 +gan +ravish [*otgan tong, o'tgan kecha*]

Fe'1 +gani (ko'makchi) + fe'1 [*ko'rgani bormoq, bilgani uchun aytmoq*]

#### **Ravishdoshli boshqaruv**

Fe'1+gach (kach, qach, guncha, kuncha, quncha, b, ib, y, ay)+fe'1 [*ko'rgach quvonmoq, kelguncha kutmoq, kulib gapirmoq, beraqochmoq*]

Yuqorida so'z birikmalarining o'zaro grammatik bog'lanishi hamda bog'lanish turiga ko'ra turli konstruksiyalarning qurilishi yuzasidan olib borilgan ilmiy tadqiqot ishlarida aks etgan LSQlarning bir qancha turlari keltirib o'tildi.

Bu o'rinda olib borilgan ilmiy tadqiqot ishlarida tilshunoslar nazaridan chetda qolgan, tilimizda mavjud bo'lgan so'z birikmalarining ayrim LSQlari turli misollar asosida tahlil qilinib, yuqoridagi LSQlar qatoriga kiritildi. Bu borada bildirilgan ko'plab ilmiy xulosalarda moslashuvli birikmalarda birikmaning hokim qismi ot va boshqa barcha otlashgan so'z turkumlari bilan ifoddalanishi nazarda tutiladi. Biroq shunday birikmalar ham borki, hokim qism sifat, ravish, son so'z turkumlari bilan ifodalanib, tobe so'z bilan qaratqich va qaralmish munosabatini shakllantirsa-da otlashish xususiyatiga ega bo'lmaydi. Masalan: ot+ning+sifat *noning issig'i*, ot +ning +ravish *gapning ozi*, ot +ning +son *gullarning biri*. Bundan tashqari *rahmat, tashakkur, sa-*

*lom* kabi soʻzlar shaxs oti va olmoshlar bilan grammatik munosabatga kirishib, boshqaruvli birikmani hosil qiladi, deya fikr bildiriladi. Masalan: *sizga rahmat, barchangizga tashakkur* va h.k. [5;31] Shuni aytish kerakki, keltirilgan misollarda tushunchaga nisbatan fikr ifodalanish darajasi kuchliroq nazarimizda.

## Lingvistik modellashtirish masalasi

Lingvistik modellashtirish til birliklarining LSQlari hamda turli konstruksiyalariga asoslanadi. Shunday ekan model struktural tilshunoslik hamda kompyuter lingvistikasi kabi yoʻnalishlarni oʻzaro bogʻlash uchun xizmat qiladi.

Modellashtirish jarayonida til sathlarining bir-biriga bogʻliq ekanligini eʼtiborga olgan holda, ularning oʻzaro munosabati hamda har bir sathni oʻziga xos xususiyatlarini oʻrganish ahamiyatlidir. Bugungi kunga kelib, sintaktik tahlilni amalga oshiruvchi sintaktik analizatorning yaratilishi dolzarb masala sifatida eʼtirof etilmoqda. R.Tillayeva ilmiy tadqiqot ishida ham boshqa til birliklari bilan bir qatorda, sintaktik birliklarni modellashtirish masalasini ham atroflicha oʻrganib oʻz fikr-mulohazalarini bildiradi: Til sistemasining elementlari oʻziga yaqin turgan yuqori sath birliklari tarkibida funksiyalashadi. Shuning uchun maʼlum sath birligining mohiyati haqida maʼlumot olish uchun bu birlikning yuqori sathdagi funksiyasini oʻrganishga ham eʼtibor berish kerak boʻladi. Mana shu nuqtayi nazardan yondashganda, sintaktik sath til sistemasini yuqori sathi sifatida oʻz ichida barcha birliklarning funksiyasini jamlaydi. Gap va uning qismlarida til birliklarining barchasi jamlangan boʻladi [Tillayeva, 2016]. Til birliklari birlamchi til birliklari va ikkilamchi til birliklariga ajratilgan. Birlamchi til birliklariga fonema, morfema, leksemalarni, ikkilamchi til birliklariga esa leksemashakl, birikmashakl va gapshakllarni kiritadi. Ikkilamchi til birliklarining qurilma ekanligini taʼkidlagan holda, ularni supersegment birliklar deb hisoblaydi. Ikkilamchi til birliklariga leksemashaklni tuzish modeli, birikmashaklni tuzish modeli va gapshaklni tuzish modellari kirishi taʼkidlanadi. Uning fikricha, birikmashakl - supersegment birlik. Miyaning til xotirasi markazida birikmashakl modellarining ramzi mavjud. Ana shu modellar leksemashakl va leksema bilan toʻldirilib, nutqiy birlik holatiga oʻtadi. R.Tillayeva Sh.Rahmatullayevning sintaktik birliklarning bu ikki holati haqidagi fikrlariga qoʻshilish bilan birga, modellarni supersegment birlik degan fikriga oʻzining eʼtirozini bildiradi. Va shunday deydi: supersegment birliklar deb, odatda, seg-

ment birliklar ustiga qo'yiladigan ohang, pauza, urg'u kabi birliklar tushuniladi [Tillayeva, 2016]. Muallif bu o'rinda segment birliklarni mavhumlashtirish asosida hosil qilingan birliklar haqida - abstrakt birliklar haqida fikr yuritar ekan, ularni abstrakt birliklar sifatida talqin etish o'rinliroq bo'ladi, deb hisoblaymiz. Sintaktik birliklarni modellashtirish deganda, asosan, so'z birikmasi va gapni modellashtirish nazarda tutiladi. Biroq ushbu tadqiqot ishida modellashtirish faqatgina gap doirasida o'rganilgan. Shunday ekan, ushbu tadqiqot ishidan farqli ravishda sintaktik birliklarni modellashtirish masalasini so'z birikmasi doirasida ko'rib o'tmoqchimiz. Ba'zi tadqiqot ishlarida so'z birikmasi modeli esa uning umumiy xususiyatidan kelib chiqib, quyidagicha izohlanadi: T- H [tobe-hokim]. J.Ibragimov o'zining "O'zbek tili matnlarida so'zlarning bevosita (kontakt) va bilvosita (distant) birikuvchanligini dastlabki dasturiy modellashtirish" maqolasida o'zbek tilidagi matnlar tarkibidagi birikmalarni modellashtirishda umume'tirof etilgan ramzlardan foydalanganligini aytib, quyidagicha misollar keltiradi. *Ustodning muborak ko'zoynaklarini taqib, xalqimning o'tmishiga qaradim* gapida:

1) *o'tmishiga qaradim* - N + SFN + V + SFv Bunda N - OT, SFN - otning sintaktik shakllari, V - fe'l, SFv - fe'lning sintaktik shakllari. *O'tmishiga qaradim* kabi birikmalar tipik modelining submodelini quyidagicha berish mumkin bo'ladi: N + SG + PS<sub>aff</sub> + Acc<sub>CS<sub>aff</sub></sub> + V + TS<sub>aff</sub> + Shs<sub>aff</sub> Bunda Not, SG - birlik, PS<sub>aff</sub> - egalik shakli, Acc<sub>CS<sub>aff</sub></sub> - jo'nalish kelishigi, V - fe'l, TS<sub>aff</sub> (zamon shaklini inglizcha "Tense" so'zi asosida shu ramz orqali ifodalashga qaror qildik), ShS<sub>aff</sub> - shaxs - son shakli (shaxs - son shaklini shunday ifodalashga qaror qildik).

2) *taqib qaramoq* - V + V + SFv Bunda V - fe'l, V - fe'l, SFv - fe'lning sintaktik shakllari. 3) *xalqimning o'tmishiga qaradim* - N<sup>d</sup> + SFN + Nk + SFN + V + SFv 4) *ustodning ko'zoynaklari* - N + SFN + N + SFN Bunda N + OT, SFN - otning sintaktik shakllari, N + ot, SFN - otning sintaktik shakllari. 5) *muborak ko'zoynaklari* - ADJ + N + PS<sub>aff</sub> Bunda ADJ + sifat, N + ot, SFN - otning sintaktik shakllari [4]. So'z birikmasini modelleshtirishda ushbu yondashuvga qo'shilgan holda shuni aytish kerakki, o'zbek tilining xususiyatlarini o'zida aks ettiruvchi ramzlar orqali modellashtirish ham bu borada, o'ziga xos yondashuv bo'ladi, nazarimizda. Yuqorida o'ziga xos yondashuvlar asosida so'z birikmalarini modellashtirish masalasini o'rganib tahlil qilgan holda, barcha mustaqil so'z turkumlar hamda so'zlarni o'zaro bog'lashda ishtrok etuvchi grammatik birliklarning bosh harflaridan

foydalanib, o‘zbek tilidagi so‘z birikmalari LSQlarining modellari ishlab chiqildi.

1. OT-OT; 2. sifat-SF; 3. son-S; 4. olmosh-O; 5. ravish-R; 6. fe‘l-F; 7. sifatdosh –FSF; 8. ravishdosh-FR.

Kelishiklar: 1. Qaratqich-q.k; 2. tushum-t.k; 3. jo‘nalish-j.k; 4. o‘rin-payt-o‘.k; 6. Chiqish-ch.k; 7. ko‘makchi-k.

### **Otli birikmalar**

Bitishuvli birikmalar:

OT<sub>0</sub> +OT<sub>0</sub> [SF, S, O, R quyidagi so‘z turkumlarining barchasi ot bilan birikib, otli birikmalarni hosil qiladi. *Kumush qoshiq, chiroyli shahar, beshta o‘quvchi, barcha bolalar, so‘nggi kun.*

Moslashuvli birikmalar:

OT<sub>q,k</sub> + OT<sub>q,q</sub> [SF, S, O, R, FSF] *kitobning varog‘i, yaxshining so‘zi, bizning maktabimiz, ko‘pning ishi, o‘qiganning fikri.*

Boshqaruvli birikmalar:

OT<sub>j,k</sub> + OT [o.k, ch.k], k.k] [C, S, O, FSF] *hayotga mehr, bilimdonga olqish, birinchida yig‘ilish, unga maktub, kutilgan mehmon, baxt uchun kurash.*

### **Sifatli birikmalar**

Moslashuvli birikmalar:

OT<sub>q,k</sub> +C *bolaning begonasi*

Boshqaruvli birikmalar:

OT<sub>j,k</sub> + C [o‘.k, ch.k, k.k] [C, O, R, FSF, M.S] *mehrga muhtoj, sovuqqa chidamli, ko‘pga mehribon, kutganga aziz, borligida beqadir.*

### **Ravishli birikmalar**

Boshqaruvli birikma:

OT<sub>j,k</sub> + [o‘.k, ch.k, k.k] [O] *ishga loqayd, yil kabi uzun, hammadan keyin*

Moslashuvli birikma

OT<sub>q,k</sub> +R<sub>e,q</sub> *gapning ozi*

Sonli birikmalar

OT<sub>ch,k</sub> + S [O] *o‘quvchilardan biri, ulardan ikkisi*

Moslashuvli birikma:

OTlar+q.k+S<sub>e,q</sub> +o‘.k *kunlarning birida, sizlarning biringiz*

### **Fe‘lli birikmalar**

Bitishuvli birikma:

OT + FSF [S, R, O] *quyoshdek porlamoq, chiroyli yozmoq, tez yurmoq, bitta ko‘rmoq, o‘zi bilmoq.*

Yuqorida o‘zbek tilida so‘z birikmalari LSQning umumiy guruhla-

rini lingvistik modellashtirishga harakat qildik. Bu masala kompyuter lingvistikasida o'rganilishi va takliflar ishlab chiqilishi kerak bo'lgan muhim masalalardan hisoblanadi. Chunki matnlarning avtomatik sintaktik tahlili jarayonidagi muhim bosqichlardan biri hisoblanadi.

#### ADABIYOT

1. Abdullayev F., Ibrohimova F. O'zbek tilida boshqaruv. Monografiya. – T., 1962-yil.
2. Boysinov S., Mamirov O. Kompyuter lingvistikasida modellashtirish metodidan foydalanish. [www.jdpu.uz](http://www.jdpu.uz)
3. Bushuy T., Safarov Sh. Til qurilishi tahlil metodlari va metodologiyasi. – Toshkent, 2007-yil.
4. Ibragimov J. O'zbek tili matnlarida so'zlarning bevosita (kontakt) va bilvosita (distant) birikuvchanligini dastlabki dasturiy modellashtirish. // Zamonaviy dunyoda innovatsion tadqiqotlar: Nazariya va amaliyot nomli ilmiy, masofaviy, onlayn konferensiya materiallari. <https://doi.org/10.5281/zenodo.7515431>
5. O'zbek tili grammatikasi II tom. – T., 1978.
6. Rahimov A. Kompyuter lingvistikasi asoslari. – T., 2011.
7. Sayfullayeva R., Mengliyev B. va b. Hozirgi o'zbek adabiy tili. – Toshkent, 2009.
8. Tillayeva R. Lisoniy birliklarni modellashtirish. Magistrlik dissertatsiyasi ishi. – Namangan, 2016.
9. Ревзин И.И. Модели языка. – М., 1962. – с. 9.12.

УДК 811.35

**К ФОРМАЛЬНОЙ МОДЕЛИ ТЮРКСКОГО ИМЕННОГО  
СОГЛАСОВАНИЯ: ДАННЫЕ КУМЫКСКОГО ЯЗЫКА<sup>1</sup>****О. В. Федорова, С. Г. Татевосов***МГУ имени М.В. Ломоносова, Москва, Россия**olga.fedorova@msu.ru, tatevosov@gmail.com*

В статье обсуждается структура так называемой четвертой изафетной конструкции (И4), или *possessive free genitives* в терминологии Ozturk 2016, в терском диалекте кумыкского языка, тюркском языке, распространенном в нескольких регионах Северного Кавказа. И4 представляют собой именную группу с посессором в генитиве, в которой отсутствует согласование на вершине-обладаемом (например, 'наш-GEN сын-∅'). В кумыкском языке обнаруживается следующее ограничение: дистрибуция И4 ограничена именными группами с посессором первого и второго лица множественного числа. Для всех остальных видов генитивного посессора создается стандартная тюркская конструкция, известная как третья изафетная (И3), в которой согласование с посессором обязательно (например, 'мой-GEN сын-1SG'). В этой работе мы исследуем эмпирические последствия отсутствия согласования с посессором, опираясь на два ранее не отмеченных наблюдения. Во-первых, в то время как конструкция с согласующимся посессором не допускает синтаксического вложения в другую такую же конструкцию, это становится возможным, при отсутствии согласования (например, 'наш-GEN Марата-GEN сын-3SG'). Этот факт вместе с некоторыми дополнительными допущениями указывает на то, что И4 следует анализировать как конструкцию без согласования посессора и обладаемого, а не как вариант И3 с фонологически не реализованным согласовательным показателем. Во-вторых, И4 не допускают, чтобы посессор становился фокусом частного вопроса, а также исключают информационно-структурное передвижение посессора за пределы именной группы. Если допустить, что частные вопросы в тюркских языках образуются посредством скрытого передвижения, эти два факта можно свести к одному и тому же ограничению: посессор, не вступивший в отношение согласования, остается невидим для синтаксических элементов, вызывающих передвижение составляющих.

**Ключевые слова:** тюркские языки, кумыкский язык, синтаксис именной группы, конструкции с посессором, согласование.

---

<sup>1</sup> Данные получены в ходе полевого исследования терского диалекта, проводившегося в августе 2023 года в с. Предгорное Моздокского р-на РСО-Алания лингвистической экспедицией филологического факультета МГУ имени М.В. Ломоносова.

**TOWARDS A FORMAL MODEL OF TURKIC NOMINAL  
AGREEMENT: EVIDENCE FROM KUMYK**

*Olga Fedorova, Sergei Tatevosov*  
*Lomonosov Moscow State University*  
*Moscow, Russia*

olga.fedorova@msu.ru, tatevosov@gmail.com

This paper explores the structure of possessive free genitives, to use the terminology of Ozturk 2016, in the Terski dialect of Kumyk, a Turkic language spoken in the Caucasus, Russia. Possessive free genitives are possessive DPs where the possessor occurs in the genitive without triggering agreement on the head/possessee (e.g. ‘our-GEN son-Ø’). In Kumyk, the following restriction holds: possessive free genitives are restricted to 1PL and 2PL possessors; with any other type of possessor overt agreement is obligatory and a standard Turkic genitive-possessive construction obtains (e.g. ‘my-GEN son-1SG’). In this paper, we examine further empirical consequences of the lack of agreement with the possessor, focusing on two novel observations. First, while a genitive possessive construction cannot be embedded within another genitive possessive construction, it can be part of a possessive free genitive construction (e.g. ‘our-GEN son-3SG of John’s’). This fact, with certain additional assumptions, points towards analyzing free genitive possessives as genuinely lacking agreement between the possessor and the head rather than involving silent agreement with no phonological signature. Second, possessive free genitives disallow *wh*-questions and cannot be scrambled out of a DP. If *wh*-question formation in Turkic involves covert movement, these two facts can reduce to the same constraint: a possessor that has not established agreement with the head is syntactically inert for movement processes.

**Keywords:** Turkic languages, Kumyk, nominal syntax, possessive constructions, agreement.

Существенный компонент грамматики тюркских языков – именная согласовательная система, которая представлена в именных посессивных конструкциях (adnominal possessive construction, [Aikhenvald, 2019]). Ее основные элементы – посессор (possessor, обладатель) и обладаемое (possessum, possessee). Основные значения, выражаемые именной посессивной конструкцией, перечислены в (1) (см. [Aikhenvald, 2019; Коптѣвская-Тамм, 2002] среди многих других):

- (1) Значения посессивной конструкции
1. владение чем-то: дом Ивана;
  2. отношение «часть-целое»: ножка стула, голова коровы;
  3. родственные отношения: сестра Пети;
  4. ассоциативные отношения: парикмахер Кати;
  5. атрибут человека / объекта: женская работа;



6. время, материал, свойство: каменный дом, литр молока;

7. характеристики человека / объекта: рождение сына, высота горы.

Задача построения общей формальной модели посессивной конструкции для тюркских языков связана с проблемой значительного межъязыкового варьирования в этой области. **С точки зрения синтаксического оформления** именные посессивные конструкции можно с некоторой долей условности разделить на четыре типа так называемых **изафетов**, традиционно выделяемых в отечественной тюркологии<sup>1</sup>. Они различаются двумя параметрами: наличием генитивного показателя у посессора и посессивным показателем, реализуемым на обладаемом – вершине ИГ. Для **первого изафета** (И1, Ascriptive construction по [Öztürk, 2016]) характерен безгенитивный посессор и отсутствие посессивного показателя, для **второго изафета** (И2, Possessive compounds по [Öztürk, 2016]) – безгенитивный посессор и посессивный показатель на вершине, для **третьего изафета** (И3, Genitive-possessive constructions по [Öztürk, 2016], Canonical possessive construction по [Csató, 2019]) – наличие обоих показателей, для **четвертого изафета** (И4, Possessive free genitives по [Öztürk, 2016], Non-canonical possessive construction по [Csató, 2019]) – наличие генитивного показателя при отсутствии посессивного показателя. В таблице 1 приведены примеры каждой конструкции из терского диалекта кумыкского языка:

---

<sup>1</sup> Термином изафет, впрочем, принято обозначать только такие конструкции, оба члена которой являются именами существительными; кроме того, чаще выделяют три типа изафета, а не четыре. Ср., в частности, фрагмент статьи в БРЭ: «В тюркологии термин «И.» обозначает именные определит. сочетания, оба члена которых выражены существительными. Выделяется 3 типа И.: для первого характерно отсутствие морфологич. показателей связи компонентов (напр., азерб. *dəmir qapı* ‘железные ворота’, букв. – ‘железо ворота’), для второго – наличие при определяемом аффикса принадлежности 3-го лица (напр., тур. *türk dil-i* ‘турецкий язык’), для третьего – аффикс принадлежности 3-го лица при определяемом и аффикс род. п. при определении (напр., туркм. *at-uyň ölüm-i* ‘смерть коня’)... Распределение функций между типами И. в разных тюрк. языках не совпадает; в ряде случаев в одном и том же языке разл. его типы могут употребляться как синтаксич. синонимы.»

**Таблица 1. Типы посессивных конструкций в терском диалекте кумыкского языка**

| показатель |            | назва-<br>ние | примеры: ‘наш сын’; ‘моря берег’      |
|------------|------------|---------------|---------------------------------------|
| посессор   | обладаемое |               |                                       |
| ∅          | ∅          | И1            | *биз улан; денгиз ягъа                |
| ∅          | Poss       | И2            | *биз улан-ыбыз; ?денгиз ягъа-сы       |
| Gen        | Poss       | И3            | (биз-ин) улан-ыбыз; денгиз-ни ягъа-сы |
| Gen        | ∅          | И4            | биз-ин улан; *денгиз-ни ягъа          |

Цель этой статьи – обсудить несколько ранее не отмеченных в литературе фактов на материале **терского диалекта кумыкского языка**, понимание которых существенно для построения формальной модели, описывающей дистрибуцию каждой из конструкций. Основное внимание в последующем изложении будет уделено И4, в которой наблюдается нетривиальное лично-числовое ограничение.

Согласно [Абдуллаева и др. 2014] в литературном кумыкском языке могут употребляться конструкции И1, И2 и И3, а также несколько топонимов с И4. В некоторых, относительно редких, случаях одно и то же значение можно выразить всеми тремя конструкциями, различия между которыми не очевидны, (2):

- (2) а. мен [денгиз ягъа]-да яша-й-ман И1  
я море берег-LOC жить-IPFV-1SG  
*Я живу на берегу моря. {a=b=c}*
- б. ?мен [денгиз ягъа-сы]-нна яша-й-ман И2  
я море берег-3-LOC жить-IPFV-1SG
- с. мен [денгиз-ни ягъа-сы]-нна яша-й-ман И3  
я море-GEN берег-3-LOC жить-IPFV-1SG

Наиболее употребительными в современном терском диалекте являются конструкции И1 и И3. Конструкция И2 широко используется только в сочетании с посессором-личным местоимением, а примеры типа (2b) признаются грамматичными не всеми носителями. Конструкции с И4 зафиксированы в терском диалекте исключительно в сочетании с посессором-личным местоимением, см. ниже, а примеры из кумыкского литературного языка, представленные в [Абдуллаева и др., 2014, с. 174] признаются неграмматичными всеми опрошенными носителями, (3).

- (3) а. Март-**ны** сын-**ы** И3  
 Март-**GEN** памятник-3  
*памятник Марту {a=b}*
- б. \*Март-**ны** сын И4  
 Март-**GEN** памятник

В отличие от конструкций, в которых посессор представляет собой именную группу третьего лица единственного числа, конструкции с посессором-личным местоимением имеют следующую особенность. И4 в них допускается, однако, только для двух комбинаций признаков лица и числа у посессора: 1PL и 2PL<sup>1</sup>, см. таблицу 2.

**Таблица 2. Личная посессивность в сочетании с существительным ‘сын’**

|     | И4          | И3                 | Русский перевод |
|-----|-------------|--------------------|-----------------|
| 1SG | *мени улан  | (мени) улан-ым     | ‘мой сын’       |
| 2SG | *сени улан  | (сени) улан-ынг    | ‘твой сын’      |
| 3SG | *ону улан   | (ону) улан-ы       | ‘его сын’       |
| 1PL | бизин улан  | (бизин) улан-ыбыз  | ‘наш сын’       |
| 2PL | сизин улан  | (сизин) улан-ыгъыз | ‘ваш сын’       |
| 3PL | *оланы улан | (оланы) улан-ы     | ‘их сын’        |

|     | И4               | И3                      | Русский перевод |
|-----|------------------|-------------------------|-----------------|
| 1SG | *мени улан-ылар  | (мени) улан-ылар-ым     | ‘мои сыновья’   |
| 2SG | *сени улан-ылар  | (сени) улан-ылар-ынг    | ‘твои сыновья’  |
| 3SG | *ону улан-ылар   | (ону) улан-ылар-ы       | ‘его сыновья’   |
| 1PL | бизин улан-ылар  | (бизин) улан-ылар-ыбыз  | ‘наши сыновья’  |
| 2PL | сизин улан-ылар  | (сизин) улан-ылар-ыгъыз | ‘ваши сыновья’  |
| 3PL | *оланы улан-ылар | (оланы) улан-ылар-ы     | ‘их сыновья’    |

<sup>1</sup> Отметим, что в мишарском диалекте татарского языка конструкция «местоимение-посессор в генитиве + существительное без показателя посессивности» запрещена только в третьем лице обоих чисел (Татевосов и др. ред. 2017).

Полная модель посессивной конструкции должна, несомненно, содержать объяснение того, почему именно эти две комбинации допускаются, если посессору не соответствует лично-числовая морфология на обладаемом. Не предлагая такого объяснения в пределах этого исследования, мы, однако, покажем, в И4 местоименный посессор 1/2PL проявляет иное синтаксическое поведение, чем посессоры с другой комбинацией лично-числовых характеристик.

Первое свойство местоименного посессора 1/2PL состоит в отсутствии эффектов, вызванных конфликтом различных контролеров согласования. Конфликт иллюстрируется примерами типа (4), в которых посессивная ИГ имеет собственный посессор. В (4) это ‘мой’ – посессор при ИГ ‘сын Расула’.

(4) \*(мени) [Расул-ны улан]-ым гел-ди тж. сени  
 я.GEN Расул-GEN сын-1SG приходить-PST  
*Мой сын Расула пришел.*

(5) \*мени [Расул-ны улан-ы] гел-ди тж. сени  
 я.GEN Расул-GEN сын-3SG приходить-PST  
*Мой сын Расула пришел.*

Как видно из (4)–(5), такая конструкция неграмматична при любой лично-числовой морфологии вершины – 1SG, как в (4), и 3SG, как в (5). Можно предположить, что неграмматичность обусловлена тем, что и в (4) и в (5) отсутствует согласование с одним из двух посессоров.

Посессор 1/2PL, однако, показывает другой паттерн:

(6) \*(бизин) [Расул-ны улан]-ыбыз гел-ди тж. сизин  
 мы.GEN Расул-GEN сын-1PL приходить-PST  
*Наш сын Расула пришел.*

(7) бизин [Расул-ны улан-ы] гел-ди<sup>1</sup> тж. сизин  
 мы.GEN Расул-GEN сын-3SG приходить-PST И4  
*Наш сын Расула пришел.*

(6) неграмматичен, как и его аналог в (4). Вероятно, это обусловлено той же причиной: отсутствием согласования с посессором 3SG ‘Расул’. Однако (7) оказывается приемлемым, в отличие от (5). Это показывает, что посессор 1/2PL по меньшей

<sup>1</sup> Похожее предложение [(бизин) Расул-ыбыз]-ны улан-ы гел-ди, в котором личное местоимение выступает посессором при посессоре ‘Расул’, имеет ожидаемую интерпретацию ‘Сын нашего Расула пришел’.

мере не нуждается в согласовании с вершиной. Если бы это было не так, а конструкция в (7) представляла бы собой структуру с согласованием, не имеющим фонологической реализации, предложение в (7) было бы так же неграмматично, как и (6).

Второе свойство посессора 1/2PL в составе И4 состоит в том, что возможности фокусирования этой ИГ существенно ограничены. (8) показывает, что посессор в составе И3 грамматичен в качестве фокуса общего вопроса, но такая же возможность недоступна для посессора в И4:

- (8) а. (**бизин**) ким-ибиз гел-ди?  
 мы.GEN кто-1PL приходит-РST  
*Наш кто пришел? {a=b}*  
 б. \***бизин** ким гел-ди?  
 мы.GEN кто приходит-РST

Неграмматичность (8) может иметь более одного объяснения. Для теорий, которые опираются на идею, что образование общего вопроса предполагает скрытое передвижение в неаргументную позицию даже в так называемых языках *wh in situ*, было бы привлекательно связать ограничение в (8) с невозможностью передвижения посессора в отсутствие согласования с вершиной.

Косвенно о том, что такую возможность следует рассматривать серьезно, свидетельствует следующий факт, связанный с дистрибуцией посессора 1/2PL: скрэмблинг посессора допускается только для согласующихся с вершиной посессоров. Это иллюстрируется в (9):

- (9) а. (**бизин**) тюне (**бизин**) улан-ыбыз (**бизин**)  
 мы.GEN вчера мы.GEN сын-1P мы.GEN  
 гел-ди  
 приходит-РST  
*Вчера наш сын пришел. {a=b=c=d}*  
 б. тюне **бизин** улан гел-ди  
 вчера мы.GEN сын приходит-РST И4  
 в. \***бизин** тюне улан гел-ди  
 мы.GEN вчера сын приходит-РST И4  
 д. \*улан тюне гел-ди **бизин**  
 сын вчера приходит-РST мы.GEN И4

Хотя ограничения на скрэмблинг могут быть не полностью идентичны ограничениям на вопросительное передвижение, невозможность появления посессора вне именной группы, которая наблюдается в (9c-d), но не в (9b), возможно, указывают на параллелизм с примерами типа (8a-b).

Завершая этот очерк, суммируем наши наблюдения. В терском диалекте кумыкского языка личные местоимения 1/2PL допускают, но не требуют согласование с обладаемым. При отсутствии согласования становится возможным вложение в посессивную конструкцию другой посессивной конструкции вида ИЗ. Однако посессор, не вступивший в отношение согласования, не видим для синтаксических элементов и операций, вызывающих передвижение составляющих.

#### ЛИТЕРАТУРА

1. Абдуллаева А.З., Гаджихамедов Н.Э., Кадыраджиев К.С., Керимов И.А., Ольмесов Н.Х., Хангишиев Д.М. 2014. Современный кумыкский язык. Махачкала: ИЯЛИ ДНЦ РАН. 557 с.
2. Татевосов С.Г., Пазельская А.Г., Сулейманов Д.Ш. (ред.). 2017. Элементы татарского языка в типологическом освещении. М.: Буки Веди. 761 с.
3. Aikhenvald A.Y. 2019. Expressing 'possession': motivations, meanings, and forms // L. Johanson, L.F. Mazzitelli, I. Nevskaya (eds.) *Possession in languages of Europe and North and Central Asia*, pp. 7–25.
4. Csató E.A. 2019. On Turkish non-canonical possessives // L. Johanson, L.F. Mazzitelli, I. Nevskaya (eds.) *Possession in languages of Europe and North and Central Asia*, pp. 85–102.
5. Коптјевскаја-Тамм М. 2002. Adnominal possession in the European languages: form and function // *STUF*, 55(2), pp. 141–172.
6. Öztürk B., Taylan E.E. 2016. Possessive constructions in Turkish // *Lingua*, 182, pp. 88–108.

УДК 811.35

О СКАЛЯРНОМ ХАРАКТЕРЕ ЧИСЛОВОЙ НЕЙТРАЛЬНОСТИ<sup>1</sup>*С. Г. Татевосов**МГУ имени М. В. Ломоносова, Москва, Россия*

tatevosov@gmail.com

В статье обсуждаются структура и интерпретация существительных в тюркских языках до их соединения с граммемой числа. Основной исследовательский вопрос состоит в том, обозначают ли существительные единичные индивиды, множественные индивиды или и то и другое. Последний случай принято называть числовой нейтральностью. В работе намечается два аргумента в пользу гипотезы, согласно которой множество существительных распадается на два класса – те, которые обозначают единичные сущности, и те, которые показывают числовую нейтральность. Оба аргумента опираются на схожую линию рассуждений. Чтобы определить характеристики лексических единиц до соединения с граммемами тех или иных категорий, следует исследовать их поведение в таких конфигурациях, где отсутствует по меньшей мере часть функциональной структуры. В работе предлагается две такие конфигурации: не маркированное аккузативом прямое дополнение и элемент, инкорпорированный в отыменной глагол. В карачаево-балкарском языке немаркированные прямые дополнения демонстрируют два паттерна с точки зрения дискурсивной анафоры. Некоторые могут выступать антецедентом анафорических местоимений как в единственном, так и во множественном числах и тем самым характеризуются числовой нейтральностью. Другие допускают только местоимения в единственном числе, то есть являются семантически единичными. Распределение существительных между двумя классами определяется их позицией на шкале индивидуруемости. Большинство артефактов, крупные животные и люди проявляют семантическую единичность, прочие существительные нейтральны по числу. Второй аргумент опирается на предельность неэргативных отыменных глаголов в татарском языке, которые обозначают инкрементальное отношение (обычно созидание или эмиссию) между событиями и индивидами из экстенционала производящего существительного. Некоторые существительные создают предельные глаголы и тем самым являются квантованными. Другие приводят к неопределенности, то есть оказываются неквантованными. Поскольку квантованность характеризует только именные предикаты, обозначающие единичные сущности, предельность производного глагола дает возможность определить числовые характеристики инкорпорированного существительного. Как и в балкарском языке, распределение существительных между двумя классами определяется их положением на шкале индивидуруемости.

**Ключевые слова:** тюркские языки, карачаево-балкарский язык, татарский язык, именное число, числовая нейтральность, вариативное маркирование прямого дополнения, инкорпорация существительного

<sup>1</sup> Исследование поддержано грантом РФФИ № 22-18-00285, реализуемом в МГУ имени М.В. Ломоносова.

## ON THE SCALAR CHARACTER OF NUMBER NEUTRALITY

*Sergei Tatevosov**Lomonosov Moscow State University, Moscow*  
tatevosov@gmail.com

This paper aims at exploring the structure and interpretation of uninflected numberless nominals in Turkic languages. Specifically, it addresses the question of whether such nominals denote singular individuals, plural individuals or both, i.e. are number neutral. It outlines two arguments in favor of the hypothesis that the entire set of nominals falls into two classes: lexical items that denote singularities and number neutral ones. Both arguments rely on the same strategy. In order to identify characteristics of uninflected items, one needs to look at their distribution and interpretation in structurally deficient configurations, where at least some of the functional structure is not projected and their true characteristics are more transparently visible. I propose two such configurations: direct objects with no accusative marking and nominals incorporated into denominal verbs. In Karachay-Balkar, unmarked direct objects show two patterns with respect to discourse anaphora. Some can antecede both singular and plural pronouns (i.e. have a number neutral denotation), others only license singular ones (i.e. denote singularities). Class membership of individual lexical items is predictable from their projection on the scale of individuation. Most artifacts, big animals and humans are semantically singular, while the rest of nouns (substances, granular aggregates, collective aggregates, ...) are number-neutral. The other argument comes from telicity of unergative denominal verbs in Tatar which denote an incremental relation (typically, creation or emission) between eventualities and individuals from the extension of the incorporated noun. Some of the nouns produce telic verbs suggesting that they have a quantized denotation. Others give rise to atelicity, thus indicating the lack of quantization. Since nominal predicates of singular individuals are quantized while number-neutral predicates are cumulative, telicity of corresponding verbs is indicative of these properties of their denotations. Crucially, just as with plural anaphora in Balkar, nouns in Tatar are distributed between the two classes based on their projection of the scale of individuation.

**Keywords:** Turkic languages, number neutrality, scale of individuation, differential argument marking, denominal verbs.

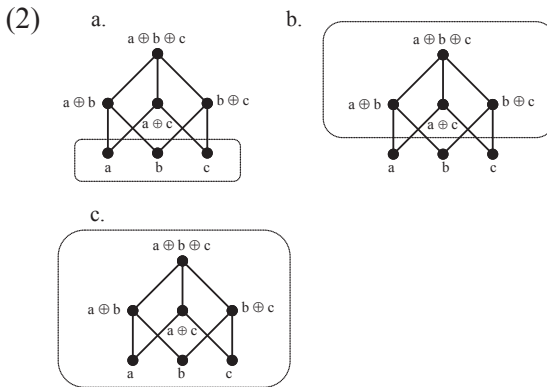
Формальные модели грамматики естественных языков существенным образом опираются на допущения о том, какого рода грамматическая информация и в каком объеме хранится в словаре. Важный вопрос в этой связи – как выглядят словарные единицы до того, как они соединяются с граммами тех или иных категорий, например, что представляет собой глагол никакого вида и времени или существительное никакого числа.

В этой статье обсуждаются проблема словарной числовой нейтральности существительных в тюркских языках, которую в первом приближении можно сформулировать в виде вопроса (1):



(1) Содержит ли экстенционал существительного как словарной единицы единичные сущности, их мереологические суммы или и то и другое?

Если придерживаться точки зрения, согласно которой существительное обозначает алгебраическую структуру с отношением суммы (Link 1983 и последующие работы), есть ровно три возможности представить себе денотат существительного никакого числа – тот денотат, который оно имеет до соединения с числовым показателем. Эти возможности показаны в (2a-c):



В (2a) существительное обозначает единичные сущности, в (2b) – множественные, в (2c) – и те и другие, то есть весь домен индивидов, подпадающих под соответствующее описание. Последний случай – случай числовой нейтральности, когда существительное словарно не содержит информации о числе.

Цель этой работы – предложить два аргумента в пользу тезиса о скалярном характере числовой нейтральности в тюркских языках:

(3) Тезис о скалярном характере числовой нейтральности

Тюркские исчисляемые существительные распадаются на два класса. Часть из них словарно имеет структуру в (2a), т.е. обозначает единичные сущности. Другая часть показывает числовую нейтральность, как в (2c). Принадлежность к одному из двух классов определяется положением существительного на шкале индивидуированности.

Таким образом, далее мы попытаемся обосновать, что формальная модель семантики тюркских исчисляемых существитель-

ных должна распределять их на два класса – единичные и числово нейтральные.

Общая стратегия выявления характеристик языковых объектов, которые они имеют до соединения с разнообразными граммеммами, состоит в следующем. Чтобы увидеть такие характеристики, необходимы конфигурации, в которых представлено возможно меньшее количество функциональной структуры. Для глаголов это, в частности, не полностью развернутые клаузы. Существительные мы будем рассматривать в составе структур двух типов: именной группы, занимающей позицию не маркированного винительным падежом прямого дополнения, и именной вершины, инкорпорированной в отыменную глагол.

Прямое дополнение, не маркированное винительным падежом, представляет собой не полностью развернутую именную группу [Dayal 2011, van Urk 2019, Levin 2019, Driemel 2020], структурный размер которой меньше, чем DP, и которая имеет неаргументный семантический тип.

Если в структуру такой именной группы не входит элемент, обозначающий граммему числа, то это ровно та конфигурация, в которой возможно эмпирически различить случаи из (2a-c). В Головина 2021 представлены релевантные факты из малкарского диалекта карачаево-балкарского языка, в котором числовая нейтральность проверялась при помощи диагностики, иллюстрируемой в (4):

- (4) Alim alma zĭj-a e-di. Ala tatli-la  
 Алим яблоко собирать-IPFV AUX-PST.3 он.PL  
 вкусный-PL  
 e-di-le.  
 быть-PST.3-PL  
 ‘Алим собирал яблоки. Они были вкусные.’

В (4) немаркированное прямое дополнение выступает антецедентом анафорического местоимения во множественном числе. Грамматичность такой анафорической отсылки показывает, что денотат существительного *alma* в (4) содержит в себе множественные сущности. (На это же, кроме того, указывает допустимость *alma* ‘яблоки’ в качестве прямого дополнения глагола ‘собирать’, нуждающегося во множественном актанте.) Допустимость анафорического местоимения в единственном числе в (5), кроме того, показывает, что единичные сущности также выступа-

ют элементом экстенционала этого существительного в позиции немаркированного прямого дополнения:

- (5) Alim har kün alma aša-j e-di. OI  
 Алим каждый день яблоко есть-IPFV AUX-PST.3 он  
 tatli e-di.  
 вкусный быть-PST.3  
 ‘Алим каждый день съедал яблоко. Оно (каждый раз) было вкусное.’

В совокупности примеры (4) и (5) показывают, что денотат существительного *alma* ‘яблоко’ содержит и единичные сущности и их суммы, то есть устроен в соответствии с (2с).

Не все существительные, однако, демонстрируют такой же паттерн. В (6) показано существительное *zašciq* ‘мальчик’:

- (6) Alim zašciq kör-e e-di.  
 Алим мальчик видеть-IPFV AUX-PST.3SG  
 ‘Алим видел мальчика.’  
 a. OI midax e-di.  
 он грустный быть-PST.3SG  
 ‘Он был грустный.’  
 b. \*Ala midax-la e-di-le.  
 он.PL грустный-PL быть-PST.3-PL  
 ‘Они были грустные.’

В отличие от (4), в (5) прямое дополнение не может быть антецедентом дискурсивного местоимения во множественном числе. Это показывает, что денотат существительного ‘мальчик’ содержит только единичные сущности.

Возникает естественный вопрос, как существительные распределяются между двумя паттернами. А.А. Головнина [Головнина 2021] приводит данные, согласно которым это определяется положением существительного на шкале индивидуируемости в (7) ([Grimm, 2018] и другие работы) или ее более дробном варианте (8) с подразделением индивидов на подкатегории:

- (7) substance < granular aggregates < collective aggregates < individuals  
 (8) substance < granular aggregates < collective aggregates < natural objects < small animals < pair/grouped body parts < artifacts < middle-sized animals < humans

В карачаево-балкарском языке граница проходит, по видимому, по категории артефактов. Названия веществ, субстанций, дробных совокупностей ('песок', 'дробь', 'рис'), составных совокупностей ('вишня', 'инструмент'), некрупных животных ('волк', 'баран'), парных частей тела ('руки') ведут себя так же, как *alma* 'яблоко'. Крупные животные ('корова') и большинство артефактов ('книга', 'стол') и т.п. похожи на *zašsciġ* 'мальчик'.

Другое явление, которое дополнительно подтверждает это обобщение, – интерпретация отыменных глаголов. В татарском языке самый продуктивный способ их образования – показатель *-la-*. Нас интересует конкретная разновидность таких глаголов, когда посредством инкорпорации существительного создается неэргатив. Одна иллюстрация показана в (9)<sup>1</sup>. Структура глагольной группы этого предложения выписана в (10); она отражает анализ, впервые предложенный в [Hale, 2002].

- (9) sɤjɤt bɤzau-la-dɤ.  
 корова теленок-VRB-PST  
 'Корова отелилась.'

- (10) [<sub>VP</sub> sɤjɤt [<sub>V</sub> [<sub>N</sub> bɤzau]+[<sub>V</sub> la]] [<sub>NP</sub> [<sub>N</sub> bɤzau ] ] ]

Согласно (10), глагольная вершина представляет собой вербализирующий аффикс *-la-*, а исходная позиция именного компонента глагола – комплемент этой вершины; два элемента соединяются посредством передвижения вершины. У неэргивных глаголов рассматриваемого класса, согласно [Татевосов, 2017], глагольная вершина интерпретируется как событие созидания или эмиссии сущности, обозначенной комплементом.

Предикаты, описывающие процессы созидания или эмиссии ('строить', 'писать' и т. п.) предполагают инкрементальное отношение между процессом и его участником. Это означает, что свойства событийного предиката, который возникает после заполнения аргументной позиции в (10), зависят от кумулятивности / квантованности именного предиката. Перед нами, таким образом, тот редкий случай акциональной композиции, когда в ней участвует не синтаксически реализованный аргумент, а аргумент, подвергающийся семантической инкорпорации.

С квантованным именным предикатом мы ожидаем возник-

<sup>1</sup> Пример из Татевосов (ред.) 2017, представляющий мишарский диалект татарского языка.

новения предельной интерпретации. По всей видимости, именно это происходит в случаях типа (9). Если верно, что предикат *bezau* ‘теленок’ квантован, предельность (9) вытекает из общих принципов теории акциональной композиции и не нуждается ни в каком дополнительном объяснении.

Если инкорпорируемый предикат неквантован, у него ожидается возникает неопределенная интерпретация. Так, например, происходит при инкорпорации неисчисляемого, а следовательно и неквантованного существительного *maNka* ‘сопли’ в (11)<sup>1</sup>:

- (11) *marat maŋka-la-dx.*  
 Марат сопли-VRB-PST  
 ‘Марат сопливел.’

В этом месте становится существенным следующий факт: предикаты, экстенционал которых имеет структуру (2a), квантованы, а предикаты с экстенционалом вида (2b)-(2c) нет. Следовательно, инкорпорация существительного в глагол, (10), – это еще одна конфигурация, благодаря которой можно увидеть свойства именного предиката. Если в момент инкорпорации существительное имеет структуру в (2a), мы получаем предельный глагол, а если структуру в (2b) или (2c) – неопределенный. Предельность отыменного глагола, таким образом, становится индикатором, указывающим на семантическое содержание существительного.

Данные, обсуждаемые в [Татевосов, 2017], а также примеры из [Закиев, 1993: 412–419] указывают, по всей видимости, на ту же закономерность, что и в карачаево-балкарском языке. Часть существительных имеет экстенционал типа (2a), в который входят только единичные сущности, и создает предельные предикаты. Так происходит, например, с существительным *brzau* ‘теленок’ в (9) и аналогичными. Другая часть существительных оказывается числово нейтральной, и результирующий глагол становится неопределенным. Несколько экземпляров этого последнего класса, упомянутых в Закиев (ред.) 1993, показано в (12):

- (12) *болытла* ‘заволакивать тучами, болыт ‘туча, облако’  
 хмуриться’  
*тавышла* ‘голосить’ тавыш ‘голос, звук’

<sup>1</sup> Такие глаголы могут иметь инцептивную интерпретацию (‘засопливеть’), нерелевантную для текущего обсуждения; подробнее см. Татевосов 2017.

|                                       |                           |
|---------------------------------------|---------------------------|
| <i>ымла</i> ‘подавать знаки’          | <i>ым</i> ‘жест, мимика’  |
| <i>фикерле</i> ‘мыслить’              | <i>фикер</i> ‘мысль’      |
| <i>тулгакла</i> ‘тужиться’            | <i>тулгак</i> ‘потуги,    |
| родовые                               | схватки’                  |
| <i>кортла</i> ‘червиветь’             | <i>корт</i> ‘червь’       |
| <i>үгетлә</i> ‘наставлять, увещевать, | <i>үгет</i> ‘наставление’ |
| уговаривать’                          |                           |
| <i>такмакла</i> ‘петь частушки’       | <i>такмак</i> ‘частушка’  |

Все существительные, от которых образуются глаголы в (12), – исчисляемые; каждое из них способно присоединять показатель множественного числа. Однако в составе глаголов ни одно из них не обозначает единичный объект ‘облако’, ‘червь’, ‘частушка’ и т.д. Напротив, все они имеют множественную, то есть неквантованную, интерпретацию, а образованные от них глаголы оказываются, в полном соответствии с предсказаниями теории акциональной композиции, неопредельными. Самое существенное наблюдение, связанное с (12), состоит в том, что все перечисленные существительные занимают относительно невысокое положение на шкале индивидуируемости. В карачаево-балкарском языке этой зоне на шкале также соответствует числовая нейтральность.

Таким образом, данные об интерпретации существительных в составе немаркированных прямых дополнений (карачаево-балкарский язык) и отыменных глаголов (татарский язык) дают основание сделать предварительный вывод о том, что экстенционал именных предикатов варьирует между (2а) и (2с) в зависимости от положения существительного на шкале индивидуируемости.

## ЛИТЕРАТУРА

1. Головнина А.А. 2021. Число и числовая нейтральность: данные карачаево-балкарского языка. Технический отчет №21-02-6. МГУ имени М.В. Ломоносова.
2. Татевосов С.Г. 2017. Событийная структура некоторых отыменных глаголов // Татевосов С.Г., Пазельская А.Г., Сулейманов Д.Ш. (ред.) 2017. Татарский язык в типологическом освещении. Мишарский диалект. М. С. 249–280.
3. Татевосов С.Г., Пазельская А.Г., Сулейманов Д.Ш. (ред.) 2017. Татарский язык в типологическом освещении. Мишарский диалект. М.
4. Dayal, Veneeta. 2011. Hindi pseudo-incorporation // *Natural Language and Linguistic Theory* 29: 123–167.

- 
5. Driemel, Imke. 2020. Pseudo-incorporation and its movement patterns // *Glossa: a journal of general linguistics* 5(1): 106.
  6. Grimm, Scott. 2018. Grammatical number and the scale of individuation // *Language* 94(3): 527–574.
  7. Hale, Ken L., Keyser, Samuel J. 2002. *Prolegomenon to a theory of argument structure*. Cambridge.
  8. Levin, Theodore. 2019. On the nature of differential object marking: Insights from Palauan // *Natural Language and Linguistic Theory* 37: 167–213.
  9. Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach // Bäuerle R., Schwarze Ch and von Stechow A. (eds.) *Meaning, use, and interpretation of language*. Berlin, New York: De Gruyter. P. 302–323.
  10. van Urk, Coppe. 2019. Object Licensing in Fijian and the role of adjacency // *Natural Language and Linguistic Theory* 38: 313–364.

УДК 81'44

**СЛОВОИЗМЕНТЕЛЬНЫЕ ПАРАДИГМЫ РУССКИХ  
ЗАИМСТВОВАНИЙ В ТУВИНСКОМ ЯЗЫКЕ,  
ОКАНЧИВАЮЩИЕСЯ НА СТЕЧЕНИИ СОГЛАСНЫХ -PT<sup>1</sup>****Э. К. Аннай., Б. Ч. Ооржак, Ч. Г. Ондар,  
Н. М. Монгуш***Тувинский институт гуманитарных и прикладных  
социально-экономических исследований  
при Правительстве Республики Тыва,  
Кызыл, Тыва, Российская Федерация*eannaj@mail.ru, oorzhak.baylak@mail.ru, choygandi@mail.ru,  
mongusnachyn@mail.ru

В данной статье анализируется парадигма словоизменения русских заимствований в тувинском языке, оканчивающиеся на стечении согласных -рт. Выявлено, что заимствованные из русского языка лексемы, оканчивающиеся на стечения согласных -рт, и тувинские лексемы, оканчивающиеся на те же стечения согласных, имеют аналогичные словоизменительные парадигмы. На основе обнаруженных особенностей парадигмы заимствованных существительных будет составлена база данных заимствованных слов из русского языка, их словоизменительные и словообразовательные парадигмы, составят базу формализованной грамматики, разрабатываемого в настоящее время Национального корпуса тувинского языка.

**Ключевые слова:** тувинский язык, русский язык, заимствование, словоизменительная парадигма, согласные звуко сочетания

**INFLECTIONAL PARADIGMS OF RUSSIAN BORROWINGS  
IN THE TUVAN LANGUAGE ENDING IN A CONCATENATION  
OF CONSONANTS -RT*****Ellada Annai, Bailak Oorzhak, Choigan Ondar,  
Nachyn Mongush****Tuvan Institute of Humanities and Applied Social and Economic  
Research under the Government of the Republic of Tuva  
Kyzyl, Tuva, Russian Federation*eannaj@mail.ru, oorzhak.baylak@mail.ru, choygandi@mail.ru,  
mongusnachyn@mail.ru

---

<sup>1</sup> Исследование выполнено за счет гранта Российского научного фонда № 23-28-10294 «Заемствования из русского языка в зеркале развития тувинского» (при паритетной финансовой поддержке Республики Тыва), ТИГПИ.



This article analyzes the paradigm of inflection of Russian loanwords in the Tuvan language, ending in the confluence of consonants -rt. It was revealed that lexemes borrowed from the Russian language ending in consonantal combinations -rt, and Tuvan lexemes ending in the same consonantal combinations have similar inflectional paradigms. Based on the discovered features of the borrowed noun paradigm, a database of borrowed words from the Russian language will be compiled, their inflectional and word-formation paradigms will form the basis of the formalized grammar of the National Corpus of the Tuvan language currently being developed.

**Keywords:** Tuvan language, Russian language, borrowing, inflectional paradigm, consonantal sound combinations

## Введение

Влияние русского языка на тувинский язык стало наиболее активным с начала XX в. и проявляется во всех формах его функционирования. Это обусловило изменения в области лексики и грамматики тувинского языка. Последние фундаментальные исследование влияния русского языка на тувинский относятся к 70-м годам XX века [Татаринцев, 1974]. Отдельные сведения содержатся в исследованиях [Чадамба, 1974; Гансух, 2009; Бавуу-Сюрюн, 2000, 2013, 2015; Аннай, 2021].

В словарный фонд тувинского языка в огромном количестве проникает заимствованная лексика из русского и посредством русского из других языков. Попадая в тувинский язык, они адаптируются в разной степени. Поэтому возникают с одной стороны, объективные трудности в их отражении на письме в соответствии с морфо-фонетическими правилами тувинского языка с одной стороны, с другой – уровень владения русским языком современных носителей тувинского языка отличается от того, как это было в 60-е и 70-е гг. XX в., когда разрабатывались правила орфографии и орфоэпии. Тогда же были разработаны правила правописания русских заимствований в тувинском языке.

В последние годы проводятся исследования по направлению «Корпуса языков народов России», в рамках которого идёт работа по грамматической разметке в корпусах тюркских языков южной Сибири: хакасского [Дыбо, Шеймович 2014], тувинского [Oorzhak, Khertek, 2015; Салчак, Байыр-оол, 2015; Washington, Bayir-ool, Salchak, Tyers, 2016]. Работа по лингвистическому аннотированию тувинских текстов ведётся в разных программах и

по разным методикам. В Тувинском институте гуманитарных и прикладных социально-экономических исследований при Правительстве Республики Тыва ведется работа по разработке программы проверки правописания тувинского языка на основе Hunspell, в рамках которой создаются словарная и аффиксальная базы тувинского языка.

Наиболее сложной для формального описания является словоизменительная парадигма склонения заимствованных слов. И целью данной статьи является описание особенностей освоения заимствованной из русского языка существительных с точки зрения их морфологической адаптации, отражающейся в грамматической парадигме.

Материалом для исследования послужила составленная авторами база данных заимствованных слов из русского языка в тувинский, общим объемом 6 324 ед.

### **Особенности словоизменения заимствований в тувинском языке, оканчивающихся на стечения согласных *-рт***

Особенности в словоизменительных парадигмах имен существительных проявляются у русских заимствований, оканчивающихся на следующие стечения *-рт*. Это такие лексемы как: *стандарт*, *концерт*, *курорт* и др. Всего подобных заимствованных лексем в электронной базе около 30 ед.

Электронная программа Notepad++ позволяет составить парадигмы всех слов. В первой строке представлен: подсчет всех единиц под обозначением N3, на второй строке в квадратных скобках прописываются гласные буквы последнего слога в слове, в следующей квадратной скобке прописываются предпоследние согласные буквы слов и в последней квадратной скобке прописываются конечные согласные буквы, за скобками следуют грамматические показатели имени: число и падеж (+Sg+Nom). В представленном ниже рисунке 1 в строках с 3-го по 22-й расписываются формы родительного падежа разных словоформ в соответствии с их фонетической адаптацией, требуемой законом сингармонизма в тувинском языке (+Sg+Gen).

|    |     |    |   |      |              |                      |                       |  |         |
|----|-----|----|---|------|--------------|----------------------|-----------------------|--|---------|
| 1  | SFX | N3 | У | 8641 | #            |                      |                       |  |         |
| 2  | SFX | N3 | 0 | 0    | [аевиооууэя] | [бвгджклымнпртфхцшщ] | [бвгджклымнпртфхцшщь] |  | +Sg+Nom |
| 3  | SFX | N3 | 0 | иник | [еизэ]       | [лр]                 | [ф]                   |  | +Sg+Gen |
| 4  | SFX | N3 | 0 | ууну | [оуу]        | [м]                  | [б]                   |  | +Sg+Gen |
| 5  | SFX | N3 | 0 | иник | [е]          | [р]                  | [в]                   |  | +Sg+Gen |
| 6  | SFX | N3 | 0 | ынны | [авя]        | [лн]                 | [р]                   |  | +Sg+Gen |
| 7  | SFX | N3 | 0 | иник | [еи]         | [н]                  | [р]                   |  | +Sg+Gen |
| 8  | SFX | N3 | 0 | ууну | [оу]         | [лн]                 | [р]                   |  | +Sg+Gen |
| 9  | SFX | N3 | 0 | тиң  | [еэ]         | [р]                  | [р]                   |  | +Sg+Gen |
| 10 | SFX | N3 | 0 | туң  | [оу]         | [р]                  | [р]                   |  | +Sg+Gen |
| 11 | SFX | N3 | 0 | тың  | [ая]         | [р]                  | [д]                   |  | +Sg+Gen |
| 12 | SFX | N3 | 0 | ынны | [а]          | [йлн]                | [д]                   |  | +Sg+Gen |
| 13 | SFX | N3 | 0 | иник | [еи]         | [эйр]                | [д]                   |  | +Sg+Gen |
| 14 | SFX | N3 | 0 | ууну | [о]          | [эн]                 | [д]                   |  | +Sg+Gen |
| 15 | SFX | N3 | 0 | туң  | [оу]         | [р]                  | [д]                   |  | +Sg+Gen |
| 16 | SFX | N3 | 0 | тың  | [а]          | [днр]                | [ж]                   |  | +Sg+Gen |
| 17 | SFX | N3 | 0 | тиң  | [еи]         | [др]                 | [ж]                   |  | +Sg+Gen |
| 18 | SFX | N3 | 0 | туң  | [о]          | [р]                  | [ж]                   |  | +Sg+Gen |
| 19 | SFX | N3 | 0 | тиң  | [е]          | [нр]                 | [э]                   |  | +Sg+Gen |
| 20 | SFX | N3 | 0 | тың  | [а]          | [вл]                 | [к]                   |  | +Sg+Gen |
| 21 | SFX | N3 | 0 | тиң  | [е]          | [йл]                 | [к]                   |  | +Sg+Gen |
| 22 | SFX | N3 | 0 | ынны | [а]          | [н]                  | [к]                   |  | +Sg+Gen |
| 23 | SFX | N3 | 0 | ынны | [а]          | [л]                  | [л]                   |  | +Sg+Gen |

Рис 1. Образец словоизменительной парадигмы заимствованных слов в родительном падеже (+Sg+Gen)

Словоизменительные парадигмы некоторых заимствованных слов: *курорт* ‘курорт’ и тувинского слова *өрт* ‘огонь’ в родительном (+Gen.) и дательном (+Dat.) падеже совпадают, например, на Рис. 2: *курортка* ‘курорту’, *өртке* ‘огню’ см. 234 строку на Рис. 2.

|     |     |    |   |      |          |               |     |  |         |
|-----|-----|----|---|------|----------|---------------|-----|--|---------|
| 232 | SFX | N3 | 0 | игре | [еи]     | [вйкнпфхц]    | [т] |  | +Sg+Dat |
| 233 | SFX | N3 | 0 | уга  | [оуу]    | [гйклымнпфхц] | [т] |  | +Sg+Dat |
| 234 | SFX | N3 | 0 | ка   | [авоуэя] | [дирьэ]       | [т] |  | +Sg+Dat |
| 235 | SFX | N3 | 0 | ке   | [еизэ]   | [рсть]        | [т] |  | +Sg+Dat |
| 236 | SFX | N3 | 0 | ка   | [авоуэ]  | [ств]         | [т] |  | +Sg+Dat |
| 237 | SFX | N3 | 0 | ыга  | [а]      | [н]           | [ф] |  | +Sg+Dat |
| 238 | SFX | N3 | 0 | ке   | [е]      | [й]           | [ф] |  | +Sg+Dat |
| 239 | SFX | N3 | 0 | уга  | [у]      | [м]           | [ф] |  | +Sg+Dat |
| 240 | SFX | N3 | 0 | ка   | [ауу]    | [р]           | [ф] |  | +Sg+Dat |
| 241 | SFX | N3 | 0 | ка   | [ао]     | [нр]          | [х] |  | +Sg+Dat |
| 242 | SFX | N3 | 0 | ке   | [еи]     | [кйр]         | [х] |  | +Sg+Dat |
| 243 | SFX | N3 | 0 | ка   | [а]      | [р]           | [ц] |  | +Sg+Dat |
| 244 | SFX | N3 | 0 | ке   | [е]      | [р]           | [ц] |  | +Sg+Dat |
| 245 | SFX | N3 | 0 | ыга  | [а]      | [нр]          | [ч] |  | +Sg+Dat |
| 246 | SFX | N3 | 0 | игре | [еизэ]   | [нр]          | [ч] |  | +Sg+Dat |
| 247 | SFX | N3 | 0 | уга  | [оу]     | [т]           | [ч] |  | +Sg+Dat |
| 248 | SFX | N3 | 0 | ка   | [авоу]   | [нрь]         | [ш] |  | +Sg+Dat |
| 249 | SFX | N3 | 0 | ке   | [еиу]    | [нрь]         | [ш] |  | +Sg+Dat |
| 250 | SFX | N3 | 0 | ка   | [о]      | [р]           | [щ] |  | +Sg+Dat |
| 251 | SFX | N3 | 0 | ка   | [авоуэя] | [вджсплщш]    | [ь] |  | +Sg+Dat |
| 252 | SFX | N3 | 0 | ке   | [еи]     | [джсплщш]     | [ь] |  | +Sg+Dat |
| 253 | SFX | N3 | 0 | га   | [аоуэя]  | [лнр]         | [ь] |  | +Sg+Dat |
| 254 | SFX | N3 | 0 | ке   | [аоу]    | [лнр]         | [ь] |  | +Sg+Dat |

Рис 2. Образец словоизменительной парадигмы заимствованных слов, оканчивающихся на *-рт* в дательном падеже

В словоизменительной парадигме нами расписываются полные две формы направительного падежа в тувинском языке: фор-

ма на *-че / -же* (+Lat1) и форма на *-дыва / -тыве* (+Lat2) со всеми фонетическими вариантами. Например, см. строки 416 на Рис. 3: *курортче* ‘на курорт’, *өртче* ‘на огонь’.

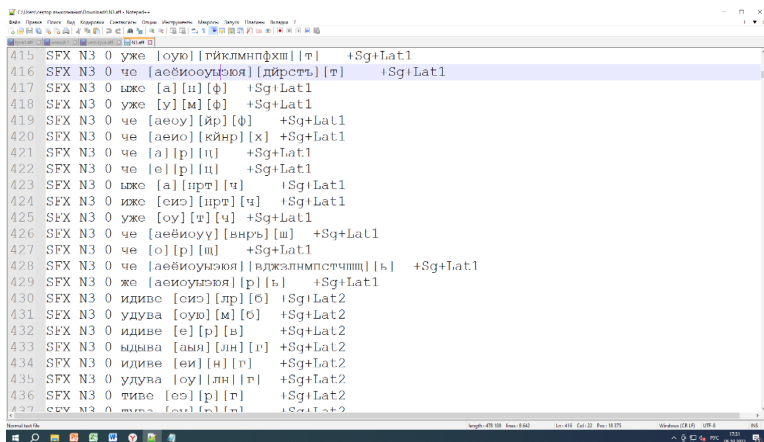


Рис 3. Образец словоизменительной парадигмы заимствованных слов в тувинском языке в направительном падеже *-че / -же* (+Lat1)

И также форма в направительном падеже на *-дыва / -тыве* (+Lat2) см. строку 480 *курорттыва* ‘на курорт’ и в строке 490 *өрттыве* ‘на огонь’. Рис. 4

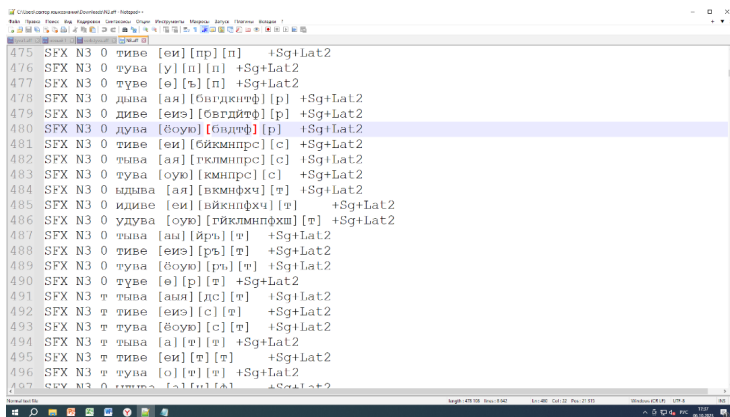


Рис 3. Образец словоизменительной парадигмы заимствованных слов в тувинском языке в направительном падеже *-дыва / -тыве* (+Lat2).

Парадигма множественного числа также отличается, так как аффикс множественного числа зависит от гласной буквы последнего слога спрягаемого слова. Например, см. строки на Рис. 5: *курорттар* ‘курорты’, *өрттер* ‘огни’.

|     |     |    |   |      |          |              |     |         |
|-----|-----|----|---|------|----------|--------------|-----|---------|
| 568 | SFX | N3 | 0 | улар | [oʊʊ]    | [Гийсмипфхш] | [т] | +P1+Nom |
| 569 | SFX | N3 | 0 | тар  | [абоуыа] | [дйрсь]      | [т] | +P1+Nom |
| 570 | SFX | N3 | 0 | тер  | [оио]    | [реть]       | [т] | +P1+Nom |
| 571 | SFX | N3 | т | тер  | [ей]     | [сг]         | [т] | +P1+Nom |
| 572 | SFX | N3 | т | тар  | [абоуо]  | [сг]         | [т] | +P1+Nom |
| 573 | SFX | N3 | 0 | ылар | [а]      | [н]          | [ф] | +P1+Nom |
| 574 | SFX | N3 | 0 | өтер | [ө]      | [й]          | [ф] | +P1+Nom |
| 575 | SFX | N3 | 0 | улар | [у]      | [м]          | [ф] | +P1+Nom |
| 576 | SFX | N3 | 0 | тар  | [аоу]    | [р]          | [ф] | +P1+Nom |
| 577 | SFX | N3 | 0 | тар  | [ао]     | [нр]         | [х] | +P1+Nom |
| 578 | SFX | N3 | 0 | тер  | [ои]     | [кыр]        | [х] | +P1+Nom |
| 579 | SFX | N3 | 0 | тер  | [а]      | [р]          | [ш] | +P1+Nom |
| 580 | SFX | N3 | 0 | тер  | [ө]      | [р]          | [ш] | +P1+Nom |
| 581 | SFX | N3 | 0 | ылар | [а]      | [нр]         | [ч] | +P1+Nom |
| 582 | SFX | N3 | 0 | илер | [өиэ]    | [нр]         | [ч] | +P1+Nom |
| 583 | SFX | N3 | 0 | улар | [оу]     | [т]          | [ч] | +P1+Nom |
| 584 | SFX | N3 | 0 | тар  | [абоу]   | [нр]         | [ш] | +P1+Nom |
| 585 | SFX | N3 | 0 | тер  | [өиу]    | [нр]         | [ш] | +P1+Nom |
| 586 | SFX | N3 | 0 | тар  | [ө]      | [р]          | [ш] | +P1+Nom |
| 587 | SFX | N3 | 0 | тар  | [абоуа]  | [нджэсгчш]   | [ы] | +P1+Nom |
| 588 | SFX | N3 | 0 | тар  | [өи]     | [джэсгшш]    | [ы] | +P1+Nom |
| 589 | SFX | N3 | 0 | дар  | [аоукэн] | [а]          | [ы] | +P1+Nom |
| 590 | SFX | N3 | 0 | оур  | [оим]    | [о]          | [ы] | +P1+Nom |

Рис 4. Образец словоизменительной парадигмы заимствованных слов в тувинском языке во множественном числе

Притяжательная форма: см. строки на Рис. 5: *курорту* ‘его курорт’, *өртү* ‘его огонь’, где ожидалось бы озвончение конечного согласного.

|      |     |    |    |      |        |             |     |               |               |
|------|-----|----|----|------|--------|-------------|-----|---------------|---------------|
| 1138 | SFX | N3 | 0  | им   | [өио]  | [бнлдйф]    | [р] | +Sg+PxSg1+Nom |               |
| 1139 | SFX | N3 | 0  | ум   | [ёоуэ] | [бвдтф]     | [р] | +Sg+PxSg1+Nom |               |
| 1140 | SFX | N3 | с  | изим | [өи]   | [бйкмнр]    | [с] | +Sg+PxSg1+Nom |               |
| 1141 | SFX | N3 | сс | зим  | [өй]   | [с]         | [с] | +Sg+PxSg1+Nom |               |
| 1142 | SFX | N3 | с  | ызым | [ая]   | [гкльнр]    | [с] | +Sg+PxSg1+Nom |               |
| 1143 | SFX | N3 | сс | зым  | [ая]   | [с]         | [с] | +Sg+PxSg1+Nom |               |
| 1144 | SFX | N3 | с  | узум | [оуэ]  | [кмлнр]     | [с] | +Sg+PxSg1+Nom |               |
| 1145 | SFX | N3 | сс | зум  | [оу]   | [с]         | [с] | +Sg+PxSg1+Nom |               |
| 1146 | SFX | N3 | 0  | ым   | [ая]   | [вкймрфхч]  | [т] | +Sg+PxSg1+Nom |               |
| 1147 | SFX | N3 | 0  | им   | [өиэ]  | [вйкмрфхч]  | [т] | +Sg+PxSg1+Nom |               |
| 1148 | SFX | N3 | 0  | ум   | [ёоуэ] | [гйкмнрфхш] | [т] | +Sg+PxSg1+Nom |               |
| 1149 | SFX | N3 | 0  | им   | [өио]  | [р]         | [т] | +Sg+PxSg1+Nom |               |
| 1150 | SFX | N3 | т  | дым  | [а]    | [ы]         | [т] | +Sg+PxSg1+Nom |               |
| 1151 | SFX | N3 | т  | дим  | [э]    | [ы]         | [т] | +Sg+PxSg1+Nom |               |
| 1152 | SFX | N3 | т  | дум  | [ө]    | [ы]         | [т] | +Sg+PxSg1+Nom |               |
| 1153 | SFX | N3 | 0  | үм   | [ө]    | [р]         | [т] | +Sg+PxSg1+Nom |               |
| 1154 | SFX | N3 | т  | рүм  | [G]    | [о]         | [р] | [т]           | +Sg+PxSg1+Nom |
| 1155 | SFX | N3 | т  | ым   | [а]    | [д]         | [т] | +Sg+PxSg1+Nom |               |
| 1156 | SFX | N3 | 0  | ым   | [ая]   | [с]         | [т] | +Sg+PxSg1+Nom |               |
| 1157 | SFX | N3 | ст | зим  | [өй]   | [с]         | [т] | +Sg+PxSg1+Nom |               |
| 1158 | SFX | N3 | ст | зум  | [ёоуэ] | [с]         | [т] | +Sg+PxSg1+Nom |               |
| 1159 | SFX | N3 | тт | дым  | [а]    | [т]         | [т] | +Sg+PxSg1+Nom |               |
| 1160 | SFX | N3 | тт | дум  | [ө]    | [т]         | [т] | +Sg+PxSg1+Nom |               |

Рис 5. Образец словоизменительной парадигмы заимствованных слов в тувинском языке в притяжательной форме

## Заключение

Обнаружено, что заимствованные из русского языка лексемы, оканчивающиеся на стечения согласных *-рт*, и тувинские лексемы, оканчивающиеся на те же стечения согласных (*өрт* ‘огонь’, *дөрт* ‘четыре’, *курт* ‘червь’, *чурт* ‘страна’, сложные слова *үш-дөрт* ‘три-четыре’), имеют аналогичные словоизменительные парадигмы. Так, в выборке парадигм на стечения согласных всего насчитывается 955 ед. и 8 641 строк, которые определены системе Hunspell в группу N3. На основе обнаруженных особенностей парадигмы заимствованных существительных будет составлена база данных заимствованных слов из русского языка, их словоизменительные и словообразовательные парадигмы, составят базу формализованной грамматики, разрабатываемого в настоящее время Национального корпуса тувинского языка.

### Условные обозначения

Nom – именительный падеж

Gen – родительный падеж

Dat – дательный падеж

Lat1 – направительный падеж -

Lat2 – направительный падеж -

Sg – единственное число

Pl – множественное число

PxSg1 – притяжательная форма единственного числа –ым

PxPl1 – притяжательная форма множественного числа –ывыс

## ЛИТЕРАТУРА

1. Аннай Э. К. Экспрессивная лексика тувинского языка, характеризующая человека (в сопоставительном аспекте) : автореферат диссертации на соискание ученой степени кандидата филологических наук. Новосибирск, 2021. – 28 с.
2. Бавуу-Сюрюн М. В. Влияние русского языка на образование современных тувинских фамилий и имен / М. В. Бавуу-Сюрюн // Этносоциальные процессы в Сибири. Новосибирск, 2000. – С. 204–205.
3. Бавуу-Сюрюн М. В. Русизмы в диалектах тувинского языка [Электронный ресурс] / М. В. Бавуу-Сюрюн, М. В. Ондар // Новые исследования Тувы. – 2013. – № 4. Режим доступа: <http://www.tuva.asia>. (Дата обращения 24.02.2021 г.). – С. 39–44.
4. Бавуу-Сюрюн М. В. Современные словообразовательные процессы, обусловленные языковыми контактами (на материале тувинского языка) // Сибирский филологический журнал. 2015. №2. URL: <https://>

cyberleninka.ru/article/n/sovremennye-slovoobrazovatelnye-protsessy-obuslovlennye-yazykovymi-kontaktami-na-materiale-tuvinskogo-yazyka (дата обращения: 07.09.2023). – С.114-123.

5. Гансух Х. Особенности тувинской речи жителей Цэнгэла : автореферат автореферат диссертации на соискание ученой степени кандидата филологических наук. Новосибирск, 2009. – 20 с.

6. Дыбо А. В., Шеймович А. В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. 2014. № 2 (36). С. 20–26.

7. Татаринцев Б. И. Русские лексические заимствования в современном тувинском языке. Кызыл: Тувинское книжное издательство, 1974. – 113 с.

8. Чадамба З. Б. Тоджинский диалект тувинского языка. Кызыл: Тувинское книжное издательство, 1974. – 136 с.

9. Oorzhak B., Khertek A. Development of semantic mark-up for the corpus of Tuvan language // Proceedings of the International Conference “Turkic languages processing” Turklang-2015, Казань, Россия, 17–19 сентября 2015 года / Tatarstan Academy of Sciences L.N. Gumilyov Eurasian National University Ministry of Education and Science of the Republic of Kazakhstan Kazan Federal University Institute of Philology and Intercultural Communication. – Казань, Россия: Академия наук Республики Татарстан, 2015. – P. 351–362.

10. Салчак А. Я., Байыр-оол А. В. Электронный корпус тувинского языка: состояние, проблемы // Мир науки, культуры, образования. – 2013. – № 6(43). – С. 408–409.

11. Washington J. N., Bayyr-ool A., Salchak A., Tyers F. M. Development of a finite-state model for morphological processing of Tuvan // Родной язык: лингвистический журнал. – 2016. – No. 1(4). – P. 156–187.

УДК 81'322.2

ЛЕКСИЧЕСКО-ГРАММАТИЧЕСКИЕ СТРУКТУРЫ  
АГГЛЮТИНАТИВНЫХ ЯЗЫКОВ ДЛЯ РАЗРАБОТКИ НОВЫХ  
ТЕХНОЛОГИЙ ОБРАБОТКИ ИНФОРМАЦИИ  
(НА ПРИМЕРЕ ТАТАРСКОГО ЯЗЫКА)

*Дж. Ш. Сулейманов<sup>1</sup>, Р. А. Гильмуллин<sup>1</sup>,  
И. Р. Мухаметзянов<sup>1</sup>, А. Я. Фридман<sup>2</sup>*

*<sup>1</sup>Институт прикладной семиотики Академии наук  
Республики Татарстан, Казань, Россия*

*<sup>2</sup>Институт информатики и математического моделирования  
им. В. А. Пугилова Кольского научного центра РАН  
Апатиты, Россия*

*dvdt.slt@gmail.com, rinatgilmullin@gmail.com,  
ilnur.mukhametzyanov@gmail.com, fridman@iimm.ru*

В этой статье представлена модель для анализа агглютинативных языковых систем, основанная на их лексических и грамматических свойствах. Это включает в себя рекурсию, морфологический эллипс, функциональное разнообразие и семантическую многозначность аффиксов (в частности, аффиксов, кодирующих недостаточно определенную информацию и нечеткие команды). Это попытка решить главную проблему современных инструментов обработки и накопления данных в этих языках, а именно их низкую производительность. По нашему мнению, это обусловлено сложностями в интеллектуализации этих инструментов, которые были созданы на основе упрощенных искусственных языков программирования. Они, по сути, являются подмножеством аналитических словоизменительных языков или искусственными структурами на их основе. В качестве решения предлагается создание моделей агглютинативных языков на базе децентрализованных механизмов проверки и идентификации значений с учетом вышеупомянутых свойств этих языков. В оценке и иллюстрации наших идей будет использоваться татарский язык, так как авторы имеют большой опыт его изучения.

**Ключевые слова:** агглютинативный естественный язык, интеллектуальный инструмент для вербализации и распознавания значений, татарский электронный корпус “Туган Тел”.



LEXICAL AND GRAMMATICAL STRUCTURES  
OF AGGLUTINATIVE LANGUAGES FOR THE DEVELOPMENT  
OF NEW INFORMATION PROCESSING TECHNOLOGIES  
(ON THE EXAMPLE OF THE TATAR LANGUAGE)

*Dz. S. Suleimanov<sup>1</sup>, Ri. A. Gilmullin<sup>1</sup>,  
I. R. Mukhametzyanov<sup>1</sup>, A. Y. Fridman<sup>2</sup>*

*<sup>1</sup>Institute of Applied Semiotics, Academy of Sciences of the Republic of  
Tatarstan, Kazan, Russia.*

*<sup>2</sup>Institute for Informatics and Mathematical Modelling, Kola Science  
Center RAS, Apatity, Russia*

*dvd.t.slt@gmail.com, rinatgilmullin@gmail.com,  
ilnur.mukhametzyanov@gmail.com, fridman@iimm.ru*

This text presents a model for the analysis of agglutinative language systems based on their lexical and grammatical properties. This includes recursion, morphological ellipse, functional diversity and semantic ambiguity of affixes (in particular, affixes encoding insufficiently defined information and fuzzy commands). This is an attempt to solve the main problem of today's data processing and accumulation tools in these languages, namely their low performance. In our opinion, this is due to the difficulties in the intellectualization of these tools, which were created on the basis of simplified artificial programming languages. They are, in fact, a subset of analytical inflectional languages or artificial structures based on them. As a solution, it is proposed to create models of agglutinative languages based on decentralized mechanisms for checking and identifying values, taking into account the above-mentioned properties of these languages. The Tatar language will be used in the evaluation and illustration of our ideas, as the authors have extensive experience in studying it.

**Keywords:** Agglutinative Natural Language, Intelligent Tool for Verbalization and Meaning Recognition, Tatar Electronic Corpus "Tugan Tel".

## Введение

Изучение естественных языков состоит из трех основных компонентов: когнитивного, коммуникативного и технологического [Сулейманов, 2010]. Когнитивный компонент описывает способность языка описывать модель мира, процессы мышления и представлять структурные и функциональные аспекты знания. Коммуникативный компонент отражает способность языка кодировать, принимать и передавать информацию, обрабатывать семиотические данные и организовывать диалоги. Технологический компонент определяет формальный и концептуальный возможности естественного языка для создания эффективных инструментов обработки и хранения информации, а также разработки интеллек-

туального программного обеспечения, включая операционные системы.

Современные инструменты накопления и обработки информации на естественном языке недостаточно эффективны и плохо справляются с задачами поиска и отбора информации в распределённых базах данных и извлечения знаний. Причиной тому является то, что они по своей сути не являются интеллектуальными, поскольку создаются на основе простых искусственных языков программирования. Последние представляют собой подмножество флективных аналитических языков или искусственно созданных структур на основе таких естественных языков.

Другой проблемой в системах обработки естественных языков является то, что их модели строятся на основе формализованных систем, например, универсальных грамматик (например, [Хомский, 1957]), Это создает две основные проблемы: монотонность выводов и пассивность средств логико-семантического анализа данных. В работе [Цейтин, 1980] такие модели естественного языка названы глобальным подходом к исследованию естественного языка. Однако эти методы не подходят для изучения агглютинативных языков и татарского в частности, так как они не способны адекватно описать особенности этих языков и их структуру.

## II. ДЕЦЕНТРАЛИЗОВАННАЯ МОДЕЛЬ АГГЛЮТИНАТИВНОГО ЕСТЕСТВЕННОГО ЯЗЫКА

В своей статье, Г.С.Цейтин [Цейтин, 1980] предугадал будущие сложности в моделировании языка, исходящие из традиционного глобального подхода к изучению языков. На основе своего обширного опыта в сфере логики и программирования, ученый предложил иной подход – рассматривать язык как множество отдельных подсистем, которые взаимодействуют друг с другом без выделения какой-либо единой системы. Однако такое направление, по всей видимости, не получило продолжения, хотя оно хорошо согласуется с современными распределёнными и многоагентными системами [Дорри, 2018], а также с прагматическим подходом к созданию лингвистических моделей [Сулейманов, 1998].

Мы считаем, что при децентрализованном подходе к созданию моделей естественного языка можно использовать более гибкий и прагматический подход к их интеллектуальному развитию. Это может быть достигнуто за счет применения сложных семиотиче-

ских моделей, которые изначально предназначены для решения нетривиальных семантических задач.

### III. СЕМИОТИЧЕСКИЕ МОДЕЛИ ЛЕКСИЧЕСКИХ И ГРАММАТИЧЕСКИХ СРЕДСТВ АНГЛИЙСКОГО ЯЗЫКА

Децентрализованный подход основан на традиционной иерархической структуре языка. Обычно выделяют четыре уровня: фонетический (самый высокий), морфологический, синтаксический и семантический (самый глубокий). Аналитическая модель обеспечивает лингвистические знания для алгоритмов анализа, переводя информацию с более высокого уровня на более низкий. Инструменты в системе, которые соответствуют обрабатываемым данным, активизируются в зависимости от их типа и уровня. То есть инструменты вторичны по отношению к изучаемым данным и не обладают собственным интеллектом.

Мы предлагаем использовать интеллектуальные инструменты, которые обрабатывают языковую информацию и выполняют семантически ориентированные задачи. Эти инструменты работают аналогично агентам в многоагентных системах, однако они не создаются самостоятельно, а разрабатываются совместно с учетом их взаимодействия и общей цели системы. Для координации работы этих инструментов мы предлагаем использовать модели, которые содержат семиотические модели Поспелова-Полякова [Поляков, 2017]. Эти модели обеспечивают целостность и полноту использования информации на каждом этапе работы системы моделирования. Благодаря этому свойству, система сможет реализовать все основные характеристики ситуационного подхода [Поспелов, 1986], такие как строгость логического вывода, однозначная классификация ситуаций, согласованность задач между моделями. возможность однозначной классификации и обобщения ситуаций [Поспелов, 1986; Фридман, 2018]; согласованность задач для всех моделей, участвующих в решении текущей задачи [Фридман, 2016].

### IV. ПРОГРАММНО-АЛГОРИТМИЧЕСКОЕ НАПОЛНЕНИЕ СИСТЕМ МОДЕЛИРОВАНИЯ АГГЛЮТИНАТИВНОГО ЕСТЕСТВЕННОГО ЯЗЫКА

В данном случае мы в основном говорим о некоторых способах интеллектуального развития разрабатываемой модели агглютинативного языка.

Как и в большинстве современных инструментов для работы с естественными языками, основная структура формализации и представления информации будет базироваться на системе онтологий. Специфика ее построения будет заключаться в поддержке ситуативного подхода [Артемьева, 2018]. У авторов имеется бэклог по данной теме, представленный в [Зуенко, 2013]. Однако, безусловно, он требует дополнительной проверки и доработки для каждого вида задач в рамках текущего проекта. Инструменты координации взаимодействия между средствами анализа информации и обеспечения корректности ее обработки будут строиться в рамках концепции ситуативной осведомленности, в то время как общая модель моделирования будет позиционироваться как сетцентрическая структура [Фридман, 2020].

Предполагается, что тестирование и верификация модели моделирования будет проводиться с использованием обширной электронной базы “Туган Тел” [Гильмуллин, 2017], которая включает около 200 млн морфологически маркированных слов.

## V. О ПОТЕНЦИАЛЕ ГРАММАТИКИ ТАТАРСКОГО ЯЗЫКА ДЛЯ РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Важными особенностями систем принятия интеллектуальных решений являются активность знаний и способность работать с неоднозначной информацией.

Активность знаний определяется структурой предложения татарского языка  $\langle S, O, V \rangle$ , которое сначала воспринимает данные, затем обрабатывает их и только потом начинает действовать. Английский язык имеет структуру  $\langle S, V, O \rangle$ , которая сначала принимает решение, а затем воспринимает данные. Татарский язык имеет богатую аффиксальную систему, включающую около 90 аффиксов. Они служат операторами контекстного управления и могут переводить слова из одного типа в другой. Например, аффикс “-ГА” может управлять более чем 15 различными контекстами.

S означает субъект, O обозначает объект, а V обозначает глагол.

В следующем примере татарские слова приведены с морфологическим разбором, переводы на русский даны в угловых скобках.

(бакчаГа  $\langle \text{сад} + \text{ГА} \rangle$   $\langle \text{сад} + \text{существительное} + \text{единственное число} + \text{падеж (ГА)} \rangle$  бара  $\langle \text{идет} \rangle$ ) означает “идет в сад” .;

(китапКА(ГА) <книга+ГА> <книга+существительное +единственное число+ падеж (ГА)> алмаштырды <изменено>) означает “поменял на книгу”;

(кулГА <рука+ГА> <рука+существительное+единственное число+ падеж (ГА)> ала <берет>) означает “берет в руку”.

Это пример неопределенной контекстно-зависимой информации: (урамДАГЫДАГЫНЫКЫНЫКЫ <улица+ДАГЫ+ДАГЫ+НЫКЫ+НЫКЫ > <улица+Существительное+Единственное число+ Атрибут-Локатив(ДАГЫ)+ Единственное число + Атрибут- Локатив(ДАГЫ)+Единственное число+ Атрибут-Генитив(НЫКЫ)+Единственное число+ Атрибут-Генитив(НЫКЫ)>) обозначает “то, что принадлежит тому, что принадлежит тому, что находится на том, что находится на улице”. Эта словоформа приобретает значение в контексте фразы, например, “цвет пера птицы, которая сидит на дереве на улице”. Примером нечеткой команды является следующая: (баргалаштыргалаштыр <бар+ГАлА+штЫр+ГАлА+штЫр> <иди+глагол+ уменьшение частоты\_1(ГАлА)+ уменьшение частоты\_2(штЫр)+ уменьшение частоты\_1(ГАлА)+ уменьшение частоты\_2(штЫр)>) означает “ходи редко, редко, редко, редко”.

В последнее время активно развивается область искусственного интеллекта, в частности, технологии нейронных сетей и машинного обучения. В связи с этим возникает необходимость в исследованиях по управлению искусственным интеллектом и созданию систем, способных интерпретировать решения и быть интерпретируемыми. На первый план выходит задача “объяснения” решений, принимаемых искусственным интеллектом. Также важно, чтобы машина могла понимать и адекватно воспринимать человека как “старшего”.

Современные системы искусственного интеллекта на основе нейронных сетей, машинного обучения и big data эффективно решают многие интеллектуальные задачи, такие как машинный перевод, распознавание изображений и речи. Однако они не способны создавать новые знания и “объяснять” свои решения, чтобы сделать их понятными для человека.

Одним из решений этой проблемы может стать разработка новых методов и подходов к обучению искусственного интеллекта, которые позволят ему не только решать задачи, но и “объяснять” свое решение. Это может включать в себя разработку новых алгоритмов обучения, использование более сложных моделей и мето-

дов анализа данных, а также создание новых подходов к обработке и интерпретации информации.

## VI. ЗАКЛЮЧЕНИЕ И ПЛАНЫ БУДУЩИХ ИССЛЕДОВАНИЙ

Изучение технологического аспекта агглютинативных языков может привести к выявлению естественных лексико-грамматических структур, которые могут быть использованы для создания новых языков программирования. Эти языки, в свою очередь, могут расширить возможности интеллектуальной обработки информации.

Децентрализованная структура системы моделирования агглютинативного языка позволяет некоторым из ее подсистем приобрести общезыковое значение. Это аналогично тому, как синтаксис может относиться ко всему языку за некоторыми исключениями.

В области семантики возможно формализовать отдельные подсистемы, например, обозначения родственных отношений или моменты времени. Однако, не было обнаружено доминирующей системы семантики для всего языка. Вместо этого, наиболее развитой формализованной системой семантики является теоретико-множественная семантика, заимствующая аппарат математической логики из математики.

Принимая во внимание профессиональные интересы и нарабатанный материал, которое имеется у авторов (например, [Хусаинов, 2018; Гатауллин, 2015; Сулейманов, 2018]), предполагается, что эта структура будет реализована на примере татарского языка. С высокой вероятностью представленный здесь подход может позволить продвинуться в моделировании языка для систем искусственного интеллекта.

В будущем мы планируем изучить и построить математические модели, отражающие лексический и грамматический потенциал татарского языка как основы интеллектуальных технологий, включая такие морфологические свойства, как рекурсия, морфологический эллипс, функциональное разнообразие и семантическая многозначность аффиксов (в частности, аффиксов, кодирующих недостаточно определенную информацию и нечеткие команды). Также представляется вполне разумным разработать синтаксическую структуру, обеспечивающую реализацию свойства активности знаний, которое является важным показателем интеллекта прикладной системы.

В результате будет разработан фокусно-децентрализованный подход, при котором реализуется не только децентрализация, но и определяются фокусы деятельности в соответствии с исследованием и моделированием в фокусной области, рассматривая эту “выбранную” область как своего рода целостность, возможно, с неопределенностями по краям. Такая территория должна соответствовать особенностям и критериям ситуационного подхода.

## СПИСОК ЛИТЕРАТУРЫ

1. Д. Сулейманов. К вопросу об изучении технологического аспекта естественных языков. В кн.: *Обработка текста и когнитивные технологии: Материалы XI Международной научной конференции (Констанца, 7–14 сентября 2009 г.)*. Изд-во Государственного университета, Казань, с. 232–245 (2010).
2. Н. Хомский. *Синтаксические структуры*. Мутон, Гаага (1957).
3. Г. Цейтин. О взаимосвязи между естественным языком и формальной моделью. В: *Архив Академии наук СССР. Работа в Научном совете по комплексной проблеме “Кибернетика”* (1980).
4. А. Дорри, С. Канхере, Р. Джурдак. Мультиагентные системы: обзор. Доступ к IEEE, 6, стр. 28573–28593 (2018).
5. Д. Сулейманов. Обработка естественных языков-текстов на основе прагматически ориентированных лингвистических моделей. В кн.: *Обработка текста и когнитивные технологии. Выпуск 3. Материалы научного семинара “Когнитивное моделирование”* (Пушино, октябрь 1998 г.), стр. 205–212 (1998).
6. Т. Медведева. *Формальные модели в лингвистике: Учебное пособие*. Научная книга, Саратов (2010).
7. В. Поляков. Проблемы представления, приобретения и использования знаний в свете обработки естественного языка. В кн.: *Когнитивно-семиотические аспекты моделирования в гуманитарных науках*. Издательство Академии наук Республики Татарстан, Казань, стр. 145–163 (2017).
8. Д. Поспелов. *Ситуационное управление: теория и практика*. Медицинский институт Бателле, Колумбус, Огайо (1986).
9. А. Фридман. Когнитивная категоризация в иерархических системах под ситуационным контролем. В: *Достижения в области исследований интеллектуальных систем*, том 158, Atlantis Press, стр. 43–50 (2018).
10. А. Фридман. Планирование и координация в иерархиях интеллектуальных динамических систем. *ТЕЛКОМНИКА*, 14(4), 1408–1416 (2016).

11. И. Артемьева, А. Фридман. Онтологии в задаче автоматизации ситуационного моделирования. В: 3-я Российско-Тихоокеанская конференция по компьютерным технологиям и приложениям (RPC), IEEE, стр. 48–53 (2018).

12. Б. Кулик, А. Фридман. Сложные методы логического анализа, основанные на простой математике. Издательство Cambridge Scholars, Ньюкасл-апон-Тайн (2022).

13. А. Зуенко, А. Фридман, Б. Кулик. Генерация последовательности вопросов в интеллектуальных системах обучения, основанных на алгебраическом подходе. Информатика и информационные технологии, 1(2), 125–131 (2013).

14. М. Эндсли. Заключительные размышления: модели и меры осознания ситуации. Когнитивная инженерия и принятие решений, 9(1), 101–111 (2015).

15. А. Фридман, А. Олейник. Моделирование ситуационной осведомленности в сетевых коммерческих системах. В: Материалы 34-й ежегодной Европейской конференции по имитационному моделированию LAAS-CNRS Tou-louse - Франция, 21–23 октября, стр. 64–67 (2020).

16. Татарский национальный корпус “Туган Тел”, <http://tugantel.tatar>, последняя обработка 2022/04/11.

17. Р. Гильмуллин, Б. Хакимов, Р. Гатауллин. Нейросетевой подход к устранению морфологической неоднозначности, основанный на архитектуре LSTM в Национальном корпусе татарского языка. Материалы семинара CEUR, 2303 (2018).

18. Р. Гильмуллин, Р. Гатауллин. Система морфологического анализа татарского языка. В: Конспекты лекций по информатике (включая подсерии Конспектов лекций по искусственному интеллекту и конспекты лекций по биоинформатике), 10449 LNAI, стр. 519–528 (2017).

19. А. Хусаинов, Д. Сулейманов, Р. Гильмуллин, А. Гатиатуллин. Построение татарско-русской системы НМТ на основе повторного перевода многоязычных данных. В: Конспекты лекций по информатике (включая подсерии Конспектов лекций по искусственному интеллекту и конспекты лекций по биоинформатике), 11107 LNAI, стр. 163–170 (2018).

20. А. Гатауллин, Р. Гильмуллин, Д. Сулейманов. Методы и программные средства устранения морфологической неоднозначности в текстах на татарском языке. 10(24), 44795–44800 (2015).

21. Д. Сулейманов, Д. Якубова. Лексический и грамматический потенциал тюркских языков для разработки новых технологий обработки информации. В: Материалы XV Международной конференции по компьютерной и когнитивной лингвистике TEL-2018, в 2-х томах. Серийный. “Интеллект. Язык. Компьютер”, стр. 361–372 (2018).



УДК 81'33

## КОМБИНАТОРНЫЕ СВОЙСТВА ЛЕКСИЧЕСКИХ ЕДИНИЦ

**Б. А. Юнусова**

*Самаркандский государственный университет  
имени Шарофа Рашидова, Самарканд, Узбекистан*  
yunusovabakhora@gmail.com,

В статье рассматривается и анализируется отличие слова от лексемы, их сходство, виды лексического сочетания в процессе употребления слова в контексте текста, комбинаторная лексикология.

**Ключевые слова:** лексема, синтагматика, лексическая единица, парадигма, лексическое сочетание, комбинатор.

## COMBINATORIAL PROPERTIES OF LEXICAL UNITS

**Bahora Yunusova**

*Sharof Rashidov Samarkand State University,  
Samarkand, Uzbekistan*  
yunusovabakhora@gmail.com

The article examines and analyzes the difference between a word and a lexeme, their similarity, types of lexical combinations in the process of using a word in the context of a text, combinatorial lexicology.

**Keywords:** lexeme, syntagmatics, lexical unit, paradigm, lexical combination, combinator.

Известно, что все слова, существующие в языке, называются словарным составом или лексикой. При этом изучаются проблема слова, являющегося основной единицей языка, построение словарного состава, применение, обогащение, развитие словарных единиц и другие аспекты. Несмотря на многолетнее изучение слова и связанных с ним явлений языка и речи, в настоящее время оно остается основным источником исследования в языкознании. Основной причиной этого является постоянное изменение и обновление слова и связанных с ним явлений и тот факт, что слово, связанные с ним понятия занимают важное место в качестве средства общения в обществе. Поэтому проблема слова является основным источником изучения лексикологии. Мы постараемся осветить тему, опираясь на суждения ученых о различии, общих чертах слова и лексемы. Лексема реализуется в речи в слове. Наряду с тем, что лексема является готовой, общей и обязательной для всех членов общества, она обладает также следующими иными

свойствами: 1. Член общества не создает лексему, принимает ее в готовом виде. 2. В сознании члена общества лексема «живет» в одном ряду со схожими лексемами (в парадигмах). Например: [daftar] ~ [bloknot]; [daftar] ~ [oynoma] ~ [ro'znoma]; [daftar] ~ [qissa] ~ [roman]; [daftar] ~ [miqova] ~ [varaqa] ~ [bet] ~ [bob] ([тетрадь] ~ [блокнот]; [тетрадь] ~ [журнал] ~ [газета]; [тетрадь] ~ [повесть] ~ [роман]; [тетрадь] ~ [обложка] ~ [лист] ~ [страница] ~ [глава]). Слово *daftar* на основе этих отношений имеет несколько смыслов. 3. Лексемы в сознании человека «живут» также в соседских (синтагматических) отношениях. Например: [тетрадь] ~ [пиши] ~ [возьми] ~ [качественный] ~ [математика] ~ [родной язык]; [тетрадь] ~ [числовые дополнения] ~ [притяжательные аффиксы] ~ [надежные дополнения]... Эти сходственные и соседские отношения, возможности смысла и задач проясняются, уточняются в речи. Следовательно, лексемы являются также совокупностью речевых возможностей, реализованных и реализуемых в сознании носителей языка [Влавацкая, 2010]. В процессе применения слова в окружении текста существует 2 вида лексических комбинации: *внутренняя комбинация и внешняя комбинация*. *Внутренняя комбинация* – имеет целостный смысл, состоит из стабильных отношений двух и более слов до процесса речи, привносится в речь в готовом виде, образуя переносный смысл, реализуется посредством фразеологической или лексической единицы. При этом ярко проявляются, в основном, в *описательном выражении (перифразах), фразах (фразеологизмах), паремологических единицах (поговорах и поговорках), мудрых словах (афоризмах)*.

Смысл слова «подняться» от слова «высокий» скрыт в значении «отличаться, побеждать», «превзойти друг друга, превзойти друг друга, не опуститься ниже». Он пил воду из высокого корыта. Прийти с высоты 1) высокомерно говорить, высокомерно поступать; 2) установить большую, высокую цену, завесить цену. Не поднимайся с такой высоты, спускайся. Хочешь продать, возьми (в Торге). Его нос (или клюв) высокий. димог. Высокомерный Очень самоуверенный, высокомерный. Рука высоко 1) повезло, крупный бизнес, повезло. Не забывай, моя дорогая, мы будем рады, если в этом году получим хороший урожай. Ш. Рашидов, Сильнее бури; 2) победитель, победитель. Но в то время, когда сугдийцы были в приподнятом настроении, я получил дополнительную поддержку из Мароканда в размере одного округа.

Из приведенных примеров видно, что через взаимодействие действия и его результата, отношение действия и его исполнителя, взаимодействие материала и сделанной из него вещи в словах может обретаться новый смысл. выражать вещи. Чтобы назвать вещи и предметы в человеческом существовании, необходимо выявить их важные признаки, знать изменение отношения к этим вещам и предметам в повседневной жизни, а также понимать, что одно слово может сочетаться с другим словом или сочетанием. То есть, анализируя качественные свойства предмета или события, используя его для выделения и описания важного признака, воспринимать, замечать, воспринимать, понимать, знать и в уме воспринимать набор признаков, принятый группой говорящих. должен уметь воплотить в жизнь описываемую в его воображении вещь или событие.

Поэтому говорящий, опираясь на свои знания, основанные на языковых и жизненных обобщениях, замечает в характеристиках определенной вещи или человека некоторые общие черты между другой вещью или человеком, характер связи между ними, т. е. путем их соединения находят общий доминирующий характер и назвать его на своем языке. В результате у названия первой вещи (первичного референта) появляется новый смысл и на его основе появляется новое имя у второй вещи (вторичный референт). При создании нового имени говорящий должен обладать высоким уровнем мышления, то есть человеческий разум должен быть способен выносить суждения и выводы. Потому что, если слушатель или читатель не сможет вынести суждение о наличии сходства между признаками предмета, если он не увидит общности признаков, он не поймет смысла. В этом процессе важно, чтобы слова говорящего выступали в разных значениях в потоке речи, сочетании слов и положении слова в возникновении таких значений.

В именовании выделяют три аспекта: именуемый объект, субъект именованного и элементы выбираемого языка. Объектом для наименования может быть отдельное понятие, предмет, знак (красота, книга, скажем, зелень), предмет с конкретными признаками (зеленое дерево) или целое событие (Весна! Птицы полетели). Содержание символа, выбранного в процессе именованного в качестве основы для наименования, является основой для образования внутренней комбинации. Итак, один и тот же объект может называться по-разному в зависимости от его разных знаков. В общий словарный запас обычно включаются имена, соответствующие

законам внутреннего развития языка и способные удовлетворить потребности представителей того или иного языка.

*Внешняя комбинация* – реализуется посредством лексических единиц, образованных на основе переноса в процессе речи в качестве имени слов в прямом смысле и означающих прямой смысл из предметов и явлений в другие предметы и явления. Это ярко проявляется в *словосочетаниях и метафорах*. Паремиологические единицы и афоризмы, составляющие структуру, рассматриваются как объект изучения литературы в целом, а поскольку содержание и цель обучения в этих единицах занимают первостепенное место, то давая определения и пояснения лингвистически и принося их в классификации как лингвистическое явление сбивает с толку и может привести к ложным выводам [Абузалова, 2008].

Комбинаторная лексикология изучает взаимодействие слов в потоке языка и речи [Кадилова, 2022]. Комбинаторная лексикология – это оптимальная интерпретация сочетания лексических единиц, которая осуществляется в словарях. Разработка двух язычных словарей осуществляется посредством изучения комбинаторных свойств уровней языка. Комбинаторное языкознание является особой областью узбекского языкознания, изучающее отношения между различными единицами языка в качестве системы признаков языка. Комбинаторное языкознание в соответствии с предметом изучения и в качестве самостоятельного направления охватывает масштабные проблемы соответствия (сочетания) единиц языка. С этой точки зрения возникает проблема существования определенного метаязыка и возможность описания с его помощью. Результат исследования показывает, что в XX веке возникло множество понятий соответствия, которые используются в качестве синонимов. Приведение к определенной норме этих понятий в терминологической системе и выбор приемлемого варианта зависит от развития комбинаторного языкознания.

#### ЛИТЕРАТУРА

1. Влавацкая М.В. Комбинаторная лингвистика в структуре науки о языке // Вестник Ленинградского государственного университета имени А.С.Пушкина. – (Ленинград, 2010. – С. 23–31).
2. Абузалова М., Назарова С. Систем тилшунослик асослари. – (Бухоро, 2008. – Б. 20).
3. Кадилова З.З. Алишер Навоийнинг насрий асарларида перифразлар: Фил. фан. бўйича фалс. докт. (PhD). – (Термиз, 2022. – Б. 13).

УДК. 491.3:809

**НЕВЕРБАЛЬНЫЕ КОМПОНЕНТЫ КОММУНИКАЦИИ  
В ПРОИЗВЕДЕНИЯХ ФОЛЬКЛОРА***Д. Б. Уринбаева**Национальный центр обучения педагогов новым методикам  
Самаркандской области, Кафедра методики преподавания  
языков, доктор филологических наук, профессор.**Узбекистан, Самарканд  
dilbarxon@inbox.ru*

В статье рассматриваются основные типы невербальных компонентов коммуникации, встречающихся в текстах фольклора. Отмечаются наиболее характерные для коммуникации виды невербальных компонентов.

**Ключевые слова:** семиотика, коммуникация, вербальный, невербальный, фольклор, сказки, дастан, песни, пословицы, загадки.

**NONVERBAL COMPONENTS OF COMMUNICATION  
IN WORKS OF FOLKLORE***Urinbayeva Dilbar Bazarovna**National center for training teachers in new methods  
of Samarkand region**dilbarxon@inbox.ru,*

The main types of non-verbal communication components found in folklore texts are considered. The most characteristic types of non-verbal components are noted.

**Key words:** semiotics, communication, verbal, non-verbal, folklore, fairy tales, dastan, songs, proverbs, riddles.

В сегодняшние дни самой популярной новой научной дисциплиной стала – «невербальная семиотика», целью которой является изучение невербального поведения человека и решение проблемы, связанной с соотношением вербального и невербального кодов [Махина, 2017; Khasanova, 2020]. Вербальный и невербальный компоненты коммуникации можно рассматривать как знак. Таким образом, коммуникация может иметь знаковый характер и представлять собой совместную деятельность субъектов. Семиотический подход к знаку позволяет выявить закономерности функционирования невербального и вербального компонентов как типов знаков в коммуникации.

Во многих научных трудах выделены следующие невербальные компоненты поведения человека: 1) *кинетические: жесто-*

вые, мимические, пантомимические, тактильные; 2) миремические; 3) паралингвистические: фонационные, респираторные.

Мы взяли для анализа фольклорные тексты, так как для них характерно самое полное функциональное проявление прагматического потенциала всех языковых единиц, в том числе и эмоций. Исходным объектом для наших исследований послужили два жанра узбекского устного народного творчества, такие как дастан (эпос) (1), узбекские народные сказки (2).

#### Анализируемые произведения и их объем

|    | Произведения                           | Выборка N | Лс/ф  |
|----|--|-----------|-------|
| 1. | «Алпомиш» (Ташкент, 1998)              | 14029     | 96011 |
| 2. | «Ўзбек халқ эртақлари» (Ташкент, 2007) | 14502     | 76666 |

В результате проведённых исследований были составлены частотные словари этих жанров, затем объединенные в один единый словарь, в котором была указана частотность употребления каждого слова. Применяя статистико-информационную методику, мы провели сравнение между жанрами фольклора – дастан, сказки, народные песни, пословицы, загадки. Так как именно в произведениях фольклора уделено пристальное внимание детальному изображению героев: их поведению, чувствам и эмоциям в различных ситуациях. Анализ фактического материала показал, что традиционная поэтика узбекской сказки характеризуется прагматической сдержанностью, конкретикой и логикой изложения, стремлением к правдоподобию ведению повествования. Узбекской же системе стереотипов языка и стиля свойственны такие черты, как легкость, пародийность, шутливость, ироничность, по возможности уход от реального ведения повествования.

Рассмотрим пример, в котором невербальный компонент жестового характера выражает положительную эмоциональную реакцию: Enasi *kulib* borganini ko'rib, *vaqti xush bo'lib*, bir so'z aytib turgan ekan:

Men yig'ladim yaratganga zori-zor,  
Zorimni eshitsu qudratli jabbor,  
Duogo'ying (menman), enang mushtipar,  
Ne sababdan kulib kelding, Yodigor? (Д. 357).

– Shunda tulki *qah-qah urib kulib yuboribdi*, cho'ponga: Shu kichkina xaltaga katta ilonning sig'ganiga men hayron bo'lib, kulib yubordim, – debdi. Ilon “unga men sig'dim”, – debdi (С. 17).

В процессе речевого акта коммуникант совершает невербальное действие, которое описано в тексте с помощью выражения “*kulib* (Д.32; С.22<sup>1</sup>) / смеяться”, “*vaqti xush bo’lib* (Д.19) / веселое время провождение”, “*qah-qah urib kulib yuboribdi* (С.2) / смеялся от души”. Анализ дефиниции лексемы и ее интерпретации в рамках контекста позволяет сделать вывод, что данные жесты, совершенные коммуникантами дают разные оттенки. В жесте «*kulib*» выражается счастье коммуникатора, в «*vaqti xush bo’lib*» уровень счастья значительно выше, в «*qah-qah urib kulib yuboribdi*» не виден смысл счастья, в котором значение презрения, сарказм и недоверие фокусируются с голосом. В примере 1 положительная эмоциональная реакция выражается через жестовый компонент, а во втором - негативная эмоция. Значит, слово “*qah-qah urib kulib yuboribdi*” в тексте используется 22 раза, но не все значения дают позитивный оттенок. Исследования показали, что в текстах сказок это слово использовалось 8 раз с негативным оттенком, 14 раз с позитивным значением. А в текстах дастана 32 раза используются только позитивные оттенки. Как показывает анализ практического материала, наряду с жестовым невербальным компонентом в текстах фольклора широко представлены фонационные компоненты, характеризующие голос коммуниканта (говорить раздражительно, говорить весело, закричать от удовольствия и др.). Фонационные компоненты различаются по своей эмоциональной направленности.

Анализ практического материала позволяет нам сделать вывод, что невербальные компоненты, характеризующие любые движения глаз коммуниканта, следует отнести к менее употребительным. При этом наиболее интересным и важным для нас представляется то, что эти паралингвистические средства, будучи по своей эмоциональной направленности как положительными, так и отрицательными, выражают высшую степень проявления эмоции.

Проявление столь же сильной эмоции встречаем и в следующей коммуникативной ситуации: *Oh urib, tiqilib ko’zingdan yoshing,*

*Nega xafa bo’lib kelding emikdoshim* (Д.57).

Однако здесь из контекста ярко выражена отрицательная эмоция коммуниканта. На это указывает и дефиниция глагола «*oh urib*

---

<sup>1</sup> Д-дастан, С – сказка. Дана частота слов употребленных в произведениях.

/охая от страданий» (Д.29; С.5), «*tiqilib ko'zingdan yoshing / глаза наполнены слезами*» (Д.2), «*xafa bo'lib / расстроился*» (Д.42; С.53), и анализ контекста. Яркая эмоция коммуниканта обусловлена неожиданными переменами в его жизни.

Как было выявлено в ходе анализа фактического материала, группа невербальных компонентов, основу которой составляют респираторные компоненты (*hursinmoq / вздохнуть, yig'lamoq / всхлипывать, piqillamoq / сопеть и др.*), немногочисленна. Кроме того, разбор фольклорных текстов показывает, что данные паралингвистические средства по своей эмоциональной направленности отрицательны, и их реализация в большинстве случаев наблюдается либо у того коммуниканта, который стоит на более низкой ступени социальной лестницы, нежели его собеседник, либо у того, кто в силу сложившихся обстоятельств находится в менее выгодном положении.

Bu gapni eshitgan Sherzodning yegan ovqatlari *tomog'iga tiqilib, nafasi ichiga tushib, shaytonlagan odamday bo'lib qoldi*. Birozdan so'ng *o'zini tutib olib, ko'zlari yoshlik va dili g'ashlik holatda, hasrat-unadomatlar bilan*:

– Ey, otajon, biz og'amiz bilan ovchilikda yuraversak, siz bilan kampirni boqolmay qolarmidik? – deb boboning *ko'nglini ko'tardi* (Сказка, 85).

Ключевым для понимания значения указанного невербальным компонентом является анализ контекста, подчеркивающий чувство страха, волнения, возмущение, испуг которое испытывает коммуникант. Это чувство выражено словосочетанием «*tomog'iga tiqilib* (С.1) / *застрял ком в горле*», «*nafasi ichiga tushib* (С.3) / *переводя дыхание*», «*shaytonlagan odamday bo'lib qoldi* (С.1) / *он стал подобен человеку, одержимому дьяволом*», «*o'zini tutib olib* (С.3) / *сдерживая себя*», «*ko'zlari yoshlik va dili g'ashlik holatda* (С.1)», «*hasrat-u nadomatlar bilan ko'nglini ko'tardi* (С.1) / *раскрыв свою печаль, развеял свою тоску*». Это ответная реакция коммуниканта на известие о том, что рыбак услышал вести о принце, а он был рядом с ним, но рыбак не знал об этом. В этом случае через респираторный невербальный компонент выражается эмоциональная реакция.

Следовательно, через данный вид невербальным компонентом эксплицируется отрицательная эмоциональная реакция, ядро которой составляет чувство печали, грусти, уныния. Вероятно, объ-



яснение этому следует искать в нормах этикета, которые имеют свою культурную специфику. Поэтому при общении коммуниканты старались не использовать контактных жестов, прибегая к ним в редких случаях, в которых, что является немаловажным, они имели явно выраженную положительную эмоциональную направленность.

Qallig'i juda sog'inganidan yorini quchoqlab, yuzlaridan o'pib, qo'lini yelkasiga tashlab yig'lab turar ekan (Сказка, 64). Bu so'zdaytib bir-birini ko'radi // Barchinni quchoqlab endi jiladi (Дастан, 392). Согласно дефиниции, глагол "*quchoqlab*" выражает привязанность к человеку. Контекст показывает, что характер этого жеста обусловлен взаимоотношениями двух людей, являющихся супругами. Следовательно, через свой жест коммуникант выражает положительную эмоцию, основу которой составляют любовь, забота, переживание. Рассмотрев различные виды невербальной коммуникации, отражающие эмоции разной направленности, мы выявили их характерные особенности на материале произведений фольклора.

#### ЛИТЕРАТУРА

1. Алпомиш. Фозил Йўлдош ўғли. Тошкент: «Шарк» нашриёти-матбаа концерни бош тахририяти, 1998.
2. O'zbek xalq ertaklari.1 tom. Toshkent: "O'qituvchi" nashriyot-matbaa ijodiy uyi, 2007.
3. Крейдлин Г. Е. Невербальная семиотика. – Москва, 2004.
4. Махина Л.А. Вербальные и невербальные средства выражения коммуникативно-прагматической категории «враждебность» в конфликтных текстах: на материале современного немецкого языка: Автореф. дисс. канд. филол.наук. – Санкт-Петербург, 2017. С.30.
5. Белозеров А.В. Языковая репрезентация коммуникативного поведения инициатора конфликта в англоязычном художественном тексте: гендерный аспект: Автореф. дисс. канд. филол.наук. – Нижний Новгород, 2017. С. 30.
6. Khasanova, G. Kh. The Functions of the Nonverbal Means in Dialogic speech// Journal of Critical Review. Vol 7, Issue 15, 2020. ISSN-2394-5125. – Pp. 6174–6183.

УДК 81'322.2

**REVIEW: «INTONATION MODEL FOR SIMPLE SENTENCES  
IN THE KAZAKH LANGUAGE FOR A KAZAKH SPEECH  
SYNTHESIZER»**

*Yenglik Kadyr, Bibigul Sh. Razahova, Aizhan Nazyrova*

*L. N. Gumilyov Eurasian National University,  
Nur-Sultan, Kazakhstan*

The proposed intonation model offers the promise of considerably enhancing applications such as voice assistants, language learning tools, and assistive technologies for the Kazakh-speaking community, fostering improved communication and access. This research contributes to the better field of speech synthesis and demonstrates its potential impact in the realm of linguistic diversity and technology-driven communication.

**Keywords:** Intonation modeling, Kazakh language, speech synthesis, simple sentences.

**ОБЗОР: «ИНТОНАЦИОННАЯ МОДЕЛЬ ДЛЯ ПРОСТЫХ  
ПРЕДЛОЖЕНИЙ НА КАЗАХСКОМ ЯЗЫКЕ  
ДЛЯ КАЗАХСКОГО РЕЧЕВОГО СИНТЕЗАТОРА»**

*Енлик Кадыр, Бибигуль Разахова, Айжан Назырова*

*Евразийский национальный университет им. Л. Н. Гумилева,  
Нурсултан, Казахстан*

Предложенная модель интонации предоставляет обещание значительно-го улучшения приложений, таких как голосовые помощники, средства обучения языку и ассистивные технологии для казахскоязычного сообщества, способствуя улучшению коммуникации и доступа. Это исследование вносит вклад в область синтеза речи и демонстрирует его потенциальное влияние в области лингвистического разнообразия и технологически ориентированной коммуникации.

**Ключевые слова:** Интонационная модель, казахский язык, синтез речи, простые предложения.

## **Introduction**

In recent years, speech synthesis, or the artificial generation of human-like speech from text or other linguistic input, has made remarkable strides. One area of particular interest is the development of intonation models tailored to specific languages, with the goal of producing more natural and expressive synthetic speech. In this section, we will focus on the development of an intonation model for simple sentences in Ka-

zakh, which will be a key component of a Kazakh speech synthesizer. This endeavor is very important for both technological and linguistic reasons. The importance of precise intonation modulation in speech synthesis cannot be overstated. Intonation is a fundamental aspect of human speech that includes fluctuations in pitch, rhythm, and stress. In spoken communication, it conveys nuanced meanings, emotions, and intentions. As a result, the capacity to replicate these intonational patterns in synthesized speech is essential for creating really authentic and contextually appropriate vocalizations.

At the moment, the utilization of AI technology in medicine is one of the most essential trends in the world. AI and neural networks can not only improve medical services, but also change, for example, diversify the diagnostic system, influence the emergence of new drugs, in a word, provide quality medicine and reduce costs. The Detectron2 libraries allow us to implement the intended program, for example, a program that analyzes skin videos and then detects skin cancer using that added video data.

A variety of sources support the crucial role of intonation in speech synthesis. H. Zen and T. Nose investigate the role of prosody in improving the expressiveness of synthesized speech, emphasizing that the correct application of intonation is essential for conveying emotions and intentions effectively [Zen, 2008].

B. Wang et al. emphasize the need for language-specific intonation models in the context of multilingual speech synthesis, arguing that generic models frequently fail to capture the subtleties of individual languages [Wang, 2015].

The importance of precise intonation modulation in speech synthesis cannot be overstated. Intonation is a fundamental aspect of human speech that includes fluctuations in pitch, rhythm, and stress. In spoken communication, it conveys nuanced meanings, emotions, and intentions.

B. Wang et al. highlight the need for language-specific intonation models in the context of multilingual speech synthesis, arguing that generic models frequently fail to capture the subtleties of particular languages [Wang, 2015]. Furthermore, the progress in the creation of intonation models for less widely spoken languages, such as Kazakh, has broader ramifications. It helps with linguistic preservation and promotes diversity in technology. According to M. Adler et al., the development of speech technology for less-represented languages is critical for ensuring that these communities can access and interact with modern technological advancements [Adler, 2018].

## Transforming Communication with an Intonation Model for Kazakh Speech Synthesis

An intonation model for the Kazakh language intends to transform the way Kazakh speech is synthesized, much like how smart devices revolutionized healthcare by supplying individuals and healthcare professionals with real-time health data. It improves communication by integrating naturalness, experience, and linguistic richness into synthesized Kazakh speech.

### Using Linguistic Sensors for Better Communication

Similarly to how smart devices use various sensors to collect health-related data, our intonation model employs linguistic sensors to analyze the intricacies of pitch, rhythm, and stress in Kazakh speech. These sensors serve as the foundation for authentic intonation patterns, allowing synthesized Kazakh speech to effectively convey emotions, intentions, and contextual nuances.

### Seamless Integration and Remote Access

High-tech medical devices send data to mobile apps or cloud-based systems with ease. Similarly, our intonation model allows for remote access and analysis of synthesized Kazakh speech. This access benefits not just individuals seeking improved communication, but also professionals, educators, and assistive technology users who can remotely access and fine-tune speech synthesis for specific contexts.

### Improving Language Technology Adoption and Inclusivity

The development of an intonation model for the Kazakh language promotes linguistic inclusion, much as health tracking through smart devices has gained popularity across diverse demographic groups. It ensures that the Kazakh-speaking populace may access and interact with technology in their native language, regardless of age or region. This corresponds to the broader trend of technology becoming more inclusive and responsive to linguistic diversity.

## **Data Collection and Preprocessing**

The acquisition of high-quality audio recordings of native Kazakh speakers uttering simple sentences across varied contexts and emotions is required for our Kazakh intonation model data collection. This corpus is the basis for training and validating our model.

We draw inspiration from successful data collection approaches used in similar language and intonation modeling projects to ensure data quality and diversity. The Multilingual Intonation Corpus (MINT) [Avesani, 2017] used a diverse set of speakers and sentence types to

capture distinct intonational patterns, whereas the Kazakh National Corpus [Beisembayev, 2013] provided valuable linguistic resources for our text data.

The preprocessing of the collected data entails several steps, including:

**Audio Segmentation:** We divide the continuous audio recordings into smaller, more manageable segments and align them with corresponding text transcriptions using forced alignment tools like the Penn Forced Aligner [Yuan, 2008].

**Extraction of FEATURES:** From the segmented audio data, we extract relevant acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs) and fundamental frequency (F0) contours. These characteristics are critical for modifying intonation patterns [Eyben, 2010].

**Normalization of Text:** We normalize and tokenize the text transcriptions to ensure consistency in sentence structure and format. Text Normalization Challenge [Zampieri, 2017] techniques can be adapted for this purpose.

**Data Augmentation:** To improve model robustness, we use data augmentation techniques such as pitch shifting and speed perturbation on our audio data [Ko, 2015].

These data processing steps seek to create a clean and standardized dataset that aligns audio and text, hence easing the training of our intonation model.

## Algorithms

We use the most recent machine learning and neural network algorithms to learn the intricate patterns of intonation from processed audio and text data for our Kazakh intonation model. These ML and NN algorithms form the basis for training our intonation model, with the goal of reflecting the complex interplay between linguistic content and acoustic features in Kazakh speech.

**LSTM (Long Short-Term Memory):** LSTMs are an essential component of our model, effectively capturing sequence dependencies in both audio and language features [Hochreiter, 1997]. This architecture has been used successfully in a variety of speech synthesis tasks.

**CNN (Convolutional Neural Networks):** CNNs are used to extract features from spectrograms or MFCC representations of audio data [LeCun, 2015]. These networks improve at capturing local patterns and are especially effective in image-like data such as spectrograms.

**Mechanisms of Attention:** Attention mechanisms [Vaswani, 2017] are used to capture fine-grained dependencies between text and audio,

allowing the model to focus on relevant areas of the input data during synthesis.

### **Key Findings and Achievements**

Our research not only advances the field of speech synthesis, but it also has ramifications for language preservation and increased access to technology for the Kazakh-speaking population. It emphasizes the importance of language-specific intonation modeling in achieving genuine and expressive synthesized speech.

**Efficient Intonation Modeling:** Our research has successfully developed an intonation model tailored to the Kazakh language, considerably enhancing the expressiveness and naturalness of synthesized speech. This achievement is consistent with the findings of H. Zen and T. Nose [Zen, 2008], who emphasized the importance of intonation in improving the expressiveness of synthesized speech.

**Language-Specific Method:** We addressed the need for language-specific models by adapting our intonation model for Kazakh, as described by B. Wang et al. [Wang, 2015]. Generic models frequently fall short of capturing the subtleties of individual languages, making language-specific modeling essential.

**Linguistic Preference:** Our work contributes to linguistic preservation by advancing speech technology for less widely spoken languages such as Kazakh. This is consistent with M. Adler and M. Sumner's [Adler, 2018] goals of improving and promoting indigenous languages using technology.

**Increased Accessibility:** The development of a Kazakh intonation model promotes inclusion by allowing Kazakh-speaking communities to access and interact with modern technological advances. This reflects the broader implications highlighted in M. Adler and M. Sumner's work [Adler, 2018], where technology plays a critical role in bridging linguistic differences.

### **Challenges and Future Directions:**

**Data Diversity:** Collecting a sufficient quantity of dataset for Kazakh intonation modeling continues to be a challenge. Future efforts should be focused on expanding data sources, especially for under-represented dialects and speakers of different ages and genders.

**Constraints on Resources:** The development of robust intonation models necessitates large computational resources. Future research should investigate optimizations for efficient training in order to make the technology more accessible.

**Cross-Linguistic Adaptation:** Due to the unique prosodic peculiar-

ities of each language, adapting intonation models across languages poses challenges. Future plans include cross-linguistic transfer learning and the development of universal intonation models.

Incorporating emotion recognition into intonation models for more expressive speech synthesis is a potential avenue. For enriched speech synthesis, research should delve into emotion-aware intonation modeling.

Extending intonation models for real-time applications such as voice assistants necessitates low-latency solutions. Addressing these limits is of the utmost importance for improving the experience for users.

Collaborations with linguists and language experts can further enhance intonation models, providing accurate representation of linguistic nuances specific to the Kazakh language.

End-User Feedback: Collecting feedback from end-users and incorporating their preferences into intonation models can lead to more user-centric speech synthesis systems.

Multimodal Synthesis: Integrating intonation models with facial movement and gesture synthesis can create more immersive and natural communication interfaces.

### **Conclusion**

In conclusion, our data gathering and processing efforts have laid a solid foundation for the development of the Kazakh intonation model. The diverse corpus, aligned audio-text data, and painstaking preprocessing ensure that our model is well-equipped to capture the nuances of Kazakh intonation. Moving forward, the use of advanced machine learning techniques will be critical in achieving our objective of creating expressive and natural-sounding synthesized speech in Kazakh. This research is a big step toward improving speech technology for Kazakh speakers while also encouraging linguistic diversity and inclusivity.

### REFERENCES

1. Zen, H., & Nose, T. (2008). HMM-based expressive speech synthesis. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 3905–3908.
2. Wang, B., Lu, Y., & Li, H. (2015). Language-specific intonation modeling for Mandarin-English code-switch speech synthesis. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 4989–4993.

3. Adler, M., & Sumner, M. (2018). Building ASR and TTS systems for the indigenous African language Xitsonga. Proceedings of the Eleventh International Conference on Language Resources and Evaluation, 163–168.
4. Avesani, C., et al. (2017). MINT: The Multilingual INTonation corpus. Proceedings of the 8th International Conference on Speech Prosody, 808–812.
5. Beisembayev, R., et al. (2013). Kazakh National Corpus: Creation and research applications. Language Resources and Evaluation, 47(4), 1107–1125.
6. Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. Proceedings of Interspeech, 1749–1752.
7. Eyben, F., et al. (2010). Towards the extraction of expressive dimensions in the singing voice. Proceedings of the International Society for Music Information Retrieval Conference, 599–604.
8. Zampieri, M., et al. (2017). Findings of the 2017 Conference on Machine Translation (WMT17): ACL 2017. Proceedings of the Second Conference on Machine Translation, Volume 2: Shared Task Papers, 169–214.
9. Ko, T., et al. (2015). A study on data augmentation of reverberant speech for robust speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(11), 1939–1949.
10. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735–1780.
11. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
12. Vaswani, A., et al. (2017). Attention is all you need. Advances in neural information processing systems, 30.
13. Wang, B., Lu, Y., & Li, H. (2015). Language-specific intonation modeling for Mandarin-English code-switch speech synthesis. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 4989–4993.

**This research has been/was/is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No.AP19678613)**



УДК 811.11.112.

## НЕКОТОРЫЕ СООБРАЖЕНИЯ О НЕВЕРБАЛЬНЫХ КОММУНИКАЦИЯХ

*Г. Х. Хасанова*

*Самаркандский государственный университет  
ветеринарной медицины, Самарканд, Узбекистан  
gulruh\_88@mail.ru*

Отношения, возникающие в рамках общения, различны и имеют определенное информационное наполнение. Невербальное общение – это вид невербального общения между говорящим и слушающим, который используется для дополнения вербальных средств, бывает произвольным и непроизвольным. Невербальные средства являются вспомогательными к вербальным средствам, повышающими эффективность речи. Во многих случаях при обмене мыслями людей вместе с языковыми единицами используются различные невербальные средства, то есть движения тела, мимика, взгляд глаз, движения рук и тон голоса. Каждый из этих инструментов подчинен прагматической цели в речевой ситуации. Значения невербальных средств передачи сообщения лицу, к которому обращена речь, участвуют только в той речевой ситуации, имеющей одно значение. Его функционально-прагматическая функция может меняться в зависимости от речевого процесса. Невербальное общение используется для передачи информации о расовых и социальных характеристиках человека, его психическом, физическом и эмоциональном состоянии, его отношении к конкретной ситуации или конкретному человеку или объекту, а также о психологическом климате в общении. Контекстная ситуация важна при использовании жестов в соответствии с речевыми средствами. Все участники речи должны быть осведомлены о контекстуальной ситуации для правильного понимания предложения, выраженного через предложение и сопровождающие его жесты. В частности, если говорящий использует в процессе общения невербальные средства в соответствии с вербальными средствами, речь говорящего будет эффективной и осмысленной с помощью невербальных средств.

**Ключевые слова:** невербальные средства, телодвижения, жесты, языковые единицы, сообщение, зрительный канал, общение, лингвистическое, паралингвистическое воздействие.

## SOME CONSIDERATIONS ON NONVERBAL COMMUNICATION

*Khasanova Gulrukh*

*Samarkand State University of Veterinary Medicine  
Samarkand, Uzbekistan  
gulruh\_88@mail.ru*

Relations that arise within the framework of communication are different and have a certain informational content. Non-verbal communication is a type of

non-verbal communication between the speaker and the listener, which is used to supplement the verbal means, it is voluntary and involuntary. Nonverbal means are auxiliary means to verbal means that increase the effectiveness of speech. In many cases, in the exchange of people's thoughts, various nonverbal means are used together with linguistic units, that is, body movements, facial expressions, eye gaze, hand movements and tone of voice is used. Each of these tools is subordinated to a pragmatic goal in a speech situation. The meanings of the non-verbal means of conveying a message to the person to whom the speech is directed participate only in that speech situation with one meaning. Its functional pragmatic function can change depending on the speech process. Non-verbal communication is used to convey information about a person's racial and social characteristics, their mental, physical and emotional state, their attitude to a specific situation or a specific person or object, as well as the psychological climate in the community. The contextual situation is important in the use of gestures in accordance with verbal means. All the participants of the speech should be aware of the contextual situation for the correct understanding of the proposition expressed through the sentence and the accompanying gestures. In particular, if the speaker uses non-verbal means in the process of communication in accordance with verbal means, the speech of the speaker will be effective and meaningful with the help of non-verbal means.

**Key words:** non-verbal means, body movements, gestures, language units, message, visual channel, communication, linguistic, paralinguistic influence.

There are two main components that make up communication: verbal and non-verbal. According to the nature of the tools used, various methods of information transfer can be divided into two groups: verbal (verbal) and non-verbal (nonverbal). In the first case, the message is transmitted through language units through auditory or visual channels. In the second case, information is transmitted using paralinguistic means - non-linguistic units "included in the voice message and conveying semantic information"[ Хлыстова, 2005, с. 151]. Verbal means are very important in the transmission of cognitive information. In linguistics, various terms such as "non-verbal means", "extralinguistic means", "paralinguistic means" are used for the participation of gestures, hands, eyes, eyebrows, and body movements without the participation of words in human communication. In linguistics, extralinguistics refers to the addition of pauses to speech, as well as various mental states of a person, such as crying, coughing, laughing, sighing[15]. Paralinguistic influence is the surrounding factors that disturb the speech, strengthen it or weaken it. These include high or low pitch, articulation, sounds, pauses, stuttering, coughing, tongue movements, and exclamations[14].

It is known that language performs three important tasks in the process of communication - communication, message, influence (V. Vino-

gradov). It is natural that language strives for brevity in the process of communication. Based on this nature of the language, there is a need to use additional tools, i.e. non-verbal tools, in the process of communication. For example, to indicate the size, volume, length and other external signs of an object, a person often uses gestures in verbal communication. As a rule, in this case, the gesture is combined with verbal means and acts as a visual accompaniment, confirming or clarifying the relevant content of the speech segment [Хлыстова, 2005, с. 151]:

– “*Shaftoli olib keling, shaftolidan shuncha olib keling, – Robiya nozik barmoqchalarini yozib ko‘rsatdi*”.

– *Xo‘p, shuncha olib kelaman, – deb ikki qo‘limning barmoqlarini tengdan yozib ko‘rsatgan edim, Robiya xuddi to‘yib shaftoli yegandek o‘zida yo‘q sevinib ketdi, ko‘zlarini g‘alati suzib opasiga maqtanib ham qo‘ydi (X.To‘xtaboyev, “Besh bolali yigitcha”).*

*“Bring peaches, bring as many peaches as you can,” said Rabiya, indicating with her delicate fingers.*

*“Okay, I’ll bring this much,” I said, showing the fingers of my two hands, Rabiya was overjoyed as if she had eaten enough peaches, her eyes were strangely floating and she even boasted to her sister (Kh. Tokhtaboyev, “Young man with five children”).*

As can be seen in the above example, the participants of the dialogue can determine the amount of what they are expressing using linguistic means (bring as many peaches as possible) and visual hand gestures (*he showed his thin fingers, I showed the fingers of both hands equally*) will give. Also, the addressee is sending information (also boasting) to the addressee (illocutive) through his eyes (swimming his eyes strangely) without using verbal means. This shows that it is convenient to use non-verbal means in the speech process.

According to Dixon and Hargy, we use non-verbal means in the following situations [Dickson, 2003, p.50]: 1) to replace verbal communication in situations where it is inconvenient or impossible to speak; 2) to complete verbal communication and transfer information; 3) to change the spoken word; 4) voluntarily or involuntarily oppose the expressed opinion; 5) when expressing feelings and interpersonal relationships; 6) regulating the conversation by correcting mistakes in speech; 7) in relationships such as dominance, control and liking; 8) when determining personal and social status through clothing and appearance; 9) in contextualizing the interaction by creating a certain social environment.

The importance of non-verbal means in the communication process is important, among other things, according to B. Akhmedov, paralin-

guistic means (non-verbal means) smooth out linguistic units that are not necessary for verbal communication, which seem methodologically stupid, that is, speech expression with the help of necessary gestures will be labeled. In this case, the verbal speech unit (act) involved in a specific speech communication - a word, a phrase, a sentence, sometimes a sentence - is identified with gestures that can be replaced in this situation [Axmedov, 2019, p.79]. In addition to being an auxiliary tool for speech in expressing thoughts, gestures are a clear evidence of the unique mentality of the Uzbek people. In this respect, the consistent study of Uzbek paralinguistic tools is gaining relevance.

M. Saidkhanov said that in the course of communication, a person, who is a whole biological organism, uses various ways and methods to convey his thoughts to the person to whom he is speaking. As a result, along with verbal means, mimic non-verbal means are also used in communication [Saidkhanov, 2020, p.238].

Scientists have long been interested in the role of non-verbal means in human speech communication. 20 centuries ago, Cicero studied the role of gestures in the communication process and taught speakers to use gestures correctly. Also, the first dictionary of gestures, especially hand movements, was compiled by the Roman orator Quantilion [Arslanov, 2019, p. 50].

A. Nurmonov was one of the first in Uzbek linguistics to conduct research in the field of paralinguistics [Nurmonov, 1980]. Scientist Uzbek paralinguistic tools, paralinguistic signs and language structure, the relationship between linguistic and paralinguistic signs, the origin of gestures, gestures and gesture verbs, paralinguistic tools of the Uzbek language that express negation, the need to use gestures in speech was a study of issues. M. Saidkhanov revealed the essence of non-verbal means on the basis of Uzbek language texts and worked in the somatic aspect [Saidkhanov, 1993].

B. Akhmedov reveals the gender aspect of paralinguistic tools in his research [Axmedov, 2021]. In her monograph "Olfactory Linguistics", M. Burkhonova reveals the role of non-verbal semiotic tools in the communication process, the linguopoetics of non-verbal tools in the creolizing text, the representation and linguistic expression of concepts related to the olfactory system, deictic and linguopoetic features [Burxonova, 2022].

Also, in addition to separate monographic works on non-verbal means in Uzbek linguistics, it has also been partially discussed in the scientific researches of linguists. In particular, the linguist Kh. Ismailov,

in the second chapter and third part of his doctoral dissertation entitled “Sociolinguistic and Psycholinguistic Aspects of Forensic Linguistics”, showed the importance of non-verbal means in the court process, and showed the importance of facial expressions of defendants during their participation in the court session, in particular, when they spoke. observes that observation is essential in determining the truth [Ismoilov, 2021, b. 17-18]. Also, Z.Karimova in her scientific work “Sociopragmatic features of utterances in Uzbek and English languages” describes the use of non-verbal means in speech along with words and the effect of these means on the speaker’s age, gender, social status, even, showed that it can be differentiated according to the region in the Uzbek and English languages [ Karimova, 2021, b. 23–24].

I. Aslanov, in the third chapter of the methodical manual entitled “Psychology of social activity and behavior”, showed the types of non-verbal means and their importance in interpersonal relations [Aslanov, 2019, b. 50].

It should also be noted that people believe that non-verbal units are not deceiving, moreover, they trust non-verbal units more than verbal ones. Non-verbal means communicate the real state of mind of a person. It has been tested in practice, for example, when a person who is angry, disliked or upset speaks with a frown, it is more effective:

– *Qalay, bolalar bilan tanishib oldilaringmi? – Sekingina so‘radi O‘ris xola.*

– *Hammasi odam yovvoyiga o‘xshaydi-ku, – nordon narsa chaynab olgandek aftini jiyirib dedi Qoravoy.*

– *Hechqisi yo‘q, keyin oralaringdan qil o‘tmaydigan o‘rtoq ham bo‘lib ketasizlar (X. To‘xtaboyev, “Besh bolali yigitcha”).*

*Have you met the children? – Aunt Russian asked slowly.*

*“Everything looks like a wild man,” said Karavoy, **snorting** as if he had chewed something sour.*

– *It’s okay, then you will become inseparable friends (Kh. Tokhtaboyev, “A boy with five children”).*

In every communication, the speaker aims to influence the person to whom the speech is directed. Non-verbal means further enhance this effect. Most people are more likely to deceive others with their words than with their actions. Verbal units are easy to control, but body movements, facial expressions, and tone of voice are difficult to control. If such non-verbal means are given importance, it is possible to avoid deception or understand the sincerity of the speaker. If the speaker contrasts the non-verbal units while speaking, the listener will focus on the non-verbal units. *For example,*

– *Baqqol, guruchni qayerdan oldingiz? – deb so‘radi.*

– *Keliningizga tug‘dirib kelyapman, – rangi allanechuk oqarib, bit ko‘zlari pirpirab ketdi baqqolning.*

*Baqqolning g‘alati bo‘lib ketganini sezgan Parpi buvam chap qo‘li- ni marzaga tirab sekin o‘rnidan turdi :*

– *Bu zormanda objuvozniki-ku!*

– *Yo‘g‘-e, mozori sharif ursin.*

– *Bu zormanda kolxozning guruchi-ku!*

– *Amaki, tepamda xudo bor, og‘zingizga qarab gapiring (X. To‘xtaboyev, “Besh bolali yigitcha”).*

– *Grocer, where did you get the rice? he asked.*

*“I’m giving birth to your daughter-in-law,” said the grocer, his face turned pale and his eyes twinkled.*

*Sensing that the grocer had become strange, Grandmother Parpi stood up slowly, putting her left hand on the floor:*

*“It’s the objuvoz’s!”*

– *No, let Sharif dig the grave.*

– *This is the rice of the collective farm!*

– *Uncle, there is a god above me, speak with your mouth open (H. Tokhtabayev, “A boy with five children”).*

In this dialogue, the speaker (the grocer) revealed the truth to the listener (Grandma Parpi) with a change in his face. Although he tried to convince the audience with words, he expressed his inner secret with his facial expressions. If the listener correctly understands the speaker’s non-verbal information, he will learn his inner attitude and purpose and respond appropriately.

Also, F. Delsarte observed the organic connection between body movement and character (“Harmonic Gymnastic and Pantomimic Expression”, 1895) and noted as follows: “There is nothing better than a gesture without meaning and reason the thing is not sad. Gesture is more than verbal communication, it is a way of conveying information without words. Verbal speech is weaker than gestures, since gestures are an agent of the heart, a persuasive tool. In fact, non-verbal means have a stronger effect than verbal means in a speech situation. We can know the inner feelings of the speaker and the listener through gestures and tone of voice. We usually plan our words, but in non-verbal communication we convey information unconsciously. We do not deliberately raise an eyebrow or blush, but these situations happen involuntarily, naturally:

...– *Choyxonadagilarga borib aytsam, buning janozasiga chiqish shart emas, deyishyapti.*

– *Nega shart bo‘lmas ekan? – Bu gap Akbar domlaga yoqmay, qoshlarini chimirdi. – Kim aytdi, kallasi joyidami?*

– *Hasan chillak aytdi, Komil bangi aytdi. Mahallaga aralashmagan odamning janozasiga ham, to‘yiga ham chiqilmaydi, deyishyapti (T.Malik, “Halovat”). ... - When I go to the people in the teahouse, they say that it is not necessary to go to the funeral.*

*“Why not?” - Teacher Akbar did not like this and frowned. - Who said, is the head okay?*

– *Hasan said chillak, Kamil said bangi. They say that people who are not involved in the neighborhood cannot go to funerals or weddings (T. Malik, “Halovat”).*

In this dialogue, we can see that the listener (Akbar Domla) involuntarily shows his inner displeasure with the message the speaker is conveying with an eyebrow movement.

It can be noted that people use non-verbal communication for the following reasons:

1. Non-verbal communication is in some cases more effective than verbal means, in particular, words can sometimes have limitations (when explaining the form, directions, the speaker expresses his thoughts with more non-verbal means).

2. Non-verbal means have a strong influence: non-verbal means, first of all, express the inner feelings of a person (verbal messages are mainly related to the outside world).

3. Nonverbal means are more difficult to control than verbal means, and they convey more real information to the listener.

4. Non-verbal means can be used for situations where the use of verbal means is inappropriate (when verbal speech is limited): sometimes, when speaking is limited due to social etiquette, non-verbal means can convey a message.

5. Non-verbal means are necessary to help send complex messages: the subject of the speech can convey an illocutionary expression to the addressee of the speech by using simple non-verbal means simultaneously with a very complex verbal message.

Linguist N. Mahmudov writes about the connection of non-verbal means with the speech situation, the relationship between the speaker and the listener in speech communication: “Usually, there are three main elements of the communicative situation, that is, the speaker, the listener and the topic or information . In order to convey certain infor-

mation to the listener, the speaker chooses a medium - an appropriate channel. Naturally, the main channel is the language itself. However, other channels will be launched in accordance with the general situation and purpose for full information delivery. Paralinguistic and extralinguistic tools are meant here. Indeed, various factors, such as various gestures, facial expressions, head nods, body movements, proximity of space, nature of voice, clothes, social or other status of the speaker and the listener have a special value in the communication process. The communication channel is selected in accordance with the content, purpose and nature of the information to be transmitted" [Maxmudov, 2007, b. 40].

In conclusion, the purpose of communication is to influence the person to whom the speech is directed. In this process, the influence of non-verbal means may be stronger than verbal means, and non-verbal communication may be more effective than verbal communication (non-verbal means are involved as additional means when speakers express the shape or directions of something). Non-verbal means can be conditionally divided into two groups: companions of linguistic means and groups such as specific manifestations of linguistic means. It is known that non-verbal means can convey the message without verbal means through direct visual-signal representations of linguistic means, but since it is not possible to directly see non-verbal action in written texts, it is possible to get information about non-verbal with the help of verbal means.

#### REFERENCES

1. Abjalova M. O‘zbek tili o‘zlashma so‘zlarining urg‘uli lug‘ati [Matn]: o‘quv-uslubiy lug‘at / M.Qurbonova, M.Abjalova, N.Axmedova, R.To‘laboyeva. – Toshkent: Nodirabegim, 2021. – 988 b. ISBN 978-9943-6940-9-5
2. Aslonov I.N. Ijtimoiy faoliyat va muomila psixologiyasi. Metodik qo‘llanma. – Toshkent, 2019. – 105 b.
3. Axmedov B. Paralingvistik vositalarning genderologik va pragmatik tadqiqi. Filol.fan.bo‘yicha falsafa d-ri (PhD)... diss. – Andijon, 2021. – 145 b.
4. Axmedov B.R. Nutqiy muloqotda paralingvistik vositalarning o‘rni // O‘zbekistonda xorijiy tillar.journal.fledu.uz. 2019. – 269 b.
5. Burxanova M. Olfaktor lingvistika. Monografiya. – Farg‘ona, 2022. – 126 b.
6. Dickson David & Hargie Owen. Skilled interpersonal communication: research, theory and practice, Routledge. – London, 2003. – 50 p.



7. Ismoilov X. Sud lingvistikasining sosiolingvistik va psixolingvistik aspektlari: Filologiya fanlari bo'yicha falsafa doktori (phd) dissertasiyasi avtoreferati. – Andijon, 2021. – 151 b.

8. Karimova Z.G'. O'zbek va ingliz tillarida so'z-gaplarning sosiopragmatik xususiyatlari: Filologiya fanlari bo'yicha falsafa doktori (phd) dissertasiyasi avtoreferati. – Toshkent, 2021. – 60 b.

9. Mahmudov N. O'qituvchi nutqi madaniyati. – Toshkent, 2007. – 40 b.

10. Nurmonov A. O'zbek tilining paralingvistik vositalari haqida. – Andijon, 1980. – 42 b.

11. Saidxonov M. Noverbal vositalar va ularning o'zbek tilida ifodalanishi. Filol.fan.nom...diss. avtoreferat. – Toshkent, 1993. –24 b.

12. Saidxonov M. Mimik noverbal vositalar. Nutq madaniyati va o'zbek tilshunosligining dolzarb muammolari // xalqaro ilmiy-amaliy konferensiya materiallari. – Andijon, 2020-yil, 4-may. – 238 b.

13. Хлыстова, Вероника Геннадьевна. Функционально-структурная и семантическая характеристика кинематических речений, отражающих коммуникативный аспект кинесики: На материале английского языка: диссертация ... кандидата филологических наук : 10.02.04. – Нижний Новгород, 2005. – 151 с.

14. <http://hozir.org/zbekiston-respublikasi-olij-va-rta-mahsus-talim-vazirligi-tosh-v142.html?page=7>.

15. <https://aim.uz/referaty/59-psikhologiya/19679-mulo-otning-verbal-noverbal-paralingvistik-ta-sir-vositalari.html>.

УДК 81'286

**ДИАЛЕКОМЕТРИЯ И АЗЕРБАЙДЖАНСКИЙ ЯЗЫК:  
ПРОБЛЕМЫ, РЕШЕНИЯ И ПЕРСПЕКТИВЫ***Афруз Гурбанова, Мехрибан Багирова**Институт Информационных Технологии, Баку, Азербайджан  
afruz1961@gmail.com, mehriban.amea@gmail.com*

Диалектология – это изучение диалектов, а диалектометрия - измерение диалектных вариаций, т. е. языковых различий, распространение которых определяется прежде всего географией. Новые источники информации и аналитическое программное обеспечение расширяют сферу применения диалектометрии. В исследовательской работе были проанализированы диалектометрические методы. Исследованы возможности использования диалектометрических методов в Азербайджане и даны рекомендации по направлению их применения.

**Ключевые слова:** диалектология; диалектометрия; диалектологический атлас; диалектная вариация; лингвистическая география.

**DIALECTOMETRY AND THE AZERBAIJANI LANGUAGE:  
PROBLEMS, SOLUTIONS AND PERSPECTIVES***Afruz Gurbanova, Mehriban Bagirova**Institute of Information Technology, Baku, Azerbaijan  
afruz1961@gmail.com, mehriban.amea@gmail.com*

Dialectology is the study of dialects, while dialectometry is the measurement of dialect differences, i.e. linguistic differences, the distribution of which is determined primarily by geography. New sources of information and analytical software are expanding the scope of dialectometry. Dialectometric methods were analyzed in the research work. The possibilities of using dialectometric methods in Azerbaijan were investigated and recommendations were given in the direction of their application.

**Key words:** dialectology; dialectometry; dialectological atlas; dialect variation; linguistic geography.

**Introduction**

The changes occurring in society in modern times, strengthening the integration of scientific fields necessitate a new approach to dialectological research.

The relevance of studying dialects is determined by the following features:

- dialects serve to understand the process of the historical development of a language: dialects often retain archaisms necessary to recreate a broad language movement;
- dialects serve to establish a mutual relationship between the history of the language and the history of the people, as dialect facts allow us to trace how tribes and peoples migrated in ancient times;
- dialects serve to understand the variety of words, sounds and forms of the modern language, to practically consider the features of local speech.

The largest areas of study of dialects are dialect lexicography and linguistic geography [Levina, 2016]. The oldest branch of dialectology, today called “dialect geography”, studies the geographical distribution of language varieties.

Methodological innovation in the field of linguistic geography is related to the development of computer programs that allow direct analysis of large volumes of data and graphical visualization in a simple way, starting from the 1980s.

The systematic study of the dialects of a language by the linguistic-geographical method began in Europe in the second half of the 19th century. This method consists of collecting data from a large number of settlements through a single questionnaire. Answers to each question of the questionnaire form the basis of a dialect map reflecting their territorial distribution, and a set of dialect maps compiled for the same settlements is combined in a dialectological atlas [Arkhangelsky, 2021]. The first such atlas was G. Wenker’s atlas of German dialects (1876–1881) [Herrgen, 2010]. A French language atlas and an atlas of other European languages were published by J. Gillieron (1902–1910) [Goebel, 2010]. All major European languages now have language atlases, including hundreds or thousands of maps. The Linguistic Atlas of Europe, published since 1975, is a project that collects data on many languages. This atlas is the largest scientific project for the study of languages using the linguistic-geographical method. The Atlas covers six language families present on the European continent: Altaic, Basque, Indo-European, Caucasian, Semitic and Uralic. These families are divided into 22 language groups comprising 90 languages and dialects [Viereck, 2006].

Today, the rapid development of mass media and new technologies has greatly impacted local languages and dialects, putting them in danger of extinction [Gurbanova, 2023]. In order to preserve local languages and dialects, it is necessary to carry out dialectometric stud-

ies and to determine the dialect variations of the language, to assess the similarities and differences between them.

The necessity of a deeper study of the dialect differences of national languages by using mathematical methods, especially dialectometric research, in the current era of rapid globalization in the research work, was justified, and proposals were developed in this direction.

## **2. Related work**

In [Mehrabani & Hansen, 2015], the main differences between dialects or closely related languages are explored based on the available speech data of those dialects/languages. A method is proposed to measure spectral acoustic differences between dialects based on the analysis of volumetric space within a 3D model using log probability distributions derived from traditional cepstral Mel Frequency Coefficient and Gaussian Mixture Models. The proposed dialect affinity measures are evaluated and compared on a corpus of Arabic dialects as well as a corpus of closely related South Indian languages.

In [Goebel, 2010], documented the taxometric and cartographic achievements of the Salzburg school of dialectometry. Problems related to Romance dialectology and Romance language geography, historical linguistics, numerical classification, statistics and statistical cartography were investigated. Issues such as measuring linguistic atlas data, creating a data matrix, choosing a similarity index, creating appropriate similarity and distance matrices, similarity maps, parameter maps and cartographic email of dendrograms were analyzed and their visualization was carried out using “Visual Dialectometry” software.

In [Donoso & Sánchez, 2017], an information-theoretical approach to geographic language variation is proposed using a corpus based on Twitter. Dialectometric measurements (cosine similarity and Jensen-Shannon difference) are used to quantify the linguistic distance between the hollows created in a single grid (set) on the map. The authors believe that social networks can be used qualitatively for dialectometric analysis.

In [Asadpour, 2011], the method of measuring the cumulative degree of lexical, phonological, morphological and syntactic differences between the dialect variations of the Azerbaijani language is proposed. Using hierarchical cluster analysis, dialect distances are analyzed, the benefits of applying the methods developed and tested in Turkic languages to the Azerbaijani language are shown.

### 3. Dialectology

A dialect is a regionally or socially distinctive variety of a language defined by a particular set of words and grammatical structures. Spoken dialects are also usually associated with a different pronunciation or accent.

A dialect is a variety of language used in a particular geographical location. So dialects are changed and influenced by a group of people who use it. Other social factors such as class, occupation, and age can also create and influence dialects.

The aims and methods of dialect geography are as follows:

- Pure form – Dialect geography studies the relationship between language and geography, identifying the local dialect, especially in villages, where the dialect is in its purest form before being polluted, weakened and completely lost. He looks for the most “original” and the most “typical” form of speech used in a certain field.

- Non-educated Old Rural Males – NORM – The purest form of a dialect is mostly taken from old rural males.

- Rural area – In order to obtain the purest form of the dialect, teaching should be concentrated in villages where the language is less “contaminated” by foreign elements.

- Raw data – Data collected for research is presented in raw form.

- Linguistic Mapping – After the interviews are completed, data is collected, responses are tabulated, and linguistic maps are constructed to show dialect variations.

This method of studying dialects is also known as traditional dialectology (Chapter 2: Literature Review).

Dialects of a particular language keep the history, culture, ethnography and folklore of the people alive in the language of their great-grandparents, while preserving the intricacies of the language and conveying it to future generations. Therefore, the study of dialects and dialects of the language is one of the most important issues of linguistics. The collection and study of the materials of each dialect or dialect plays an important role in the study of the language and history of the people.

Dialect differences are the main research object of dialectology. Dialectology is the branch of linguistics that studies regional dialects, dialect differences, and dialect language in its present state and history.

In traditional dialectology, sources of information are dialects, dictionaries, dialect atlases, and other materials.

As in the case of other languages, further strengthening of the

state care for the Azerbaijani language has opened wide opportunities for the development of various fields of Azerbaijani linguistics. The inclusion of the issue of “Ensuring the study of various dialects and dialects of the Azerbaijani language in accordance with the requirements of the modern era” into the State Program “On the use of the Azerbaijani language in accordance with the requirements of the time and the development of linguistics in the country” put the study of the dialects of the Azerbaijani language as an important task [Decree of the President of the Republic of Azerbaijan, 2013]. In the direction of the implementation of the program, “Nakhchivan Dialectological Atlas of the Azerbaijani Language”, “Karabakh and Eastern Zangezur Dialectological Atlas of the Azerbaijani Language”, “Nakhchivan Dialectological Dictionary of the Azerbaijani Language”, “Karabakh Dialectological Atlas of the Azerbaijani Language”, “Karabakh Dialects of the Azerbaijani Language” monograph and “Karabakh Dialect of the Azerbaijani Language” dialectological dictionary” was published.

“Nakhchivan dialectological atlas of the Azerbaijani language” contains 278 maps covering about 1000 settlements of the Nakhchivan Autonomous Republic, and a CD with the voices of the informants during the expedition. In the atlas, the characteristics of dialects that create isogloss are given with special signs, and their distribution area is determined.

In the “Karabakh and Eastern Zangezur dialectological atlas of the Azerbaijani language”, 692 of those settlements were coded and reflected in 278 maps. In each map of the atlas, the specific characteristics of the dialect words of the region, changes, usage forms and variants, replacement cases, and synonyms of the words used in the settlements are also given.

The Azerbaijani language has a rich dialect system. Dialects and dialects of our language have been studied from different aspects and grouped according to their geographical location and level characteristics. Dialects and dialects of the Azerbaijani language are divided into 4 groups according to the principle of geographical area:

- Eastern group dialects and dialects of the Azerbaijani language – Guba, Baku, Shamakhi dialects and Mughan, Lankaran dialects;
- Western dialects and dialects of the Azerbaijani language – Kazakh, Karabakh, Ganja dialects and Ayrym dialects;
- Dialects and dialects of the northern group of the Azerbaijani language – Sheki dialect and Zagatala-Gakh dialects;
- Southern dialects and dialects of the Azerbaijani language –

Nakhchivan, Ordubad, Tabriz dialects and Iravan dialect. [Shiraliyev, 2008].

An in-depth study of dialects is necessary for proper assessment of variability at all levels of language structure, development of optimal grammatical orthoepic and other norms. The main differences between the dialects are investigated based on the available speech data of those dialects.

From sociolinguistic methods in the study of dialects: long-term panel observation, various types of speech writing, diaries, questionnaires, interviews, surveys, secret tests, etc. It is used.

Dialects are the unwritten form of language, dialectologists use the questionnaire method and direct observation method to study them.

The collection of information about the linguistic features of the dialect by the questionnaire method is carried out by receiving written answers to the questions of a specially designed questionnaire from linguistically competent persons (teachers, rural intellectuals, etc.).

One of the methods of studying modern dialects is direct observation, when the researcher identifies linguistic features while listening to live dialect speech. With the method of direct observation, dialectologists record the live speech of dialect speakers on the basis of a pre-designed questionnaire program. In order to determine which generations and speakers have more dialects, it is necessary to measure dialectics. Using special computer programs, it is possible to obtain more information than tape recordings, which allow you to repeatedly repeat and analyze speech segments. A fund of such writings will preserve modern dialect speech for future researchers [Methods for studying dialects, 2015].

#### **4. Dialectometry**

Obtaining an appropriate measure of the distance between two pronunciations is important not only for dialectologists interested in establishing relationships between different dialects, but also for sociolinguists studying the effects of boundaries on spoken language.

The presence of a measure of distance between word pronunciations necessitates quantitative analysis to investigate geographic and sociolinguistic factors [Wieling et al., 2014].

In the second half of the 20th century, a large number of collected dialectological materials and the development of information technologies led to the emergence of a new approach to the study of dialects – dialectometry. Dialectometry studies language variation using

statistical methods. Dialectometric methods allow dialectologists to quantify dialect differences and measure language change based on this [Nerbonne & Kretzschmar, 2006; Pickl & Rumpf, 2012].

Dialectometry is a branch of geolinguistics that deals with the measurement, visualization, and analysis of common dialect similarities or distances depending on the characteristics of a geographic location. Dialectometric studies build computational approaches to identify “general, apparently hidden structures from large numbers of features” and focus on quantitative, cartographic visualization, and exploratory data analysis to extract patterns. Dialectometry provides methods for estimating the linguistic distance between two arbitrary dialects in dialectological atlas projects. The main goal is to determine the degree of closeness of dialects, confirm existing dialect classifications, and solve problems related to dialect division [Vozenilek, 2022].

In addition to dialectological studies, dialectometry has made theoretical contributions to comparative dialectology and the study of dialect distribution.

## **5. Recent advances in dialectometry**

Dialectometric methods are constantly improving, opening new possibilities for explaining linguistic variations:

- focus on identifying the most important (diagnostic) individual language elements that form the basis of the general dialect variation;
- understanding that lexical and social factors can determine geographical variation;
- new methods for evaluating linguistic change in dialects;
- dialectological theory;
- more attention to dialect grammar and morphosyntax;
- to use new data sources in addition to the traditional dialect atlas data;
- to create new (online) software that allows dialect researchers to use dialectometric tools more easily [Wieling, 2015].

The changes occurring within the language are not only related to the geographical distance of the population speaking the dialect variants. Factors such as education, access to audiovisual media, self-affirmation, and cultural expansion can influence language use.

The main concepts of dialectometry are:

- measures of difference or similarity between linguistic varieties for one or more linguistic functions;



- aggregation and clustering algorithms for organizing large data sets by proximity/difference;
- tools to present changes measured across time, space and social groups.

Dialectometric methods began to develop actively in the 1970s and 1980s after Jean Séguy (France) and Hans Goebel (Austria) applied statistical approaches to the study of Romani dialects using previously collected atlas data [Arkhangelsky, 2021]. Goebel's main contribution was the development of various methods for combining individual distances into global distances and global distances into global clusters.

At the beginning of the 21st century, John Nerbonne (American) and Wilbert Heeringa, following the research started by Goebel, developed and tested new analytical methods based on various statistical procedures, in addition, they included a quantitative measure of articulation distance. These studies contributed to mastering the theoretical foundations of dialectometry, expanding its application areas and perspectives.

Dialectometry began to develop further with the application of computer tools for the analysis of geolinguistic data. The most common and used statistical analysis programs can be applied to the study of language variations. Widely used by statisticians and data collectors, the R open-source statistical package is commonly used to obtain figures involving distance measurement, similarities, cluster analysis, and many other complex analyses. R packages such as rMaps make it easy to create and share interactive maps from R.

In recent years, research in the field of dialectometry has used many methods of data visualization that represent the main methods of GIScience (geographical information science). GIS software systems offer many software products that can be used directly to represent geolinguistic data, the results obtained from the statistical analysis of this data in a simple way [Dubert & Sousa, 2016].

Dialectometric methods are based on the concept of “distance” between settlements. By distance here we mean the mathematical function  $d(X, Y)$  that calculates the value based on the atlas data for any two settlements ( $X$  and  $Y$ ), showing how the answers to the survey questions at one point differ from the answers to the questions at another point. In practice, different functions can be applied, but in any case they must meet at least three requirements [Arkhangelsky, 2021]:

- If the answers to all questions of the questionnaire coincide at two points, the distance between them is equal to 0  $d(X, Y) = 0$ ;

- The function is symmetric, i.e.  $d(X, Y) = d(Y, X)$ ;
- The more the questionnaires X and Y differ on a large number of questions, the greater the distance between them.

Dialectometrics is a new direction emerging from classical dialectology, where the differences between different dialects in a region are statistically calculated and presented using dialect maps and atlases.

The origins of classical dialectometry are associated with the search for dialect boundaries. Since dialect boundaries are made up of groups of isoglosses, simply adding a few isoglosses is enough to determine where enough of them are grouped together to form a true dialect boundary. The idea of combining isoglosses to create boundaries of varying “thickness” between points on a map was first implemented by Carl Haag in 1898.

Recent studies have shown that the Hilbert-Schmidt independence test (HSIC) is effective in measuring spatial autocorrelation of different types of linguistic variables. The purpose of the method is to provide statistical evidence of the existence of dialect boundaries. These boundaries are called “dialectons” [Rodriguez-Diaz, 2018].

A linguistic map is a thematic map showing the geographical distribution of speakers of a language or isoglosses of a language family. Linguistic atlases serve as empirical databases documenting in detail the dialect profile of a large number of locations. A variety of well-known numerical classification methodologies are used to summarize and visualize the underlying pattern from the vast amount of data contained in linguistic atlases.

The matrix of double points calculated on the basis of the data of the dialectological atlas can be used in several ways. First, it makes it possible to find important groups of isoglosses, i.e. lines, where the boundary of variants is crossed at the same time in many dialect maps. Secondly, the classification of dialects and dialects can be obtained with the help of the automatic clustering method. In this case, clusters (sub-dialects) will connect points that are slightly different from each other, but significantly different from neighboring clusters [Batagel], 1992].

Since it took many years to create dialect atlases, researchers have also tried to use other sources of information to study dialect variation. Szmrecsanyi called these research approaches “Corpus dialectometry”. When studying pronunciation variation, they compare using the data collected by the compilers of linguistic atlases – the pronunciations of the same word in different places.

Corpus dialectometry is a younger field, where statistical methods are applied to Corpus data rather than dialect maps [Szmrecsanyi, 2011].

The most reliable form of storing dialect texts and the optimal source database is a corpus of electronic texts supported by software. The form of electronic representation of dialect texts increases the preservation of this unique material and provides linguists of various scientific fields with freer access to primary dialect material, allowing them to observe the real relationships between units in the course of dialect speech.

The general principle of forming the text base of the corpus is the principle of full reflection of the features of dialect communication in the corpus. To implement this principle, it is necessary to create several subcorporations. Creating a subcorpus involves including a variety of significant textual materials representing:

- the most important types of dialect speech (everyday speech, folklore, official speech, ritual communication);
- different speech forms (dialogue, polylogue, monologue);
- various topics of village communication;
- social differentiation of dialect speakers (by gender, age, profession, education level).

Assessing the closeness between several dialects of the language is an interesting but complex research topic. This type of assessment shows how often dialects are mixed or differentiated in the given language space. Thus, the data obtained determine dialect differences, isogloss, clusters, etc., and it opens up ample opportunities for visualization.

Dialectometry is a quantitative methodology for calculating linguistic distances between linguistic varieties. The most commonly used methods of dialectometry can be divided into the categories of traditional and computational methods.

The main issue of dialectometric analysis is the acquisition of a modern “map of similarity” of linguistic idioms with each other. In this case, the similarity can arise not only by the relationship of idioms, but also by the recent migration or the general influence of another language. Evaluation can be carried out in the following sections [Mehrabani, 2015]:

- differences in the physical sounding of speech;
- differences in the linguistic expression of speech;
- signs of perception assessment;
- differences of classifiers of automatic speech sounding systems.

There are a number of ways to compare languages, dialects, or other

types of speech. Various string distances such as Levenstein, Euclidean [Jeszczynski, 2017] and Manhattan [Heeringa, 2009] distance are used to account for pronunciation differences between dialects.

Levenstein distance is a measure of the difference between two sequences or strings. Such sequences can be words of the language being studied. When measuring the Levenshtein distance between two pronunciation variants of a word, the minimum number of operations (insertion, deletion and replacement of characters) that one variant must go through in order to transform it into another is calculated [Heeringa, 2004].

The Wagner-Fischer algorithm is a dynamic programming algorithm that measures the Levenstein distance between two character strings.

A dialect continuum is a group of language dialects that vary within an area. On the dialect continuum, the further apart two dialects are, the more different they are, the more difficult they are to understand each other, or not at all. People in close proximity on the dialect continuum can understand each other when they speak.

Multivariate scaling (MS) is a statistical method used to study dialect continua. MS converts complex distance data into interpretable low-dimensional images [Klis, 2020].

Clustering algorithms are applied to classify the studied set of objects by identifying the closest clusters in that set. The source material for the analysis is the matrix of distances between the studied objects, and the result of the algorithm can be presented as a hierarchical structure that shows the sequence of clusters [Galdino, 2019].

The UPGMA (Unweighted Arithmetic Mean Pair Grouping Method) method is one of the simplest and most widely used hierarchical clustering algorithms for creating a dendrogram from a distance matrix. Here, the local topological relationships are obtained in descending order of similarity and the dendrogram is constructed stepwise. That is, the two closest data points are identified first and grouped in the dendrogram. After the first grouping, the two closest data points are treated as one data point (composite) and new distances are calculated using the mean of the distances between the simple data point and the components of the composite data point. Then the next closest data points are added to the dendrogram until all data points are included.

The WPGMA (Weighted Pair Group Method using Arithmetic Averages) algorithm is similar to its unmeasured variant, the UPGMA algorithm. In the WPGMA algorithm, the distance between clusters is calculated as a simple average. WPGMA gives a simple average

weighted result, while in UPGMA it gives a proportional average weightless result [Garcia-Vallvé, 2009].

Hans Gebl and Edgar Heimerl developed special software called “VisualDialectometry (VDM)”.

VDM was developed as a dialectometry project at the University of Salzburg between 1998 and 2000 and implements algorithms that support dialectometric analyses of Dialectological Atlas data. It offers functionality for managing pre-classified Atlas data, various dialectometric approaches to data analysis, and various methods for visualizing the results of such analyzes (dendrograms, diagrams, or maps) is one of the most used tools for dialectometric analysis of various languages of VDM [Galdino, 2019; Goebel, 2006].

Gabmap is a web application for dialectometry and cartography. Allows for comparison and statistical analysis of dialect data. Gabmap is a graphical user interface that performs not only comparison of vocabulary or other categories of information, but also comparison of pronunciation using editing distance. Gabmap allows researchers in dialectology to perform computer-assisted exploration and calculations [Website “Dialektometrie Projekt” – Salzburg; Nerbonne, 2011].

Quantitative analysis methods make it possible to reveal the relationship that exists between the two distributions of data. The method used in this case is called correlation dialectometry. The method allows you to visualize and compare geolinguistic relationships between the distribution of phonetic data and other morphological data, as well as analyze the relationship between linguistic and geographical distances [Montemagni, 2008].

## **6. Conclusions**

Dialectometric methods were analyzed in the research work. It was determined that dialectometric methods are constantly being improved, significant progress has been made in this field. So that:

- Various methods have been developed by specialists in the field of dialectometry to simultaneously analyze the linguistic and social factors behind geographical differences and to assess their relative strength;
- Dialectometry has been greatly improved to assess linguistic changes in dialects;
- Dialectometry uses data sources other than traditional dialect atlases, especially dialect corpora built from online sources, to study dialect variation;

– With the creation of new (online) applications, many dialectologists use dialectometric tools.

The importance of in-depth study and application of dialectometric methods in Azerbaijan was determined. The possibility of using dialectometric methods for calculating the dialect differences of the Azerbaijani language is investigated and the following is recommended:

- It is necessary to create a web portal to enter the data of the “Dialectological Atlas of the Azerbaijani language” into the database and ensure the availability of this data for everyone. The portal will allow the creation of a single dialectal environment of the Azerbaijani language and the operative search of dialects.

- It is possible to use modern tools, including mathematical methods, especially algorithms widely used in dialectometric research - cluster analysis and multidimensional scaling – to analyze the data entered into the database.

- As a part of the ecosystem of the Azerbaijani language on the e-state platform [Alguliyev, 2021], the creation of a single dialect system of the Azerbaijani language will create a technological basis for deeper study, classification and evolution of these dialects.

## REFERENCES

1. Alguliyev R., Yusifov F., Gurbanova A. (2021). Protection of Azerbaijani Language in e-government platform, E-Journal of Linguistics, Vol. 15(2), pp. 155–161. <https://ojs.unud.ac.id/index.php/eol/issue/view/3948>
2. Arkhangelsky, T. (2021). Application of the dialectometric method to the classification of Udmurt dialects. Ural-Altai studies, V 2, 7–20.
3. Asadpour, H. (2011). A Survey of Language Varieties in Azerbaijan-e Qærbi through Dialectometric Analysis. Journal of Persian Academy of Language, Dialectology. V.1, 173–202.
4. Batagelj, V. et al. (1992). Automatic Clustering of Languages. Computational Linguistics, V. 18(2), 339–352.
5. Chapter 2: Literature Review [http://studentsrepo.um.edu.my/5575/3/3.Chapter\\_2.pdf](http://studentsrepo.um.edu.my/5575/3/3.Chapter_2.pdf)
6. Decree of the President of the Republic of Azerbaijan on the approval of the “State Program on the use of the Azerbaijani language in accordance with the requirements of the time and the development of linguistics in the country”, (2013). Baku, <https://e-qanun.az/framework/25537>

7. Donoso, G. & Sánchez, D. (2017). Dialectometric analysis of language variation in Twitter. arXiv preprint arXiv:1702.06777, <http://aclanthology.lst.uni-saarland.de/W17-1202.pdf>
8. Dubert, F. & Sousa, X. (2016). On quantitative geolinguistics: an illustration from Galician dialectology. *Dialectologia: revista electronica*, 191–221.
9. Galdino, S. & Maciel, P. (2019). Weight Pair Group Average Mean Clustering for Interval-valued Data. *IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–7.
10. Garcia-Vallvé, S. & Puigbo, P. (2009). DendroUPGMA: a dendrogram construction utility. *Universitat Rovira i Virgili*, 1–14.
11. Goebel, H. (2006). Recent advances in Salzburg dialectometry. *Literary and linguistic computing*, V. 21(4), 411–435.
12. Goebel, H. (2010). Dialectometry: theoretical pre-requisites, practical problems, and concrete applications (mainly with examples draw from the "Atlas linguistique de la France", 1902-1910). *Dialectologia: revista electrònica*, 63-77.
13. Gurbanova A.M. (2023). Problems and Prospects for Minority Languages in the Age of Industry 4.0. *Lecture Notes on Data Engineering and Communications Technologies (LNDECT) – Springer Publisher*, V. 158, pp. 722-734. [https://link.springer.com/chapter/10.1007/978-3-031-24475-9\\_59](https://link.springer.com/chapter/10.1007/978-3-031-24475-9_59), DOI: 10.1007/978-3-031-24475-9\_59
14. Heeringa, W. (2004). Measuring dialect pronunciation differences using Levenshtein distance. 323 p. <http://www.wjheeringa.nl/thesis/thesis.pdf>.
15. Heeringa, W. et al. (2009). Measuring Norwegian dialect distances using acoustic features. *Speech Communication*, V. 51(2), 167–183.
16. Herrgen, J. (2010). The digital wenker atlas ([www.diwa.info](http://www.diwa.info)): an online research tool for modern dialectology. *Dialectologia: revista electrònica*, 89-95.
17. Jeszenszky, P. et al. (2017). Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German. *Journal of Linguistic Geography*, V. 5(2), 86–108.
18. Levina, M. (2016). Linguistic geography as a basis for areal studies of the Mordovian languages. *Finno-Ugric world*, V. 3 (28), 50–59.
19. Martijn van der Klis & Tellings, J. (2020). Multidimensional scaling and linguistic theory. arXiv, org., <https://www.researchgate>.

net/publication/346857889\_Multidimensional\_scaling\_and\_linguistic\_theory.

20. Mehrabani, M. & Hansen, J. (2015). Automatic analysis of dialect/language sets. *International Journal of Speech Technology*, V. 18, 277–286.

21. Methods for studying dialects, <https://studfile.net/preview/2378303/page:4/>

22. Montemagni, S. (2008). The space of Tuscan dialectal variation: A correlation study. *International Journal of Humanities and Arts Computing*, V. 2(1-2), 135–152.

23. Nerbonne, J. & Kretzschmar, W. (2006). Progress in dialectometry: toward explanation. *Literary and Linguistic Computing*, V. 21(4), 87–397.

24. Nerbonne, J. et al. (2011). Gabmap-a web application for dialectology. *Dialectologia: revista electronica*, 65–89. <https://www.raco.cat/index.php/Dialectologia/article/view/245345>.

25. Pickl, S. & Rumpf, J. (2012). Dialectometric concepts of space: Towards a variant-based dialectometry. *Dialectological and folk dialectological concepts of space: Current methods and perspectives in sociolinguistic research on dialect change*, V. 17, 199–214.

26. Rodriguez-Diaz et al. (2018). Dialectones: Finding Statistically Significant Dialectal Boundaries Using Twitter Data. *Computación y Sistemas*, 22(4), 1213–1222.

27. Shiraliyev, M. (2008). *Basics of Azerbaijani dialectology*. Baku, “East-West”, 416 p.

28. Szmrecsanyi, B. (2011). Corpus-based dialectometry: a methodological sketch. *Corpora*, V. 6(1), 45–76.

29. Viereck, W. (2006). The linguistic and cultural significance of the *Atlas Linguarum Europae*. *Gengojohogaku Kenkyuhokoku (Memoir for Linguistic Informatics)*, V. 9, 58–80.

30. Vozenilek, V. et al. (2022). Mapping, synthesis and visualization of Czech dialects. *International Journal of Cartography*, V. 8(1), 148–163.

31. Website “Dialektometrie Projekt” – Salzburg. URL: <http://www.dialectometry.com/>

32. Wieling M. et al. (2014). A cognitively grounded measure of pronunciation distance. *PloS one.*, 9(1), e75734

33. Wieling M. & Nerbonne, J. (2015). Advances in dialectometry. *Annual Review of Linguistics*, V. 1, 243–264.



УДК 81'322.2

**ПОСЛЕДСТВИЯ ФЕНОМЕНА СИНКРЕТИЗМА  
ДЛЯ СИНСЕТОВ ЛИНГВИСТИЧЕСКИХ ОНТОЛОГИЙ*****М. А. Абжалова***

*Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои, Ташкент, Узбекистан  
abjalova.manzura@gmail.com*

Лингвистическая онтология – это лексическая база данных на естественном языке, которая не только служит набором лексических единиц, но и устанавливает семантические связи между ними. Сперва создаются синонимичные наборы лексических единиц. Вводится толкование слова-синонима в этой наборе, а затем по его наличию определяются его семантические ветви, такие как род (гипероним), тип (гипоним), целостность (холоним), часть (мероним), антоним. Поэтому для лингвистической онтологии и словарей тезауруса важно изучение таких лингвистических явлений, как синонимы, гиперонимы, партонимы, антонимы, омонимы и многозначность. В языке также существует лингвистическое явление, называемое синкретизмом, которое не было глубоко изучено в узбекской лингвистике. В данной статье рассматривается синкретизм и его проявление как лингвистическое явление, его значение в онтологии узбекского языка.

**Ключевые слова:** лингвистическая онтология, тезаурус, синкретизм, семантические отношения, языковое событие.

**IMPLICATIONS OF THE SYNCRETISM PHENOMENON FOR  
LINGUISTIC ONTOLOGY SYNSETS*****Abjalova Manzura***

*Tashkent State University of Uzbek Language and Literature.  
Tashkent, Uzbekistan  
abjalova.manzura@gmail.com*

Linguistic ontology is a lexical database in a natural language that not only serves as a collection of lexical units, but also establishes semantic relationships between them. First, synonymous sets of lexical units are created. An explanation of the synonym word in this collection is entered, and then its semantic branches are determined by its presence, such as gender (hyperonym), type (heponym), wholeness (holonym), part(s) (meronym), antonym. Therefore, for linguistic ontology and thesaurus dictionaries, the study of linguistic phenomena such as synonyms, hyperonyms, partonyms, antonyms, homonyms, and polysemantics is important. There is also a linguistic phenomenon in language called syncretism, which has not been studied in depth in Uzbek linguistics. This article discusses syncretism and its manifestation as a linguistic phenomenon, its importance in the ontology of the Uzbek language.

**Keywords:** linguistic ontology, thesaurus, syncretism, semantic relationship, language event / linguistic phenomenon.

## INTRODUCTION

In recent years, the scope of research on the phenomenon of linguistic syncreticity has been expanding. However, the results regarding syncretism remain controversial. *Syncretism* (Greek. *sygkretizmos* = “union”) 1) adhesion, joining, union; is a characteristic of the initial state in the development of something. This concept is found in almost all fields of knowledge; 2) linguistic syncretism is studied as a linguistic phenomenon. [Abjalova, 2021]. The first use of this word is associated with the ancient Greek historian Plutarch. Later, it was formed as a term and started to be used in various fields. The use of this term in linguistics and the understanding of the essence of syncretism from a linguistic point of view are determined by the fact that in 1943 L. El’mslev studied the issue of syncretism in one paragraph of his work [El’mslev, 1960; 344]. The phenomenon of linguistic syncretism is expressed by V.V. Babajceva in the process of language development, functionally different grammatical categories and forms in one form; It is clearly defined as the combination of differential structural and semantic features of units (some words, meanings, sentences, sentence fragments) in the language system [Babajceva, 2000; 446].

## LITERATURE REVIEW

Syncretism and neutralization in linguistic research [El’mslev, 1960; 343], syncretism and contamination (mixing, hybridity) [Babajceva, 1990; 446], syncretism and pun (pun) [Buzarov, 1996; 24-25] is observed. Also, in some studies, syncretism is a multi-meaning of the word [Eremin, 2001; 74], the manifestation of homonymy and ambiguity is connected with [Babajceva, 1990; Demidova, 2001]. It should be noted that the nature, nature and characteristics of syncretism are studied by many scientists, and syncretism is considered as a linguistic phenomenon. In general, in linguistics, syncretism is a phenomenon of combining several meanings in one form during language development [21]. In some sources, syncretism is called cumulation of grammatical meanings [Plungjan, 2000]. In this case, several grammes belonging to different grammatical categories are represented by one indivisible indicator. For example, the suffix *-a* in the Russian word “*zima*” (winter) cumulatively represents the head agreement and unity. The suffix

“-lar” (-s) in the Uzbek language, in addition to being a plural form, also means respect, type and variety: “*kitoblar*” (books), “*dadamlar*” (dads), “*tuzlar*” (salts), “*olma daraxtlari*” (apple trees). Also, the suffix of departure is a morphological syncretism: it is used to express such meanings as direction, cause, purpose, goal: “*ishga bormoq*” (go to work) (direction), “*o‘qishga berdim*” (gave to study) (cause or goal is spoken it will be clear in the situation), “*ukamga oldim*” (I took (noun) to my brother) (presume).

In some sources, syncretism is considered a factor of linguistic economy and compactness [Buzarov, 1996]. M.A. Pavljukovec states that one of the special cases of implementation of linguistic economy at different levels of the language system is the phenomenon of syncretism [Pavljukovec, 2009; 3]. Based on syncretism, it is considered primary that one form covers several meanings, several functions, categories and several forms [Beresneva, 2011]. It is this aspect of syncretism that motivates its research in combination with the phenomena of polysemy, homonymy, and polyfunctionality.

The phenomenon of syncretism in the grammar of Russian, German, English and other languages was announced by a scientist such as O. Jespersen (1958), J.L. El’mslev (1960), V. Bloch (1966), V. Skaličhka (1967), V.V. Babaitseva (1967, 1973), V.V. Vinogradov (1978), T. Peterson (1988), M. Aronof (1994), J.P. Blevins (1995), A. Calabres (1995) S.N. Daniel (1999), V.V. Buzarov (1998, 2001), Sh. Researched in the works of scholars such as Balli (2001). S. Luraghi (1987), G. Meizer (1992), M. Weller (1993), J. Johnston (1997), V.V. Babajceva (2000), I.V. Visotskaya (2006), B.A. Beresneva (2009) the results of his special research on the essence, nature and typology of syncretism.

While syncretism was initially studied at the lexical and morphological level, many modern studies have studied its features at the syntactic level. For example, B. Milan (1998) and T. Peterson studied the syncretic feature of agreement forms in inflectional languages more deeply and widely, while L.D. Chesnokova (1988), T. Ye. Anoshkina (1981), V.V. Babaitseva (1984, 1997), Z.V. Valjusinskaya (1992), P.V. Chesnokov (1992), L.L. Bezobrazova (1993), N.A. In the works of scientists such as Kobrina (2007), the cases related to the manifestation of syncretism at the syntactic level were studied.

## RESEARCH METHODOLOGY

According to B.A. Beresneva, linguistic syncretism is understood in two senses [3]:

1) syncretism in linguistic forms; 2) the scientific-linguistic concept of syncretism. In it, syncretism is considered and studied as a linguistic phenomenon. But defining its nature and determining its status as a linguistic phenomenon is still controversial.

In the first case, two or more semantic functions are combined in one linguistic form, and it differs from the phenomena of contamination, pun, homonymy and polysemy [Beresneva, 2008; 3].

M.A. Pavljukovec also understands syncretism in two senses [Pavljukovec, 2009; 9]:

1) syncretism, on the one hand, is the combination of two or more meanings in one form, which is reflected in the dominant of these meanings; 2) on the other hand, syncretism is a situation where a specific categorical meaning manifests itself in a specific syntactic situation.

O. I. Prosjannikova in his research emphasizes that syncretism combines polysemy and meaning transfer, homonymy and neutralization phenomena and that this phenomenon can be observed at all levels of the language [Prosjannikova, 2011; 95]. Also, A. M. Shherbak in his research explains syncretism with the phenomenon of homonymy, that is, the meanings belonging to several categories are united in one form, he says, and gives examples from homonyms.[20].

## THE IMPORTANCE OF SPEECH SITUATION IN DETERMINING SYNCRETISM

In the process of communication, participants try to use a minimum number of lexical units, but even so, the speech situation can understand information from even the shortest verbal elements. Naturally, the first replica gives the maximum information about the object of the conversation, and the next replicas are understood in a certain context. Verbal expressions that have become a habit in everyday life can find their formal reflection at the levels of language development. In this case, the “*ko 'k*” – blue (green, blue colors), “*xunuk*” ugly (“badbas-hara → turqi sovuq”, “sovuq” – cold (word with a negative effect), “*toza*” (ozoda, yangi) – clean (neat, fresh), “*yangi*” – fresh (odd, now prepared / now arrived), “*hozir*” – now (ready, now), “*uchun*” – for (reason, to presume, to intend, purpose) lexical units such as can be cited as an example.

## ANALYSIS AND RESULTS

A. Martine emphasizes that syncretism is not a random phenomenon and that it affects the structure of the language. Also, this phenomenon

complicates the functioning of the language, the cause of the confusion is considered to depend on the speech situation, that is, as a result of the economy of human speech, the understood expression appears only in speech situations, says the linguist. In general, this opinion of A. Martine is close to the truth. A simple example question in everyday life is “*Ishlaring qaley?*” (How are you doing in?), The answer “*Dahshat*” (Horror) has become a common one. When linguistic economy is not used in this speech situation, such a question-and-answer situation is restored: “*Ish (ahvol)laring qanday? – Ish (ahvol)larim juda zo ‘r!*” (What is your job? – (My job (condition)) is great!) The response that the situation is very good was given in the reply. But in the speech situation, two linguistic units were saved and one lexical unit was used: “*dahshat*” (horror). When hearing this word, in a person’s physiological state, reactive feelings in the amygdral nuclei of the visual cortex, strong panic, trembling and phobia arise in seconds. Also, the scary reality that he has heard, seen, or read is embodied before a person’s eyes. Therefore, the word “*dahshat*” (horror) is at a higher level in terms of its meaning, and therefore it has become common to use it in a speech situation instead of the lexeme “*zo ‘r*” (excellent), which is superior in terms of its semantics. Since the quality of super-amplification in both word semes is considered an integral seme, a situation of exchange has arisen.

The economist of words A. Qahhor named his story “*Dahshat*” (“Horror”) to make the reader feel the horror of the whole reality in the story, to bring it to his mind with a shudder.

In fact, the word “*dahshat*” (horror) belongs to the noun family and is a synonym of the word “*qo ‘rquv*” (fear), only in pragmatic analysis, more precisely, in a speech situation, the word “*dahshat*” (horror) occurs as a synonym of the word “*zo ‘r*” (great, excellent). As a result, in the **ontology of the Uzbek language**, the synonyms of “*dahshat*” (horror) and “*qo ‘rquv*” (fear) are placed in the first - upper line synsets as form complete synonyms, and the words “*dahshat*” and “*zo ‘r*” in the lower line synsets as the form of a synonymous set. After that, the semantic relations of these synonyms in each set are determined. Then, in the synset belonging to the noun group in the upper row, the semantic relations specific to the synonyms in this set are determined, and in the lower set, the semantic relations of the words belonging to the adjective group are determined.

T.V. Kolesnikova says that syncretism is expressed as a complex of contrasting lexical and / or grammatical meanings, and the universality

of syncretism is determined by its occurrence in different periods of language development and its manifestation in different language levels and different speech styles [Kolesnikova, 2009; 47]. This opinion of T.V. Kolesnikova clarifies the opposite semantics at the heart of syncretism and clarifies the situation of manifestation of syncretism. The conflicting semantics present in syncretism creates enantiosemey.

According to S.L. Charekov, during the development of the language, meanings polarized in one semantic structure arose as a result of the indiscretion of human perception, which later stimulated the development of syncretism [Charekov, 2009; 116]. It is stated in the studies on the occurrence of enantiosemey that in the diachronic aspect, enantiosemey is the result of two opposite semantic syncretisms in the word, such a contradiction of meanings within one word reflects the first, primitive stage of the development of thinking [Makarova, 2010; Pimenova, 2007; Charekov, 2009].

Examples of enantiosemey in words of a syncretic nature [Romanchuk, 2017; 53]:

***to draw***

- 1) to open the blinds, curtains, etc.
- 2) to close the blinds, curtains, etc.

***dust***

- 1) to clean furniture, a room, etc. by removing dust from surfaces with a cloth;
- 2) to cover something with fine powder, flour, etc.

***to rent***

- 1) to regularly pay money to somebody so that you can use something that they own;
- 2) to allow somebody to use something that you own in exchange for regular payments;

***seed***

- 1) to plant seeds in an area of ground;
- 2) remove the seeds from vegetables, etc;

***awesome***

- 1) rather frightening;
- 2) very good, enjoyable, etc;

***some***

- 1) a large number or amount of something;
- 2) a small number or amount of something.

***dahshat*** (horror)

- 1) very scary;

2) it's great

**bebaho** (priceless)

1) no price

2) has a very high price

**aylanib chiqmoq** (go around)

1) wander around without going inside;

2) go inside and look at all the places;

**aylanmoq** (turn around)

1) stand still and turn around

2) explore many points

**diqqatli bo'lmoq** (be careful) (!)

1) with his whole body

2) lost in thought

**sanksiya** (sanction)

1) to allow;

2) restrict / prohibit

In Turkic languages: *suchik* (bitter, delicious), *o'r* (depth, hill).

Based on research sources, it can be said that the phenomenon of syncreticity in the language is similar to the phenomenon of polysemantic and homonymy, and syncretic forms appear during the development of the language, and the opposite meanings at its core are the factors in the formation of enantiosemy.

## CONCLUSION / RECOMMENDATIONS

1. Syncretism is a linguistic phenomenon that includes several meanings, several functions, categories, and several forms in one form, and it occurs in the process of language development.

2. Syncretism is a phenomenon with form and expression.

3. For linguistic ontology, the phenomenon of syncretism in Uzbek linguistics requires in-depth research, and it is necessary to identify lexical units with syncreticity and form them as a base.

4. In-depth study of the phenomenon of syncretism in the system of Uzbek language lexicons increases the possibility of tagging lexicons in natural language processing.

## REFERENCES

1. Abjalova M.A. Ontology of the Uzbek language: technology and concept of creation. [Text]: monograph / M.A. Abjalova. – Tashkent: Nodirabegim, 2021. – 215 p. ISBN 978-9943-7804-5-3

2. Babajceva V. V. Sinkretizm // *Lingvističeskij jenciklopedičeskij slovar'* / Glavnyj redak. V. N. Jarceva. – M.: Sovetskaja jenciklopedija, 1990. – 685 s. – ISBN 5-85270-031-2.
3. Babajceva, V. V. Bol'shoj jenciklopedičeskij slovar' / V. V. Babajceva. – 2000. – S. 446.
4. Beresneba V.A. Teorija vseedinstva L.P. Karsavina kak filosofskij fon ligvističeskogo sinkretizma.
5. Beresneva V. A. Sinkretizm vremennyh form sovremennogo nemeckogo jazyka. Kirov: Izd-vo VjatGGU, 2008. – S. 3.
6. Beresneva V.A. Lingvističeskij sinkretizm: Ontologija i gnoseologija. – Kirov: Izd-vo Kirov. gos. un-ta, 2011. – 246 s.
7. Buzarov V. V. Sinkretizm kak raznourovnevoe sredstvo realizacii jazykovoju jekonomii // *Lingvističeskie kategorii v sinhronii i diahronii*. – Pjatigorsk, 1996. – S. 19–42.
8. Demidova K. I. Sinkretičnye javlenija v leksike sovremennogo russkogo jazyka // *Jazykovaja dejatel'nost': perehodnost' i sinkretizm: sb. st. nauch.-metod. seminaru «TEXTUS»*. – Vyp. 7 / pod red. K. Je. Shtajn. – M.; Stavropol': Izd-vo SGU, 2001. – S. 71.
9. Drugovejko S. V. Sinkretizm jazykovogo znaka v poezii postmodernizma // *Vestn. S.-Peterb. un-ta*. – Ser. 2. Istorija, jazykoznanie, literaturovedenie. – SPb., 2000. – Vyp. 2, № 10. – S. 58–61.
10. El'mšev L. Prolegomeny k teorii jazyka: per. s angl. Ju. K. Lekomceva // *Novoe v lingvistike: sb. st.* – M.: Inostr. lit., 1960. – Vyp. 1. – S. 264–389.
11. Eremin A. N. Perehodnost' i sinkretizm v leksičeskoj semantike prostorečnogo slova // *Jazykovaja dejatel'nost': perehodnost' i sinkretizm: sb. st. nauchometod. seminaru «TEXTUS»*. – M.; Stavropol': Izd-vo SGU, 2001. – Vyp. 7 / pod red. K. Je. Shtajn. – S. 74.
12. Kolesnikova, T. V. K voprosu o vydelenii vidov sinkretizma / T. V. Kolesnikova // *Gumanitarnye issledovanija*. – 2009. – № 3 (31). – S. 47.
13. Makarova, E. M. O pricinah i projavlenijah jenantiosemi v russkom jazyke v mezhsлавjanskom aspekte / E. M. Makarova // *Vestnik Nizhegorodskogo universiteta im. N. I. Lobachevskogo*. – 2010. – № 4 (2). – S. 631–635.
14. Pavljukovec, M. A. Sinkretizm na morfologičeskom i sintaksičeskom urovnjah anglijskogo jazyka kak projavlenie jazykovoju jekonomii: funkcional'nyj aspekt: avtoref. dis. ... kand. filol. nauk: 10.02.04 / M. A. Pavljukovec. – Rostov-na-Donu, 2009. – 22 c.
15. Pimenova, M. V. Semantičeskij sinkretizm v diahronii / M. V. Pimenova // *Russkij jazyk v kontekste nacional'noj kul'tury*. – Saransk: Izd-vo Mord. un-ta, 2007. – S. 161–166.
16. Plungjan V. A. Additivnaja model' morfologii i otklonenija ot neju // *Obshhaja morfologija: Vvedenie v problematiku: Učebnoe posobie*. – Izd.



2-e, ispravlennoe. – M.: Jeditorial URSS, 2003. – S. 42. – 384 s. – (Novyj lingvisticheskij uchebnik). – 2000 jekz. – ISBN 5-354-00314-8.

17. Prosjannikova, O. I. Semanticheskie izmenenija v sinkreticheskikh formah «sushhestvitel'noe / glagol» v anglijskom jazyke / O. I. Prosjannikova. – Vestnik Leningradskogo gosudarstvennogo universiteta im. A. S. Pushkina. – 2011. – S. 95.

18. Romanchuk, Ju. V. Sinkretizm v jazyke. Jenantioseimija kak chastnyj sluchaj projavlenija sinkretizma v jazyke // Filologicheskie nauki v Rossii i za rubezhom : materialy V Mezhdunar. nauch. konf. – Sankt-Peterburg, 2017. – S. 53.

19. Charekov, S. L. Semanticheskaja struktura slovoobrazovanija v russkom i altajskih jazykah: monogr. – 2-e izd. ispr. i dop. – SPb.: LGU im. A. S. Pushkina, 2009. – 116 s.

20. Shherbak A.M. Oчерki po sravnitel'noj morfologii tjurkskih jazykov (glagol). – M. – S. 8–12.

21. [https://ru.wikipedia.org/wiki/Sinkretizm\\_\(lingvistika\)](https://ru.wikipedia.org/wiki/Sinkretizm_(lingvistika))

УДК 004.891

**ВСПОМОГАТЕЛЬНЫЙ МЕТОД ОЦЕНКИ СТЕПЕНИ  
ИСТИННОСТИ ЭТИМОЛОГИЙ ОГУЗСКИХ ЭТНОНИМОВ  
СПИСКА М. КАШГАРИ*****И. А. Исмаилов***

*“Институт космических исследований природных ресурсов”  
Национального Аэрокосмического Агентства Азербайджана  
Азербайджан, г. Баку,  
tokuzoghuz@gmail.com*

Предлагается новый метод оценки степени истинности этимологий этнонимов на примере этимологизации огузских этнонимов списка М. Кашгари. Метод опирается на результаты анализа экспертных этимологических классификаций тюркских этнонимов и факты существования и упадка ранних тюркских государств, выявивших, что гипотетическими этимонами для огузских этнонимов могут быть этимоны определённых типов. Суть метода состоит в разбиении множества гипотетических этимонов на категории и присвоении каждой категории определённого числа от 0.1 до 0.9. В зависимости от категории этимона, к которому принадлежит конкретный гипотетический этимон для целевого этнонима, полученное число будет являться оценкой степени истинности гипотетического этимона (этимологической цепи). Достоинством метода является то, что при отсутствии вне лингвистических (исторических, географических и иных) подтверждающих (или опровергающих) этимологию этнонима фактов (традиционный метод оценки), с помощью нового метода можно в какой-то мере оценить степень истинности (или ложности) гипотетического этимона (этимологической цепи). В качестве приложения метода демонстрируется оценка степени истинности этимона и этимологической цепи предложенной экспертной системой “Oghuz Ethnonyms ES” для одного из огузских этнонимов списка М. Кашгари, которая подтверждается результатом оценки с помощью традиционного метода.

**Ключевые слова:** оценка этимологий этнонимов, экспертная система, база знаний, историческая фонетика, этимология

**“AN AUXILIARY METHOD FOR ASSESSING THE DEGREE OF  
TRUTHFULNESS OF THE ETYMOLOGIES OF THE OGHUZ  
ETHNONYMS IN THE M. KASHGARI LIST”*****Ismayilov Ismayil Arif oglu***

*“Institute for Space Research of Natural Resources”  
of Azerbaijan National Aerospace Agency  
Azerbaijan, Baku  
tokuzoghuz@gmail.com*

A new method for assessing the degree of truthfulness of the etymologies of ethnonyms is proposed on the example of the etymologization of the Oghuz ethnonyms in the list of the 11th century scholar Mahmud Al-Kashgari. The evaluation method is based on the results of the analysis of expert etymological classifications of Turkic ethnonyms and the facts of the existence and decline of the early Turkic and Oghuz states, which preceded the union of 24 Oghuz tribes, which revealed that the following types of etymons: ethnonyms, titles, anthroponyms, toponyms, totems and some common nouns (in descending order of priority) could be etymons for 22 Oghuz ethnonyms. The essence of the method consists in dividing the set of hypothetical etymons into categories or types of etymons (ethnonyms related to the confederation of Oguz tribes - Tokuz-Oguz from Old Turkic texts; Turkic ethnonyms from Old Turkic texts; ethnonyms from the book *Divanu-lugat at-Turk* by M. Kashgari; reconstructed ethnonyms Tokuz-Oguz from Chinese sources; names of groups of tribes or titles from Old Turkic texts; non-Turkic ethnonyms from Old Turkic texts; proper names or terms disputed between an ethnonym and an anthroponym; anthroponyms; toponyms or zoonyms; common nouns) with a certain number assigned to each category from 0.1 to 0.9. Depending on the category of the etymon to which the specific hypothetical etymon belongs for the target ethnonym from the list of M. Kashgari, the assigned number will be an assessment of the degree of truth of the specific hypothetical etymon (etymological chain). The advantage of the proposed assessment method is that in the absence of facts outside the linguistic (historical, geographical and other) confirming (or refuting) the etymology of the ethnonym (which form the basis of the traditional method for assessing etymons (etymologies) of ethnonyms), using the proposed new method, it is possible to measure the degree of truth (or falsity) of a hypothetical etymon (or a hypothetical etymological chain) for the target ethnonym. In case there are facts outside linguistic confirming (or refuting) the etymology of the ethnonym, the proposed method can be used as an auxiliary tool for assessing the degree of truth (or falsity) of a hypothetical etymon (or a hypothetical etymological chain). The paper demonstrates heuristic historical-linguistic expert rules and an objective fact, on the basis of which the inference machine Rule-based ES of the "Oghuz Ethnonyms ES" proposed a hypothetical etymon and a hypothetical etymology for one of the Oguz ethnonyms of the M. Kashgari list. The truth of the proposed hypothetical etymon (hypothetical etymology) is evaluated using the proposed new auxiliary evaluation method. The assessment obtained by the new method is confirmed by the result of assessing the degree of truth of the etymology of the given ethnonym using the traditional assessment method.

**Keywords:** evaluation of etymologies of ethnonyms, expert system, knowledge base, historical phonetics, etymology

### **Введение и постановка задачи**

Известно, что для правильной этимологизации слова (а также в частности этнонима – вставка моя) кроме собственно лингвистики, необходимы знания из разных наук, особенно из истории, эпиграфики, литературы и географии, которые принято называть

«внелингвистическими факторами» [Введенская, 2004, 26]. Обозначим «внелингвистические факторы» или «внелингвистические факты» (мы вместо термина «фактор» используем термин «факт» (fact) как более ясный логически и интуитивно термин) - (outside the Linguistics facts) аббревиатурой (OLF).

Таким образом, если гипотетический этимон (или гипотетическая этимологическая цепь) имеет подтверждающий «вне лингвистический факт», то данный этимон (или этимологическая цепь) может считаться близкой к истинному этимону (или истинной этимологической цепи).

В случае, когда у гипотетического этимона или гипотетической этимологической цепи нет какого-либо подтверждающего (или наоборот опровергающего) вне лингвистического факта, то возникает проблема оценки степени истинности (или ложности) данного гипотетического этимона (или гипотетической этимологической цепи).

С целью разрешения этой проблемы на примере этимологизации огузских этнонимов списка знаменитого учёного XI века Махмуда Аль-Кашгари [Atalay, 1985, s. 55-58], возникла задача создания какого-либо иного отличного от традиционного метода (имеется в виду использование для оценки этимологии этнонимов OLF), вспомогательного метода оценки степени истинности (или ложности) гипотетического этимона или гипотетической этимологической цепи.

### Решение

Этимологизировать – значит устанавливать первоначальное (истинное, основное) значение слова, т.е. отыскивать то исходное слово (этимон), от которого произошло рассматриваемое слово [Введенская, 2004, с. 10]. Известно, что этноним, будучи словом, подчиняется законам языка. Форма этнонима за время его существования может измениться. Самые всеобщие замены – фонетические. [Никонов, 1970, с. 25–30].

Как известно языковыми изменениями занимается специальная наука - историческая лингвистика и её часть - историческая фонетика. Согласно этой науке существуют следующие видоизменения звуков в потоке речи:

I. Комбинаторные изменения (в зависимости от соседства других звуков);

II. Позиционные изменения (связанные с положением в неуданном слоге, в конце слова и т.д.).

К комбинаторным изменениям относятся например: приспособление артикуляции (движения произносительных органов при образовании звуков) согласных под влиянием гласных и гласных под влиянием согласных; ассимиляция с её видами – уподобление согласного согласному или гласного гласному; диссимиляция – обратное ассимиляции – расподобление артикуляции двух одинаковых или подобных звуков; метатеза (греч. перестановка) – взаимная перестановка звуков или слогов в пределах слова и т.д.

К позиционным изменениям относятся, например: редукция – изменение (ослабление) звуков по качеству и количеству; отпадение звуков; оглушение – потеря звонкости звуков, паразитические звуки и т.д. [Бондаренко, 2007, с. 114–118].

Помимо фонетических изменений, на формирование и эволюцию этнонимов оказывают влияние и процессы суффиксации. На важность учета этнонимобразующих аффиксов (или формантов) при этимологическом анализе этнонимов указывал выдающийся ономастик В.А. Никонов [Никонов, 1970, с. 25–27].

На пути от первоначального этимона до конечного этнонима могут происходить различные перечисленные выше фонетические изменения, словообразовательные изменения (в частности в связи с тюркскими этнонимами, присоединение или выпадение аффиксов, образование слов – композитов или наоборот разделение композита с дальнейшим выпадением бывшей части композита) а также не учитываемые нами семантические изменения.

Для задач этимологизации огузских этнонимов списка М. Кашгари нами была разработана “Основанная на правилах экспертная система” (Rule-based ES) – ‘Oghuz Ethnonyms ES’ [Абдуллаева, Исмаилов, 2016, с. 127-128]. Для представления знаний в базе знаний ЭС нами выбрана логическая модель, точнее как более рентабельный обратный логический вывод [Negnevitsky, 2005, p. 38–40], [Endriss, 2014, p. 5–6], [Рассел, 2006, с. 311–316], [Исмаилов, 2022, с. 75–76]. В процессе использования этой системы для целевого этнонима “Tüger” (или “Tögär” по Ерджиласун [Ercilasun, 2008, с. 14–15], которого придерживаемся и мы) из списка 22 огузских этнонимов приведённых знаменитым учёным – лингвистом XI-го века Махмудом Аль-Кашгари [Atalay, 1985, s. 55–58], сработали (fired) следующие историко-фонетические эвристические экспертные правила, которые приводятся далее.

**Правило-1: (вариативность гласных  
(back\_vowels>front\_vowels))**

- ТО** *в этимологической цепи гласные заднего ряда > в соответствующие гласные переднего ряда*  
**ЕСЛИ** *этимон содержит только гласные заднего ряда.*

При конструировании этого эвристического правила были использованы экспертные знания [Кононов, 1980, с. 66–67], [Тенищев, 1984, с. 52–55, 67, 69], [Atalay, 1985, с. 56–57], [Алиева, 2006, с. 3–14].

**Правило-2: (деназализация “η”)**

- ТО** *в этимологической цепи звук ‘η’ > ‘g’*  
**ЕСЛИ** *этимон содержит звук “η” И звук “η” находится в середине слова И (звук “η” находится между гласными звуками ИЛИ между гласным и согласным звуками).*

В процессе работы над конструированием этого эвристического правила (деназализация “η”) мы воспользовались экспертными знаниями [Текин, 1968, р. 92–93], [Щербак, 1970, с. 179], [Кононов, 1980, с. 104], [Тенищев, 1984, с. 339–341], [Erdal, 2004, р. 80].

**Правило-3: (Звуковая метатеза (r+гласная) > (гласная+r))**

- ТО** *в этимологической цепи (r+гласная) > (гласная+r)*  
**ЕСЛИ** *этимон содержит сочетание звуков (r+гласная)”.*

При конструировании данного эвристического экспертного правила (метатеза) мы руководствовались экспертными знаниями [Пальмбах, 1955, с. 293–297], [Щербак, 1961, с. 65], [Кононов, 1980, с. 72–73], [Тенищев, 1984, с. 368–369], [Erdal, 2004, с. 113–114].

Факт предметной области - этноним “Тоңға (Tongra)” (один из этнонимов конфедерации племён Токуз-Огуз), который упоминается в древнетюркских текстах несколько раз: в надписи Кюль-тегина (северная сторона, строка 7) [Текин, 1988, с. 22], в надписи Бильге кагана (восточная сторона, строка 31) [Текин, 1988, с. 48], в надписи Тоньюкука (южная сторона, строка 9) [Малов 1951, с. 61] удовлетворил сработавшим экспертным правилам данным выше.

Таким образом, система “Oghuz Ethnonyms ES” предложила следующую гипотетическую этимологию для целевого этнонима

*Tögär: Тоҗра > Төҗрә (вариативность гласных о>ö и а>ä) > Төҗрә (деназализация җ>g) > Төгәр (метатезис rä>är).*

Прежде всего инженеру по знаниям (в моем лице) необходимо было ответить на вопрос, какие слова могут быть источниками (или этимонами) для этнонимов вообще и для 22 этнонимов огузов в частности.

Чтобы получить ответ на этот вопрос было необходимо провести поиск и исследование существующих этимологических классификаций этнонимов вообще и Тюркских в частности.

Анализ популярных этимологических классификаций этнонимов (предложенных со стороны лингвистов Эрдманн, Г. Лангельфельд, В.И. Супрун, В.А. Никонов, А.И. Попов, Н.А. Баскаков, Д.Е. Еремеев, Е.З. Ахмедова выявил, что антропонимы, титулы, тотемы, топонимы, этнонимы, заимствования, имена нарицательные со значениями человек, человек, говорящий, друг, родство, интеграция, чужие, немые, враги, животные, внешние характеристики, ландшафт, религия, духовное качества могут быть этимонами этнонимов вообще.

Подробный анализ этимологических классификаций тюркских этнонимов, данных Н.А. Баскаковым, Д.Е. Еремеевым, Э.З. Ахмедовой, мнение известного эксперта-ономаста А.В. Суперанской, что основная часть собственных имён образована не непосредственно от имён нарицательных, а от других собственных имён, более ранних по времени своего возникновения [Суперанская, 1986, с. 81] и особенно существование и упадок (до периода образования Сырдарьинского Огузского государства) ряда тюркских государств (два Гёк-Тюркских и два Токуз-Огузских (Уйгурских) каганатов), племена которых могли быть предками 22 огузских племен, даёт основание, что следующие типы этимонов: этнонимы, титулы, антропонимы, топонимы, тотемы и упомянутые выше имена нарицательные (в порядке убывания приоритета) могли быть этимонами для 22 Огузских этнонимов [Исмаилов, 2022, с. 143–144].

Опираясь на этот вывод, был разработан метод оценки степени истинности (или ложности) этимона или гипотетической этимологической цепи для целевых Огузских этнонимов списка М. Кашгари, который мы назвали “ЕО” (от английских слов “etymon” и “origin” – происхождение этимона)). Псевдокод алгоритма предлагаемого метода представлен ниже на рисунке 1.

```

If TC Is TOE Then
    EO = 0.9
ElseIf TC Is OTE Then
    EO = 0.8
ElseIf (TC Is KTE) Or
    (TC Is RTOE) Then
    EO = 0.7
ElseIf TC Is GT Then
    EO = 0.6
ElseIf TC Is NT Then
    EO = 0.5
ElseIf TC Is PN Then
    EO = 0.4
ElseIf TC Is A Then
    EO = 0.3
ElseIf TC Is TZ Then
    EO = 0.2
ElseIf TC Is CN Then
    EO = 0.1
End If

```

Рис. 1. Псевдокод алгоритма метода EO  
(The algorithm's pseudo code of the method of EO)

Где TC (term's categories) – категория или тип этимона; TOE (Toquz Oghuz Ethnonym) – этноним относящийся к конфедерации огузских племён Toquz Oghuz из древнетюркских текстов; OTE (Ethnonym from Old Turkic texts) – тюркский этноним из древнетюркских текстов; KTE (Ethnonym from M. Kashgari's book Diwan Lugat at-Turk) – этноним из книги Дивану-лугат ат-Тюрк М. Кашгари; RTOE (Reconstructed from Chinese Toquz Oghuz Ethnonym) – реконструированный из китайских источников Токуз-Огузский этноним; GT (Name of the group of tribes or Title) – название группы племён или титул из древнетюркских текстов; NT (Non-Turkic Ethnonym from Old Turkic texts) – не тюркский этноним из древнетюркских текстов; PN (Proper Name or disputed term between ethnonym and anthroponym) – имя собственное или спорный термин между этнонимом и антропонимом; A (Antroponym) – антропоним; TZ (Toponym or Zoonym) – топоним или зооним; CN (Common Noun) – имя нарицательное.

Сущность предложенного метода состоит в разбиении множества гипотетических этимонов на выше приведённые категории и присвоении каждой категории определённого числа от 0.1 до 0.9. В зависимости от категории этимона, к которому принадлежит конкретный гипотетический этимон для целевого этнонима,



полученное число будет являться оценкой степени истинности гипотетического этимона (этимологической цепи).

В результате применения метода ЕО к этимону “**Toŋra**” и этимологии “**Toŋra>Töŋrā>Tögrā>Tögär**” для целевого огузского этнонима “**Tögär**” строковая переменная ТС получила значение ТОЕ, и следовательно одноимённая с методом числовая переменная ЕО получает максимальное значение 0.9.

Оценка степени истинности (или ложности) данного гипотетического этимона или данной гипотетической этимологической цепи традиционным методом OLF, также показала высокий результат на основании следующего предметного факта: Этноним “ttagara” из хотаноязычного свитка Stäel-Holstein был исследован ученым В. Хеннинг (V. Henning), который сначала поставил этот этноним в один ряд с этнонимом индоевропейского народа “Тохар” (Tokhar), но затем поменял своё мнение в пользу Токуз-Огузского этнонима “Тоŋра”. Согласно новому мнению В. Хеннинга, “ttagara” представляет \*tögere (\*toʏar пишется как \*Ttauħ:ari) [Henning, 1938, p. 545–571]. Примечательно, что репрезентация хотанского “ttagara” как \*tögere очень схожа с огузским этнонимом “Tögär” из списка М. Кашгари.

Таким образом, в результате применения двух методов оценки степени истинности (или ложности) гипотетического этимона “Тоŋра” и соответственно гипотетической этимологической цепи (**Toŋra >Töŋrā (вариативность гласных o>ö и a>ä) > Tögrā (деназализация ŋ>g) > Tögär (метатезис rā>är)**) для этнонима “Tögär” из списка 22 огузских этнонимов списка М. Кашгари, были получены высокие оценки, что является достаточно убедительным по нашему мнению, аргументом в пользу этимона и этимологии, предложенной системой “Oghuz Ethnonyms ES”.

## Заключение

Большим достоинством предложенного метода оценки степени истинности (или ложности) гипотетического этимона (гипотетической этимологической цепи) является то, что при отсутствии вне лингвистических подтверждающих (или опровергающих) этимологию этнонима фактов, с помощью данного метода можно в какой-то мере оценить степень истинности (или ложности) гипотетического этимона (или гипотетической этимологической цепи) для задач этимологизации огузских этнонимов списка М. Кашгари.

В случае же наличия вне лингвистических подтверждающих (или опровергающих) этимологию этнонима фактов (т.е. при срабатывании традиционного метода OLF), предложенный новый метод может использоваться как вспомогательный метод для оценки степени истинности (или ложности) гипотетического этимона (или гипотетической этимологической цепи) для этимологизации огузских этнонимов списка М. Кашгари.

В результате применения двух методов оценки степени истинности (или ложности) одного гипотетического этимона предложенной системой “Oghuz Ethnonyms ES” для этнонима “Tögär” из списка 22 огузских этнонимов списка М. Кашгари, были получены высокие оценки, что является достаточно убедительным по нашему мнению, аргументом в пользу истинности предложенной системой “Oghuz Ethnonyms ES” этимологии.

#### ЛИТЕРАТУРА

1. Абдуллаева Г.Г., Исмаилов И.А. Конструкция батареи экспертных систем для установления этимологий этнонимов (на примере огузских этнонимов) // Transactions of Azerbaijan National Academy of Sciences. Series of Physical-Technical and Mathematical Sciences. Informatics and Control Problems, V. XXXVI, 2016, № 3, p. 123 - 130.
2. Алиева Т. К. Вариантность слова и литературная норма (на материале современного карачаево-балкарского языка) : авт. дис. ... канд. филол. Наук / Кабар.-Балкар. гос. ун-т. Нальчик, 2006.
3. Бондаренко М.А. Курс лекций «Введение в языкознание», Тула, 2007, 391 с.
4. Введенская Л.А., Колесников Н.П., Этимология: Учебное пособие.-СПб.: Питер, 2004. – 221 с.
5. Джексон П. Введение в экспертные системы. М.: Вильямс, 2001, 623 с.
6. Исмаилов И.А. Разработка структурной экспертной системы // Вестник Компьютерных и Информационных Технологий. 2018, № 10, с. 48–58.
7. Исмаилов И.А. Применение структурной экспертной системы в этимологических изысканиях. PROCEEDINGS of the X International Conference on Computer processing of Turkic Languages “TURKLANG 2022” с. 69–79.
8. Исмаилов И.А., Исмаилоглы Г.И. Выявление приоритетностей этимонов этнонимов с помощью статистической обработки этимологических классификаций // Известия Кыргызского Государственного

Технического Университета имени И. Раззакова. 2022, №3 (63), с. 138–144.

9. Кононов А. Н. Грамматика языка тюркских рунических памятников VII–IX вв. Л. : Наука, 1980.

10. Малов С. Е. Памятники древнетюркской письменности. М.-Л. : АН СССР, 1951.

11. Никонов В.А. Этнонимия // Этнонимы. М., 1970.

12. Пальмбах А. А., Исхаков Ф. Г. Явления метатезы в Тувинском и в некоторых других тюркских языках // Исследования по сравнительной грамматике тюркских языков. Фонетика. Москва : АН СССР, 1955.

13. Рассел С., Норвиг П. Искусственный интеллект: современный подход. М.: Вильямс, 2006, 1408 с.

14. Суперанская, А.В. Теория и методика ономастических исследований. Наука : Москва, 1986, 255 с.

15. Тенищев Э. Р. Сравнительно-историческая грамматика тюркских языков. Фонетика. М. : Наука, 1984.

16. Щербак А. М. Грамматический очерк языка тюркских текстов X–XIII вв. из восточного туркестана. М.-Л. : АН СССР, 1961.

17. Щербак А. М. Сравнительная фонетика тюркских языков. Л. : Наука, 1970.

18. Atalay Besim. Divanü Lüğat-it-Türk tercümesi. Ankara : Türk Tarih Kurumu Basım evi, 1985.

19. Erdal M. A. Grammar of Old Turkic. Leiden : Brill, 2004.

20. Ercilasun, A. B. (2008). Oğuz Boy Adlarının Etimolojisi [The Etymology of Oghuz Tribe Names]. Dil Arashtirmalari Dergisi, 3, 9–25. (In Turkish).

21. Henning W. Argi and the ‘Tokharians’. Bulletin of the School of Oriental and African Studies 1938; 9(3): p. 545–571.

22. Negnevitsky Michael, Artificial Intelligence, Addison-Wesley. England. 2005, 407 p.

23. Ulle Endriss. (2014) Lectures Notes. An Introduction to prolog programming. Institute for Logic, Language and Computation, University of Amsterdam, 2014.

24. Tekin T. (1968) A Grammar of Orkhon Turkic. Indiana University, Bloomington, 1968.

25. Tekin T. (1988) Orhon yazıtları. Ankara: Türk Tarih Kurumu Basım evi, 1988 (in Turkish).

---

## ЛИНГВОПРОЦЕССОРЫ

УДК 81'322.2:004.912:811.512.157

### МОРФОЛОГИЧЕСКИЙ ПРЕОБРАЗОВАТЕЛЬ (АНАЛИЗ И СИНТЕЗ) ДЛЯ ЯКУТСКОГО ЯЗЫКА

*В. Н. Кортесосо, В. П. Захаров*

*Санкт-Петербургский государственный университет*

*Санкт-Петербург, Россия*

st082534@student.spbu.ru, v.zakharov@spbu.ru

Морфологический анализ играет решающую роль в конвейере NLP, особенно когда мы имеем дело с агглютинативными языками, такими как якутский. В данной статье даются общие рекомендации по построению морфологического преобразователя для якутского языка. Морфологические преобразователи обычно работают в двух режимах: анализ текста и генерация. В режиме анализа преобразователь принимает словоформу в качестве входных данных и выделяет лексический корень вместе с соответствующими словообразовательными и флективными аффиксами, составляющими структуру словоформы. И наоборот, в режиме генерации преобразователь на основе заданного лексического корня (основы) генерирует флективную или словообразовательную словоформу. Исходный код преобразователя находится в открытом доступе и постоянно расширяется и совершенствуется.

**Ключевые слова:** морфологический преобразователь, агглютинативный язык, якутский язык, открытый исходный код.

### A MORPHOLOGICAL TRANSDUCER FOR YAKUT LANGUAGE

*Cortegoso Vissio Nicolás, Zakharov Victor Pavlovich*

*Saint Petersburg State University,*

*Saint Petersburg, Russia*

st082534@student.spbu.ru, v.zakharov@spbu.ru

Morphological analysis plays a crucial role in the NLP pipeline, particularly when dealing with agglutinative languages like Yakut. This paper provides general guidelines for constructing a morphological transducer for the Yakut language. Morphological transducers typically operate in two modes: analysis and text generation. In the analysis mode, the transducer takes a word form as input and identifies the lexical root, along with the associated derivational and inflectional affixes contributing to its structure. Conversely, in generation mode, the transducer generates an inflected or derived word form based on a given lexical root. The source code for this transducer is publicly accessible and subject to ongoing expansion and refinement efforts.

**Keywords:** morphological transducer, agglutinative language, Yakut, open-source.

## Introduction

Yakut (also known by its endonym ‘Sakha’) belongs to the Turkic family, a group of approximately 40 languages spoken by 200 million people across a vast geographical region spanning from Eastern and Southern Europe to East and North Asia. Turkic languages present interesting challenges for language processing due to features such as agglutinating morphology, vowel harmony, and free constituent order in syntax.

Among the languages in this family, Turkish has the most tools and resources for transforming a text into general-purpose linguistic structures, which can be used to support various NLP applications. The research conducted on Turkish and the solutions proposed to address its linguistic complexities can serve as a foundation for developing similar tools for less-resourced members of the Turkic family, such as Yakut.

Like Turkish, Yakut constructs words through a highly productive affixation process that involves attaching multiple suffixes to a lexical root. For example, the word *суруйааччыларбыт* ‘our writers’ is formed from the root *суруй* ‘to write’ by adding three affixes: *-ааччы* (agent), *-лар* (plural), and *-быт* (first person plural possessive). Each affix conveys a distinct meaning, and the overall meaning of the word is a combination of the root meanings of all the affixes [Ubryatova et al., 1982, p. 32].

The surface realization of the affixes, i.e., how they effectively appear in a word form, is influenced by various regular morphophonological processes such as vowel harmony and consonant assimilation. Consequently, nearly all affixes exhibit systematic allomorphs, including variations in vowel patterns and boundary consonants. For instance, the plural affix mentioned earlier can take on up to 16 different forms {*лар, лэр, лор, лөр, тар, тэр, тор, төр, дар, дэр, дор, дөр, нар, нэр, нор, нөр*}, depending on the stem to which it attaches.

To perform morphological analysis on a word form, the transducer described in this paper decomposes it into a sequence of tags representing the identified affixes. This analysis involves the application of a set of morphographemic rules that map the surface form of the word to its lexical representation, effectively segmenting the word into morphemes. For example, given the input *суруйааччыларбыт*, the transducer generates the following output:

суруй^verb^part5+plur+p#pl&1

In this representation, the symbols ‘^’ and ‘+’ indicate the boundaries between affixes, and the nomenclature that follows represents the type of affix. The symbol ‘^’, when occurring immediately after the root, refers to the class of the lexical root; otherwise, it indicates that the affix is derivational. In the given example, ^verb designates that the lexical root *суруй* is a verbal root, and the following affix ^part5 stands for one of many participial forms that can be derived from a verbal root. On the other hand, the symbol ‘+’ indicates that the affix is inflectional, such as +plur or +p&pl&1 in the example. The symbol ‘#’ inside an inflectional affix specifies number agreement, while ‘&’ represents grammatical person.

The tags that the transducer outputs in analysis mode are the same that can be used as input in generation mode to produce the desired word form. For instance, if in generation mode, we change the person in the last inflectional affix to:

суруй^verb^part5+plur+p#pl&2

the output will be *суруйааччыларгыт* ‘your writers’.

The role of the morphological transducer is to trace the origin of each productive affix back to the lexical root. The lexical root is a stand-alone word with a core meaning, which is further completed by subsequent affixes. The transducer’s generation mode can be used to produce intermediate representations of an analyzed word form. For example:

суруй^verb^part5 → *суруйааччы*

суруй^verb^part5+plur → *суруйааччылар*

Hence, employing a morphological analyzer on a text corpus can effectively address the challenge of word form sparsity commonly encountered when working with agglutinative languages.

## Previous work

In comparison to well-studied languages like English or Russian, the available linguistic research for Yakut is relatively limited. However, significant contributions have been made to the field since pioneering work in the 17th century, including the creation of grammars and dictionaries dedicated to Yakut. For the development of the morphological transducer, our primary reference was [Ubratova, 1982],

as it represents the most comprehensive work on Yakut morphology available to us. For an in-depth exploration of the history of Yakut linguistic studies, readers are encouraged to consult [Filippov, 2006].

As mentioned in the introduction, the majority of research in the field of NLP for Turkic languages has predominantly focused on Turkish. An overview of research specifically related to the Turkish language can be found in [Ofłazer, Saraçlar, 2018].

Yakut has also made significant strides in its online presence. As of September 2023, the Yakut Wikipedia<sup>1</sup> hosts 16,782 articles in the Yakut language. Additionally, there are dedicated news agencies regularly producing content in Yakut, such as *КЫЫМ* ‘Spark’<sup>2</sup> and *ЯСИА* ‘Yakut-Sakha Information Agency’<sup>3</sup>.

In terms of NLP resources, the Universal Dependencies project introduced a treebank in 2022 [Merzhevich, Ferraz Gerardi], and Fast-Text<sup>4</sup> includes sub-word model embeddings for the Yakut language.

Regarding available electronic tools, the site sakhatyla.ru<sup>5</sup> provides an online dictionary for Yakut, and Yandex Translator<sup>6</sup> offers a beta version for translating this language.

In 2022, a morphological analyzer and generator for Yakut was published [Ivanova, 2022]. This analyzer/generator, built using the Helsinki Finite State Technology (HFST) framework, has open-source code and an available online demo<sup>7</sup>. The authors compiled the morphotactic and morphophonological rules by hand following a methodology similar to the one that they used previously for a Turkish analyzer [Washington, 2019].

The existence of this morphological analyzer/generator for Yakut recently came to our attention. Therefore, the research phase was carried out with TRMOR [Kayabaş et al., 2019] as a reference. TRMOR is

---

<sup>1</sup> Wikipedia in Yakut language. URL: <https://sah.wikipedia.org> (last access: 28/09/2023)

<sup>2</sup> Kyym Online news agency. URL: <https://kyym.ru> (last access: 28/09/2023)

<sup>3</sup> Yakut-Sakha Information Agency. URL: <https://sakha.ysia.ru/> (last access: 28/09/2023)

<sup>4</sup> Fasttext word vectors for Yakut (Sakha). URL: <https://fasttext.cc/docs/en/crawl-vectors.html> (last access: 28/09/2023)

<sup>5</sup> SakhaTyla.Ru. URL: <https://sakhatyla.ru> (last access 28/09/2023)

<sup>6</sup> Yandex Translator. URL: <https://translate.yandex.com/> (last access: 28/09/2023)

<sup>7</sup> Online demo Yakut morphological analysis. URL: <https://beta.apertium.org/index.eng.html#analysis?aLang=sah> (last access: 28/09/2023)

an open-source morphological analyzer specifically designed for Turkish, implemented within the Stuttgart Finite State Transducer (SFST) framework [Schmid, 2005]. The SFST is a collection of software tools used for the generation, manipulation, and processing of finite-state automata and transducers. The availability of TRMOR's source code and the opportunity to study it motivated the decision to use the same framework for developing the morphological transducer for Yakut. Initially, the intention was to closely adhere to the guidelines established in TRMOR. However, later on, it was decided to organize the code in a way that reflects the structure presented in the subsequent section of this paper.

### **Transducer's development cycle**

A morphological transducer generally consists of three components:

1. a lexicon containing lexical roots and affixes, along with essential information about them;
2. a set of rules that determines the permissible sequence of morphemes in a word form;
3. a set of phonological rules that describe the changes occurring in a word form when morphemes combine.

The development of the morphological transducer for Yakut aims to model these three elements following the description of Yakut phonology and morphology by Ubryatova et al. [1982].

The construction of the transducer involves six main blocks, that also structure the following sections of this article:

1. alphabet definition: specifies the input symbols accepted by the transducer;
2. stem definition: specifies the types of primary and derived stems;
3. derivational affixes: lists the derivational affixes;
4. inflectional affixes: lists the inflectional affixes;
5. morphotactics: refers to the rules and models that govern the process of concatenating affixes to stems;
6. phonotactics: models morphophonemics, including phenomena like vowel harmony, consonant assimilation, and alteration in stems after affixation.

The transducer's source code is organized to reflect this structure, resulting in six files named after the listed items. The lexicon, containing the roots to which the suffixes attach, is stored separately. Users can expand the lexicon by adding new lexical roots to that file.



The development process begins with the compilation of testing data, primarily consisting of declension and inflection tables, as well as example sentences found in Ubryatova's work. This approach offers two main advantages: comprehensive coverage of Yakut morphology and the presence of glosses and explanations that assist non-Yakut speakers.

The declension and inflection tables from the grammar are used to model general morphotactic and phonological rules. Since these tables represent abstractions of affixation processes, example sentences are employed to further test and refine these rules. Lexical roots are added to the vocabulary as needed.

While this development process may progress slowly, it allows for a meticulous approach. The constraints for affix concatenation are modeled based on the descriptions found in the source grammar. When the grammar does not fully specify these constraints, we prefer a less restrictive implementation that may occasionally produce incorrect outputs in generative mode. This approach is favored over a more restrictive one that might fail to analyze valid inputs.

## 1. Alphabet definition

This section describes the written representation of sounds in the Yakut language. Since 1939, the Yakut language has adopted an alphabet based on the Cyrillic script, which includes the entire Russian alphabet along with five additional letters {Ҕ, Ҥ, Ө, Һ, Ү}, and two digraphs {дь, нь}. The letters {в, ж, з, ф, ц, ш, щ, ь, я, е, ё} are primarily found in Russian loanwords.

The alphabet defines the set of characters that will be recognized by the transducer. For the sake of simplicity, only lowercase characters were considered. Consequently, the input string must be converted to lowercase, and any characters not included in the alphabet definition, such as punctuation marks and numbers, must be removed. For example, the transducer will fail to analyze the input string *Caxa!* 'Yakut!' unless it is given as *caxa*.

Formally, we define the alphabet as the concatenation of vowels and consonants:

$$\text{alphabet} \rightarrow \{ \text{vowels, consonants} \}$$

Vowels and consonants are defined separately as supersets of more specific nested subsets, as described in the following subsections 1.1

and 1.2. The minimal unit is a single letter character; therefore, the digraphs {дь, нь} from the Yakut alphabet are not interpreted as single characters but as the concatenation of ‘д’ or ‘н’ and ‘ь’.

The alphabet definition includes a group of special symbols known as placeholders, which convey underspecified vowels and consonants. These placeholders are realized in context based on the application of phonetic rules. Using placeholders to describe allomorphs and apply phonetic rules is a common practice in the design of transducers. In the source code, placeholders are represented by one or more Latin capital letters enclosed in angle brackets, and they are used extensively in the affix definition. For example, the plural affix, which in Yakut has as many as 16 allomorphs is defined as ‘<D2><O>p’, where ‘<D2>’ stands for dental-alveolar, covering {л, т, д, н}, and ‘<O>’ for long open vowel, representing {а, э, о, ө}. The selection of affix allomorphs is carried out through a combination of the rules of progressive consonant assimilation and vowel harmony (see section 6).

### 1.1 Yakut vowels

Yakut vowels have both short and long counterparts. Long vowels are not represented by special characters but are orthographically indicated by doubling the vowel.

**Table 1. Classification of vowels [Kharitonov, 1947, p51]**

|           | back      |         | front     |         |
|-----------|-----------|---------|-----------|---------|
|           | unrounded | rounded | unrounded | rounded |
| open      | а, аа     | о, оо   | э, ээ     | ө, өө   |
| closed    | ы, ыы     | у, уу   | и, ии     | ү, үү   |
| diphthong | ыа        | уо      | иэ        | үө      |

The transducer recognizes Yakut vowels {а, о, э, ө, ы, и, у, ү}, as well as vowels from Russian loanwords {е, я, ю, ё}. These sets are subdivided into sets of narrower scope, which are used to specify the application context of the phonological rules in section 6. For example:

back-unrounded-vowels: { а, ы }

back-rounded-vowels: { о, у }

back-russian-vowels: { е }

front-unrounded-vowels: { э, и }

front-rounded-vowels: { ө, ү }

front-russian-vowels: { ё, ю, я }

back-vowels: { back-unrounded-vowels, back-rounded-vowels, back-russian-vowels }

front-vowels: { front-unrounded-vowels, front-rounded-vowels, front-russian-vowels }

open-vowels: { а, э, о, ө, <O>, <LO> }

closed-vowels: { ы, и, у, ү, <C>, <LC> }

russian-vowels: { back-russian-vowels, front-unrounded-vowels }

vowels: { open-vowels, closed-vowels, russian-vowels, <DI> }

The placeholders ‘<O>’ and ‘<LO>’ represent open vowels and long open vowels, respectively. Similarly, ‘<C>’ stands for closed vowels, and ‘<LC>’ represents long close vowels. The placeholder for diphthongs is ‘<DI>’.

## 1.2. Yakut consonants

Yakut consonants are classified according to both place and manner of articulation. They are categorized by place of articulation into bilabial, dental/alveolar, palatal, velar/uvular, and glottal, and by manner of articulation into plosives, fricatives, nasals, laterals and trills [Ubrayeva, 1982, p. 54].

**Table 2. Classification of consonants [Kharitonov, 1947, p. 58]**

|            |            | by manner of articulation |     | by place of articulation |                 |         |              |         |
|------------|------------|---------------------------|-----|--------------------------|-----------------|---------|--------------|---------|
|            |            |                           |     | bilabial                 | dental/alveolar | palatal | velar/uvular | glottal |
| obstruents | plosives   | voiceless                 | п   | т                        |                 | к       |              |         |
|            |            | voiced                    | б   | д                        |                 | г       |              |         |
|            | fricatives | voiceless                 | (ф) | с, (ш)                   |                 | х       | h            |         |
|            |            | voiced                    | (в) | (з, ж)                   | й               | б       |              |         |
| sonorants  | nasals     |                           | м   | н                        | [нь]            | н       |              |         |
|            | laterals   |                           |     | л                        | [ль]            |         |              |         |
|            | trills     |                           |     | р                        |                 |         |              |         |

**Table 3. Compound consonants [Kharitonov, 1947, p. 58]**

| voiceless | voiced |
|-----------|--------|
| ч (ц, щ)  | дь     |

Formally, the transducer recognizes the following consonants: {п, б, м, т, д, с, н, л, р, й, к, г, х, б, н, h, ч}, along with consonants bor-

rowed from Russian loanwords {ф, в, ш, з, ж, ц, щ}, and the symbols {ь, ъ}. Similar to the vowel definition, this set of consonants can be arranged into smaller subsets for use with phonetic rules. For example:

unvoiced-stops: { п, т, к }  
 voiced-stops: { б, д, г }  
 stops: { unvoiced-stops, voiced-stops }  
 unvoiced-russian-fricatives: { ф, ш }  
 voiced-russian-fricatives: { в, з, ж }  
 unvoiced-fricatives: { с, х, h, unvoiced-russian-fricatives }  
 voiced-fricatives: { ь, voiced-russian-fricatives }  
 fricatives: { unvoiced-fricatives, voiced-fricatives }  
 nasals: { м, н, ң }  
 approximants: { л, р, й }  
 russian-composites: { ц, щ }  
 composites: { ч, russian-composites }  
 symbols: { ь, ъ }  
 consonant-placeholders: { <B>, <D1>, <D2>, <V>, <H>, <K> }  
 consonants: { stops, fricatives, nasals, approximants, composites, symbols, consonant-placeholders }

The placeholder '<B>' stands for Yakut bilabial consonants {п, б, м}. Dental-alveolar consonants {т, д, н, л} are represented by the placeholders '<D1>' and '<D2>', each with two different realizations. For example, after vowels, '<D1>' is replaced by т, while '<D2>' is replaced by 'л'. Velar-uvular consonants {ь, г, к, х, ң} use '<V>' as a placeholder. The placeholder '<K>' represents the velar-uvulars {г, к, ң}, and '<H>' stands for the letters {h, с, ч} and diagraphs {дь, нь}.

## 2. Stem definition

As mentioned in the introduction, the Yakut language features a pure concatenative morphology, represented as a sequence of affixes attached to a lexical root. The lexical root is a non-divisible form that conveys the core meaning of a word. In terms of analyzability, the lexical root remains intact after all productive affixes have been removed and cannot undergo further morphological analysis. On the other hand, a stem is the unit to which affixes are attached, consisting, at a minimum, of a lexical root.

Lexical roots are classified into types and listed in the lexicon along with their corresponding base stem types. Stem types are enclosed with-

in angle brackets in the code. For example, the Yakut word for ‘fish’ could be included in the lexicon as:

балык<noun-2c>

The stem types found in the lexicon are considered primary, unlike derived stems, which result from affixation. Derived stems are always formed from primary or other derived stems with the assistance of productive affixes.

Stem types encode linguistic information that plays a crucial role during the affixation process and the subsequent application of phonetic rules. The definition of stems is valuable for modeling both the ordering restrictions on morphemes (morphotactics) and how affixes change when attached to the base (phonotactics).

The stem definitions (both primary and derived) implemented here follow a three-element pattern, consisting of:

1. the class of the base stem, which is mandatory for all types of stems. The class can refer to a part-of-speech (e.g., adjective, adverb), a type of stem (e.g., nominal, verbal), or a type of affix (e.g., possessive, plural).

2. the number of syllables in the root, which applies optionally and exclusively to the roots in the lexicon (primary stems). Derivational stems do not track the number of syllables.

3. the type of stem ending, indicating whether it ends in a consonant, a vowel, a glide, etc. This also applies optionally to both primary and derived stems.

In the previous example of the lexical root *балык* ‘fish’, the stem ‘<noun-2c>’ is described as a two-syllable noun ending in a consonant. This nomenclature is purely conventional and could be replaced by other terms. However, the specifications regarding the class of the stem, the number of syllables, and the stem ending are essential for the current implementation of the transducer.

Formally, the definition of the stems consists of a set of all primary and derived stems. Similar to the alphabet definition in the previous section, stem types can be grouped into nested subsets:

stems → {primary stems, derived stems}

Stems impose certain restrictions on the type of affixes that can be attached to them. For example, nominal stems cannot take voice affixes, which can follow verbal stems. Conversely, affixes also present constraints in terms of the stems they can be attached to. For instance, the

affix '<LC>' forms nouns from verbal stems (both primary and derived) ending in a consonant.

Affixes have both a surface and an analysis form. The surface form models the realization of the affix in generation mode, while the analysis form represents the affix output in analysis mode.

Some affixes have different allomorphs depending on the stem to which they are attached. For instance, the accusative case affix takes the form 'н<C>' after vowels and '<C>' after consonants:

сир<sup>^</sup>noun+acc → *сири* (accusative form of land/earth)

паарта<sup>^</sup>noun+acc → *паартаны* (accusative form of student's desk)

Each new affix that is attached changes the stem and redefines the set of new affixes that can eventually be attached after it. The entire process of affixation can be described as a one-directional domino concatenation (to the right), where the last joined piece introduces new constraints for the pieces that can be attached after it. Similar to a domino piece, which has two square ends marked with a number of spots determining its concatenation capabilities, an affix also possesses two joints that determine its affixation possibilities to the left and to the right. Therefore, affixes are formally represented as a tuple of three elements:

({previous stems}, analysis form:surface form, new stem)

In this tuple, the first element contains the set of stems to which the affix can be attached, the second represents the analysis and surface forms of the given affix, and the third is the new type of stem that is formed after its attachment. The first element of the tuple serves to model concatenation constraints: the broader the scope in the stems set, the greater the variety of stems that the affix in question can join.

Some affixes are concatenated after consonants by adding a closed consonant '<C>' to the stem. This has been modeled as two variants of the same affix. For example, the affix corresponding to the instrumental case can be modeled as:

variant 1:

({nominal-v, cardinals-v, ...}, instrumental:н<O>н, <case-c>)

variant 2:

({nominal-c, cardinals-c, ...}, instrumental:<C>н<O>н, <case-c>)

In these examples, the first variant defines a set of previous stems ending in vowels ('-v'), while the second does the same with conso-

nants ('-c'). The surface form of the second variant starts with the symbol '<C>'.  
This three-element tuple is employed to define both the derivational affixes in section 3 and the inflectional affixes in section 4.

### 3. Derivational affixes

This section defines the affixes that form derivative stems. Derivational affixes include affixes that complement or modify the lexical meaning of the stem. One salient aspect of Yakut morphology is the extensive use of derivational affixes in word formation. The transducer identifies several types of derivational affixes.

- Pronoun derivational affixes: according to Ubryatova [1982, pp. 187–213], there are nine types of pronouns: personal, demonstrative, interrogative, definitive, indefinite, personal-reflexive, collective, generalizing, and possessive. The first three types are included in the lexicon, while the rest are derived forms from them.

- Negative affix: this affix attaches to primary and derived verbal stems.

- Possessive affix: it can be attached to a wide variety of nominal and participial stems, conveying the meaning of 'possession of something by someone/something' [Ubryatova, 1982, p. 163].

- Participial affixes: these affixes form verbal nouns and are considered as a lexical and grammatical category of words that can function as both nouns and verbs.

- Gerunds: gerunds represent the form of the verb that indicates the secondary nature of the action and its correlation in one way or another with another action.

- Mood affixes: these affixes are used in the construction of certain verbal moods.

- Nominalization affixes from nouns: these affixes form nouns from other nouns. For example, the affix '<H><C>г' denotes a person or figure for whom the object or concept represents the object of their occupation or profession, while affixes like 'ч<C>к', 'к<O>', 'ч<LO>н' form diminutives.

- Nominalization affixes from verbs: these affixes form nouns that represent the name or the result of an action.

- Adjectivation affixes from nominals: these affixes form adjectives denoting a special inclination, a sense of closeness, a feature of behavior, place, and time.

- Adjectivation affixes from verbals: these affixes express tendencies, habituality, special abilities related to the stem's meaning, or signs resulting from an action.
- Numeral affixes from other numerals: ordinal numbers serve as the basis for forming the names of all other type of numerals.
- Adverbialization affixes from verbals: these affixes create adverbs from verbal stems and participles.
- Adverbialization affixes from nominals: similar to the previous type, these affixes form adverbs, but they are derived from nominals.
- Verbalization affixes from nominals: these affixes transform nouns and adjectives into verbs. The specific meaning of the derived verb depends on the nominal root's meaning.
- Verbalization affixes from onomatopoeias: onomatopoeic verbs are formed from onomatopoeic roots with the help of various affixes.
- Modal expression affixes: these verbal affixes attach to any verbal stem, conveying a modal-emotional connotation of affection, pity, regret, or emotions like humiliation, irony, annoyance, or contempt towards someone or something.
- Voice affixes: these affixes attach to verbal stems to specify the relationship of the action to the grammatical subject. They add to the lexical meaning of the primary verb stem, expressing the concept of the action in its most general form.
- Aspect affixes: aspect affixes attach to both primary and derived verbal stems, specifying aspects of an action, such as multiple or repeated actions.

#### **4. Inflectional affixes**

This section lists inflectional affixes. In Yakut, inflectional affixes are relatively fewer in number compared to derivational ones. They primarily consist of affixes that do not alter the lexical meaning of the stem but instead serve to express the connection or relationship of a given word to other words within a sentence. This category encompasses declension and conjugation affixes, as well as the predicate form of nominal stems [Ubryatova, 1982, p. 35]. The inflectional affixes in Yakut include the following:

- Plural: this affix attaches not only to nominals and verbal stems but also to various other stems, such as interrogative pronouns.
- Predicative affixes: these affixes are available for the first and second persons in both singular and plural forms.



- Possessive affixes: these affixes are used to show that one thing belongs to or is associated with another. They encode information about person, number, and grammatical case.
- Case affixes: Yakut utilizes eight cases to express the syntactic function and logical relationships between words.
- Interrogative affixes: this type of affix can be attached to the end of several stems to convey an interrogative meaning.

## 5. Morphotactics

In this section, derivational and inflectional affixes are effectively combined with the lexical roots defined in the lexicon. In Yakut, affixes are always attached to the stem at the back (to the right) since prefixes (affixes located ahead of the stem) do not occur [Ubryatova, 1982, p. 33].

Therefore, a Yakut word form can be formally defined as the concatenation of zero or more suffixes to the right of a lexical root:

root suffix\* (optional)

The stem definitions determine which concatenations are allowed and which are not. For example, the reflexive voice affix ‘<C>H’ might join a verbal stem but not a nominal one:

$k\acute{o}p\langle\text{verb-1c}\rangle + \{\text{verb-1c}, \dots\}\langle\text{C}\rangle\text{H}\langle\text{refl}\rangle \rightarrow k\acute{o}p\acute{u}n$

$at\langle\text{noun-1c}\rangle + \{\text{verb-1c}, \dots\}\langle\text{C}\rangle\text{H}\langle\text{refl}\rangle \rightarrow \emptyset$  (no result)

A valid joint results from a match between a stem type within the set of previous stems specified for an affix  $\{\text{stem1}, \text{stem2}, \dots\}$  and the last stem to the right in the concatenation chain up to that point. This is implemented in the code as a filter that excludes non-valid concatenations.

## 6. Phonotactics

This section outlines the application of phonological rules to both stems and affixes. It is essential to note that this aspect of transducer development is particularly challenging as the rules must not conflict with each other to achieve the intended outcomes. Crafting these rules involves considerations of functionality, optimization, and readability. Therefore, striking a balance between writing minimal rules while maintaining code readability is often a non-trivial task.

Optimizing this code segment requires continual revision and testing, which makes it less amenable to a static description compared to the previous sections. Consequently, we will provide a general overview of the rules to minimize the risk of content obsolescence.

In essence, phonetic rules encompass three fundamental processes: insertion, deletion, and replacement of symbols. In all cases, the rules must specify the context in which they apply. The implemented transducer rules adhere to the following structure:

‘x’ is always realized as ‘y’ in context ‘z’

Here ‘x’ is the analysis and ‘y’ is the surface form. These transformations differ in scope; some are generally applied, like vowel harmony, while others are specific to transforming a particular lexical root. The rule’s scope is determined by the context ‘z’.

Since in this implementation phonetic rules are applied in cascade, the succession order is crucial. Rules with narrower scopes are applied first since they can change the context in which more general rules operate. The basic cascade order is as follows:

- Root changes
- Stem changes
- Progressive consonant assimilation
- Regressive consonant assimilation
- Vowel harmony

The following subsections provide some examples of what these rules may consist of.

### 6.1 Root changes

These rules concern lexical roots that undergo significant alterations when affixes are attached. For instance, the roots of the demonstrative pronouns *бы* and *ол* change in oblique cases (when inflectional case affixes are attached), becoming *ман-* and *он-*, respectively [Ubryatova, 1982, p. 192].

### 6.2 Stem changes

This subsection models changes in the stem that occur when specific affixes are attached. Changes may involve the insertion or deletion of vowels, diphthong shortening, consonant devoicing, and more.

### 6.3 Progressive Consonant Assimilation

Progressive assimilation determines the first consonant of an affix based on the stem to which it is being attached. This consonant is represented by special placeholders like ‘<D1>’, ‘<D2>’ and ‘<V>’, as detailed in the alphabet definition section. For example, the dental consonant ‘<D2>’ in the plural affix ‘<D2><O>p’ is realized as shown in the table 4.

**Table 4. Surface realization of plural affix ‘<D2><O>p’**

| Lexical root | Terminal letter of the stem          | Rule     | Surface form        | Gloss           |
|--------------|--------------------------------------|----------|---------------------|-----------------|
| паарта       | Vowels and diphthongs                | <D1> → л | <i>паарталар</i>    | student’s desks |
| харандаас    | Voiceless consonants {к, п, с, т, х} | <D1> → т | <i>харандаастар</i> | pencils         |
| сарай        | й, р                                 | <D1> → д | <i>сарайдар</i>     | barn            |
| аан          | Nasals {м, н, ҥ}                     | <D1> → н | <i>ааннар</i>       | doors           |

Progressive consonant assimilation rules must be applied after the rules that govern stem changes because the latter can alter the stem ending, affecting the consonant assimilation process.

#### 6.4 Regressive Consonant Assimilation

Regressive consonant assimilation rules describe alterations in the word ending when an affix is attached. For instance, voiceless consonants {п, к, х} change to their voiced counterparts {б, г, ҕ} at the end of a stem before an affix starting with a vowel. In the case of the sibilant ‘с’, which does not have a voiced counterpart, it changes to the voiced guttural consonant ‘h’ in intervocalic positions.

In this implementation, regressive consonant assimilation occurs after progressive assimilation because it requires the latter to be applied first. For example, when attaching the dative affix ‘<V><O>’ to the word for horse *am*, the velar-uvular ‘<V>’ is realized as the voiceless plosive ‘к’. This voiceless plosive causes the end of the stem to change from the dental voiceless plosive ‘т’ to the uvular voiceless plosive ‘к’:

ат + <V><O>

ат + ка

ак + ка

#### 6.5 Vowel Harmony

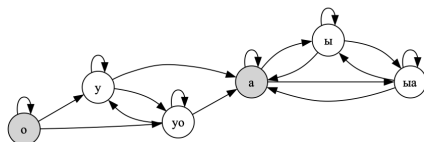
Vowel harmony in Yakut can be explained by two key principles:

1. palatal vowel harmony: word forms consist of either front or back vowels. If the first syllable of a word contains a back vowel, subsequent syllables must also have back vowels, and vice versa. Front vowels in the first syllable require all following vowels to be front as well.

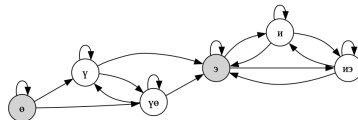
2. abial vowel harmony: within a word form, whether it is composed of back or front vowels, there are specific restrictions on transitions be-

tween syllables with open and closed vowels. Schemas 1 and 2 describe the allowable transitions for back and front vowels, with closed vowels represented in white and open vowels in gray.

Schema 1. Back vowels transition



Schema 2. Front vowels transition



Vowel harmony applies regularly to Yakut roots, which are internally harmonious. This is often not the case with Russian loanwords, which require differential treatment. Therefore, lexical roots corresponding to non-harmonious Russian loanwords should be identified in the lexicon using a special stem type.

### Limitations of the current approach

One of the main limitations of the current approach is lexicon coverage. The transducer fails to analyze a word form if the lexical root is absent from the lexicon. Even a substantial lexicon cannot account for every word form that may appear in a text. Proper names theoretically form an infinite set, and Yakut texts often include many words from Russian. Additionally, the presence of typos and misspelled words poses challenges. Oflazer [2018, p. 38] proposes a method to infer lexical roots based on a sequence of suffixes. This approach employs a transducer with no lexicon, where a lexical root is defined as a string of  $n$  characters. The drawback of this method is that, without establishing a clear boundary for the lexical root, various different segmentations of the given word into morphemes become possible. For example, a word form like *абалара* will produce at least the following analyses, assuming the primary stem of the lexical root is interpreted as a noun:

абалара<sup>^</sup>noun  
 абалар<sup>^</sup>noun+p&3  
 аба<sup>^</sup>noun+plur+p&3

However, if the type of the primary stem is not circumscribed to nouns, the number of possible analyses for the given word form increases further, as the lexical root could be identified as a verb, adverb,

adjective, or proper noun. The trade-off of not having unknown words is a transducer that outputs many spurious analyses (even though at least one of them should be correct). When parsing a text, we may still want to use the transducer without a lexicon as a last resort if the regular transducer fails to analyze a word form. The context of the sentence can be taken into account to resolve the morphological ambiguities in the analyses.

The transducer with the lexicon is also not exempt from producing ambiguities in the analysis. There are two types of ambiguity that may arise. One type of ambiguity could result from an insufficient implementation of the transducer rules. This source of ambiguity must be removed. The other type of ambiguity is intrinsic to the language and should be present. A common cause of natural ambiguity occurs systematically when a suffix has a homograph. For example, the predicate and possessive affixes in Yakut have the same form for the first and second person of the plural. As a result, word forms like *барыххыт* and *барыххыт* have two different interpretations:

*барыххыт* → бар<sup>verb^part3+a#pl&1</sup>  
бар<sup>verb^part3+p#pl&1</sup>

*барыххыт* → бар<sup>verb^part3+a#pl&2</sup>  
бар<sup>verb^part3+p#pl&2</sup>

Another source of ambiguity can be found within the lexicon itself when a lexical root is associated with more than one primary stem. For instance, the root *санаа* can be interpreted, before any suffixes are added, as a verbal stem (meaning ‘to think’) or as a nominal stem (meaning ‘thought’). In such cases, each lexical root is listed in the lexicon with its corresponding stem type:

санаа<verb-2v>  
санаа<noun-2v>

Further affixation will help determine the appropriate root type and resolve the initial ambiguity in the lexicon.

On the other hand, in certain analyses, more than one equally valid surface form may result. For example, when considering the accusative and instrumental cases for the interrogative pronoun *туох*, two alternative variants are legit, and the transducer rules should be capable of both generating and analyzing them:

*туох* + accusative case → *туоҕу* | *тугу*  
*туох* + instrumental case → *туобунан* | *тугунан*

Generally speaking, the development of the morphological transducer should aim to simultaneously minimize two problems: undergeneration and overgeneration. Undergeneration occurs when the transducer's rules are overly restrictive, leading to the failure to produce valid word forms in generation mode. Consequently, it also hinders successful analysis. Overgeneration, on the other hand, results from overly permissive rules that allow for the generation of non-existent word forms. In analysis mode, overgeneration may lead to an excessive number of parses.

Dealing with overgeneration is challenging because the constraints modeled here are primarily formal, and the implemented lexicon offers limited means to encode semantic information. Therefore, in generation mode, the transducer may often produce formally correct yet meaningless or non-existent words.

### Conclusions and further work

In this paper, we have outlined the foundational principles for a morphological transducer for the Yakut language. The transducer's code repository is now accessible to the public and open for use, adaptation, and distribution<sup>1</sup>. As an ongoing project, our commitment extends to expanding the existing lexicon, conducting rigorous testing, and addressing any challenges that may emerge. Additionally, our future endeavors will include the development of a mechanism to handle unknown words and effectively resolve any morphological ambiguities that may arise during sentence analysis.

### REFERENCES

1. Filippov G.G. Istoriya izucheniya yakutskogo yazyka i ee perspektivy [History of Yakut Language Study and its Prospects]. *Vestnik YaGU* [YaGU Bulletin], 2006, vol 3, no. 4, pp. 58–62.
2. Ivanova S., Washington J.N., Tyers F.M. A Free/Open-Source Morphological Analyser and Generator for Sakha. In: *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*. Marseille, 2022, pp. 5137–5142.
3. Kayabaş A., Schmid H., Topcu A., Kilic O. TRMOR: a Finite-State-based Morphological Analyzer for Turkish. In: *Turkish Journal of Electrical Engineering and Computer Sciences*. 2019, pp. 3837–3851.

---

<sup>1</sup> Code repository to Yakut morphological transducer. URL: <https://github.com/nicolasortegoso/yakutmorph> (last access 28/09/2023)

4. Kharitonov L.N. Sovremennyy yakutskiy yazyk. Fonetika i morfologiya [Modern Yakut Language. Phonology and Morphology]. Yakutsk, Gosizdat YaASSR, 1947. 313 p.

5. Merzhevich T., Ferraz Gerardi F. Introducing YakuToolkit. Yakut Treebank and Morphological Analyzer. In: *Proceedings of the 1st Annual Meeting of the ELRA/ISCA. Special Interest Group on Under-Resourced Languages*. Marseille, France. European Language Resources Association, 2022, pp. 185–188.

6. Oflazer K. Morphological Processing for Turkish. In: *Turkish Natural Language Processing*. Oflazer K., Saraçlar M. (eds.). Springer, 2018, pp 21–52.

7. Oflazer K., Saraçlar M. Turkish and its Challenges for Language and Speech Processing. In: *Turkish Natural Language Processing*. Oflazer K., Saraçlar M (eds.). Springer, 2018, pp. 1–22.

8. Schmid H. A Programming Language for Finite State Transducers, In: *Proceedings of the 5th International Workshop on Finite State Methods in Natural Language Processing*. Helsinki, Finland, 2005, pp. 308–309.

9. Ubryatova E.I. (ed.) Grammatika sovremennogo yakutskogo literaturnogo yazyka. Tom 1: Fonetika i morfologiya [Grammar of Modern Yakut Literary Language. Vol. 1: Phonology and Morphology]. Moscow, Nauka Print, 1982, 496 p.

10. Washington J.N., Salimzianov I., Tyers F. M., Gökırmak M., Ivanova S., Kuyrukçu O. Free/Open-Source Technologies for Turkic Languages Developed in the Apertium Project. In: *Proceedings of TurkLang*, 2019, pp. 30–71.

УДК 81'322.3

**МОРФОЛОГИЧЕСКИЙ АНАЛИЗ СЛОВОФОРМ  
В УЗБЕКСКОМ, КАРАКАЛПАКСКОМ  
И КЫРГЫЗСКОМ ЯЗЫКАХ, ПРИНАДЛЕЖАЩИХ  
К ТЮРКСКОЙ СЕМЬЕ ЯЗЫКОВ**

**Э. Назирова<sup>1</sup>, Н. Абдурахмонова<sup>2</sup>,  
Ш. Абидова<sup>1</sup>, М. Узакова<sup>3</sup>**

<sup>1</sup>*Ташкентский университет информационных технологий  
имени Мухаммада аль-Хорезми, Ташкент, Узбекистан*

<sup>2</sup>*Национальный университет имени Мирзо Улугбека  
Ташкент, Узбекистан*

<sup>3</sup>*Ташкентский университет информационных технологий  
имени Мухаммада аль-Хорезми  
Самарканд, Узбекистан*

elmira\_nazirova@mail.ru, abdurahmonova.1987@mail.ru,  
shaxnoza23@mail.ru, yulduzxon2626@gmail.com

В данной статье представлен морфологический анализ словообразования в узбекском, каракалпакском и кыргызском языках, принадлежащих к тюркской языковой семье. По данным морфологического анализа выявлены специфические черты тюркских языков. В языке возникает явление сингармонизма.

**Ключевые слова:** морфологический анализ, агглютинация, сингармонизм, грамматическая структура.

**MORPHOLOGICAL ANALYSIS OF WORD FORMATION  
IN UZBEK, KARAKALPAK AND KYRGYZ LANGUAGES  
BELONGING TO THE TURKIC LANGUAGES FAMILY**

**Elmira Nazirova<sup>1</sup>, Nilufar Abdurakhmonova<sup>2</sup>, Shakhnoza Abidova<sup>1</sup>,  
Mamura Uzakova<sup>3</sup>**

<sup>1</sup>*Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan*

<sup>2</sup>*National University of Uzbekistan named after Mirzo Ulugbek  
Tashkent, Uzbekistan*

<sup>3</sup>*Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Samarkand, Uzbekistan*

elmira\_nazirova@mail.ru, abdurahmonova.1987@mail.ru,  
shaxnoza23@mail.ru, yulduzxon2626@gmail.com



This article presents a morphological analysis of word formation in Uzbek, Karakalpak and Kyrgyz languages belonging to the Turkic language family. According to the morphological analysis, the specific features of the Turkic languages have been revealed. The phenomenon of synharmonism occurs in the language.

**Key words:** morphological analysis, agglutination, synharmonism, grammatical structure.

In today's modern world, the role of language is becoming very important for the development of society, and we have included the level of language learning as one of the urgent issues. Language is a social phenomenon formed not by a certain group, but by the whole society and its members in the entire historical process of human society over the centuries. [Абдурахманова, 2013]

It is known that there are different opinions about the role of word formation processes in linguistics. If a certain group of scientists believes that lexicology should study the objects of word formation, another group of scientists believes that these studies belong to the field of morphology. Both of these opinions are valid. As mentioned above, N.A. Baskakov shares word formation systems in morphology. In linguistics, the word-formation system acts as a syntactic department, and the word-formation system acts as a vocabulary department, thereby directly participating in the enrichment of the lexical content of the language. The classification of languages in the world will be observed, and in addition we will consider the family of Turkic languages. [Полатов, 2009]

Turkic languages are not only genetically related, but also typologically the same. According to the morphological classification of languages, it belongs to the group of agglutinative languages. The characteristic feature of Turkic agglutinative languages is as follows:

1. The word always begins from a root.
2. The core is basically unchanged. Any affix added after the root does not change the root phonetically.
3. Word forms are formed mainly by means of affixes.
4. Suppletive forms do not participate in the formation of word forms, that is, different forms of the same word are formed from only one root.
5. The core and the affix are not organically combined, the border between them is clear and obvious in most cases.

For example, in the form of the word *bog'dorchilik*,

*bog'* - root,  
*-dor* – word-forming suffix that forms the noun,  
*-chi* – shape-formating suffix  
*-lik* – word-forming suffix :

these are clearly distinguished from each other.

6. A separate affix is added to express each grammatical meaning, so several affixes appear in a row in one word form. In Turkic languages, words are divided into core and affixes.[2]

Having considered the morphological analysis of word formation in Uzbek, Karakalpak and Kyrgyz languages belonging to the Turkic languages family, we will compare these analyzes with each other. According to the typological classification, Turkic languages belong to the family of agglutinative languages, and the meaning of the word agglutinative in languages is the formation of a new word or grammatical form by adding affixes with grammatical meaning to the root of the word.

Based on the above, we assume that the system of word formation in the language should be considered not as a separate field of linguistics, but as a special system of morphology.

Language units – phoneme, morpheme, word, sentence are integrally connected with logical concepts (emotion, perception, thinking).

If we consider the grammatical structure of word formation in the Uzbek language, the words in the Uzbek language are morphologically divided into 2 types: simple and complex words.

There are methods of word formation in the Uzbek language such as affixation, composition, semantics, and the following 2 methods are the main and leading methods[4]:

1. Morphological method
2. Syntactic method

Forming a new word by adding word-forming suffixes to the core and root is called morphological method. Since suffixes are involved in word formation, this method is an affixation method. (Figure-1)

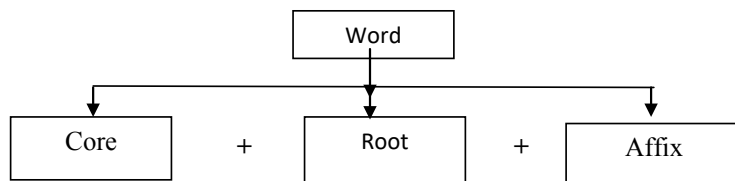


Figure-1. Morphological word formation scheme in the Uzbek language

One of the important conditions in the morphological method of word formation is the connection between the word and the new formation. Figure-1 shows the general scheme of word formation in Uzbek. There is no phonetic change when a suffix is added to the morpheme and base.

For example, kitobim = kitob+im  
 sinfdoshlarim = sinf+dosh+lar+im  
 bog'dorchilik = bog'+dor+chi+lik  
 paxtakorchilar = paxta+kor+chi+lar  
 bodomzor = bodom+zor

In this example, word-forming, word-changing, and shape-forming suffixes are added to the root, and morphemes and suffixes remain phonetically unchanged.

We will consider the morphological analysis of word formation in the Kyrgyz language. In this language, the rule of word formation is repeated as above, but in the process of word formation in Kyrgyz, the phenomenon of synharmonism occurs. Synharmonism (greek syn together and harmonia), harmony of vowels is a phonetic phenomenon that occurs mainly in Turkic languages; in which a phonetic change of a certain word form (both core, root and affixes) occurs. The simplest example of the occurrence of synharmonism is the addition of the plural suffix - **lar**.

So,

– words ending with «y» or «r» added –**lar** it will changed to - *lar* (-*ler*, -*lor*, -*lör*)

– words ending with voiced consonants added – **lar** it will changed to –*dar*, -*der*, - *dor*, - *dör*

– words ending with voiceless consonants added – **lar** it will changed –*tar*, -*ter*, -*tor*, *tör*

For example, **biy** + **lar** = **biyiklar**

**toktom** + **lar** = **toktomdor**

**student** + **lar** = **studentter**

Besides, китеп+им=ките**б**им ,

конок+ум=коно**г**ум

ысык+ыраак=ысы**г**ыраак

тап+уу=та**б**уу in these examples, the phonetic principle prevails over the morphological principle.

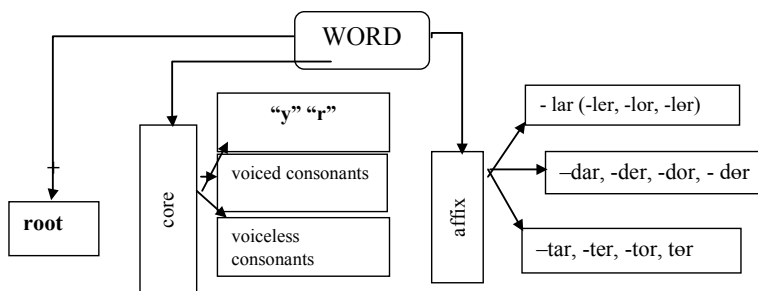


Figure-2. Morphological method of word formation scheme in Kyrgyz language (phenomenon of synharmonism)

This picture shows the word formation scheme in the Kyrgyz language.

The Karakalpak language is similar to the Kyrgyz language, and the word formation process is similar.

The Karakalpak language belongs to agglutinative languages. “One of the effective methods of forming grammatical forms for agglutinative languages is affixation, that is, attaching or not adding grammatical particles - suffixes to the root of the word, through which the word is formed or inverted. [Баскаков, 1952]

Characteristic features of the Karakalpak language (Figure-3):

- harmony of vowels – there is synharmonism: atlarshimiz (our horses), kunler – (days);
- consonant *ch* exchanged to *sh*; consonant *sh* exchanged to *s*, mas, qash (qoch), tas (tosh), bas (bosh);
- in some words, instead of the consonant *g'* exchanging *v*; instead of the consonant *g* exchanging *y*, mas, tav (tog<sup>o</sup>), tiy (teg) and e.t.

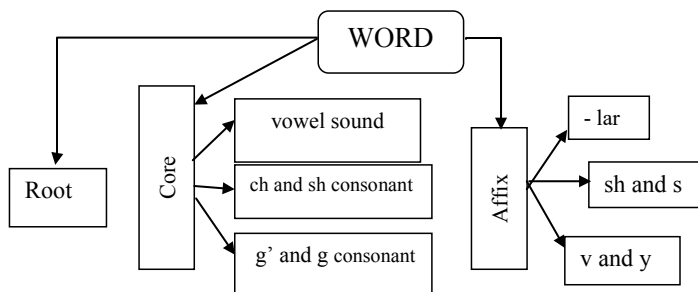


Figure-3. Morphological method of word formation scheme in Karakalpak language (phenomenon of synharmonism)

Synharmonism occurs in the process of word formation in the Karakalpak language. As a result of the change of sounds in the process of morphological analysis, the phonetic category shows its role.

In conclusion, the morphological analysis of word formation in Uzbek, Kyrgyz and Karakalpak languages belonging to the Turkic language family is very important in language learning. Morphemic analysis shows how words are formed, that the word is the basis of the sentence, and that this analysis is very important.

### LITERATURES

1. Абдурахманова Н., Хакимов М. Логико-лингвистические модели слов и предложений английского языка для многоязычных ситуаций компьютерного перевода. / Компьютерная обработка тюркских языков. Латинизация письменности. 1-я Международная конференция. – Астана, 2013. – С. 297–302. Nazirova E.Sh., Abidova Sh.B., Uzakova
2. M.A. Turkiy tillar uchun ikki tili elektron tarjimaning model va algoritmi. – T. 2023 yil
3. Po'latov A., Muhamedov S. Kompyuter lingvistikasi (o'quv qo'llanma) . – T., 2009
4. Пулатов А.К. Тексты лекций по математической и компьютерной лингвистике
5. Баскаков Н.А. Каракалпакский язык. – Т. II.1-qism. M., 1952 yil.

УДК 811.512.157

**МОРФОЛОГИЧЕСКИЙ АНАЛИЗАТОР ЯЗЫКА САХА:  
ОПЫТ РАЗРАБОТКИ И АПРОБАЦИИ****Г. Г. Торотов<sup>1</sup>, В. В. Бочкарев<sup>2</sup>***<sup>1</sup>Северо-Восточный федеральный университет  
им. М. К. Аммосова,**<sup>2</sup>Институт гуманитарных исследований и проблем  
малочисленных народов Севера Сибирского отделения  
Российской Академии наук  
torgav@mail.ru, uus-aldan@mail.ru*

В статье освещается многолетний опыт работы авторов данной статьи по разработке компьютерной программы «Морфонологический анализатор языка саха», начиная от создания лингвистической базы данных «Индексированные словоизменяемые аффиксы языка саха» и заканчивая апробацией работы морфонологического анализатора. Морфонологический анализатор языка саха при решении некоторых задач технического характера станет эффективным инструментом «Национального корпуса языка саха» и будет использован учителями родного языка в своей педагогической практике. Данный проект осуществляется в рамках государственной программы Республики Саха (Якутия) «Сохранение и развитие государственных и официальных языков в Республике Саха (Якутия) на 2020-2024 годы».

**Ключевые слова:** язык саха, национальный корпус, база данных, морфонологический анализатор.

**MORPHOLOGICAL ANALYZER OF THE SAKHA LANGUAGE:  
EXPERIENCE OF DEVELOPMENT AND TESTING<sup>1</sup>****Torotoev G. G.<sup>1</sup>, Bochkarev V. V.<sup>2</sup>***<sup>1</sup>Ammosov North-Eastern Federal University**<sup>2</sup>Institute for Humanitarian Research and Problems of Minor Peoples  
of the North of the Siberian Branch of the Russian Academy  
of Sciences**torgav@mail.ru, uus-aldan@mail.ru*

**The abstract:** The article describes the many years of experience of the authors of this article in developing the computer program “Morphological Analyzer of the Sakha Language”, starting from the creation of the linguistic database “Indexed

---

<sup>1</sup> Проект финансируется за счет средств государственной программы Республики Саха (Якутия) «Сохранение и развитие государственных и официальных языков в Республике Саха (Якутия) на 2020-2024 годы».

inflectional affixes of the Sakha language” and ending with testing the work of the morphological analyzer. The morphological analyzer of the Sakha language, when solving some technical problems, will become an effective tool of the “National Corpus of the Sakha Language” and will be used by teachers of the native language in their teaching practice. This project is carried out within the framework of the state program of the Republic of Sakha (Yakutia) “Preservation and development of state and official languages in the Republic of Sakha (Yakutia) for 2020–2024.”

**Key words:** Sakha language, national corpus, database, morphological analyzer.

Современное развитие якутского языкознания требует от исследователя оперативной обработки огромного количества лингвистической информации. Поэтому как никогда остро встает вопрос об автоматизации процесса работы языковедов. Для таких целей во всем мире используются национальные корпуса - электронные массивы текстов, которые подвергаются автоматическому анализу согласно выбранным параметрам.

С 2022 г. в рамках реализации государственной программы Республики Саха (Якутия) «Сохранение и развитие государственных и официальных языков в Республике Саха (Якутия) на 2020–2024 годы» начата работа по созданию «Национального корпуса языка саха» (рис. 1). На данный момент объем электронного корпуса составляет около 10 млн. словоупотреблений.

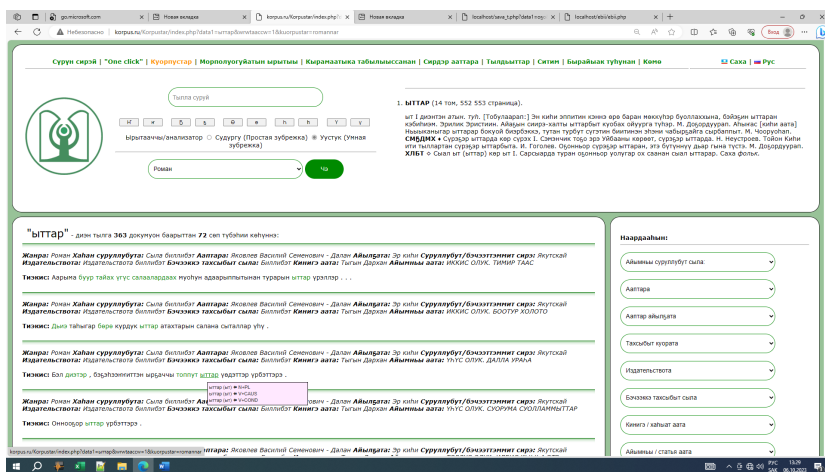


Рис. 1. Скриншот интерфейса «Национального корпуса языка саха»

Одним из важнейших инструментов корпуса языка саха является морфологический анализатор. Для того, чтобы компьютер мог автоматически анализировать языковой материал из электронного корпуса, необходимо было, в первую очередь, описать унифицированными тегами все актуальные граммы языка саха. Поэтому с 2014 г. начата планомерная работа по созданию базы данных по словоизменительным аффиксам языка саха. Результаты исследований апробированы в журналах ВАК, Scopus, материалах научных конференций [Ducksoo Kang, Торотов Г.Г., 2016, С. 66–90], [Леонтьев Н.А., Торотов Г.Г., 2017, С. 101–104], [Торотов Г.Г., Ноговицына А.Н. 2017, С. 108–120], [Торотов Г.Г., Торотова С.Г., 2019, С. 137–142], [Торотов Г.Г., Бочкарев В.В., 2022, С. 137–142], [Торотов Г.Г., Торотова С.Г., 2021, С. 329–336].

База данных «Индексированные словоизменительные аффиксы языка саха» (Торотов Г.Г., Торотова С.Г.), которая представляет собой систему индексов (тегов), отображающую весь словоизменительный потенциал языка саха, в 2020 г. получила свидетельство о государственной регистрации базы данных в Реестре баз данных (рис. 2).



Рис. 2. Свидетельство о государственной регистрации базы данных



Из этой базы данных в качестве примера обратимся к таблице «Залоговые формы якутского глагола» (табл.1). База данных «Индексированные словоизменительные аффиксы языка саха» состоит из 24 таблиц подобного формата.

**Табл. 1. Залоговые формы якутского глагола**

| Индекс | Расшифровка      | Название категории       | Алломорфы  | Морфемы                                   |
|--------|------------------|--------------------------|--|---|
| ACT    | Active voice     | Основной залог           | –  | –   |
| PASS   | Passive voice    | Страдательный залог      | -н<br>-[ы]лын/-[и]лин/<br>-[у]лун/-[ү]лүн  | -н<br>-[Ы]лЫн                             |
| REFL   | Reflexive voice  | Возвратный залог         | -[ы]н/-[и]н/-[у]н/<br>-[ү]н  | -[Ы]н                                     |
| CAUS   | Causative voice  | Побудительный залог      | -т<br>-тар/-тэр/-тор/-төр<br>-дар/-дэр/-дор/-дөр<br>-нар/-нэр/-нор/-нөр<br>-лар/-лэр/-лор/-лөр<br>-ар/-эр/-ор/-өр<br>-ыар/-иэр/-уор/-үөр | -т<br>-ТАр<br><br><br><br><br>-Ар<br>-ЫАр |
| RECP   | Reciprocal voice | Совместно-взаимный залог | -[ы]с/-[и]с/-[у]с/<br>-[ү]с  | -[Ы]с                                     |

Параллельно с этой базой также составлена база данных «Система идентификаторов (тегов), отображающая граммы и квазиграммы якутского языка». Сейчас ведется работа по созданию базы данных «Индексированные словообразовательные аффиксы языка саха», что позволит в будущем эффективно проводить этимологические исследования с целью воссоздания исторической связи языка саха с древнетюркским языком, а также выявления тюрко-монгольских языковых параллелей.

В целях компьютерной обработки языковых материалов, представленных в национальном корпусе языка саха, с 2018 г. начата разработка «Морфонологического анализатора языка саха» (разработчики – Торотоев Г.Г., Бочкарев В.В.) в стенах Института языков и культуры народов Северо-Востока Российской Федера-

ции СВФУ им. М.К. Аммосова. Сейчас IT-продукт находится на стадии доработки и апробации, и мы можем утверждать, что она способна в автоматическом режиме производить фонетический и морфологический анализ лексем.

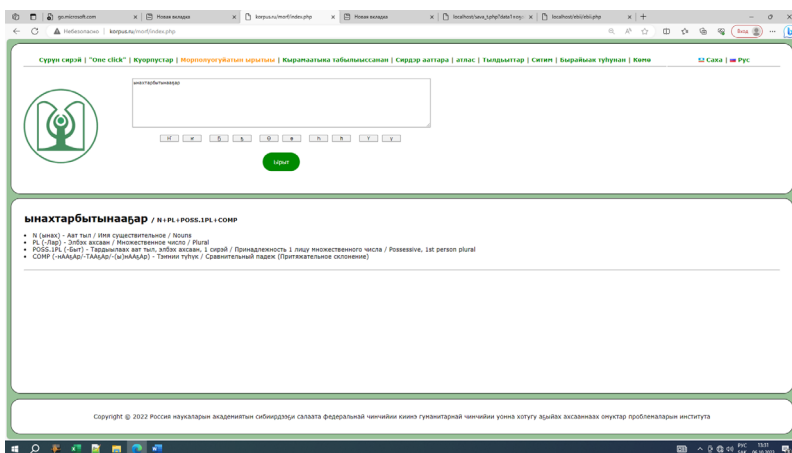
Морфонологический анализатор имеет сложную разветвленную модульную конструкцию и состоит из следующих компонентов:

- Модуль определения леммы;
- Модуль определения аффиксов;
- Модуль определения совместимости аффиксов с учетом законов сингармонизма языка саха;
- Модуль вывода результатов анализа.

В целях оптимизации автоматической обработки данных с большим объемом разработан алгоритм, основанный «на параллельных моделях». В данном случае продемонстрирована одна формула, отображающая имя обладания во множественном числе с 16 вариантами:

N-лардаахтар /-/ N-PL-PROPR- PRED.3PL  
 N-лордоохтор /-/ N-PL-PROPR- PRED.3PL  
 N-лэрдээхтэр /-/ N-PL-PROPR- PRED.3PL  
 N-лөрдөөхтөр /-/ N-PL-PROPR- PRED.3PL  
 N-нардаахтар /-/ N-PL-PROPR- PRED.3PL  
 N-нордоохтор /-/ N-PL-PROPR- PRED.3PL  
 N-нэрдээхтэр /-/ N-PL-PROPR- PRED.3PL  
 N-нөрдөөхтөр /-/ N-PL-PROPR- PRED.3PL  
 N-дардаахтар /-/ N-PL-PROPR- PRED.3PL  
 N-дордоохтор /-/ N-PL-PROPR- PRED.3PL  
 N-дэрдээхтэр /-/ N-PL-PROPR- PRED.3PL  
 N-дөрдөөхтөр /-/ N-PL-PROPR- PRED.3PL  
 N-тардаахтар /-/ N-PL-PROPR- PRED.3PL  
 N-тордоохтор /-/ N-PL-PROPR- PRED.3PL  
 N-тэрдээхтэр /-/ N-PL-PROPR- PRED.3PL  
 N-төрдөөхтөр /-/ N-PL-PROPR- PRED.3PL

Обратимся к скриншот (рис.3), где запечатлен результат работы морфонологического анализатора, в частности, ответная реакция на запрос по лексеме *ынахтарбытынааҕар* ‘по сравнению с нашими коровами’. И мы можем констатировать, что анализатор справился со своей задачей и адекватно разобрал данную лексему.

Рис. 3. Анализ лексемы *ынахтарбытынааҕар*

Как известно, снятие омонимии - одна из часто возникающих проблем при морфонологическом анализе. В данном скриншоте (рис.4) анализатор по омоформе *ыттар* выдал 3 результата '1) *собаки*; 2) *дай (ему) выстрелить*; 3) *если бы (он) выстрелил*', причем они все проанализированы правильно. Однако анализатор не смог распознать 4 вариант данной омоформы *ыттар* '(он) *забирается вверх*'. В таких случаях приходится решать проблему вручную.

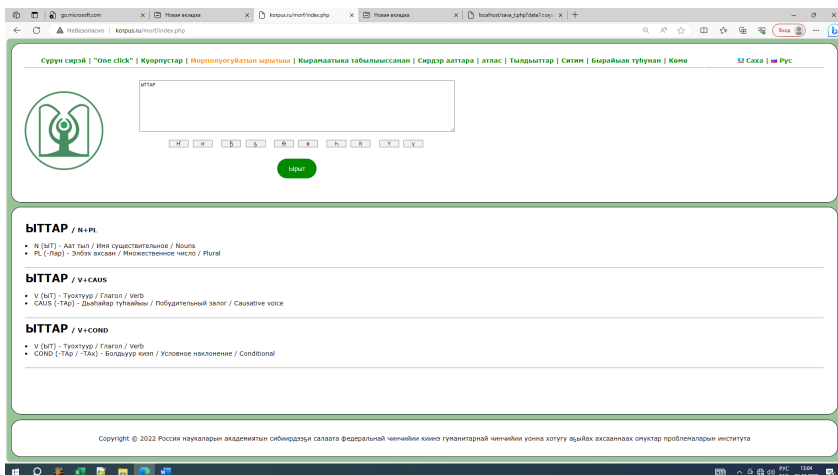


Рис. 4. Снятие омонимии

Копируем из корпуса простое предложение *Кыыс дьэтигэр кэлбит 'Девушка пришла домой'* (рис. 5) и вставляем в окно разбора. Морфонологический анализатор разбирает не только лексемы, но и отдельные фразы, а также целые абзацы.

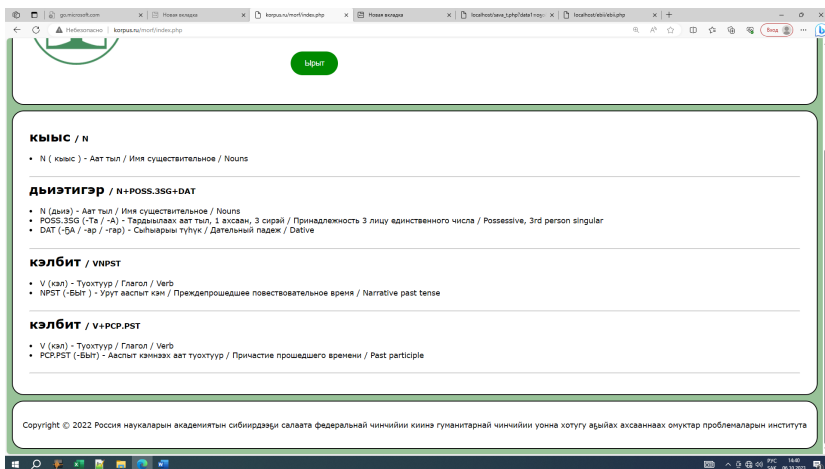


Рис. 5. Анализ простого предложения

При условии решения некоторых задач технического характера морфонологический анализатор языка саха станет эффективным инструментом национального корпуса языка саха. Мы надеемся, что наш IT-продукт найдет свое применение в обучении подрастающего поколения и внесет свой скромный вклад в дело сохранения и развития языка саха как родного и государственного языка республики.

## ЛИТЕРАТУРА

1. Леонтьев Н.А., Торотов Г.Г. Многопользовательская морфологическая разметка корпуса якутского языка // Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы (16–17 марта 2017 г., Сыктывкар): сб. материалов Международной науч. конф. – Сыктывкар: ГОУ ВО КРАГСиУ, 2017. – С. 101–104.
2. Торотов Г.Г., Ноговицына А.Н. Лингвистическое аннотирование наклонений глагола якутского языка // Вестник СВФУ. – №3(59). – 2017. – С. 108–120.
3. Торотов Г.Г., Торотова С.Г. Поморфемная нотация как эффективный метод в интерпретации переводческих трансформаций // Пере-

вод в поликультурном языковом пространстве Российской Федерации: потенциал и перспективы [Электронный ресурс]: сборник тезисов по материалам Всероссийской научно-практической конференции с международным участием (г. Якутск, 12-13 апреля 2019 г.) / [редкол.: Е.С. Герасимова и др.]. – Якутск: Издательский дом СВФУ, 2019. – С. 137–142.

4. Торотов Г.Г., Бочкарев В.В. Саха тылын ырытар бырагырааманы оноруу кедьүүһэ=Актуальность разработки универсального анализатора языка саха // Наследие предков и современный тюркский мир: языковые и культурные аспекты [электронный ресурс] : материалы II Международной научно-практической конференции, посвященной 100-летию со дня рождения известного якутского ученого-тюрколога, доктора филологических наук, профессора Якутского государственного университета им. М.К. Аммосова Николая Климовича Антонова, г. Якутск, 13 декабря 2019 г. / [под ред. Д.И. Чирковой, И.Ю. Васильева, Н.А. Ефремовой и др.]. – Якутск: Издательский дом СВФУ, 2020. – 1 электр. опт. диск. – С. 51–54.

УДК 81'322.2

**РАЗРАБОТКА СИСТЕМЫ ПРОВЕРКИ ОРФОГРАФИИ  
ТУВИНСКОГО ЯЗЫКА НА ОСНОВЕ HUNSPELL****Ч. Г. Ондар<sup>1</sup>, А. В. Чемышев<sup>2</sup>, Ч. Б. Хуурак<sup>1</sup>***<sup>1</sup>Тувинский институт гуманитарных и прикладных  
социально-экономических исследований при Правительстве  
Республики Тыва, Кызыл, Тыва, Российская Федерация**<sup>2</sup>Марийский научно-исследовательский институт языка,  
литературы и истории им. В.М. Васильева, Йошкар-Ола,**Республика Марий Эл, Россия**choygandi@mail.ru, chemyshev.andrey@gmail.com,  
huurak-chingis@yandex.ru*

В настоящее время ведется несколько работ по созданию программы проверки орфографии для тувинского языка. В данной статье описывается работа над проектом по созданию системы автоматической проверки орфографии на основе программы Hunspell. В статье приводится описание структуры лингвистической базы данных и технические принципы работы онлайн сервиса проверки правописания <https://tyvalab.ru/spelling/>: настоящее время расписаны парадигмы склонения для местоимений, имен существительного, числительного, ведется работа над формированием парадигмы глагола и прилагательного. Даются рекомендации разработчикам систем проверки правописания для тюркских языков и другие возможные области применения полученной базы данных.

**Ключевые слова:** тувинский язык, тюркские языки, система проверки правописания, Hunspell, орфография.

**DEVELOPMENT OF A TUVAN SPELL CHECKING SYSTEM  
BASED ON HUNSPELL*****Choigan Ondar<sup>1</sup> Chemyshev A. V.<sup>2</sup> Khuurak Ge. B.<sup>1</sup>****<sup>1</sup>Tuvan Institute of Humanities and Applied Social and Economic  
Research under the Government of the Republic of Tuva, Kyzyl,  
Tuva, Russian Federation**<sup>2</sup>Mari Scientific Research Institute of Language, Literature and  
History named after V.M. Vasiliev, Yoshkar-Ola, Mari El**choygandi@mail.ru, chemyshev.andrey@gmail.com,  
huurak-chingis@yandex.ru*

Currently, the Tuvan language is actively used in social networks and messenger apps; however, typos and mistakes frequently occur in texts written by native Tuvan speakers. They are also found in texts published online by governmental organizations, books, newspapers, and magazines. When we discuss

literacy in the traditional sense, we often imply that the language in question is a written language with a system of spelling rules. In the digital world, this implies that literacy needs to be supported by digital systems. Letters of the alphabet need to be available in Unicode and via keyboards. Moreover, spelling rules need to be reflected in spell checkers. These products are necessary for the digital development of any language. For the Tuvan language, keyboard layouts are available for Windows, Linux, MacOS, Android, and iOS; however, digital Tuvan spell checkers have not yet been created.

Spell checking software may be created on the basis of two methods: using a text corpus and using a set of rules; currently, several Tuvan projects are being developed. In the following article, we describe the creation of an automatic spell checking software on the basis of the Hunspell software. It specializes in languages with complex morphological systems, including the Tuvan language. In the article, we describe the structure of its linguistic database (vocabulary and the system of affixes) and technical principles of the online spell checker available at <https://tyvalab.ru/spelling>. In the linguistic database, paradigms of declension of nouns, pronouns and numerals are currently available; verb and adjective paradigms are also being developed, as well as a complete vocabulary of Tuvan lemmas. In this article, we offer recommendations to developers of spell checking softwares for Turkic languages, as well as other possibilities of usage of this linguistic database.

**Keywords:** Tuvan language, Turkic languages, spell checking system, Hunspell, spelling

## 1. Введение

По мнению известного венгерского лингвиста и математика Андреша Корная есть три основные показателя оценки неминуемой смерти естественного языка в цифровом пространстве: во-первых, происходит потеря функциональности, наблюдаемая всякий раз когда другие языки берут на себя целые функциональные области, во-вторых, потеря престижа языка, что особенно явно отражается в настроениях молодого поколения, в-третьих, потеря компетентности, когда носители языка принимают резко упрощенную версию грамматики при использовании языка в интернет-коммуникации [Kornai, 2013, с. 1].

Когда мы говорим о грамотности в традиционном смысле, мы предполагаем, что для этого языка есть письменность и система норм орфографии. В цифровом мире это означает, что в первую очередь необходимо, чтобы письменность поддерживалась в компьютерных системах. Буквы регулярного алфавита должны иметь юникод и методы ввода букв алфавита (раскладки клавиатур). Во-вторых, наличие стандартизированной орфографии соответствует необходимости в средствах проверки орфографии, *spell-checkeraх*. Эти два продукта являются основой цифрового развития

любого языка [Чемышев, 2021, с. 59]. И если для тувинского языка уже существуют раскладки клавиатуры для систем Windows, Linux, MacOS, Android, iOS<sup>1</sup> [Ондар, Донгак, Монгуш, 2023, с. 199, 201], то программы автоматической проверки орфографии для тувинского языка до сих пор не существуют.

Говоря о языковых компетенциях носителей тувинского языка, мы смело можем утверждать, что уровень грамотности тувинцев на русском языке выше, чем на тувинском языке. По крайней мере, об этом можно судить по многочисленным ошибкам в записях на тувинском языке в социальных сетях и мессенджерах, а также в текстах официальных сайтов учреждений [Ондар, Донгак, Монгуш, 2023, с. 204]. Во-первых, до появления и широкого распространения Интернета носители тувинского языка редко писали на тувинском языке. С появлением социальных сетей и мессенджеров значительная часть тувиноязычного населения, активно пользующаяся Интернетом, каждый день пишет на тувинском языке. Несмотря на то, что старшее поколение изучало правила орфографии и пунктуации тувинского языка в школах, со временем без практики навыки грамотного письма забывается. А молодое поколение в большинстве своем не изучали тувинский язык в школах или изучали совсем мало, чтобы в полной мере овладеть правилами правописания.

Таким образом, несмотря на то, что тувинский язык в настоящее время достаточно активно используется в социальных сетях и мессенджерах, носители делают большое количество опечаток и ошибок в текстах. Такая ситуация наблюдается в текстах официальных сайтов госучреждений, и даже в текстах книг, газет и журналов.

Программу по проверке правописания можно сделать двумя способами: на основе корпуса текстов и на основе правил.

Спеллчекер на основе корпуса текстов для тувинского языка разрабатывает тувинский разработчик Валерий Иргит на основе метода «Расстояние Левенштейна»<sup>2</sup>. Им был взят корпус на осно-

---

<sup>1</sup> Тыва танал (Тувинская раскладка клавиатуры) [Электронный ресурс] // Тувинский раздел Википедии. URL: [https://tyv.wikipedia.org/wiki/Тыва\\_танал](https://tyv.wikipedia.org/wiki/Тыва_танал) (дата обращения: 29.09.2023).

<sup>2</sup> Иргит В. Разбор кода на Kaggle: сбор корпуса и разработка системы проверки орфографии тувинского языка [Электронный ресурс] // YouTube. 2022, 25 декабря. URL: <https://www.youtube.com/live/JYJ6YTQ-IS-g?feature=share> (дата обращения: 29.09.2023).



ве текстов статей сайта Правительства Республики Тыва<sup>1</sup> и сайта газеты «Шын»<sup>2</sup>. Это корпус размером примерно 4,9 млн словоупотреблений, общее количество уникальных слов – 186406. Проект находится на стадии доработки.

В Институте гуманитарных и прикладных социально-экономических исследований при Правительстве Республики Тыва (далее – ТИГПИ) сектором языкознания совместно с отделом цифрового развития и Андреем Валерьевичем Чемышевым разрабатывается система проверки орфографии тувинского языка на основе Hunspell<sup>3</sup>. Данный проект также находится в состоянии разработки, продолжается процесс создания файлов «aff» и «dic», о которых речь пойдет ниже.

## **2. Разработка системы проверки орфографии тувинского языка на основе Hunspell**

Hunspell – это свободная программа для проверки орфографии, который больше предназначен для языков агглютинативной структуры. Hunspell используется офисным пакетом LibreOffice, некоторыми браузерами, такими как Mozilla Firefox и Google Chrome. Также его можно реализовать как отдельную программу по проверке орфографии. Возможности применения такой системы очень широки. Однако, надо понимать, что это программа предназначена для определения орфографических ошибок, т. е. пунктуационные, синтаксические или стилистические ошибки ею не устанавливаются.

### **2.1. Структура Hunspell для тувинского языка**

Hunspell состоит из двух файлов. Первый файл – это словарь («dic»), содержащий все слова (основы слов, т. е. леммы), второй файл – это файл аффиксов («aff»), который определяет значения специальных меток (флагов) в словаре (по сути, это модель систе-

<sup>1</sup> Официальный портал Республики Тыва. Главная страница [Электронный ресурс] // Официальный портал Республики Тыва. URL: <https://tyva.ru> (дата обращения: 29.09.2023).

<sup>2</sup> Сайт газеты «Шын» [Электронный корпус] // Общественно-политическая газета «Шын». URL: <https://shyn.ru> (дата обращения: 29.09.2023).

<sup>3</sup> Тимирханов Т., Любимов И., Словесник А., Губанов А. Hunspell (Описание Hunspell) [Электронный ресурс] // Mozilla Россия. URL: <https://mozilla-russia.org/projects/dictionary/hunspell.html> (дата обращения: 29.09.2023).

мы словоизменения, формообразования, включая некоторую часть словообразовательной системы языка).

Файл «dic» мы создаем на основе еще не изданного орфографического словаря тувинского языка (34 тыс. слов), подготовленного сектором словарей и языкознания ТИГПИ в 2022 году. Туда же мы планируем добавить список собственных имен и фамилий (тувинских, русских и др.), список топонимов Тувы, России, список стран, городов, валют, словарь интернационализмов т.п., чтобы охватить как можно больше слов, используемых в тувинском языке.

|    |                |        |
|----|----------------|--------|
| 1  | a/C1           |        |
| 2  | aa/N1Y1        |        |
| 3  | аагайлаар/V1   | [Verb] |
| 4  | аагар/V2       | [Verb] |
| 5  | аадам/A1N1     |        |
| 6  | аадама/A1N1    |        |
| 7  | аадамнаар/V1   | [Verb] |
| 8  | аадан/N2       | [Noun] |
| 9  | аадаңнаар/V1   | [Verb] |
| 10 | аадар/V2       | [Verb] |
| 11 | аадыг/N2       | [Noun] |
| 12 | аадышкын/N2    | [Noun] |
| 13 | аажок/A1N2     |        |
| 14 | аажы/N1        | [Noun] |
| 15 | аажылаар/V1    | [Verb] |
| 16 | аажылаашкын/N2 | [Noun] |
| 17 | аажылал/N2     | [Noun] |
| 18 | аажы-чаң/N2    | [Noun] |
| 19 | аазаар/V1      | [Verb] |
| 20 | аазаашкын/N2   | [Noun] |
| 21 | аазакчы/N1     | [Noun] |
| 22 | аазатпай/A1N2  |        |
| 23 | аазатпаяк/N2   | [Noun] |
| 24 | аай/N2         | [Noun] |

Рис. 1. Файл «dic» тувинского языка

Вообще перед тем, как приступать к формированию содержания файла «aff», лучше сразу составить полный список слов, который послужит основой для создания файла «dic». В большей степени это касается существительных, потому что обычно в список слов орфографического словаря добавляется список именованных сущностей. А список других частей речи (местоимений, прилагательных, наречий, глаголов и служебных слов) обычно полно представлен в орфографическом или двуязычных словарях. Это необходимо для того, чтобы на старте иметь все возможные варианты склонения, иначе придется добавлять парадигмы склонения вновь добавленных слов.

Из данного списка необходимо удалить слова, производные от других частей речи, с помощью аффиксов, которые могут быть применены ко всем членам части речи, а также залоговые формы у глаголов и т.п. Например, в тувинском языке словообразовательный аффикс *-лыг*, образующий относительное прилагательное от существительного, может присоединяться к каждому существительному. А в орфографическом словаре тувинского языка представлено ограниченное количество относительных прилагательных, поэтому этот и другие аналогичные аффиксы лучше вынести в файл «aff».

Сам файл «aff» тувинского языка имеет следующий вид (см. Рис. 2). Хорошее объяснение структуры этого файла имеется в описании к Hunspell<sup>1</sup>. В настоящее время расписаны парадигмы склонения для местоимений, имен существительного, числительного, ведется работа над формированием парадигмы глагола и прилагательного. Таким образом, из изменяемых частей речи остались служебные имена.

```

1 SET UTF-8
2 TRY абвгдгёжзийклмнопрстуфхцчшщъыьэюяАБВГДЕЖЗИЙКЛМНОПРСТУУФХЦЧШЩЪЫЬЭЮЯ
3
4 FLAG long
5
6 WORDCHARS -
7
8 SFX N1 Y 236
9 SFX N1 0 0 . +Sg+Nom
10 SFX N1 0 ннц [аны] +Sg+Gen
11 SFX N1 0 ниц [ениэ] +Sg+Gen
12 SFX N1 0 нун [ёоуэ] +Sg+Gen
13 SFX N1 0 нун [өү] +Sg+Gen
14 SFX N1 0 ны [аны] +Sg+Acc
15 SFX N1 0 ни [ениэ] +Sg+Acc
16 SFX N1 0 ну [ёоуэ] +Sg+Acc
17 SFX N1 0 ну [өү] +Sg+Acc
18 SFX N1 0 га [аёоуыюя] +Sg+Dat
19 SFX N1 0 ге [ениэүэ] +Sg+Dat
20 SFX N1 0 да [аёоуыюя] +Sg+Loc
21 SFX N1 0 де [ениэүэ] +Sg+Loc
22 SFX N1 0 дан [аёоуыюя] +Sg+Abl
23 SFX N1 0 ден [ениэүэ] +Sg+Abl
24 SFX N1 0 же [аеёиоёууэюя] +Sg+Lat1
25 SFX N1 0 дыва [аны] +Sg+Lat2
26 SFX N1 0 диве [ениэ] +Sg+Lat2
27 SFX N1 0 дува [ёоуэ] +Sg+Lat2
28 SFX N1 0 луве [өү] +Sg+Lat2
29 SFX N1 0 лар [аёоуыюя] +Pl+Nom
30 SFX N1 0 лер [ениэүэ] +Pl+Nom
31 SFX N1 0 ларнын [аёоуыюя] +Pl+Gen
32 SFX N1 0 лернин [ениэүэ] +Pl+Gen
33 SFX N1 0 ларны [аёоуыюя] +Pl+Acc

```

Рис. 2. Начало файла «aff» для тувинского языка

<sup>1</sup> Тимирханов Т., Любимов И., Словесник А., Губанов А. Hunspell (Описание Hunspell) [Электронный ресурс] // Mozilla Россия. URL: <https://mozilla-russia.org/projects/dictionary/hunspell.html> (дата обращения: 29.09.2023).

Для создания файла «aff» в первую очередь нужно иметь список лемм по принципу обратного словаря, поскольку выбор аффиксов зависит от конечных букв основы. Эту операцию можно осуществить в сервисе <https://tyvalab.ru/sorting/>, специально созданного для разработки системы проверки правописания тувинского языка.

а  
аа  
шалбаа  
холбаа  
сарбаа  
кодан-таваа  
чаваа  
аскыр-чаваа  
оваа  
шоваа  
эрес-шоваа  
уваа  
дагаа  
сагаа  
чагаа  
шагаа  
салгаа  
чалгаа  
богаа  
тодуг-догаа  
хальын-догаа  
ногаа  
маргаа  
чугаа  
амдыгаа

Рис. 3. Список слов по принципу обратного словаря

Таким образом, мы здесь представили некоторые важные моменты для создания основы Hunspell – файлов «aff» и «dic». На основе слияния данных файлов мы также можем получить наиболее полный словарь словоформ тувинского языка, который можно интегрировать в программы распознавания текстов, например, в АБВУУ FineReader, и в различные текстовые редакторы в качестве словаря поддержки тувинского языка.

## 2.2. Сервис проверки правописания тувинского языка на основе Hunspell

В данном разделе мы рассмотрим создание онлайн сервиса (сайта)<sup>1</sup> для проверки орфографии тувинского языка.

<sup>1</sup> Сервис проверки правописания тувинского языка [Электронный ресурс] // Сайт TyvaLab.ru (ТИГПИ). URL: <https://tyvalab.ru/spelling> [дата обращения: 04.10.2023].

Процесс взаимодействия интернет-пользователя и сайта основан на постоянной обработке запросов на стороне сервера и выдаче результатов на стороне пользователя. А чтобы обеспечить интерактивность и динамичность сервиса проверки правописания, то было решено в архитектуре использовать специальный шаблон (паттерн) под названием MVC на языке программирования PHP.

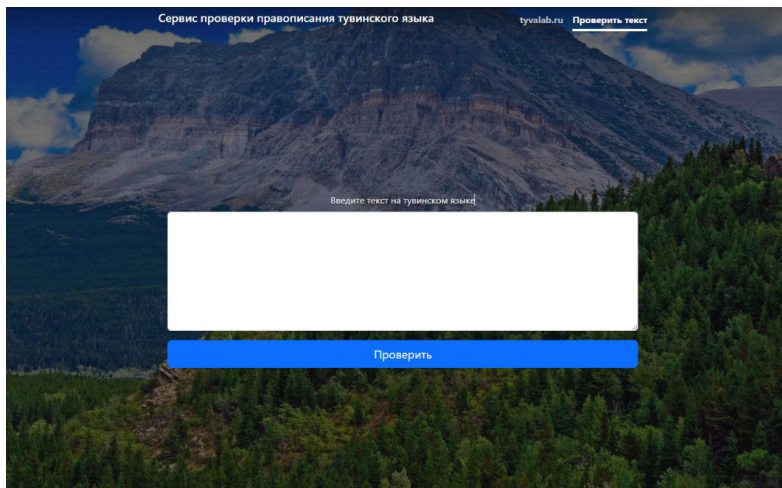


Рис. 4. Главная страница веб-сайта

Model-View-Controller (MVC, «Модель-Представление-Контроллер», «Модель-Вид-Контроллер») – схема разделения данных приложения и управляющей логики на три отдельных компонента: модель, представление и контроллер – таким образом, что модификация каждого компонента может осуществляться независимо.

- **Модель (Model)** предоставляет данные и реагирует на команды контроллера, изменяя своё состояние.
- **Представление (View)** отвечает за отображение данных модели пользователю, реагируя на изменения модели.
- **Контроллер (Controller)** интерпретирует действия пользователя, оповещая модель о необходимости изменений<sup>1</sup>.

<sup>1</sup> Рогачев С. Обобщённый Model-View-Controller [Электронный ресурс] // RSDN – сайт, посвященный разработке программного обеспечения, 23.03.2007 г. URL: <http://rstdn.org/article/patterns/generic-mvc.xml> [дата обращения: 04.10.2023].

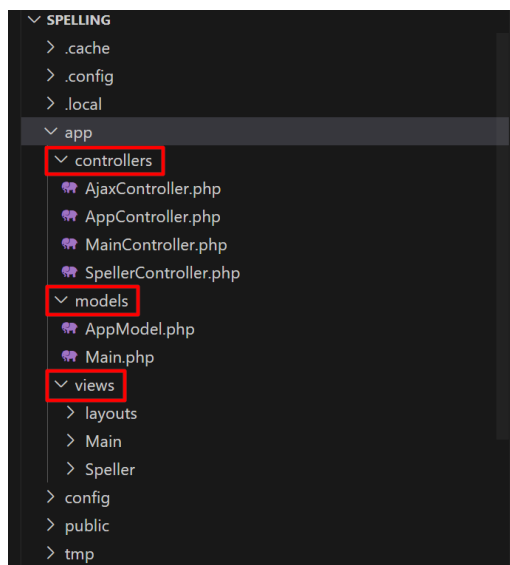


Рис. 5. Структура веб-сайта по принципу MVC

При разработке сервиса, вторым решением было принято подключение сторонней готовой Hunspell-PHP-библиотеки с лицензией открытого и свободного программного обеспечения в [github.com](https://github.com)<sup>1</sup>. Такая библиотека представляет собой оболочку для проверки орфографии Hunspell. На первых порах наш выбор пал на открытый репозиторий [@mekras/php-speller](https://github.com/mekras/php-speller) от программиста Михаила Красильникова из Московской области.

Принцип работ:

1) Сперва размещаем на сервере готовый «aff» и «dic» файлы, чтобы Hunspell-библиотека смогла сравнить введенные пользователем слова.

2) На главной странице (View) пользователь добавляет или вставляет свой текст. После нажатия через HTTP метод POST данный текст отправляется на страницу вывода [tyvalab.ru/spelling/speller](https://tyvalab.ru/spelling/speller).

<sup>1</sup> PHP spell check library [Электронный ресурс] // GitHub.com, 14 апреля 2023 г. URL: <https://github.com/mekras/php-speller> [дата обращения: 04.10.2023].

3) Но перед тем как вывести введенный текст на странице пользователю, то в Контроллере (Controller) этот текст сравнивается с «dic»-словарем с помощью соответствующих объектов StringSource() и Hunspell() из Hunspell-библиотеки.

4) Затем полученный из Hunspell-библиотеки массив с ошибками через цикл приводим к виду «значение|значение|значение», чтобы через функцию preg\_replace() произвести замену текста со специальным html-тегом и css-классом.

5) Если текст не будет содержать ошибок, то массив с ошибками будет пустым.

```
SpellerController.php X
app > controllers > SpellerController.php > PHP Intelephense > SpellerController > indexAction
8 class SpellerController extends ApplicationController
9 {
10     // Метод для индексной страницы
11     public function indexAction(){}
12
13     // Если POST запроса не существует,
14     // то перенаправляем на главную
15     if (!isset($_POST['tyvan_text'])) {
16         redirect('/');
17     }
18
19     // Метаданные в страницу
20     $this->setMeta('Проверка текста');
21
22     // Текст с ошибок на POST запроса
23     $text_check = verify_text($_POST['tyvan_text']);
24
25     $source = new StringSource($text_check);
26     $speller = new Hunspell();
27     $speller->setDictionaryPath('/hunspell/tyv');
28     // $speller->setCustomDictionaries(['мен', 'тыма']);
29     $issues = $speller->checkText($source, ['tyv']);
30
31     // Если при проверке текста были найдены слова с ошибками,
32     // то выделим эти слова
33     if (!empty($issues)) {
34
35         // Пустой массив для выделения текстов с ошибками
36         $array_replace = [];
37
38         $i_array = 0;
39         for($i = 1; $i <= count($issues); $i++) {
40             $array_replace[] = $issues[$i_array]->word;
41             $i_array++;
42         }
43
44         // Приведение массива со значениями в вид: value|value|value
45         // для вставки в preg_replace("#(value|value|value)#i", "<span>\&#160;</span>", $variable)
46         $array_replaceimplode = implode('|', $array_replace);
47
48         // Выполняет поиск и замену по регулярному выражению
49         $text_result = preg_replace("#(" . $array_replaceimplode . ")#i", "<span class='text-error'&#160;</span>", $text_check);
50     }
51
52     // Отправляем массив в страницу
53     $this->set(compact('text_check', 'text_result', 'issues'));
54
55 }
```

Рис. 6. Исходный код класса SpellerController() для обработки текста с ошибками пункта 3 и 4

6) Далее текст выводится на странице с соответствующей html-разметкой.

```

index.php X
app > views > Speller > index.php > ...
 7 <div class="row g-5">
 8
 9 <div class="col-md-5 col-lg-4 order-md-last">
10 <h3 class="d-flex justify-content-between align-items-center mb-3">
11 <span class="text-primary">Найденные ошибки</span>
12 <span class="badge bg-primary rounded-pill"><?>count($issues)</span>
13 </h3>
14
15 <?php
16 // Если массив с ошибками не пустой, то показываем пронумерованный список
17 // Иначе текст сообщения
18 if ( !empty($issues) ) :
19 >
20 <ol class="list-group list-group-numbered mb-3">
21 <?php foreach($issues as $v) :>
22 <li class="list-group-item d-flex align-items-start">
23 <div class="ms-2">
24 <div class="fw-bold"><?>$v->word</div>
25 <?php if ( !empty($v->suggestions) ) :>
26 <small class="fst-italic">Возможные варианты:</small>
27 <div>
28 <?php foreach($v->suggestions as $v2) :>
29 <button type="button" class="btn btn-primary btn-sm mb-1"><?>$v2</button>
30 <?php endforeach ;>
31 </div>
32 <?php endif ;>
33 </div>
34 </li>
35 <?php endforeach ;>
36 </ol>
37 <?php else :>
38 <ul class="list-group mb-3">
39 <li class="list-group-item">
40 <span>Ошибка не найдена</span>
41 </li>
42 </ul>
43 <?php endif ;>
44 </div>
45
46 <div class="col-md-7 col-lg-8">
47 <h4 class="mb-4">Исходный текст</h4>
48 <div id="txtchck">
49 <div class="text">
50 <?>$text_result<?>
51 </div>
52 </div>
53 </div>
54 </div>

```

Рис. 7. Исходный код html-страницы вывода текста с ошибками

## Заключение

Таким образом, в данной статье мы описали процесс разработки системы проверки орфографии тувинского языка на основе Hunspell, который находится в процессе разработки.

Значение системы проверки орфографии трудно переоценить: программы распознавания ошибок и опечаток помогают не только получить грамотный текст, но и ежедневно учит носителей языка правилам правописания, ведь когда нам предлагают правильные варианты написания слов, мы рано или поздно можем их запомнить. И у носителей в целом повышается уровень грамотности лишь благодаря наличию систем проверки правописания.

В перспективе необходимо разработать комбинированную систему проверки правописания тувинского языка на основе корпуса текстов и Hunspell, объединив усилия разработчиков и лингвистов.



## СПИСОК ЛИТЕРАТУРЫ

1. Ондар, Ч. Г., Донгак, В. С., Монгуш, Д. Ш. Тувинский язык в Интернете: представленность, проблемы и перспективы // Новые исследования Тувы. 2023, № 1. С. 186–207. DOI: <https://doi.org/10.25178/nit.2023.1.11>
2. Чемышев, А. В. Подготовка исходных данных для обучения нейросетей (на примере марийского языка) // Сборник материалов «Языковая политика в Российской Федерации». – М., 2022. С. 58–63. 180 с.
3. Kornai, A. (2013) Digital language death // PLoS ONE, 8 (10): e77056. DOI: <https://doi.org/10.1371/journal.pone.0077056>

УДК 81-33

**МОРФОЛОГИЧЕСКИЙ АНАЛИЗ ДЛЯ СБОРА  
ТЕКСТОВЫХ ДАННЫХ В НАЦИОНАЛЬНОМ КОРПУСЕ  
КЫРГЫЗСКОГО ЯЗЫКА*****Т. С. Садыков<sup>1</sup>, Т. Туратали<sup>2</sup>, А. Б. Турдубаева<sup>3</sup>****<sup>1</sup>Бишкекский государственный университет,  
Бишкек, Кыргызстан**<sup>2</sup>Сити Банк, Бишкек, Кыргызстан**<sup>3</sup>КГТУ им. И. Раззакова, Бишкек, Кыргызстан  
tash\_sadykov@mail.ru, timur.turat@gmail.com,  
aida.baktybekovna1212@gmail.com*

В этой статье мы рассмотрим сбор текстовых данных с помощью морфологии и синтаксиса кыргызского языка. Морфология языка изучает взаимодействие с грамматической структуры слов с их функцией в предложении. Как и в других языках, в кыргызском языке существует морфемы, грамматические категории и синтаксические отношения между словами.

**Ключевые слова:** Кыргызский язык, обработка естественного языка, машинный перевод, автоматический анализ текста, набор данных (датасет).

**MORPHOLOGICAL ANALYSIS FOR COLLECTING TEXT  
DATA IN THE NATIONAL CORPORA OF THE KYRGYZ  
LANGUAGE*****Sadykov T. S.<sup>1</sup>, Turatali T.<sup>2</sup>, Turdubaeva A. B.<sup>3</sup>****<sup>1</sup>Bishkek State University, Bishkek, Kyrgyzstan**<sup>2</sup>City Bank, Bishkek, Kyrgyzstan**<sup>3</sup>Kyrgyz State Technical University named after I.Razzakov  
Bishkek, Kyrgyzstan**tash\_sadykov@mail.ru, timur.turat@gmail.com,  
aida.baktybekovna1212@gmail.com*

In this article, we will look at collecting text data using the morphology and syntax of the Kyrgyz language. Morphology of language studies the interaction of the grammatical structure of words with their function in a sentence. Like other languages, the Kyrgyz language has morphemes, grammatical categories, and syntactic relationships between words.

**Keywords:** Kyrgyz language, natural language processing, machine translation, automatic text analysis, dataset.

Кыргызский язык является агглютинативным языком, в котором слова образуется путем добавления деривационных и реля-

цинных морфем к корню или основе слова. Так, например, слово *тоо* (гора) может быть изменено в зависимости от грамматической формы. Добавление морфемы *-нун* формирует форму родительного падежа *тоонун*, а присоединение морфемы *-го* образует форму дательного падежа *тоого*.

Особенность кыргызского языка состоит в том, что для него характерен развитая морфологическая структура. В частности, каждое слово в этом языке может иметь множество форм, которые меняются в зависимости от лексического и грамматического контекста предложения, а именно, существительные могут иметь различные формы категорий падежа, принадлежности и числа, а глаголы – различные залоговые, временные и модальные формы. В этом ракурсе кыргызский язык характеризуется шестью падежами, такими, как основной (атооч жөндөмө), родительный (илик жөндөмө), дательный (барыш жөндөмө), винительный (табыш жөндөмө), местный (жатыш жөндөмө) и исходный (чыгыш жөндөмө) падежа. Каждый падеж имеет свою функцию в предложении. Например, в кыргызском подлежащее и сказуемое могут иметь разные падежи в зависимости от контекста. Кроме того, в глаголах присутствуют грамматические категории рода, числа, времени, наклонения и залога. Например, глаголы в кыргызском образует три временные формы: форм прошедшего, настоящего и будущего времени. Синтаксические отношения между словами в кыргызском языке также могут быть выражены разными способами. Например, порядок слов в кыргызском языке достаточно фиксирован, а изменение его может менять синтаксическую функцию слов в предложении.

Что касается сбора данных, который представляет важным этапом для реализации любого исследования в области переработки текстов на естественном языке, системах машинного перевода и искусственного интеллекта. При этом одним из самых популярных методов сбора данных является алгоритмизация морфологических процессов, участвующих в склонении именных и в спряжении глагольных частей речи. Склонение – это процесс изменения именных слов в случае, когда к ним присоединяются окончания падежа, числа, принадлежности и лица. Спряжение – это процесс изменения глагольных слов в том случае, когда к ним присоединяются окончания наклонения, времени, залога и лица.

Таким образом, учет как склонения именных частей речи, так и спряжения глагольных частей речи является важным этапом при

построении систем морфологического анализа естественного языка с агглютинативной типологией, каковым является кыргызский язык.

Данный набор данных разработан с целью обеспечить более точный анализ текстов на кыргызском языке. Каждая словоформа в датасете содержит информацию о ее морфологических характеристиках, таких как часть речи, падеж, число, род и т.д. и используются следующие теги [Садыков, 2015]:

| Теги-эптекер- tags | Части речи           | Сөз түркүмү  | Parts of speech |
|--------------------|----------------------|--------------|-----------------|
| N                  | имя существительное  | зат атооч    | noun            |
| ADJ                | имя прилагательное   | сын атооч    | adjective       |
| V                  | глагол               | этиш         | verb            |
| ADV                | наречие              | тактооч      | adverb          |
| NUM                | имя числительное     | сан атооч    | numeral         |
| PN                 | местоимение          | ат атооч     | pronoun         |
| CNJ                | союз                 | байламта     | conjunction     |
| POST               | предлог              | жандооч      | postposition    |
| PART               | причастие            | бөлүкчө      | particle        |
| INTRJ              | междометия           | сырдык сөз   | interjection    |
| MOD                | модальное слово      | модалдык сөз | modal word      |
| IMIT               | подражательное слово | тууранды сөз | imitative word  |

### Категории имен существительных атоочтун категориялары

1. Категория числа, имеющая форм единственного и множественного числа. Они выражаются теггами следующих видов:

а) единственное число **SG** - Ø

б) множественное число **PL** – **ЛАр**: -лар -дар -тар

-лер -дер -тер

-лор -дор -тор

-лөр -дөр -төр

2. Притяжательный падеж

а) 1 лицо единственного числа притяжательный **POSS\_1SG** - **[Ы]м**:

-ым -им -ум -үм

-м;

б) 2 лицо единственного числа притяжательный **POSS\_2SG - [Ы]ң:**

-ың -иң -уң -үң

-ң;

в) 2-е лицо единственного числа притяжательный формальный **POSS\_2SGF - [Ы]ң[Ы]з:**

-ыңыз -иңиз -уңуз -үңүз

-ңыз -ңиз -ңуз -ңүз;

г) 3-е лицо единственного числа притяжательный **POSS\_3SG - [с]Ы[н]:**

-ы -и -у -ү -ын -ин -ун -үн

-сы -си -су -сү

-сын -син -сун -сүн;

Притяжательный падеж множественного числа

а) 1-е лицо множественного числа притяжательный **POSS\_1PL - [Ы]б[Ы]з:**

-ыбыз -ибиз -убуз -үбүз

-быз -биз -буз -бүз;

б) Притяжательное местоимение 2 лица множественного числа **POSS\_2PL - [Ы]ң[A]р:**

-ыңар -иңер -уңар -үңөр

-ңар -ңер -ңар -ңөр;

в) 2-е лицо множественного числа притяжательный формальный **POSS\_2PLF - [Ы]ң[Ы]зд[A]р:**

-ыңыздар -иңиздер

-уңуздар -үңүздөр

-ңыздар -ңиздер

-ңуздар -ңүздөр;

г) 3-е лицо множественного числа притяжательный **POSS\_3PL - [с]Ы[н]:**

-ы -и -у -ү

-ын -ин -ун -үн

-сы -си -су -сү

-сын -син -сун -сүн;

3. Падежи

а) Именительный падеж **NOM - ∅**

б) Родительный падеж **GEN - [н]Ын:**

-нын -нин -нун -нүн

-дын -дин -дун -дүн

-тын -тин -тун -түн

-ын -ин -ун -үн;

в) Дательный падеж **DAT** - [Г]А:

-га -ге -го -гө

-ка -ке- ко -кө

-а -е- о -ө;

г) Винительный **ACC** - [н][Ы]:

-ны -ни -ну -нү

-ды -ди -ду -тү

-ты -ти -ту -тү

-ы -и -у -ү;

д) Творительный **LOC** – ДА:

-да -де -до -дө

-та -те -то -тө;

е) Предложный **ABL** - [Д]Ан:

-дан -ден -дон -дөн

-тан -тен -тон -төн

-ан -ен -он -өн;

4. Категория лица

а) 1 лицо единственного числа **1SG** - м[Ы]н:

-мын -мин -мун -мүн;

б) 2 лицо единственного числа **2SG** - с[Ы]ң:

-сың -сиң -суң -сүң;

в) формальный 2-е лицо единственного числа **2SGF** - с[Ы]з:

-сыз -сиз -суз -сүз;

г) 3 лицо единственного числа **3PL** [с][Ы][н]:

∅;

Множественное число

а) 1 лицо множественного числа **1PL** - б[Ы]з:

-быз -биз -буз -бүз;

б) 2 лицо множественное число **2PL** - с[Ы]ң[А]р:

-сыңар -сиңер -суңар -сүңөр;

в) формальное 2-е лицо множественного числа **2PLF** - с[Ы]

зд[А]р:

-сыздар-сиздер-суздар -сүздөр;

г) 3 лицо множественное число **3PL** - [с][Ы][н]:

∅;

5. ПРИЛАГАТЕЛЬНОЕ

а) Сравнительная степень **COMP** - [Ы]раААК:

-ыраак -ирээк -ураак -үрөөк

-раак -рээк -раак -рөөк;

#### 6. ЧИСЛИТЕЛЬНОЕ

а) Порядковые числительные **NUM\_ORD** - [Ы]нчы:

-ынчы -инчи -унчу -үнчү

-нчы -нчи -нчу -нчү;

б) Собирательное числительное **NUM\_COLL** - **ОО[н]**:

-оо -өө

-оон -өөн;

в) Склонение числительных 1 **NUM\_APPR1** – **ЧА**:

-ча -че -чо -чө;

г) Склонение числительных 2 **NUM\_APPR2** – **ДАЙ**:

-дай -дей -дой -дөй

-тай -тей -той -төй;

д) Склонение числительных 3 **NUM\_APPR3** – **ДАГАН**:

-даган -деген -догон -дөгөн

-таган -теген -тогон -төгөн.

#### ГЛАГОЛ

7. Категория залога

а) Основной залог **ACT** -  $\emptyset$

б) Средневозвратный **PASS** - [Ы]л||н:

-ыл -ил -ул -үл

-л;

-ын -ин -ун -үн

-л;

в) Страдательный **REFL** - [Ы]н:

-ын -ин -ун -үн

-л;

г) Косвенно-возвратное 1 **CAUS\_1** – **ДЫР**:

-дыр -дир -дур -дүр

-тыр -тир -тур -түр

д) Косвенно-возвратное 2 **CAUS\_2** – **т**

-т

е) Взаимно-возвратное **RECP** - [Ы]ш:

-ыш -иш -уш -үш

-ш;

8. Наклонение глагола

а) Изъявительное: 1-е лицо единственного числа **HOR\_SG** -

**[А]йЫн:**

-айын -ейин -ойун -өйүн

-йын -йин -йун -йүн;

б) Изъявительное: 1 лицо множественного числа **HOR\_PL - [А||й]ЛЫ[к]:**

- алык -елик -олук -өлүк
- йлык -йлик -йлук -йлүк;
- алы -ели -олу -өлү
- йлы -йли -йлу -йлү;

в) Повелительное: 2 лицо единственного числа **IMP\_SG – ГЫн:**

- гын -гин -гун -гүн
- кын -кин -кун -күн;

г) Повелительное: второе лицо множественного числа **IMP\_PL – ГЫЛА:**

- гыла -гила -гула -гүла
- кыла -кила -кула -күла;

д) Изъявительное: 2-е лицо единственного числа формальный **IMP\_SGF - [Ы]ңЫз:**

- ыбыз -ибиз -убуз -үбүз
- быз -биз -буз -бүз;

е) Повелительное: 2-е лицо множественного числа формальный **IMP\_PLF - [Ы]ңЫзДар:**

- ыңыздар-иңиздер-унуздар-үнүздөр
- ңыздар -ңиздер -ңуздар -ңүздөр;

ё) Повелительное: 3-е лицо единственного числа **JUS\_SG – сЫн:**

- сын -син -сун -сүн;

ж) Повелительное: 3-е лицо множественного числ **JUS\_PL - [Ыш]сЫн:**

- ышсын -ишсин -ушсун -үшсүн
- сын -син -сун -сүн;

з) **PREC\_1 – чЫ:**

- чы -чи -чу -чү;

9. Времена глаголов

а) Настоящее **PRES - [А||й]:**

- а -е -о -ө

-й;

б) Прошедшее определенное **PST\_DEF – ДЫ:**

- ды -ди -ду -дү

-ты -ти -ту -тү;

в) Прошедшее неопределенное **PST\_INDF - ГА[н]:**

- ган -ген -гон -гөн



-кан -кен -кон -көн;

-га -ге -го -гө

-ка -ке -ко -кө;

г) Прошедшее **PST\_EVID** - [Ы]п[тыр]:

-ыптыр -иптир -уптур -үптүр

-птыр -птир -птур -птүр;

-ып -ип -уп -үп

-п;

д) Прошедшее **PST\_ITER** – чУ:

-чу -чү;

е) Будущее определенное **FUT\_DEF** - А||й:

-а -е -о -ө

-й;

ё) Будущее неопределенное **FUT\_INDF** - [А]р

-ар -ер -ор -өр

-р;

ж) Будущее неопределенное отрицательное **FUT\_INDF\_NEG**

- БAc

-бас -бес -бос -бөс

-пас -пес -пос -пөс;

10. Причастия

а) Настоящее **PCP\_PR** - [УУ]чУ

-уучу -үүчү

-чу -чү;

б) Прошедшего времени **PCP\_PS** - ГAn

-ган -ген -гон -гөн

-кан -кен -кон -көн;

в) Будущее **PCP\_FUT\_DEF** - [А]р

-ар -ер -ор -өр

-р;

г) Будущее отрицательное **PCP\_FUT\_NEG** - БAc

-бас -бес -бос -бөс

-пас -пес -пос -пөс;

11. Наречия

а) **ADVV\_ACC** - [Ы]п:

-ып -ип -уп -үп

-п;

б) **ADVV\_CONT** - А||й:

-а -е -о -ө

-й;

- в) **ADV\_V\_INT – ГАнЫ:**  
 -ганы -гени -гону -гөнү  
 -каны -кени -кону -көнү;
- г) Наречное (отрицательная форма) **ADV\_V\_NEG – ГЫЧА:**  
 -гыча -гиче -гуча -гүчө  
 -кыча -киче -куча -күчө;
- д) Наречное (последовательное значение) **ADV\_V\_SUC – ГЫЧА:**  
 -гыча -гиче -гуча -гүчө  
 -кыча -киче -куча -күчө;
- е) Наречное (ограничивающий) **ADV\_V\_SUC – ГАНЧА:**  
 -ганча -генче -гончо -гөнчө  
 -канча -кенче -кончо -көнчө.
12. Отглагольные существительные
- а) Инфинитив1 **INF\_1 – ОО:**  
 -оо -өө;
- б) Инфинитив2 **INF\_2 – УУ:**  
 -уу -үү;
- в) Инфинитив3 **INF\_3 - [Ы]ш:**  
 -ыш -иш -уш -үш  
 -ш;
- г) Инфинитив4 **INF\_4 – МАГ:**  
 -мак -мек -мок -мөк  
 -маг -мег -мог -мөг;
- д) Инфинитив5 **INF\_5 – ГЫ:**  
 -гы -ги -гу -гү  
 -кы -ки -ку -кү;
13. Модальная форма
- а) Условный **COND – сА:**  
 -са -се -со -сө;
- б) Желательный (намерение) **DESIDE - МАк[ЧЫ]:**  
 -макчы-мекчи-мокчу-мөкчү  
 -мак -мек -мок -мөк;
- в) Желательный1 **OPT - ГЫ+POSS келет||келди:**  
 -гы -ги -гу -гү  
 -кы -ки -ку -кү;
- г) Желательный2 **OPT - ГАй эле+PERS:**  
 -гай -гей -гой -гөй  
 -кай -кей -кой -көй;
- д) Премонитив **PREM - БАГАй эле+PERS:**

-багай -бегей-богой -бөгөй  
-пагай-пегей-погой-пөгөй;

Актуальность темы в том, что морфологический анализ языка играет важную роль в обработке естественного языка в машинном обучении. Например, он может быть использован для автоматического разбора предложений на части речи и определения их грамматических характеристик. Это может быть полезно при создании программ для машинного перевода или для автоматического анализа текста. Кроме того, может помочь улучшить качество поисковых систем, помогая в определении смысла запроса пользователя. Например, если пользователь ищет «книги про Кыргызстан», поиск может использовать знание о том, что «книги» – это существительное во множественном числе и «про Кыргызстан» – это предлоговая группа, чтобы предложить более точные результаты. В целом, имеет широкий потенциал для использования в обработке естественного языка. Наш датасет может быть использован для создания более точных систем распознавания речи, так как он в дальнейшем будет определять правильную форму слова в контексте предложения.

Собранный нами датасет в открытом доступе на GitHub. Он разделен на тестовую и тренировочную выборки для обучения и проверки модели машинного обучения. Это важный шаг в разработке модели, так как он позволяет оценить ее эффективность на новых данных, которые модель ранее не видела.

В тестовой выборке содержится 11565 строк данных, а в тренировочной выборке содержится 25259 строк данных, которые были предварительно обработаны и приведены к одному формату. Мы также провели анализ и очистку данных для удаления выбросов, дубликатов и пропущенных значений. Мы надеемся, что наш датасет будет полезен и для других исследователей, а также для разработчиков, которые будут использовать эти данные в своих проектах.

Ссылка на датасет на GitHub: <https://github.com/Aida-eje/kyrgyz-language-dataset-2>

## ЛИТЕРАТУРА

1. Т. Садыков, Г.Э. Жумалиева, М.Ж. Түмөнбаева, Б. Шаршембаев «КЫРГЫЗ ТИЛИ: компьютердик лингвистиканын негиздери», Бишкек – 2015;
2. Т. Абдиев “КОНСТРУКЦИИ С КАУЗАТИВНЫМИ ГЛАГОЛАМИ В КИРГИЗСКОМ ЯЗЫКЕ”, Бишкек – 2009;
3. К. Сейдакматов “Кыргыз тилинин кыскача этимологиялык сөздүгү” “илим” БАСМАСЫ, Фрунзе 1988;
4. И. Абдувалиев “КЫРГЫЗ ТИЛИНИН МОРФОЛОГИЯСЫ”, Бишкек-2008;
5. С. Үсөналиев “Кыргыз тилинин справочниги”, “Турар” БАСМАСЫ, Бишкек-2010;

---

# КОРПУСНАЯ ЛИНГВИСТИКА И КОРПУСНЫЕ ИССЛЕДОВАНИЯ

УДК 81'322.2

## ISSUES OF KYRGYZ SYNTACTIC ANNOTATION WITHIN THE UNIVERSAL DEPENDENCIES FRAMEWORK

*Aida Kasieva<sup>1</sup>, Gulnura Dzhumalieva<sup>1</sup>, Anna Thompson<sup>2</sup>,  
Murat Jumashiev<sup>3</sup>, Bermet Chontaeva<sup>4</sup>, Jonathan Washington<sup>5</sup>*

*<sup>1</sup>Kyrgyz\_Turkish Manas University, Bishkek, Kyrgyzstan,*

*<sup>2</sup>Independent researcher, Leeds, UK*

*<sup>3</sup>Independent researcher, Bishkek, Kyrgyzstan*

*<sup>4</sup>Universität Tübingen, Tübingen, Germany,*

*<sup>5</sup>Swarthmore College, Swarthmore, USA,*

*aida.kasieva@manas.edu.kg, gulnur.jumalieva@manas.edu.kg,*

*thompsonannad@gmail.com, jumashieff@gmail.com,*

*bermet.chontaeva@student.uni-tuebingen.de,*

*jonathan.washington@swarthmore.edu*

This paper examines key issues encountered in syntactic annotation work for a forthcoming Universal Dependencies (UD) corpus of Kyrgyz. It presents an overview of the corpus creation process, including sentence sampling from the Manas-UdS Kyrgyz corpus and manual annotation using UD guidelines. The corpus contains over 1600 tokens across 230 sentences sampled from literary and news domains. Four central issues in Kyrgyz UD annotation are then discussed in-depth: copula tokenization, categorization of “small words” like *да* and *керек*, null-headed clauses (including relative clauses, and -DAGI and -NIKI constructions), and differentiating inflection vs. derivation. For each issue, multiple analysis options are weighed, including contrasting the approach in prior Turkic UD treebanks. Copula analysis compares subject agreement morphology as dependent subtokens vs independent words. The discourse and intensifier functions of *да* are examined to determine optimal POS and dependency labels. Strategies for representing implicit nominal heads in relative clauses and genitive constructions are evaluated. Criteria for categorizing productive derivational morphology as inflectional cases vs separate words are outlined. Throughout, examples illustrate annotation decisions and dependency graphs. Comparisons are made to the analysis of related phenomena in existing UD treebanks for Kazakh [Tyers & Washington 2015, Makazhanov et al. 2015], Turkish, and the small Kyrgyz UD corpus [Benli, 2023]. The work identifies ongoing challenges in representing Kyrgyz syntax within UD, while developing an improved annotated resource. It highlights issues where UD guidelines exhibit limitations for Turkic languages, providing analysis to advance understanding of best practices for Kyrgyz and related languages.

**Keywords:** Kyrgyz, syntax, annotation, Universal Dependencies.

**ПРОБЛЕМЫ КЫРГЫЗСКОЙ СИНТАКСИЧЕСКОЙ АННОТАЦИИ  
В ФРЕЙМВОРКЕ UNIVERSAL DEPENDENCIES**

*Касиева А.<sup>1</sup>, Джумалиева Г.<sup>1</sup>, Томпсон А.<sup>2</sup>, Юмашев М.<sup>3</sup>,  
Чонтаева Б.<sup>4</sup>, Джонатан В.<sup>5</sup>*

<sup>1</sup>*Кыргызско-Турецкий университет Манас, Бишкек, Кыргызстан,*

<sup>2</sup>*Независимый исследователь, Лидс, Великобритания*

<sup>3</sup>*Независимый исследователь, Бишкек, Кыргызстан*

<sup>4</sup>*Тюбингенский университет, Тюбинген, Германия,*

<sup>5</sup>*Суортморский колледж, Суортмор, США,*

*aida.kasieva@manas.edu.kg, gulnur.jumalieva@manas.edu.kg,*

*thompsonannad@gmail.com, jumasheff@gmail.com,*

*bermet.chontaeva@student.uni-tuebingen.de,*

*jonathan.washington@swarthmore.edu*

В данной статье рассматриваются основные вопросы, возникающие в процессе работы над синтаксической аннотацией для разрабатываемого корпуса кыргызского языка на базе универсальных зависимостей (УЗ). Представлен обзор процесса создания синтаксического корпуса, предложения для которого отобраны из корпуса кыргызского языка “Manas-UdS”. Синтаксическая аннотация корпуса выполняется вручную в соответствии с рекомендациями УЗ. Данный корпус содержит более 1600 токенов в 230 предложениях, отобранных из текстов художественных произведений и новостей. Подробно рассматриваются четыре основные проблемы аннотирования УЗ кыргызского языка: токенизация копулы (глаголы-связки), категоризация «служебных слов», таких как “да” и “керек”, предложения с отсутствующим главным элементом (null-head), включая относительные предложения, конструкции с -ДАГЫ и -НЫКЫ, а также разграничение между флексией и деривацией. Для каждого случая рассматриваются различные варианты анализа, включая сравнение подходов, применяемых в существующих тюркских УЗ-трибанках. Анализ копулы позволяет сравнить морфологию согласования подлежащего в качестве зависимых субтокенов с независимыми словами. Функции дискурса и усилителя “да” рассматриваются для определения подходящих для него частей речи и соответствующих аннотаций зависимостей. Оцениваются стратегии представления имплицитных номинативных элементов в относительных предложениях и генитивных конструкциях. Изложены критерии, определяющие, следует ли классифицировать продуктивную деривационную морфологию как случай словоизменения или как отдельные слова. Примеры иллюстрируют используемые модели аннотаций и диаграмм зависимостей. Проводятся сравнения анализа подобных случаев в существующих УЗ трибанках для казахского [Tuers & Washington 2015, Makazhanov, 2015], турецкого и небольшого корпуса УЗ для кыргызского языка [Benli, 2023]. В работе выявляются проблемы, связанные с представлением синтаксиса кыргызского языка в рамках УЗ при разработке улучшенного варианта аннотированного ресурса. Наряду с выявлением моментов, в которых руководство по УЗ демонстрирует ограничения в рекомендациях по использованию

УЗ для тюркских языков, также предлагаются более оптимальные варианты анализа для кыргызского и смежных языков.

**Ключевые слова:** Кыргызский язык, синтаксис, аннотация, универсальные зависимости

## I. Introduction

This paper examines issues that have arisen in syntactic annotation work for a forthcoming Universal Dependencies (UD) corpus of Kyrgyz, a Turkic language of Central Asia. We lean on prior research on Kyrgyz syntax and existing UD corpora of Turkic languages as a foundation, and use existing Kyrgyz textual analysis tools and UD annotation tools for our work. Manually annotated syntactic data is an invaluable resource for understanding the grammatical patterns and constructions of a language.

The creation of a UD Kyrgyz treebank will support more in-depth investigation into the syntax and morphology of Kyrgyz within a cross-linguistically consistent framework. In addition, annotated syntactic data is essential for developing accurate natural language processing tools like part-of-speech taggers and parsers. The release of a high-quality, comprehensive UD treebank for Kyrgyz will fill a crucial gap, enabling the training of NLP models for syntactic and morphological analysis, machine translation, information retrieval and more. Currently the resources available for syntactically parsing Kyrgyz text are limited. This work seeks to address this need and provide a valuable annotated dataset that can serve as training and evaluation data for Kyrgyz language technologies.

Our treebank draws sentences from a broader range of domains contained in the Manas-UdS Kyrgyz corpus. Second, we provide an in-depth analysis of major syntactic issues that have arisen during the annotation of copula tokenization, null-headed clauses, and differentiating derivation from inflection. We extensively compare potential solutions for each issue, weighing the probability of their occurrences in Kyrgyz. Third, we contrast our analysis and annotation decisions with those made previously, particularly the existing UD treebanks for Kyrgyz [Benli, 2023]. The creation of a larger treebank and thorough examination of ongoing annotation challenges advances understanding of applying UD conventions to Kyrgyz and builds a higher quality resource to support future parsing and NLP applications.

Section 2 discusses background on UD, its application to Turkic languages, and syntactic research into Kyrgyz. Section 3 overviews the annotation work of the authors to date.

In Section 4, a range of issues encountered in annotation are discussed. These include copula tokenisation, the treatment of difficult-to-categorise «small» words, null-headed clauses (including relative clauses, and *-DAGI* and *-NIKI* constructions), and decisions regarding inflection versus derivation. Section 5 concludes and proposes future work.

## II. Background

The Universal Dependencies project aims to develop cross-linguistically consistent treebank annotation for many of the world's languages [Nivre, 2016]. It represents predicate-argument structure through labeled dependency parses, providing common guidelines for annotation across languages. The consistency enables cross-linguistic learning and analysis. The sentences in the US are represented through directed acyclic graphs, with words as nodes and grammatical relations as labeled edges. UD guidelines strive for consistency across languages, while allowing language-specific extensions; the quality and coverage of UD resources varies across languages.

Several UD treebanks have been developed for Turkic languages, including Turkish, Uyghur, and Kazakh, as well as a recent treebank for Kyrgyz.

Tyers and Washington [Tyers, 2015] describe the development of the first free and open-source dependency treebank for Kazakh, which they released using UD v1 annotation standards. At the time of publication, the treebank contained 402 sentences from open-source and public domain texts to ensure free availability and extensibility (it is now larger). The texts were first morphologically analysed and disambiguated using existing resources for Kazakh [Washington et al. 2014], and were then manually annotated for dependency syntax. The authors further discuss several linguistic issues in Kazakh focusing on their analysis in UD, including functions of case morphemes, derivations, non-finite clauses, and copulas. The decisions of annotation are outlined, like marking the copula as a dependent and last conjunct as the head in coordination. Verbal nouns, adjectives, and adverbs are annotated for their functions as subjects, modifiers, or clausal complements. Their preliminary parsing experiments showed 63.9% LAS and 74.7% UAS with structural features, comparable to other small treebanks. This treebank has since been converted to be in line with UD v2.



Makazhanov et al. [Makazhanov, 2015] conducted their study based on 300 sentences randomly selected from the closed source Kazakh Language Corpus [Makhambetov et al. 2013]. Their work on syntactic annotation revealed several challenges. First, they had difficulty categorizing the analytic negation markers *жоқ* and *емес*, ultimately opting to classify them as copulas. Second, their dataset did not contain non-relative (acl:relcl) examples of clausal noun modifiers, resulting in annotations with no specified clausal noun modifier relation. Lastly, they faced challenges in ensuring accurate dependency relations, particularly in complex phrases like *үлкен үй-дегілер* ‘those in the big house’. In such cases, directly attaching the adjective ‘big’ to ‘those in the house’ led to a misrepresentation of the intended meaning, highlighting the need for consistent tokenisation and annotation conventions.

Tyers et al. [Tyers, 2017] present an early assessment of UD guidelines for Turkic languages. They highlight areas of cross-linguistic consistency, and note discrepancies between guidelines for Turkish, Kazakh, and Uyghur. Open issues discussed include tokenization, differentiating core arguments, complex predicates, and copula usage. Our work builds on their assessment, tackling similar issues for our Kyrgyz UD treebank.

Aili et al. [Aili, 2018] took steps to extend Universal Dependencies (UD) resources to Uyghur. They mapped the treebank’s labeling scheme to UD labels, making structural changes like marking auxiliaries and copulas as dependents. Some UD relations were introduced for Uyghur-specific syntax like modifier emphasis (advmod:emph). Aili et al. also defined new labels needed to represent complex Uyghur structures concisely within UD, including compound reduplication (compound:redup). Their work demonstrates both adapting UD’s universal principles to Uyghur and extending UD conventions as required for the language.

Four treebanks have also been published for Turkish, along with a number of academic papers associated with them [e.g., Sulubacak et al. 2016, Çetinoğlu and Çöltekin 2022].

Sulubacak et al. [Sulubacak, 2016] converted the IMST Treebank (Turkish), originally available in the CoNLL-X data format, to the CoNLL-U format in compliance with UD standards. Utilizing the Inflectional Group (IG) formalism [Oflazer 1999; Hakkani-Tür et al. 2002], the authors segment orthographic tokens into morphosyntactic words at derivational boundaries. They provide comprehensive

mapping rules for converting both morphological features and dependency relations to align with UD standards. The paper also discusses the challenges of annotating non-projective sentences, which led to a slight drop in labeled attachment scores. The authors test their methodology on the UD version of the IMST Treebank, providing valuable metrics on its effectiveness, achieving a labeled attachment score (LAS) of 81.41% and an unlabeled attachment score (UAS) of 85.48%.

Çetinoğlu and Çöltekin [Çetinoğlu, 2022] present the Turkish-German SAGT code-switching treebank. It contains rich linguistic annotations including language IDs, lemmas, POS tags, features, and dependency relations. The SAGT treebank is one of the few publicly available resources for studying code-switching between German and Turkish. Special care was taken during annotation to handle multilingual consistency and informal language. Features like CSID indicate code-switching type (intra-lexical, intrasentential). Data was collected from conversations between 20 Turkish-German bilingual students and annotated using UD monolingual treebanks. The treebank comprises 2,184 sentences and 37,233 tokens after segmentation. Most annotation differences result from divergent grammatical traditions, not linguistic discrepancies. Challenges identified include consistent multilingual annotation and informal language. Proposed solutions involve tailored guidelines, multiple annotation layers, and contextualized annotation. Iteratively identifying and fixing errors is important since code-switching complexity produces more annotation errors than monolingual treebanks. Overall, this paper introduces an invaluable annotated resource to spur advances in code-switching analysis.

A UD corpus has also been released for Tatar [Taguchi 2022].

These existing UD corpora provide a useful starting point, as models. However, many open questions remain regarding UD annotation for Turkic languages.

As an agglutinative Turkic language, Kyrgyz exhibits flexibility in its word order, including both head-initial and head-final structures. Kyrgyz syntax has been the focus of few studies, with some examination of relative clauses [Imanalieva 2015] and other constructions. The Universal Dependencies framework aims to represent syntactic variation across diverse languages. While UD has some bias toward head-initial order, it can also model head-final structures where attested. Capturing the word order variation found in

Kyrgyz thus presents an interesting test case for dependency annotation under UD guidelines. Further research will be valuable for assessing how well UD accommodates the syntactic patterns of Kyrgyz.

Kyrgyz, as a morphologically rich language, pushes the limits of the guidelines for phenomena like non-canonical word order and complex predicate formations [Thompson 2021].

The Kyrgyz language currently has limited syntactic resources available in the UD framework. As of the UD v2.12 release, the recently added Kyrgyz UD treebank [Benli 2023] contains only 781 sentences and its domain is mainly news headlines and stories selected from Kyrgyz novels and news websites. Details of annotation decisions are not discussed in depth.

Dzhumalieva et al. [Dzhumalieva, 2023] investigate the challenges and opportunities of syntactic annotation for the Kyrgyz language within the UD framework. They propose the adaptation of relevant terminology into Kyrgyz and outline their initial steps in manual tagging of tokens, lemmas, and POS-tags, laying the groundwork for future automated Natural Language Processing tasks. A central focus is syntactic analysis and treebank annotation using the Universal Dependencies framework.

Musazhanova et. al [Musazhanova, 2023] discuss an effort in the syntactic annotation of the Kyrgyz language using UD. The paper offers annotation examples of the Kyrgyz sentences and reveals that the Kyrgyz language's grammatical categories haven't been fully explored within the UD framework. The syntactic analysis examples provide insight into adapting Universal Dependencies standards for Kyrgyz. It highlights a significant gap in computational linguistics for Kyrgyz and lays the groundwork for future research on annotated corpus development.

Washington et al. [Washington, 2012] present a finite-state morphological transducer for Kyrgyz. At publication, the lexical foundation covered over 8,000 stems across major word classes; it now covers over 15,000 stems. While intended for machine translation, the transducer may also be used to aid morphological analysis for syntactic parsing. Our UD annotation experience suggests that this resource may need extension to handle complex verbs and other phenomena.

While valuable related work exists on Turkic and specifically Kyrgyz UD, many issues persist in developing high-quality UD-annotated resources for Kyrgyz. Our paper aims to advance understanding of these syntactic annotation challenges through analysis

of a larger, more diverse UD Kyrgyz treebank. To this end, we take steps towards creating an improved Kyrgyz UD corpus through manual annotation of new data.

### **3. Corpus Development**

This section describes our corpus creation process, including sentence sampling, and annotation workflow, as well as corpus statistics and metadata. It also presents the current state and future plans for the corpus.

For completion of undergraduate linguistics coursework, Thompson [Thompson, 2021] completed a thesis on the relationship between syntactic structure and syntactic parallelism using 85 randomly selected Kyrgyz proverbs, building on work done previously for a course project in Washington’s Structure of Kyrgyz course. The workflow began with the analysis of the proverbs using the Apertium Kyrgyz morphological transducer [Washington, 2012]. The proverbs were then manually annotated using Universal Dependencies guidelines. This allowed for analysis of the corpus of proverbs to engage with the common terms and categories from previous research about syntactic parallelism and proverb structure. The analysis categorized proverb syntax and identified patterns, like the association between parataxis relations and syntactic parallelism. This corpus of 85 proverbs was made publicly available under an open-source license, constituting the first freely available dependency-annotated corpus of Kyrgyz, and constitutes part of our corpus as well.

Building on this work, Kasieva and Dzhumalieva, along with their students in the Translation department of Kyrgyz-Turkish Manas University, extracted and manually annotated sentences from the 2M-word Manas-UdS Kyrgyz corpus [Kasieva et al. 2020], which was compiled from Kyrgyz literary works and the state newspaper “Erkin-Too.”<sup>1</sup> The literary portion was drawn from 12 short stories and 3 novels, selecting sentences with a range of syntactic constructions. The news portion sampled sentences from 15 articles spanning different topics. Care was taken to extract sentences covering diverse lexical content and syntactic phenomena.

New Kyrgyz sentences were manually annotated using the UD Annotatrix interface [Tyers et al. 2018]. Annotations were completed by various combinations of the authors of this paper, often building

---

<sup>1</sup> <https://erkin-too.kg/>

on student work. The open-source tool UD Annotatrix provides an interface designed for UD annotation and validation. Annotators followed the UD guidelines, referring to prior analyses of Kyrgyz structures [Thompson, 2021]. Disagreements were discussed and resolved to reach consensus. Inter-annotator agreement was over 90% by the end of the process.

Designed for manual annotation of Universal Dependencies (UD), UD Annotatrix was a valuable asset for creating this Kyrgyz syntactic corpus. The tool handles annotation guidelines such as two-level segmentation schemes, and provides validation feedback. It allows for customization of guidelines specific to the Kyrgyz language and lists and auto-completes language features (e.g., POS and dependency relations). Linguists working with Kyrgyz can utilise its built-in features, including automated parsing and dependency visualization, for a streamlined annotation process. This ensures the creation of a comprehensive and consistent Kyrgyz syntactic corpus.

The resultant corpus contains 2456 tokens across 332 sentences. Sentence lengths range from 5 to 35 tokens, with an average of 14 tokens. The vocabulary size is 829 unique words. Morphological features and universal POS tags were applied using the morphological transducer developed by Washington et al. [Washington, 2012].

Work is underway to finalize annotation, add detailed metadata, expand the corpus size, and prepare submission to the Universal Dependencies project.<sup>1</sup> In future work, we hope to increase the domain diversity by sampling scientific articles, spoken dialogues, and social media text. This will provide a robust annotated corpus to support NLP research on the Kyrgyz language.

#### 4. Issues of interest

This section identifies challenges in applying UD to Kyrgyz by presenting recurring issues that have arisen during annotation. Not all challenges encountered are discussed in this paper.

In our data, we encountered challenges like copula tokenisation (4.1), the treatment of difficult-to-categorise «small» words (4.2), and null-headed clauses (including relative clauses, and *-DAGI* and *-NIKI* constructions) (4.3). We also had to make decisions regarding inflection versus derivation (4.4).

---

<sup>1</sup> The corpus is currently available at <https://github.com/apertium/apertium-kir/tree/main/corpora>.

#### 4.1. Copula tokenisation

In Kyrgyz there are several strategies used to form copula sentences.

In non-past-tense copula sentences, the normal strategy is to add a subject agreement morpheme to the predicate, as in (1).

- (1) *Мен сенин үйүңдөмүн.*  
 men senin üy-(I)η-DA-MIn.  
 I your house-POSS.2SG-LOC-COP.NPST.1SG  
 ‘I m at your house.’

In the past direct tense, an apparently irregular form of a defective verb *э-* is used, as in (2).

- (2) *Мен сенин үйүңдө элем.*  
 men senin üy-(I)η-DA ele-m.  
 I your house-POSS.2SG-LOC COP.PST.DIR-1SG  
 ‘I was at your house.’

There is additionally an irregular-looking past verbal noun form of *э-*: *экен* (cf. expected \**эген*), as shown in (3).

- (3) *Мен сенин үйүңдө экенимди билиптирсиң.*  
 men senin üy-(I)η-DA eken-(I)m-NI bil-(I)ptIr-sIŋ.  
 I your house-POSS.2SG-LOC COP.VN-POSS.1SG-ACC  
 know-PST.IDF-2SG

‘You knew that I was at your house.’

No other forms of this defective verb exist; missing forms include various tenses and non-finite forms, as well as negative forms.

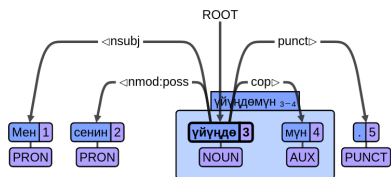
We treat the non-past copula subject agreement morphemes as if they were cliticised forms of the defective copula verb, with lemma *э* and POS tag AUX, and as a subtoken of the space-delimited «word» that they are part of. Despite the fact that they have an unrelated etymology from the defective copula verb, there are several reasons we believe this approach is advantageous:

1. The defective copula verb does not have non-past forms. This approach allows the non-past agreement morphemes to fill that gap in the paradigm.

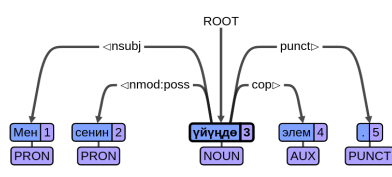
2. This approach allows non-past and direct past to have similar analyses, as shown in Graphs 1 and 2.

3. This approach prevents the problem of having multiple person/number/formality marking on a singular noun, as otherwise would be necessitated in (1).

4. This approach allows the morphemes to be labelled as copula.



Graph 1. UD graph of sentence (1) depicting a non-past copula construction



Graph 2. UD graph of sentence (2) depicting a direct past copula construction

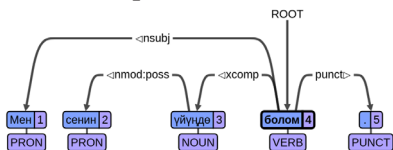
For the sake of consistency, we have chosen to analyse a separate copula subtoken even in the third person (singular and plural), where it has no orthographic content. It would be possible to leave this subtoken out of the annotation, as Tyers and Washington [Tyers, 2015] did for Kazakh, but it would have the disadvantage of then having no indication of subject agreement.

In parts of the paradigm of э- where forms are non-existent, the verb бол- ‘be, become’ is used instead. In fact, the verb бол- can be used in certain contexts in place of forms of э-. For example, sentences (4) and (5) can have the same meanings as (1) and (2), respectively.

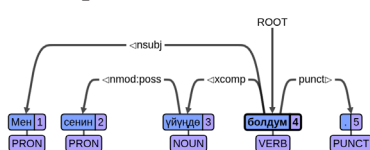
(4) Мен сенин үйүңдө болом.  
 men senin üy-(I)ŋ-DA бол-Е-м.  
 I your house-POSS.2SG-LOC be-NPST-1SG  
 ‘I am (/will be) at your house.’

(5) Мен сенин үйүңдө болдум.  
 men senin üy-(I)ŋ-DA бол-DI-м.  
 I your house-POSS.2SG-LOC be-PST.DIR-1SG  
 ‘I was at your house.’

While these are essentially copula constructions, бол- is a regular non-defective lexical verb, and so we treat it as such. In annotation of these sentences, then, бол- is annotated as a VERB, and the predicate as an xcomp dependent of it, as shown in Graphs 3 and 4.



Graph 3. UD graph of sentence (4) depicting a non-past бол- verbal construction



Graph 4. UD graph of sentence (5) depicting a direct past бол- verbal construction

This has the disadvantage of semantically and morphologically very similar structures being treated as having different syntax. Tyers and Washington [Tyers, 2015] opt instead to treat *бол-* constructions the same as *э-* copula constructions.

Benli [Benli, 2023] annotates the following types of copula constructions in the following ways:

- When subject-agreement morphemes occur on non-verbal predicates, the noun or adjective comprising the final word in the predicate is analysed as having person features of the subject, and is sometimes misanalysed as a verb.

- Forms of *эле* are given the POS tag VERB and are treated as compound:svc (elements of serial-verb constructions) dependents of the non-verbal predicate.

- Complements of *бол-* are analysed as amod dependents.

It is not clear what the reasoning for these analyses might be.

## 4.2. «Small» words

This section addresses the analysis of several «small» words that originally presented difficulties for annotation: *да* (§4.2.1), *эле* (§4.2.2), *бар* and *жок* (§4.2.3), and *керек* (§4.2.4).

### 4.2.1. *да*

In Kyrgyz, there are several distinguishable uses of the word *да*, likely constituting several distinct lexical words. These are the uses, as delimited by the authors:

**1. Post-predicate «modal particle».** In this use, *да* indicates that the speaker(s) is making a statement whose truth value they believe to be evident to the interlocutor(s), but which needs to be asserted to explain something else. An example of this from the corpus is given in (6).

(6) *Натыйжалар жарыяланыптыр да, ээ?*

‘The results have been announced, haven’t they?’

**2. Conditional intensifier.** In this use, *да* adds intensity to a conditional adverbial clause, translating to English roughly as «even» in uses as «even if». An example of this from the corpus is given in (7).

(7) *Оозу кыйшык болсо да, байдын уулу сүйлөсүн.*

‘Even if his mouth is crooked, let the rich person’s son speak.’

**3. General contrastive intensifier.** In this use, *да* adds a contrastive focus to the preceding element, which may constitute a wide range of phrase types. It can be translated to English as «even». An example of this from the corpus is given in (8).



(8) *Тамашада да чындыктын үлүшү бар.*

‘Even in a joke is some element of truth.’

**4. General conjoining adverb.** In this use, *да* adds the sense that what is being said about the preceding phrase is true in addition that same situation regarding a parallel phrase. It can be translated to English as «also» or «too». This meaning and the preceding one may often both be interpreted in a single example. An example of this from outside the corpus is given in (9).

(9) *Атам да каршы болду.*

‘My father was also against it.’ (or: ‘Even my father was against it.’)

**5. Correlative conjunction.** In this use, *да* is used twice, with two parallel phrases, to conjoin them, translating to English as «both ... and». An example of this from outside the corpus is given in (10).

(10) *Атам да, апам да каршы болду.*

‘Both my father and my mother were against it.’

The meanings and distributions of many of these uses are similar. While the first use has a very distinct meaning and distribution, meanings 2 and 3 are very similar, as are 3 and 4; additionally, meaning 5 seems like it could be understood as a repeated use of 4, or possibly 3.

Benli [Benli, 2023] analyses *да* in all uses as a coordinating conjunction (CCONJ), attached to its head with a mark dependency. According to Universal Dependencies guidelines [Zeman, 2023], however, coordinating conjunctions conjoin two syntactic constituents with no subordination relationship, and mark is the dependency for a word that is used to subordinate one clause to another. Neither of these types of relationships hold in any of these examples.

In Kazakh, the first use of *да* does not exist, and an additional use to conjoin two parallel constituents is found (e.g., *Астана елімізге қайырлы да құтты қала болды*). As for the remaining uses, Tyers & Washington [Tyers, 2015] and Makazhanov et al. [Makazhanov, 2015] mostly annotate these as ADV, with an advmod dependency on not the preceding element, but the root. For example, in (4), *да* ‘also’ would be an advmod dependency on *болду* ‘was’, as opposed to *атам* ‘my father’. Given the analysis as an adverb, this dependency attachment is somewhat sensible, as the general guidelines for advmod state that it indicates modifier to a predicate or modifier word.

However, the subtype advmod:emph appears to be dedicated specifically to indicating an intensifier or emphasising word that can modify various parts of speech, including nouns and prepositional

phrases. This dependency relation is used in Tatar [Taguchi 2022] and Turkish<sup>1</sup> treebanks for uses similar to *da*.

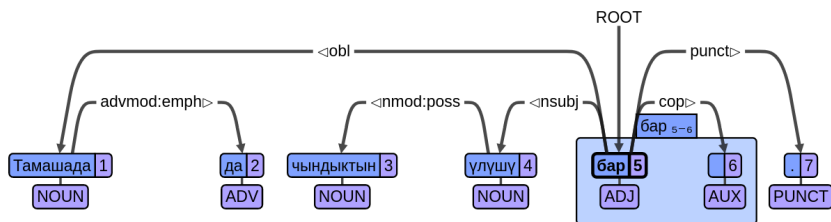
A Kazakh word with a similar distribution to the first use of *da* (although different meaning), *зоу/қоу* (corresponding to Kyrgyz *зо*), is annotated by Tyers & Washington [2015] and Makazhanov et al. [2015] as PART, with a discourse relation to the root.

Given all of this, we opt to analyse *da* in the following ways:

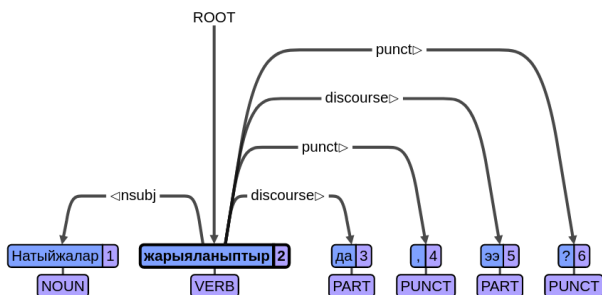
1. Post-predicate modal particle *da* is PART, with a discourse relation to the root.

2-5. Intensifier / emphasis uses of *da* are ADV, with an *advmod:emph* relation to the word it intensifies / emphasises.

Examples of this from our corpus are provided in Graphs 5 and 6.



Graph 5: Example of annotation of *advmod:emph* dependent of NOUN, corresponding to sentence (8).



Graph 6: Example of annotation of discourse *да*, corresponding to sentence (6).

#### 4.2.2. эле

The word *эле*, translating as ‘only, just’ (ignoring copula uses per §4.1), may occur after nearly any part of speech or phrase type in Kyrgyz:

<sup>1</sup> Per UD documentation of the existing four Turkish treebanks: <https://universaldependencies.org/tr/dep/advmod-emph.html>

- after nouns: *бала эле* ‘just a child’
- after adjectives: *кичинекей эле* ‘not that big’
- after numbers: *эки эле* ‘just two’
- after adverbs: *кечээ эле* ‘just yesterday’
- after adverbial clauses: *үч күн өткөндөн кийин эле* ‘only after three days had passed’

Benli [2023] analyses *эле* as ADV, with an *advmod* relation (except in cases like *чын эле* ‘really’ where it is given a fixed or compound relation), and Tyers & Washington [2015] and Makazhanov et al. [2015] do the same with the Kazakh word *зана/қана*, which has a similar distribution and meaning. However, the distribution of *эле*, including after nouns and numbers, makes it difficult to consider it a true adverb. However, we feel that like *да*, these uses of *эле* fit the intended use of the dependency relation *advmod:emph*. Hence, we annotate it this way, along with the POS tag ADV.

#### 4.2.3. *бар* and *жок*

The Kyrgyz words *бар* and *жок* are used in constructions that translate into English roughly as ‘there is/are’ and ‘there is not / are not’, respectively. With either possession or locative morphology, they can translate into ‘has/have’ and ‘do(es) not have’ constructions. Despite these verb-based translation, the fact that these words occur in copula constructions (11) and *be* verbs (12) is strong evidence that they are in fact either adjectives or nouns.

- (11) *Сен турганда мен бармын.*  
 sen tur-GAn-DA men bar-MIn  
 you stand-VN-LOC I present-COP.NPST.1SG  
 ‘I’m there when you get up.’  
 (literally: ‘I’m present’)

- (12) *Эртең Бишкектин айрым жерлеринде суу жок болот.*  
 erteŋ Biŝkek-NIn ayрым jer-LAr-(s)I(n)-DA suu  
 joq bol-E-t  
 tomorrow Bishkek-GEN some place-PL-POSS.3-LOC water  
absent be-NPST-3  
 ‘Tomorrow there will not be water in some places in Bishkek.’  
 (literally: ‘water will be absent’)

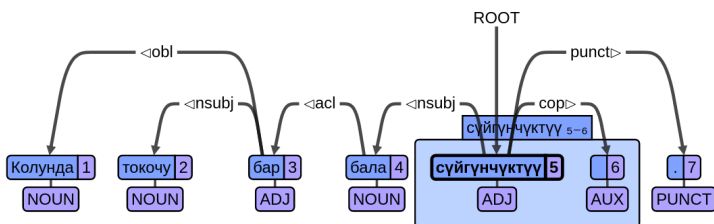
Examples like (8), from the corpus, push us to consider these words adjectives, translating literally as «present» and «absent», respectively.

- (13) *Колунда токочу бар бала сүйгүнчүктүү.*  
 qol-(s)I(n)-DA toqoç-(s)I(n) bar bala süygünçük-LUU.  
 hand-POSS.3-LOC loaf-POSS.3 present child darling.

‘The child with / who has the loaf (of bread) in their hands is darling.’

(literally: ‘[in their hand their loaf (being) present] child’ or ‘the child [whose loaf is present in their hand]’)

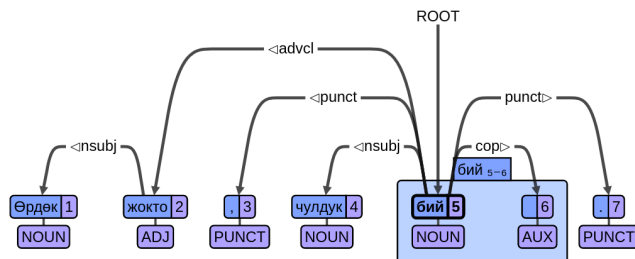
In (13), *бар* is the predicate of nominal subject *токочу* in a sort of copula construction, but the entire phrase is an adjective clause modifying *бала*. This is depicted in Graph 7.



Graph 7: UD dependency graph for sentence (8).

Furthermore, *бар* and *жок* can be used as nouns in Kyrgyz, receiving regular nominal morphology. An example from the corpus is (14), where *жокто* forms the head of an adverbial clause dependent on the main copula construction. A dependency graph for (14) is shown in Graph 8.

- (14) *Өрдөк жокто, чулдук бий.*  
 ördök joq-DA çulduq biy  
 duck absent-LOC sandpiper bey  
 ‘When the duck is absent, the sandpiper is king.’



Graph 8: UD dependency graph for sentence (14).

While *бар* and *жок* are often used in predicates, these previous two examples show their uses in other contexts. We understand these

words to be categorised as adjectives in Kyrgyz no matter what kind of construction they are encountered in.

4.2.4. *керек*

In Kyrgyz the word *керек* is used in ‘need to’ phrases, like that in (15).

- (15) Мен китепти тапшырышым керек.  
 men kitep-NIt apşır-(I)ş-(I)m керек.  
 I book-ACC turn.in-VN-POSS.1SG needed

‘I need to return the book.’

(literally: ‘me returning the book is needed/necessary.’)

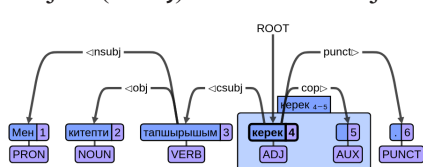
Due to this kind of translation in English as well as the distribution of cognates in some other Turkic languages (e.g., Turkish), it may be tempting to analyse *керек* as a verb, as Benli [2023] mostly does.<sup>1</sup> However, unlike the behaviour of said cognates, *керек* in Kyrgyz does not take any verbal morphology, suggesting that it is not a verb. Instead, it has a morphological and syntactic distribution more like that of a noun or adjective, as in sentences like (16); in this example, it comprises a non-finite predicate, and has a copula morpheme (§4.1) attached to it.

- (16) Мен үй-бүлөмө керекмин.  
 men üy-bülö-(I)mA керек-мIn  
 I family-POSS.1SG:DAT needed-COP.NPST.1SG

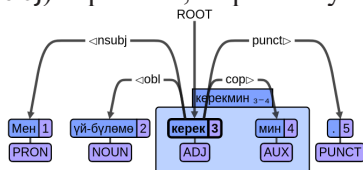
‘My family needs me.’

(literally: ‘I am needed/necessary to my family.’)

As with *бар* and *жок*, we opt to analyse *керек* as an adjective (ADJ) with a literal paraphrase of ‘needed’ or ‘necessary’, although an analysis as a noun (with a reading like ‘a needed/necessary thing’) might also be possible. We annotate these sentences as shown in Graphs 9 and 10, with *керек* as a copular predicate, having a clausal subject (csubj) or nominal subject (nsubj) dependent, respectively.



Graph 9: Dependency graph of sentence (15), showing *керек* annotated as ADJ with a csubj dependent.



Graph 10: Dependency graph of sentence (16), showing *керек* annotated as ADJ with an nsubj dependent.

<sup>1</sup> In a few instances Benli (2023) instead analyses *керек* as a NOUN.

### 4.3. Null-headed clauses

Turkic languages exhibit a number of phenomena where null or empty heads are posited. This term refers to a phrase operating as if a lexical head is present, despite one not being overtly realised. Three different instances of this process are discussed here: substantivised verbal adjectives (§4.3.1), substantivised relativised locative expressions (§4.3.2), and substantivised genitive expressions (§4.3.3).

#### 4.3.1. Substantivised verbal adjectives

An example of this phenomenon is «substantivised» verbal adjectives [see Washington et al. 2022]. In these constructions, a verbal adjective modifies a noun that is not present, but is understood through the nominal morphology that is in turn attached to the verbal adjective. Verbal adjectives in Turkic are used to form relative clauses, so these may also be considered «headless» relative clauses. These may be read in English as «(the) person/thing/one who/that». Examples of this type of construction are presented in (17) and (18), sentences drawn from the corpus.

(17) *Колуң менен кылганды, мойнуң менен тартасың.*  
 qol-(I)ŋ menen qıl-GAn-NI moyun-(I)ŋ menen tart-E-sİŋ.  
 hand-POSS.2SG with make-VADJ-ACC neck-POSS.2SG  
 with pull-NPST-2SG

‘You will pull with your neck what you make with your hands.’

(18) *Балалуу болбогон кубанганды билбеген.*  
 bala-luu bol-BA-GAn quban-GAn-NI bil-BA-GAn.  
 child-ORN be-NEG-VADJ be.happy-VN-ACC know-NEG-  
 PST;3

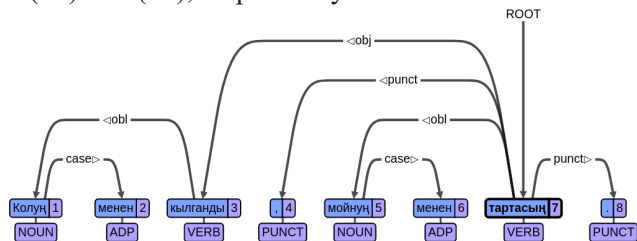
‘One who has not had children has not known being happy.’

Both of these sentences could be stated with an additional word added after the verbal adjective suffix and have almost exactly the same meaning, e.g. *кылган нерсени* ‘make-VADJ thing-ACC = the thing you make’ and *болбогон киши* ‘the person who has not had’, respectively. Hence, one possibility is to add an additional null node to the UD analysis of these sentences. However, UD standards are strongly against adding null nodes if at all possible.

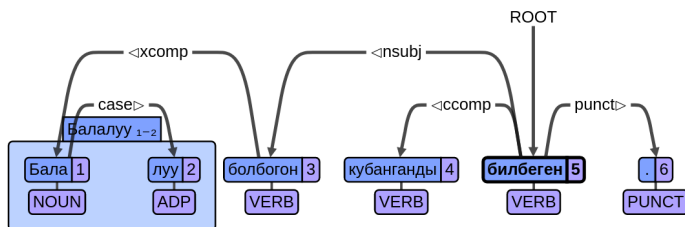
Even without adding null nodes, the most obvious way (to us) of annotating these verbal adjective clauses still shows that they are (tacitly) dependent on a nominal head. Specifically, treating them as a nominal object (obj) (17) or nominal subject (nsubj) (18) instead of as a clausal complement (ccomp) (17) or clausal subject (csubj) (18) makes

it clear that the verb is not the head of these phrases. Since they are verbal adjectives (VerbForm=Part), they are not really able to operate as objects or subjects (clausal or nominal) on their own anyway.

This approach is shown in the annotated versions of (17) and (18) in Graphs (11) and (12), respectively.



Graph 11: UD annotated version of sentence (17), without a null head explicitly represented, and showing the empty-headed verbal adjective as an object, as opposed to a clausal complement.



Graph 12: UD annotated version of sentence (18), without a null head explicitly represented, and showing the empty-headed verbal adjective as a nominal subject, as opposed to as a clausal subject.

Benli [2023] inconsistently treats empty-headed verbal adjectives as verbal nouns (VerbForm=Vnoun) and verbal adjectives (VerbForm=Part). In nearly every instance that they were analysed as verbal nouns, they include plural morphology, which is not semantically compatible with verbal nouns in most Turkic languages. Tyers & Washington (2015) treat these forms fairly consistently as gerunds (VerbForm=Ger), equivalent to verbal nouns, and as clausal dependents instead of nominal dependents.

#### 4.3.2. Substantivised relativised locative expressions

The locative case in Kyrgyz is *-DA*, and can only be used adverbially. A derived form of the locative case used attributively is *-DAGI*. An example from corpus is presented in (19).

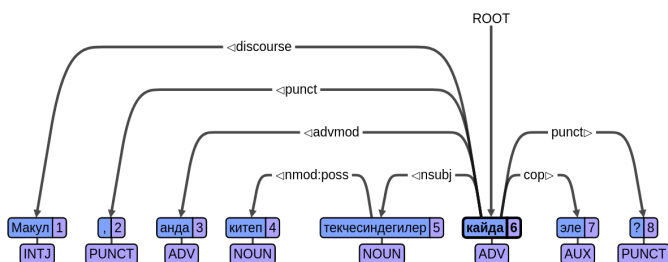
(19) *Алыстагы душмандан аңдып жүргөн дос жаман.*  
 alıs-DAGI duşman-DAn aңdı-(I)p жүr-GAn dosjaman.  
 far-LOC;ATTR enemy-ABL spy-INF go.around-VADJ friend bad.  
 ‘A friend who spies on you is worse than an enemy who is far away.’

Here *алыстагы* is an nmod:loc dependent on the noun *душман*.

Forms in -DAGI can also have an empty head, and hence can function as nominal heads and take nominal morphology. An example from the corpus is presented in (20).

(20) *Макул, анда китеп текчесиндегилер кайда эле?*  
 maqul anda kitep tekçe-(s)In-DAGI-LAr qayda ele?  
 okay then book shelf-POSS.3-LOC;ATTR-PL where were?  
 ‘Okay, then where were the ones on the bookshelf?’

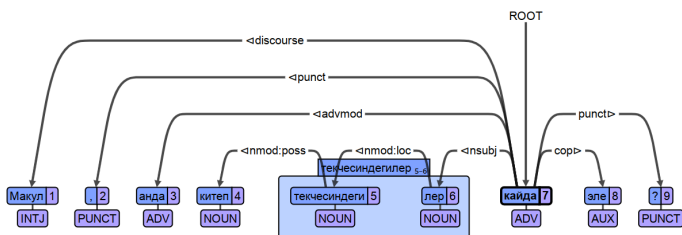
In one UD annotation of this sentence, presented in Graph 13, the subject misleadingly appears to simply be an inflected form of *китеп текчеси* ‘bookshelf’, despite the existence of another participant, hidden from the analysis due to it not having a surface realisation. Additionally, in morphological features, there are two distinct items: a singular bookshelf, and a plural set of items on the shelf—which number to annotate this form with is not clear according to UD guidelines.



Graph 13: A UD annotation of sentence (20), without an extra token for the additional «empty» participant.

The only other way to annotate such structures, as we see it, would be to break the problematic form into two subtokens, as in Graph 14. It is not clear to us that this is preferable, but it solves the issue of associating features for multiple participants with one form. This approach also clarifies that there are multiple participants. For now, we have gone with this approach in the corpus.





Graph 14: A UD annotation of sentence (20) with an extra token for the additional «empty» participant.

### 4.3.3. Substantivised genitive expressions

Also wrapped up in discussions of «*ki*» in Turkish are substantivised genitive expressions. In Kyrgyz these are formed with *-NIKI*. Despite the similarity in Turkish and potential etymological unity of the *KI* element, the way *-NIKI* works is conceptually different from *-DAGI*: the former creates a substantivised form of the genitive, whereas the latter creates an attributive form of an adjective, and, on occasion, a substantivised form as discussed in §4.3.2. An example of Kyrgyz *-NIKI* is presented in sentence (21), which is drawn from our corpus.

(21) *Жубайым дачадагы балдар бөлмөсүнүн терезесин*  
 jubay-(I)m daça-DAGI bala-LAR bölmö-(s)I(n)-nIn tereze-(s)In  
 spouse-POSS.1SG summer.house-LOC.ATTR child-PL room-POSS.3-GEN window-POSS.3:ACC

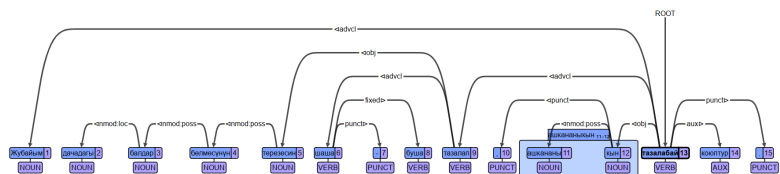
*шаша-буша таз лап, ашкананыкын тазалабай коюптур.*

*şaş-E-buş-E tazala-(I)p, aşqana-NIKIn tazala-BA-E qoy-(I)ptlr.*

*rush-VADV-rush-VADV clean-VADV kitchen-GEN.SUBST:ACC*  
*clean-NEG-INF put-PST.IDF;3*

‘My spouse cleaned the windows of the children’s room at the summer house hastily, but didn’t clean the kitchen’s.’

Our approach to dealing with *-NIKI* is the same as for substantivised forms of *-DAGI*: we explicitly add an empty head to our dependency graph, and give it the features of the head of the phrase, with an *nmod:poss* relationship to the noun with *-NIKI* morphology. We segment the subtokens in the middle of the morpheme, e.g. *ашкананыкын*, where we treat the first part as a genitive (the full equivalent of which is *ашкананын*) and the second part consists of *KI* and any additional morphology—here, accusative case. A dependency graph showing this is provided in Graph 15.



Graph 15: Dependency graph for sentence (21), with the null head a separate token.

Tyers and Washington [Tyers, 2015] parse cognate forms as two separate tokens, but make the second one a case dependent on the first, preventing the analysis of two separate participants in the event. Benli [Benli, 2023] mysteriously treats some *-NIKI* forms as being ADJ and amod dependents; however, one form in the corpus is treated as having two case markings, one being accusative (instead of genitive) and the other being the case of the empty head.

#### 4.4. Inflection versus derivation

A number of suffixes that occur with nouns in Kyrgyz are not commonly analysed as case suffixes among Kyrgyz linguists. For example *-LUU* could be considered an ornative case suffix, but it is not. Similarly, *-sIz* could be considered abessive or privative case, *-DAy* could be considered semblative case, and *-çA* could be considered an adverbial case.

There are at least three different ways to analyse nouns with such morphology:

1. One way is to treat the nouns as deriving adjectives or adverbs; e.g., *балалуу* in (18) could be treated simply as an adjective ‘having child(ren)’ that happens to be derived from the noun *бала* ‘child’. This analysis is unsatisfying because these morphemes act quite productively.

2. Another possible analysis is to analyse these morphemes as case marking, as is done for other cases in Turkic. So, the lemma of *балалуу* would be *бала*, and there would be annotation for ornative case in the grammatical features. One disadvantage of this approach is that unlike other case morphemes, these morphemes both do not enable the noun to act as a core argument of a verb (i.e., one that is involved in case demotion or promotion related to grammatical voice inflection), and can be used to create attributive (adjective-like) forms. Additionally, the names for these cases are not standardised in Kyrgyz grammar, and

the names that seem the most obvious to use are not all standard within Universal Dependencies.

3. These morphemes can be treated as cliticised postpositions. This has the advantage of highlighting their productivity and their difference from case suffixes. The main disadvantage of this approach is that tokenisation then does not line up with spaces. Another disadvantage is that the other case morphemes could also be annotated this way, and not doing so introduces some level of arbitrariness into the corpus.

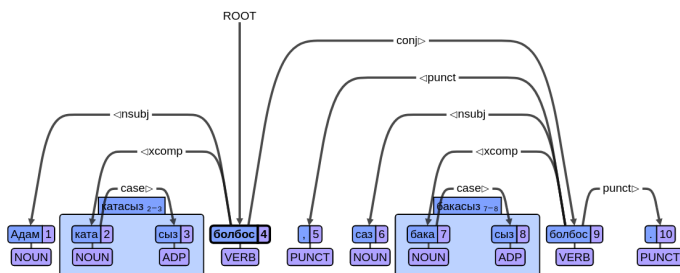
We decided to go with the third approach for the most part, as shown in (22) and Graph 16, depicting a sentence drawn from our corpus.

(22) *Адамкатасыз болбос, саз бакасыз болбос.*

adam qata-sIz bol-BAs, saz baqa-sIz bol-BAs.

person error-ABE be-NEG;FUT.IDF marsh frog-ABE b e -  
NEG;FUT.IDF

‘A person won’t be without errors, a marsh won’t be without frogs.’



Graph 16: Annotated version of sentence (22), showing case-like morphology treated as postpositions.

Benli [Benli, 2023] uses a mixture of the first two approaches. For example, *-LUU* and *-sIz* forms are sometimes annotated as NOUN, other times as ADJ, and other times as ADV, and often with the lemma being the noun lemma which the morpheme is attached to, regardless of the part of speech annotated. Tyers and Washington [Tyers, 2015] use a mixture of the first and third approaches; for example, *баласыз* ‘without children’ is treated as having two subtokens, with *сыз* an ADP, while *сансыз* ‘without count’ is treated simply as an adjective.

## 5. Conclusion

In this paper we have presented several issues of syntactic annotation relevant to the annotation of a forthcoming Universal Dependencies

annotated corpus of Kyrgyz text. We have compared these issues to the existing UD corpora of Kazakh [Tyers and Washington, 2015; Makazhanov et al., 2015] and Kyrgyz [Benli 2023]. We have weighed the advantages and disadvantages of various approaches, and argue for specific solutions to these issues. Compared to the existing Kyrgyz treebank [Benli 2023], we aim to present a more comprehensive analysis of ongoing annotation issues and to build a treebank of larger size and domain coverage.

The UD Kyrgyz corpus presented here significantly contributes to the syntactic resources available for the Kyrgyz language. This corpus will serve as a valuable resource for studying the syntax and grammatical structure of the Kyrgyz language, as well as for developing language technologies such as dependency parsers and machine translation systems. It provides higher-quality annotated data compared to the previously available UD Kyrgyz treebank [Benli 2023], addressing the need for expanded Kyrgyz resources to support natural language processing applications. The inclusion of a new syntactic corpus in the Universal Dependencies framework for Kyrgyz will not only enhance the quality of linguistic research but also contribute to the broader goal of enhancing the representation of underrepresented languages in language technology. By addressing existing limitations and inaccuracies, this endeavour enables the Kyrgyz language to be better understood, studied, and utilized in various language-related applications along with promoting the development of resources to support Kyrgyz natural language processing.

### **Acknowledgments**

We gratefully acknowledge the foundational work of Prof. Elke Teich and MSc. Jörg Knappen at the Universität des Saarlandes in developing the initial Manas-UdS Kyrgyz corpus that provided much of the textual data for this UD project. Their efforts in compiling a representative sample of texts across various domains helped ensure a strong underlying syntactic corpus. We are also grateful to the participants of the 2023 UD Turkic Workshop, who offered insightful feedback and discussion. We also sincerely thank our talented students Aidai Abitova, Alina Iskenderova, Alina Nijazbekova, Azima Naamatbekova, Bermet Ulukbekova, Cholpon Kultaeva, Kurmanjan Ydyrysova, Meerim Taalaibekova, Suyun Tostonova, and Zuura Mirlanova at Kyrgyz-Turkish Manas University who diligently performed a first pass of syntactic annotation for many of the Kyrgyz sentences according to the Universal Dependencies guidelines. The students' careful application of UD principles during the annotation process, while also thoughtfully handling

ambiguities and inconsistencies, resulted in a high-quality base for the resource under development. Their contributions have made the annotated Kyrgyz UD corpus a valuable asset for future research and tool development for the natural language processing community working with the Kyrgyz language.

## REFERENCES

1. Aili M., Mushajiang W., Yibulayin T., Yan Liu K. A. In Proceedings of the Third International Workshop on Worldwide Language Service Infrastructure and Second Workshop on Open Infrastructures and Analysis Frameworks for Human Language Technologies (WLSI/OIAF4HLT2016). Osaka, Japan. 2016. Pp. 44–50.
2. Benli İ. UD\_Kyrgyz-KTMU: Universal Dependency treebank for Kyrgyz. 2023 [https://github.com/UniversalDependencies/UD\\_Kyrgyz-KTMU](https://github.com/UniversalDependencies/UD_Kyrgyz-KTMU), [https://universaldependencies.org/treebanks/ky\\_ktmu/index.html](https://universaldependencies.org/treebanks/ky_ktmu/index.html).
3. Çetinoğlu Ö., Çöltekin Ç. Two languages, one treebank: building a Turkish-German code-switching treebank and its challenges. In: Language Resources and Evaluation. 2023. Vol. 57, pp. 545–579. <https://doi.org/10.1007/s10579-021-09573-1>.
4. Hakkani-Tur D., Oflazer K., Tur G. Statistical Morphological Disambiguation for Agglutinative Languages. *Computers and the Humanities*. 2002. Vol. 36. 381–410. <http://doi.org/10.1023/A:1020271707826>.
5. Imanalieva Zh. Кыргыз жана орус тилдеринде синтаксистик катыштардын синтаксистик өзгөчөлүктөрү, Бишкек, 2015. С. 197–200. <http://www.science-journal.kg/media/Papers/nntiik/2015/11/197-200.pdf> (Accessed: 20 October 2023).
6. Kasieva A., Knappen J., Fischer S., and Teich E. A new Kyrgyz corpus: sampling, compilation, annotation. Poster at: 42. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Hamburg (Germany), March 2020. <https://www.zfs.uni-hamburg.de/dgfs2020/programm/abstracts/dgfs2020-clp-kasieva.pdf>, [https://corpora.clarin-d.uni-saarland.de/cqpweb/kyrgyz\\_2022\\_03\\_08](https://corpora.clarin-d.uni-saarland.de/cqpweb/kyrgyz_2022_03_08).
7. Makazhanov A., Sultangazina A., Makhambetov O., and Yessenbayev Zh. Syntactic Annotation of Kazakh: Following the Universal Dependencies Guidelines. A report. In: Proceedings of the 3rd International Conference on Computer Processing in Turkic Languages (TurkLang 2015). Kazan, Tatarstan, 2015. Pp. 338–350. <http://www.turklang.org/en/turklang-2015-2/>.
8. Nivre J., Marneffe M., Ginter F., Goldberg Y., Hajic J., Manning Ch., McDonald R., Petrov S., Pyysalo S., Silveira N., Tsarfaty R., Zeman D. Universal Dependencies v1: A Multilingual Treebank Collection. In Proc. of LREC 2016. Pp. 1659–1666.
9. Oflazer K., Say, B., Zeynep, D., Tur, G. Building a Turkish Treebank. Abeillé, 2003. [http://doi.org/10.1007/978-94-010-0201-1\\_15](http://doi.org/10.1007/978-94-010-0201-1_15).

10. Sulubacak U., Gokirmak M., Tyers F., Çöltekin Ç., Nivre J., Eryiğit G. Universal Dependencies for Turkish. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 3444–3454, Osaka, Japan. 2016. <https://aclanthology.org/C16-1325>.
11. Taguchi Ch. UD Tatar-NMCTT: Universal Dependency corpus for Tatar. 2022. [https://github.com/UniversalDependencies/UD\\_Tatar-NMCTT](https://github.com/UniversalDependencies/UD_Tatar-NMCTT), [https://universaldependencies.org/treebanks/tt\\_nmctt/index.html](https://universaldependencies.org/treebanks/tt_nmctt/index.html).
12. Thompson A. Syntactic Parallelism and Structure in Kyrgyz Proverbs (Bachelors thesis). Bryn Mawr College, Pennsylvania. 2021.
13. Tyers F., Sheyanova M., Washington J. UD Annotatrix: An annotation tool for Universal Dependencies. In: Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories (TLT). Praha, Česko, 2017. Pp. 10–17. <https://aclanthology.org/W17-7604>.
14. Tyers F., Washington J. Towards a free/open-source universal-dependency treebank for Kazakh. In: Proceedings of the 3rd International Conference on Computer Processing in Turkic Languages. TurkLang 2015. Kazan, Tatarstan. 2015 Pp. 276–289. <http://www.turklang.org/en/turklang-2015-2/>.
15. Tyers F., Washington J., Çöltekin Ç., Makazhanov A. An assessment of Universal Dependency annotation guidelines for Turkic languages”. In: Proceedings of the Fifth International Conference on Turkic Language Processing. TurkLang 2017. Vol. 1. Pp. 276–297. <http://www.turklang.org/en/turklang-2017-2/>.
16. Washington J., Ipasov M., Tyers F. A finite-state morphological transducer for Kyrgyz. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12). European Language Resources Association (ELRA). Istanbul, Turkey. 2012. Pp. 934–940 <https://aclanthology.org/L12-1642/>.
17. Washington J., Salimzyanov I., Tyers F. Finite-state morphological transducers for three Kypchak languages. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14). European Language Resources Association (ELRA). Reykjavik, Iceland. 2014. Pp. 3378–3385, <https://aclanthology.org/L14-1143/>.
18. Washington J., Tyers F., Salimzyanov I. Non-finite verb forms in Turkic exhibit syncretism, not multifunctionality. In: Shagal, Ksenia, Pavel Rudnev, and Anna Volkova (eds.), *Folia Linguistica*, vol. 56, no. 3, Special Issue: Multifunctionality and syncretism in non-finite forms, 2022. Pp. 693–742. <https://doi.org/10.1515/flin-2022-2045>.
19. Zeman D. et al. Universal Dependencies 2.12». In: LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. 2023 <http://hdl.handle.net/11234/1-5150>.

20. Джумалиева Г.К., Касиева А.А., Мусажанова С.Ж. [Dzhumalievа G.K., Kasieva A.A., Musazhanova S.J.]. Адаптация терминов веб-проекта универсальные зависимости на кыргызский язык [Adaptation of Web Project Terms for Universal Dependencies in the Kyrgyz Language]. In: Вестник КРСУ [Bulletin of KRSU]. Bishkek, 2023. Vol. 23, № 6, pp. 71–75. <http://doi.org/10.36979/1694-500X-2023-23-6-71-75>.

21. Мусажанова С. Ж., Касиева А. А., Джумалиева Г. К. [Musazhanova S. J., Kasieva A. A., Dzhumalievа G. K.]. Синтаксическая аннотация кыргызского языка на основе новосозданного корпуса [Syntactic Annotation of the Newly-Created Kyrgyz Corpus]. Вестник Иссык-Кульского университета [Bulletin of the Issyk-Kul University], Karakol, 2023. №54. Pp. 140–148.

УДК 81'322.2

**АННОТАЦИЯ ГРАММАТИЧЕСКИХ ОШИБОК В ТЕКСТАХ  
НА ТАТАРСКОМ ЯЗЫКЕ С ПОМОЩЬЮ ИНСТРУМЕНТОВ  
КОРПУС-МЕНЕДЖЕРА****Б. Э. Хакимов<sup>1</sup>, Д. Р. Мухамедшин<sup>2</sup>, З. И. Садыкова<sup>2</sup>**<sup>1</sup>*Институт прикладной семиотики Академии Наук Республики Татарстан, Казанский федеральный университет, Казань, Россия*<sup>2</sup>*Институт прикладной семиотики Академии Наук Республики Татарстан, Казань, Россия*khakeem@yandex.ru, damirmuh@gmail.com,  
ziliasadykova@mail.ru

В данной статье описаны первые результаты аннотации грамматических ошибок в текстах на татарском языке, выполненной с использованием корпус-менеджера “Туган Тел”. Датасет для аннотации был собран из социальных сетей (Telegram, VK) и информационных веб-сайтов. Он состоял примерно из 20000 текстов, сообщений и комментариев. Была разработана классификация грамматических ошибок в татарском языке, учитывающая такие аспекты, как орфография, пунктуация, грамматика, выбор слов и другие ошибки, которые характерны для татарских пользователей Интернета. Был разработан модуль для ручного аннотирования и специальный набор тегов. Разработанный модуль позволяет после аннотации экспортировать параллельный корпус предложений с аннотацией грамматических ошибок, который может быть использован для автоматизации задачи исправления ошибок в татарских текстах.

**Ключевые слова:** татарский язык, автоматическое исправление грамматических ошибок, классификация ошибок, корпус ошибок, аннотация ошибок.

**GRAMMATICAL ERROR ANNOTATION IN TATAR TEXTS  
USING CORPUS MANAGEMENT TOOLS****B. E. Khakimov<sup>1</sup>, D. R. Mukhamedshin<sup>2</sup>, Z. I. Sadykova<sup>2</sup>**<sup>1</sup>*Institute of Applied Semiotics of the Tatarstan Academy of Sciences, Kazan Federal University, Kazan, Russia*<sup>2</sup>*Institute of Applied Semiotics of the Tatarstan Academy of Sciences, Kazan, Russia*khakeem@yandex.ru, damirmuh@gmail.com,  
ziliasadykova@mail.ru

This paper describes the first results of the grammatical error annotation in Tatar texts performed using “Tugan Tel” corpus management system. The dataset



was collected from the social media (Telegram, VK) and informational web-sites. It consisted of about 20,000 comments. A classification of grammatical errors in Tatar was developed, taking into account such issues as spelling, punctuation, grammatical, word choice and other errors, which are typical for the Tatar users of the Internet. A module for manual annotation and the special tagset was developed. As a result of the annotation we can export a sentence-level parallel corpus with grammatical error annotation which can be used to perform the grammatical error correction task.

**Keywords:** Tatar language, grammatical error correction, GEC, error classification, error corpus, error annotation.

## 1. Введение

Электронный корпус языка как комплекс интегрированных структурно-функциональных лингвистических моделей, с одной стороны, дает возможность проводить широкий спектр исследований в области филологии и автоматической обработки текста, а с другой стороны, является содержательным и технологическим ядром различных прикладных разработок. Данное направление также включает в себя подготовку и разметку специализированных подкорпусов для задач автоматического анализа текстов, в частности, таких как автоматическое обнаружение и корректировка ошибок.

В понятие автоматического исправления грамматических ошибок (Grammatical Error Correction) в современной практике автоматической обработки естественного языка включаются не только собственно грамматические ошибки в узком смысле, но и орфографические, семантические и другие типы ошибок. За последнее десятилетие в этой области был достигнут значительный прогресс, разработаны методы, основанные на правилах, статистических классификаторах, статистическом машинном переводе и, наконец, нейронном машинном переводе [Bryant et al., 2022]. Для решения данной задачи на современном этапе требуется наличие размеченного корпуса ошибок. Для татарского языка таких ресурсов до настоящего времени не существовало. Наше исследование ориентировано на частичное заполнение этого пробела.

## 2. Подготовка данных

В рамках подготовки размеченных подкорпусов для задач автоматического анализа текстов для языков в Республике Татарстан, в данном исследовании осуществлялся сбор текстовых мате-

риалов на татарском языке из сети Интернет, социальных сетей и мессенджеров для пилотной версии из сети Интернет для составления параллельного корпуса ошибок и последующей разметки датасета в соответствии с разработанной классификацией частотных ошибок с использованием инструментов корпус-менеджера «Туган тел».

План работы включал следующие этапы:

- Поиск источников для датасета в социальных сетях
- Парсинг, скачивание и предобработка выборки комментариев
- Фильтрация выборки по языку комментариев и удаление нерелевантных примеров
- Разработка классификации частотных типов ошибок и тэгов для разметки
- Разметка и исправление ошибок

В результате был собран датасет объёмом 20000 комментариев с Интернет-ресурсов на татарском языке (Телеграм-канал “Сөйләсем килә”, сообщества СМИ, звёзд татарской эстрады ВКонтакте, сайт журнала “Сөембикә”). Ограниченное количество сайтов и страниц в социальных сетях, которые ведутся преимущественно на татарском языке, обуславливает главную особенность датасета: в основном он включает в себя тексты социально-бытовой направленности.

Для сбора датасета использовались следующие инструменты: экспорт чатов в Telegram, готовые инструменты для парсинга (<https://pepper.ninja/> – для страниц сообществ ВКонтакте), библиотеки Python requests и BeautifulSoup (для парсинга сайтов).

Датасет комментариев был предобработан: удалены дубликаты, пустые ячейки, комментарии на нерелевантных языках (в частности, – комментарии на башкирском, арабском языках), при помощи регулярных выражений и библиотек (pandas, NumPy) удалены нетекстовые символы (эмоджи), нерелевантные наборы символов (номера телефонов, адреса веб-сайтов, адреса электронной почты, id пользователей и т. д.).

### 3. Классификация ошибок

Аннотация грамматических ошибок в собранном корпусе требовала разработки классификации наиболее частотных ошибок, совершаемых при написании текстов на татарском языке и фор-

мальных правил для автоматизации пополнения корпуса ошибок.

На основе анализа собранного датасета Интернет-комментариев нами была составлена классификация частотных собственно грамматических, орфографических и лексических ошибок в татарском языке:

1. Замена специфических символов татарского алфавита
2. Случайная вставка/удаление/замена/перестановка
3. Необоснованное употребление заимствований из русского (при наличии эквивалента из татарского языка)
4. Ошибки в написании заимствований из русского языка (как ассимилированных, так и не ассимилированных)
5. Необоснованное повторение одной буквы
6. Морфема как отдельный токен
7. Написание частицы слитно с предыдущим словом
8. Имена собственные на русском языке
9. Замена символа по регулярному правилу
10. Раздельное/слитное написание вместо дефиса
11. Вставка символа по регулярному правилу
12. Удаление символа по регулярному правилу
13. Другие ошибки

В данной классификации представлены как ошибки, являющиеся следствием несоблюдения регулярных правил, существующих в татарском языке (к примеру, пункты 6, 7, 11, 12), так и те, которые объясняются внеязыковыми факторами (невнимательность пользователя – пункты 2, 5; отсутствия татарской клавиатуры на электронном устройстве пользователя – пункт 1).

Один тип ошибки может включать в себя сразу несколько языковых правил. Например, пункт “Замена символа по регулярному правилу” включает в себя следующие ошибки:

- замена “в” в интервокальной позиции на “у” (“тауы, ауыл, сөйләуе» вместо «тавы, авыл, сөйләве»);
- «-не» вместо суффикса «-ле» («тәмне» вместо «тәмле»);
- написание звонкой согласной вместо его глухой пары в конце слова («китаб» вместо «китап»);
- и другие.

#### 4. Разметка ошибок

Распределенная система корпус-менеджера “Туган тел” с расширяемым функционалом и специализированными модулями обеспечивает автоматизацию рутинных процессов в лингвистиче-

ских исследованиях татарского языка [Сулейманов, Мухамедшин, 2018]. Размеченные коллекции лингвистических данных создают основу для дальнейших исследований в области автоматического анализа текстов.

С помощью онлайн-инструмента разметки корпуса “Туган тел” (<https://tugantel.tatar>) была проведена разметка комментариев: размечены границы и типы ошибок, сохранены исправленные версии предложений.

Для разметки использовался модуль разрешения морфологической неоднозначности системы управления корпусными данными [Мухамедшин, Сулейманов, 2020]. Общий вид интерфейса для разметки примеров с ошибками представлен на Рисунке 1. Данный модуль позволяет сохранять результаты разметки в виде дискретных действий: исправление словоформы, исправление морфологической разметки.

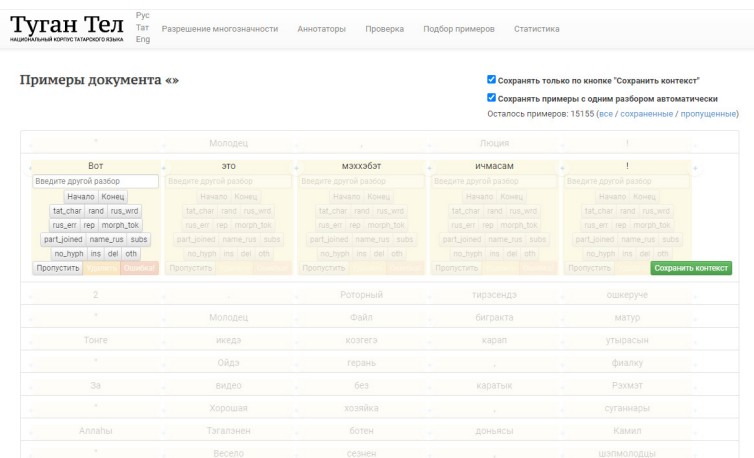


Рис. 1. Интерфейс модуля разрешения морфологической неоднозначности для разметки примеров с ошибками

Для решения задачи по разметке ошибок в текстах в морфологическую разметку были добавлены дополнительные теги типов ошибок:

- `tat_char` – замена специфических символов татарского алфавита;
- `rand` – случайная вставка/удаление/замена/перестановка;

- `rus_wrd` – необоснованное употребление заимствований из русского (при наличии эквивалента из татарского языка);
- `rus_err` – ошибки в написании заимствований из русского языка (как ассимилированных, так и не ассимилированных);
- `rep` – необоснованное повторение одной буквы;
- `morph_tok` – морфема как отдельный токен;
- `part_joined` – написание частицы слитно с предыдущим словом;
- `name_rus` – имена собственные на русском языке;
- `subs` – замена символа по правилу;
- `no_hyph` – раздельное/слитное написание вместо дефиса;
- `ins` – вставка символа по регулярному правилу;
- `del` – удаление символа по регулярному правилу;
- `oth` – другое.

Также для разметки ошибок, состоящих из нескольких словоформ, теги могут быть расширены метками начала и завершения участка текста (“`start_`” и “`end_`” соответственно), состоящего из нескольких словоформ.

В интерфейсе модуля разрешения была добавлена возможность выбора типа ошибки для соответствующей словоформы, обновленный интерфейс представлен на Рисунке 2.

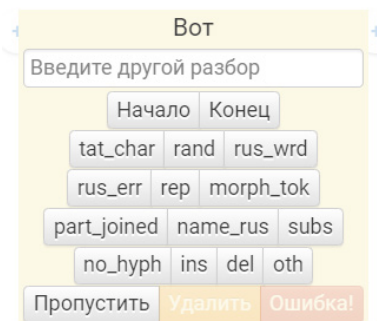


Рис. 2. Выбор типа ошибки в интерфейсе модуля разрешения морфологической неоднозначности

В случае, если внутри одного токена можно было найти более одной ошибки, и даже если эти ошибки были одного типа (к примеру, все специфические символы татарского алфавита были заменены символами русского алфавита), размечалась каждая из них.

После выполнения разметки ошибок модуль позволяет выгрузить результаты в виде таблицы с исходными текстами, размеченными текстами и исправленными текстами. Примеры выгрузки результатов представлены в Таблице 1.

**Таблица 1. Примеры результатов разметки ошибок**

| Исходный текст   | Размеченный текст  | Исправленный текст   |
|--|--|--|
| Вот это мэххэбэт ичмасам !   | Менә{{{end_start_Менә;rus_wrd}}} бу{{{end_start_бу;rus_wrd}}} мэхэббэт{{{end_start_мэхэббэт;tat_char;tat_char;tat_char;rep;rand}}} ичмасам !   | Менә бу мэхэббэт ичмасам!  |
| 2 . Роторный тирэсендэ ошкерүче Радиф дигән кешенен нормерын белүче юк микән ? | 2{{{start_end_rand}}}. Роторный тирэсендэ{{{start_end_тирэсендэ;tat_char;tat_char}}} ошкерүче{{{start_end_ошкерүче;tat_char;tat_char}}} Радиф дигән{{{start_end_дигән;tat_char}}} кешенен{{{start_end_кешенен;tat_char}}} нормерын белүче{{{start_end_нормерын_белүче;tat_char}}} юк микән{{{start_end_микән;tat_char}}} ? | 2. Роторный тирэсендэ ошкерүче Радиф дигән кешенен нормерын белүче юк микән? |

## 5. Заключение

В результате работы был создан параллельный размеченный корпус ошибок на татарском языке. Разметка выполнена на основе специально разработанной классификации ошибок. Такой корпус создается для татарского языка впервые, после расширения и дополнения может быть использован при разработке прикладных систем исправления ошибок.

Создание классификаций частотных ошибок и их формальных описаний способствует разработке систем автоматического исправления ошибок в текстах на татарском языке. Подобные системы могут быть полезны как при изучении языка, так и в работе различных государственных организаций и ведомств.

## ЛИТЕРАТУРА

1. Bryant, Christopher & Yuan, Zheng & Qorib, Muhammad & Cao, Hannan & Ng, Hwee & Briscoe, Ted. (2022). Grammatical Error Correction: A Survey of the State of the Art. 10.48550/arXiv.2211.05166.
2. Сулейманов Д. Ш., Мухамедшин Д. Р. Система корпус-менеджер: архитектура и модели корпусных данных // Программные продукты и системы. – 2018. – Т. 31. – №. 4. – С. 653–658.
3. Мухамедшин Д. Р., Сулейманов Д. Ш. Модуль разрешения морфологической неоднозначности: архитектура и организация базы данных // Программные продукты и системы. – 2020. – Т. 33. – №. 1. – С. 38–46.
4. Nevzorova O., Mukhamedshin D., Gataullin R. Developing corpus management system: architecture of system and database // Proceedings of the International Conference on Information and Knowledge Engineering (IKE). – The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2017. – С. 108–112.
5. Suleymanov, D.; Nevzorova, O.; Gatiatullin, A.; Gilmullin, R.; Khakimov, B. National Corpus of the Tatar Language “Tugan Tel”: Grammatical Annotation and Implementation. In Procedia-Social and Behavioral Sciences, 2013, Vol.95, pp. 68–74. <https://doi.org/10.1016/j.sbspro.2013.10.623>
6. <https://pypi.org/project/requests/>
7. <https://pypi.org/project/beautifulsoup4/>
8. <https://pandas.pydata.org/>
9. <https://numpy.org/>

УДК 81'322.2

**ИССЛЕДОВАНИЕ АВТОМАТИЧЕСКОГО  
ФОРМИРОВАНИЯ СИНТЕТИЧЕСКИХ КОРПУСОВ РЕЧИ  
ТЮРКСКИХ ЯЗЫКОВ****У. А. Тукеев, Ж. Толеубеков, Т. Балабекова,  
Ж. Жуманов, Б. Темирханов***Казахский Национальный Университет  
им. Аль-Фараби, Алматы, Казахстан*  
ualsher.tukeyev@gmail.com, zhandoslp@gmail.com,  
t.balabekova@mail.ru, z.zhake@gmail.com,  
temirkhanov.bigeldi@gmail.com

В настоящее время основным инструментом исследования машинного перевода речи в речь являются технологии нейронных сетей. Однако для многих языков исследование прямого машинного перевода речи затруднено из-за сложности формирования параллельных речевых корпусов для обучения нейронных сетей. Исследования машинного перевода речи тюркских языков практически отсутствуют из-за сложности создания параллельных речевых корпусов для обучения нейронных моделей. Поэтому актуальной проблемой является автоматическое формирование синтетических параллельных корпусов для обучения нейронного машинного перевода речи тюркских языков. В данной статье предлагается технология формирования синтетических параллельных речевых корпусов для тюркских языков с использованием каскадной схемы машинного перевода на примере казахско-татарской, казахско-турецкой, казахско-узбекской языковых пар. Особенностью этой каскадной схемы является то, что она использует реляционную модель, основанную на морфологической модели CSE (Полный набор окончаний) для этапа преобразования текста в текст. Научным вкладом данной работы является исследование технологии формирования синтетических параллельных речевых корпусов тюркских языковых пар с каскадной схемой машинного перевода речи на реляционных моделях. В дальнейшем полученные синтетические корпуса параллельной речи будут использоваться для обучения нейронного машинного перевода речи тюркских языков.

**STUDY OF AUTOMATIC FORMATION OF TURKIC  
LANGUAGES SYNTHETIC SPEECH CORPORA****Ualsher Tukeyev, Zhandos Tolebekov, Tolganay Balabekova,  
Zhandos Zhumanov, Bigeldi Temirkhanov***Al-Farabi Kazakh National University, Almaty, Kazakhstan*  
ualsher.tukeyev@gmail.com, zhandoslp@gmail.com,  
t.balabekova@mail.ru, z.zhake@gmail.com,  
temirkhanov.bigeldi@gmail.com



Currently, the main tool for researching machine speech-to-speech translation is neural network technologies. However, for many languages, research into direct machine translation of speech is difficult due to the difficulty of generating parallel speech corpora for training neural networks. Research into machine translation of speech between Turkic languages is in practice absent due to the difficulty of creating parallel speech corpora for training neural models. Therefore, an urgent problem is the automatic generation of synthetic parallel corpora for training neural machine translation of speech in Turkic languages. This article proposes a technology for generating synthetic parallel speech corpora for Turkic languages using a cascade machine translation scheme using the example of Kazakh-Tatar, Kazakh-Turkish, Kazakh-Uzbek language pairs. The special feature of this cascading scheme is that it uses a relational model based on the CSE (Complete Set of Endings) morphological model for the text-to-text conversion step. The scientific contribution of this work is the study of the technology for generating synthetic parallel speech corpora of Turkic language pairs with a cascade scheme for machine speech translation on relational models. In the future, the resulting synthetic corpora of parallel speech will be used to train neural machine translation of speech in Turkic languages.

**Keywords:** Turkic languages, synthetic, parallel, speech corpora, machine translation

## 1. Введение

В настоящее время основным инструментом исследования машинного перевода речи в речь являются технологии нейронных сетей. Для многих языков исследование прямого машинного перевода речи затруднено из-за сложности формирования параллельных речевых корпусов для обучения нейронных сетей.

Исследования параллельных речевых корпусов для тюркских языков практически отсутствуют из-за сложности создания параллельных речевых корпусов для обучения нейронных моделей.

Поэтому в данной работе предлагается исследовать подход построения машинного перевода речи по каскадной схеме для формирования синтетических параллельных речевых корпусов, которые могут быть использованы для обучения нейронного машинного перевода речи в речь.

Научным вкладом данной работы является исследование технологии формирования синтетических параллельных речевых корпусов тюркской языковой пары по каскадной схеме машинного перевода речи на реляционных моделях. В дальнейшем полученные синтетические параллельные речевые корпуса будут использоваться для обучения нейронному машинному переводу речи тюркских языков.

## 2. Связанные работы

Машинный перевод речи осуществляется с использованием технологий обучения нейронных сетей. Однако нейронный машинный перевод имеет свои ограничения. Для получения качественных результатов нейронного машинного перевода (текстов или речи) необходимы большие объемы высококачественных параллельных корпусов для обучения нейронному машинному переводу. Сбор подобных параллельных корпусов, подготовленных профессиональными переводчиками на многие языки, очень трудоемкий. Хотя нейронный машинный перевод показал впечатляющие результаты для многих языков мира, проблема качественного машинного перевода для малоресурсных языков так и не решена. Поэтому разработка и исследование методов и средств, повышающих качество машинного перевода текста и речи для малоресурсных языков, остается весьма актуальной.

Проблема параллельных корпусов для обучения машинному переводу речи S2ST (Speech to Speech Translation) в настоящее время интенсивно изучается [Karakanta, 2020; Jia, 2022; Bentivogli, 2022]. Параллельные корпуса для S2ST (Speech to Speech Translation) создаются специально для этой задачи и требуют специального оборудования для записи речи и значительных финансовых затрат.

Особенно актуальна эта проблема для тюркских языков, поскольку параллельных речевых корпусов для пар тюркских языков практически нет. Однако одноязычные речевые корпуса уже появляются для отдельных языков тюркской группы: казахского, узбекского [Musaev, 2020; Mussakhoyayeva, 2021]. Существует несколько отечественных исследований распознавания речи для казахского, узбекского языков [Musaev, 2021; Мамырбаев, 2020; Мамырбаев, 2022; Мамырбаев, 2022; Khassanov, 2021].

Исследования нейронного машинного перевода речи для тюркских языков практически отсутствуют из-за трудностей создания параллельных речевых корпусов S2ST (Speech to Speech Translation) для обучения.

Поэтому, в данной статье предлагается автоматическое формирование параллельных корпусов речи путем построения машинного перевода речи S2ST (Speech to Speech Translation) по каскадной схеме, где этап ТТТ (Text-To-Text) предлагается решать для тюркских языков с использованием реляционных моделей машинного перевода. перевод текстов на основе новой моде-

ли морфологии по полному набору окончаний (CSE-модели). Затем эта система машинного перевода речи S2ST (Speech to Speech Translation) по каскадной схеме используется для формирования параллельного корпуса речи тюркских языков. Далее можно на полученных параллельных корпусах речи обучать нейронную модель машинного перевода речи тюркских языков.

### 3. Метод

Для разработки технологии речевого машинного перевода тюркских языков по каскадной схеме формирования параллельных корпусов речи тюркских языков в работе рассматриваются следующие задачи:

- разработка CSE-модели морфологии для каждого из выбранных тюркских языков;
- разработка реляционных моделей, алгоритмов и программ машинного перевода текстов на тюркские языки;
- подбор средств распознавания речи в текст (СТТ – Speech-To-Text) для казахского языка;
- подбор средств синтеза речи из текста (ТТС – Text-To-Speech) для татарского, турецкого и узбекского языков;
- разработка каскадной схемы машинного перевода текста с использованием реляционных моделей для этапа ТТТ.

Разработку CSE-модели морфологии рассмотрим на примере татарского языка.

#### 3.1 Создание лингвистических ресурсов для казахско-татарского машинного перевода текста в текст

Этапы машинного перевода текста в текст казахско-татарской языковой пары на основе модели CSE следующие:

1. разработка полного набора окончаний для казахско-татарских языков с использованием морфологической модели CSE
2. проведение морфологического анализа связей казахско-татарских языков, составление морфологической таблицы связей казахского языка и татарского языка
3. идентификация морфологической таблицы казахско-татарских языков.
4. составить список основополагающих слов казахско-татарского языка.
5. Создать список базы стоп-слов казахско-татарского языка.
6. создание алгоритма машинного перевода на казахско-татарский язык на основе модели CSE

7. создание программного обеспечения на основе алгоритма  
Перечисленные выше шаги анализируются отдельно ниже.

Вывод полного набора окончаний татарского языка. Вывод полного набора окончаний татарского языка. Для казахского языка разработан полный набор окончаний с использованием морфологической модели CSE [Tukeyev, 2020]. Слова в татарском языке имеют 3 больших союза, которые присоединяются к словам существительным: окончания множественного числа (К), притяжательные окончания (Т), падежные окончания (С), основы слов (S). Количество размещений определяется по формуле (1):

$$A_{nk} = n!/(n-k)! \quad (1)$$

Давайте рассмотрим все возможные варианты размещения суффиксных типов: один тип, два типа и три типа. Количество размещений на татарском языке – 7.

3 местоположения одного типа окончаний (К, Т, С) являются допустимым определением с точки зрения значения. Две разные концовки имеют 3 смысловых выгодных местоположения (КТ, ТС, КС). Из трех типов окончаний существует 1 семантически приемлемый тип размещения (КТС). Мы рассмотрели окончания в татарском языке как именные основы (существительные, прилагательные и числительные) и глагольные основы (глаголы, наречия, наречия). Полный набор татарских окончаний: всего – 2249.

Примеры вывода окончаний для татарского языка размещения КТ, КС для татарских окончаний представлены в таблицах 1–2.

**Таблица 1. Вывод окончаний размещений типа КТ  
(Plural-Possessive)**

| Tatar    | Suffixes type K | Suffixes type T |            | Number |
|----------|-----------------|-----------------|------------|--------|
|          |                 | Singular        | Plural     |        |
| Examples | нар-            | ым, ем          | ыбыз, ебез | 4*5=20 |
|          | нәр-            | ың, ең          | ыгыз, егез |        |
|          | лар-            | ы, е            | ы, е       |        |
|          | -нар-           | ым, ың, ы       | ыбыз, ыгыз | 5      |
|          | -нәр-           | ем, ең, е       | ебез, егез | 5      |
|          | -лар-           | ым, ың, ы       | ыбыз, ыгыз | 5      |
|          | -ләр-           | ем, ең, е       | ебез, егез | 5      |

Количество окончаний для размещения КТ равно 20.

**Таблица 2. Вывод окончаний размещений типа КС (Plural-Case)**

|          | Suffixes type K              | Suffixes type C   |  | Number |
|----------|------------------------------|---|--|--------|
| Examples | нар-<br>нәр-<br>лар-<br>ләр- | 1. nom.<br>2. gen.<br>3. dat.<br>4. acc.<br>5. loc.<br>6. abl.<br>7. gen. | –<br>ның, нең<br>га, ге<br>ны, не<br>да, дә<br>дан, дән<br>белән | 4*6=24 |
|          | -нар-                        | -ның, га, ны, да, дан, белән  |  | 6      |
|          | -нәр-                        | -нең, ге, не, дә, дән, белән  |  | 6      |
|          | -лар-                        | -ның, га, ны, да, дан, белән  |  | 6      |
|          | -ләр-                        | -нең, ге, не, дә, дән, белән  |  | 6      |

Разработка таблицы соответствия окончаний, стемов, стоп слов. С использованием коллекции казахско-татарских языковых окончаний на основе морфологической модели CSE была разработана соответствующая таблица морфологических признаков окончаний двух языков (табл. 3).

**Table 3. Morphological table of Kazakh-Tatar endings**

| Kazakh Endings | Kazakh Morph  | Tatar Morph  | Tatar Endings |
|----------------|---|--|---------------|
| dar            | <NB>*dar<pl>  | <NB>*нар<pl>   | нар           |
| m              | <NB>*m<pos><sg><p1>                                     | <NB>*м<pos><sg><p1>  | м             |
| ğa             | <NB>*ğa<dat>  | <NB>*га<dat>   | га            |
| myn            | <NB>*myn<per><sg><p1>                                   | <NB>* <per><sg><p1>  | empty         |
| darym          | <NB>*dar<pl><br>*ым<pos><sg><p1>                        | <NB>*нар<pl><br>*ым<pos><sg><p1>                           | нарым         |
| dary-<br>mamyn | <NB>*dar<pl><br>*ым<pos><sg><p1><br>*a<dat>*myn<sg><p1> | <NB>*нар<pl><br>*ым<pos><sg><p1><br>*a<dat>* <per><sg><p1> | нарыма        |

Согласно этому подходу выработано соответствие окончаниям каждого языка. Всего для соответствия казахских и татарских окончаний выявлено 5217 рядов окончаний.

Разработаны соответствия стемов казахских и татарских слов (табл. 4).

**Таблица 4. Стемы в казахском и татарском языках**

| stem in Qazaq | stem in Tatar |
|---------------|---------------|
| bir           | бер           |
| bala          | бала          |
| belgi         | билге         |

Разработана таблица соответствия стоп слов казахского и татарского языков (Table 5).

**Таблица 5. Стоп слова в казахском и татарском языках**

| SW in Qazaq | SW in Tatar |
|-------------|-------------|
| men         | мин         |
| jäne        | һәм         |
| ne          | яки         |

### 3.4 Выбор средств распознавания речи и синтеза речи.

Для инструментов распознавания речи в текст (STT – Speech-To-Text) для казахского языка взята тонко настроенная модель на основе предварительно обученной модели на основе преобразователя wav2vec2-large-xlsr-53 с коннекционистской временной классификацией (СТС) [Baevski, 2020]. Эта модель основана на идее самоуправляемого обучения, что означает, что она обучается на большом количестве немаркированных аудиоданных, а не на маркированных данных (данных с соответствующей транскрипцией).

Для синтеза речи использовался инструмент синтеза речи TurkicTTS с открытым исходным кодом [Mussakhojajeva, 2021].

## 4. Эксперименты и результаты

Ниже в таблице 6 представлены результаты машинного перевода речи с казахского языка на турецкий. Таблица соответствия окончаний для казахско-турецкой языковой пары содержит 36 слов и 9135 строк. Таблица соответствия стемов казахско-турецкой языковой пары состоит из 19846 слов. Допол-

нительно таблица соответствия стоп-слов казахско-турецкой языковой пары включает 198 слов.

**Table 6. Исходный текст речи, текст распознанной речи, оценка текста распознанной речи, стандарт перевода турецкой речи, машинный перевод речи, оценка машинного перевода речи**

| Qazaq (source text for speech)  | audio recognition   | Qazaq STT estimation                                   | Turkish(gold standard)   | Turkish machine translation   | Turkish machine translation estimation                  |
|---|---|--|--|---|---|
| 1   | 2   | 3  | 4  | 5   | 6   |
| Аян екеуіміз бір партаға отырғанбыз.  | аян екеуіміз бір партаға отырғанбыз   | WER: 5.71<br>BLEU: 50.81<br>TER: 20.00<br>chrF2: 86.71 | Ayan ve ben aynı masada oturuyorduk.   | ayan <b>ikimiz bile</b> bir <b>masaya</b> oturuyorduk.                          | WER: 41.67<br>BLEU: 13.13<br>TER: 66.67<br>chrF2: 47.43 |
| Сол күннен бастап күнде кешкісін ат қораның төбесіне жиналуды әдетке айналдырдық. | сол күннен бастап күнде кешкісін ат қораның төбесіне жиналуда әдетке айналдырды | WER: 4.94<br>BLEU: 53.88<br>TER: 18.18<br>chrF2: 90.97 | O günden sonra her akşam ahırın çatısında toplanmayı alışkanlık haline getirdik. | o günden beri itibaren hergün akşam isim ahırın çatıya toplama alışkanlık döndü | WER: 52.50<br>BLEU: 4.79<br>TER: 72.73<br>chrF2: 44.68  |
| Біз ертегілердің бәрін Аянның өзі күнделікті ойлап шығаратынын білдік.            | біз ертегілердің бәрін аянның өзі күнделікті ойлап шығаратынын білдік           | WER: 5.71<br>BLEU: 45.94<br>TER: 22.22<br>chrF2: 88.44 | bütün masaları her gün Aya'nın uydurduğunu öğrendik.                             | biz masaların her şey akortnın өз günlük düşünmek belkianı bilmekmadık          | WER: 79.25<br>BLEU: 4.20<br>TER: 128.57<br>chrF2: 29.60 |
| Қалың ойда жатқан сияқты.   | қалың ойда жатқан сияқты  | WER: 4.00<br>BLEU: 66.87<br>TER: 0.00<br>chrF2: 94.83  | kalın akılda uzanmak biraz.  | kalın düşünceda yatmakanlar biraz   | WER: 50.00<br>BLEU: 19.00<br>TER: 50.00<br>chrF2: 38.14 |
| Кешкісін жерден қазылған баракқа келеміз.   | кешкісін жерден қазылған баракқа келеміз  | WER: 7.32<br>BLEU: 19.36<br>TER: 40.00<br>chrF2: 84.71 | akşam yerden kazılmış sayfaya geleceğiz  | akşam yerden kazılmış sayfaya geleceğiz   | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 |
| Қараңғылық ерте түскен.   | қараңғылық ерте түскен  | WER: 4.55<br>BLEU: 55.03<br>TER: 0.00<br>chrF2: 94.23  | Karanlık Erken düşmüş  | Karanlık Erken düşmüş   | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 |
| Аян баракта төрт күн жатып қалды  | аян баракта төрт күн жатып қалды  | WER: 3.12<br>BLEU: 75.98<br>TER: 0.00<br>chrF2: 95.90  | Ayan sayfada dört gün uzanmak sol  | ayan sayfada dört gün uzanmak sol   | WER: 3.03<br>BLEU: 75.98<br>TER: 0.00<br>chrF2: 96.06   |

## Продолжение таблицы 6

| 1   | 2  | 3  | 4   | 5  | 6   |
|---|--|--|---|--|---|
| Осыдан кейін Аян тобығын қайтып орнына салдарған жоқ. | осыдан кейін аян тобыған қайтып орнына салдарған жоқ | WER: 9.43<br>BLEU: 30.51<br>TER: 37.50<br>chrF2: 77.02 | buradan sonrasında mafya ger yerine sebebiyle yok | buradan birazın ayan mafya geri yeriya sonuçlaranlar yok | WER: 48.00<br>BLEU: 6.57<br>TER: 85.71<br>chrF2: 36.96  |
| Көпке дейін ұйықтай алмадым                           | көпке дейін ұйықтай алмадым                          | WER: 3.70<br>BLEU: 59.46<br>TER: 0.00<br>chrF2: 95.32  | uzun süre uyuyamadım                              | uzun süre uyumak al onumadık                             | WER: 35.00<br>BLEU: 31.95<br>TER: 66.67<br>chrF2: 62.78 |
| Ағамның иісі сіңіп қапты                              | ағамның иісі сіңіп қапты                             | WER: 8.33<br>BLEU: 19.00<br>TER: 25.00<br>chrF2: 74.60 | kardeşlerim koku kırık çanta                      | kardeşmin koklamaka kırık çanta                          | WER: 39.29<br>BLEU: 31.95<br>TER: 50.00<br>chrF2: 52.10 |

Распределение ошибок по этапам. Эксперимент включал три этапа: преобразование речи в текст (STT), перевод текста в текст (TTT) и преобразование текста в речь (TTS). Ошибки наблюдались на каждом этапе, причем самый высокий процент ошибок в среднем приходится на этап TTT (58%), за которым следует этап STT (42%). На этапе TTS ошибок не было. В среднем оценки для казахско-турецкой пары: WER: 26,65; BLEU: 47,96; TER: 31,76; chrF2: 76,44.

В таблице 7 представлены результаты машинного перевода речи с казахского языка на узбекский. Таблица соответствия окончаний пары казахско-узбекский языков насчитывает 6042 строки. Таблица соответствия стемов пары казахско-узбекский языков насчитывает 20285 слов. Таблица соответствия стоп-слов для казахско-узбекской языковой пары насчитывает 198 слов.

**Таблица 7. Исходный текст речи, текст распознанной речи, оценка текста распознанной речи, стандарт перевода узбекской речи, машинный перевод речи, оценка машинного перевода речи**

| Qazaq (source text for speech) | audio recog-nition            | Qazaq STT estimation                                    | Uzbek (gold standard)             | Uzbek ma-chine transla-tion                 | Uzbek machine translation estimation                   |
|--------------------------------|-------------------------------|---|-----------------------------------|---|--|
| 1                              | 2                             | 3   | 4                                 | 5   | 6  |
| Мен шай ішкенді жақсы көремін  | мен шай ішкенді жақсы көремін | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 | Men choy ichishni yaxshi ko'raman | men choy <i>ichishkanni</i> yaxshi ko'raman | WER: 9.09<br>BLEU: 30.21<br>TER: 20.00<br>chrF2: 88.12 |



## Продолжение таблицы 7

| 1   | 2   | 3   | 4   | 5  | 6   |
|---|---|---|---|--|---|
| Ол университетте француз тилин үйренуде   | ол университетте француз тилин үйренуде   | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 | U universi-tetda fransuz tilini o'rganmoqda.        | u universi-tetda fransuz tilini <b>o'rgan-ishda</b>          | WER: 9.52<br>BLEU: 39.76<br>TER: 20.00<br>chrF2: 84.39  |
| Ол эр сенби күні футбол ойнайды           | ол эр сенби күні футбол ойнайды           | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 | U har shanba kuni futbol o'yнайdi.                  | u har shanba kuni <b>футбу</b> o'yнайdi                      | WER: 18.75<br>BLEU: 53.73<br>TER: 16.67<br>chrF2: 65.80 |
| Өткен жазда биз жагажайға бардык          | өткен жазда биз жагажайға бардык          | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 | O'tgan yozda biz plyajga bordik.                    | o'tgan yozda biz plyajga bor-dik                             | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 |
| Олар қазір кино тамашалауда               | олар қазір кино тамашалауда               | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 | Ular hozir kino tomosha qilishmoqda.                | ular hozir kino tomosha qilish-moqda                         | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 |
| Менің сүйікті түсім көк                   | менің <b>сүйікі</b> түсім көк             | WER: 8.70<br>BLEU: 31.95<br>TER: 25.00<br>chrF2: 73.91  | Mening se-vimli rangim ko'k.                        | mening <b>сүйікі</b> rangim ko'k                             | WER: 23.08<br>BLEU: 35.36<br>TER: 25.00<br>chrF2: 58.10 |
| Ол бос уақытында кітап оқуды жақсы көреді | ол бос уақытында кітап оқуды жақсы көреді | WER: 0.00<br>BLEU: 100.00<br>TER: 0.00<br>chrF2: 100.00 | U bo'sh vaq-tlarida kitob o'q-ishni yaxshi ko'radi. | u bo'sh vaq-tlarida kitob <b>o'qish-moqda</b> yaxshi ko'radi | WER: 14.29<br>BLEU: 37.68<br>TER: 28.57<br>chrF2: 77.06 |

Общее количество слов на казахском языке – 73. Общее количество ошибочных слов в машинном переводе на узбекский язык – 14. На этапе СТТ количество ошибочных слов – 6. На этапе ТТТ количество ошибочных слов – 8. На этапе ТТС количество ошибочных слов равно 0. Процент ошибок для этапа СТТ составит  $6/73 = 8\%$ , для этапа ТТТ будет  $8/73 = 10\%$ . Влияние (процент) этапов перевода речи на общее качество: СТТ –  $6/14 = 42\%$ ; ТТТ –  $8/14 = 58\%$ ; ТТС –  $0/14 = 0\%$ . В среднем оценки для казахско-узбекской пары: WER: 15.98; BLEU: 47.23; TER: 27.30; chrF2: 75.43.

Четыре метрики WER, BLEU, TER и chrF2 использовались для оценки с использованием инструмента Sacrebleu [Sacrebleu, 2023]. Неправильно распознанные и неверно переведенные слова выделены жирным шрифтом. Слова, переведенные синонимами, выделены жирным шрифтом и курсивом.

В первом столбце таблицы 6 приведены предложения на казахском языке, которые необходимо озвучить. Во втором

столбце представлен результат этапа распознавания речи. В третьем столбце таблицы представлена оценка распознавания речи по указанным метрикам. В четвертом столбце таблицы представлен «золотой стандарт» перевода предложений на турецкий, узбекский язык. В пятом столбце таблицы представлен результат машинного перевода речи на турецкий, узбекский язык. Последний столбец представляет оценку машинного перевода речи.

Результаты экспериментов казахско-турецкой и казахско-узбекской пар по четырем использованным метрикам примерно близки в среднем. В целом, эксперимент показал, что предлагаемая технология перевода речи с казахского на турецкий, узбекский языки является многообещающей, о чем свидетельствуют относительно высокие показатели BLEU и chrF2.

## 5. Заключение

В статье представлено исследование возможности автоматического формирования синтетического параллельного корпуса каскадной схемой машинного перевода речи тюркских языков на примере татарского, турецкого и узбекского языков. В данной каскадной схеме фаза машинного перевода текст-в-текст выполняется на реляционных моделях по CSE модели морфологии [Tukeyev, 2020]. Результаты экспериментов показывают о возможности автоматического формирования параллельных корпусов речи тюркских языков предложенным методом.

## ЛИТЕРАТУРА

1. Karakanta, A., Negri, M., Turchi, M. (2020) MuST-Cinema: a Speech-to-Sub-titles corpus. LREC 2020: 3727–3734
2. Jia, Y., Ramanovich, M.T., Wang, Q., and Zen, H. (2022) Cvss corpus and massively multilingual speech-to-speech translation. arXiv preprint arXiv:2201.03713.
3. Bentivogli, L., Mauro Cettolo, Marco Gaido, Alina Karakanta, Matteo Negri, Marco Turchi: Extending the MuST-C Corpus for a Comparative Evaluation of Speech Translation Technology. EAMT 2022: 359-360
4. Musaev, M., Mussakhoyayeva, S., Khujayorov, I., Khassanov, Y., Ochilov, M., & Varol, H. A. (2020). USC: An Open-Source Uzbek Speech Corpus and Initial Speech Recognition Experiments. arXiv preprint arXiv:2107.14419.

5. Mussakhoyayeva, S., Janaliyeva, A., Mirzakhmetov, A., Khassanov, Y. and Varol, H.A. (2021) KazakhTTS: An Open-Source Kazakh Text-to-Speech Synthesis Dataset. Proc. Interspeech 2021, 2786-2790, doi: 10.21437/Interspeech.2021-2124
6. Mamyrbayev, O., Alimhan, K., Zhumazhanov, B., Turdalykyzy, T., Gusmanova, F. (2020) End-to-End Speech Recognition in Agglutinative Languages. *ACIIDS* (2) 2020: 391-401
7. Mamyrbayev, O., Alimhan, K., Oralbekova, D., Bekarystankyzy A., Zhuma-zhanov, B. (2022). Identifying the influence of transfer learning method in developing an end-to-end automatic speech recognition system with a low data level. *Eastern-European Journal of Enterprise Technologies*, 19(115), 84–92. <https://doi.org/10.15587/1729-4061.2022.252801> (Scopus, процентиль 56)
8. Мамырбаев О.Ж., Оралбекова Д.О., Алимхан К., Оthman М., Жумажанов Б. Применение гибридной интегральной модели для распознавания казахской речи // *News of the National academy of sciences of the republic of Kazakhstan*. – 2022. – Vol. 1, № 341. – P. 58–68 // doi.org/10.32014/2022.2518-1726.117.
9. Khassanov, Y., Mussakhoyayeva, S., Mirzakhmetov, A., Adiyev, A., Nurpe-iissov, M., Varol, H.A.: A crowdsourced open-source Kazakh speech corpus and initial speech recognition baseline. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. pp. 697–706. Association for Computational Linguistics, 2021. <https://is-sai.nu.edu.kz/ru/%d0%b3%d0%b%d0%b0%d0%b2%d0%bd%d0%b0%d1%8f/#research>.
10. Tukeyev, U., Karibayeva, A. (2020) Inferring the Complete Set of Kazakh End-ings as a Language Resource. In: *Hernes M., Wojtkiewicz K., Szczerbicki E. (eds) Advances in Computational Collective Intelligence. ICCCI 2020. Communications in Computer and Information Science*, vol 1287, pp.741-751. Springer, Cham. [https://doi.org/10.1007/978-3-030-63119-2\\_60](https://doi.org/10.1007/978-3-030-63119-2_60)
11. Baeviski, A., Zhou H., Mohamed A., and Auli, M. “wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations.” arXiv, Oct. 22, 2020. doi: 10.48550/arXiv.2006.11477.
12. Mussakhoyayeva, S., Janaliyeva, A., Mirzakhmetov, A., Khassanov, Y. and Varol, H. A. KazakhTTS: An Open-Source Kazakh Text-to-Speech Synthesis Dataset. arXiv:2104.08459v3 [eess.AS] 16 Jun 2021.
13. Sacrebleu. <https://github.com/mjpost/sacrebleu>. Access date: March 1, 2023.

УДК 81'322.2

**ВОЗМОЖНОСТИ СИСТЕМЫ УПРАВЛЕНИЯ  
КОРПУСНЫМИ ДАННЫМИ ДЛЯ РАБОТЫ С КОРПУСОМ  
КРЫМСКОТАТАРСКОГО ЯЗЫКА**

***Д. Р. Мухамедшин<sup>1</sup>, Б. Э. Хакимов<sup>1</sup>, Л. Ш. Кубединова<sup>2</sup>***

*<sup>1</sup>Институт прикладной семиотики АН РТ,  
Россия, Татарстан, Казань*

*<sup>2</sup>Институт прикладной семиотики АН РТ,  
Россия, Крым, Симферополь*

*damirmuh@gmail.com, khakeem@yandex.ru,  
kubedinova@gmail.com*

Система управления корпусными данными разработана специально для работы с лингвистическими корпусами. Поисковый функционал, предлагаемый системой корпус-менеджер, включает в себя поиск лексических единиц, морфологический поиск, лексико-морфологический поиск, поиск синтаксических единиц, поиск n-грамм с учетом грамматики, поиск по метаданным и др. Также система позволяет производить сложные выборки из базы корпусных данных при помощи консольных утилит. Новым опытом для авторов стала интеграция корпуса крымскотатарского корпуса в систему управления корпусными данными. В статье описываются первые этапы интеграции, рассматриваются возможности системы для работы с корпусом крымскотатарского языка.

**Ключевые слова:** система управления корпусом, корпусные данные, корпусная лингвистика, поисковая система.

**CORPUS DATA MANAGEMENT SYSTEM CAPABILITIES FOR  
WORKING WITH CRIMEAN TATAR LANGUAGE CORPUS**

***Mukhamedshin D. R.<sup>1</sup>, Khakimov B. E.<sup>1</sup>, Kubedinova L. Sh.<sup>2</sup>***

*Institute of Applied Semiotics of the AS of the RT,  
Russia, Tatarstan, Kazan*

*Institute of Applied Semiotics of the AS of the RT,  
Russia, Crimea, Simferopol*

*damirmuh@gmail.com, khakeem@yandex.ru,  
kubedinova@gmail.com*

The corpus data management system is designed specifically for working with linguistic corpora. The system functionality includes search for lexical units, morphological search, lexico-morphological search, search for syntactic units, search for n-grams with grammar, search for lists of word forms or lemmas, search with taking into account the metadata of documents, search with grouping by document, context, word form, lemma, morphological features. The system

also allows you to make complex selections from the database of the corpus data using console utilities. A new experience for the authors was the integration of the Crimean Tatar corpus into the corpus data management system. The article describes the first stages of integration and discusses the capabilities of the system for working with the Crimean Tatar language corpus.

**Keywords:** corpus management system, corpus data, corpus linguistics, search engine.

**Введение.** Основной целью разработанной системы управления корпусными данными [Nevzorova, 2017], является работа с разнообразными лингвистическими корпусами. Система управления корпусными данными с 2014 года работает с электронными корпусами текстов на татарском языке и позволяет подключать лингвистические корпусы на других агглютинативных и флективных языках. Функционал, предлагаемый системой, включает в себя поиск лексических единиц, морфологический поиск, лексико-морфологический поиск, поиск синтаксических единиц, поиск n-грамм с учетом грамматики, поиск списков словоформ или лемм [Mukhamedshin, 2017], поиск с учетом метаданных документов, поиск с группировкой по документу, контексту, словоформе, лемме, морфологическим признакам. Также в системе реализован функционал формирования частотных списков на основе поискового функционала, доступен открытый API для быстрого обмена данными с другими системами. Поисковые технологии реализованы на базе современных общедоступных программных средств: система управления базой данных MariaDB и хранилище данных Redis. Благодаря концептуальной модели представления корпусных данных, поиск в корпусе производится менее, чем за 0,05 сек. в 98,71% случаев.

Разработанная система управления корпусными данными ориентирована в первую очередь на поддержку электронных корпусов тюркских языков, что является весьма актуальным для активно развивающегося направления тюркской корпусной лингвистики. В 2023 году авторами предпринята попытка интеграции корпуса крымскотатарского языка [Кубединова, 2016] в систему управления корпусными данными. В данной статье будут описаны схема работы и архитектура системы управления корпусными данными, а также первые этапы интеграции: установка системы управления корпусными данными, основные настройки, новые возможности управления корпусными данными, добавленные для более простого подключения новых корпусов.

**Схема работы системы управления корпусными данными.** Для программной реализации системы управления корпусными данными используется парадигма MVC [Leff, 2001] (Model-View-Controller, Модель-Представление-Контроллер), которая была несколько изменена для решения задач системы. В парадигме MVC пользовательский ввод, моделирование внешнего мира и визуальная обратная связь с пользователем явно разделены и обрабатываются тремя типами объектов, каждый из которых специализируется на выполнении своей задачи. Представление управляет текстовым или графическим выводом. Контроллер интерпретирует вводимые пользователем данные, отдавая команду Модели или Представлению на необходимые изменения. Наконец, Модель управляет поведением и данными приложения, отвечает на запросы информации о его состоянии и отвечает на инструкции по изменению состояния. Абстрактная схема взаимодействия между компонентами системы представлена на Рисунке 1.

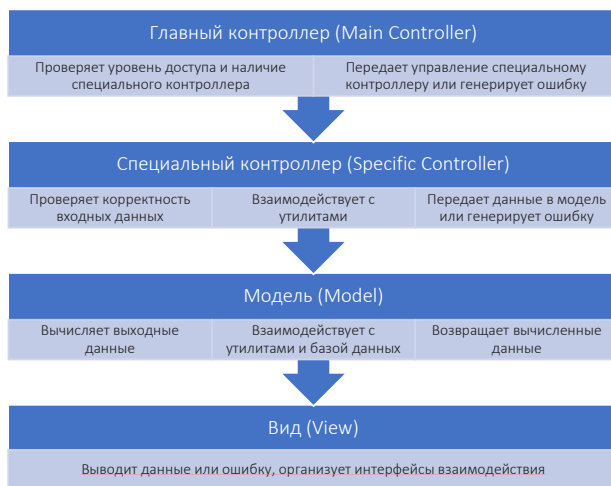


Рис. 1. Абстрактная схема выполнения задания (взаимодействия между компонентами) в системе управления корпусными данными

Такая архитектура четко разбивает решение задачи на несколько этапов, каждый из которых важен для решения поставленной перед системой управления корпусными данными задачи:

1. Контроллеры проверяют входные данные от пользователя, обеспечивая второй уровень безопасности, и фильтруют выходные данные, обеспечивая четвертый уровень безопасности. Кроме того, контроллеры являются связующим звеном между моделями и представлением.

2. Модели выполняют запросы к БД и реализуют функционал бизнес-логики, если это необходимо. Некоторые модели могут использовать другие модели и утилитарные функции. Не все модели используются непосредственно контроллерами.

3. Вид реализует функции вывода данных и ошибок и хранит в своем объекте необходимые для этого данные.

**Архитектура системы управления корпусными данными.**

Подробная архитектура системы управления корпусными данными представлена на Рисунке 2, а функционал компонентов кратко описан в Таблице 1.

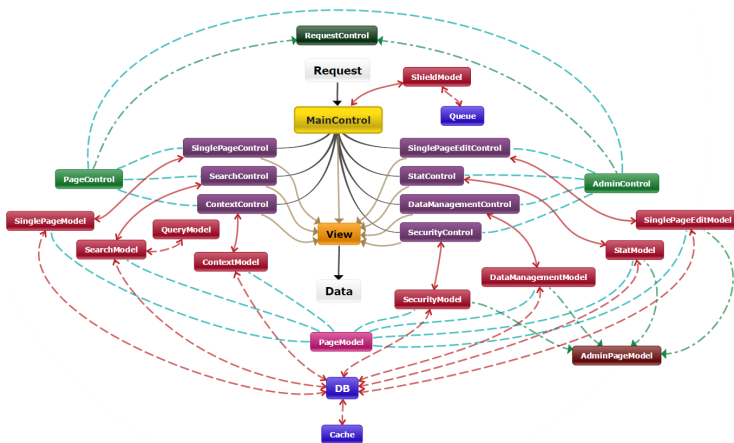


Рис. 2. Архитектура системы управления корпусными данными

**Таблица 1. Функционал компонентов системы управления корпусными данными**

| № | Название компонента               | Функционал   |
|---|-----------------------------------|--|
| 1 | 2                                 | 3  |
| 1 | MainControl (основной контроллер) | Отвечает за безопасность первого уровня, проверяя уровень доступа пользователя, а также обеспечивая контроль за паразитным трафиком. |

## Продолжение таблицы 1

| 1 | 2                     | 3  |
|---|-----------------------|--|
| 2 | PageControl           | Абстрактный контроллер PageControl наследуется всеми контроллерами, которые отвечают за действие, связанное с открытием какой-либо страницы (документа). В нем должны быть описаны общие процедуры контроля и абстрактная структура дочерних контроллеров. |
| 3 | SinglePageControl     | Контроллер статичных страниц отвечает за обработку действия «открытие статичной страницы». Проверяет наличие необходимых для этого входных данных и передает их модели SinglePageModel. Если страница существует, передает данные Виду.                    |
| 4 | SearchControl         | Обеспечивает защиту от инъекций, фильтруя входные данные (второй уровень защиты). После фильтрации безопасные данные передаются модели поиска SearchModel. Если существуют результаты поиска, передает их и метаданные Виду.                               |
| 5 | ContextControl        | Обрабатывает действие «расширение контекста». Обеспечивает защиту от получения полного содержимого документов. Использует модель контекста ContextModel, если контекст получен, передает его Виду.   |
| 6 | AdminController       | Это расширение абстрактного контроллера PageControl. Основной задачей контроллера AdminControl является предотвращение несанкционированного доступа и контроль за общими действиями администратора системы. Позволяет снизить количество ошибок оператора. |
| 7 | SinglePageEditControl | Контроллер SinglePageEditControl обрабатывает действие «редактирование статичной страницы». Использует модель SinglePageEditModel для выполнения задачи, поставленной администратором.   |
| 8 | StatControl           | Контроллер статистики StatControl обрабатывает входные данные от администратора и обеспечивает выполнение действий, связанных с просмотром статистики системы. Использует модель StatModel для получения данных.   |
| 9 | DataManagementControl | Этот контроллер обрабатывает действия, связанные с данными системы. Использует модель DataManagementModel для выполнения поставленной задачи.  |



## Продолжение таблицы 1

| 1  | 2                    | 3  |
|----|----------------------|--|
| 10 | SecurityControl      | Контроллер SecurityControl обрабатывает действия, связанные с защитой системы. Использует модель SecurityModel для выполнения поставленной задачи.                     |
| 11 | RequestControl       | Хранит в себе основные функции для обработки запросов по установленным правилам. Очень важный элемент для защиты от инъекций.  |
| 12 | SinglePageModel      | Получает контент статичной страницы из БД.   |
| 13 | SearchModel          | Наиболее важная модель в системе. Производит поиск по заданному запросу. Каждый запрос обрабатывается моделью запросов QueryModel, после чего производится поиск в БД. |
| 14 | QueryModel           | Модель запросов преобразует запросы пользователей системы в объект запроса, оптимизируя время его выполнения.  |
| 15 | ContextModel         | Модель контекстов ContextModel получает контекст установленного размера из БД.   |
| 16 | SinglePageEdit Model | Выполняет действия над статичными страницами: добавление, редактирование, удаление. Результат действия записывается в БД.  |
| 17 | StatModel            | Организует объект статистики системы, получая ее из БД. Выполняет действия над статистикой: просмотр, обнуление, настройка ведения статистики.                         |
| 18 | DataManagement Model | Выполняет действия над данными системы: просмотр, добавление, редактирование, удаление. Результат действия записывается в БД.  |
| 19 | SecurityModel        | Выполняет настройку защиты системы. Основные задачи: добавление, редактирование и удаление правил по IP, подсетям, интенсивности и др.                                 |
| 20 | PageModel            | Абстрактная модель страницы PageModel организует общую структуру и процедуры, присущие странице. Основное предназначение – структурирование объекта страницы для Вида. |
| 21 | AdminPageModel       | Хранит в себе функции, присущие странице администрирования.  |

Продолжение таблицы 1

| 1  | 2     | 3  |
|----|-------|--|
| 22 | DB    | Модель базы данных полностью абстрагирует работу с БД от других элементов системы. Организует подключение и основные функции для работы с БД.  |
| 23 | Cache | Модель кэша используется в модели базы данных для быстрой обработки частых запросов. Абстрагирует от других элементов системы основные действия, связанные с кэшированием.   |
| 24 | View  | Обеспечивает вывод страниц и ошибок, либо интерфейса, если используется API. Объект Вида сильно упрощен, основные действия происходят в шаблонах Вида. Шаблоны не используют какой-либо шаблонизатор для увеличения скорости вывода страниц. |

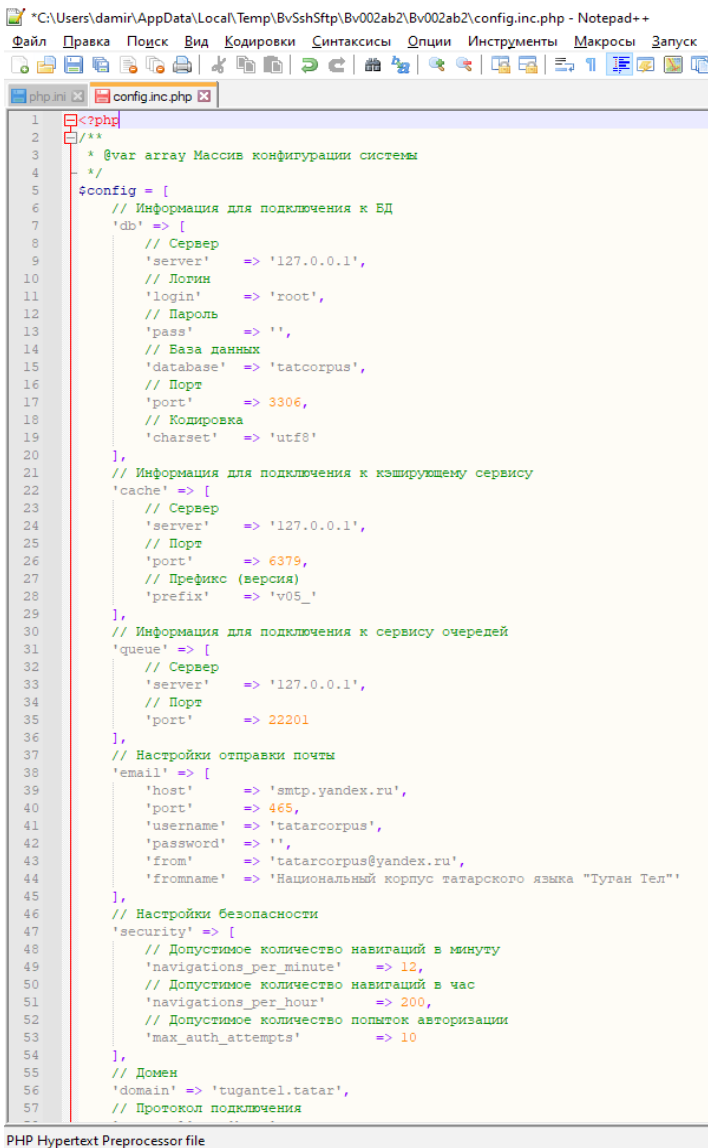
**Установка системы управления корпусными данными и основные настройки.** Благодаря постоянной оптимизации системы [Mukhamedshin, 2020], она нетребовательна к вычислительным ресурсам и может быть запущена при выполнении следующих минимальных требований:

- CPU не менее 2 ядер;
- RAM не менее 4 Гб;
- SSD-накопитель со свободным пространством не менее 10 Гб (для корпуса объемом не более 1 000 000 словоформ);
- Установленная Unix-подобная операционная система;
- Установленный пакет ПО: Apache (версии не ниже 2.0), MySQL (версии не ниже 5.6) (или MariaDB версии не ниже 10.0), PHP (версии не ниже 5.6), Redis (версии не ниже 5.0).

После загрузки исходного кода на сервер, соответствующий минимальным требованиям, необходимо установить архитектуру базы данных. Для этого нужно создать базу данных посредством выполнения SQL-запроса в СУБД MySQL:

```
CREATE DATABASE tatcorpus CHARACTER SET utf8 COLLATE utf8_general_ci;
```

Затем необходимо импортировать структуру базы данных из имеющегося файла `structure.sql`. Для этого рекомендуется воспользоваться графическим клиентом СУБД MySQL (например, phpMyAdmin).



```
1 <?php
2 /**
3  * @var array Массив конфигурации системы
4  */
5 $config = [
6     // Информация для подключения к БД
7     'db' => [
8         // Сервер
9         'server' => '127.0.0.1',
10        // Логин
11        'login' => 'root',
12        // Пароль
13        'pass' => '',
14        // База данных
15        'database' => 'tatcorpus',
16        // Порт
17        'port' => 3306,
18        // Кодировка
19        'charset' => 'utf8'
20    ],
21    // Информация для подключения к кэшируемому сервису
22    'cache' => [
23        // Сервер
24        'server' => '127.0.0.1',
25        // Порт
26        'port' => 6379,
27        // Префикс (версия)
28        'prefix' => 'v05_'
29    ],
30    // Информация для подключения к сервису очередей
31    'queue' => [
32        // Сервер
33        'server' => '127.0.0.1',
34        // Порт
35        'port' => 22201
36    ],
37    // Настройки отправки почты
38    'email' => [
39        'host' => 'smtp.yandex.ru',
40        'port' => 465,
41        'username' => 'tatarcorpus',
42        'password' => '',
43        'from' => 'tatarcorpus@yandex.ru',
44        'fromname' => 'Национальный корпус татарского языка "Туган Тел"'
45    ],
46    // Настройки безопасности
47    'security' => [
48        // Допустимое количество навигаций в минуту
49        'navigations_per_minute' => 12,
50        // Допустимое количество навигаций в час
51        'navigations_per_hour' => 200,
52        // Допустимое количество попыток авторизации
53        'max_auth_attempts' => 10
54    ],
55    // Домен
56    'domain' => 'tugantel.tatar',
57    // Протокол подключения
58    'protocol' => 'https'
```

RHP Hypertext Preprocessor file

Рис. 3. Вид файла конфигурации системы. Выделены основные настройки, которые необходимо изменить

Далее необходимо открыть для редактирования файл `engine/config.inc.php` (Рисунок 3). Согласно комментариям в файле, нужно указать настройки сервера MySQL [Christudas, 2019], Redis [Liu, 2019], домен, протокол подключения. В большинстве случаев достаточно изменить логин, пароль и название базы данных, остальные настройки можно оставить по умолчанию. Если требуется, можно также настроить соответствие между обозначениями морфологических свойств и соответствующим номером в битовом векторе морфологических свойств. Система поддерживает до 128 различных морфологических свойств.

**Возможности управления корпусными данными.** Загрузка корпусных данных в систему управления является самым продолжительным процессом. Ранее система управления корпусными данными поддерживала ручную загрузку корпусных данных при помощи специализированной подсистемы, запускаемой в консоли сервера. В целях упрощения этого процесса и обеспечения пользователей возможностью редактирования документов корпуса непосредственно в интерфейсе системы управления корпусными данными, был разработан административный раздел управления документами корпуса. Интерфейс данного раздела для корпуса крымскотатарского языка представлен на Рисунке 4.

The screenshot shows the 'Tugan Tel' corpus management interface. At the top, there is a navigation bar with links for 'Рус', 'Тат', 'Енг', 'Корпусы', 'Поиск', 'Публикации', 'Инструкции', and 'Документы'. A search bar is on the right. Below the navigation bar, there is a section titled 'Документы корпуса [+]' with a 'Фильтры:' button and several dropdown menus for 'Все статусы', 'Все источники', and 'Название файла или его часть'. A search button 'Искать' is also present. The main content is a table with the following columns: 'ID', 'Файл', 'Источник', 'Статус метаданных', 'Дата создания', and 'Дата изменения'.

| ID ↓ | Файл ↓   | Источник     | Статус метаданных | Дата создания ↑ ↓   | Дата изменения ↑ ↓  |
|------|--|--------------|-------------------|---------------------|---------------------|
| 1    | 24.04.2023 22:40:45.txt                                | Другие       | Нет метаданных    | 24.04.2023 22:40:45 | 24.04.2023 22:41:37 |
| 3    | 1. A.M.Gorkiy Umi.corpus.crt.docx                      | Сканирование | Проверены         | 23.05.2023 14:58:43 | 23.05.2023 14:58:43 |
| 5    | 1.txt  | Сканирование | Проверены         | 23.05.2023 16:25:45 | 23.05.2023 16:25:45 |
| 6    | 1. A.M.Gorkiy Umi.corpus.crt.txt                       | Другие       | Проверены         | 26.06.2023 13:09:39 | 26.06.2023 13:09:39 |
| 7    | 2. A.M.Gorkiy Kyzchik(Krymskiye eskizy).corpus.crt.txt | Другие       | Проверены         | 26.06.2023 13:13:23 | 26.06.2023 13:13:23 |
| 8    | 3. A.M.Gorkiy Bahyt.corpus.crt.txt                     | Другие       | Проверены         | 26.06.2023 13:15:00 | 26.06.2023 13:15:00 |

Рис. 4. Интерфейс управления документами в системе управления корпусными данными

В интерфейсе показаны существующие документы корпуса с указанием источника (поддерживаются источники «Сканирование», «Web» и «Другие»), статуса проверки метаданных («Нет метаданных», «Ожидающие проверки», «Проверенные»), даты создания и последнего изменения документа. Также имеется воз-

возможность поиска по статусу проверки метаданных, источнику и названию файла документа. При нажатии на название необходимого документа, пользователь может перейти к его редактированию. Если нужно добавить новый документ, необходимо нажать на кнопку добавления документа «[+]».

Интерфейс редактирования документа представлен на Рисунках 5 и 6. Система позволяет редактировать содержимое документа с разметкой, загружать новую версию документа, а также хранить документы прошлых версий, обеспечивая возможность вернуться к предыдущим версиям документа для поиска возможных ошибок и уточнения метаданных. Если документ был загружен в формате PDF или TXT, его содержимое будет отображено непосредственно в интерфейсе системы. Если формат документа отличается от вышеуказанных, то будет предложено загрузить документ на компьютер для дальнейшей работы с ним в сторонних приложениях.

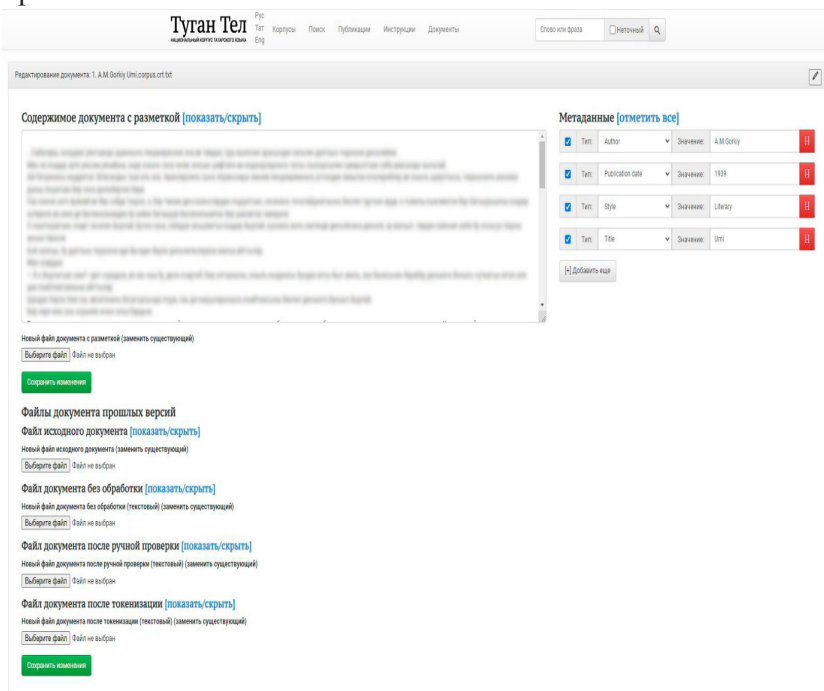


Рис. 5. Интерфейс редактирования документа. Основная часть

Системой поддерживаются следующие версии документа:

- Исходный документ – эта версия содержит первоначальную версию документа, как правило, без морфологической разметки.
- Документ без обработки – эта версия документа актуальна для сканированных документов с автоматическими распознанным текстом. В этой версии могут содержаться технические ошибки, связанные с алгоритмами распознавания, морфологическая разметка также в большинстве случаев отсутствует.
- Документ после ручной проверки – эта версия документа сохраняется после ручной проверки и исправления ошибок экспертом.
- Документ после токенизации – эта версия документа сохраняется после автоматической токенизации и разделения на предложения.
- Документ с разметкой – версия документа, участвующая в поиске по корпусу. Эта версия документа должна включать в себя морфологическую разметку [Гатауллин, 2018], если поддерживается поиск по морфологическим свойствам.

Также в интерфейсе редактирования документа имеется возможность указания метаданных документа [Nevzgorova, 2016] (Рисунок 6). Пользователь может добавить свой тип метаданных и указать значение, соответствующее редактируемому докумен-

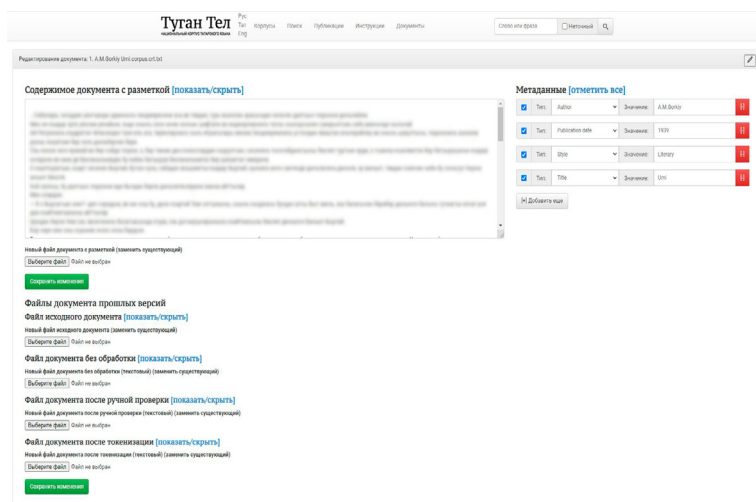


Рис. 6. Интерфейс редактирования документа.  
Редактирование метаданных

ту. Этот же функционал используется для проверки метаданных (например, автоматически распознанных в web-документах). Правильность заполнения метаданных отмечается флажком рядом с той или иной парой «тип-значение» метаданных. Если пара «тип-значение» метаданных является некорректной, имеется возможность удалить ее при помощи кнопки «[-]». Пользователь может добавлять любое необходимое количество метаданных для каждого документа, при этом созданные ранее типы метаданных сохраняются в системе и предлагаются для выбора при заполнении метаданных новых документов.

Сохраненный документ сразу же начинает участвовать в поиске по соответствующему корпусу. При каждом сохранении происходит переиндексация документа, что обеспечивает такую же скорость поиска, как и после загрузки документов при помощи специализированной подсистемы, запускаемой в консоли сервера.

**Заключение.** Система управления корпусными данными, представленная в данной статье, применяется для работы с электронным корпусом татарского языка «Туган Тел» [Сулейманов, 2014]. Новые возможности подключения корпусов других языков позволяют существенно расширить исследовательские и прикладные функции системы управления корпусными данными. Интеграция корпуса крымскотатарского языка уже на текущем этапе открывает обширные возможности для исследования крымскотатарского языка. Авторами запланировано дальнейшее расширение возможностей для работы с лингвистическими корпусами в первую очередь тюркских языков.

Несмотря на расширение функционала системы управления корпусными данными, время, необходимое для обработки и выполнения поискового запроса системой, не превышает 0,05 сек. в 98,71% случаев для лексического поиска, в 77,71% случаев для морфологического поиска и в 98,08% случаев для лексико-морфологического поиска. Во многом таких показателей удалось достичь благодаря использованию предложенных автором новых методов и технологий хранения и обработки корпусных данных, впервые примененных в системах управления корпусными данными.

#### ЛИТЕРАТУРА

1. Гатауллин Р. Р. Гибридный морфологический анализатор татарского языка на основе правил и статистики // Научно-технический вестник Поволжья. – 2018. – №. 9. – С. 89–92.

2. Кубединова Л. Ш., Гатиатуллин А. Р. Морфологическая разметка крымскотатарского электронного корпуса (на опыте татарского) // Ученые записки Крымского федерального университета имени В.И. Вернадского. Филологические науки. – 2016. – Т. 2. – №. 3. – С. 380–384.
3. Сулейманов Д. Ш. и др. Размеченный корпус татарского языка» Туган тел»: аспекты реализации //Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2014. – 2014. – С. 88–93.
4. Christudas B. MySQL. – Apress, 2019. – С. 877–884.
5. Leff A., Rayfield J. T. Web-application development using the model/view/controller design pattern // Proceedings fifth ieee international enterprise distributed object computing conference. – IEEE, 2001. – С. 118–127.
6. Liu Q., Yuan H. A High Performance Memory Key-Value Database Based on Redis // J. Comput. – 2019. – Т. 14. – №. 3. – С. 170–183.
7. Mukhamedshin D., Nevzorova O., Kirillovich A. Using FLOSS for Storing, Processing and Linking Corpus Data // IFIP International Conference on Open Source Systems. – Cham : Springer International Publishing, 2020. – С. 177–182.
8. Mukhamedshin D., Nevzorova O., Khusainov A. Complex Search Queries in the Corpus Management System // Computational Collective Intelligence: 9th International Conference, ICCCI 2017, Nicosia, Cyprus, September 27-29, 2017, Proceedings, Part II 9. – Springer International Publishing, 2017. – С. 407–416.
9. Nevzorova O., Mukhamedshin D., Gataullin R. Developing corpus management system: architecture of system and database // Proceedings of the International Conference on Information and Knowledge Engineering (IKE). – The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2017. – С. 108–112.
10. Nevzorova O., Mukhamedshin D., Kurmanbakiev M. Semantic aspects of metadata representation in corpus manager system // Open Semantic Technologies for Intelligent Systems (OSTIS-2016). – 2016. – С. 371–376.



УДК 81'322.5

**TATSC: ПЕРВЫЙ БОЛЬШОЙ ОТКРЫТЫЙ РЕЧЕВОЙ КОРПУС ТАТАРСКОГО ЯЗЫКА*****Р. А. Гильмуллин<sup>1</sup>, Б.Э. Хакимов<sup>1,2</sup>, М. Р. Галимов<sup>1</sup>****<sup>1</sup>Институт прикладной семиотики Академии Наук Республики Татарстан, Казань, Россия**<sup>2</sup>Казанский федеральный университет, Казань, Россия*rinatgilmullin@gmail.com, khakeem@yandex.ru,  
magl.galimov@gmail.com

TatSC представляет собой первый речевой корпус татарского языка большого объема с открытым доступом для использования в задачах автоматического распознавания речи (ASR). До настоящего времени отсутствовали татарские речевые корпуса достаточного объема для разработки ASR моделей. Датасет TatSC был разработан в академической коллаборации с Институтом умных систем и искусственного интеллекта в Астане, Казахстан. Корпус состоит из трех частей, собранных из различных источников: предложения из сети Интернет, аудиокниги и данные, собранные методом краудсорсинга. В частности, с целью расширения корпуса за счет разнообразных голосов и фоновых шумов, был разработан бот Telegram с функциями озвучивания и оценки записей. Предложения были озвучены пользователями разного пола и возраста. Лексическое и грамматическое разнообразие в датасете обеспечивалось путем использования поисковых запросов в Татарском национальном корпусе “Туган тел” и анализа N-грамм. Планируется продолжить работу по сбору и проверке данных с целью увеличения объема и репрезентативности корпуса.

**Ключевые слова:** татарский язык, речевой корпус, автоматическое распознавание речи, лингвистическая репрезентативность, подготовка датасета, краудсорсинг

**TATSC: THE FIRST LARGE OPEN-ACCESS SPEECH CORPUS FOR TATAR LANGUAGE*****R. A. Gilmullin<sup>1</sup>, B. E. Khakimov<sup>1,2</sup>, M. R. Galimov<sup>1</sup>****<sup>1</sup>Institute of Applied Semiotics of the Tatarstan Academy of Sciences, Kazan, Russia**<sup>2</sup>Kazan Federal University, Kazan, Russia*rinatgilmullin@gmail.com, khakeem@yandex.ru,  
magl.galimov@gmail.com

Tatar speech corpus (TatSC) is the first large open-access speech corpus for Tatar, which may be used in automatic speech recognition (ASR) tasks. There are no open Tatar speech corpora of sufficient size to develop modern ASR models. The TatSC dataset has been developed as a part of the joint project in academic collaboration with Institute of Smart Systems and Artificial Intelligence, Astana, Kazakhstan. TatSC consists of three parts collected from different sources: web-based, crowdsourcing and audiobooks. In order to extend the dataset by including various voices and background noises, a Telegram bot with narration and evaluation functions was developed. The Tatar National Corpus “Tugan Tel” was utilized as one of the sources by using specific search queries and N-gram analysis to include various grammatical features and vocabulary. The sentences were narrated by speakers of different age and gender. We plan to continue the data collection and checking process to increase the volume and linguistic representativeness of the corpus.

**Keywords:** Tatar language, speech corpus, speech recognition, ASR, linguistic representativeness, dataset preparation, crowdsourcing

## 1. Введение

Татарский речевой корпус TatSC был разработан с целью обеспечения потребности в больших открытых речевых коопусах для использования в задачах автоматического распознавания речи. Существующие корпуса со свободным доступом имеют незначительный объем, так, татарский корпус на платформе Common Voice [Ardila et al., 2020] составляет немногим более 26 часов. Татарский язык является малоресурсным языком, что обуславливает проблему сбора достаточного количества языковых данных. Чтобы частично решить эту проблему и добиться большей лингвистической репрезентативности, мы использовали специально отобранные поисковые запросы в Татарском национальном корпусе «Туган тел» на основе N-граммного анализа распределения буквосочетаний, лексических и грамматических единиц.

## 2. Подготовка данных

Разработанный речевой корпус TatSC состоит из трех подкорпусов, собранных из различных источников: предложения из сети Интернет, аудиокниги и данные, собранные методом краудсорсинга. Всего было собрано 269.1 часов данных, состоящих из 217,914 речевых записей. Подробный состав корпуса приведен в Таблице 1.

Таблица 1. Состав речевого корпуса TatSC

| Показатель      | Интернет | Краудсорсинг | Аудиокниги | Всего   |
|-----------------|----------|--------------|------------|---------|
| Длительность    | 99,5 ч   | 146,1        | 23,5       | 269,1   |
| Кол-во записей  | 87425    | 110683       | 19806      | 217914  |
| Кол-во слов     | 540584   | 881168       | 171117     | 1592869 |
| Уникальных слов | 50719    | 12957        | 28214      | 68623   |

Данные корпуса были автоматически проверены с помощью модели распознавания речи многоязычной модели автоматического распознавания речи для тюркских языков ISSAI [Mussakhojayeva et al., 2023] на основе метрики CER (доля ошибочно распознанных символов). Записи с высоким значением CER были перепроверены и исправлены вручную. Кроме того, были применены и другие действия по редактированию, такие как преобразование цифр в слова и др.

Отдельные подкорпуса в итоговом датасете TatSC были поделены на тренировочный (training set) и оценочный набор (evaluation set), который в свою очередь состоит из набора разработки (development set), используемого для настройки и усовершенствования процесса обучения модели, и тестового набора (test set), используемого для получения результатов по метрикам. Каждый оценочный набор состоит примерно из семи часов данных.

### 3. Интернет-подкорпус

Данная часть датасета создана на основе набора предложений, собранных методом автоматического обхода с различных татароязычных веб-сайтов в рамках предыдущих экспериментов по созданию моделей для распознавания речи на татарском языке [Suleymanov et al., 2021]. В настоящий момент эта часть с общей продолжительностью 99,5 часов состоит из 87425 записей, содержащих 50719 уникальных слов. Распределение предложений по длине и записей по длительности представлено на рисунках 1а и 1б.

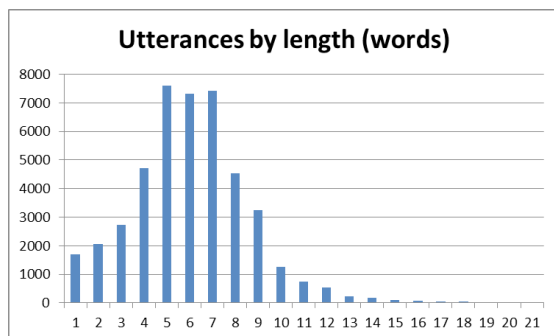


Рис. 1а. Распределение предложений по длине в Интернет-подкорпусе

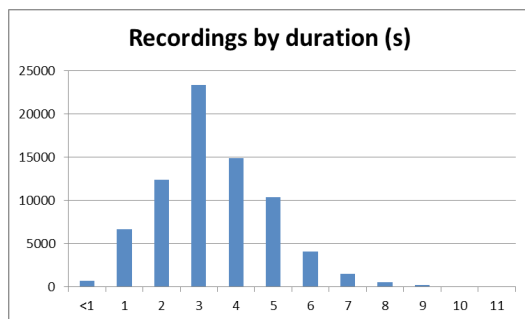


Рис. 1б. Распределение записей по длительности в Интернет-подкорпусе

Отобранные предложения были прочитаны и записаны носителями языка разного возраста и пола, а также с разным уровнем владения языком. Записи были выполнены на ноутбук с использованием гарнитуры без посторонних шумов. Данная часть корпуса TatSC организована в виде архива с папками, содержащими соответствующие файлы TXT и WAV с одинаковыми именами.

#### 4. Подкорпус аудиокниг

Общий объем данной части датасета составляет 23,5 часов и 19806 предложений, содержащих 28214 уникальных слов. Распределение предложений по длине и записей по длительности представлено на рисунках 2а и 2б.

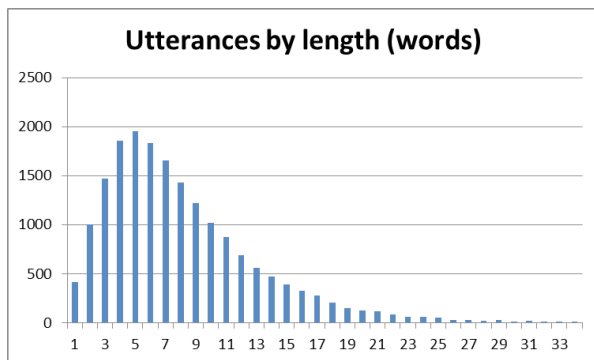


Рис. 2а. Распределение предложений по длине в подкорпусе аудиокниг

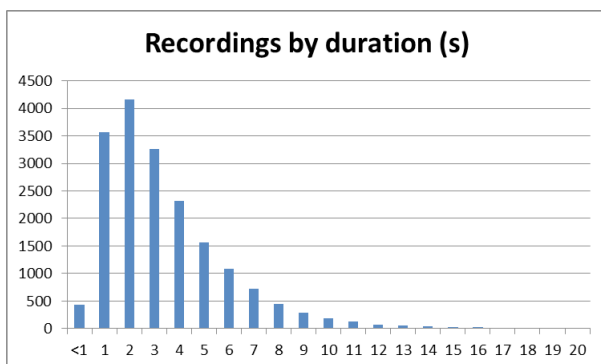


Рис. 2б. Распределение записей по длительности в подкорпусе аудиокниг

Для сбора данных из аудиокниг использовалось несколько источников. Основным источником послужила коллекция Татарского книжного издательства [Татарстан китап нәшрияты], опубликованная на официальном веб-сайте. Дополнительно привлекались материалы веб-сайтов с аудиокнигами в открытом доступе. Книги отбирались с учетом стиля, жанра и года публикации. В соответствии с экспертным выбором лингвиста, в выборку были включены книги, опубликованные с начала XX века по настоящее время, что соответствует периодизации современного татарского литературного языка [Бәширова һ.б., 2015]. Большинство книг

являются оригинальными сочинениями, небольшая часть представлена переводами. Все аудиокниги были записаны в студии профессиональными дикторами-женщинами и мужчинами, преимущественно театральными актерами. Процесс нарезки, аннотации и выравнивания был автоматизирован с помощью открытой платформы Label Studio [Label Studio]. К ручной проверке и выравниванию фраз из аудиокниг были привлечены носители языка с лингвистическим образованием.

Процесс подготовки подкорпуса аудиокниг состоял из следующих этапов:

- Выбор подходящей аудиокниги из списка доступных аудиокниг в онлайн-доступе;
- Поиск текстовой версии книги (в онлайн-библиотеках);
- Загрузка аудио и текста с помощью специальных скриптов для разных веб-сайтов;
- Предварительная нарезка аудио на отрезки по 4-5 минут и загрузка в соответствующий проект Label Studio с добавлением текста книги;
- Ручная нарезка, выравнивание и проверка фраз;
- Выгрузка финальной базы данных.

Подкорпус аудиокниг представлен в виде архива с папками, каждая из которых соответствует определенной книге и содержит вложенный папки TXT и WAV. Соотнесенные TXT и WAV файлы имеют одинаковые имена.

## **5. Краудсорсинговый подкорпус**

С целью расширения корпуса за счет разнообразных голосов и фоновых шумов, был разработан бот Telegram с функциями озвучивания и оценки записей. С помощью бота было собрано 146,1 часов записанной речи, в записи приняло участие более 650 носителей языка разного возраста и пола. Выборка предложений для записи содержала 12957 уникальных слов. Распределение предложений по длине и записей по длительности представлено на рисунках 3а и 3б.

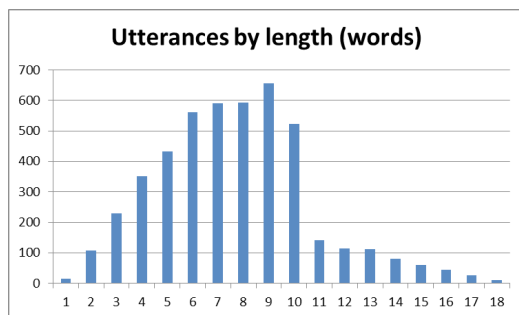


Рис. 3а. Распределение предложений по длине в краудсорсинговой части датасета

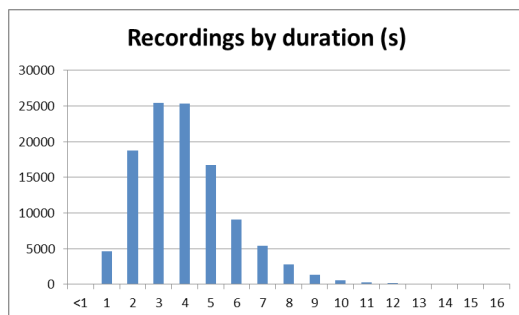


Рис. 3б. Распределение записей по длительности в краудсорсинговой части датасета

Одной из целей являлось построение репрезентативного подкорпуса на основе ограниченного, но лингвистически сбалансированного набора предложений. Предложения были экспортированы из Татарского национального корпуса “Туган тел”, который представляет собой репрезентативный текстовый корпус татарского языка [Suleymanov et al., 2013]. Были использованы поисковые запросы с включением различных грамматических признаков и определенных лексем. Далее был произведен биграммный анализ экспортированной выборки предложений в сравнении с полным корпусом “Туган тел” на уровне символов и словоформ, а также составлены частотные списки лемм и аффиксальных цепочек. Как показано в таблице 2, распределение в целом идентично, что позволяет оценить набор предложений как репрезентативный и сбалансированный.

**Таблица 2. Распределение наиболее частотных биграмм на уровне символов в краудсорсинговом подкорпусе TatSC и Татарском национальном корпусе “Туган тел”**

|    | Выборка краудсорсингового подкорпуса TatSC |       | Корпус “Туган тел” в целом |       |
|----|--|-------|----------------------------|-------|
|    | биграмма                                   | Доля  | Биграмма                   | доля  |
| 1  | ар   | 0.027 | ар                         | 0.028 |
| 2  | ла   | 0.021 | ан                         | 0.026 |
| 3  | лэ   | 0.020 | ла                         | 0.023 |
| 4  | ан   | 0.020 | га                         | 0.022 |
| 5  | ен   | 0.019 | лэ                         | 0.021 |
| 6  | эр   | 0.016 | эн                         | 0.019 |
| 7  | ын   | 0.016 | ел                         | 0.017 |
| 8  | ер   | 0.016 | ын                         | 0.016 |
| 9  | га   | 0.014 | ер                         | 0.016 |
| 10 | эн   | 0.027 | эр                         | 0.016 |

Характеристики записанных пользователей в данном датасете приведены на рис. 4–6. В целом, в записи с помощью бота Telegram приняло участие 667 носителей языка (437 женщин и 230 мужчин), Средни было некоторое число с диалектными особенностями и акцентом.

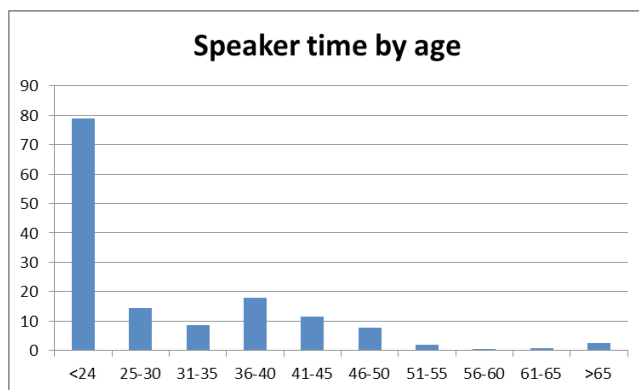


Рис. 4. Распределение суммарного времени записи в зависимости от возраста спикера



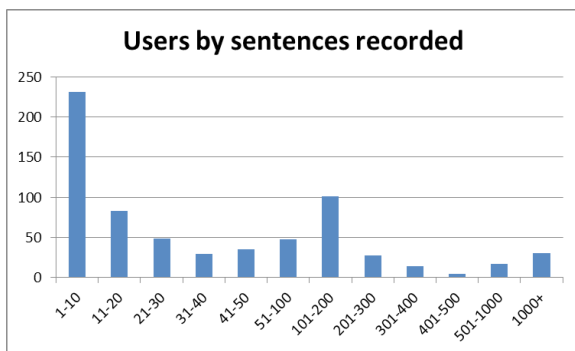


Рис. 5а. Распределение пользователей бота по числу записанных предложений

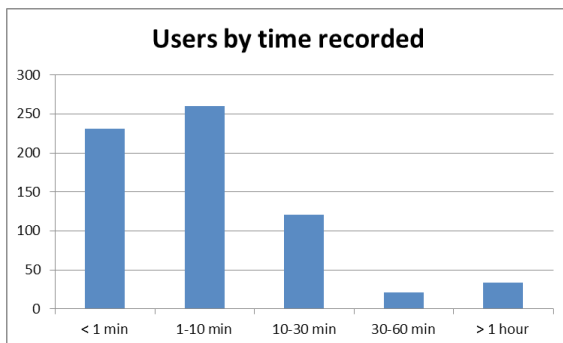


Рис. 5б. Число пользователей бота по суммарному времени записи

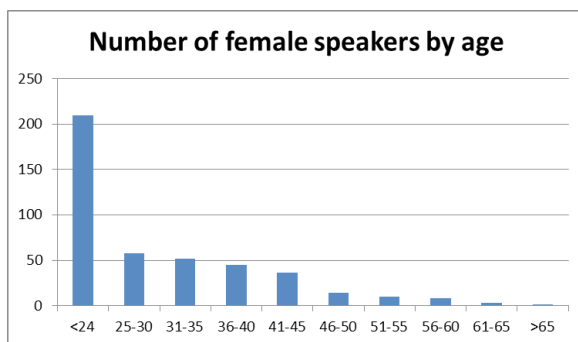


Рис. 6а. Распределение спикеров-женщин по возрасту

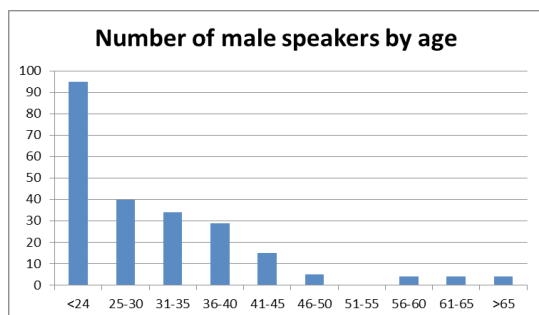


Рис. 6b. Распределение спикеров-мужчин по возрасту

Краудсорсинговый подкорпус организован в виде архива с папками, каждая из которых соответствует определенному предложению и содержит TXT и WAV файлы. Имена TXT файлов соответствуют именам папок, имена WAV файлов содержат имя папки и идентификатор пользователя в базе данных бота Telegram.

## 6. Разработка бота Telegram

Разработанный бот Telegram имеет функции записи и оценки записанных предложений. Для начала работы пользователю необходимо зарегистрироваться и сообщить о себе такую информацию, как пол и возраст, служащую для оценки сбалансированности и репрезентативности подкорпуса с социально-демографической точки зрения. Бот присылает пользователю случайное предложение для записи. После записи пользователь имеет возможность прослушать записанное, перезаписать или пропустить предложение (рис. 7a).

Функционал оценки становится доступным после записи 200 предложений. Пользователь получает от бота записи других пользователей в случайном порядке (рис. 7b). Для оценки используется 5-балльная шкала, рейтинг пользователя образуется их средней оценки его записей. Всего было получено более 21000 оценок, что дало покрытие 27% записей. В дальнейшем записи пользователей с высоким средним рейтингом (>4) могли подтверждаться автоматически, тогда как записи пользователей с низким средним рейтингом (<2,5), как и все записи с таким же низким рейтингом отклонялись. Записи со средним рейтингом (2,5-4) были частично проверены вручную. Общий средний рейтинг всех оцененных

записей составил 4,4. Это означает, что субъективная оценка качества записей с точки зрения носителей языка является высокой.



Рис. 7а. Функция записи в боте

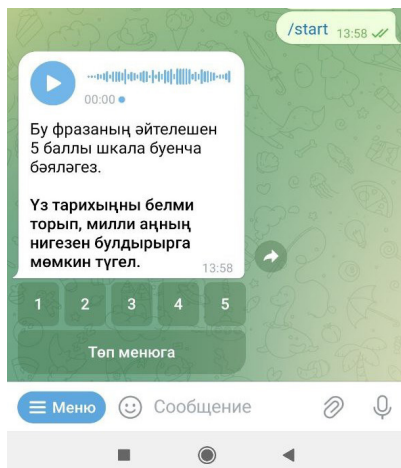


Рис. 7б. Функция оценки в боте

При разработке бота Telegram использовалась библиотека Django и были созданы следующие модели (таблицы): OriginalString, TelegramUser, UserAnswers, UserSkips, FirstBatch и AnswerRate.

OriginalString – это модель, которая хранит в себе данные об исходном тексте, используемом для озвучивания. Она позволяет боту отправить пользователю текст с наименьшим количеством собранных на данный момент голосов.

TelegramUser – модель, предназначенная для идентификации пользователей Telegram.

UserAnswers – модель, которая позволяет боту сохранять ответы пользователей на конкретные OriginalString сущности.

Userskips – модель, сохраняющая в себе пропуски определенных предложений. Возможность пропуска некоторых ответов позволяет определить наиболее тяжелые для озвучивания предложения.

FirstBatch – модель, позволяет определить набор предложений обязательных к озвучиванию предложений. Она обеспечивает возможность в срочном порядке получить минимально необхо-

димый набор озвучиваний на определенный набор предложений вне очереди.

AnswerRate – модель, позволяющая боту сохранять информацию об оценках на ответы. Это полезно для определения рейтинга конкретного ответа, что в дальнейшем используется для подсчета рейтинга пользователя.

Для случаев, когда пользователь намеренно занижает или завышает оценки реализован модуль, который подсчитывает среднюю оценку, которую выставляет пользователь. Если средняя оценка, при большом количестве оцениваний ниже 3, и при этом сверка каждой оценки с средней оценкой имеет большую разницу в 1,5 балла, то данная пользователем оценка не учитывается в рейтинге остальных.

На основе созданных моделей была разработана админ панель для анализа и обработки полученных данных (ответов, оценок, пользователей, предложений) (рис. 8).

| BACKEND                |            |            |
|------------------------|------------|------------|
| Запросы пользователей  | + Добавить | ✎ Изменить |
| Оригиналы строк        | + Добавить | ✎ Изменить |
| Ответы пользователей   | + Добавить | ✎ Изменить |
| Оценки озвучиваний     | + Добавить | ✎ Изменить |
| Первоочередный набор   | + Добавить | ✎ Изменить |
| Пользователи           | + Добавить | ✎ Изменить |
| Пропуски пользователей | + Добавить | ✎ Изменить |
| Розыгрыши              | + Добавить | ✎ Изменить |
| Участники конкурса     | + Добавить | ✎ Изменить |

Рис. 8. Админ-панель бота.

## 7. Результаты первичных экспериментов

В рамках проекта многоязычного распознавания речи Söyle [Söyle] в сотрудничестве с Институтом умных систем и искусственного интеллекта в Астане (ISSAI) был проведен ряд экспериментов с использованием датасета TatSC. Подробно данные эксперименты описаны в [Mussakhojayeva, Gilmullin et al.].

Результаты экспериментов показывают высокое качество распознавания по метрикам CER и WER. В частности, обученная на корпусе TatSC многоязычная модель Söyle демонстрирует на тестовом наборе Common Voice (CVC) 9,1% WER, в то время как для модели распознавания речи для турецких языков, в которой использовался татарский корпус Common Voice объемом около 26 часов, лучшим результатом было 16,5% WER на том же тестовом наборе [Mussakhojajeva et al., 2023].

Высокие результаты WER также были получены для разных уровней шума (SNR, Signal-to-Noise Ratio), в том числе не использовавшихся в процессе обучения модели. Результаты из [Mussakhojajeva, Gilmullin et al.] для корпуса TatSC представлены в таблице 3. Уровни SNR 5 и 1, не использовавшиеся в дообучении модели, выделены серым фоном.

**Таблица 3. Показатели WER для моделей Söyle и дообученной на аудио с помехами Söyle-NR на разных уровнях шума SNR**

| SNR   | Söyle | Söyle-NR |
|-------|-------|----------|
| Clean | 7,06  | 6,49     |
| 100   | 7,06  | 6,49     |
| 50    | 7,07  | 6,49     |
| 25    | 7,51  | 6,89     |
| 10    | 11,5  | 10,6     |
| 5     | 17,6  | 16,4     |
| 1     | 30,2  | 27,1     |

В целом, можно констатировать, что результаты первичных экспериментов с обучением моделей распознавания речи с использованием датасета TatSC демонстрируют эффективность разработанного речевого корпуса.

## 8. Заключение

Общий объем речевого корпуса TatSC превышает 269 часов. Всего он содержит 217914 записей и 68623 уникальных слов.

TatSC представляет собой первый большой речевой корпус татарского языка, ориентированный на задачи автоматического распознавания речи. Планируется продолжить работу по сбору и

проверке данных с целью увеличения объема и репрезентативности корпуса.

#### ЛИТЕРАТУРА

1. Ardila, R.; Branson, M.; Davis, K.; Kohler, M.; Meyer, J.; Henretty, M.; Morais, R.; Saunders, L.; Tyers, F.; Weber, G. Common Voice: A Massively-Multilingual Speech Corpus. In Proceedings of the Language Resources and Evaluation Conference (LREC), Marseille, France, 11–16 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 4218–4222.
2. Mussakhojayeva, S.; Dauletbek, K.; Yeshpanov, R.; Varol, H.A. Multilingual Speech Recognition for Turkic Languages. *Information* 2023, 501 14, 74.
3. Suleymanov, D.; Khusainov, A.; Mukhametzhanov, I. Self-Supervised Training for the Tatar Speech Recognition System. *Russian Advances in Fuzzy Systems and Soft Computing: Selected Contributions to the 8th International Conference on “Integrated Models and Soft Computing in Artificial Intelligence (IMSC-2021) // CEUR Workshop Proceedings, 2021, Vol. 2965*, pp. 220–225.
4. Татарстан китап нәшрияты (Татарское книжное издательство), <https://tatkniga.ru/> (дата доступа 15.07.2023)
5. Бәширова, И.Б.; Нуриева, Ф.Ш.; Кадыйрова, Э.Х. Татар әдәби теле тарихы (XIII гасыр – XX йөз башы), 1 том, Казан: “ТӘҺСИ”, 2015; 696 б.
6. Label Studio, <https://labelstud.io/> (дата доступа 15.07.2023)
7. Татарский национальный корпус “Туган тел”, <https://tugantel.tatar> (дата доступа 15.07.2023)
8. Suleymanov, D.; Nevzorova, O.; Gatiatullin, A.; Gilmullin, R.; Khakimov, B. National Corpus of the Tatar Language “Tugan Tel”: Grammatical Annotation and Implementation. In *Procedia-Social and Behavioral Sciences*, 2013, Vol.95, pp. 68–74. <https://doi.org/10.1016/j.sbspro.2013.10.623>.
9. Telegram, <https://telegram.org>
10. Söyle, <https://github.com/IS2AI/Soyle> (дата доступа 29.09.2023)
11. Mussakhojayeva, S., Gilmullin, R., Orel, D., Khakimov, B., Abilbekov, A., Galimov, M., Varol, H.A. Söyle: Noise Robust Multilingual Speech Recognition with Long Transcription Featuring the Tatar Speech Corpus (в печати)

УДК 811.512.141'42

**ОСОБЕННОСТИ ДИАЛЕКТНОГО СИНТАКСИСА  
БАШКИРСКОГО ЯЗЫКА  
(НА МАТЕРИАЛЕ ТЕКСТОЛОГИЧЕСКОЙ БАЗЫ  
ДИАЛЕКТОЛОГИЧЕСКОГО ПОДФОНДА МАШИННОГО  
ФОНДА БАШКИРСКОГО ЯЗЫКА)**

*Л. А. Бускунбаева*

*Ордена Знак Почета Институт истории,  
языка и литературы Уфимского федерального  
исследовательского центра РАН, Уфа, Россия  
buskl@yandex.ru*

В данной статье рассматриваются синтаксические особенности башкирских диалектов на материале Текстологической базы Диалектологического подфонда Машинного фонда башкирского языка. Синтаксическая система башкирских говоров в меньшей степени отличается друг от друга, чем в фонетическом, лексическом или морфологическом плане. В то же время можно наблюдать синтаксические явления, которые не свойственны современному литературному языку и составляют специфику того или иного говора или диалекта. Обнаруженные синтаксические диалектизмы показывают уникальность каждого говора и диалекта башкирского языка, сохранившего реликтные черты в своем составе.

**Ключевые слова:** башкирский язык, текстологическая база, Машинный фонд башкирского языка, синтаксис, диалект

**FEATURES OF THE DIALECT SYNTAX OF THE BASHKIR  
LANGUAGE  
(BASED ON THE MATERIAL OF THE TEXTUAL BASE OF  
THE DIALECTOLOGICAL SUB-FUND OF THE BASHKIR  
LANGUAGE MACHINE FUND)**

*Buskunbaeva Lilia Aisovna*

*The Institute of History, Language and Literature of the Ufa Federal  
Research Center of the Russian Academy of Sciences  
Ufa, Russia  
buskl@yandex.ru*

This article discusses the syntactic features of Bashkir dialects based on the material of the Textual base of the Diacological Subfund of the Bashkir Language Machine Fund. The syntactic system of Bashkir dialects differs less from each other than in phonetic, lexical or morphological terms. At the same time, it is possible to observe syntactic phenomena that are not peculiar to the modern

literary language and constitute the specifics of a particular dialect or dialect. The discovered syntactic dialectisms show the uniqueness of each dialect and dialect of the Bashkir language, which has preserved relic features in its composition.

**Keywords:** Bashkir language, textual base, Machine fund of the Bashkir language, syntax, dialect

Развитие компьютерной и корпусной лингвистики позволило создать репрезентативные и сбалансированные базы данных, электронные корпуса и по звучащей речи, что намного расширило возможности лингвистов исследовать те аспекты устной речи, которые были ранее недоступны из-за скудного объема нужного фактического материала.

Разработка Текстологической базы, интегрированная в Дialeктологический подфонд Машинного фонда башкирского языка, открывает доступ к богатому диалектному материалу, собранному во время многочисленных экспедиций в разные годы сотрудниками ИИЯЛ УФИЦ РАН.

В данной базе представлены транскрибированные тексты, отражающие живую речь носителей говоров башкирского языка и снабженные экстралингвистической разметкой. Экстралингвистическая разметка включает в себя следующие экстралингвистические параметры: диалект, говор, год записи, образование, пол, возраст, национальность, ФИО, место жительства информанта, тематика текста [Сиразитдинов, 2016, 216].

Все параметры, за исключением года записи, выбираются из ниспадающего меню. Год записи пользователь вводит сам. Образование включает следующие поля фиксации: без образования, начальное, неполное среднее, среднее, среднее техническое, высшее.

Возраст информанта задается в виде шкалы лет: 5–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80, 81–90, 91–100.

Программное обеспечение позволяет производить поиск текстов по любому из 7 параметров или по комбинации этих параметров. Текстологическая база не является корпусной базой, но тем не менее позволяет производить поиск словоформ или фрагментов текста по полю “текст”.

Башкирские диалекты изучены достаточно полно и разнообразно – изданы монографии с подробным описанием фонетических, лексических и морфологических диалектных особенностей [Баишев, 1955; Ишбулатов, 1974; Максютлова, 1976;



Максютова, 1996; Миржанова, 1991; Миржанова, 1979; Шакуров, 2012], составлены словари [Башкорт һөйләштәренен һүзлеге, 1967; Башкорт һөйләштәренен һүзлеге, 1970; Башкорт һөйләштәренен һүзлеге, 1987] и диалектологический атлас [**Диалектологический атлас башкирского языка**, 2005], охватившие лексику всех диалектных зон башкирского языка.

Однако несмотря на многочисленные исследования по диалектологии башкирского языка, остается немало нерешенных проблем, требующих новых подходов и инструментов для их успешного решения. Одним из наименее изученных и актуальных проблем современной диалектологии остается диалектный синтаксис. Отсутствие работ по данной проблематике было связано, по нашему мнению, прежде всего с ограниченным количеством диалектных текстов.

Текстологическая база позволяет восполнить данный пробел и на богатом диалектном материале начать анализ синтаксических особенностей башкирских говоров.

Башкирские диалекты в синтаксическом плане в меньшей степени отличаются друг от друга, чем в фонетическом, лексическом или морфологическом. В то же время можно наблюдать синтаксические явления, которые не свойственны современному литературному языку и составляют специфику того или иного говора или диалекта.

Некоторые различия в синтаксисе диалектной речи наблюдаются на уровне словосочетаний. Одной из отличительных черт является употребление в именных словосочетаниях (имя существительное + имя существительное) безаффиксального вида, т.е. вместо изафета второго типа, принятого в литературном языке, изафета первого типа. Данное явление можно объяснить тем, что первый изафет исторически является более древней формой, нежели производные, возникшие в более позднюю эпоху. Те диалекты башкирского языка, в которых встречается данный способ словосочетания, сохранили архаические особенности древнебашкирского языка. Например, в большинстве говоров северо-западного и восточного диалектов изафет второго типа применяется без аффикса принадлежности [Миржанова, 2006, с. 92]: *Йылга арйакта йеләк бешкән* лит. *Йылга аръягында еләк бешкән* 'За рекой поспели ягоды'; *Әмир бичә килә* лит. *Әмир бисәһе килә* 'Идет жена Амира'.

В диалектах можно наблюдать и обратное явление, когда вместо литературного первого изафета в диалектах (в таныпском говоре) употребляется второй тип изафета: *ир кешесе* лит. *ир кеше* ‘мужчина’; *шайтан ашыгы* лит. *шайтан ашык* ‘щиколотка’.

Особенности в падежном управлении ярко проявляются в диалектах башкирского языка. Как правило, в литературном языке при данном способе связи словосочетаний грамматически подчиненные слова ставятся в той или иной форме косвенных падежей в зависимости от синтаксических потенциалов управляющего слова [Грамматика современного башкирского литературного языка, 1981, с. 367]. В башкирских диалектах часто можно наблюдать безаффиксальную форму подчиненного слова.

Например, в демском говоре южного диалекта, восточном и северо-западном диалектах при сочетании с глаголами движения вместо дательного падежа употребляется основной падеж: *Уфа бараңмы?* лит. *Өфөгә бараһыңмы?* ‘В Уфу едешь?’; *Базар киттеңме?* лит. *Базарға киттеңме?* ‘На базар едешь?’; *Сугыш ваҡытында Ташкин сығып китеп, Дәүләкәндең бере кешегә барҙым* лит. *Һугыш ваҡытында Ташкентка сығып китеп, Дәүләкәндең бер кешегенә барҙым* ‘Когда началась война, уехала в Ташкент, вышла замуж за человека родом из Давлеканово’.

В аргаяшском говоре восточного диалекта падежная система отличается неоформленностью, т.е. в большинстве случаев косвенные падежи выступают без грамматических показателей, их функции выполняют основной падеж [Максютова, 1976, с. 108]. Например, употребление безаффиксального винительного падежа: *Нисә йыл кала күргәнем йук* лит. *Нисә йыл каланы күргәнем юк* ‘Сколько лет не бывал в городе’; безаффиксального родительного падежа со значением дополнения: *Мин эштән калган йук* лит. *Минең эштән калғаным юк* ‘Я никогда от работы не отказываюсь’.

Среди особенностей словосочетания в диалектах можно отметить и функциональный параллелизм падежей, когда один косвенный падеж может быть замещен другим. Например, в гайнинском говоре северо-западного диалекта дательный падеж употребляется вместо основного: *аң атала* лит. *ул атала* ‘это называется’. В сальютском говоре восточного диалекта наблюдается употребление дательного падежа вместо винительного: *Буранбайға матур йырҙай* лит. *Буранбайҙы матур йырлай* ‘Хорошо поет песню Буранбай’.

Синтаксические диалектизмы наиболее ярко проявляются на уровне предложений.

**Вопросительные предложения** в башкирских диалектах, как и в литературном языке, образуются с помощью вопросительных местоимений (*кем?* ‘кто’ *нимә?* ‘что’, *кайза?* ‘где’, *нисек?* ‘как’ и т.д.), вопросительных частиц *-мы/-ме* или вопросительной интонации.

В то же время в плане выражения вопросительных предложений в диалектах наблюдаются некоторые различия.

В аргаяшском, айском и сальютском говорах восточного диалекта вопросительная частица *-мы/-ме* располагается перед показателями лица и числа: *Килдемегес?* лит. *Килдегезме?* ‘Пришли?’, *Калага китәмең?* лит. *Калага китәһеңме?* ‘В город собираешься?’ Н. Х. Максютова объясняет такое расположение аффикса-частицы архаической формой, которая была характерна и для древнетюрского языка [Максютова, 1996, с. 149].

Одним из основных средств выражения вопросительности в литературном языке выступают вопросительные частицы *-мы/-ме*, которые присоединяются к сказуемому. Однако в диалектах наблюдаются случаи, когда вопросительная частица может быть присоединена к любому члену предложения. Такое явление наблюдается в гайнинском говоре северо-западного диалекта башкирского языка: *Ул мындамы керә?* ‘Он сюда идет?’; *Әйбәтме йәшәйбес?* ‘Хорошо ли живем?’; *Кыз кәртүкме бешерә?* ‘Девушка варит картофель?’. Чаще всего вопросительная частица присоединяется к тому слову с целью подчеркивания того, о чем спрашивается.

В айском говоре восточного диалекта наблюдается употребление удвоенной вопросительной частицы, присоединенной посредством форманта *-да*: *караныңмыдама?* ‘уже посмотрел?’. Образованные подобным образом слова, помимо вопроса содержат и семантику удивления [Максютова, 1976, с. 62].

**Побудительные предложения** в башкирских диалектах в основном строятся по тем же моделям, что и в литературном языке. Однако можно обнаружить некоторые отличительные черты, находящие отражение во многих говорах башкирского языка. Например, в большинстве говоров восточного и северо-западного диалектов функционирует древнейшая форма повелительного наклонения на *-ың/-ең* (2 л., мн. ч.) вместо литературного аффикса *-гыз/-гез*: *Тизерәк кайтың* лит. *Тизерәк кайтыгыз* ‘Скорее возвращайтесь!’, *Сайлабылың!* лит. *Һайлап алыгыз!* ‘Выбирайте!’

Обнаруженные синтаксические диалектизмы показывают уникальность каждого говора и диалекта башкирского языка, сохранившего реликтные черты в своем составе.

#### ЛИТЕРАТУРА:

1. Баишев Т.Г. Башкирские диалекты в их отношении к литературному языку. М.: МГУ, 1955.
2. Башкорт һөйләштәренәң һүзлеге. Т.1. Көнсығыш диалект / Н.Х.Мәксүтова һ.б. Өфө, 1967.
3. Башкорт һөйләштәренәң һүзлеге. Т.2. Көнъяк диалект / Н.Х.Мәксүтова һ.б. Өфө, 1970.
4. Башкорт һөйләштәренәң һүзлеге. Т.3. Көнбайыш диалекты / Н.Х.Мәксүтова һ.б. Өфө, 1987.
5. Диалектологический атлас башкирского языка / Составители: Н. Х. Максүтова, С. Ф. Миржанова, У. Ф. Надергулов, М. Н. Дильмухаметов, С. Г. Сабирьянова, Г. Г. Гареева. Уфа: Гилем, 2005.
6. Грамматика современного башкирского литературного языка. М., 1981.
7. Ишбулатов Н. Х. Диалектная система башкирского языка в сравнительно-историческом освещении: Автореф. дисс. ... д-ра филол. наук. Уфа, 1974.
8. Максүтова Н. Х. Восточный диалект башкирского языка в сравнительно-историческом освещении. М.: Наука, 1976.
9. Максүтова Н.Х. Башкирские говоры, находящиеся в иноязычном окружении. Уфа: Китап, 1996.
10. Миржанова С.Ф. Северо-западный диалект башкирского языка. Уфа, 1991. 295 с.
11. Миржанова С. Ф. Южный диалект башкирского языка. М.: Наука, 1979.
12. Сиразитдинов З. А., Бускунбаева Л. А., Каримова Р. Н. Диалектологическая база Машинного фонда башкирского языка как инструмент исследования башкирского диалектного континуума // Вестник КИГИ РАН. 2016; № 1 (23). С. 212–219.
13. Шакуров Р. З. Диалектная система башкирского языка // Ватандаш. 2012. № 8. С. 40–61.

УДК 81'322.2

**АНАЛИТИЧЕСКИЙ ОБЗОР ПО СМЫСЛОВОМУ  
РАСПРЕДЕЛЕНИЮ СЛОВ В УЗБЕКСКОМ КОРПУСЕ***Исроилов Жасур<sup>1</sup>, Абдурахмонова Нилуфар<sup>2</sup>**<sup>1</sup>Наманганский государственный университет  
Наманган, Узбекистан**<sup>2</sup>Национальный университет Узбекистана  
Ташкент, Узбекистан, 100174*

jasurbek9109@gmail.com, n.abduraxmonova@nuu.uz

Омонимия является одним из наиболее проблемных вопросов компьютерной лингвистики. Проблемы, возникающие в процессе работы с омонимами в области перевода, морфо анализа и лексикологии, мотивируют развитие связанных с ними исследований. Двусмысленность и омонимия являются проблемными случаями в лингвистике. Если слова семантически связаны, то это полисемия, т. е. явление многозначности, а если значения неродственны и представляют собой совершенно разные понятия, то такие лексемы являются омонимами. Различие между омонимией и полисемией в семантическом анализе было проблематичным аспектом для лингвистов. Омонимию иногда путают с явлением полисемии. Полисемия рассматривается как перенос значения, который происходит, когда одно слово ассоциируется с разными словами, тогда как омонимия представляет собой набор слов, не имеющих отношения ни в письменной, ни в устной форме.

В данной статье опубликованы результаты исследования по применению устранения смысловой неоднозначности слов в узбекском корпусе, что является сегодня одной из наиболее актуальных проблем обработки естественного языка.

**Ключевые слова:** NLP, WSD, disambiguation, Knowledge-based, approach, LESK, WordNet, corpus linguistics, Uzbek corpus.

**ANALYTICAL REVIEW ON WORD SENSE DISAMBIGUATION  
FOR UZBEK CORPUS***Jasur Isroilov<sup>1</sup>, Nilufar Abdurakhmonova<sup>2</sup>**<sup>1</sup>Namangan State University, Namangan, Uzbekistan,**<sup>2</sup>National University of Uzbekistan  
Tashkent, Uzbekistan*

jasurbek9109@gmail.com, n.abduraxmonova@nuu.uz

Homonymy is one of the most problematic issues in computational linguistics. The problems that arise in the process of working with homonyms in the field of translation, morphological analysis and lexicology motivate the development of related research. Ambiguity and homonymy are problematic cases in linguistics. If

the words are semantically related, then this is polysemy, that is, the phenomenon of polysemy, and if the meanings are unrelated and represent completely different concepts, then such lexemes are homonyms. The distinction between homonymy and polysemy in semantic analysis has been a problematic aspect for linguists. Homonymy is sometimes confused with the phenomenon of polysemy. Polysemy is seen as a transfer of meaning that occurs when one word is associated with different words, while homonymy is a set of words that have no relationship in either written or spoken form.

This article publishes the results of a study on the use of word sense disambiguation of words in the Uzbek corpus, which is one of the most pressing problems in natural language processing today.

**Keywords** – NLP, WSD, disambiguation, Knowledge-based, approach, LESK, WordNet, corpus linguistics, Uzbek corpus.

## I. INTRODUCTION

In today's world, we are surrounded by information coming from many different sources. These sources include verbal and non-verbal communications, as well as various textual forms. Take short messages, emails, tweets or newspapers as some of the many text-based communication examples. What is more, we are both the recipients, and the producers of these pieces of information, a big part of which is generated through social media. However, even in the shortest messages we produce, we are prone to making spelling, punctuation or grammar mistakes. [Loïc Vial, 2017]

We know that words make up the vocabulary of the language, interact with each other and perform various tasks. In conversation, words are semantically close to each other or, on the contrary, have opposite meanings. There are such words, although they are not close in meaning, but similar in appearance, form, spelling and pronunciation. Such words are called synonymous words. The goal of word sense determination (WSD) is to correctly determine the meaning of a word in a general sentence. All natural languages exhibit ambiguity in the meaning of words, and resolving this automatically remains a pressing problem. Hence, WSD is considered an important natural language processing (NLP) problem. The motivation behind our current research stems from the need for new WSD methods and tools. Standard assessment resources are required to develop, evaluate, and compare WSD methods. A number of initiatives have led to the development of WSD reference corpora for a wide range of languages from different language families. [Saeed, A. 2019] It underlies the understanding of language and has already been studied from different points of view. In

general, there is no clear path in this area, in part because identifying real improvements over current approaches becomes a challenge with current evaluation benchmarks. This is mainly due to the lack of a unified structure, which prevents a direct and fair comparison of systems.

Under the leadership of Professor Nilufar Abdurakhmonova, work is being carried out on the corpus of the Uzbek language. To date, the corpus contains more than 10 million words. Morpho analyzer, thesaurus, linguistic analyzer has been implemented in this corpus. This article analyzes WSD applications, approaches, problems of the Uzbek language in WSD and recommended algorithms for solving them. [Abdurakhmonova, 2021]

## II. RELATED WORKS

One of the most complex problems in NLP is WSD, and research in this area dates back to the 1940s. Zipf (1949), who brought the concept of the “Law of Understanding” to this field in the early years, Masterman (1957), who proposed the theory of finding the true meaning of a word using the catalogs listed in the Roget International Thesaurus, in 1975, Wilks gave the exact meaning of an ambiguous word. Wilks (1975), who developed the “Preference semantics” model, and Rieger and Small (1979), who developed the idea of “vocabulary experts” conducted research to find the name. By the 1980s, a number of linguistic corpora and linguistic databases were formed, and scientific research in the fields of WSD also developed. As a result, researchers have begun to use various automated knowledge extraction procedures (Wilkes et al., 1990) in parallel with manual methodologies. Lesk (1986) developed an algorithm for similar glosses of words. Later, this algorithm was used in Knowledge-based WSD as LESK algorithm. [Lesk, M., 1986] As a result of the research carried out, by the 1990s, great results were obtained in the field of NLP. [Ranjan Pal, 2015] In particular, Wordnet was created by Miller (1990) [Miller G.A., 1990], statistical methodologies were created in the field of WSD, Today, hand-labeled corpus learning methods (i.e., supervised learning methods) have become the mainstream approach to WSD. Corpus-based WSD (WSD) is a type of WSD that uses a corpus of text to disambiguate the senses of words. The corpus is used to learn statistical relationships between words and their meanings. These relationships can then be used to disambiguate the senses of words in the new text. The first Corpus-based WSD idea for NLP was developed by Brown (1991). Senseval is an international series of evaluations of

computational semantic analysis systems. It was started in 1997 proposed by Reznik and Jarowski, following a workshop, Tagging with Lexical Semantics: Why?, What?, and How?, held at the conference on Applied Natural Language Processing. The first three evaluations, Senseval-1 through Senseval-3, were focused on WSD (WSD), each time growing in the number of languages offered in the tasks and in the number of participating teams. [Cucerzan, 2002]

WSD in the Turkish language family is a challenging task due to the high degree of polysemy in Turkish words. Turkish is an agglutinative language, which means that words are formed by adding suffixes to a root word. This can lead to words having multiple meanings, depending on the suffixes that are attached. WSD was carried out by also Kazakh, Kyrgyz, Tatar, Turkish, Turkmen, Bashkir, Khakas, and Uzbek scientists. R. Bekdjanova (Kyrgyz language), T. Kerim (Kyrgyz language), Z. Sattarova (Crimean-Tatar language), V. Abayeva, O.S. Akhmanova A.I. Smirnisky, V.V. Vinogradov, K. Akhanov (Kazakh language), M.Kh. Researches of Turkic scientists such as Akhtyamov (Bashkir language), A.A. Gasanov (Azerbaijani language), A. Geldimuradov (Turkman language) serve as the basis for studying WSD in computer linguistics. In every language, WSDs are a historical phenomenon of lexicology, and they are not an indicator of the poverty of the language, but an indicator of the richness and diversity of the lexical stock.

In corpus linguistics, there have been many studies on solving the problem of WSD, tagging homonyms, and eliminating WSD during automatic text reading. Among them, G. I. Kustova, O. N. Lyashevskaya, Ye. V. Paducheva, Ye. V. Rakhilina, B. P. Kobrisov, T. I. Reznikova, V. V. Kukanova, A. A. Kretov conducted a number of research works aimed at eliminating WSD in computer linguistics. WSD elimination is important in many computational linguistics applications, particularly in search engines. Because it can increase the accuracy of processing certain classes of requests or reduce the amount of information stored.

In a number of studies carried out in the field of computer linguistics of the Uzbek language, there are efforts to create analysis programs designed for computer memory to recognize and read homonymous units. Some comments on the problems of tagging homonyms in the Uzbek language, preliminary efforts to create a WSD detection algorithm were made. A clear example of this is the research of a number of researchers such as M. A'lamova, N. Abdurakhmonova, [Dependency Parsing Based On Uzbek Corpus. Language technology for all, 2019] Sh. Gulyamova, H. Akhmedova, G. Abduvakhobov.



### III. APPROACHES AND APPLICATIONS OF WSD

#### A) Applications of WSD:

WSD (WSD) is the process of determining the correct meaning of a word in a given context. It is a challenging task, as many words have multiple meanings. However, WSD is essential for many natural language processing (NLP) applications, such as:

**Machine translation:** WSD is used in machine translation to ensure that the correct meaning of a word is translated into the target language. For example, in the Uzbek sentences “Ahmad doira chertib qo‘shiq hirgoya qilardi”, “Samad daftariga doira rasmini chizdi”, the word “doira” has different meanings, which is a big problem in translation.

**Information retrieval:** WSD is used in information retrieval to improve the accuracy of search results. For example, a user might search for the word “bank”. Without WSD, the search results would include documents that contain both the meaning of “financial institution” and the meaning of “riverbank”. With WSD, the search results would be more likely to include documents that are relevant to the user’s query.

**Question answering:** WSD is used in question answering systems to understand the meaning of the user’s question. For example, a user might ask the question “Bozordan nok, olma, uzum oldingmi?”. Without WSD, the question answering system would need to consider both the meanings of “olma” (as in “don’t take”) and “olma” (as in “fruit”). With WSD, the question answering system will be able to correctly determine what the user is asking about buying fruit.

**Text analysis:** WSD is used in text analysis to extract the meaning of text. For example, a text analysis system might be used to analyze a news article to determine the article’s main topic. With WSD, the text analysis system would be able to correctly identify the meaning of the words in the article, which would help it to better understand the article’s overall meaning.

#### B) WSD Approaches

There are three main approaches to WSD (WSD):

**Knowledge-based WSD** approaches use a knowledge base of word senses to disambiguate words. A knowledge base of word senses is a database that contains information about the different senses of a word, such as its definition, its examples, and its relations to other words.

We reviewed and compared some knowledge-based WSD approaches:

**Lesk algorithm** is a simple but effective algorithm that disambiguates words by finding the sense of a word that has the highest similarity score with the context in which the word is used. The similarity score is calculated by comparing the context to the definitions of the different senses of the word. [Lesk, 1986]

**WordNet::Similarity** library is a Python library that provides a number of methods for calculating the similarity between words. These methods can be used to disambiguate words by finding the sense of a word that has the highest similarity score with the context in which the word is used. [A. Agostini, 2021]

**WordNet::SenseRelate** library is a Python library that provides a number of methods for finding the relationships between words and their senses. These methods can be used to disambiguate words by finding the sense of a word that is most closely related to the context in which the word is used.

**Linguistic Knowledge Sources** such as WordNet and FrameNet provide information about the morphology, syntax, and semantics of words. This information can be used to disambiguate words by finding the sense of a word that is most consistent with the linguistic features of the word. [Loïc Vial, 2017]

Knowledge-based WSD approaches are typically more accurate than statistical WSD approaches, but they can be slower and require more manual effort to build and maintain.

As a result of our research, we have found that knowledge-based WSD approaches have the following advantages:

- They are more accurate than statistical WSD approaches, especially for words with multiple senses that are closely related.
- They can be used to disambiguate words in new contexts, even if the word has not been seen before in a corpus of text.
- They can be used to disambiguate words that are ambiguous due to their morphology or syntax.

Here are some of the disadvantages of knowledge-based WSD approaches:

- They can be slower than statistical WSD approaches.
- They require more manual effort to build and maintain.
- They may not be able to disambiguate words that are ambiguous due to their semantics.

**Statistical approaches** use statistical methods to learn the most likely sense of a word from a corpus of text. A corpus of text is a collection of text that has been annotated with word senses. [Yarowsky, 1994]

We reviewed and compared some statistical WSD approaches:

**Naive Bayes** is a simple but effective algorithm that uses a statistical model to calculate the probability of a word having a particular sense in a particular context. The probability is calculated based on the frequency of the word in different senses in the corpus of text. [Voorhees, 1993]

**Support vector machine** is a more complex algorithm that can be used to learn a more sophisticated model of the relationship between words and their senses. The SVM model is trained on a corpus of text that has been annotated with word senses. The model is then used to disambiguate words in new contexts.

**Ensemble methods** combine the results of multiple WSD systems to improve WSD performance. Ensemble methods can be used to combine the results of knowledge-based, statistical, and hybrid WSD systems.

Statistical WSD approaches are typically faster than knowledge-based WSD approaches, but they can be less accurate.

As a result of our studies, we have identified some advantages of Statistical WSD approaches:

- They are faster than knowledge-based WSD approaches.
- They are less dependent on manually-curated knowledge bases.
- They can be used to disambiguate words in new contexts, even if the word has not been seen before in a corpus of text.

Disadvantages of statistical WSD approaches:

- They can be less accurate than knowledge-based WSD approaches, especially for words with multiple senses that are closely related.
- They may not be able to disambiguate words that are ambiguous due to their morphology or syntax.
- They require a large corpus of text that has been annotated with word senses.

**Hybrid approaches** combine knowledge-based and statistical WSD approaches to achieve better performance than either approach on its own.

We reviewed and compared some statistical WSD approaches:

**Bootstrapping** is a machine learning technique that can be used to improve the performance of WSD systems. Bootstrapping starts with a small set of manually disambiguated words. The WSD system is then used to disambiguate the remaining words in the corpus. The results of the WSD system are then used to update the manually disambiguated words. This process is repeated until the WSD system converges.

**Co-training** is a semi-supervised learning technique that can be used to improve the performance of WSD systems. Co-training uses two WSD systems that are trained on different parts of the corpus. The two WSD systems then share information with each other to improve their performance.

**Ensemble methods** combine the results of multiple WSD systems to improve WSD performance. Ensemble methods can be used to combine the results of knowledge-based, statistical, and hybrid WSD systems.

Hybrid WSD approaches are typically more accurate than knowledge-based or statistical WSD approaches on their own. However, they can also be more complex and require more training data.

We compared the hybrid approach with other approaches and found the following advantages of the hybrid WSD approach:

- They can achieve better performance than knowledge-based or statistical WSD approaches on their own.
- They can be more robust to noise and ambiguity in the data.
- They can be more flexible and can be adapted to different tasks and domains.

Disadvantages of hybrid WSD approaches:

- They can be more complex and require more training data.
- They can be more difficult to develop and tune.
- They may not be as accurate as knowledge-based WSD approaches for words with closely related senses.

#### **IV. APPLICATION OF WSD IN THE UZBEK LANGUAGE CORPUS**

##### **A) Linguistic models of homonyms within part of speech**

Such phenomena as homonymy, synonymy, antonymy and polysemy in world linguistics have been widely studied from the point of view of computational linguistics, which is considered new for Uzbek linguistics, and its practical results are used for such purposes as the creation and improvement of corpora. The semantic analyzer is the basis of analysis and understands all the rules associated with the programming language. The semantic analyzer serves to distinguish between the meanings of homonymous, polyfunctional, polysemantic words in a language.

Modeling homonyms observed within different parts of speech is to determine which categories they belong to and their morphological,

syntactic and semantic properties. If the noun is among the groups that make up the noun, they are often indistinguishable by grammatical forms. Sometimes a category that is not part of a noun can receive its lexical and syntactic forms, just like nouns. In this case, the syntactic factor comes to the fore, is differentiated based on the principle of cohesion and is modeled accordingly.

Perfect semantic analysis of texts requires modeling of these elements. First, the elements are linguistically modeled. Mathematical models are developed based on linguistic models. In corpus linguistics, a number of studies have been carried out on solving the problem of WSD, tagging homonymous units, and eliminating WSD in the process of automatic text reading. Including G.I. Kustova, O.N. Lyashevskaya, Ye.V. Paducheva, Ye.V. Rakhilina, T.I. Reznikova, B.P. Kobrisov, V.V. Kukanova, A.A.Kretov, A.Ye. Alexander, Y.E. The Yermolayeva developed linguistic and mathematical models for eliminating side words.

For the corpus of the Uzbek language, the presence of homonyms in one or more categories creates problems in dividing words into tokens and determining their categories. In order to work with homonyms in the corpus, it is desirable to first determine in which word groups they form WSD and develop their models.

In creating the national corpus of the Uzbek language, the issue of forming a base of grammatical homonyms and their linguistic modeling is considered important. In Uzbek linguistics, words form homonyms within one or more categories. Words can form homonyms within one, two and three categories. In order to develop models of homonymous words in computer linguistics, it is necessary to first determine which categories of words form homonymy and give them conditional symbols. In modeling, it is customary to use globally accepted abbreviations for words.

Below we consider the conventional abbreviations accepted for each word group:

Ot – Noun → N

Fe'l – Verb → V

Sifat – Adjective → Adj.

Ravish – Adverb → Adv.

Olmosh – Pronoun → Pron

Son – Numeral → Num.

Taqlid – Mimic → Mim.

Undov – Exclamation → Exl.

Yuklama – Particle → Part.  
 Bog‘lovchi – Conjunction → Conj.  
 Ko‘makchi – Preposition → Pr.  
 Modal fe‘l – Modal verb → MV

**a) WSD between noun and verb**

WSD between noun and verb is often observed in linguistics. It is formed as a result of grammatical addition to the base of the word.

$$V + ma = HW \rightarrow [N \leftrightarrow V] \quad (1)$$

Let's look at the example of the word *Qovurma*. As a result of adding -ma to the stem of the word *Qovur* (verb), WSD appears in two-word groups. (1)

*Qovurma* – Noun+ma (former): a dish prepared with meat and vegetables;

*Qovurma* – Verb + ma (infinitive form): a verb form that expresses the absence of an action.

$$V + moq = N \leftrightarrow V \quad (2)$$

In the example of the word *Ilmoq*. A noun and a verb WSD are formed as a result of the addition of the word base *il* (Verb) verb + moq suffix. (2)

*Ilmoq* – Noun + moq(maker): a tool used for hanging.

*Ilmoq* – Verb + moq (action noun form): to hang on hooks, to hang

**b) WSD between adjective and verb.**

In this case, grammatical homonyms are defined according to the grammatical forms they take or the units they connect before or after them.

$$\text{Adj} + \text{ish} = HW \rightarrow [\text{Adj.} \leftrightarrow V] \quad (3)$$

Let's see the example of the word *Oqish*. The word oq also forms a WSD in the base case as an adjective and a verb: oq kabutar (white dove)/suv to'qinlanib oqdi (water rippled);

Oq (Verb) + ish (noun form of action): oqish (to flow);

When it comes in the form of Adj(ish) + Noun Oqish – Adj. Oqish rang (Flowing color), Oqish gul (flowing flower);

When it comes Noun, Vr(gerund), + Verb (ish) – Verb. Suvda oqish (Flowing in water).

**c) WSD between noun and adverb.**

A WSD between a noun and an adverb result from the addition of a suffix to the stem.

$$N + cha = Adj. \leftrightarrow Noun \quad (4)$$

Let's see the example of the word *Yigitcha*.

*Yigit+cha* (diminutive form of noun): an endearing term for little boys. *Yosh yigitcha* (A young man), *aqlli yigitcha* (a smart young man).

*Yigit+cha* (formative suffix): *erkaklardek*(to behave like men).

Here the suffix *-cha* is a mutual homonymous suffix. Joining the word also forms a WSD. This WSD is determined by the context of the sentence or text, not by its association with words that precede or follow it. The reason is that homonyms in both categories can be connected to verbs after the word itself. But *yigit+cha=Noun* is differentiated by whether it can be connected to an adjective before it. *Madaniyatli yigitcha* (Cultured young man)/Adj.+Noun.

#### a) WSD between noun and verb

The WSD of certain word forms is one of the most pressing problems for computational linguistics. The result of adding some words with a homonymous base is another homonymous form. For example, the lexeme of *terim*.

$$HW \rightarrow [N_1, N_2, N_3] \quad (5)$$

HW  $\rightarrow$  N<sub>1</sub>. *ter*(Verb. The act of picking something) + *im* (noun-forming suffix) = *terim* (picking work)

HW  $\rightarrow$  N<sub>2</sub>. *ter* (Noun. Liquid from a person's body) + *im* (possessive form) = *ferim* (skin on my forehead)

HW  $\rightarrow$  N<sub>3</sub>. *ter* (Noun. Part of the human body) + *m* (possessive form) = *terim* (my skin).

As we have already seen, all three homonymy words are nouns. Even when we divide such words into their meaningful parts, i.e., in the form of base and suffix, it preserves its WSD. We can determine such homonymous forms based on the context of the sentence only with the help of conjunctions that come before or after it. But if other grammatical forms are added to the base of such homonyms, the WSD between them disappears.

In the Uzbek language, homonyms often form WSD within the three-word group. Determining which categories of homonyms form WSD and developing models will prevent errors in the process of tokenizing homonymous lexemes.

Below are the words that make up WSD within the three-word families:

$$Noun \leftrightarrow Adj \leftrightarrow Verb$$

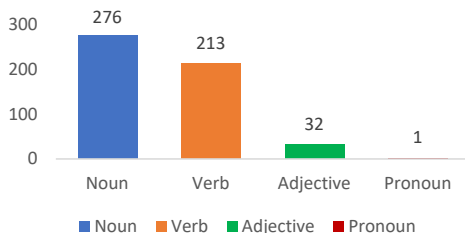
Num↔Noun↔Verb

Adv↔Noun↔Adj

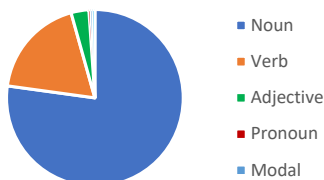
Noun↔Pro↔Verb

Noun↔Adv↔Verb

In the process of collecting the database of homonymous words in the Uzbek language, the statistics of words forming WSD within the Noun, Verb, Adjective, Pronoun part of speech were calculated. The results are shown in the chart below:



**Diag 1: Statistics of homonyms database by part of speech**



**Diag 2: Statistics of words that generate WSD within a part of speech**

The results show that in the Uzbek language nouns are found in the largest number within a word group. No words forming WSD were found within the number and ravish word groups.

As a result of distinguishing homonyms in the Uzbek language by categories, it became known that problematic situations arise in the process of distinguishing homonyms in the corpus because names can take the same grammatical forms. Our task is to develop appropriate models for lexemes that take the same grammatical forms. But we also observed such words that the meaning and category they represent can be determined only through the context.

### **C) Creating a WSD approach for the Uzbek language corpus**

Corpus linguistics is a branch of computational linguistics that deals with the development of general principles of construction and use of a linguistic corpus (text corpus) with the help of computer tech-



nologies. A linguistic corpus or text corpus is a large machine-readable, combined, tagged, formatted, philologically sound collection of language data created to solve specific linguistic problems.

It is known that the study, research and analysis of language phenomena has its importance in every era. Phenomena such as homonymy, synonymy, antonymy and polysemy in world linguistics have been studied in a wide range from the point of view of computational linguistics, which is considered new for Uzbek linguistics, and its practical results are used for purposes such as corpus creation and improvement.

In the few researches carried out in the field of Uzbek computer linguistics, there are efforts to create analysis programs designed for the computer memory to “recognize” and “read” homonymous units, and some comments on the problems of tagging homonyms in the Uzbek language and initial efforts to create a WSD detection algorithm have been made.

Currently, <http://uzbekcorpus.uz>, which contains more than 50 million words and is being researched by professor Nilufar Abdurakhmanova and her students, is a relatively perfect corpus for the Uzbek language. [Agostini, 2021] Since 2018, research work is being carried out on this corpus. Despite the research work carried out on the corpus, the corpus still has its shortcomings, in particular, the problems related to homonyms in the interface have not yet been fully resolved. This article discusses some of the shortcomings and their solutions. [Abdurakhmonova, 2018]

In order to investigate the issue of WSD in the computer intelligence system, it is necessary to perform the following tasks:

- extract homonyms from Uzbek language works, explanatory dictionaries and homonyms dictionary;
- determining the level of distribution of homonyms;
- to reveal the systemic character of WSD;
- show symmetry/asymmetry (dissymmetry) aspects of homonyms;
- classification of homonyms according to word groups;
- sort homonyms in text processing;
- creating a matrix of homonyms;
- modeling homonyms.

In order for the semantic analyzer to distinguish homonyms, their linguistic filter and models should be developed first. Morphological and syntactic factors play an important role in distinguishing homonyms. Morphological in WSD between noun and verb; morphological-syntactic in WSD between adjective and verb; syntactic-morpho-

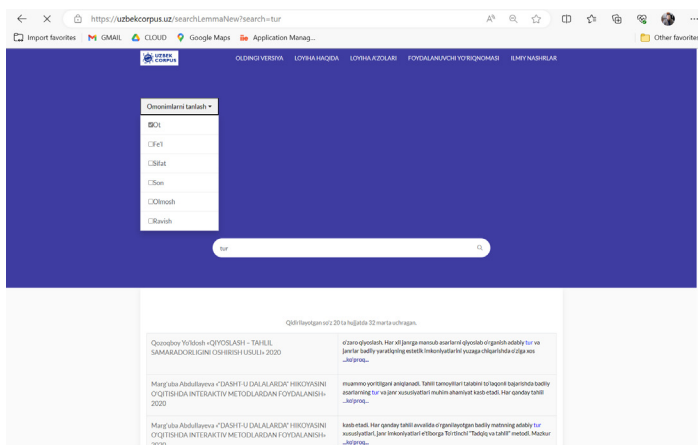
logical in WSD within the group of only nouns and only verbs; syntactic only in WSD within the adjective category; it was found that morphological, syntactic, morphological-syntactic and syntactic-morphological factors lead in WSD within different categories.

On the basis of the SQL database, the collected homonyms are separated by categories and their models are created.

The tables below show the base of suffixes they take to create homonym models. (Pic. 1) With the help of this database of grammatical forms, it is intended to eliminate the shortcomings that arise in the process of searching by token. When the user searches for the required lexeme, the program analyzes this lexeme by separating its grammatical forms from right to left and identifies the token.

| NOUN+N.gsh (ot lug') |               |                               |
|----------------------|---------------|-------------------------------|
| <b>N.gsh</b>         |               | m, im, miz, imiz, ng, ing, ng |
| Aff.eg               | egalik        | iz, ingiz, i, si              |
| Aff.kl               | kelishik      | ning, ni, da, dan, ga, ka, qa |
| Aff.kp               | ko'plik       | lar                           |
| Aff.xs               | xoslik        | dagi [da+gi= ə]               |
| Aff.qr               | qarashlilik   | niki [ni+ki= ə]               |
| Aff.kch              | kichraytirish | cha, chak, loq, choq          |
| Aff.ek               | erkalash      | oy, xon, jon, gina, bek       |
| Aff.chg              | chegara       | gacha [ga+cha= ə]             |
| Aff.bm               | bog'lamalar   | man, san, miz, siz, dir       |

Pic. 1: Base of grammatical forms.



Pic. 2: Interface of the WSD tool in Uzbek Corpus

Using the homonym database, we created a tool that identifies WSD in the Uzbek Corpus, algorithmizing the use of homonyms in grammatical forms and parts of speech. (Pic. 2) The interface of this tool has a search box with the ability to filter by part of speech. Homonyms searched by the user will be found among more than 50 million words in the corpus.

### **CONCLUSION:**

In this paper, a comparative study of WSD in different international and Turkish languages was conducted. Different approaches used in research in these languages were studied. In order to solve the WSD problem in the Uzbek corpus, as a first step, a WSD tool was created using the knowledge-based WSD approach.

In the future scientific and research works, it is aimed to create WSD for the Uzbek Corpus using Statistical approach and Hybrid approaches.

### **REFERENCES**

1. Loric Vial, Benjamin Lecouteux, Didier Schwab. Sense Embeddings in Knowledge-Based WSD. 12th International Conference on Computational Semantics (IWCS), 2017.
2. Ranjan Pal, Diganta Saha, WSD: A Survey. International Journal of Control Theory and Computer Modeling (IJCTCM) Vol.5, No.3, July 2015.
3. Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. An Enhanced Lesk WSD Algorithm through a Distributional Semantic Model. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 1591–1600, Dublin, Ireland.
4. Vasile Rus, Mihai Lintean, Rajendra Banjade, Nobal Niraula, and Dan Stefanescu, SEMILAR: The Semantic Similarity Toolkit. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 163–168, Sofia, Bulgaria. Association for Computational Linguistics, 2013.
5. Patrick, Y. and Timothy, B., “Verb Sense Disambiguation Using Selectional Preferences Extracted with a State-of-the-art Semantic Role Labeler”, Proceedings of the 2006 Australasian Language Technology Workshop (ALTW2006), pages 139–148.
6. Yarowsky D. Decision Lists for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French’, in Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, pp. 88–95. 1994.
7. Voorhees, E.M., Using WordNet to Disambiguate Word Senses for Text Retrieval, In Proceedings of SIGIR-93, pages 171–180, Pittsburgh, PA, USA, 1993.

8. Durga Prasad Palanati, Ramakrishna Kolikipogu. Decision List Algorithm for WSD for TELUGU Natural Language Processing. International Journal of Electronics Communication and Computer Engineering Volume 4, Issue (6), 2013 NCRTCST-2013, ISSN 2249-071X
9. Jean Veronis and Nancy M. Ide. WSD with Very Large Neural Networks Extracted from Machine Readable Dictionaries. In COLING 1990 Volume 2: Papers presented to the 13th International Conference on Computational Linguistics.
10. Hwee Tou Ng. Exemplar-Based WSD Some Recent Improvements. In Second Conference on Empirical Methods in Natural Language Processing. 1997.
11. Martin-Wanton, T., Berlanga-Llavori, R., “A clustering-based Approach for Unsupervised WSD”, *Procesamiento del Lenguaje Natural*, Revista no 49 septiembre de 2012, pp 49-56. [http://rua.ua.es/dspace/bitstream/10045/23919/1/PLN\\_49\\_05.pdf](http://rua.ua.es/dspace/bitstream/10045/23919/1/PLN_49_05.pdf) date: 14/05/2015.
12. A.Agostini, T.Usmanov, U.Khamdamov, N.Abdurakhmonova, M.Mamasaidov. Uzwordnet: A lexical-semantic database for the uzbek language. Proceedings of the 11th Global Wordnet conference. pp 8–19, 2021/1.
13. N.Abdurakhmonova, J.Isroilov,. Personal names spell-checking—a study related to Uzbek. *Journal of Social Sciences and Humanities Research*. Volume 6, 2018.
14. N.Abdurakhmonova, U.Tuliyev, A.Gatiatullin, Linguistic functionality of Uzbek Electron Corpus: *uzbekcorpus. uz*. 2021 International Conference on Information Science and Communications Technologies (ICISCT). 2021/11/3.
15. N.Abdurakhmonova, Dependency Parsing Based On Uzbek Corpus. *Language technology for all (LT4all)*. 2019.
16. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J., (1990) “WordNet An on-line Lexical Database”, *International Journal of Lexicography*, 3(4): 235–244.
17. Lesk, M.,(1986) “Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone”, *Proceedings of SIGDOC*.
18. Cucerzan, R.S., C. Schafer, and D. Yarowsky, (2002) “Combining classifiers for word sense disambiguation”, *Natural Language Engineering*, Vol. 8, No. 4, Cambridge University Press, Pp. 327-341.
19. N.Abdurakhmonova, J.Isroilov,. O‘zbek tili lotin alifbosi uchun klaviatura yoki “Yunikod” masalasi, *Til va adabiyot ta’limi*. #6, pp 6-7, 2018.
20. <https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>.
21. Saeed, A., Nawab, R.M.A., Stevenson. A word sense disambiguation corpus for Urdu. *Language Resources and Evaluation*, 53 (3). 2019.

УДК 81'322.2

**ОБРАБОТКА НЕСТРУКТУРИРОВАННЫХ ДАННЫХ  
ДЛЯ КОРПУСА КАЗАХСКОГО ЯЗЫКА.***А. Н. Шормакова, Д. Р. Рахимова**Казахский Национальный Университет имени аль-Фараби  
Алматы, Казахстан*

Shormakovaassem@gmail.com, di.diva@mail.ru

Данная статья связана с неструктурированными данными для корпуса казахского языка. Основная цель сделать классификацию, иерархию больших, разных типов документов одной отрасли. В данном случае для анализа и сбора и классификации были взяты юридические документы Республики Казахстан. Был сделан анализ небольшого объема данных на казахском языке. Было разработано использование кроулинговой системы для скачки. Были использованы технологии обработки морфологического анализатора (стоп слова, нормализация) и категоризация этих данных. Связи с этим сразу были откорректированы данные и собраны по иерархии для дальнейшего использования. Перечислены дальнейшие доработки для улучшения работы. Выражаем благодарность за поддержку при разработке этой работы проекта : AP19677835 Исследование моделей и разработка интеллектуальной вопросно-ответной системы на основе семантических подходов для государственного языка в сфере законодательства Республики Казахстан.

**Ключевые слова:** юридические документы, неструктурированные данные, законы, классификация документов

**PROCESSING OF UNSTRUCTURED DATA FOR THE KAZAKH  
LANGUAGE CORPUS***Shormakova Assem, Rakhimova Diana**Al-Farabi Kazakh National University, Almaty, Kazakhstan*

Shormakovaassem@gmail.com, di.diva@mail.ru

This article deals with unstructured data for the Kazakh language corpus. The main goal is to make a classification, a hierarchy of large, different types of documents of the same industry. In this case, legal documents of the Republic of Kazakhstan were taken for analysis, collection and classification. An analysis was made of a small amount of data in the Kazakh language. The use of a crawling system for horse racing was developed. Processing technologies of a morphological analyzer (stop words, normalization) and categorization of this data were used. In connection with this, the data was immediately corrected and collected in a hierarchy for further use. Further improvements to improve performance are listed. We would like to thank the following project for their support in the development of this topic: AP19677835 Research of models and development of an intelligent

question-answer system based on semantic approaches for the state language in the field of legislation of the Republic of Kazakhstan.

**Key words:** legal documents, unstructured data, laws, classification of documents.

**Review of unstructured data.** Structured data is formatted into tables, rows, and columns that follow a clearly described schema with specific data types, relationships, and rules. A persistent schema means that the structure and organization of data is predetermined and agreed upon. Such data is typically stored in database management systems (DBMS) such as SQL Server, Oracle, and MySQL and managed by data analysts and database administrators. Analysis of structured data is usually performed using SQL queries and data mining techniques.

Unstructured data is unpredictable and lacks a consistent pattern, making it difficult to analyze. Without a consistent schema, data may vary in structure and organization. These include formats such as text, images, audio and video. File systems, data lakes, and Big Data processing frameworks like Hadoop and Spark are often used to manage and analyze unstructured data[1].

To collect unstructured data in the Kazakh language, a corpus of the Kazakh language was used, collected from Internet texts within the framework of this project. This task was performed using an automated collection of text data from the Internet. The Scrapy<sup>1</sup> framework was used to automatically collect a corpus of sentences in the Kazakh language. Scrapy is a framework for crawling websites and extracting unstructured data that can be used for applications such as data mining, information processing, historical archiving, data searching.

The site <https://adilet.zan.kz/> and the site <http://online.zakon.kz/> were chosen as a source of text data. The structuring of documents and files is different, so it is necessary to create a hierarchy of documents. Collecting data from the Internet using Scrapy consists of the following steps:

- creating a project for data extraction;
- specifying a list of addresses from which data will be retrieved;
- description of selectors containing the data to be retrieved;
- debugging and launching the data collector;
- saving collected data;
- processing of collected data.

As stated above, unstructured data needs to be classified and divided into a more understandable form. There are many types of areas of

---

<sup>1</sup><https://scrapy.org/>

activity and the legal documents of the Republic of Kazakhstan were reviewed. In this work, in order to expand on the topic of unstructured data, we examined legal documents of the Republic of Kazakhstan. In general, according to last year, more than 2.5 thousand laws were adopted in the Republic of Kazakhstan. Of these, 301 are independent laws. The remaining 90% are laws on amendments and additions, as well as ratified international treaties.

The laws of the Republic of Kazakhstan can be briefly seen as follows<sup>1</sup>:

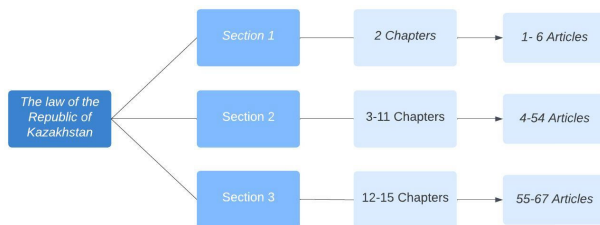


Figure 1. General structure of the law of the Republic of Kazakhstan

Basic and derivative types of normative legal acts are written in Section 2, Chapter 3, Article 7<sup>2</sup>:

Regulatory legal acts are divided into basic and derivative.

The main types of regulatory legal acts include[2-4]:

1) The Constitution of the Republic of Kazakhstan, constitutional laws of the Republic of Kazakhstan, codes of the Republic of Kazakhstan, consolidated laws of the Republic of Kazakhstan, laws of the Republic of Kazakhstan, temporary resolutions of the Government of the Republic of Kazakhstan having the force of law;

2) Regulatory legal decrees of the President of the Republic of Kazakhstan;

2-1) Regulatory legal acts of the Chairman of the Security Council of the Republic of Kazakhstan;

3) regulatory legal decisions of the Parliament of the Republic of Kazakhstan and its Chambers;

4) regulatory legal decisions of the Government of the Republic of Kazakhstan;

<sup>1</sup> [https://online.zakon.kz/Document/?doc\\_id=37312788&pos=3;-106#pos=3;-106](https://online.zakon.kz/Document/?doc_id=37312788&pos=3;-106#pos=3;-106)

<sup>2</sup> [https://online.zakon.kz/Document/?doc\\_id=37312788&pos=247;-43#pos=247;-43](https://online.zakon.kz/Document/?doc_id=37312788&pos=247;-43#pos=247;-43)

5) normative decisions of the Constitutional Court of the Republic of Kazakhstan, the Supreme Court of the Republic of Kazakhstan;

6) regulatory legal decisions of the Central Election Commission of the Republic of Kazakhstan, the Supreme Chamber of Auditors of the Republic of Kazakhstan, the National Bank of the Republic of Kazakhstan and other central government bodies;

7) regulatory legal orders of the ministers of the Republic of Kazakhstan and other heads of central government bodies;

8) regulatory legal orders of heads of departments of central government bodies;

9) regulatory legal decisions of maslikhats, regulatory legal decisions of akimats, regulatory legal decisions of akims and regulatory legal decisions of audit commissions.

2. Derived types of regulatory legal acts include:

1) position;

2) technical regulations;

3) rules;

4) instructions.

5) other forms

Legal documents were classified for further analysis. For each item, materials were collected to collect a corpus of legal documents in the Kazakh language.

### **Data processing method**

These groups of methods are associated with the processing of received information and its interpretation. There are many types and forms of information:

– false and true;

– documentary and supported by documents;

– written and oral, graphic and verbal;

– scientific, political, technical.

Of the types listed, documentary is used. The classification method was chosen. In this method, we consider the distribution of objects, concepts, and phenomena into classes depending on their common characteristics, and sometimes according to the presence of their dissimilarities.

### **Technical implementation of unstructured data processing**

As a result, 67 legal documents in the Kazakh language were processed. The following data was extracted: publication date, publication



title, publication text (additions, changes to documents). After downloading, the data was exported to json [Geddis, 2019; Zlatopol'skiy, 2017; Rikhter, 2013] format with UTF-8 encoding. The data in json format was processed using the jq utility, a json processor for the command line. As a result, a monolingual corpus of the Kazakh language was obtained consisting of 14,500 sentences and 217,500 words. The resulting corpus will be used to study the features of the Kazakh language that affect the quality of search.

Algorithm for collecting unstructured data in Kazakh language

Figure 2. consists of the following steps: data analysis process; crawling systems; data normalization; data categorization.

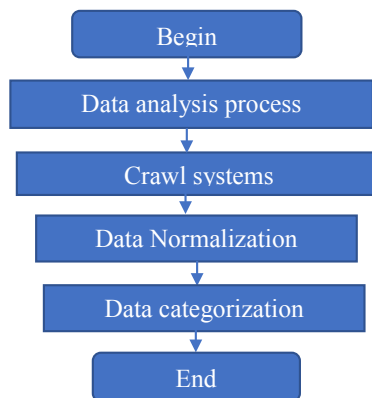


Figure 2. Algorithm for collecting unstructured data in Kazakh language

### Future works

In the future, it is planned to supplement the corpus on the Kazakh language with processed materials. Also consider the semantic connection of documents with each other. Update the software accordingly and update the data.

### Gratitude

We express our gratitude for the support in developing this topic to the project with grant funding : AP19677835 Research of models and development of an intelligent question-answer system based on semantic approaches for the state language in the field of legislation of the Republic of Kazakhstan.

## Conclusion

An analysis was made of a small amount of data in the Kazakh language. The use of a coupling system for horse racing was developed. Processing technologies of a morphological analyzer (stop words, normalization) and categorization of this data were used. Due to this, we immediately corrected the data and collected it in a hierarchy for further use. As a result, the features of documents in the Kazakh language were described. Legal documents have been classified for further more convenient processing when implementing software. A general algorithm for processing unstructured data is demonstrated. Further refinements and collection of information from legal documents were considered.

## REFERENCES

1. [<https://habr.com/ru/articles/756454/>] Date of the application: 29.09. 2023.
2. Zakonodatel'stvo // Kazakhstan. Natsional'naya entsiklopediya. – Almaty: Қазақ entsiklopediyasy, 2005. – Т. II. – ISBN 9965-9746-3-2. [Legislation // Kazakhstan. National Encyclopedia. – Almaty: Kazakh encyclopedias, 2005. – Т. II. – ISBN 9965-9746-3-2.
3. 7 stat'ya Zakona Respubliki Kazakhstan «O normativnykh pravovykh aktakh». [Article 7 of the Law of the Republic of Kazakhstan “*On normative legal acts*”]. [https://online.zakon.kz/Document/?doc\\_id=37312788&pos=247;-43#pos=247;-43](https://online.zakon.kz/Document/?doc_id=37312788&pos=247;-43#pos=247;-43). Date of the application: 29.09. 2023.
4. Zakona Respubliki Kazakhstan «*O normativnykh pravovykh aktakh*». [Law of the Republic of Kazakhstan “*On normative legal acts*”] Date of the application: 29.09. 2023.
5. Geddis T. Nachinaem programmirovat' na Python. – 4-e izd.: Per. s angl. – SPb.: BKhV-Peterburg, 2019. – 768 s. [Gaddis T. Getting started programming in Python. – 4th ed.: Transl. from English – St. Petersburg: BHV-Petersburg, 2019. – 768 p.]
6. Zlatopol'skiy D.M. Osnovy programmirovaniya na yazyke Python. – M.: DMK Press, 2017. – 284 s. [Zlatopolsky D.M. Basics of programming in Python. – M.: DMK Press, 2017. – 284 p.]
7. Rikhter, Dzheffri CLR via C#. Programmirovanie na platforme Microsoft .NET Framework 4.0 na yazyke C# / – M.: Piter, 2013. – 928 c. [Richter, Jeffrey CLR via C#. Programming on the Microsoft .NET Framework 4.0 platform in C# / – M.: Peter, 2013. – 928 c.]

УДК 81-25

**О СТРУКТУРНО-СЕМАНТИЧЕСКОМ МОДЕЛИРОВАНИИ  
НА МАТЕРИАЛЕ КОРПУСНЫХ ПРОЕКТОВ  
БАШКИРСКОГО ЯЗЫКА****З. А. Сиразитдинов***Уфа, Башкортостан, Россия  
ИИЯЛ УФИЦ РАН*

В статье анализируются обобщенно-личные пословицы из базы Машинного фонда башкирского языка. На основе элементарно простых предложений выделяются модели и структурно-семантические типы пословичных паремий. Подробно рассматриваются двухактантные структуры.

**Ключевые слова:** башкирский язык, пословицы, синтаксис, простое предложение, обобщенно-личные предложения, структурная схема, модель, семантические роли.

**ABOUT STRUCTURAL AND SEMANTIC MODELING  
BASED ON THE MATERIAL OF CORPUS PROJECTS  
BASHKIR LANGUAGE****Z. A. Sirazitdinov***Ufa, Bashkortostan, Russia  
RIHLL UFIC RAS*

The article analyzes generalized personal proverbs from the base of the Bashkir language Machine Fund. On the basis of elementary simple sentences, models and structural-semantic types of proverbs are distinguished. Two-act structures are considered in detail.

**Keywords:** Bashkir language, proverbs, syntax, simple sentence, generalized personal sentences, block diagram, model, semantic roles.

Выявление моделей синтаксических структур языка имеет большое практическое значение в разработке автоматизированных систем обработки языка. Такие исследования активно проводились тюркологами в последние 20 лет прошлого столетия [Ахматов, 1983; Тыбыкова, 1991; Кетенчиев, 1993]. В частности, на материалах башкирского языка осуществлено монографическое исследование Д. С. Тикеевым [Тикеев, 2004].

Однако подходы авторов к моделированию приводили к большому количеству структурных схем синтаксических конструкций тюркских языков. Понятие минимальной структуры предложения (ЭПП), введенное основателем новосибирской синтаксической

школы Черемисиной М. И. [Черемисина. 1989], позволяет создавать компактные модели и структурные схемы предложений.

Моделирование на основе ЭПП широко используется лингвистами, но существуют разные подходы к определению ее компонентов [6: 26; 7: 65; 8: 125; 9: 284; 10: 97].

К облигаторным актантам мы вслед за синтаксической школы Черемисиной относим предметных участников ситуации и локальные распространители при глаголах движения, перемещения, местонахождения [11-14].

В данной статье при описании синтаксических моделей мы рассматриваем только пословичные выражения с простыми финитными глаголами.

Материалом для исследования являются односоставные глагольные обобщенно-личные пословицы фольклорного корпуса башкирского языка. В паремиях рассматриваемой группы двухактантные синтаксические структуры составляют большинство (191 единица).

В таблице 1 даны статистические данные по двухактантным структурным схемам пословичных выражений. Таблица показывает, что схема (S) + N<sub>ACC</sub> + V<sub>f</sub> с объектом в винительном падеже является наиболее высокоупотребительной.

Таблица 1.

**Структурные схемы пословичных выражений, представленных двухактантными обобщенно-личными клаузами**

| №     | Структурная схема                         | ед. | %   |
|-------|---|-----|-----|
| 1     | (S) + N <sub>ACC</sub> + V <sub>f</sub>   | 117 | 62  |
| 2     | (S) + N <sub>DAT</sub> + V <sub>f</sub>   | 51  | 26  |
| 3     | (S) + N <sub>Abl</sub> + V <sub>f</sub>   | 14  | 7   |
| 4     | (S) + N <sub>instr</sub> + V <sub>f</sub> | 7   | 4   |
| 5     | (S) + N <sub>loc</sub> + V <sub>f</sub>   | 2   | 1   |
| Итого |   | 191 | 100 |

В схеме (S) + N<sub>ACC</sub> + V<sub>f</sub> находит отражение 8 моделей ЭПП, каждая из них характеризуется своим набором семантических ролей и типовым значением.

1. Модель физического воздействия на объект [(S) + N<sup>Pat</sup><sub>ACC</sub> + V<sup>Act</sup><sub>f</sub>] (62, здесь и далее в скобках указаны численность





‘Волка не испугаешь, закидав шапкой’.

8. Модель социального отношения к объекту [(S) + N<sup>obj</sup><sub>ACC</sub> + V<sup>social</sup>] с пропозицией “кто кого/что защищает/надеется” (4).

Пропозиция социального отношения в пословичных выражениях реализуется глаголами *һакла* ‘хранить, беречь’ (3) и *көт* ‘ждать’ (1). Актантами при глаголе *һакла* выступают абстрактные слова (*һамыс* ‘честь’ и *дан* ‘слава’), конкретное существительное *юлдаш* ‘спутник’:

*Һамысыңды* *йәштән* *һакла.*  
Совість - N(POSS1SGP2ACC) смолоду - +ADV хранить -  
V(IMPP2Sg).

‘Храни совесть смолоду’.

### Заключение

Структурная схема (S) + N<sub>ACC</sub> + V<sub>f</sub> элементарного простого предложения является наиболее частотной среди обобщенно-личных пословичных выражений, поскольку эта структурная схема является базовой в реализации представления о действии с указанием самого глагола и двух необходимых участников: субъекта и объекта в винительном падеже [Черемисина. 2008: 121]. Данная схема является простым типом синтаксических конструкций и можно предположить, что синтаксические особенности, выделенные в конструкции данного типа, с большой вероятностью проявятся в современном литературном языке и разговорной речи с более сложной синтаксической структурой.

### Сокращения, использованные в статье при глоссировании

| Глосса           | Расшифровка                    |
|------------------|--------------------------------|
| Abs              | категория лишительности        |
| Acc              | аккузатив определенной формы   |
| Acc <sub>0</sub> | аккузатив неопределенной формы |
| Dat              | датив                          |
| Fut.Indf         | будущее неопределенное время   |
| Gen              | генитив                        |
| Ger1             | деепричастие на -n             |

## Продолжение таблицы

| Глосса    | Расшифровка  |
|-----------|--|
| Ger2      | деепричастие на <i>-гас/-гэс</i>                             |
| Imp       | императив  |
| Loc       | локатив  |
| Loc.Atr   | категория атрибутивного локатива (с аф. <i>-дагы/-даге</i> ) |
| Neg       | аффикс глагольного отрицания                                 |
| P1        | показатель множественного числа                              |
| P2        | показатель 2-го лица   |
| P3        | показатель 3-го лица   |
| Poss. Atr | категория притяжательности (с аф. <i>-дыкы/-деке</i> )       |
| Poss      | категория принадлежности                                     |
| Pst.Ptcp  | причастие, прошедшего неопределенного времени                |
| Sg        | показатель единственного числа                               |

## ЛИТЕРАТУРА

1. Ахматов И. Х. Структурно-семантические модели предложения в современном карачаево-балкарском языке. Нальчик: Эльбрус, 1983. 360 с
2. Тыбыкова А. Т. Исследования по синтаксису алтайского языка. Простое предложение. Новосибирск: изд-во НГУ, 1991. 228 с.
3. Кетенчиев М. Б. Формально-семантические модели именных предложений в современном карачаево-балкарском языке. Нальчик: Изд-во КБГУ, 1993. 80 с
4. Тикеев Д. С. Основы синтаксиса современного башкирского языка. М.: Наука, 2004. 312 с.
5. Черемисина М. И. О теоретических вопросах модельного описания предложения // Предложение в языках Сибири. Новосибирск: Наука, 1989. С. 3–18.
6. Шведова Н. Ю. Спорные вопросы описания структурных схем простого предложения и его парадигм // Вопросы языкознания. 1973. № 4. С. 25–36.
7. Распопов И. П. Что же такое структурная схема предложения? // Вопросы языкознания, 1976. № 2. С. 65–70.
8. Никитин М. В. Основы лингвистической теории значения. М.: Высшая школа, 1988. 168 с.



9. Аппоев А. К. Семантическая структура паремических высказываний в карачаево-балкарском языке // Вестник Челябинского государственного педагогического университета. 2014. № 2. С. 281–290.
10. Алексанова С. А. Обстоятельственные распространители в структуре простого предложения // Вестник Адыгейского университета. Серия 2. Филология и искусствоведение. 2009. № 1. С. 93–98.
11. Чугункова А. Н. Глаголы движения и формируемые ими модели простого предложения (на материале хакасского языка): автореф. дисс... канд. филол. наук. Новосибирск, 1998. 20 с.
12. Байжанова Н. Р. Модели элементарных простых предложений в алтайском языке: структурная схема ЭПП  $N_1-V_r$ . Новосибирск: Наука, 2004. 176 с.
13. Ойноткинова Н. Р. Алтайские пословицы и поговорки: поэтика и прагматика жанров / отв. ред. О. Н. Лагута. Новосибирск: Редакционно-издательский центр НГУ, 2012. 354 с.
14. Черемисина М. И. Итоги исследования простого предложения в языках Сибири // Языки коренных народов Сибири. Новосибирск: 1998. Вып. 4. С. 3–31.
15. Черемисина М. И., Озонова А. А., Тазранова А. Р. Элементарное простое предложение с глагольным сказуемым в тюркских языках Южной Сибири. Новосибирск: Любава, 2008. 205 с.

УДК 811.512.141'342

**ФУНКЦИОНИРОВАНИЕ ИНОЯЗЫЧНЫХ  
ФАРМАКОФИТОНИМОВ В БАЗЕ ДАННЫХ МФБЯ  
И В УСТНОМ ПОДКОРПУСЕ СМИ<sup>1</sup>**

*А. Ш. Ишмухаметова*

*Ордена Знак Почета Институт истории, языка и литературы  
Уфимского федерального исследовательского центра РАН*

*Уфа, Россия,*

*ishmuhametova\_anita@mail.ru*

В статье рассматриваются функционирование иноязычных фармакофитонимов в башкирском языке. Материалом исследования послужили словари, а также материалы базы данных Машинного фонда башкирского языка (далее – МФБЯ) и устного подкорпуса средств массовой информации башкирского языка, разработанный лабораторией лингвистики и информационных технологий (ныне – отделом прикладной лингвистики и диалектологии) Института истории, языка и литературы УФИЦ РАН. Объектом исследования стали иноязычные названия лекарственных растений в башкирском языке. Цель данной статьи заключается в описании заимствованных слов, вошедшие в состав фармакофитонимов в башкирском языке.

**Ключевые слова:** башкирский язык, машинный фонд башкирского языка, база данных, растительная лексика, лекарственные растения, заимствования, иноязычные слова, фармакофитонимы.

**FUNCTIONING OF FOREIGN-LANGUAGE  
PHARMACOPHYTONYMS IN THE DATABASE OF THE  
BASHKIR LANGUAGE MACHINE FUND AND IN THE ORAL  
SUBCORPUS OF MASS MEDIA**

*Ishmukhametova A. Sh.*

*Order of the Badge of Honor Institute of History, Language  
and Literature of the Ufa Federal Research Center of the Russian  
Academy of Sciences, Ufa, Russia*

*ishmuhametova\_anita@mail.ru*

The article discusses the functions of foreign-language pharmacophytonyms in the Bashkir language. The research material was dictionaries, as well as material databases of the Bashkir Language Machine Fund (hereinafter referred to as MFBL) and the oral subcorpus of the Bashkir language mass media, developed by

---

<sup>1</sup> Исследование выполнено за счет гранта Российского научного фонда № 23-28-01343 «Кодовые переключения в условиях башкирско-русского двуязычия (на материале диалектных дискурсов)».

the Laboratory of Linguistics and Information Technologies (hereinafter referred to as the Department of Applied Linguistics and Dialectology) of the Institute of History, Language and Literature of the Ufa Scientific Center of the Russian Academy of Sciences. The object of the study was the foreign-language names of medicinal plants in the Bashkir language. The purpose of this article was to describe loanwords included in the composition of pharmacophytonyms in the Bashkir language.

**Keywords:** Bashkir language, Bashkir language machine fund, database, plant vocabulary, medicinal plants, borrowings, foreign words, pharmacophytonyms.

Использование корпусов является полезным инструментом для исследования языка, поскольку позволяет изучать язык в более широком контексте. Корпусы предоставляют доступ к большим объемам текста, что обеспечивает больше данных для изучения языка. Кроме того, использование корпусов может помочь в анализе различных аспектов языка, таких как фразы, структура предложений, использование определенных слов и грамматические конструкции. Поэтому использование корпусов нашло применение в современной исследовательской практике и стало важным достижением лингвистической науки.

Разработанный лабораторией лингвистики и информационных технологий (ныне – отдел прикладной лингвистики и диалектологии) Института истории, языка и литературы Уфимского федерального исследовательского центра Российской академии наук Машинный фонд башкирского языка (далее – МФБЯ) является хранилищем практически всего лексического богатства башкирского языка. Данная информационная система на сегодняшний день включает в себя 7 крупных баз данных (генеральную картотеку, лексикографическую базу, грамматическую базу, каталога рукописных книг, каталога старопечатных книг, экспериментально-фонетическую базу, диалектологическую базу) и 3 корпусных проекта (корпуса прозаических текстов, корпуса публицистики, корпуса фольклорных текстов) (Рис.1) [Сиразитов, 2020, с. 76].

В данной статье будут рассмотрены некоторые иноязычные фармакофитонимы на материале данных МФБЯ.

Среди фармакофитонимов в составе лексики башкирского языка значительное место занимают арабские, персидские, русские заимствования и иноязычные слова, заимствованные через русский язык.

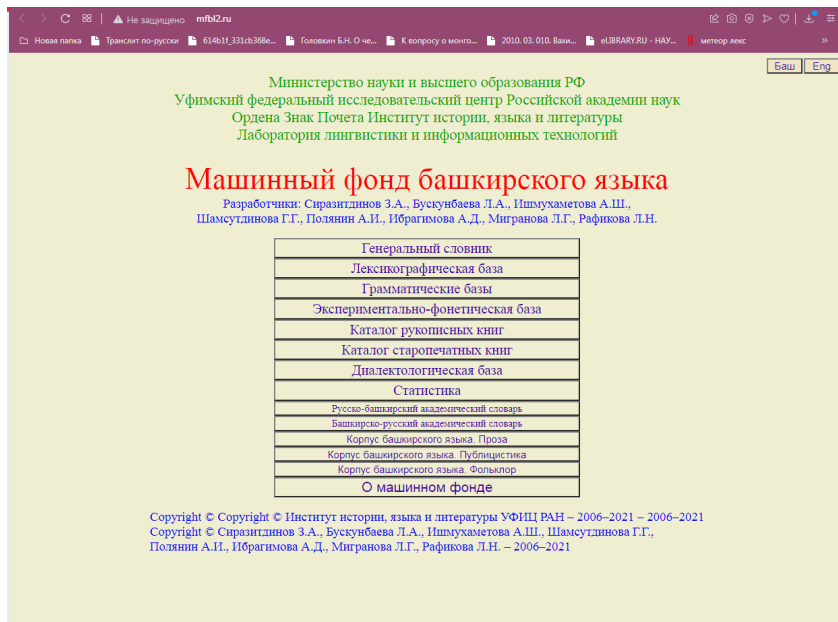


Рис.1. Интерфейс Машинного фонда башкирского языка: **mfb12.ru**

Многие языки мира берут наименования растений из латинского языка. Это связано с тем, что в долгие века латынь была языком науки и познания, и большинство научных терминов и названий было сформировано именно на латыни. К примеру для наименования алоэ используется заимствованное слово *алоэ* / *алоэ* лат. *Aloë*. В Лексикографической базе МФБЯ, в которой представлены более 50 словарей башкирского языка, даются синонимы этой лексемы – *йөзйәшәр* (досл. долгожитель), *шеш гөлө* (досл. цветок от опухоли) [МФБЯ: Лексикографическая база].

Для обозначения данного слова в башкирских диалектах используются следующие слова: в среднем говоре южного диалекта фонетический вариант *алуи*, в демском говоре южного диалекта – *алуигөл* (досл. алоэ+цветок), в кызыльском говоре восточного диалекта – *сәнескеле гөл* (досл. колючий цветок), в караидельском говоре северо-западного диалекта, среднем говоре южного диалекта – *сәнескәкле гөл* (досл. колючий цветок), в миасском, айском говорах восточного диалекта – *сәскәкте гөл* (досл. колючий цветок) [МФБЯ: Диалектологическая база].

В народной медицине алоэ применяется при гастрите и других заболеваниях желудка, кашле, запоре, туберкулезе, воспалении легких, при аллергических и воспалительных заболеваниях кожи и др. Народные рецепты, в составе которых имеется данное растение, зафиксированы в текстах газет, вошедших в Публицистический корпус МФБЯ. Ср.: *Халык медицинаһында алоэ шулай ук туберкулезды дауалағанда, йүтәлләүҙән, йөкөһөзлөктән кулланыла. Киске Өфө/15.05.2010/20 Өлкәндәр өсөн дә алыштырғыһыз ризыҡ ул – һөт/мөхәрририәт/5* [МФБЯ: Корпус башкирского языка. Публицистика] – ‘В народной медицине алоэ применяется также при лечении туберкулеза, кашли, от бессонницы’.

*Яран* ‘герань’ лат. *Geranium* – относится *яран һымактар ғаиләһе* к семейству гераниевые лат. *Geraniaceae* Juss.

В «Русско-башкирском словаре башкирского языка» отмечено несколько разновидностей этого растения: *һаз яраны* ‘герань болотная’, *кызыл яран* ‘герань кровянокрасная’, *урман яраны* ‘герань лесная’, *туғай яраны, саған сәскә* ‘герань луговая’, *кысыр сәскә* ‘герань Роберта’ [МФБЯ: Лексикографическая база].

В Диалектологической базе МФБЯ зафиксированы следующие диалектные варианты данного цветка: в среднеуральском говоре восточного диалекта – *мәриймана* (досл. Марьямана ← Марьям + ана (мать)), в миасском говоре восточного диалекта – *мәриймана күз йәше* (досл. слезы Марьямана), а в том же говоре кровянокрасная герань – *алмайара* (досл. яблоко+рана) [МФБЯ: Диалектологическая база].

Герань, которую выращивают в домашних условиях, называется *ярангөл* (досл. герань+цветок). Это растение, имеющее антисептическое, антивирусное, противовоспалительное, противоотечное, ранозаживляющее и др. действия, отмечено в газетных текстах Публицистического корпуса МФБЯ. Ср.: *Арығанда, депрессиянан, баш ауыртқанда ярангөлдө ескәргә кәрәк* [МФБЯ: Корпус башкирского языка. Публицистика] – ‘При усталости, депрессии, головной боли, нужно понюхать герань’.

Башкирское слово *кәнәфер* заимствовано из арабского *لفنرق* ‘с обильными цветками вечно зеленое тропическое дерево и его семена’ [Бейшев, 2009, с. 64] → ‘гвоздика’. Для обозначения данной лексемы в башкирских диалектах используются следующие слова: *күпертке* – в сакмарском говоре южного диалекта, *кәләмфер* – в аргаяшском говоре восточного диалекта, *кәләмфер* – в сакмарском говоре южного диалекта и во всех говорах восточного

диалекта: кызыльском, аргаяшском, миасском, сальютском, айском [МФБЯ: Диалектологическая база].

Народные рецепты, в составе которых имеется данное растение, зафиксированы в текстах газет, вошедших в Публицистический корпус МФБЯ: *Яман шеш ауырыузарына каршы көрәшеу өсөн сәйгә корица, бер нисә кипкән кәнәфер һәм балгалак осонагына элп имбирь кушып, бешекләп эшһән, файза булып. Киске Өфө/23.10.2010/43 Халык дауаһы/Ф. Бикембәтова/2* [МФБЯ: Корпус башкирского языка. Публицистика] – ‘Для борьбы против злокачественными опухолями будет полезен заваренный чай с корицей, несколькими высушенными гвоздиками и на кончике чайной ложки имбирем’.

*Миләүшә* ‘фиалка’ лат. *Viola*. В северо-западном диалекте башкирского языка это слово отмечено как *күкбаиш*. Данную лексему М. Ряснен возводит к персидскому *bynařsa ~ bänäřsä* ‘фиалка’ [Ряснен, 1969, с. 69]. Слово функционирует во всех тюркских языках: тур. *menekşe*, гаг. *menevša*, аз. *bänövšä*, турк. *bynařsa*, ктат. *menewše*, тат., каз. *miläwšä*, кирг. *binarša* [СИГТЯ, 1997, с. 143]. Также бытует в диалекте уйгурского языка *bänpäšä*.

До революции заимствовалась главным образом хозяйственно-бытовая терминология, подвергавшаяся значительной звуковой обработке в соответствии с произносительными нормами башкирского языка. Сюда можно отнести названия следующих фармакофитонимов: лит. *арыш* ‘рожь’, диал. *пирэй* (демск. гов. южн. диал.) лит. *актамыр* ‘пырей’, диал. *сәснүк / сөснөк* (сакмарск. гов. южн. диал.) лит. *һарымһак* ‘чеснок’ лит. *картуф* ‘картофель’, диал. *шарфый* (средн. гов. южн. диал.) / *шәлфей* (кызыльск. гов. вост. диал.) ‘шалфей’, *маркуф* (средн. гов. южн. диал.) ‘морковь’ и др. [МФБЯ: Диалектологическая база] Эти заимствования подвергнуты изменению согласно структурным особенностям языка. Такое изменение претерпели все ранние заимствования из русского языка, поскольку в этот период русские слова в башкирский язык проникали в основном через разговорный язык. Анализ фармакофитонимов показывает, что поздние заимствования используются без изменений: *петрушка, календула, хризантема* и т.д. Об этом свидетельствуют материалы, которые вошли в устный подкорпус средств массовой информации башкирского языка: [*рәфия буранбаева*] *гәлийемеш / # календула # / # чабрец # / # малина # / кара карагат һәм еләк япрактарында / балтырғандың һабағында һәм япрактарында / с витамини күп //* [рафия буранбаева] в

шиповнике / календуле / чабреце / малине / черной смородине и на листьях ягоды / на стебле и листьях борщевика / много витамина с' [МФБЯ: Текстологическая база].

Текстологическая база МФБЯ первоначально создавалась для представления иллюстративного материала по говорам на основе «Образцов башкирской разговорной речи», но в дальнейшем база наполнилась текстами транскрипций фонетической речи по говорам восточного диалекта, собранным и транскрибированным в рамках проведенной лабораторией НИР «Создание корпуса диалектных текстов башкирского языка» (2017-2021). По материалам базы данных можно наблюдать кодовое переключение, т.е. «переход говорящего в процессе речевого общения с одного языка на другой в зависимости от условий коммуникации» [Багана, 2010, с. 64], в данном случае использование в башкирской речи русских терминов. В ходе беседы информант использует русское название растений, т.е. происходит переключение кодов. Ср.: # *гваздикалар* # # *шуттан* # *бөткөһөз* / *сәскәләрем нык күп* // лит. *кәнәферзәр һанап бөткөһөз* / *сәскәләрем нык күп* // 'гвоздик не сосчитать / у меня очень много цветов //' (ж., 48, среднее, кызыльский)<sup>1</sup>; # *и* # *әсәйзең йараткан сәскәләр* # *роза* # *гортензийалар үсә* // лит. *һәм әсәйзең йараткан сәскәләре рауза гортензиялар үсә* // 'и мамыны любимые цветы розы гортензии растут' // (ж., 14, среднее, кызыльский) [МФБЯ: Диалектологическая база]. В вышеприведенных примерах кодовое переключение с башкирского на русский язык мотивировано тем, что в процессе коммуникации по каким-либо причинам информант не может вспомнить нужное слово в родном языке и переключается на русский.

Материалы баз данных МФБЯ становятся полезными для изучения фармакофитонимов в башкирском языке. Выявленные диалектные и иноязычные фармакофитонимы представляют собой неотъемлемой частью развития языка, которое позволяет расширить лексические возможности языка, отразить межкультурные связи. Таким образом, заимствование фармакофитонимов является эффективным способом обогащения растительной лексики любого языка и является важным элементом языковой и культурной коммуникации.

<sup>1</sup> На примерах представлены пол, возраст, образование и говор информанта.

## ЛИТЕРАТУРА

1. Багана Ж., Блажевич Ю.С. К вопросу о переключении кодов // Научные ведомости. Серия гуманитарные науки. 2010. № 12(83). Выпуск. 6. С. 64.63-68.
2. Бейешев Ә. Ғ. Башкорт телендә йөрөгән ғәрәп һәм фарсы һүззәре / Ә. Ғ. Бейешев. – Өфө, 2009. – 137 с.
3. МФБЯ: Диалектологическая база. URL:<http://mfbl2.ru/mfbl/bashdial> (дата обращения: 30.08.2023).
4. МФБЯ: Корпус башкирского языка. Публицистика. URL:<http://212.193.134.139:8080/bashcorp/korpubp4> (дата обращения: 30.08.2023).
5. МФБЯ: Лексикографическая база. URL: <http://mfbl2.ru/mfbl/bashlex> (дата обращения: 30.08.2023).
6. МФБЯ: Текстологическая база. URL:<http://mfbl2.ru/mfbl/bashdial> (дата обращения: 30.08.2023)
7. Образцы башкирской разговорной речи / ответ. ред. Н.Х.Максютова. Уфа, 1988. 224 с.
8. Сиразитдинов З., Бускунбаева Л., Ишмухаметова А., Шамсутдинова Г. Диалектологический ресурс башкирского языка // Ұлы дала тұлғалары: академик Шора Сарыбаев және ұлттық тілтаным тағылымы. Алматы, 2020. С. 75–81.
9. Сравнительно-историческая грамматика тюркских языков: Лексика. – М.: Наука, 1997. – 800 с.
10. Räsänen M. Versuch eines etymologischen wörterbuchs der türksprachen. Helsinki. 1969. 533 s.



УДК 81'322.2

**СОЗДАНИЕ АВТОРСКОГО КОРПУСА ЗАХИРИДДИНА  
МУХАММАД БАБУР – ТРЕБОВАНИЕ ПЕРИОДА*****М. А. Абжалова***

*Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои  
Ташкент, Узбекистан  
abjalova.manzura@gmail.com*

Авторский корпус – это электронная система, охватывающая творческие тексты конкретного автора, с грамматическими, семантическими и стилистическими тегами его стиля письма, языка и особенностей текста в его произведениях. Такой корпус является важной электронной системой для более глубокого понимания творчества конкретного Творца, исследования его текстов, изучения мастерства и стиля письма создателя, понимания языка той эпохи, в которой жил создатель. В связи с этим актуальным вопросом является создание авторского корпуса Захириддина Мухаммада Бабура. Такие корпуса, в которых концентрируются произведения зрелых творцов, считаются ценными в связи с их исторической, воспитательной, социальной, духовной значимостью в образовательном процессе. В данной статье были рассмотрены возможности и актуальность авторских корпусов.

**Ключевые слова:** Захириддин Мухаммад Бабур, авторский корпус, семантический тег, семантическая база, образовательный процесс.

**CREATION OF THE AUTHOR'S CORPUS OF ZAKHIRIDDIN  
MUKHAMMAD BABUR – PERIOD REQUIREMENT*****Manzura Abjalova,***

*Tashkent State University of Uzbek Language and Literature.  
Tashkent, Uzbekistan.  
abjalova.manzura@gmail.com*

The author's corpus is an electronic system covering the creative texts of a particular author, with grammatical, semantic, and stylistic tags of his writing style, language and features of the text in his works. Such a corpus is an important electronic system for a deeper understanding of the creativity of a particular Creator, the study of his texts, the study of the skill and style of writing of the creator, and understanding the language of the era in which the creator lived. In this regard, the creation of the author's corpus of Zakhiriddin Mukhammad Babur is an urgent issue. Such buildings, in which the works of mature creators are concentrated, are considered valuable due to their historical, educational, social, and spiritual significance in the educational process. In this article, the possibilities and relevance of the author's corpora were considered.

**Keywords:** Zakhiriddin Mukhammad Babur, author's Corpus, semantic tag, semantic base, educational process.

A linguistic corpus is a complex of certain language units stored electronically, which is considered a source of solving various problems for linguists, and a system of lingua-didactic and educational value for users. The object of language corpora is natural language texts, and the subject is language corpora. The presence of a special search system (corpus manager) in the corpus, the grammatical and semantic interpretation of lexical units in the texts shows its difference and even superiority over other types of philological systems.

The practical efforts to create the corpus began in the 60s of the 20th century, and today it is both a scientific and a practical process that is gaining great interest all over the world and is rapidly accelerating and developing. The corpus allows you to quickly and conveniently find unknown words and phrases and provides an opportunity to check and determine the grammatical features of the word. Depending on the preservation of language units, with the help of special programs, it is possible to immediately determine an array of examples of its use, variants of spelling, and synonymous lines for a specific word or phrase. This possibility exists in the educational corpus of the Uzbek language created by the team of specialists at the Tashkent State University of Uzbek Language and Literature named after Alisher Navoi serves to enrich the observation [<http://uzschoolcorpara.uz/>]. The possibilities of the corpus are used in the development of lexicographic practice, translation studies, linguidactics, terminology, and speech competence [Abjalova, 2022:69-72].

According to the totality of the corpora, there are common and single authors. Common corpora are created on one or more issues of a specific language. It can provide grammatical or semantic tags of texts, provide parallel translation, and display metadata about the source of texts. Single-authored corpora mainly cover material belonging to a specific creator and are considered authored corpora.

**Corpus of authorship** is a system that provides information about a specific author's language, writing style, language, and text features in his works. In the field of world linguistics, research on the creation of author corpora is currently being carried out. In particular, from 1956 to 1961, a dictionary of the Pushkin language was published, as well as dictionaries of individual works, for example, in 1978, a frequency dictionary of Leo Tolstoy's novel War and Peace [<http://www.ruscorpora.ru/search>]. This line of research is undoubtedly considered

relevant, and nowadays it is increasingly attracting the attention of researchers and writers in the field of linguistics to the issue of creating modern author's corpora.

In general, the basis for creating author corpora is related to the structure of author dictionaries, and the formation of author corpus creation works has its own tendency. First of all, author's dictionaries were created, and special dictionaries were created for the words or author's expressions used in the works of some writers. Building a vocabulary is hard work. It is required to work with thousands of cards. In this arduous process, the body is a great helper. This is why language corpora were created. A reference to the source of the contexts in which that word has participated (concordance) is provided as soon as a specific word is typed in the search bar. It is also possible to collect phrases from the works of a particular artist. In order to further optimize this opportunity, it is desirable to create author corpora.

What are the factors of creation of authorship corpora and what are the requirements for the texts included in their database? As mentioned, the national corpus is a set of texts that show all the possibilities of the language, therefore it is considered a general linguistic system. Authorship corpora are characterized by the fact that they contain the texts of the works of a certain creator.

Finding the author of unknown works as a result of studying the works of classic writers, digitizing information about the writer's personality and writing style, widely promoting the work of certain famous artists, and using their works in the research process. Author corpora have emerged in order to find an array of examples of units. More precisely, with the development of information technologies, the possibility of working with large-scale materials has increased, and author corpora have also been created.

The corpus of works belonging to a particular writer provides the following opportunities:

- 1) study the writer's style and his skill in using words;
- 2) linguopoetic analysis of literary works;
- 3) analysis of stable combinations (phrases) used by the creator;
- 4) analysis of the types of poetic art that are widely used to enhance the artistry of the work;
- 5) identify samples of folk oral creativity (proverbs, riddles) hidden among lyrical genres;
- 6) distribution of the works of the creator according to the age of the audience;

7) to be able to distinguish the type of the text of the work (for example, the oriphonic, romantic, and rindona types of the ghazal)

8) deeply understand the interpretation of words, the idea, and the content of the text based on the semantic explanation in the corpus;

9) provides a number of convenient opportunities, such as speeding up information acquisition with extensive use of digital technology.

We started the work of creating and presenting the author's corpora with these possibilities in a convenient way for users and in a beautiful design by creating the author's corpus of Alisher Navoi, a genius of Turkish literature, a unique representative of the whole world [Abjalova M., 2022]. This corpus consisting of 8 columns was created as a result of semantic tagging of genres [G'ulomova N., 2022; Abjalova M., 2023] in Navoi's "Badoye ul-wasat" book.

All the works written by Zakhiriddin Mukhammad Babur, one of the great representatives of Uzbek classic literature, are tagged grammatically and semantically, creating his author's corpus is an urgent issue.

There are works of famous writers in different genres, and all of them contain explanatory words that are incomprehensible to today's reader, that is, the user. This situation prevents the student from regularly engaging with Babur's work. In order to understand Babur's work, it is permissible to determine the semantics of the explanatory words in it in the same context, so the reader is forced to refer to the explanatory dictionaries of Mukhammad Babur's works and additionally to other dictionaries. This multi-step process requires patience and perseverance from the student. Unfortunately, these difficulties lead the reader to stop reading classical literature. In the corpus of Zakhiriddin Mukhammad Babur, it is possible to give an explanation of the word in the work in accordance with the context, to find the scope of use of a certain word in the works, to observe its statistics and the place of use. As it is understood, the modern user does not go through the process of working with dictionaries from the beginning but understands the content of the work by looking at the ready-made annotations formed by the creators of the corpus using semantic tags in the database. It is these possibilities that are not available on other sites or systems, only in the author's corpora. As a result of the creation of the corpus of Zakhiriddin Mukhammad Babur:

- to study Babur's personality;
- study of the poet's style;
- lingua poetic analysis of his work;

- researching the ability and skill of the creator to use words;
- creation of authorship dictionaries;
- summarizing the expressions of authorship;
- author's paraphrase, parema, summary of wisdom; the scope of use of figurative expressions can be determined from the context of the creator.

In short, the creation of Zakhiriddin Mukhammad Babur's corpus of authorship creates the following opportunities:

1. AC of Babur is considered a modern pedagogical technological tool that enriches the scientific research and educational process with convenient opportunities. is a system.

2. The life and work of Zakhiriddin Mukhammad Babur is studied at all levels of secondary schools. Therefore, the teacher will have the opportunity to quickly prepare fresh, meaningful, reliable, educational material for the training session from the corpus of Babur's authorship.

3. In order to understand the language of Babur's era in Zakhiriddin Mukhammad Babur's author corpus, to increase the readership of classical sources related to the 15th century, it is important to provide users with the semantic explanation of the explanatory words in the works of Z.M. Babur in the modern Uzbek literary language. becomes important.

4. The author corpus of Zakhiriddin Mukhammad Babur serves the user to obtain accurate and complete information about all the linguistic features of the word units used by the thinker. It helps to understand and understand the content of the work by easily finding the explanation of the words found in the works.

5. When a specific word is searched in the author's corpus of Zakhiriddin Mukhammad Babur, metadata about its place of use in Babur's works, contextual meaning, general usage statistics, and source of the word are provided. Such an opportunity increases the efficiency of scientific research and educational process.

6. The creation of Zakhiriddin Mukhammad Babur's author corpus, the effective use of words used in classic works in the educational system, the introduction of the national-literary heritage of the Uzbek people of the 15th century into digital technology, the creation of a formal form of the state language that has been living for centuries, a linguistic translator and text analysis programs provide opportunities to create a parallel corpus of works by Zakhiriddin Mukhammad Babur.

## REFERENCES

1. Abjalova M. Corpus Linguistics./ M.A. Abjalova. – Tashkent: Bookmany print, 2022. – P. 69–72.
2. Abjalova M., Gulomova N. Author’s Corpus of Alisher Navoi and its Semantic Database. // IEEE – UBMK – 2022: 7th International Conference on Computer Science and Engineering. 24–26 September 2022. Istanbul – Turkey. – pp. 182–187. Impakt Factor 5.5
3. Abjalova M., Gulomova N., Rashidov H. Semantic base of ghazals in Navoi’s “Badoe’ ul-vasat” divan for Alisher Navoi’s author corpus. Certificate No. BGU 00583. - Tashkent, 2022. (authorship certificate).
4. Abjalova M., Gulomova N., Sadullayeva Sh. Author corpus of Alisher Navoi. Certificate No. DGU 18544. – Tashkent, 2022. (authorship certificate).
5. Гуломова Н. Создание базы данных авторского корпуса Алишера Навои и ее семантических тегов (на основе девана «Бадое ул-васат»): автореф. кан. дисс. – Ташкент, 2023. – 58 с. [Gulomova N. Sozdanie bazy dannyh avtorskogo korpusa Alishera Navoi i ee semanticheskikh tegov (na osnove devana «Badoe ul-vasat»): avtoref. kan. diss. – Tashkent, 2023. – 58 s.]
6. Захаров В., Богданова С. Корпусная лингвистика: учебник. 3-е изд., перераб. – СПб.: Изд-во С.-Петербур. ун-та, 2020. – 234 с. [Zaharov V., Bogdanova S. Korpusnaja lingvistika: uchebnik. 3-e izd., pererab. – SPb.: Izd-vo S.-Peterb. un-ta, 2020. – 234 s.]
7. Зубов А. В., Зубова И. И. Информационные технологии в лингвистике: учеб. пос. М.: Издательский центр «Академия», 2004. [Zubov A. V., Zubova I. I. Informacionnye tehnologii v lingvistike: ucheb. pos. M.: Izdatel’skij centr «Akademija», 2004.]
8. <http://navoiykorpusi.uz/> (access time 09/09/2023)
9. <http://uzschoolcorpara.uz/> (access time 29/09/2023)
10. <http://www.ruscorpora.ru/search> – National corpus of the Russian language). Electronic resource (access time 22/09/2023)

УДК 81'322.2

**ВОЗМОЖНОСТИ АНАЛИЗА КЛАССИФИКАЦИИ  
ДИАЛЕКТОВ И ЯЗЫКОВ НА ЛИНГВОДОКЕ  
(на примере говоров северо-западного наречия  
башкирского языка)**

**Ю. В. Норманская**

*Институт языкознания РАН,  
Институт системного программирования  
им. В. П. Иванникова РАН, Москва, Россия  
julianor@mail.ru*

В настоящее время вопрос о классификации диалектов и даже языков стоит достаточно остро. Для тюркских языков интересно рассмотреть проблему принадлежности диалектов на северо-западе Башкортостана к башкирскому или татарскому языку. В статье будут рассмотрены результаты фонетико-этимологического и глоттохронологического анализа для **литературного башкирского и татарского языков** и семи башкирских диалектов (2 северо-западных, 3 южных и 2 восточных). Анализ показал, что данные глоттохронологии и фонетико-этимологического анализа могут давать различные результаты. Фонетически оба башкирских диалекта более близки к татарскому языку, при этом лексически нижнебельско-ыкский говор от него значительно отличается. Одновременно, хочется отметить, что два северо-западных говора значительно отличаются и между собой как лексически (89% совпадений, время распада IX в. н.э.), так и фонетически.

**Ключевые слова:** Диалекты, тюркские языки, лингвистическая платформа

**ANALYSIS AND CLASSIFICATION CAPABILITIES OF DIALECTS  
AND LANGUAGES ON LINGVODOC**

**Normanskaja Julia**

*Institute of Linguistics RAS,  
Institute of System Programming named after V.P.Ivannikov RAS  
Moscow, Russia  
julianor@mail.ru*

Currently, the question of classifying dialects and even languages is evermore acute. For Turkic languages, it is interesting to consider the issue of whether the dialects in the northwest of Bashkortostan belong to the Bashkir or to the Tatar language. In the article the results of the phonetic-etymological and glottochronological analysis of Tatar, Bashkir and 7 dialects of Bashkir will be discussed. An analysis of two modern Northwestern dialects of the Bashkir language on LingvoDoc shows that the data from glottochronology and phonetic-ethnological analysis can produce different results because lexicon and phonetics can change at different

rates under the influence of language contacts. We can see that phonetically, both Northwestern Bashkir dialects are closer to the Tatar language, while lexically, the Nizhnebel'sko-Ikskaya dialect differs significantly from it. At the same time, it is worth noting that the two Northwestern dialects differ significantly from each other both lexically (89% correspondence, time of divergence estimated to be around 9th century AD) and phonetically.

**Keywords:** Dialects, Turkic languages, linguistic platform

В настоящее время вопрос о классификации диалектов и даже языков стоит достаточно остро. Для тюркских языков интересно рассмотреть проблему принадлежности диалектов на северо-западе Башкортостана к башкирскому или татарскому языку, которая обсуждается очень активно как в работах ученых, так и на лингвистических конференциях. Одни учёные ср. [Миржанова 2006, Шакуров 2012] считают эти говоры башкирскими, другие, например, [Киекбаев 1958, Рамазанова 1968, Булатова 2021] не признают существование третьего башкирского диалекта, в их классификации есть только восточный и южный.

Важным для анализа классификационной принадлежности говоров на северо-западе Башкортостана в XIX в. стало описание фонетических и морфологических особенностей рукописного «уфимского» словаря белебеевского говора башкирского языка, записанного Н.Ф. Катановым, и найденного нами в Государственном архиве Республики Татарстан. Этот словарь доступен он-лайн на платформе ЛингвоДок <http://lingvodoc.ispras.ru/dictionary/2691/1138/perspective/2691/1139/view?page=1>, он, вероятно, был записан Н.Ф. Катановым летом 1897 г. и 1898 г., когда он по поручению историко-филологического университета Императорского Казанского университета совершил поездки в Белебеевский уезд Уфимской губернии, см. подробнее [Катанов 1900].

Разбор графики рукописного белебеевского словаря показывает наличие в нем архаических пратюркских черт, общих с татарским языком, в частности, сохранение ПТ \*č-, \*s-, и некоторых несвойственных татарскому языку, например, сохранение ПТ \*i, так и особых инноваций, не встретившихся в татарском: (\*-gi- > u). Особенно важным для определения классификационной принадлежности северо-западных говоров являлось наличие в белебеевском диалекте в XIX в. особого типа морфонологических чередований, свойственных восточнобашкирскому диалекту (например, аффикс множественного числа имеет варианты в северо-западном диалекте по Н.Ф. Катанову: «*lap – läp, lap – läp*,



*тар – тәр, тар – тәр, дар – дәр, дар – дәр, зар – зәр, зар – зәр»*), и, видимо, прабашкирскому языку. Эффект внешнего сходства с татарским языком в то время присутствовал за счет сохранения пратюркских архаизмов на уровне фонетики по сравнению с другими фонетически более инновационными восточными и южными башкирскими диалектами.

В настоящее время в современных говорах северо-западного диалекта уже утрачен морфонологический тип чередования, характерный для восточного диалекта башкирского языка, поэтому для анализа современной классификационной принадлежности северо-западного диалекта башкирского языка были применены программы платформы ЛингвоДок ([lingvodoc.ispras.ru](http://lingvodoc.ispras.ru)).

## Материалы и методы

В настоящее время наиболее разработанной с точки зрения наличия четких алгоритмов, реализованных в виде компьютерных программ, является классификация языков по количеству родственных слов, между которыми можно установить регулярные соответствия в 100-словном или в 110-словном списке Сводеша. С.А.Старостин предложил специальную формулу скорости распада языков в зависимости от количества различий в списках базисной лексики, на основании которой создан алгоритм построения деревьев языкового родства, реализованный в СУБД Starling ([starling.rinet.ru](http://starling.rinet.ru)), см. подробнее [Бурлак, Старостин 2005].

На платформе ЛингвоДок любой пользователь после регистрации может создавать свои словари и/или корпуса и анализировать материалы других пользователей, авторы которых из разместили в открытом доступе. В 2023 году создана опция «Глоттохронологический анализ языков/диалектов», которую можно запустить в словарях во вкладке «Инструменты». Эта опция может применяться к любому набору языков, в словарях которых доступно более 50 слов из стословного списка М.Сводеша. Список Сводеша – Старостина [Старостин 2007: 784] выбран, поскольку лишь для него обоснованы и разработаны строгие семантические спецификации [Kassian et al. 2010], позволяющие получить достаточно точные сравниваемые данные для разных языков. Согласно предложенной С. А. Старостиним глоттохронологии, см. подробнее [Старостин 1989], из подсчета по формуле, представленной на рис. 1, сначала удаляются заим-

ствования, родственные слова соединяются на платформе ЛингвоДок этимологическими связями, затем подсчитывается процент совпадений между списками двух идиомов и вычисляется время распада.

$$t = \sqrt{\frac{\ln\left(\frac{Nn(t)}{N_0}\right)}{-n\lambda^n \sqrt{Nn(t)}}$$

**Рис. 1.** Формула С. А. Старостина для обчета близости языков и диалектов встроенная в ЛингвоДок

Эта формула, для которой С.А.Старостин подобрал экспериментальным путем лямбду, равную 0,05, дает возможность определить время распада, любого набора языков, см. подробнее [Старостин 1989].

На ЛингвоДоке также встроена функция создания графиков близости языков в форматах 2D и 3D.

Но большинство компаративистов, особенно на Западе, не вполне доверяет классификации языков, созданной на основе анализа стословных списков (о противоречиях, которые возникают при классификации по 100-словным списка по сравнению с традиционным подходом, подробнее см. в [Беликов 2009]). При этом традиционный подход к генетической классификации языков, основанный, в первую очередь, на анализе общих фонетических и морфологических инноваций и используемый большинством ученых-компаративистов до сих пор, не был алгоритмизирован и компьютеризирован, а потому его применение могло варьироваться в зависимости от личности ученого.

На платформе ЛингвоДок в настоящее время в открытом доступе находятся около 2 000 словарей и корпусов, созданных на основании аудиословарей, собранных в полевых условиях в формате .wav, и архивных записей по уральским и алтайским языкам. Этот материал позволил нам приступить к разработке программ анализа данных для уточнения транскрипций языковых данных (ранее транскрибирование для коми языков было выполнено «на слух», без привлечения фонетических программ), этимологического анализа и построения классификаций на основании обчета фонетических инноваций в близкородственных языках и диалектах.

Для определения степени близости диалектов друг к другу с точки зрения фонетических инноваций была разработана программа «Анализ когнатов в разных диалектах одного языка / в разных языках», которую тоже можно найти в любом словаре во вкладке «Инструменты». На первом этапе для каждого символа из транскрипции алгоритм обчисляет его соответствия в словах из других диалектов этого же языка, связанных этимологиями с настоящим словарем. Высвечивается меню, в котором автор отмечает, с какими словарями он хочет проанализировать сравнения:

а) обчисляются корни, заранее соединенные этимологическими связями, исходя из того, что в них первый гласный (сочетание гласных) соответствует первому гласному (сочетанию гласных), первый согласный (сочетание согласных) соответствует первому согласному (сочетанию согласных), второй согласный – второму. На выходе получаем для каждой пары идиомов список соответствий. У автора словаря есть возможность его скачать, проанализировать, проверить правильность транскрипций и этимологий, которые привели к нестандартным рядам соответствий, и внести корректировки в транскрипцию и этимологию. Далее алгоритм перезапускается повторно уже на материале, выверенном автором;

б) алгоритм оценивает, есть ли фонемы, у которых два и более соответствий во втором диалекте. Если две или более фонем из рассматриваемого словаря соответствуют одной фонеме из другого, рассматривается, нет ли позиционного распределения между ними, не учтенного на 1-м этапе. По факту этого обчета система выдает в формате Excel список соответствий между двумя диалектами с возможными правилами распределения;

Эта программа позволяет в полуавтоматическом режиме обрабатывать большие массивы словарных данных (15–20 тыс. единиц) для выделения рядов соответствий и дополнительных распределений между ними на материале фонетических словарей диалектов одного языка и языков близкородственных. Эта функция необходима при обработке диалектных материалов; ее обычно очень не хватает диалектологам для выяснения полного набора рядов соответствий в однотипном материале. Пока что эта функция ЛингвоДока, насколько нам известно, не имеет аналогов в других системах.

## Исследование и результаты

### I часть. Глоттохронологический анализ

На ЛингвоДоке мы провели глоттохронологический анализ литературного башкирского и татарского языков и семи башкирских диалектов:

#### двух восточных:

– словарь книги Бессонова А. Г. Первая после букваря книжка для чтения и первоначальные уроки русского языка для юго-восточных башкир. Казань, 1907;

– аудиословарь диалект села Байназарова кызыльской группы восточного диалекта;

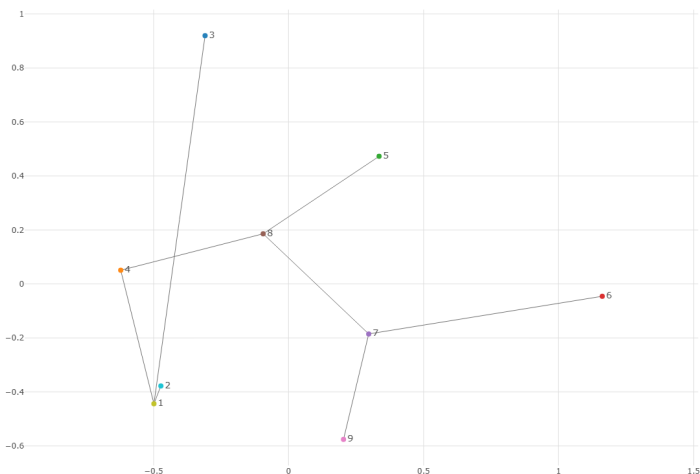
#### трёх южных:

– аудиословарь диалект д. Кинзябулатово ик-сакмарской группы южного диалекта;

– аудиословарь диалект д. Макяш, дёмский южного диалекта;

– аудиословарь диалект д. Хусаиново, дёмский южного диалекта;

Минимальное связующее дерево (встраивание относительного расстояния 2d)



- 1) 1. Словарь книги Бессонова А. Г. Первая после букваря книжка для чтения и первоначальные уроки русского языка для юго-восточных башкир. Казань, 1907
- 2) 2. Словарь диалект села Байназарова кызыльской группы восточного наречия, башкирский язык - носитель: Минзала Хадитова, рожд. 1951; запись и ст
- 3) 3. Словарь диалекта д. Кинзябулатово ик-сакмарской группы южного наречия (Зила Салихьяновна Жунзеева, рожд. 1952). Ишанбайский район, Башкорт
- 4) 4. Словарь диалекта д. Макяш, дёмский говор южного диалекта, носитель: Клара Мухамедиевна Ханидуллина, рожд. 1946 г. Давлекановский район, Башкорт
- 5) 5. Словарь диалекта с. Хусаиново, дёмский диалект южного наречия (Буляккул Миниярович Булатов, рожд. 1935). Давлекановский район, Башкортостан
- 6) 6. Словарь диалекта д. Нижнечереккулево, нижнебельско-ыкской группы северо-западного наречия (носитель - Мадина Миниязовна Рамазанова, рожд. 1
- 7) 7. Словарь говора с. Кузьмьярово гайнинского говора северо-западного диалекта (носитель - Ишанова Дамира Талгатовна, 1957 г.р. Бардымский район
- 8) 8. Словарь 200-словного списка башкирского языка
- 9) 9. фрагмент словаря татарского литературного языка

Рис. 2. График глоттохронологической близости татарского и башкирского языков и башкирских диалектов на ЛингвоДоке

**двух северо-западных:**

– аудиословарь диалекта д. Нижнечереккулево, нижнебельско-ыкской группы северо-западного диалекта;

– аудиословарь говора с. Куземьярово, гайнинского говора северо-западного диалекта.

Результаты подсчета их близости по формуле С.А.Старостина представлены ниже на графике см. Рис. 2. В результате обчета был получен следующий результат близости рассмотренных идиомов, см. Таблица 1 первая цифра обозначает, сколько тысячелетий прошло с момента распада, вторая – процент совпадений у двух идиомов в списках базисной лексики.

**Таблица 1. Время распада языков и диалектов  
и процент совпадающих слов базисной лексики**

|          | 1.<br>книга<br>Бессонов<br>1907 | 2.<br>кызыль-<br>ский<br>говор<br>восточ-<br>ного ди-<br>лекта | 3.<br>ик-<br>сакмар-<br>ский<br>говор<br>южного<br>диалекта | 4. дёмский<br>говор<br>южного<br>диалекта<br>(Макаш) | 5.<br>дёмский<br>говор<br>южного<br>диалекта<br>(Хусаи-<br>ново) | 6. нижне-<br>бельско-<br>ыкский<br>говор<br>северо-<br>западного<br>диалекта | 7. гайнин-<br>ский говор<br>северо-<br>западного<br>диалекта | 8. литера-<br>турный<br>башкир-<br>ский | 9.<br>литера-<br>турный<br>татар-<br>ский |
|----------|---------------------------------|--|---|--|--|--|--|---|---|
| <b>1</b> | n/a                             | -0.00<br>(100%)  | 1.41<br>(83%)   | 0.51<br>(97%)  | 0.99<br>(91%)  | 1.26<br>(86%)  | 0.81<br>(93%)  | 0.95<br>(91%)                           | 0.82<br>(93%)                             |
| <b>2</b> | -0.00<br>(100%)                 | n/a  | 0.94<br>(91%)   | 0.52<br>(97%)  | 0.84<br>(93%)  | 1.46<br>(82%)  | 0.90<br>(92%)  | 0.77<br>(94%)                           | 0.78<br>(94%)                             |
| <b>3</b> | 1.41<br>(83%)                   | 0.94<br>(91%)  | n/a   | 1.18<br>(87%)  | 1.07<br>(89%)  | 1.22<br>(87%)  | 1.01<br>(90%)  | 1.03<br>(90%)                           | 0.95<br>(91%)                             |
| <b>4</b> | 0.51<br>(97%)                   | 0.52<br>(97%)  | 1.18<br>(87%)   | n/a  | 0.73<br>(94%)  | 1.43<br>(82%)  | 0.80<br>(93%)  | 0.70<br>(95%)                           | 1.02<br>(90%)                             |
| <b>5</b> | 0.99<br>(91%)                   | 0.84<br>(93%)  | 1.07<br>(89%)   | 0.73<br>(94%)  | n/a  | 1.31<br>(85%)  | 0.81<br>(93%)  | 0.73<br>(95%)                           | 0.91<br>(92%)                             |
| <b>6</b> | 1.26<br>(86%)                   | 1.46<br>(82%)  | 1.22<br>(87%)   | 1.43<br>(82%)  | 1.31<br>(85%)  | n/a  | 1.10<br>(89%)  | 1.21<br>(87%)                           | 1.22<br>(87%)                             |
| <b>7</b> | 0.81<br>(93%)                   | 0.90<br>(92%)  | 1.01<br>(90%)   | 0.80<br>(93%)  | 0.81<br>(93%)  | 1.10<br>(89%)  | n/a  | 0.68<br>(95%)                           | 0.48<br>(97%)                             |
| <b>8</b> | 0.95<br>(91%)                   | 0.77<br>(94%)  | 1.03<br>(90%)   | 0.70<br>(95%)  | 0.73<br>(95%)  | 1.21<br>(87%)  | 0.68<br>(95%)  | n/a                                     | 0.78<br>(94%)                             |
| <b>9</b> | 0.82<br>(93%)                   | 0.78<br>(94%)  | 0.95<br>(91%)   | 1.02<br>(90%)  | 0.91<br>(92%)  | 1.22<br>(87%)  | 0.48<br>(97%)  | 0.78<br>(94%)                           | n/a                                       |

Помимо таблицы времени распада процентного совпадения ЛингвоДок также выводит полные списки когнатов в базисной лексике (таблица серо-белого цвета), см. Рис. 3.

| Башкортостан, с. 1973 г. Хидитов, рожд. 1951, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Домброва» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Домлекановский» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Домброва» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой |
|--|---|--|--|---|--|---|---|
| Белый [ак] Белый   | Белый [ак] Белый  | Белый [ак] Белый   | Белый [ак] Белый   | Белый [ак] Белый  | Белый [ак] Белый   | Белый [ак] Белый  | Белый [ак] Белый  |
| вода [пу] вода   | вода [пу] вода  | вода [пу] вода   | вода [пу] вода   | вода [пу] вода  | вода [пу] вода   | вода [пу] вода  | вода [пу] вода  |
| все [бары] все<br>все [бары] все<br>глаз [куд *рә *оңо]<br>] глаз                    | глаз [кәб] глаз   | глаз [кәб] глаз  | глаз [куб] глаз  | глаз [кәб] глаз   | глаз [кәб] глаз  | глаз [кәб] глаз   | глаз [кәб] глаз   |
| гора [пу] гора   | гора [пу] гора  | гора [пу] гора   | гора [пу] гора   | гора [пу] гора  | гора [пу] гора   | гора [пу] гора  | гора [пу] гора  |
| грудь [күрәк] грудь  | грудь [күрәк] грудь   | грудь [күрәк] грудь  | грудь [күрәк] грудь  | грудь [күрәк] грудь   | грудь [күрәк] грудь  | грудь [күрәк] грудь   | грудь [күрәк] грудь   |
| два [кә] два<br>два [кә] два,<br>дождь   | два [кә] два  | два [кә] два   | два [кә] два   | два [кә] два  | два [кә] два   | два [кә] два  | два [кә] два  |
| дождь [ауен] осадки, дождь   | дождь [ауен] дождь  | дождь [ауен] дождь   | дождь [ауен] дождь   | дождь [ауен] дождь  | дождь [ауен] дождь   | дождь [ауен] дождь  | дождь [ауен] дождь  |
| дорога, тропя [пу] дорога  | дорога, тропя [пу] дорога   | дорога, тропя [пу] дорога  | дорога, тропя [пу] дорога  | дорога, тропя [пу] дорога   | дорога, тропя [пу] дорога  | дорога, тропя [пу] дорога   | дорога, тропя [пу] дорога   |

Рис. 3. Фрагмент таблицы когнатов в стословных списках башкирских диалектов и татарском литературном языке

Все слова, которые ни имеют этимологические параллели в лексике стословных списков рассматриваемых идиомов, приводятся ниже во второй таблице зелёного цвета, см. Рис. 4

| «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой | «Ишимово» район, Башкортостан, запись и составление А.Ф. Абубакировой |
|---|---|---|---|---|---|---|---|
| горло [тамак] горло   |   |   |   |   |   |   |   |
| эти, это [ашо] эти  |   |   |   |   |   |   |   |
| сухой [пу, науан] жареный, сухой                                      |   |   |   |   |   |   |   |
| эти, это [шано] это   |   |   |   |   |   |   |   |
| кушать [арыу]   |   |   |   |   |   |   |   |
| кара [арыу] брат, рот   |   |   |   |   |   |   |   |
| кара [арыу] кара  |   |   |   |   |   |   |   |
| молочный [тәпә]   |   |   |   |   |   |   |   |
| молочный, небомашой й   |   |   |   |   |   |   |   |
| иля [ат] иля  |   |   |   |   |   |   |   |
| сказать [тәпә]  |   |   |   |   |   |   |   |
| сказать, говорить   |   |   |   |   |   |   |   |
| полный [тәпә]   |   |   |   |   |   |   |   |
| полный  |   |   |   |   |   |   |   |
| корень [тәпә]   |   |   |   |   |   |   |   |
| иля, корень   |   |   |   |   |   |   |   |
| знать [тәпә] знать, узнавать  |   |   |   |   |   |   |   |
| кушать [тәпә]   |   |   |   |   |   |   |   |
| кушать, искусывать  |   |   |   |   |   |   |   |
| все [бар] все   |   |   |   |   |   |   |   |
| не [оңо] не   |   |   |   |   |   |   |   |
| то, то [шә] то  |   |   |   |   |   |   |   |
| огонь [аңуан]   |   |   |   |   |   |   |   |

Рис. 4. Фрагмент слов без этимологии в стословных списках башкирских диалектов и татарском литературном языке

Эти данные можно также скачать в виде файла эксель.

О чем же свидетельствуют данные по проценту совпадающих слов в базисной лексике, собранные в Таблице 1? Считается, что совпадение более 90% является указанием на то, что два диалекта считаются диалектами одного языка. Рассмотрим с какими языками северо-западные говоры **№6 нижнебельско-ыкский**, **№7 гайнинский** имеют более 90% совпадений. Оказывается, что нижнебельско-ыкский ни с какими из рассмотренных языков и диалектов в базисной лексике такого процента совпадений не имеет. Программа ЛингвоДока выявила в нем 11 лексем, которые не имеют когнатов в списках базисной лексики других башкирских диалектов и литературных башкирского и татарского языков: *txf* 'семья', *qojrʒ* 'кора', *kuj* 'гореть', *sirak* 'нога', *qizu* 'гулять', *balsik* 'земля', *katu* 'умирать', *emij* 'грудь', *un* 'тот', *kulvaf* 'плечевая кость, грудь', *toju* 'слышать, чувствовать'. А у гайнинского, наоборот наблюдается больше 90% совпадений со всеми языками и диалектами за исключением нижнебельско-ыкского и иксакмарского. При этом с татарским процент совпадений больше (97%), чем с башкирским (95%), то есть, говор носительницы из с. Куземьярово как будто является одновременно и татарским, и башкирским. Конечно, эти результаты нельзя считать окончательными, необходим анализ большего количества списков базисной лексики от разных носителей и из разных населенных пунктов на северо-западе Башкортостана, но все же полученные материалы заставляют задуматься о том, что северо-западные диалекты не являются гомогенной группой, а очень значительно отличаются между собой, и для каждого из них вопрос о сегодняшней принадлежности к татарскому или башкирскому языку должен решаться отдельно.

## II часть. Фонетико-этимологический анализ

В результате фонетико-этимологического анализа, методология которого была описана выше, для выбранных идиомов были проанализированы транскрипции, объединенные в этимологии частично в автоматическом режиме с последующей ручной проверкой. Количество проанализированных лексем система показывает перед графиком близости, см. Рис. 5.

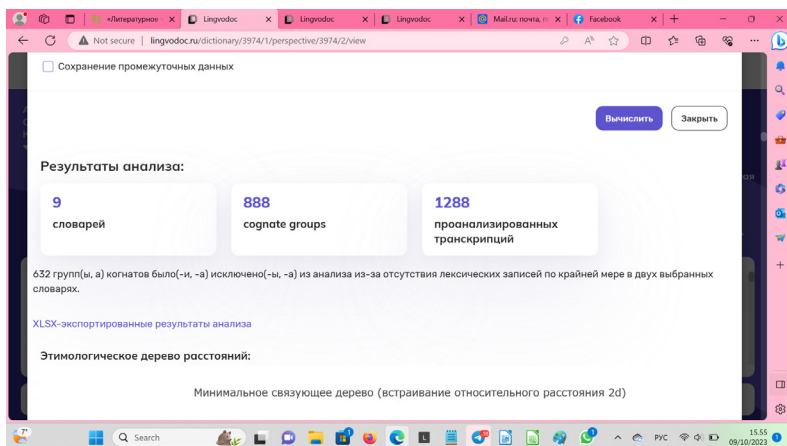


Рис. 5. Количество проанализированных словарей, этимологических групп и конкретных транскрипций

В результате их анализа бы получен следующий график близости, см. Рис. 6.

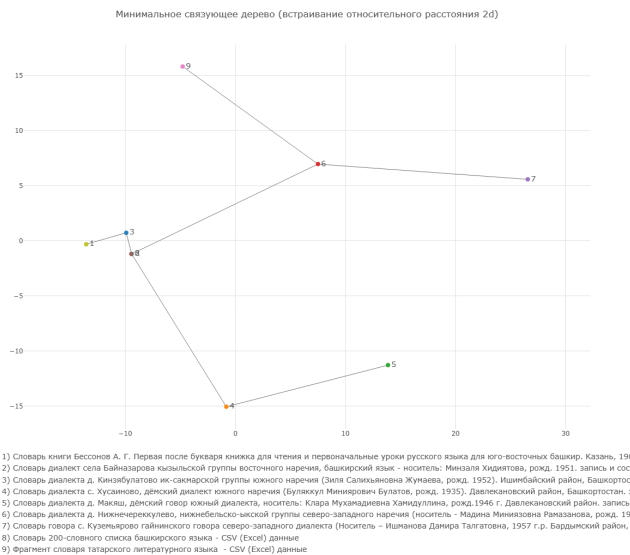


Рис. 6. График степени фонетико-этимологической близости между башкирскими диалектами, литературным башкирским и татарским языками



Ниже в ЛингвоДоке приводятся основные ряды соответствий гласных и согласных в разных позициях и полные списки примеров, которых эти соответствия представлены, см. Рис. 7, 8.

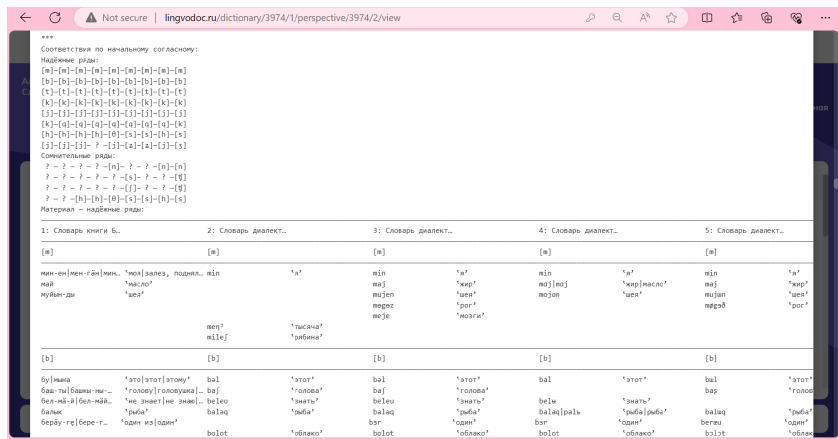


Рис. 7. Фрагмент таблицы соответствий, построенной на ЛингвоДоке с рядами соответствий для анлаутных согласных

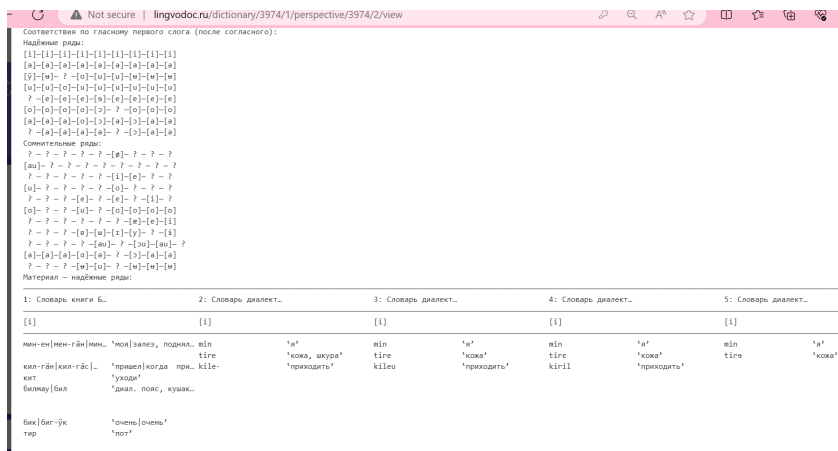


Рис. 8. Фрагмент таблицы соответствий, построенной на ЛингвоДоке с рядами соответствий для гласных первого слога

На рисунках можно видеть, что есть «Надёжные» и «сомнительные» ряды соответствий. «Надёжными» считаются те, у которых в каждом идиоме насчитывается не менее трех примеров

с такой рефлексацией, если примеров меньше, то появляется знак вопроса, если знаков вопроса значимое количество (для 9 словарей более 2), то ряд попадает в категорию «сомнительные».

Все списки слов и соответствий можно также скачать в виде файла Эксель, в конце этого файла приводятся матрицы различий между звуками в рядах соответствий по разным позициям и в заключении суммарная матрица, см. Таблица 2.

**Таблица 2. Матрица фонетико-этимологических различий между башкирскими диалектами, литературным башкирским и татарским языками**

|          | 1. книга Бессонов 1907 | 2. кызыльский говор восточного диалекта | 3. иксакмарский говор южного диалекта | 4. дёмский говор южного диалекта (Макаш) | 5. дёмский говор южного диалекта (Хусаинов) | 6. нижнебельский говор северо-западного диалекта | 7. гайнинский говор северо-западного диалекта | 8. литературный башкирский | 9. литературный татарский |
|----------|------------------------|---|---------------------------------------|--|---|--|---|----------------------------|---------------------------|
| <b>1</b> | 0                      | 4                                       | 4                                     | 25                                       | 25  | 26   | <b>42</b>                                     | 4                          | 17                        |
| <b>2</b> | 4                      | 0                                       | 2                                     | 12                                       | 23  | 13   | 29  | 0                          | 17                        |
| <b>3</b> | 4                      | 2                                       | 0                                     | 20                                       | 23  | 22   | <b>40</b>                                     | 2                          | 19                        |
| <b>4</b> | 25                     | 12                                      | 20                                    | 0  | 22  | 19   | 21  | 21                         | 32                        |
| <b>5</b> | 25                     | 23                                      | 23                                    | 22                                       | 0   | 25   | 27  | 23                         | 24                        |
| <b>6</b> | 26                     | 13                                      | 22                                    | 19                                       | 25  | 0  | 20  | 22                         | 17                        |
| <b>7</b> | <b>42</b>              | 29                                      | <b>40</b>                             | 21                                       | 27  | 20   | 0   | <b>47</b>                  | 33                        |
| <b>8</b> | 4                      | 0                                       | 2                                     | 21                                       | 23  | 22   | <b>47</b>                                     | 0                          | 17                        |
| <b>9</b> | 17                     | 17                                      | 19                                    | 32                                       | 24  | 17   | 33  | 17                         | 0                         |

Из таблицы видно, что наибольшая степень различий (более 40 пунктов, они отмечены полужирным шрифтом) отмечается между южными и восточными диалектами, башкирским литературным языком vs. гайнинским северо-западным говором. Нижнебельский северо-западный говор также отличается от башкирских диалектов, но менее значительно. Оба северо-западных говора фонетически более близки татарскому языку. График на Рис. 6 и цифры в Табл. 2 также показывают наличие особых фонетических изменений и в южных дёмских говорах башкирского языка (21–23 пункта отличий от литературного башкирского языка), от татарского они отличаются еще более значительно (24–32 пункта). С точки зрения глоттохронологии, они не имели более значительного процента отличий от других южных, восточных башкирских диалектов и литературного башкирского языка.

Пример анализа двух современных северо-западных говоров башкирского языка на ЛингвоДоке показывает, что данные глоттохронологии и фонетико-этимологического анализа могут давать различные результаты, поскольку лексика и фонетика могут изменяться с разной степенью скорости под влиянием языковых контактов. Мы видим, что фонетически оба башкирских диалекта более близки к татарскому языку, при этом лексически нижнебельско-ыкский говор от него значительно отличается. Одновременно, хочется отметить, что два северо-западных говора значительно отличаются и между собой как лексически (89% совпадений, время распада IX в. н.э.), так и фонетически (20 пунктов различий, в то время как между восточными башкирскими диалектами и литературным языком всего 4 пункта отличий). Если фонетически такая степень различий является обычной для диалектов одного языка, то лексически они отличаются так сильно, что формально должны считаться разными языками. Таким образом, анализ первых текстов на северо-западном диалекте, где в XIX в. еще были представлены морфонологические чередования, характерные для восточных башкирских диалектов, глоттохронологический анализ нижнебельско-ыкского говора, который значительно отличается и от башкирского, и от татарского языков, и фонетический анализ, выявивший определенную близость северо-западных говоров к татарскому языку, показывают, что классификация северо-западных говоров не является очевидной и не демонстрирует явной принадлежности к татарскому языку, а доказывает необходимость дальнейшего формального исследования говоров из различных населенных пунктов северо-запада Башкирии.

Безусловно, важно также анализировать и степень сходства морфологии языков и диалектов. На ЛингвоДоке для этого в закладке «Инструменты» уже существует опция «Степень морфологической близости между диалектами/языками», которую можно применять к словарям морфем, которые сделаны из глоссированных корпусов. Для современных башкирских диалектов таких корпусов пока нет, но есть корпуса первых башкирских текстов. К морфологическим словарям, которые сделаны по ним, ср. словарь морфем по «Учение... 1899» <http://lingvodoc.ru/dictionary/7532/2/perspective/7532/3/view>, «Евангелие... 1902» <http://lingvodoc.ru/dictionary/7534/2/perspective/7534/3/view>, эта опция уже применима.

Таким образом, на ЛингвоДоке есть возможность строить три графика степени близости диалектов и языков: по фонетико-этимологическому сходству, глоттохронологии, количеству этимологически родственных аффиксов.

#### ЛИТЕРАТУРА

1. Булатова 2021 – *Булатова М.Р.* Татарские говоры Башкортостана: ареальный аспект. Казань: ИЯЛИ, 2021.
2. Бурлак, Старостин 2005 – *Бурлак С.А., Старостин С.А.* Сравнительно-историческое языкознание. М., 2005
3. Киекбаев 1958 – *Киекбаев Дж.Г.* Башкирские диалекты и краткое введение в их историю // Уч. зап. Башк. госуниверситета: сер. филол. – Уфа, 1958.
4. Миржанова 2006 – *Миржанова С. Ф.* Северо-западный диалект башкирского языка (формирование и современное состояние). Уфа, 2006.
5. Старостин 1989 – *С.А. Старостин* Сравнительно-историческое языкознание и лексикостатистика // Лингвистическая реконструкция и древнейшая история Востока. Ч. I. М., 1989: 3–39.
6. Рамазанова 1968 – *Рамазанова Д.Б.* Татар теленен Урта Кама тирәсендә таралган сөйләшләр: дис.канд. филол. наук. Казан, 1968.
7. Шакуров 2012 – *Шакуров Р. З.* Диалектная система башкирского языка // Ватандаш 2012. № 8. С. 40–61.
8. Starostin 2010 – *Starostin G.* Preliminary lexicostatistics as a basis for language classification: A new approach // Journal of Language Relationship, No. 3 (2010). P. 79–116.
9. Kassian et al. 2010 – *Kassian A., Starostin G., Dybo A., Chernov V.* The Swadesh wordlist. An attempt at semantic specification // Journal of Language Relationship, No. 4 (2010), p. 46–89.

УДК: 81`1:004=512.133 81`322.2

## СОЗДАНИЕ И ЗНАЧЕНИЕ ЯЗЫКОВОГО КОРПУСА В УЗБЕКИСТАНЕ

*Г. И. Тоирова*

*Бухарский государственный университет, Бухара, Узбекистан  
tugulijon@mail.ru*

В статье рассматривается трансформация языка в язык Интернета, компьютерные технологии, математическая лингвистика, ее продолжение, а также становление и развитие компьютерной лингвистики, в частности вопрос моделирования естественных языков для искусственного интеллекта. Узбекский национальный корпус играет важную роль в повышении международного статуса узбекского языка. Работа, проводимая в области компьютерной лингвистики, играет важную роль в решении существующих проблем в узбекском языке. В частности, исследован вопрос лингвистического и экстралингвистического разделения специальных тегов для обозначения текстов и их компонентов. Определены требования к кодированию важной текстовой информации. Состояние анализирует лингвистический модуль и алгоритм и его типы из независимых компонентов лингвистического программного кода. Научно обоснована необходимость алгоритмов фонологических, морфологических и орфографических правил формирования лексико-грамматического кода. Подчеркивается значение таких языковых модулей, как фонология, морфология и орфография, в формировании языковой базы национального корпуса узбекского языка. В статье рассматривается основное назначение корпуса как сложного лингвистического источника, а также тот факт, что он содержит преимущественно два вида информации и ее типы.

Ключевые слова. возможностями корпуса, согласно статье, являются сокращение времени, затрачиваемого на процесс анализа текста, и возможность объяснить свойства языковых единиц в речи на тысячах примеров. Национальный корпус, образовательный корпус и параллельный корпус обсуждаются в предмете компьютерной лингвистики. Подчеркнуто, что их лингвистическая и экстралингвистическая маркировка, разработка алгоритмов формирования корпусов, налаживание корпусного лингвистического обеспечения являются общественной потребностью. Признается актуальность разработки основ для создания корпуса узбекского языка, проведения исследований в области компьютерной лингвистики как научно-теоретического источника.

**Ключевые слова:** корпус, искусственный интеллект, лексическая информация, морфологический признак, словесный алгоритм, формульный алгоритм, табличный алгоритм, графический алгоритм.

## CREATION AND IMPORTANCE OF LANGUAGE CORPS IN UZBEKISTAN

**G. I. Toirova**

*Bukhara State University, Bukhara, Uzbekistan*  
tugulijon@mail.ru

The article discusses the transformation of language into the language of the Internet, computer technology, mathematical linguistics, its continuation and the formation and development of computer linguistics, in particular the question of modeling natural languages for artificial intelligence. The Uzbek National Corpus plays an important role in enhancing the international status of the Uzbek language. The work carried out in the field of computer linguistics plays an important role in resolving existing problems in the Uzbek language. The question of the linguistic and extralinguistic separation of special tags for marking texts and their components is studied in particular. The coding requirements for important text information are defined. The state analyzes the linguistic module and the algorithm and its types from independent components of the linguistic program code. The need for algorithms for phonological, morphological and spelling rules for the formation of the lexical and grammatical code is scientifically substantiated. The importance of such linguistic modules as phonology, morphology and spelling in the formation of the linguistic base of the national corpus of the Uzbek language is emphasized.

The article examines the corpus's primary purpose as a complex linguistic source, as well as the fact that it primarily contains two sorts of information and its types. The key effective capabilities of the corpus, according to the paper, are reducing time spent on the text analysis process and being able to explain the properties of language units in speech with thousands of instances. The national corpus, the educational corpus, and the parallel corpus are all discussed in the subject of computer linguistics. It was stressed that linguistic and extralinguistic tagging of them, the development of corpus formation algorithms, and the establishment of corpus linguistic support are all societal need. It recognizes the urgency of developing the basis for the creation of the Uzbek language corpus, conducting research in the field of computer linguistics as a scientific and theoretical source.

**Keywords:** corpus, artificial intelligence, lexical information, morphological sign, word algorithm, formula algorithm, tabular algorithm, graphical algorithm.

### 1. Introduction

Artificial intelligence has enabled a wide range of benefits in the use of language thanks to modern information technology. He is capable of doing a variety of things that the human intellect is capable of. Electronic sources, which are the result of artificial intelligence, are designed to keep humans safe and reduce their weight. Among the most pressing problems are the conversion of the Uzbek language to

the Internet and electronic language, as well as the enhancement of national language electronic resources (Uzbek language corpus, electronic dictionaries, and website contents).

### **Research Question(s)**

We've previously mentioned that languages that have attained world linguistic civilisation have already done work on information processing using computer technology, machine translation, electronic lexicography, the establishment of thesauruses, and the creation of the language corpus. English, Russian, Arabic, French, German, Spanish, and Tajik are just a few of them. The scientific and theoretical aspects of creating a language corpus in the Internet system in these languages have also been established, emphasizing the necessity to speed up efforts to turn the Uzbek language into one that is "understood" by the Internet.

In world linguistics, the generation of language corpora on the Internet is the primary means of maintaining a particular language by the second decade of the twenty-first century, broadening the scope of its research, and demonstrating language skills. Computer technology, in particular, which is a great invention of the twentieth century, opens the door to a wide range of opportunities for linguistics as well as other fields, and imposes enormous tasks on computer language, the emergence of computer linguistics is crucial for the success of natural languages.

In global language studies, the study of linguistic modeling of language, the development of algorithms for word lemming and tags, as well as the electronic use of oral and written monuments, samples of spiritual heritage created in a specific language, in order to increase the use of national and cultural heritage. Particular emphasis is placed on information processing via computer technology, the development of necessary software and methodological software for the introduction of information resources, the development of the language corpus on the Internet, and, on this basis, scientific and theoretical aspects of the national language corpus.

A variety of studies on automatic translation, development of linguistic bases of the author's corpus, processing of lexicographic texts, and linguostatistical analysis have been conducted in Uzbek linguistics. Special emphasis was placed on "enhancing the education system and increasing the capacity of quality educational services." Given that raising the international status of the Uzbek language, elevating it to

the level of a world language of communication, learning and teaching Uzbek abroad, expanding opportunities, and polishing our national language can all be accomplished directly through the national corpus, “theoretical and practical issues of Uzbek national corpus.” Solution is relevant. In this sense, there is a need to further deepen research on the linguistic basis of the text corpus and the national corpus, the technology of creating its software.

## 2. Literature Review

The corpus is the subject of corpus linguistics. This term is variously defined in the scientific literature. For example, it is used in English with terms such as linguistic corpus or text corpus. Recognition of the scientific research of A.N. Khomsky, G.N.Luch, Ch.F.Meer, J.Sinkler, M.Z.Kurdi in solving such problems as creation of the national corpus of a certain language, its analytical technology, development of the field of corpus linguistics should [Mohamed Zakaria Kurdi, 2016; Toirova, 2020, p.57; Charlez, 2004, p. 7; Shomsku, 1962].

John Sinclair defines the term “corpus” as follows: “The corpus consists of a fragment of texts in electronic form selected according to visible criteria for the study of language or linguistic diversity, to be presented as a source of information” (Sinclair, 2004).

Large set of massive texts in Russian corpus linguistics, principles of corpus formation, linguistic database VG Britvin, VP Zakharov, IA Melchuk, AB Kutuzov, RG Kotov, LI Belyaeva, Reflected in the targeted research of E.V.Nedoshivina, V.V.Rykov, V.Plungyan [Britvin, 1983; Bloomfield, 1968; Belyaeva & Chizhakovsky, 1983; Zakharov, 2011; Nedoshivina, 2006; Rykov, 2005; Plungyan, 2005; Kutuzov, 2017; Kotov, 1977].

Russian scientist V.P. Zakharov explains the term “corpus” as follows: “corpus – a set of linguistic data units of language, compiled on the basis of oral and written texts” [Zakharov, 2011].

H.Iskhakova, S.Muhammedov, S.Riza on the linguistic-statistical analysis of the text in Uzbek linguistics, lexicographic processing, linguistic support of the automatic editing program, linguistic modules of the editing and analytical program, synonymous vocabulary of the national corpus, linguistic bases of the author’s corpus. S.Muhammedova, B.Mengliev, D.Urinbaeva, A.Pulatov, U.Dysimova, G.Valieva, G.Jumanazarova, N.Abdurahmonova, Sh.Hamroeva, M.Abjalova, A.Eshmominov, O.Kholiyorov, R. Karimov’s work is noteworthy.



Our scientists, such as S.Karimov, S.Muhammedova, Sh.Hamroeva, conducted research on the specialty 10.00.01 of corpus linguistics.

Uzbek linguists define the term “corpus” as follows: Uzbek linguists interpret the term “corpus” as follows: “corpus is a set of linguistic units that make up a set of texts collected for a specific purpose” [Eshmuminov, 2019], a set of written or oral texts stored in electronic form in a language, placed in a computerized search engine” [Bongers, 1947]. Research in Uzbek linguistics describes the essence of the corpus as follows: “The corpus is the ability to present existing information in the form of text; the ability to provide as much information as possible depending on the size of the case; it is an opportunity to use the data of a once-created corpus repeatedly to solve various problems” [Pulatov, 2011].

“A corpus is a set of texts that are subject to a search engine in order to determine the characteristics of language units, written or oral, stored in electronic form in a natural language, placed on a computer-based search engine software-based on-line or off-line system” [Mengliev, Bobojonov & Hamroeva, 2018] source.

O.Khaliyorov, who conducted research on the “educational corpus”, in his work states the following: The educational corpus of the Uzbek language is a corpus designed to teach the possibilities of the Uzbek language, has a linguodidactical character, contains electronic texts, acts as a special site” [Hamroeva, 2018].

Regarding the parallel corpus, R. Karimov says: “parallel electronic analogue of translated texts; consists of several “original texts and one / several translations of them” [Karimov, 2021].

Language corpora can be divided into different forms in terms of structure, purpose, stability, variability. For example, V.P. Zakharov lists the following forms: “according to the form of data storage: audio, written, mixed; according to the language of the text: monolingual and multilingual; by genre: literary, dialectal, oral, journalistic, mixed; according to the access to the building: free, commercial buildings, closed; by purpose: research, illustrative; according to variability: dynamic and stable; marked and unmarked according to the possession of additional information (annotated)” [Zakharov, 2011].

V.V. Rykov, on the other hand, focuses on the following aspects in the classification of corpus types: “According to the level and structure of the data, according to the chronological sign (position) of the language, according to the language of use, according to the purpose of use” [Rykov, 2005].

In her prohibition, Sh. Hamroeva divides the corpus into the following types: “According to a certain period of language or a certain type of its occurrence (genre, style, a social or age group, the language of a writer or scientist); according to the type of linguistic mark; by type of speech: written, oral, mixed; they look like a multimodal corpus, a corpus of special texts” [Hamroeva, 2018].

“Specialized corpus: a group of texts of a specific type: newspaper text, scientific articles; common building; comparative corpus; parallel corpus; educational building; didactic corpus,” says U.Kholiyorov [Kholiyorov, 2021].

### **3. Methodology**

Each academic defined linguistic corpora from his or her own perspective and categorised them in various ways. What features of the Uzbek linguistic corpus are mirrored in it, and what corpora are now being created?

The creation of the Uzbek language national corpus is a relatively new direction in both Uzbek linguistics and modern information technology. The language corpus is a major source and powerful information resource for compiling large-scale dictionaries. The language corpus allows for the rapid creation and processing of dictionaries using a computer. The importance of the corpus in the field of lexicography is that no tool can match the corpus in determining the period and frequency of use of a word. In the near future, the need for a dictionary today for a student learning a language or a researcher exploring any aspect of it will undoubtedly shift to the corpus.

### **4. Results and Discussion**

Linguists at Tashkent State University of Uzbek Language and Literature named after Alisher Navoi are now working on a project dubbed “Educational Corpus” that is both scientific and practical. The creation of the Uzbek language educational corpus aims to gradually form data based on foreign experience, and includes an electronic textbook containing modern vocabulary of the Uzbek literary language, multilingual speakers, and non-translated lexical units of the Uzbek language, as well as a set of multimedia products, including audio and video materials, as well as a mobile application, aimed at the formation of correct pronunciation skills in Uzbek.

The educational system permits students to study Uzbek as a state language, a second language, and a foreign language in depth. Users may study the Uzbek language freely thanks to the educational building's electronic material, which includes audio, video, multimedia apps, pronunciation and spelling programs, and e-learning dictionaries. Unlike other curriculum, this complex focuses on developing the capacity to utilize the Uzbek language in unusual contexts. Beginners, students, parents, instructors, and students of the Uzbek language may all benefit from it. This will contribute to the formation of scientific and technological resources that will ensure the economic growth and social development of the republic.

As a result of the focus on scientific research in the field of Uzbek computer linguistics on the processing of the Uzbek language using modern information technologies, the first appearances of the national corpus are emerging as practical work.

Participation in the international scientific-practical conference "Theoretical and practical issues of creating Uzbek national and educational corpus" in May 2021, initiated by scientists of the Tashkent State University of Uzbek Language and Literature named after Alisher Navoi, with practical proposals not only domestic but also international did. At the end of the conference, specific tasks were identified to create excellent national and educational corpus, drawing on international experience.

Scientists of Samarkand State University are also working on a project called "Design and development of the national corpus of the Uzbek language" and scientists of Bukhara State University are conducting research on theoretical and practical issues of creating a national corpus of the Uzbek language. The results of the research are of great importance in raising the international status of the Uzbek language, raising it to the level of a world language of communication, learning and teaching the Uzbek language abroad, expanding the capabilities of our national language.

The creation of a national corpus of Uzbek language will make it possible to "digitize" the Uzbek language and turn it into an Internet language.

It opens the door to new opportunities to increase learning effectiveness. It is very easy to find a word, phrase or phrase that is rarely used through the corpus, or the problem with its use and spelling (spelling) is solved in a very short time. Today, more than a grammar scholar, the average researcher needs to know the status, level of appli-

cation of a particular word, phrase, or construction, who used it when, when, and for what style. The corpus is focused on solving similar problems. The National Corpus is necessary to study the lexicon and grammar of the existing language. Another function of the corpus is to provide relevant information in the specified areas (lexicon, grammar, accentology, history of language). The National Corpus is a comprehensive universal information retrieval system that can be used not only by linguists, but also by all those who use the Uzbek language: experts in various fields, scientists, politicians, dictionary designers, researchers and others.

The formation and study of language corpus began some time before the development of the field of corpus linguistics. Examples include eighteenth-century biblical studies (e.g., Cruden), dictionaries (Johnson, Oxford English Dictionary, Webster Dictionary), language teaching (frequency corpus to Thorndike, 1921), and the Quirk Corpus (Survey of English Usage).

With the advent of computer technology, corpus linguistics began to develop rapidly. The Brown Corpus includes texts published in 1961 in the United States. Its capacity is from 2,000 to 500 plates per word. According to the special hierarchy of genres, 5 plates from the daily edition, 2 samples from the weekly editions, 4 plates of detective stories and 20 samples of novels were taken. The original version was presented as a plain text format with no characters<sup>47</sup>. In the selection of the texts, the authors Nelson Francis and Henry Kusera first developed the criteria for corpus formation:

- The origin and content of the text (the author, of course, the English version of the American version, not less than half the volume of the dialogue text);
- synchronization (it is known that it includes texts first published in 1961);
- Selection of individual texts on the basis of the presentation of different genres, the ratio of their numbers and special probability operations;
- Convenience of texts for computer analysis (inserting special characters to convey the originality of the text, etc) [Zakharov, 2011].

The full name of the Brown University is the Brown University Standard Corpus of Present-Day American English. The corpus consists of written versions of the American version of the English language, with a use of 1 million words. The Brown Corpus set the stan-

dard with 1 million words and became a benchmark for creating such corpus in other states. It later emerged that such a standard was unsatisfactory. Appropriate representation of the entity during the application of statistical methods requires only representative selection and large volumes of texts.

The Upsal Corpus (University of Uppsala, Sweden), based on Brown's principles, is also 1 million words in size, which is quite limited in terms of the number of genres it contains [Zakharov, 2011].

As the power of computers increased, it became possible to create relatively large and powerful enclosures. In the UK, the Bank of England Project (BANC) and the British National Corpus (BNC) emerged, which changed the corpus representation standard to 100 million words. It should include full texts, sample words that are common in oral speech patterns, and be easy to access via the Internet.

National corpus of many European languages (Spanish, Italian, Croatian) were created on the basis of British principles. In the Czech Republic, for example, the Czech National Corpus, which includes 100 million-word forms, has been open since 2000.

The representativeness of the corpus is very important. The corpus will not only have to study the variety of events being studied, but it will also have to correctly determine the place of this phenomenon in the lives of the speakers of that language.

The following criteria can be distinguished in the selection of the text for the corpus and the correct assessment of its representativeness:

- Text corpora, which seeks to fully reflect the diversity of the objective existence of speech;
- Cases designed for specific purposes of interest to the researcher [Zakharov, 2011].

Corpus linguists select the representative corpus based on specific conditions. This is mainly due to the fact that the corpus consists of "10–20 million word usages." At the same time, they point out that a "much more alternative by genre" has been created to ensure corpus completeness. In particular, it should cover a wide range of artistic, dramatic, poetic and other texts.

The existing buildings in the global network are characterized by large volume (abundance of materials), as well as deep sockets. For example, Pushkin's and Chekhov's corpuses are morphologically and semantically annotated corpuses. One of the most perfect corpuses. Only modern software does not meet the design requirements. Be-

cause they were 8–10 years old. The next is the perfection of the search system of Shakespeare's authorial corpus, distinguished by the brilliance of the whole design, but not morphologically and semantically marked. The Pushkin and Chekhov corpuses do not have syntactic markings. Many of the world's recognized languages have their own national corpus, which differs in its level of excellence and ability to scientifically process the text. There are about 70 language corporations currently operating on the Internet, including English, Spanish, Chinese, Arabic, French, Russian, German, Polish, Polish-Ukrainian, Czech, Slovak, Serbian, Croatian, Bosnian, Bulgarian, Bulgarian-Russian, Macedonian, Scottish, , Netherlands, Dutch-French, Swedish, Dutch, Norwegian, Icelandic, Faroese, Medieval French, Spanish, Italian, Portuguese, Romanian, Lithuanian, Latvian, Greek, Eastern Armenian, Ossetian, Albanian, Indian, Hittite, Finnish, Uralic languages, Estonian, Veps, Hungarian, Udmurt, Georgian, Anglo-Georgian, Lezgin, Turkish, Tatar, Tajik, Bashkir, Crimean Tatar, Kalmyk, Buryat, Mongolian, Arabic, Hebrew, Amharic, Japanese, ancient Japanese, Baman, Esperanto corpora of languages. Each of the corpus named above has its own advantages and disadvantages. For example, the Tajik language corpus can only act as an electronic library and a photo and video gallery. It is not marked at all, the search engine is imperfect and inconvenient. He can only point to a whole work.

Existing corporations are used for purposes such as statistical analysis of language use, natural language processing (NLP) software, lexical resource creation, language teaching or learning. The texts presented in the corpus are important in the study of the dynamic state of language or in the analysis of the subject of various branches of linguistics. For example, computer analysis and database creation of linguistic resources is one of the tasks of corpus linguistics. It therefore serves as an important electronic resource for the creation of software such as data classification, data processing, machine translation, sentiment analysis.

“Uzbek computer linguistics is formed on the basis of features of the Uzbek language that are completely different from the English language. This shows that before the creation of Uzbek computer linguistics, there is a need to perfectly systematize and formalize the Uzbek language. Bringing rich, broad and deeply developed language issues, such as Uzbek, to the level of computer-based solutions requires a greater amount of work than English” said Pulatov. Agreeing with the

scientist, it is possible to rely on his main ideas, although it is not possible to use English computer linguistics directly in the creation of Uzbek computer linguistics. In the preparation of the linguistic base and the bank of national texts for the formation of the language corpus of the Uzbek language, reference is made to the research work on the national corpus of the Russian language. The national corpus of the Uzbek language should function as both an electronic library and a linguistic corpus, have morphological, semantic, syntactic markings and meet the latest requirements of design, with the advantage of improved search system.

Below we present the plan for the formation of the Central Bank to create a national corpus of the Uzbek language:

MB in Microsoft Access There are two ways to create a table in MB MS Access MBBT. The simplest way to create an MB is to create all the necessary tables, forms, and reports using the MB Wizard. However, you can create a blank MB and then add tables, forms, reports, and other objects to it - this is the most convenient method, but it requires a separate defined object of the MB. In both cases there is a possibility to modify and expand the created MB. To create a new MB, choose Create, New Database, and then Create from the File menu (Figure 1 (a)).

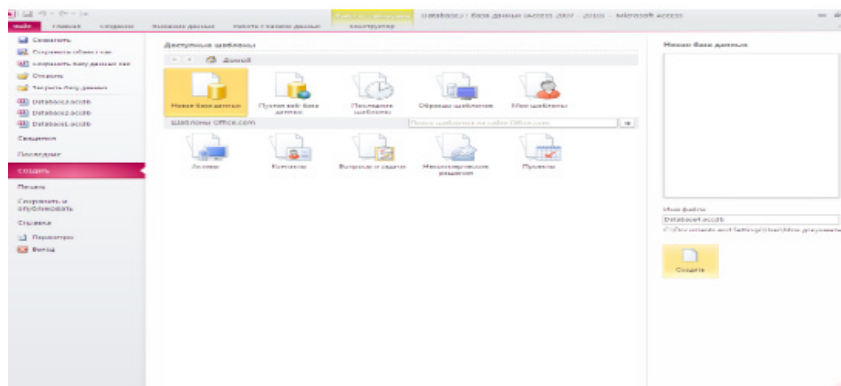


Figure 1. (a). MB window

The Create MB window will then appear on the screen. In the Name field, enter the names of the fields and select the appropriate types, then save the table and switch to table mode (Figure 1 (a)).

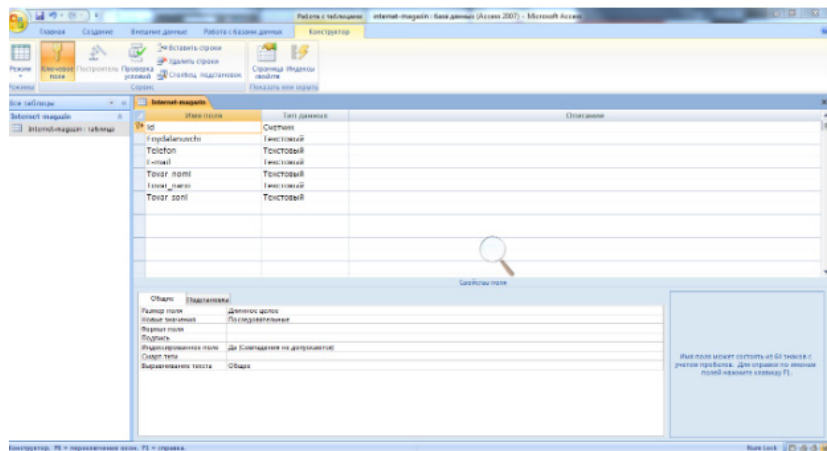


Figure 2 (a). MB window

Then in the resulting window we enter the table values, paying attention to the type of field. Only the appropriate types of data should be entered. For example, <force> must be done in one column, in the second column; noiloj. It is obligatory for a person to do something against his will, by force, or out of necessity”(Figure 2 (b)).

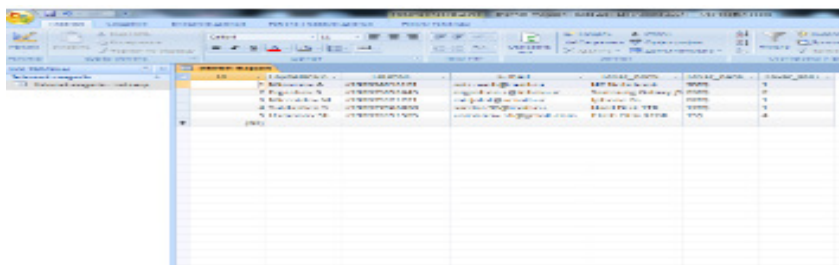
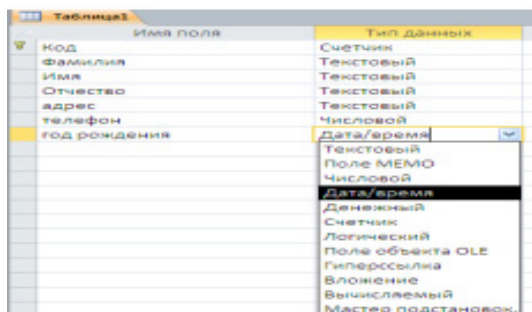


Figure 3 (b). MB window

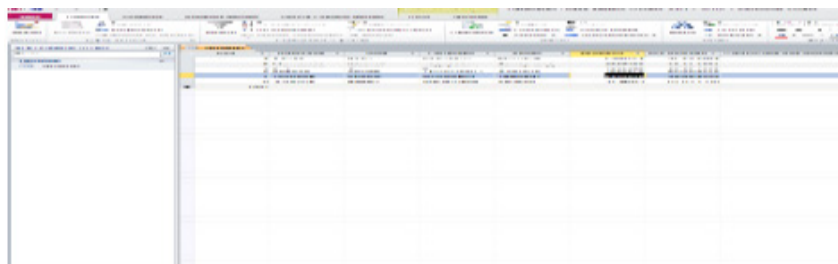
Access allows you to edit table fields and records. You can also change, add, and delete fields in wizard and table mode. Entering and editing data in the table is done in table mode. Access has the following types of data: Basic Type, Number, Data / Time, Da / Net, and Quick Start. When creating a MB, it is necessary to pay attention to the type of data and enter the appropriate data when entering (Figure 4).





**Figure 4 (a).** Data entry and editing window

To add a new record, a table or form opens first. New data is entered in the last line. For example, [f], [status. f.]. Editing the data is done like a simple table (Figure 5).



**Figure 5 (b).** Data entry and editing window

To delete entries, select the appropriate record, right-click and select Delete Record.

MB stores millions of records, from which it is possible to find the necessary information at any time. The data in the MB tables should have simple tools in searching for the required information. Search and sorting is done in tabular mode and by special queries. A matching query is created, resulting in the required records.

The search for information is done through queries, and as a result of the query we have a new table that satisfies the given conditions (Figure 6).

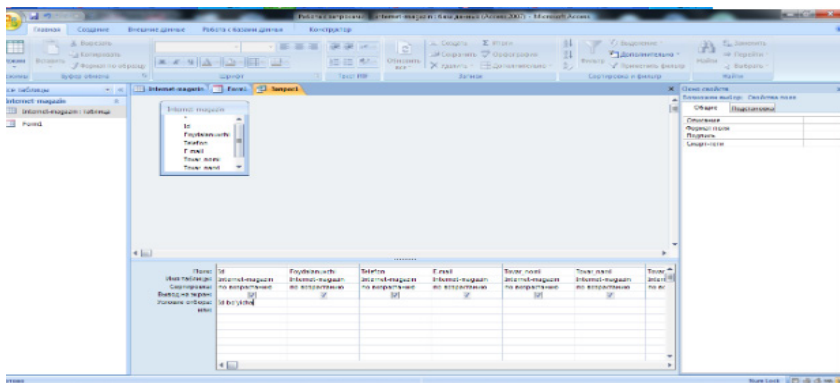


Figure 6. Information search query window

In MB, information can be sorted, words can be arranged alphabetically or numbered. Sorting is done for ease of data search. Typically, the table is sorted by key field value. Sorting can be done on one or more fields. To do this, select the required fields and select the sort condition. Database modeling is done step by step [Fries, 1969].

The process of examination includes the collection of materials, their design in the form of technical specifications. They justify the expediency of creating a bank and a database. The following factors have been identified and cited as key factors:

- commonly used information;
- providing users with interactive data access;
- the existence of complex connections between data;
- the need to update the system.

Materials containing conclusions and proposals for the creation of a bank and database based on certain conditions and capabilities are included in the feasibility study of the project, as well as they form the basis for the formation of technical conditions for the development of the database system.

*The system.* It defines the goals and scope of problems to be solved, the scale and scope of the system, the global constraints.

At the technical design stage, development results and design decisions are formalized in the form of a technical project. It covers common issues: defining the configuration of computing tools, creating a logical database model, updating and configuring it in the form of other level models, selecting the operating system and database, and phys-

ical design. Then special programs are developed for the user database, the submodels available for each user are identified.

A technical design is a basic design document that provides development and descriptions for all components of a created database. Database modeling uses a variety of methods and tools to select a particular database. This includes the initial basic changes of data preparation and working with it, the identification of technological features for all processes associated with the creation and implementation of the database. The technical project reflects the organizational changes associated with the operation of the hardware and software with the organization of new information [Leech, 1991].

Technical design solutions are presented in more detail at the design stage and are described in detail. The working draft has a technically similar structure, but is clarified by in-depth study and verification. At this stage, the collection and pre-preparation of normative references, the development of official and technological guidelines for working with new information technologies will be carried out.

The purpose of this paper is to review the research in the field of linguistic database and to see the possibilities of using this technology in lexicographic projects. It is also possible to present a variant of such a project in the form of a lexicographic database that reflects the vocabulary of the Uzbek language with sound semantics. Database technology is used in the process of creating traditional and electronic dictionaries. Dictionary bases of special and terminological dictionaries are being actively developed.

## **5. Conclusion**

In short, Corpus linguistics is the most advanced branch of linguistics, and the corpus is a necessary tool for linguists; oral, written monuments are a source of information reflecting the national-cultural heritage. The corpus is a collection of texts subject to a search program, and a well-defined corpus serves as a stable linguistic base in ensuring the effectiveness of linguistic research. As a product of artificial intelligence, the linguistic corpus includes an electronic dictionary, a translation portal, a terminological database, a virtual (electronic) library, e-government, e-publishing, e-textbooks and manuals. The general view of the Uzbek national corpus is divided into several windows and right and left columns. It will have the following windows: "Lexical search", "Morphological search", "Syntactic search". Words

and phrases from it are automatically analyzed in a matter of seconds. Linguistic and extralinguistic markings are created in a single format of data expression in the Uzbek national corpus, as well as in the world language corporations. Reconsideration of the theoretical foundations of morphological and syntactic markup based on academic grammar, practical work related to the reduction of the system of semantic markup tags will be carried out. The importance of the socket in the case is incomparable, because the width or narrowness of access to the case depends on the socket of the case. Perfect layout is a guarantee of a wide range of options, universal housing.

#### REFERENCES

1. Belyaeva, L.I., Chizhakovsky, V.A., (1983). Thesaurus in automatic text processing systems., Chisinau.
2. Bloomfield, L., (1968). Language. Moscow, "Progress".
3. Bongers, H., (1947). The history and principles of Vocabulary control, Woerden: WOCOPI.
4. Britvin, V.G., (1983). Applied modeling of syntagmatic semantics of scientific and technical text (by the example of automatic indexing), Moscow State University.
5. Charlez, Meyer, (2004). English corpus linguistics: An introduction. *Cambridge University Press*, UK, 168 p.
6. Eshmuminov, A., (2019). Synonymous database of the Uzbek language national corpus. *Dissertation of PhD in Philology*, Tashkent.
7. Fries, Ch.C., (1969). The structure of English. *An introduction to the construction of English sentences*, London.
8. Hamroeva, Sh., (2018). Linguistic bases of creation of the author's corpus of the Uzbek language: *Author's Abstract of the Dissertation of PhD in Philology*, Tashkent.
9. Karimov, R., (2021). Linguistics and programming issues of creating a parallel corpus of Uzbek and English, *Author's Abstract of dissertation of PhD*, Bukhoro, 151 p.
10. Kholiyorov, O., (2021). Linguistic bases of formation of educational corpus of Uzbek language. *Author's Abstract of the Dissertation of PhD in Philology*, Termiz.
11. Kotov, R.G., (1977). Linguistic aspects of automated control systems. Moscow, Nauka.
12. Kutuzov, A.B., (2017). Corpus linguistics. Retrieved from: <http://lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf> ;
13. Leech, G., (1991). The State of Art in Corpus Linguistics, *English Corpus Linguistics*, London.

14. Melchuk, I.A., (1985). Word order in the automatic synthesis of the Russian word (preliminary messages), *Scientific and technical information*, 12:12-36.

15. Mengliev, B., (2018). Is the Uzbek language corpus being created? *Ma'rifat newspaper*. April 3, 2018, retrieved from: [http://marifat.uz/marifat/v\\_pomosh\\_uchitelu-marifat/savol/1142.htm](http://marifat.uz/marifat/v_pomosh_uchitelu-marifat/savol/1142.htm).

16. Mengliev, B., Bobojonov, S., Hamroeva Sh., (2018). Uzbek National Corpus. April 26, 2018, retrieved from: <http://marifat.uz/marifat/ruknlar/fan/1241.htm>.

17. Mohamed Zakaria Kurdi, (2016). Natural Language Processing and Computational Linguistics: *Speech, Morphology and Syntax*, Great Britain, USA: Wiley-ISTE, 300 p.

18. Nedoshivina, E.V., (2006). Programs for working with text corpora: an overview of the main corpus managers. *Study guide*, St. Petersburg, 26 p.

19. Plungyan, V., (2005). Why are we making the National Corpus of the Russian language? [Electronic resource], *Notes of the Fatherland*, 2:20, retrieved from: [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html).

20. Pulatov, A. Q., (2011). Computer Linguistics. Tashkent, Akademnashr, 520 p.

21. Rykov, V.V., (2005). A course of lectures on corpus linguistics. URL: <http://rykov-cl.narod.ru/c.html>.

22. Shomsku, N., (1962). The logical basis for linguistic theory, *Proceedings of the IX International Congress of Linguists*.

23. Sinclair, D., (2004). How to use corpora in teaching a foreign language, *Preface to the book*, Studies in Corpus Linguistics, 12, VIII, 308 pp. retrieved from: <http://www.ruscorpora.ru/corpora-info.html>.

24. Toirova, G., (2019). The Role of Setting in Linguistic Modeling. *International Multilingual Journal of Science and Technology*, 4(9): 722-723, available at: <http://imjst.org/index.php/vol-4-issue-9-september-2019/>.

25. Toirova, G., (2020). About the technological process of creating a national corpus. *Foreign languages in Uzbekistan*, 2(31):57-64, available at: <https://journal.fledu.uz/uz/2-31-2020>.

26. Toirova, G., (2019). Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019. – P. 352–355.

27. Zakharov, V.P., (2011). Corpus linguistics: a textbook for students of humanitarian universities, Irkutsk, 161 p.

## УДК 81'322.2

UZBECORPORA.UZ: СОЗДАНИЕ КОНКОРДАНСА  
И ЕГО АНАЛИЗ*А. Б. Каршиев<sup>1</sup>, С. А. Каримов<sup>2</sup>, М. С. Турсунов<sup>1</sup>**<sup>1</sup>Самаркандский филиал Ташкентского университета  
информационных технологий имени Мухаммада аль-Хорезми  
Самарканд, Узбекистан**<sup>2</sup>Самаркандский государственный университет  
Самарканд, Узбекистан**abduvalikarshiyev@gmail.com, suyun1950@rambler.ru,  
muhammadsolih927@gmail.com*

Языковые корпуса содержат тысячи текстов, и мы используем поисковую систему, чтобы найти нужное слово и получить его лингвистические характеристики. Как и другие корпорации, эта система действует в *uzbekcorpora.uz*. В результате работы поисковой системы представляется не только грамматическая структура слова, но и совокупность предложений, в которые входит слово – конкорданс.

При строительстве корпуса используются имеющиеся ресурсы и имеющиеся возможности. Разработана корпусная система управления сбором, хранением, обработкой и мониторингом данных. Преимущество программного обеспечения имеет важное значение на всех этапах разработки проекта и управления данными. Эпос «Алпомиш» используется как предварительный результат при создании корпуса узбекского языка.

**Ключевые слова:** корпус, тексты, согласованность, разметка, грамматика, контекст, поисковая система, метаразметка.

UZBECORPORA.UZ: CREATING A CONCORDANCE AND  
ANALYZING IT*A. Karshiev<sup>1</sup>, S. Karimov<sup>2</sup>, M. Tursunov<sup>1</sup>**<sup>1</sup>Samarkand branch of Tashkent University of Information  
Technologies named after Muhammad al-Khwarezmi  
Samarkand, Uzbekistan**<sup>2</sup>Samarkand State University, Samarkand, Uzbekistan  
abduvalikarshiyev@gmail.com, suyun1950@rambler.ru,  
muhammadsolih927@gmail.com*

Language corpora contain thousands of texts, and we use a search engine to find the desired word and get its linguistic characteristics. Like other corpora, this system is active in *uzbekcorpora.uz*. As a result of the search system, not only the grammatical structure of the word, but also the set of sentences in which the word is included – the concordance – is presented.

The available resources and available opportunities are used in the construction of the corpus. A corpus management system has been developed for data collection, storage, processing and monitoring. The advantage of software is essential in all phases of case construction and data management. Alpomish epic is used as a preliminary result in the corpus of the Uzbek language being created.

**Keywords:** corpus, texts, concordance, markup, grammar, context, search engine, metamarkup.

## UZBEKCORPORA.UZ: KONKORDANS TUZISH VA UNI TAHLIL QILISH

*Qarshiyev A. B.<sup>1</sup>, Karimov S. A.<sup>2</sup>, Tursunov M. S.<sup>1</sup>*

<sup>1</sup>*Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti Samarqand filiali*

*Samarqand, O'zbekiston*

<sup>2</sup>*Samarqand davlat universiteti, Samarqand, O'zbekiston*  
*abduvalikarshiyev@gmail.com, suyun1950@rambler.ru,*  
*muhammadsolih927@gmail.com*

Til korpuslarida minglab matnlar joylashgan bo'lib, kerakli so'zni topish va uning lingvistik xarakteristikasini olish uchun qidiruv tizimidan foydalanamiz. Boshqa korpuslar singari uzbekcorpora.uzda ham bu tizim faol ishlaydi. Qidiruv tizimi natijasida nafaqat so'zning grammatik razmetkasi, balki o'sha so'z ishtirok etgan jumlarlar to'plami – konkordansi ham taqdim etiladi.

Korpusni qurishda mavjud bo'lgan resurslardan va mavjud imkoniyatlardan foydalaniladi. Ma'lumotlarni to'plash, saqlash, qayta ishlash va monitoring qilish uchun korpus boshqaruv tizimi ishlab chiqilgan. Korpus qurilishi va ma'lumotlar boshqaruvining barcha bosqichlarida dasturiy ta'minotning afzalligi juda muhimdir. Yaratilayotgan O'zbek tili korpusida dastlabki natija sifatida Alpomish dostoni qo'llaniladi.

**Kalit so'zlar:** korpus, matnlar, konkordans, razmetka, grammatika, kontekst, qidiruv tizimi, metarazmetka.

Til korpuslarida minglab matnlar joylashgan bo'lib, kerakli so'zni topish va uning lingvistik xarakteristikasini olish uchun qidiruv tizimidan foydalanamiz. Boshqa korpuslar singari uzbekcorpora.uzda ham bu tizim faol ishlaydi. Qidiruv tizimi natijasida nafaqat so'zning grammatik razmetkasi, balki o'sha so'z ishtirok etgan jumlarlar to'plami – konkordansi ham taqdim etiladi.

Korpusni qurishda mavjud bo'lgan resurslardan va mavjud imkoniyatlardan foydalaniladi. Ma'lumotlarni to'plash, saqlash, qayta ishlash va monitoring qilish uchun korpus boshqaruv tizimi

ishlab chiqilgan. Korpus qurilishi va ma'lumotlar boshqaruvining barcha bosqichlarida dasturiy ta'minotning afzalligi juda muhimdir. Yaratilayotgan O'zbek tili korpusida dastlabki natija sifatida Alpomish dostoni qo'llaniladi.

Muvozanatlashgan umumiy korpusda turli janrlardagi matnlarni o'z ichiga olishi va har bir janr uchun matn qismlari korpusga qo'shilish uchun mutanosib ravishda tanlanadi. Korpus yaratilayotganda mavjud korpuslar modeli olinadi. Rossiya korpusi, Britaniya korpusi, Amerika korpusi, Koreya korpusi va Polsha korpusi muvozanatlashgan korpus sifatida qaraladi. O'zbek tili korpusini qurishda ko'proq Rus tili korpusi kuzatilgan [Qarshiev, 2021].

“Konkordans – matnni o'rganishning an'anaviy, uzoq vaqtdan beri ma'lum bo'lgan, ammo hali ham matnni o'rganishning yetarlicha o'rganilmagan usuli. U bevosita va kengaytirilgan kontekstdagi so'zlarning to'liq indeksini beradi” [Glushakov, 2005].

Konkordans – korpusga kiritilgan matnlardagi o'rganilayotgan so'zning konteksti va ularning ro'yxati, u o'rganilayotgan so'zni uni o'rab turgan boshqa so'zlar bilan taqdim etadi. Korpus tilshunosligida o'rganilayotgan so'z odatda “*kontekstdagi kalit so'zlar*” deb tushuniladi. Uzbekcorpora.uz tizimida konkordans tuzish uchun “Tilshunoslik tadqiqotlari” menyusidan “konkordans” bo'limiga kiriladi (1-rasm).



**1- rasm.** Konkordans interfeysi

Bunda o'rganish uchun kalit so'z kiritiladi va “izlash” tugmasi bosiladi va konkordans hosil bo'ladi.

Konkordanslar quyidagi lingvistik masalalarni tahlil etishda kerak bo'ladi:



- soʻzlar roʻyxatini yaratish;
- soʻz va soʻz-shakllarni qidirish;
- soʻz va soʻz-shakllarni solishtirish;
- kalit soʻz va uning atrofidagi soʻzlarning grammatik tahlili;
- tanlangan soʻz va soʻz-shakllarning chastotasini aniqlash;
- oʻrganilayotgan soʻzning kerakli matndan tezda topish.

Konkordansning qidiruv funksiyasida iqtiboslarni tanlash, tekshirish, asl matn bilan solishtirish mumkin. Konkordansning lugʻatlardan farqi shundaki, unda qidirilayotgan soʻzning kontekstlari ham taqdim etiladi. Konkordansda odatdagi lugʻatlardan koʻra koʻproq maʼlumotlar saqlanadi, unda matndagi paradigmatic bogʻlanishlarni, ijodkorning soʻz qoʻllash mahorati, unga xos boʻlgan sintaktik konstruksiyalarni ajratib olish imkoniyati bor. Konkordansdan tilshunos ham, adabiyotshunos ham birday foydalanishlari mumkin. Shu bilan birga konkordans keyingi paytda shakllangan til korpuslarining asosiy va muhim qismiga aylangan [Tursunov, 2022].

Koʻpincha maʼlum bir soʻzning qidiruvi korpusdagi muvofiqligi tilshunosning tahlil qilishi uchun juda koʻp natijalarga olib keladi. Bizga maʼlumki, til korpuslari orqali oʻsha tilni oʻrganish mumkin, til oʻrganuvchi uchun soʻzning asl va koʻchma maʼnosi, ishlatilishi, uning etimologiyasi, qoʻllanilishi holati, hatto grammatik mazmuni ham qiziq, shu nuqtai nazardan konkordans turli toifadagi foydalanuvchilar uchun juda qoʻl keladi.

Tognini-Bonelli korpusga asoslangan tilshunoslikni quyidagicha taʼriflaydi: “Konkordans – foydalanuvchiga korpusda nima sodir boʻlishini tekshirishga, matnlarda qanday maʼno yaratilganligini, soʻzlarning qanday paydo boʻlishini va mazmunli belgilar bilan birlashtirilganligini, bu birliklarning nima ekanligi toʻgʻrisida aniq tasavvurga ega boʻlmagan holda koʻrish imkoniyatini beradi. Bu korpusga nazariy jihatdan neytral tarzda murojaat qilish usuli boʻlishi mumkin” [Пиотровский, 1988].

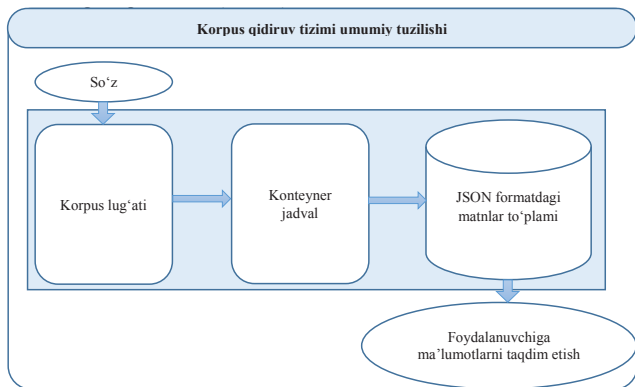
Korpusdan soʻzni qidirganimizda qanday qilib, qisqa vaqt ichida konkordans tuzish mumkin? Oʻzbek tili milliy korpusiga matnli fayllar kiritiladi. Bir nechta fayl bitta matn boʻlishi mumkin yaʼni bitta matn boʻlaklarga boʻlib korpusga kiritish imkoniyati mavjud. Bunda matn nomiga va fayllarga ID raqamlar beriladi va fayllarga matn nomining ID raqamlari birlashtiriladi. Bunda qidiruv jarayonida bir nechta faylni bitta matnga tegishli deb bilish imkonini beradi.

Oʻzbek tili milliy korpusida biror soʻz qidirilganda qidiruv tizimi natijasiga koʻra foydalanuvchilarga konkordans tuzib beradi.

Konkordans ro'yxati o'rganilayotgan so'z ishtirok etgan matn nomlari va kontekstlardan iborat bo'ladi.

O'zbek tili milliy korpusida qidiruv tizimi qanday ishlaydi? Korpusda so'zlar ro'yxati, matnlarning *json* formatdagi fayllari hamda lug'at va fayllarning o'rtasida konteyner vazifasini bajaruvchi jadvaldan iborat (2-rasm). *json* formatdagi matn faylda har bir so'zning grammatik ma'lumotlari va matnda joylashgan o'rni ya'ni satr va ustun raqamlari so'zga birlashtirilgan bo'ladi. Konteyner jadval maydonlari so'zning IDsi, *json* faylning IDsi, so'zning matnda joylashgan o'rni ya'ni so'z joylashgan ustun va satr raqamlaridan iborat bo'ladi.

Faraz qilaylik, korpusda minglab fayllar mavjud. Biz o'rganayotgan so'zni qisqa vaqt ichida fayllardan topib, foydalanuvchiga ma'lumotlarni taqdim etishi lozim. Agar so'z har bir matnli fayldan qidirilsa va har bir matnli faylning har bir so'zi bilan taqqoslanib chiqilsa, bu dasturdan juda katta yuklanish va natijaviyligi uzoq vaqt talab etishi mumkin. Biz so'zni fayllardan qidirishda boshqa yondashuv bilan algoritm ishlab chiqdik. Qidiruv tizimi yordamida so'z qidirilganda, dastavval so'zni korpus lug'atidan izlaydi, agar so'z mavjud bo'lsa, so'zning IDsi konteyner jadvaldan topiladi. Konteyner jadvaldan so'zning IDsi topilgan satrda shu so'z ishtirok etgan matnlarning ID raqamlari va matn ID raqamlarga mos so'zning ustun va satr raqamlarini oladi va matn IDsiga mos raqamli *json* faylning ichiga kirib, kerakli ustun va satrga boradi hamda so'zning grammatik ma'lumotlarini olib, foydalanuvchiga taqdim etadi (2-rasm).



2-rasm. Korpus qidiruv tizimi umumiy tuzilishi

Ishlab chiqilgan algoritmning afzalliklari:

– o‘rganilayotgan so‘zning korpusdagi matnli fayllarning aniq qaysi faylida joylashganini aniqlaydi;

– o‘rganilayotgan so‘zning mavjud matnli fayldagi o‘rniga to‘g‘ridan to‘g‘ri murojaat qiladi.

O‘zbek tili milliy korpusida qidiruv natijalariga kirishni ta‘minlash alohida vazifa bo‘lib, sayt yetarlicha ma‘lumotga ega bo‘lishi kerak, ammo bir vaqtning o‘zida keraksiz ma‘lumotlar bilan ortiqcha yuklanmasligi lozim. Ushbu funksiya tadqiqot uchun murojaat qilgan mutaxassis filologlar uchun mo‘ljallangan. Foydalanuvchiga axborot berishning umumiy tamoyillari [5] o‘rganildi va ular asosida dastlabki qidiruv tizimi loyihasi va dasturiy ta‘minoti tayyorlandi. Qidiruv tizimidan foydalanish tushunarli bo‘lishi uchun sahifalar imkon qadar sodda ko‘rinishga keltirilgan. Shu sababli qidiruv sahifasining asosiy qismida ortiqcha ma‘lumotlar qo‘yilmagan. Bu ish maydonini vizual ravishda tengaytirish imkonini berdi. Filologlar tomonidan ilgari surilgan talablardan kelib chiqqan holda, yaratilgan axborot resursi foydalanuvchiga ma‘lumot olishning bir nechta variantlarini taqdim etadi: 1. Qidirilayotgan so‘z ishtirok etgan gapni kontekst sifatida olish. 2. Qidirilayotgan so‘zni old tomonidan va orqa tomonidan olinishi kerak bo‘lgan so‘zlar miqdorini ko‘rsatgan holda kontekstni olish. Kontekst – tanlangan so‘zni atrofini o‘rab turgan so‘zlar bilan birgalikda olish.

*Texnik talablar.* Korpusni tahlil qilish vositalari korpusdagi matnlarni json formatdagi fayllar to‘plami sifatida qidiradi yoki matnlar oldindan indekslangan bo‘lishi mumkin, bu esa tezroq qidirish va yanada kuchli so‘rovlarga imkon beradi. Belgilar va matnlarni kodlashning alohida shakllari, fayl formati va ma‘lumotlari dastur tomonidan oqilona talqin qilinishini ta‘minlash uchun ehtiyot bo‘lish kerak. Agar korpusning o‘zi yetarlicha standart tarzda qurilgan bo‘lsa hamda korpus tuzilishi va kodlanishi yaxshi hujjatlashtirilgan bo‘lsa, bu yanada aniqroq bo‘ladi.

Konkordanslar foydalanuvchiga odatda ekranda ko‘rinadi, ammo ularni qayta ishlash, o‘zgartirish imkoni yo‘q. Konkordans alohida faylga saqlab olinadi va bunga boshqa sabablar ham mavjud:

– korpusga yoki konkordansga kirish vaqtinchalik bo‘lishi mumkin;

– korpus rivojlanish bosqichida bo‘lishi va o‘zgarishi mumkin;

– vositalar yangilanishi va ularning funksiyalarini nozik usullar bilan o‘zgartirishi mumkin.

Bundan tashqari, konkordansni keltirib chiqaradigan dasturdan tashqarida hamjihatlikni boshqa vositalar yordamida qayta ishlash

yoki o‘qitishda veb-saytda yoki nashrda foydalanish zarur bo‘ladi. Shuning uchun konkordansni ba’zi bir ko‘chma formatda, masalan, *exsel*da saqlash kerak. Konkordans so‘zdan foydalanishning barcha kontekstlarini solishtirish, ularni tahlil qilish, badiiy asar matnidagi so‘zni ko‘rish imkonini beradi. Bu matnni o‘rganishning eng samarali vositalaridan biridir.

Grammatik xususiyatlar bo‘yicha qidiruvni bajarish uchun, asosan, barcha ma’lumotlar serverda qayta ishlanadi, shundan so‘ng server foydalanuvchining so‘roviga javob yuboradi. Dasturning ishlash funksiyalari *Python* dasturlash tilida yozilgan. Dasturiy ta’minot funksiyalari to‘plami so‘z va so‘z-shaklning grammatik razmetkalarni bir joyga yig‘adi.

*Matndan so‘zni topish.* Yuqorida tavsiflangan funksiyalardan birining harakatlari natijasida so‘rov yaratilgan va serverga yuborilgandan so‘ng, kerakli parametrlar orqali so‘zlarni qidirish sodir bo‘ladi.

Algoritm natijasida olingan mantiqiy ifoda so‘rovga almashtiriladi va qayta ishlashga yuboriladi. Har bir so‘z matn raqami, bob raqami, abzas raqami, jumla raqami va so‘z raqami orqali qidiriladi. Asl matnning manzili serverda aniqlanadi. Jadvallardan matn boshiga nisbatan gap boshining ofset-joylashuvi topiladi. Shundan so‘ng fayl o‘qiladi va keyin ekranda ko‘rsatiladi.

*Kontekstning chiqishi.* Bizda so‘z uchun aniq identifikator mavjud bo‘lganligi sababli, biz izlayotgan so‘zni matndagi joylashuvida o‘rab turgan oldingi va keyin so‘zlardan qanchasini olishni ko‘rsatishimiz yoki so‘z ishtirok etgan gapni olishimiz mumkin. Bunda korpusdagi matnli fayllardan topilgan kontekstlar ro‘yxatini foydalanuvchiga chiqarishda, har bir matndan ko‘pi bilan o‘ntadan kontekstini chiqaradi. Bitta matndagi to‘liq kontekstlar ro‘yxatini ko‘rish uchun ushbu matnning nomi ko‘rsatilgan maydonda kontekstlari soni ham ko‘rsatiladi. Ushbu kontekstlar sonini ustidan sichqonchanning chap tugmasi bosilsa, sahifada matnning kontekstlar ro‘yxati to‘liq taqdim etiladi (3-rasm).

Bunda kalit so‘z qizil rangda va uni o‘rab turgan so‘zlar qora rangda berilgan. “<...>” – belgisi kengaytirilgan kontekstga o‘tish va undan qaytish uchun foydalaniladi. Foni qora rangda bo‘lgan “Alpomish” yozuvi matn nomi hisoblanadi. “Alpomish” so‘zi korpusning tarkibida bitta matndan topilib, 264 marta ishtirok etganini korish mumkin.



3-rasm. Konkordansning hosil bo'lishi (kontekstning chiqishi)

So'zning chiqishi va uning morfologik parametrlari. Yuqorida aytib o'tganimizdek, biz har bir so'zni matndan ma'lumotlar bazasidagi o'xshashiga bog'lash imkoniyatiga egamiz. Bu noyob identifikator yordamida amalga oshiriladi. Foydalanuvchi sichqonchani chap tugmasi bilan ma'lum bir so'zni bosadi, buning natijasida so'zning noyob identifikatorini o'qiydigan va so'zning o'zi hamda uning o'ziga xos xususiyatlari ko'rsatiladigan oynani ochadigan prosedura chaqiriladi. Buning uchun ma'lumotlarni yig'ish bloki va parametrlarni dekodlash bloki ishga tushiriladi. Konkordansdagi har bir so'zga morfologik parametr biriktirilgan va ixtiyoriy so'zni ushbu matnda kelishi mumkin bo'lgan morfologik xususiyatlarini ko'rish mumkin (4-rasm).



4-rasm. So'zning morfologik xususiyatlari

Bunda foydalanuvchiga so'zning turkumi, lemmasi va razmetkasi haqidagi ma'lumotlar taqdim etiladi.

Belgilangan aniq identifikator tomonidan ma'lumotlarni olish bloki so'zni, uning shifrlangan xususiyatlarini topadi, shundan so'ng shifrni

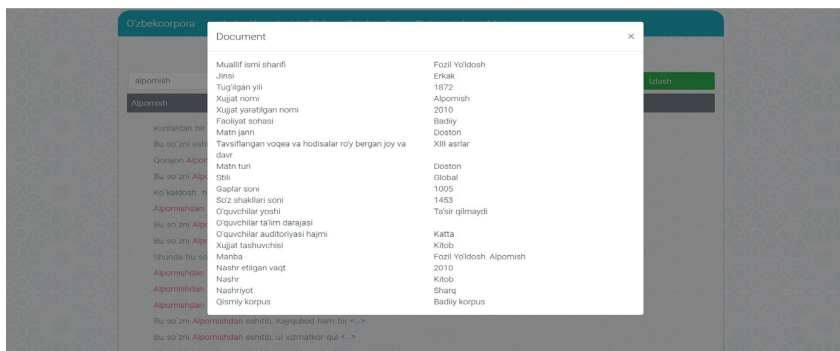
ochish bloki deb nomlanuvchi ikkita rekursiv funksiyadan iborat funksiya ishga tushadi. Birinchisi berilgan kod bo'yicha parametr nomini topadi, ikkinchisi uning qiymatini topadi. Shuningdek, ikkinchi funksiya qolgan parametrlarning kodlarini aniqlaydi, shundan so'ng u parametrni chiqish funksiyasini chaqiradi.

*Kengaytirilgan kontekst chiqishi.* Topilgan kontekstlarni ko'rsatgandan so'ng dastur foydalanuvchiga kengaytirilgan kontekstga o'tishni taklif qiladi. Kengaytirilgan kontekst sifatida kontekst ishtirok etgan abzasni chiqarib ko'rsatadi. Kengaytirilgan kontekstda ham har bir so'zning parametrlarini ko'rish imkoniyati mavjud (5-rasm).



5-rasm. Kengaytirilgan kontekst

*Metarazmetka.* Kalit so'z topilgan matn to'g'risida ma'lumo taqdim etadi. Buning uchun matnning nomi ustidan sichqonchanning chap tugmasi bosiladi (6-rasm).



6-rasm. metarazmetka

Shunday qilib, konkordans asosan ma'lumotlarni vizuallashtirish usuli hisoblanadi. O'rganilayotgan so'z va uning atrofidagi so'zlarni matnda kelgan xususiyatlarini va matnning muhitini baholash hamda matnlarda so'zlarning xususiyatlarini o'zgarishini ko'rish imkonini beradi. Bu esa foydaluvchiga matnga nisbatan turli xil xususiyatlarni ko'rish imkoniyatini beradi.

#### ADABIYOT

1. A.B. Qarshiyev, S.A. Karimov, M.S. Tursunov, O'zbek tili korpusining dasturiy ta'minotini yaratishning dastlabki natijalari // "Al-Xorazmiy avlodlari" jurnali, 2021 yil, 1-soni.
2. Glushakov, S.V. Программирование Web-страниц / S.V.Glushakov, I.A. Jakin, T.S. Xachirov. – Xarkov: Folio, 2005. – 390 s.
3. M.S.Tursunov, O'zbek tili milliy korpusini yaratishda dastlabki ma'lumotlar // Ilmiy axborotnoma, SamDU, 2022-yil 6-son (136), 82–88 s.
4. Пиотровский Р.Г. Компьютеризация преподавания языков (учебное пособие по спецкурсу). – Л.: 1988. С. 75.
5. Fedorchuk, A. Как создаются Web-сайты. Краткий курс / A.Fedorchuk – SPb.:Piter, 2000. – 224 s.

UDK: 81`1:004=512.133

**РЕФЛЕКСИЯ В ЯЗЫКОВОМ КОРПУСЕ  
АНТРОПОНИМИЧЕСКИХ ЕДИНИЦ***Г. И. Тоирова**Бухарский государственный университет**Бухара, Узбекистан**tugulijon@mail.ru*

В статье говорится о том, как антропонимы выражают в своей семантике экстралингвистическую информацию и составляют неотъемлемую часть структуры знаний людей определенного языка и культуры, как зеркало отражающую исторические, религиозные, мифологические события, обычаи и традиции этого народа. Отмечается, что имена закрепляются в их составе с течением времени и концентрируются национальные особенности, имеющие социальное значение для данного общества. Репрезентация антропонимических единиц в языковых корпусах показывает, что лингвистика вышла на новый уровень.

**Ключевые слова:** национальный корпус, корпусная лингвистика, лексикографическая основа, антропонимы, интернет-язык, база данных, языковые единицы.

**REFLECTION IN LANGUAGE CORPORA  
OF ANTHROPNYMIC UNITS***Toirova G. I.**Bukhara State University, Bukhara, Uzbekistan**tugulijon@mail.ru*

The article talks about how anthroponyms express extralinguistic information in their semantics and form an integral part of the knowledge structure of people of a certain language and culture that reflects the historical, religious, mythological events, customs and traditions of this nation like a mirror. It is noted that the names are fixed in their composition over time and national features of social importance for that society are concentrated. Representation of anthroponymic units in language corpora shows that linguistics has reached a new level.

**Keywords:** National corpus, corpus linguistics, lexicographic basis, anthroponyms, internet language, database, language units

Language is not only a means of communication between people, but also a symbol of national identity and national pride. In order to preserve the language, corpora have already been developed to ensure its safety, and this work is ongoing.

The national corpus is a collection of national language units collected in the order of reanalysis.



In the National Corpus of the Uzbek language, explanatory dictionaries also serve as the lexicographic basis. However, a number of shortcomings in this type of dictionaries have not yet been corrected and cause difficulties in creating a corpus. Because the lexicographic basis ensures the perfection of the corpus. The first field to work on the basis of the corpus is lexicography, which is the main, unique source for the compilation of voluminous dictionaries. All modern, latest dictionaries are based on the corpus, and they are evaluated by the authenticity and credibility of their examples. Because the language in the corpus reflects how it lives in society, as a result, the example in the dictionary is convincing and reasonable. [Volosnova, 2006].

Another period of corpus linguistics before the computer age is the 18th-19th centuries, which is associated with the development of lexicography and the creation of dictionaries. Most of the authors of today's popular numbered dictionaries have essentially built an illustrative corpus based on thousands of numbered index cards.

The vocabulary of any (English, Russian, Uzbek...) language will be uniquely rich. That is why native speakers have the opportunity to freely choose words in their speech. If we consider the vocabulary of linguistics as a large system, it contains lexical systems:

- system of common words;
- system of dialectal words;
- system of words related to profession - trade;
- system of social class slangs;
- the system of proper nouns.

It can be seen that adjectives also have their place in the vocabulary system of the language, which in turn forms a large system. For example, anthroponyms (names of people), toponyms (names of places), zoonyms (animal names), phytonyms (plant names), theonyms (religious names), astronoms (universal scientists), documentonyms (document names), chrononyms (historical events) names) etc. Each of the mentioned nouns has its own systemic relationship, which is reflected in their internal systems.

Each nation relied on the vocabulary and capabilities of its native language when choosing a name. A noun is made up of words that exist in the language. But the vocabulary of any language does not consist only of its own. Clans, tribes, peoples, peoples who have lived as neighbors in a certain area for a long time interacted with each other in different ways. Political-economic, cultural-educational, religious ethnic relations have also affected the languages of these peoples.

It was common for nouns to enter and assimilate words from one language to another. The names of people are the ancient names of our people, in which the way of life and thinking of our ancestors, that is, our ancestors, clearly reflected.

Therefore, the name chosen as an anthroponym has its illocutionary character. For example, there is no name without a motive. When choosing any name, we can observe that life, way of thinking is taken as a motive. For example, some people who do not know the specific laws of the appearance of names are surprised to hear the names Bo'riboy, Borioi, Kochkor, Topiboldi, Sotiboldi, Boltaboy, Bolgaboy in the Uzbek language, even condemns them as „inappropriate, old name“. Because these names contain an implicit expression. However, all these names are based on the illocutionary meaning of wishing and hoping that the child will grow up healthy and not die prematurely. It is known that the reason for the emergence of any name, including a name, is a motive in basic science, the field of scientific research of such motives is called motivation, and the naming of things and events based on specific motives is called a nomination. The fact that the name has a specific motive is called motivation („motivirovannost“, „motivation“), the sign, situation, concept that is the basis for the motive is called „motivator“, „motiveme“. [Naumov, 1985].

As can be seen from the above, the corpus can cover all areas of the language. Currently, the corpus is divided into educational and non-educational parts according to the fields of application, and relevant information is collected. The database of Anthroponyms (personal names) that we offer is intended for use as a non-educational corpus. In this E.A. „Explanation of Uzbek names“ by Begmatov dictionary serves as a lexicographic basis.[Begmatov, 2016]

Today, the approach to language research, especially national language research, is gaining practical importance. The scientific and practical nature of providing anthroponyms from linguistic units in the national corpus is shown in the following:

- 1) as a result of the reflection of various language units and idioms in language corpora in the modern era of developing information technologies, linguists, translators, teachers, journalists and researchers in various fields can use their time effectively, i.e. find and analyze the desired unit quickly and in a short period of time gives;

- 2) learning and teaching with the help of communication and informational tools is becoming more and more improved as a requirement of the time. This makes it possible to directly use the

information contained in the national corpus or language corpus in educational processes;

3) Taking into account that the Uzbek language is taught as a science in several countries of the world, by using the database of anthroponyms, it becomes possible for foreign interested parties to get acquainted with the nationally specific units of our language, and to apply them practically in all areas related to the computerization of linguistics.

In conclusion, it can be said that the representation of anthroponymic units in language corpora showing the national-cultural code of the language shows that linguistics has risen to a new level. It is impossible not to influence linguistics, like all other fields, with current technical and technological innovations. As a result of this, electronic dictionaries, educational platforms, and corpora of language data were created based on speed and accuracy. The operation of these collections is a reliable guarantee that the national fund of national anthroponyms – people's names - will be passed down from generation to generation.

#### REFERENCES

1. Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида). Филол.фан.бўйича фалсафа доктори (PhD)...дис. автореф. – Тошкент, 2018
2. Бегматов Е.А. Ўзбек исмлари изохи. –Тошкент., 2016, – 608 б.
3. Волоснова Ю.А. Корпусная лингвистика: проблемы и перспективы// Лесной вестник №7, 2006, С. 43.
4. Горбунова Е. А. Лингвокультурный комментарий прецедентных феноменов в англоязычном художественном тексте: Автореф. дис. канд. филол. наук. Самара, 2008, – Б. 3.
5. Наумов В.Ген. Явление мотивации слов в системе диалекта (лексический аспект). – Томск, 1985. – С. 23.
6. Тоирова Г. Ўзбек тили миллий корпусни яратишда интерфейснинг аҳамияти. // Сўз санъати халқаро журнали, – Тошкент, 2020, № 3, – Б. 100–105.
7. Тоирова Г. Лингвистик базани тузишда модуллаштиришнинг аҳамияти // Наманган давлат университети илмий ахборотномаси. – Наманган, 2021. – №3. – Б. 377–386.
8. Эшмўминов А. А. Ўзбек тили миллий корпусининг синонимлар базаси. Филол. фан. бўйича фалсафа доктори (PhD) дис.автореферати. – Қарши, 2019.

9. Ҳамроева СҲ. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Филол.фан.бўйича фалсафа доктори (PhD) ... дис. афтореф. – Бухоро, 2018

10. Ўзбек тили корпуси яратилиаяптими?/ Бахтиёр Менглиев. [хтпс://www/хабар.уз.>таълим](https://www.хабар.уз.>таълим)

11. Nikonov V.A. Die Periodisierung der russischen Anthroponymie von den Anfängen bis 1917 (vorläufiges Schema). // Sowjetische Namenforschung. - Berlin: Akademie-Verlag, 1975. – S. 103–115.

8. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019. – P. 352–355. (scopus)

9. Toirova G, Abdurahmonova N., Ismoilov A., Applying Web Srawler Teshnologies for Sompiling Parallel Sorpora as one Stage of Natural Language Prossessing. 2022 7th International Conference on Computer Science and Engineering (UBMK) Sep. 14–16, 2022, Diyarbakir /Turkey pp. 73–75. (scopus)

10. Toirova G. The importance of linguistic models in the development of language bases. // Buxoro Davlat universiteti ilmiy axboroti. – Buxoro, 2020. – № 6. – В. 98–106.

---

## ЛИНГВИСТИЧЕСКИЕ РЕСУРСЫ

УДК 81'272

### БАЗА СОЦИОЛИНГВИСТИЧЕСКИХ И ЯЗЫКОВЫХ ДАННЫХ ПО ТЮРКСКИМ ЭТНОСАМ РЕСПУБЛИКИ КАЗАХСТАН

*И. А. Невская*

*Институт филологии СО РАН, Новосибирск, Россия*

*Франкфуртский университет им. Й. Гете*

*Франкфурт, Германия*

*nevskayairina60@gmail.com*

В статье рассматривается база данных, представляющая результаты социолингвистического исследования тюркских этносов Казахстана и языковой документации неописанных и исчезающих тюркских идиомов в Казахстане, проведенных в период 2014-2019 г.г. в разных регионах республики в рамках проекта международного сотрудничества «Взаимодействие тюркских языков в культурах постсоветского Казахстана».

Социолингвистический опрос предоставил информацию об истории семей информантов, их самоидентификации, языковых предпочтениях в семье и в повседневном общении вне семьи, а также в ряде других аспектов. Языковая документация велась в основном по отношению языковых идиомов турок Казахстана. На них говорят так называемые турки-месхетинцы, хотя сама эта этническая группа этим именем себя не называет. Существуют различные тюркские субэтнические группы в Казахстане (ахыска, терекеме, хемшиллы и др.), языки которых ранее не были документированы.

База данных хранится и визуализируется на главной странице проекта: <https://tyurki.weebly.com>.

**Ключевые слова:** документирование языков, исчезающие варианты турецкого языка, создание базы социолингвистических данных, корпус текстов

### A DATABASE OF SOCIOLINGUISTIC AND LANGUAGE DATA ON TURKIC ETHNIC GROUPS IN THE REPUBLIC OF KAZAKHSTAN

*Irina Nevskaya*

*Institute of Philology, Siberian Division of Russian Academy*

*of Sciences, Novosibirsk, Russia*

*Johann-Wolfgang-Goethe University*

*Frankfurt, Germany,*

*nevskayairina60@gmail.com*

The article deals with a database presenting the results of a sociolinguistic survey of Turkic ethnic groups and a language documentation of undescribed and endangered varieties in Kazakhstan, conducted in the period of 2014-2019 in different regions of Kazakhstan in the framework of the international cooperation project “Interaction of Turkic languages in cultures in post-Soviet Kazakhstan”.

The sociolinguistic survey provided information on the history of informants’ families, their self-identification, language preferences in the family and in everyday communication outside the family, among further issues. The language documentation was done mostly for Turkish linguistic varieties spoken in Kazakhstan. They are spoken by the so-called Meskhetian Turks, although the people do not call themselves by that name. There are various Turkish sub-ethnic groups (Ahiska, Terekeme, Hemshilli, etc.), whose languages have not been documented earlier.

The database is stored and visualized on the homepage of the project: <https://tyurki.weebly.com>.

**Key words:** language documentation, Turkish endangered variants, sociolinguistic database creation, text corpus

## 1. Введение

В 2014-2019 г.г. группой исследователей ЕНУ им. Л.Н. Гумилева, ТарГУ им. М.Х. Дулати, Франкфуртского университета им. Й.В. Гете и Берлинского Свободного университета проводился международный научный проект “Interaction of Turkic Languages and Cultures in post-Soviet Kazakhstan”, финансирующийся Фондом Фольксвагена. Руководителями проекта были И.А. Невская, К. Шёнинг, С.Ж. Тажибаева, Н.Г. Шаймердинова. Проект был нацелен на проведение социолингвистического исследования тюркских языков современного Казахстана, документацию ранее не фиксировавшихся или исчезающих тюркских идиомов, в первую очередь идиомов так называемых месхетинских турок, а также на представление результатов проекта в Интернете. Анкетирование среди представителей тюркских этносов и документация турецких идиомов проводились по всем регионам Казахстана силами преподавателей, докторантов, магистрантов, студентов ЕНУ им. Л.Н. Гумилева и ТарГУ им. М.Х. Дулати, а также Франкфуртского Университета. Все ответы респондентов были занесены в базу данных, которая находится на сайте данного проекта: <https://tyurki.weebly.com>. Необходимо особо подчеркнуть, что данные социолингвистического исследования имеют открытый доступ и дают возможность всем заинтересованным лицам использовать результаты проекта для своего исследования.

## 1. База языковых данных

Документирование исчезающих языков – достаточно новое направление в современной лингвистике. В рамках этого направления в последние годы был опубликован ряд монографий и научных статей [Dorian, 2010; Haig и др., 2011; Grenoble и Furbee, 2010; Lehmann, 1983; Gippert и др., 2006].

Языковая документация представляет собой отдельное направление современной лингвистики со своими методами исследования, правилами, подходами и инновационными технологиями. Языковая документация не является описательной и нормативной. В ее задачи не входит создание грамматики, словаря или же стандартных норм языка. Основной целью документации языков является создание многоцелевого корпуса репрезентативных примеров использования языка в различных условиях естественной коммуникации, т.е. запись различных коммуникативных ситуаций в разнообразных социальных и культурных контекстах. В дальнейшем вся информация расшифровывается, аннотируется и комментируется; см. информацию о некоторых проектах по документации языков [Невская, 2005; Fedotov и др., 2015]. Использование естественного языка необходимо документировать для того, чтобы, во-первых, корпус примеров был доступен как для исследователей-филологов например: (фонетистов, лексикологов, литературоведов), так и специалистов разных направлений науки (историков, антропологов, культурологов и т.д.). Во-вторых, корпус репрезентативных примеров должен соответствовать требованиям, предъявляемых к данным этих дисциплин (например, несжатый формат WAV для фонетистов; антропологам предпочтительнее работать с данными видеозаписи, чем аудиозаписи, и т.д.). Корпус репрезентативных примеров собирается во время проведения полевых исследований.

Полевые лингвистические исследования в рамках нашего проекта позволили собрать уникальный языковой материал по исчезающим тюркским языкам Казахстана. Это материалы по крымско-татарскому языку, ногайскому, а также вариантам турецкого, узбекского и других языков. Записи аудио и видео материалов по турецким исчезающим идиомам были архивированы, была сделана транскрипция для отдельных отрывков устной речи.

Все варианты турецкого языка, которые мы обнаруживаем на территории Казахстана (ахыска, хемшиллы, терекеме, лазы), были оторваны от основного языкового массива более ста лет. В тече-

ние семидесяти лет во время существования СССР практически не было их контактов с турецким языком, функционирующим в Турецкой Республике. Сам турецкий язык Республики также претерпел значительные изменения после реформ Атаюрка, которыми варианты турецкого языка Казахстана не были затронуты.

Лингвистическая значимость изучения неописанных вариантов турецкого языка в Казахстане заключается, прежде всего, в сборе, классификации и описании живого языкового материала. Все варианты языка турецких субэтносов Казахстана находятся на грани исчезновения. Положение усугубляется тем, что старое поколение носителей уходит из жизни, а язык молодого поколения казахстанских турок претерпевает сильные изменения, связанные с живыми процессами взаимодействия языков в Казахстане [Nevskaya, Tazhibayeva, 2015a]. Документирование исчезающих вариантов языка турецкой диаспоры Казахстана продолжает оставаться актуальной задачей современной тюркологии и дает возможность лингвистам обратиться к конкретному, новому и никогда ранее не публиковавшемуся языковому материалу.

Корпус собранных текстов доступен мировой общественности на сайте так называемого «Архива языков» в Института имени Макса Планка в Наймегене, Голландия, на который стоит ссылка на сайте проекта<sup>1</sup> с описанием использованных нами программ компьютерной обработки языковых данных и инструкцией по их использованию; даны также выборочно примеры на отдельных идиомах.

## 2. База социолингвистических данных

Создание базы данных социолингвистического исследования по тюркским этносам в Казахстане было выполнено в рамках международного проекта в результате кооперации кафедры тюркологии Евразийского национального университета (ЕНУ) им. Л.Н. Гумилева, которой руководил известный тюрколог Мырзатай Жолдасбеков, и Таразского госуниверситета (ТарГУ) им. М.Х. Дулати. Руководителем социолингвистического опроса в Астане была д.ф.н. проф. Н.Г. Шаймердинова, а в ТарГУ им. М.Х. Дулати д.и.н. проф. А. Абдуалы. Результаты анкетирования обрабатывались и синхронизировались в базе данных, разработанной программистом проекта Жанар Бейсеевой, сотрудницей Кафедры информационных систем ЕНУ, которой руководит проф. Ж. Тусупов. База данных послужила основой для социологического исследо-

<sup>1</sup> <https://tyurki.weebly.com>. См. раздел «О проекте».



вания ряда аспектов этносов Казахстана, например: [Nevskaya, Tazhibayeva, 2015b].

Создание базы данных представляло собой планомерную работу, состоящую из четырех этапов: сбора данных, хранения информации, обработки информации, визуализации данных [Fedotov и др., 2015; Невская и др., 2016; Vatura и др., 2016]. В качестве программной среды для базы данных по международному проекту фонда Фольксвагена было использовано приложение Google Таблицы. Данные базы хранятся в табличной форме в формате EXCELL как это показано на Рисунке 1.

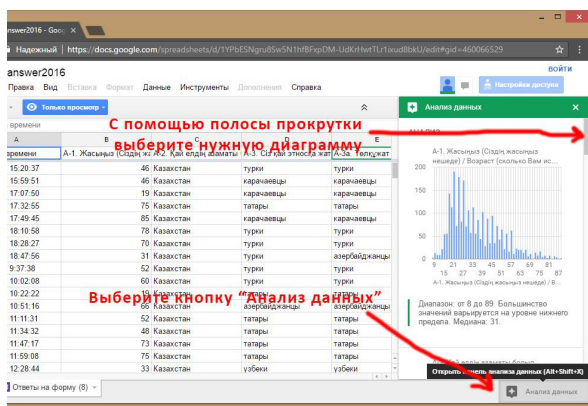


Рисунок 1. Табличная форма хранения ответов

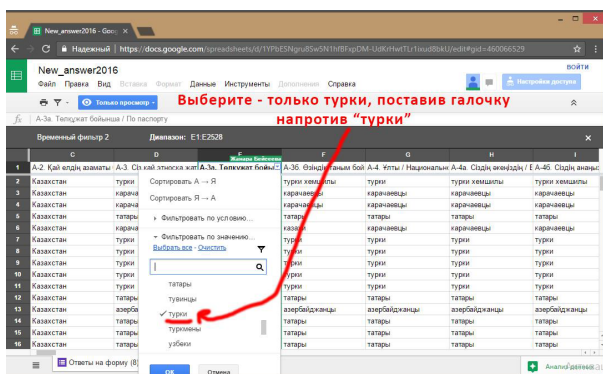


Рисунок 2. Проведение фильтрации

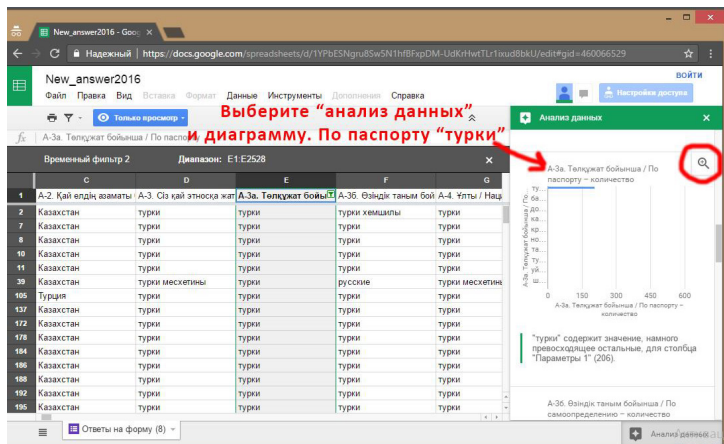


Рисунок 3. Фильтрация ответов респондентов на вопрос, кто из них записан в паспорту турком

В базе данных поиск ответов на запросы исследователя можно осуществить при помощи фильтрации. Это наиболее быстрый и простой способ нахождения подмножества данных и работы с ним в диапазоне ячеек или в столбце таблицы. В отфильтрованных данных отображаются только строки, соответствующие заданным условиям, а ненужные строки скрываются, см. Рисунок 2. При работе с базой данных можно одновременно применять несколько фильтров, при этом каждый следующий фильтр накладывается на результаты предыдущей фильтрации. Рисунок 3 иллюстрирует использование фильтра для выявления количества респондентов, записанных по паспорту турками. В отфильтрованных данных отображаются только строки, соответствующие заданным условиям.

Данные социоопроса были визуализированы в виде диаграмм, графиков, поскольку табличная форма для отображения результатов работы с фильтрами является недостаточно наглядной. Для этих целей была разработана визуализация данных при помощи инструментов Google Chart API, основанных на языке программирования Java Script. Эту работу выполнила программист Ж. Бейсева. На Рисунке 4 представлена оболочка сайта, в котором выполнена визуализация данных проекта.

О проекте    Краткая история    Культура    Заполнить анкету    Результаты опроса

**«Взаимодействие тюркских языков и культур в постсоветском Казахстане»**

Проект предполагает активное вовлечение в научное исследование студентов бакалавриата, магистрантов и докторантов. Студенты высших учебных заведений, а также волонтеры из числа профессорско-преподавательского состава вузов могут принимать участие в проекте. Студенты, участвующие в международном проекте, будут иметь возможность пройти производственную практику в рамках этого проекта. [Читать подробнее...](#)

Подгруппа 1 - Этническая принадлежность по паспорту

Выберите национальность

Подгруппа 2 - Этническая принадлежность по паспорту: только женщины

Выберите национальность

Подгруппа 3 - Этническая принадлежность по паспорту: только мужчины

Подгруппа 4 - Этническая принадлежность по самоопределению

Подгруппа 5 - Этническая принадлежность по самоопределению: только женщины

Подгруппа 6 - Этническая принадлежность по самоопределению: только мужчины

Рисунок 4. Оболочка визуализации данных

В анкете были „открытые“ и „закрытые“ вопросы. Открытые вопросы не содержат вариантов ответов, а дают возможность респонденту самому сформулировать ответ на вопрос в свободной форме. Процедура обработки открытых вопросов достаточно трудоемка. В числе открытых вопросов есть такие, которые требовали от респондента подробного описания. Например: «Какие традиции характерны для Вашего этноса?», «Как Вы празднуете различные народные праздники? Ответы на такие вопросы были представлены на сайте в виде списка.

Созданная база данных позволяют эффективно использовать социолингвистический и этнографический материал всем заинтересованным лицам. Разработанная информационная система открыта и позволяет постоянно пополнять и обрабатывать новые данные. Ниже мы приводим пример использования базы данных для исследования социолингвистической ситуации казахстанской

турецкой диаспоры. В частности, была выявлена диспропорция между этническим самоопределением казахстанских турок и их этничностью по паспорту.

#### 4. Этническая отнесенность турок Казахстана по результатам опроса

Одной из наиболее многочисленных тюркских групп, проживающих в Казахстане, являются турки.<sup>1</sup>

Турецкая этническая группа, проживающая в настоящий период в Казахстане, была депортирована из Грузии в 1944 г. Не имея возможности вернуться к себе на родину в Грузию, турки оказались разбросанными по многим странам [Nevskaya, Tazhibayeva, 2015a, с. 295]. В Казахстане по предварительным данным проживает около 180–200 тысяч турок. По официальным же данным, согласно переписи населения 2009 года, числится 105 000. Причина такого несоответствия заключается в том, что многие граждане до сих пор не могут восстановить свою истинную национальность, измененную при советской власти [Kirisci, 1996, с. 401].

Следует особо подчеркнуть, что отрыв от основного турецкого этнического массива привел, в частности, к консервации языка и определенных элементов традиционной турецкой культуры. Язык казахстанских турок сохранился в той форме, в которой он существовал до кемалистской реформы в Турецкой Республике. При этом в настоящее время имеет место достаточно быстрые процессы культурной ассимиляции казахстанских турок одновременно в казахскую и русскоязычную культуру и сохраняемые старшим поколением язык и культура нуждаются в фиксации исследователями [Nevskaya, Tazhibayeva 2015a, с. 327–329].

Представители турецкой этнической группы проживают в настоящее время в южных регионах Казахстана: в Алматинской, Жамбылской и Южно-Казахстанской областях. Наибольший же процент участвовавших в анкетировании респондентов, проживает в Жамбылской области, см. Рисунок 5.

---

<sup>1</sup> Агентство Республики Казахстан по статистике. Архив: Национальный состав населения Республики Казахстан и его областей (том 1). Численность населения по областям, городам и районам, полу и отдельным возрастным группам, отдельным этносам на 1 января 2010 года.

Местожительство в Казахстане

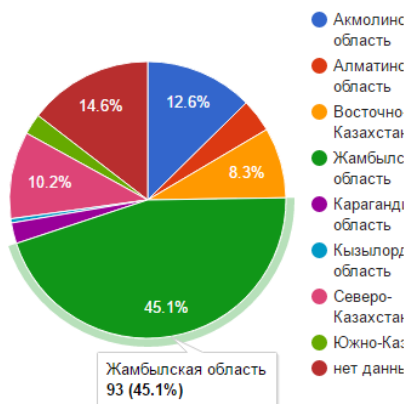


Рисунок 5. Визуализация ответов респондентов о регионах проживания в Казахстане

В анкетировании приняли участие 206 представителей турецкой этнической группы, записанных по паспорту турками (Рис.1), из них мужчин – 143 человека (69,4%) и 63 (30,6%) – женщины. Семьи 40 (19,4%) респондентов были депортированы в 1944 году в Казахстан, 49 (23,8%) участвовавших в анкетировании респондентов родились в Казахстане.

Во время обработки данных мы столкнулись с рядом проблем, одна из которых касается вопросов идентификации респондентов по паспорту и их самоидентификации. Так, 195 респондентов (94,7%) из 206 опрошенных в графе национальность в паспорте указали себя турками, по 2 человека (1,9%) имеют запись в графе национальность турки-хемшили и турки-месхетинцы, по 1-му человеку (0,5%) имеют запись узбек, русский и караим; у 4-х (1,9%) нет данных.<sup>1</sup> Анализ данных показал, что нередко представители турецкого этноса были записаны в паспортах турками, турками месхетинцами, хемшиллы, узбеками, азербайджанцами.

<sup>1</sup> <https://tyurki.weebly.com>. См. раздел «Результаты».

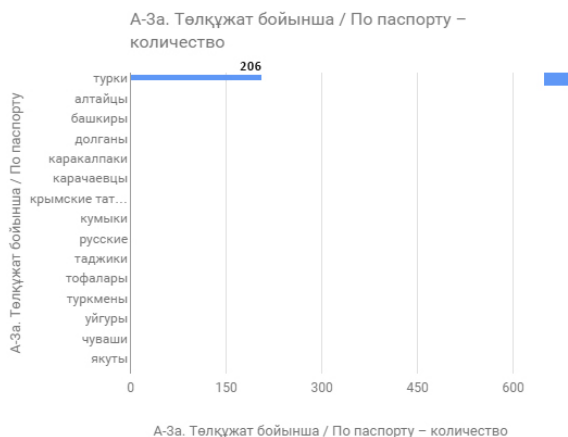


Рисунок 6. Ответы респондентов на вопрос анкеты А-3а «Национальность (этничность) по паспорту»

Что же касается этнического самоопределения, то здесь наблюдается диспропорция в количестве респондентов, записанных турками по паспорту, и теми, которые идентифицируют себя турками: 251 респондент считает себя турками, что на 45 человек больше, чем записанных турками по паспорту (Рисунок 7). Это объясняется тем фактом, что в силу различных причин ряд представителей турецкого этноса были зарегистрированы в паспортах как люди с иной этничностью.

В браке турки предпочитают жениться и выходить замуж за представителей своей этнической группы, но вступают и представители других национальностей. Анализ проведенного опроса показал следующее: у 174 респондентов (94,7%) отцы по национальности являются турками; у 23 (11,2%) – турками-месхетинцами; 2 (1%) – турками-хемшиллы и по 1-му (0,5%) – казахом и азербайджанцем, у 4-х (1,9%) – нет данных.

Национальность матерей респондентов, считающих себя турками, представлена 9-ю национальностями: у 159 респондентов (77,2%) турчанки, матери 2 (1%) респондентов записаны в паспорте как турки-хемшиллы и турки-лазы, у 11 (5,3%) – казашки; 9 (4,4%) – азербайджанки; матери 6 респондентов (2,9%) русские; 4-х (0,5%) – уйгурки; по 3 человека (3,2%) имеют матерей узбечек и киргизок; по 2 (1%) – чеченками, курдами, у 2-х (1%) – нет данных.

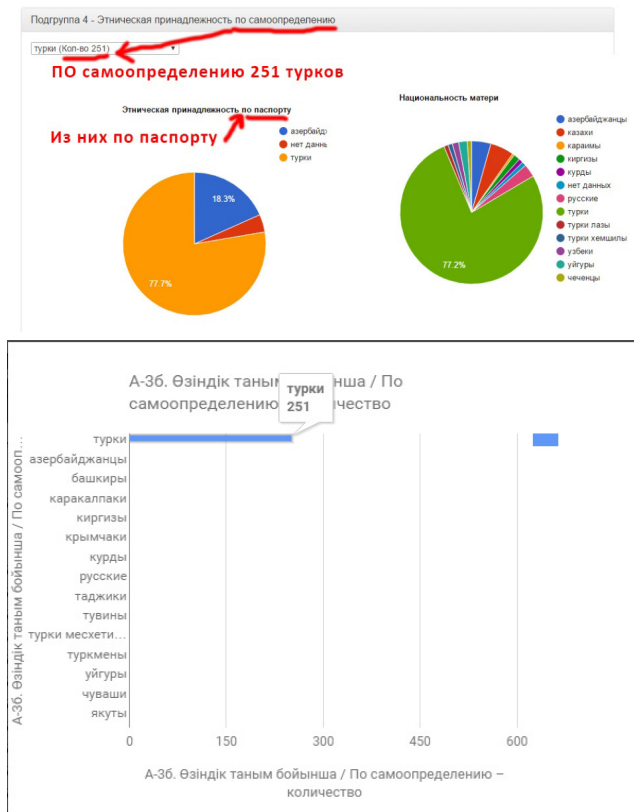


Рисунок 7. Ответы респондентов на вопрос «Ваша национальность по самоопределению?»

Таким образом, у респондентов, участвовавших в анкетировании, национальность по линии отца определена представителями своей этнической группы – турками, а национальность матери представлена принадлежностью к различным тюркским этническим группам (азербайджанки, казашки, узбечки, киргизки), так и к другим этносам (русские, курды, чеченки).

В турецких смешанных семьях главенствующее положение в семье занимает мужчина. В этническом самоопределении доминирующим маркером независимо от этноса матери является отцовская линия, поэтому национальность детей определяется по национальности отца.

Анализ наших данных показывает, что казахстанские турки, независимо от принадлежности к турецким подэтносам (ахыска, хемшили или другим) за семьдесят лет проживания на территории Казахстана идентифицируют себя турками и относят себя к одной этнической группе. Этому, по нашему мнению, способствует ряд факторов:

- укрупнение этноса за счет объединения, и, как следствие, повышение его жизнеспособности на другой территории;
- приобретение исторической этнической родины в лице Турецкой Республики вместо Грузии, где до депортации проживали турки ахыска и турки хемшили;
- создание условий для сохранения этнической культуры, языка в независимом Казахстане;
- перспектива эмиграции молодого поколения в Турцию.

#### ЛИТЕРАТУРА

1. Невская И.А. Компьютерные базы лингвистических данных как основа для сохранения и возрождения коренных тюркских языков Сибири. Образование и устойчивое развитие коренных народов Сибири. – Новосибирск, 2005. С. 90–99.
2. Невская И.А., Тажибаева С.Ж., Шаймердинова Н.Г., Тусупов .А. Тюркский мир Казахстана: исследование языков и создание базы данных Global Turk. – Астана, 2016. – С. 33–42.
3. Batura T.V., Murzin F.A., Sagnayeva S.K., Tazhibayeva S., 2016. Using the link Grammar Parser in the Study of Turkic Languages. In: Eurasian Journal of Mathematical and Computer Applications. – 2016. – № 4 (2). – P. 14–22.
4. Dorian N. Investigating Variation: The Effects of Social Organization and Social Setting – Oxford: Oxford University Press, – 2010. – 376 p.
5. Fedotov, J., Tussupov, M., Sambetbayeva, I., Idrisova, A., Yerimbetova. 2015. Development and Implementation of a Morphological Model of Kazakh Language. In: Eurasian Journal OF Mathematical AND Computer Applications (ISSN 2306–6172). Volume 3, Issue 3 (2015), 69– 79;
6. Gippert, J., Himmelmann, N., Mosel, U. Essentials of language documentation. Walter de Gruyter, 2006.
7. Grenoble L. A., Louanna Furbee N. (eds.) “Language Documentation: Practice and values. – Amsterdam/Philadelphia: John Benjamins Publishing Company”, 2010.
8. Haig G. L. J., Nau N., Schnell S., Wegener C. (eds.) “Documenting Endangered Languages: Achievements and Perspectives”. – Berlin/Boston: De Gruyter Mouton, 2011.



9. Kirisci Kemal. Refugees of Turkish origin: «Coerced Immigrant» to Turkey since 1945. – «International Migration». – Geneva, 1996. – P. 400–410.

10. Lehmann, Ch. 1983. Directions for interlinear morphemic translation. In: *Folia Linguistica*, 16. 193–224.

11. Nevskaya I., Tazhibayeva S. 2015a. Turkic Languages of Kazakhstan: Problems and research perspectives. In: Lars Johanson (ed.) *Turkic Languages*. Volume 16, 2014. Numbers 1/2. – P. 289-302. – Weisbaden: Harrassowitz, 2015.

12. Nevskaya I., Tazhibayeva S. 2015b. Sociolinguistic situation of Oguz Turks in post-Soviet Kazakhstan In: *Oguzlar. Dilleri, Tarihleri ve Kulturleri* – Ankara, 2015 – P. 321–334. – ISSN 978-975-491-405-4.

УДК 004.822

**БАЗЫ ЗНАНИЙ ПОРТАЛА «ТЮРКСКАЯ МОРФЕМА»:  
СОСТОЯНИЕ, ПЕРСПЕКТИВЫ**

*А. Р. Гатиатуллин, Н. А. Прокопьев, Дж. Ш. Сулейманов*  
*Институт прикладной семиотики АН РТ*  
*Казань, Татарстан, Россия*  
ayrat.gatiatullin@gmail.com, nikolai.prokopyev@gmail.com,  
dvdt.slt@gmail.com

В работе описываются базы знаний, представленные в портале «Тюркская морфема», представленные в виде графов знаний, технологии их создания и использования для обработки естественного языка. Особенность данных графов знаний в том, что, с одной стороны, они содержат лингвистические единицы разного языкового уровня, а с другой стороны, концепты, обозначающие значения этих лингвистических единиц, которые встроены в общую систему концептов. Данные графы знаний, используются для семантической разметки электронных корпусов, представленных в рамках портала, а также сами пополняются за счет информации, представленной в этих электронных корпусах.

**Ключевые слова:** граф знаний, интернет-портал, лингвистическая единица

**KNOWLEDGE BASE OF THE “TURKIC MORPHEME PORTAL”:  
STATUS, PROSPECTS**

*Ayrat Gatiatullin, Nikolai Prokopyev, Dzhavdet Suleymanov*  
*Institute of Applied Semiotics of TAS*  
*Kazan, Tatarstan, Russia*  
ayrat.gatiatullin@gmail.com, nikolai.prokopyev@gmail.com,  
dvdt.slt@gmail.com

The paper describes the knowledge graphs presented in the “Turkic Morpheme” web-portal, presented in the form of knowledge graphs, technologies for their creation and their use for natural language processing. Peculiarity of these knowledge graphs is that, on the one hand, they contain linguistic units of different linguistic levels, and, on the other hand, concepts denoting the meanings of these linguistic units that are built into the general system of concepts. These knowledge graphs are used for semantic annotation of electronic corpora presented within the portal, and are themselves updated with the information provided by these electronic corpora.

**Keywords:** knowledge graph, web-portal, linguistic unit

## Введение

Создание лингвистических баз знаний для тюркских языков в настоящее время особо актуально по целому ряду причин. Рассмотрим эти причины.

Все тюркские языки (кроме турецкого) - малоресурсные языки. Татарский язык также относится к данной категории, поэтому этот вопрос актуален для Республики Татарстан. Отсутствие баз данных и баз знаний с необходимыми лингвистическими ресурсами тормозит создание программного обеспечения связанного с обработкой этих языков. Это актуально, как для татарского, так и всех тюркских языков в целом.

Термин малоресурсные языки был введен еще в 2003 год Краувером [Krauwer, 2003]. Согласно его определению малоресурсные языки – это естественные языки, обладающие следующими свойствами:

1. Недостаток своей системы письменности или устойчивой орфографии;
2. Нехватка квалифицированных лингвистов и переводчиков для данного языка;
3. Ограниченное распространение в сети Интернет;
4. Нехватка электронных ресурсов для обработки языка и речи, в том числе одноязычных корпусов, двуязычных электронных словарей, орфографических и фонетических транскрипций речи, словарей произношения и т. д.

Для большинства тюркских языков первый пункт неактуален, так как они все обладают системой письменности и устойчивой орфографией, а вот 2-4 пункта достаточно актуальны. Если 2-й и 3-й пункты связаны с человеческими ресурсами, которые способны осуществлять перевод текстов и создавать контент в сети Интернет на этих языках, то 4-й пункт связан с наличием технологий и разработок для создания программного обеспечения по компьютерной обработке этих языков.

Одной из причин нехватки, соответствующего программного обеспечения является то, что лингвистические модели и ресурсы для других типов языков плохо приспособлены к структурно-функциональным особенностям тюркских языков. Тюркские языки обладают большим набором структурно-функциональных особенностей, благодаря которым многие универсальные программные продукты, созданные для других языков к ним плохо

применимы. Это подтверждает статья [Papadimitriou, 2022], где описано явление, которое, часто ухудшает работу мультязычных моделей системы BERT – по причине того, что грамматические структуры из высокоресурсных языков, типа английского, как бы перетекают в малоресурсные. Авторы пишут, что “Подобно людям, не являющимся носителями языка, мультязычные модели склонны использовать структуры родного (превалирующего) языка, нежели структуры, свойственные иностранному (малоресурсному) языку”.

Рассмотрим какие свойства характерны тюркским языкам, некоторые из них, приведены в работах [Сулейманов, 1999, Гузев, 2020]

1. Агглютинативность,
2. Сингармонизм,
3. Отсутствие грамматического выделения единственного числа,
4. Отсутствие категории рода,
5. В тюркских языках редко встречаются исключения из правил,
6. Подавляющее большинство агглютинативных аффиксов однозначно.
7. Имена существительные обладают способностью выполнять функцию определения.

Третьей причиной малоресурсности является недостаточное количество коллабораций в разработках для тюркских языков, и как следствие отсутствие единых баз данных и баз знаний, единой терминологии и системы обозначений и тегов для разметки. Это показывает анализ корпусов для тюркских языков. Данную причину можно связать с недостаточностью финансирования совместных международных и межрегиональных проектов в области компьютерной лингвистики, наличия совместных грантов.

Тюркские языки сильно отстали по количеству существующих лингвистических ресурсов от индоевропейских языков. А с учетом разницы количества специалистов работающих в разработках для индоевропейских и для тюркских языков, это отставание продолжает увеличиваться. Мнение авторов, что данный разрыв можно сократить с помощью объединения усилий разработчиков для разных тюркских языков и автоматизации сбора лингвистических ресурсов. Следует также отметить, что степень малоресурсности тюркских языков сильно различается в зависимости от того, насколько сильно поддерживаются компьютерные разработки для

этих языков государством, а также наличием квалифицированных специалистов. В связи с этим, усилия для создания языковых ресурсов и инструментов для языков с меньшим количеством ресурсов часто можно уменьшить, используя уже существующие ресурсы и инструменты для родственных, более богатых ресурсами языков. Так в работе [Тан, 2019] показывают, что использование свойства близости родственных языков может существенно повысить качество машинного перевода.

Создание комплексных (интегральных) лингвистических ресурсов и моделей для тюркских языков позволит решать прикладные задачи более экономичным способом за счет взаимодополнения разработок между языками. Так, еще в 1988 году группой авторов В.Г. Гузев, Р.Г. Пиотровский, А.М. Щербак [Гузев, 1988] была высказана идея, что для решения практических задач нужно создавать большой многоцелевой машинный фонд тюркских языков, который должен строиться, моделируя как общетюркскую языковую систему, так и систему каждого конкретного языка (функционирующего или мертвого, современного или древнего) со всеми ее инвентарными и структурными единицами, со всевозможными правилами знаковой репрезентации языковых единиц в речи, включая правила линейного развертывания речевых единиц.

Осуществить такое объединение может создание единой технологической платформы, которая одновременно будет выполнять роль некоторого координационного центра и общей ресурсной базы в области компьютерной обработки тюркских языков. Данная платформа должна реализовывать следующие функции:

1. Служить информационно-справочной системой по тюркским языкам;

2. Служить библиотекой лингвистических стандартов по тюркским языкам (термины, теги) для нового создаваемого программного обеспечения, что позволит обеспечить взаимную совместимость для разработок, создаваемых разными коллективами;

3. Служить межпрограммным интерфейсом для связывания ранее созданных лингвистических ресурсов и программного обеспечения, производящего обработку различных ресурсов на тюркских языках. Для этого в платформе должны храниться все таблицы соответствия системы обозначения посторонних разработок к системе обозначений платформы;

4. Служить библиотекой компьютерных моделей для описания структурно-функциональных особенностей тюркских языков и

информационной базой для лингвистических процессоров, которые производят компьютерную обработку тюркских языков;

5. Служить библиотекой программных модулей для создания прикладных программ, работающих с тюркскими языками;

6. Служить платформой с созданием виртуальной среды, в которой сторонние разработчики смогут реализовывать свои программные разработки по компьютерной обработке тюркских языков;

7. Служить виртуальной площадкой для общения специалистов тюркологов, которые пополняют лингвистическую базу данных и информационно-справочную систему по тюркским языкам.

Основу такой платформы будут составлять базы знаний, отражающие структурно-функциональные особенности тюркских языков, одним из наиболее популярных способов представления баз знаний для задач семантической обработки текста является создание лингвистических графов знаний и технологий их использования для эффективного решения задач обработки языка. Технологии, разработанные или адаптированные для тюркских языков, могут быть использованы для других языков агглютинативного типа, которые по другой классификации можно назвать языками с элементарно-комбинаторной грамматикой.

Для автоматизации заполнения лингвистических баз знаний данной платформы необходимы соответствующие технологии и инструментарий, прагматически-ориентированный под структурно-функциональные особенности тюркских языков. По нашему мнению, наиболее подходящими для организации баз знаний лингвистических платформ для тюркских языков являются графы знаний.

## 1. Обзор типов графов знаний

В последние годы в сфере обработки семантических данных большое внимание уделяется технологиям, называемым графы знаний [Hogan et al., 2020; Fensel et al., 2020; Ji et al., 2020], которые рассматриваются как системы представления знаний на основе графов, способных упорядочить информацию гибким и интуитивно понятным способом. Графы знаний активно используются в лингвистических сервисах таких крупных компаний как Google, Yandex, Facebook.

Несмотря на активное использование графов знаний, единого общепринятого их определения не существует [Ehrlinger et al.,

2016]. Рассмотрим одно из определений, представленное в работе [Pan et al., 2017]:

**Граф знаний** – это структурированный набор данных, собранный из разнородных источников данных, совместимый с моделью данных RDF и имеющий (OWL) онтологию в качестве своей схемы. Граф знаний (ГЗ) не обязательно связан с внешними графами знаний; однако сущности в графе знаний обычно имеют информацию о типе, определенном в его онтологии, которая полезна для предоставления контекстной информации о таких сущностях.

Одни авторы видят в графах знаний некую реинкарнацию онтологий [Lawgупowicz, 2017], другие определяют их как системы, основанные на знаниях Knowledge-based systems (KBS) [Ahmed et al., 2019]. Хотя по своей сути они представляют комбинацию онтологий и тезаурусов с множествами именованных сущностей. Как показывает обзор литературы, базовыми компонентами являются такие ресурсы, как WordNet и FrameNet, остальные ресурсы объединяются вокруг них. В более выгодном положении оказываются те языки, для которых ранее были созданы подобные ресурсы. Среди тюркских языков к таким языкам относится только турецкий.

Граф знаний с данными фреймового типа представлены в таких ресурсах, как Frame oriented Knowledge Graphs. Одним из примеров разработок является ресурс FrameSter [Gangemi et al., 2016]. Framester является концентратором между, такими ресурсами, как FrameNet, WordNet, VerbNet, BabelNet, DBpedia, Yago, DOLCE-Zero, а также другими ресурсами. Framester – это не только сильно связанный граф знаний, но в нем также применяется строгая формальная обработка семантики фреймов Филлмора, что позволяет выполнять полноценные запросы и рассуждения OWL на большом графе знаний на основе фреймов.

Каждый из описанных вариантов комбинирования графов знаний позволяет представлять разного типа лингвистическую информацию. Комбинирование с ГИС позволяет отобразить географию распространения языков и диалектов, а комбинирование с фреймовыми типами – представлять динамические языковые модели. Таким образом, для наиболее полного описания языковой семантики тюркских языков следует интегрировать в единую модель все описанные выше комбинации графов знаний.

Графы знаний также активно используются для представления лингвистической информации. Один из таких лингвистических

графов описан в работе [Basile, 2022], модель данного лингвистического графа знаний представлена на рис.1. По мнению авторов, данный граф позволяет моделировать: 1) отношения между понятиями и словами; 2) информацию о встречаемости слов; 3) диакроническая информация, как понятий, так и слов. Представленный в данной работе лингвистический граф включает такие вершины, как Концепты (Concept), Лексические концепты (Lexicon Concept), Лексемы (Lexicon Entry). Лексические концепты взаимосвязаны между собой семантическими отношениями типа гипонимии и гиперонимии. Лексемы взаимосвязаны с леммой (Lemma) и основной словоформы (Stem). Особенность данного графа знаний в том, что они не описывают ситуационную семантику, поэтому семантическое описание является достаточно ограниченным и отражает только таксономические отношения, аналогичные тем, что представлены в известном ресурсе WordNet. Для описания данной семантики подходят графы знаний фреймового типа, например такой достаточно известный ресурс, как FrameNet.

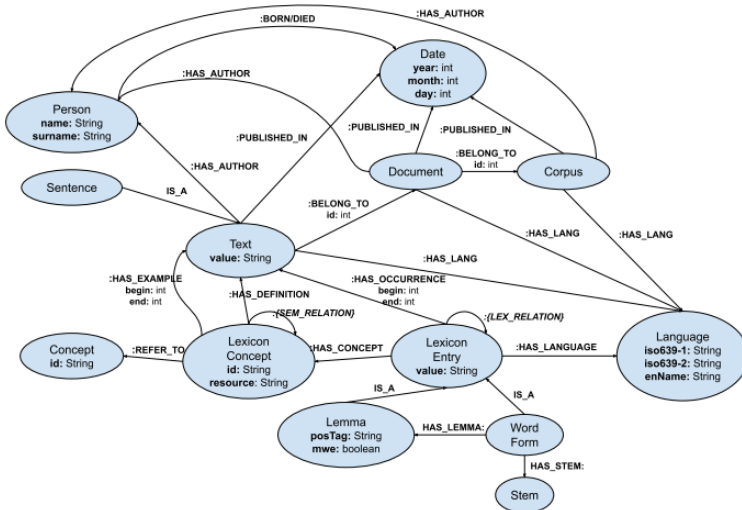


Рис.1. Модель лингвистического графа знаний

Также в этом графе знаний не представлена грамматическая (морфологическая) структура словоформ. А для турецких языков это очень важно, так как они обладают богатой морфологией



агглютинативного типа. В отличие от языков флективного типа, куда относится русский язык, в тюркских языках четкое деление на структурные компоненты слова, которые называются морфемами. Такое четкое деление позволяет и морфологическую структуру словоформы представить в виде графа, вершинами которого являются морфемы.

В другой работе описывается лингвистический граф знаний, который называется Lexicon Model for Ontologies (LeMon) [McGae, 2011], где для описания значения лингвистических единиц уже используются семантические фреймы.

Рассмотрим сущности и отношения между ними, которые представлены на рис. 2.

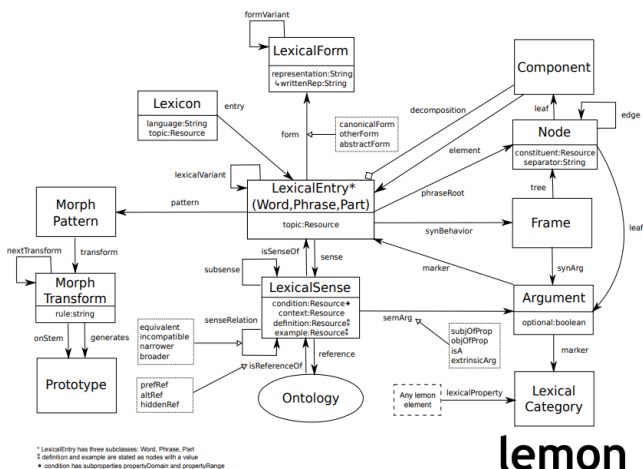


Fig. 1. The *lemon* model

Рис.2. Модель лингвистического графа знаний Lemon

Лексикон (Lexicon) – это словарь, которые имеет такие атрибуты, как язык и тема.

Центральным элементом в данной модели является Словарная статья (Lexical Entry), которая имеет наибольшее количество связей с другими сущностями и может являться словом, фразой или частью слова. Поскольку морфосинтаксическая информация связана со словарной статьей, каждая статья должна иметь стандартной написание. Варианты терминов, например, такие как аб-

бревиатура, представляются в виде отдельных словарных статей и помечаются как lexicalVariants.

Словарная статья может состоять из нескольких Форм (Lexical Form), одна из которых может быть помечена, как лемма (canonical form).

Лексическое значение (Lexical Sense) представляет собой некое соответствие между словарной статьей (Lexical Entry) и сущностью онтологии (Ontology), которое устанавливается с помощью ссылки (Reference). Оно может включать дополнительную спецификацию этого соответствия, такую как контекст, условие, определения или примеры.

С помощью сущностей Фрейм (Frame) и Аргумент (Argument) в модели могут быть представлены валентность глаголов и других лексических предикаторов. Каждый аргумент также представлен как отдельная сущность, связанная как с фреймом, чтобы указать синтаксическую роль, так с лексическим значением, чтобы указать семантическую роль.

Словарная статья может представлять многословные выражения и составные слова, для этого в модели имеется элемент компонент (Component), каждый из которых ссылается на другие словарные статьи.

Для представления в модели структуры словарной статьи, являющейся фразой существуют элементы Узел (Node) Они представляют собой ряд узлов, связанных ребрами или листовыми дугами с компонентами.

Разработчики этих графов знаний утверждают, что они разработаны в соответствии со стандартами LMF (Lexical Markup Framework), который классифицируется как ISO-TC37/SC4/WG4. Данный стандарт подробно описан в работе [Francopoulo, 2006]. На рис. 3. представлена одна из диаграмм, демонстрирующих структуру Базы данных в соответствии со стандартом LMF.

Выше описанные лингвистические графы знаний подходят для описания языков типа английского с достаточно бедной морфологией и не позволяют адекватно описывать всю полноту лингвистической информации, имеющихся в агглютинативных языках с богатой морфологией. Для тюркских языков, также важно отобразить в онтологических моделях их морфологическую структуру, поскольку аффиксальные морфемы также несут отдельные значения. Особенность тюркских языков в том, что в отличие от языков флективного типа в них существует четкое деление слово-

форм на структурные компоненты, которые называются морфемами. Такое четкое деление позволяет морфологическую структуру словоформы представить в виде графа, вершинами которого являются морфемы.

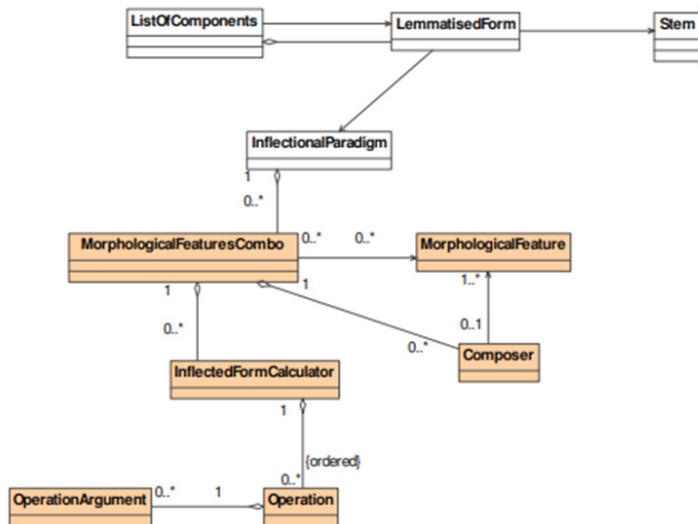


Рис.3. Структура БД для представления данных в соответствии со стандартом LMF

## 2. Графы знаний портала “Тюркская морфема”

Интернет-портал «Тюркская морфема» представляет собой web-сайт ([modmorph.turklang.net](http://modmorph.turklang.net)), который включает набор различных сервисов на базе лингвистических ресурсов с тюркскими языками и ориентирован на работу с тюркскими языками во всех аспектах: морфонологическом, морфологическом, синтаксическом, семантическом. Основной набор функций данного интернет-портала представлен на рис.4.

Как указано на рис. 4, все функции портала подразделяются на базовые и прикладные. Базовый набор этих функций является более-менее устоявшимся, к нему относятся:

1. Информационно-справочная система по тюркским языкам, предоставляющая информацию о грамматике данных языков и о межязыковом тезаурусе.



Рис 4. Функции портала

2. Ресурсная онто-лингвистическая база для решения прикладных задач, таких как создание новых программных средств для обработки естественного языка, проведения лингвистических исследований.

3. Конвейер программных модулей для обработки естественного языка, уже создаваемых на базе портала и используемых для его развития.

4. Площадка для совместной работы экспертов типологов, лингвистов, диалектологов, разработчиков, так как задача создания ресурсной базы по множеству тюркских языков требует привлечения множества заинтересованных специалистов.

Набор прикладных функций включает:

1. Инструментарий для научных исследований, формируемый на ресурсной базе портала, к примеру инструменты сводных таблиц для проведения компаративных лингвистических исследований.

2. Инструментарий для создания обучающих систем, использующих базы данных портала для создания учебных материалов и проведения автоматизированного контроля знаний.

3. Инструментарий для создания терминологии и унификации обозначений для области электронной тюркологии, к примеру унификация аннотации тюркских электронных корпусов.

Данный набор будет расширяться по мере наполнения онто-лингвистической базы портала и расширения набора про-

граммных модулей, входящих в конвейер обработки естественного языка программного инструментария портала.

## 2.1. Архитектура графов знаний портала

Лингвистическая база знаний портала представляет собой единый граф знаний, который подразделяется на несколько подграфов. Разделение на подграфы сделано в связи со структурными особенностями каждого из этих подграфов, а также с тем, что каждый из подграфов содержит наборы вершин одного типа. Вершины одного подграфа связаны между собой отношениями одного типа, а с вершинами из других подграфов отношениями другого типа. Схема разделения на подграфы представлена на рис.5. Такое разделение связано и с задачами, для решения которых используются каждый из подграфов единого графа знаний портала.

Далее рассмотрим подграфы знаний портала, объединяемые в единый граф знаний.

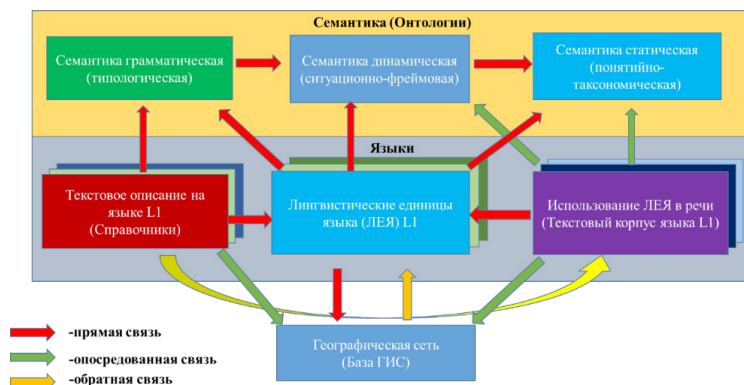


Рис.5. Архитектура подграфов графа знаний портала

Понятийно-таксономический граф – это граф знаний, образующий таксономию, аналогичную известному тезаурусу WordNet. Этот граф используется для описания значения лексических лингвистических единиц тюркских языков, описываемых в данном портале.

Таксономия строится с помощью триплетов, в которых субъектом и объектом триплета являются концепты, а ребро показывает отношение гиперонимии.

Каждый концепт имеет описание, состоящее из нескольких элементов:

1. Имя концепта на английском языке.
2. Синсет имен концепта на русском языке.
3. Описание концепта на английском языке.
4. Описание концепта на русском языке.

Подграфы знаний с описанием лингвистических единиц содержат вершины, соответствующие лингвистическим единицам тюркских языков разного языкового уровня. Это и корневые и аффиксальные морфемы, и многословные выражения и т.д. Базовой лингвистической единицей в составе графов знаний с описанием лингвистических единиц являются морфемы. Рассмотрим один тип корневых морфем, которые несут в себе лексическое значение и могут соответствовать одному или нескольким концептам таксономического графа.

### 3. Построение графов знаний

В технологиях представления знаний и работы с графами знаний главным вопросом является вопрос построения этих графов знаний. При их создании возможно извлекать данные как из структурированных, полуструктурированных, так и неструктурированных данных. Данный процесс представлен на рис.6.

Наибольшее количество графов знаний создается для английского языка, а также для целого ряда других языков. В России они не получили широкого распространения и практически отсутствуют литература на русском языке, можно лишь выделить ряд переводных работ. Например, перевод Боргестом статьи [Баклавски и др., 2020]. Возможно одна из причин в том, что в основе графов знаний находятся такие онтологические ресурсы, как WordNet и FrameNet, а для языков РФ подобных ресурсов мало (к таковым можно отнести проекты: Tatar WordNet, ...) и они довольно разрозненны, что еще раз подтверждает малоресурсность данных языков. Среди тюркских языков наличием подобных ресурсов до настоящего времени выделялся только турецкий язык.

С целью решения этой проблемы для тюркских языков в Институте прикладной семиотики Академии наук Республики Татарстан ведется разработка портала “Тюркская морфема” [Gatiatullin et al., 2020]. Главной особенностью данного портала является то, что он прагматически-ориентирован именно на структурно-функ-

циональные особенности тюркских языков и направлен на решение целого класса задач.

Семиотическому исследованию особенностей лексико-грамматических характеристик тюркских языков и их описанию посвящены работы [Suleymanov, 2010] и [Сулейманов, 2021].

Процесс построения графа знаний представлен на рис.3. К структурированным данным относятся базы данных самого портала, к полу-структурированным – в основном размеченные корпусные данные, а к неструктурированным – произвольные данные на тюркских языках. В рамках портала реализуется единый интерфейс доступа к таким данным, в том числе электронные корпуса интегрируются с порталом в едином веб-интерфейсе.

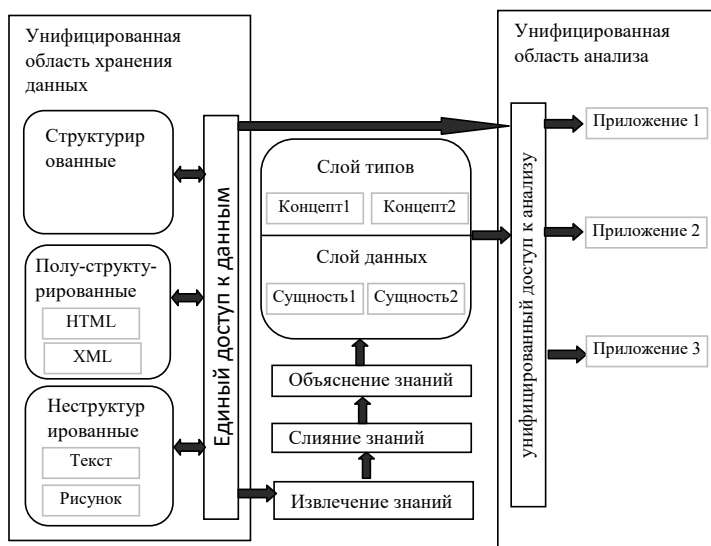


Рис. 6. Процесс построения графа знаний

Разрабатываемый инструментарий для построения графа знаний включает в себя инструменты извлечения и объяснения знаний (семантико-синтаксические анализаторы, использующие ресурсы портала), слияния знаний в онтологические структуры, которые в дальнейшем пополняют базы структурированных данных портала.

Для автоматизации заполнения графов знаний предлагается использовать технологии машинного обучения. Для реализации

данных технологий был проведен анализ существующих подходов и в качестве наиболее перспективного на данный момент предлагается контекстуализированная векторная модель BERT.

Данный метод описывается, как один из наиболее эффективных подходов для решения задач извлечения структурированных данных, в области обработки естественного языка за последнее десятилетие, который был предложен учеными компании Google. Авторы объединили подход к переносу обучения с недавно созданной архитектурой трансформер [Vaswani, 2017]. Помимо этого, была предложена задача маскированного языкового моделирования в качестве целевой задачи предварительного обучения.

В 2019 мультязычный BERT был адаптирован для русского языка и появился RuBERT. Поверх обучили CRF-голову, получился DeepPavlov BERT NER – SOTA для русского языка. На соревнованиях модель показала в 2 раза меньше ошибок, чем у ближайшего преследователя DeepPavlov NER.

В отличие от контекстуализированных векторных представлений, построенных с использованием LSTM сетей, подобная сеть намного лучше позволяет настраиваться на предметную область и целевую задачу. Для этого слои трансформер-кодировщика не замораживаются во время тонкой настройки на задачу, а обучаются вместе со слоями, необходимыми для целевой задачи.

Структура графов знаний портала “Тюркская морфема” для задач автоматизации заполнения требует решения, как извлечение концептов (concept extraction), так и извлечение отношений между ними (relation extraction).

Извлечение концептов представляет собой поиск в неструктурированном тексте и последующая интерпретация лексических обозначений некоторых ментальных конструкций, используемых в целевой модели знаний [Fu, 2020]. В рамках извлечения концептов также могут использоваться инструменты извлечения терминологии (terminology extraction) и инструменты извлечения именованных сущностей (named entity recognition).

Задача распознавания именованных сущностей может рассматриваться как задача распознавания и классификации имен собственных из корпуса. Под именованными сущностями, как правило, понимают имена собственные, выделяющие именуемый объект из ряда подобных. В зависимости от поставленной практической задачи различаются как классы распознаваемых имен



собственных. Для решения задач распознавания именованных сущностей применяется практически весь спектр различных архитектур моделей и методов МО [Lample, 2016; Li, 2020; Shen, 2018; Yadav, 2019]. Для малоресурсных языков практикуются гибридные методы, которые способны повышать общую производительность систем или преодолевать трудности, связанные с нехваткой ресурсов в контексте малоресурсных языков. В работе [Shaalán, 2014] используется комбинация методов на основе правил и машинного обучения для создания системы распознавания именованных сущностей на арабском языке. Такой подход позволил авторам повысить общую производительность предлагаемого метода, а также преодолеть проблемы, связанные с нехваткой языковых ресурсов для их языка.

Одной из важных задач в построении онтологий является построение таксономии классов, т.е. выявление отношений между более широкими (родовыми) классами и их более конкретными (видовыми) классами сущностей. В извлечении знаний из текстов данная задача ставится как извлечение гиперонимов - родовых слов для данного нового слова. Раньше извлечение гиперонимов производилось в основном на основе тезаурусов типа WordNet, содержащего представления значений более 100 тысяч слов английского языка в виде семантической сети.

В извлечении отношений аналогично методам извлечения концептов используются гибридные методы, которые представляют различные комбинации других методов для повышения производительности или преодоления каких-либо проблем, в частности, связанных с недостатком языковых ресурсов. Например, в работе [Devisree, 2016] представлен гибридный подход, комбинирующий машинное обучение и правила, для извлечения отношений между героями рассказов.

В настоящее время также популярным становится многозадачное обучение (HMTL – модель Hierarchical Multi-Task Learning) – метод, в котором единственная архитектура обучается одновременно выполнять разные задачи. Многозадачное обучение в основном используется для решения следующих задач:

- Распознавание именованных сущностей (NER)
- Обнаружение упоминаний сущностей (EMD);
- Разрешение привязки (CR);
- Извлечение отношений (RE).

## Заключение

В статье описаны перспективы использования лингвистических графов знаний в задачах обработки, накопления и изучения языковой информации. Представлены наработки и идеи развития портала “Тюркская морфема” в направлении построения графов знаний тюркских языков. В частности, рассмотрены инструменты для интеграции электронных корпусов с порталом, позволяющие, с одной стороны, пополнять базы данных портала за счет анализа корпусных данных, а с другой, использовать конвейер программных модулей для обработки естественного языка и унифицированную систему аннотации для разметки корпусов.

Таким образом, в настоящее время осуществляется формирование единого инструментария портала на основе использования разных типов интеграции графов знаний портала и электронных корпусов для тюркских языков.

## ЛИТЕРАТУРА

1. Krauwer S. The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. Proc. International workshop on speech and computer SPECOM-2003. Moscow, Russia, 2003. Pp. 8–15.
2. Tan X., Chen J., He D., Xia Y., Qin T., Liu T. Y. Multilingual Neural Machine Translation with Language Clustering // In EMNLP/IJC-NLP. – 2019.
3. Papadimitriou, Isabel, Kezia Lopez, and Dan Jurafsky. «Multilingual BERT has an accent: Evaluating English influences on fluency in multilingual models.» *arXiv preprint arXiv:2210.05619* (2022).
4. Suleymanov D.S. Natural possibilities of the Tatar morphology as a formal base of the NLP // In Proceedings of the First International Workshop “Computerisation of Natural Languages” (Varna, Sept. 3–7, 1999). – Sofia (Bulgaria): Information Services Plc, 1999. – P. 113.
5. Гузев В.Г. О некоторых экзотических особенностях тюркских языков («тюркские чудеса») // Актуальные проблемы мировой политики. Вып. 10 / под ред. Т.С.Немчиновой. СПб.: Изд-во С.-Петербур. ун-та, 2020. С. 231–245.
6. Гузев В.Г., Пиотровский Р.Г., Щербак А.М. О создании машинного фонда тюркских языков // Советская тюркология. 1988. №2. С. 92–101.
7. Fu S., Chen D., He H., Liu S., Moon S., Peterson K. J., Shen F., Wang L., Wang Y., Wen A., Zhao Y., Sohn S., Liu H. Clinical concept extraction: A methodology review. *Journal of Biomedical Informatics*, 2020, vol. 109, pp. 103526.

8. Hogan A., Blomqvist E., Cochez M., d'Amato C., de Melo G., Gutierrez C., Gayo J.E.L., Kirrane S., Neumaier S., Polleres A. Knowledge graphs // arXiv preprint arXiv:2003.02320. 2020.

9. Fensel D., Şimşek U., Angele K., Huaman E., Kärle E., Panasiuk O., Toma I., Umbrich J., Wahler A. Knowledge Graphs. Methodology, Tools and Selected Use Cases // Springer Nature, 2020

10. Ji S., Pan S., Cambria E., Marttinen P., Yu P. S. A survey on knowledge graphs: Representation, acquisition and applications // arXiv preprint arXiv:2002.00388. 2020.

11. Ehrlinger L., Wöß W. Towards a definition of knowledge graphs. SEMANTiCS (Posters, Demos, SuCCESS) // Proc. 12th International Conference on Semantic Systems – SEMANTiCS2016, CEUR Workshop Proceedings. 2016. Vol.1695.

12. Pan J.Z., Vetere G., Gomez-Perez J.M., Wu H. Exploiting Linked Data and Knowledge Graphs in Large Organizations // Springer Cham. 2017.

13. Lawrynowicz A. (2017). Semantic data mining: an ontology-based approach // Studies on Semantic Web. 2017. Vol.29.

14. Ahmed A., Al-Masri N., Abu Sultan Y.S., Akkila A.N., Almasri A., Mahmoud A.Y., Zaqout I.S., AbuNaser S.S. Knowledge-based systems survey // International Journal of Academic Engineering Research (IJAER). 2019. Vol.3 Is.7.

15. Gangemi A., Alam M., Asprino L., Presutti V., Recupero D.R. Framester: A Wide Coverage Linguistic Linked Data Hub // Lecture Notes in Computer Science. 2016. Vol.10024. – Springer Cham.

16. Pierpaolo Basile, Pierluigi Cassotti, Stefano Ferilli1, and Barbara McGillivray New Time-sensitive Model of Linguistic Knowledge for Graph Databases // Proceedings of the 1st Workshop on Artificial Intelligence for Cultural Heritage co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AIXIA 2022), CEUR Workshop Proceedings. 2022. Vol.3286.

17. J. P. McCrae, D. Spohr, P. Cimiano, Linking lexical resources and ontologies on the semantic web with lemon, in: G. Antoniou, M. Grobelnik, E. P. B. Simperl, B. Parsia, D. Plexousakis, P. D. Leenheer, J. Z. Pan (Eds.), The Semantic Web: Research and Applications – 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I, volume 6643 of Lecture Notes in Computer Science, Springer, 2011, pp. 245–259.

18. Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M., Soria, C.: Lexical markup framework (LMF). In: Proceedings of the Fifth International Conference on Language Resource and Evaluation (LREC 2006) (2006)

19. Баклавски К., Беннет М., Берг-Кросс Г., Шнайдер Т., Шарма Р., Сингер Д. Онтологический Саммит 2020. Коммюнике: Графы знаний // Онтология проектирования. 2020. Т.10, №4(38).

20. Gatiatullin A., Suleymanov D., Prokopyev N., Khakimov B. About Turkic Morpheme Portal // Proc. Computational Models in Language and Speech Workshop 2020. CEUR Workshop Proceedings. 2020. Vol.2780.

21. Suleymanov D.Sh. Natural Cognitive Mechanisms in the Tatar language // In Proc. 20th European Meeting in Cybernetics and Systems Research, Austria. 2010.

22. Сулейманов Д.Ш. Инфокоммуникационные технологии и естественный язык: региональный опыт // Труды 19 Национальной конференции по искусственному интеллекту КИИ-2021, Таганрог, 2021.

23. Vaswani A. [и др.]. Attention is all you need // arXiv preprint arXiv:1706.03762. – 2017.

24. Bahdanau D., Cho K., Bengio Y. Neural machine translation by jointly learning to align and translate // arXiv preprint arXiv:1409.0473. – 2014.

25. Neural Architectures for Named Entity Recognition / G. Lample [et al.] // arXiv:1603.01360 [cs]. arXiv, 2016.

26. A Survey on Deep Learning for Named Entity Recognition / J. Li [et al.] // arXiv:1812.09449 [cs]. arXiv, 2020.

27. Deep Active Learning for Named Entity Recognition / Y. Shen [et al.] // arXiv:1707.05928 [cs]. arXiv, 2018.

28. Yadav V., Bethard S. A Survey on Recent Advances in Named Entity Recognition from Deep Learning models // arXiv:1910.11470 [cs]. arXiv, 2019.

29. Shaalan K., Oudah M. A hybrid Approach to Arabic Named Entity Recognition. Journal of Information Science, 2014, vol. 40, pp. 67–87.

30. Devisree V., Raj P. C. R. A Hybrid Approach to Relationship Extraction from Stories: International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST – 2015) // Procedia Technology. 2016. Vol. 24. P. 1499–1506.

УДК 81'42

**СОЗДАНИЕ БАЗЫ ДАННЫХ ПОЛИТИЧЕСКОГО ДИСКУРСА  
НА КАЗАХСКОМ ЯЗЫКЕ**

*А. Д. Сайранбекова, Л. О. Орынбай, А. Ж. Укенова,  
А. А. Шарипбаев, Б. Ш. Разахова*

*Евразийский Национальный университет имени Л. Н. Гумилева  
Астана, Казахстан*

sairanbekova98@gmail.com, laura.aktobe.kz@gmail.com,  
ukenovaaru07@gmail.com, sharalt@mail.ru, razakhova\_bsh@enu.kz

Политический дискурс – это сложная исследовательская концепция, связанная с анализом языка и коммуникации, используемых в политических контекстах. Этот термин охватывает способы, которыми язык формирует, передает и влияет на политические идеи, убеждения и социокультурные нормы. В данной статье мы можем увидеть анализ интернет-ресурсов по политическому дискурсу на казахском языке и начальную работу по созданию корпуса текстов.

**Ключевые слова:** политический дискурс, анализ, казахский язык, текстовый корпус.

**DATABASE CREATION OF POLITICAL DISCOURSE IN THE  
KAZAKH LANGUAGE**

*Sairanbekova Ayaulym, Laura Orynbay, Aru Ukenova,  
Altynbek Sharipbay, Bibigul Razakhova*

*L. N. Gumilyov Eurasian National University,  
Astana, Kazakhstan*

sairanbekova98@gmail.com, laura.aktobe.kz@gmail.com,  
ukenovaaru07@gmail.com, sharalt@mail.ru, razakhova\_bsh@enu.kz

Political discourse is a complex research concept related to the analysis of language and communication used in political contexts. This term encompasses the ways in which language shapes, conveys, and influences political ideas, beliefs, and sociocultural norms. In this article we can see the analysis of Internet resources on political discourse in the Kazakh language and the initial work on the creation of a text corpus.

**Keywords:** political discourse, analysis, Kazakh language, a text corpus.

**Introduction**

Political discourse has a significant impact on the formation of public opinion, political preferences and behavior. The ability of po-

litical leaders and groups to use discourse effectively allows them to persuade, mobilize and shape social norms. The analysis of political discourse is an active area of research in the field of political science, sociology, linguistics and communications.

To begin with, we will learn the definition of political discourse from different sources. The monograph published at the Belarusian State University clearly defines the difference between text and discourse.

Despite the breadth of the approach, the concept of “discourse” is narrower than the concept of “text”. Discourse is an activity, a phenomenon and a function at the same time. But in discourse, activity is narrowed down to its socially-oriented speech manifestations.

The following definitions are proposed as a working hypothesis.

A political text is a verbalized political activity in all its manifestations: both iconic/symbolic (nominative and accumulative activity) and unfamiliar (performative texts). This concept covers the thematic scope and stylistic features of political activity implemented in the language and by means of the language.

Political discourse is a set of political discourses of society: the discourse of power, counter-discourse, public rhetoric, consolidating the existing system of public relations or destabilizing it [I.F. Ukhvanova-Shmygova, 1998, p. 12].

Political discourse in the broadest sense is a discourse in which any speech formations, subject, addressee or their content belong to the sphere of politics [Sheigal, 2004, p. 28].

Analyzing quite a lot of sources, we stopped at this definition of this term: Political discourse can be defined as a system of texts, symbols, ideas and practices used in the political field to express and organize sociopolitical realities. It includes linguistic and non-linguistic elements such as images, symbols and gestures that shape political reality and influence public opinion.

This topic was chosen because its relevance is now very high in our state. Relevance is aimed at solving the applied problem of determining the sources of political discourse, the mood of discussion in these sources, identifying hotbeds of hatred, negativity, hostility or, conversely, identifying political events, phenomena, discussions perceived by society. The solution of this urgent task has a socio-economic impact in terms of the implementation of the concept of a hearing state, and advanced artificial intelligence methods used to solve the tasks set have a positive impact on the level of scientific and technological

development of the Republic of Kazakhstan. This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19679847).

Purpose of the research: Development of methods for analyzing political texts on the Internet in the Kazakh language in order to identify official and informal information sources of political discourse, as well as determining the mood of discussion in these sources, identifying hotbeds of hatred, negativity, hostility or, conversely, highlighting political events, phenomena, discussions that are perceived by society most positively.

Now let's look at the process of work and the results of scientific work to achieve the goal.

### **Related works**

Several approaches and factors need to be taken into account when building a structured semantic database of names connected to political speech. The following methods can be used to figure out how such a database is structured: entity recognition and extraction, ontology development, database schema design, normalization and data cleaning, linking and cross-reference, taxonomies and hierarchies, metadata and annotations, geospatial dimensions and so on.

For the first time, a semantic knowledge base with roughly 100 semantic traits has been developed for Kazakh proper names [Yelibayeva, 2023, 4154]. Everyone can conduct numerous semantic searches on names in an application using semantic characteristics. [Bekmanova, 2023, 191–205] deals with the creation of an ontological model of words in Kazakh public political discourse and texts of public speeches. To ascertain the tone of the discussion in these sites, an emotional analysis of political dialogue in Kazakh social networks was conducted. Identifying the tone of conversations at the hubs of political discourse is one of the key challenges. The work on sentiment analysis of texts [Yergesh, 2019, p. 9–15; Bekmanova, 2021, p. 170–178; Bekmanova, 2022, p.1937], the development of a morphological analyzer [Zhetkenbay, 2016, p. 257–263; Bekmanova, 2017, p. 20–30], and a number of the authors' previous studies [Bekmanova, 2019, p. 717–730] on the detection of terrorist threats in social networks are prerequisites for this investigation.

## Analysis of Internet Resources on political discourse

Given the large amount of material available online, including news articles, opinion pieces, social media posts, blogs, and political party websites, TV channel websites, analyzing internet resources on political discourse calls for a critical and discerning approach. How to efficiently assess political discourse from online sources is provided here:

*Source evaluation.* By evaluating the authority of the website or platform that hosts the material, the authors have recognized reliable news sources, academic institutions, governmental websites, or individual blogs. Most of the time, reliable sources are more trustworthy. Such sources, where there is a political discourse in the Kazakh language, include the following resources:

- Online publication «Informburo.kz» [11]
- Online publication «Internet resource Tengrinews» [12]
- Online publication «www.kazpravda.kz» [13]
- Online news agency «Sputnik» [14]
- News agency «BAQ.KZ» [15]
- International news agency «Kazinform» [16]
- Egemen Qazaqstan [17]
- Kznews [18]
- Information agency ElKz [19]
- News agency Minber.kz [20]

An information, social and political portal The Qazaq Times [21]

Moreover, we have chosen these resources with the author(s) of the content by checking their qualifications, expertise, and potential biases.

*Bias and Objectivity.* By examining the language and tone employed in the resource, we were able to ascertain the political slant or affiliation of the source. Is it emotionally charged and lopsided, or does it seem objective and balanced? Political party websites refer to online platforms or web pages created and operated by political organizations or parties. There are the following parties in Kazakhstan that have web pages:

- Official site of the party «AMANAT» [22];
- Official website of the Nation's Party of Kazakhstan [23];
- Democratic Party of Kazakhstan PC «Ak Zhol» [24];
- «Auy!» Nation's Democratic Patriotic Party [25];
- «Baitaq» green party [26].

*The relevance and date.* By asking “is it up-to-date and relevant to the current political discourse?” We have verified the resources' publication date. Since political environments are subject to quick change, recent knowledge is frequently more useful.



**Audience and Intent.** The authors have taken into account whom the material is designed for. Is it aimed at a particular racial or ethnic group, political party, or ideology? Knowing the target audience can help us gain understanding of their motivations and biases. Investigate the motivation behind the resource. Is it intended to educate, convince, amuse, or provoke? Everyone can determine the credibility of a source by comprehending the author’s objectives.

Political science journals are primarily intended for academic and research purposes, targeting specialized readers within the field of political science, while newspapers are aimed at the general public with the intent to inform, engage, and influence political discourse in society. Both journals and newspapers [17, 27-34] play important roles in shaping political discussions, but they do so with different audiences and levels of depth.

**Cross-Reference.** To substantiate assertions and compile various viewpoints on the subject, contrast the material with that from other reliable sources. It may be possible to get a more complete picture from several sources.

| List of sources   |               |   |                      |                                |                  |         |   |
|---|---------------|---|----------------------|--------------------------------|------------------|---------|---|
| Title information   | Resources     | Links   | Election advertising | Speech of political candidates | Election debates | Authors | Comments  |
| Қызық екен бе, бұл кісінің қызықтары: Топан                                       | imgnews.kz    | https://kz.imgnews.kz/kozhimbe...<br>era/104e-ade-104e-0000010a-<br>b0b0d4f-104e-145110                                   |                      |                                | 1                | Laura   | Searched for "election" on the site             |
| Президенттіңе ұяйық Түркістан облысында қол жеткізгеніңіз                         | imgnews.kz    | https://kz.imgnews.kz/kozhimbe...<br>era/104e-ade-104e-0000010a-<br>0b1a104e-b0f-104e-b0a0d10                             | 1                    |                                |                  | Laura   | Searched for "election" on the site             |
| Топан қызық екен бе, бұл кісінің қызықтары: Түркістан облысында қол жеткізгеніңіз | imgnews.kz    | https://kz.imgnews.kz/kozhimbe...<br>era/104e-ade-104e-0000010a-<br>0b1a104e-b0f-104e-b0a0d10                             |                      |                                | 1                | Laura   | Searched for "election" on the site             |
| Президенттіңе ұяйық Түркістан облысында қол жеткізгеніңіз                         | imgnews.kz    | https://kz.imgnews.kz/kozhimbe...<br>era/104e-ade-104e-0000010a-<br>0b1a104e-b0f-104e-b0a0d10                             |                      |                                | 1                | Laura   | Searched for "election" on the site             |
| Сайлану 2023: Партиялар ұйымдастырып отырған сайлау кезеңі                        | www.inform.kz | https://www.inform.kz/az/news/2023-<br>partylar-uyymdastyryp-otyrghan-<br>saylau-kezegi-40199914                          | 1                    |                                |                  | Art     | Found according to the date of political events |
| Сайлану 2023: Партия өкілдері өмірлері туралы жыр                                 | almaty.tv     | https://almaty.tv/kezarsa-af%201811-<br>af%201812-20231810-segim-2023-<br>partiya-okilder-ozirdegi-azalyq-dear            | 1                    | 1                              |                  | Art     | Found according to the date of political events |
| Сайлану 2023: Партиялар қалайда бағдарламаларын таныстырып жүр                    | imgnews.kz    | https://imgnews.kz/ke-nig-<br>koncert-keleshe-14770-azalyq-2023-<br>partylar-halya-badep-lamalaryn-<br>tanystyryp-dur-hal | 1                    |                                |                  | Art     | Found according to the date of political events |

Figure 1 – Analysis of contextual comprehension

**Critical Analysis.** Critical thinking techniques were also used to challenge presumptions, spot logical fallacies, and spot any contradictions or inconsistencies in the material.

Within the contemporary information society, information publications and agencies assume pivotal roles as fundamental constituents of the media industry, fulfilling crucial functions in the propagation of news, information, and analytical content. Among the resources in which political discourse is conducted in the Kazakh language, the following sources can be identified:

- Information resource ATPress.kz [35];
- Alash ainyasy [36];
- Radio Azattyq [37].

*Multiple sources.* The authors have developed a comprehensive understanding of the political issue or topic by looking for a variety of online information from various sources and perspectives.

The websites of television channels [11, 38–39] also serve as a valuable academic resource for getting acquainted with political discourse and obtaining information related to this subject. In addition, individuals have the option to tune in to full-length election debates on these television channels.

Being an informed user of internet resources is crucial for participating in political debate in the age of a wealth of online information. While navigating the digital world, it's essential to exercise skepticism and critical thinking in order to make educated decisions about the legitimacy and dependability of online content.

## Creation of a text corpus of political discourse in the Kazakh language

To analyze political processes and opinions in Kazakhstan, initial work was carried out to create a text corpus of political discourse in the Kazakh language.

The result of the TFP «BR11765535 Development of scientific and linguistic foundations and IT resources for expanding the functions and improving the culture of the Kazakh language» will be used, namely, the synonymizer of standard samples of synonymous series of words in the texts of socio-political discourse and public speech (Figure 2),

| Word      | Part of speech | Status      | Meaning   | Example   | Synonym1                    | Synonym2       | Synonym3           | Synonym4                | Periphrase1                 | Periphrase2   | Periphrase3                        |
|-----------|----------------|-------------|---|---|-----------------------------|----------------|--------------------|-------------------------|-----------------------------|---|------------------------------------|
| сайлау    | noun           | homonym     | ақпаратты меніңгісі белгілі бір уақытта басқару және шешім қабылдау қызметтерін жүзеге асыратын адамдарды сайлау жүйесіндегі с-қозғалыстың қимылғы. | Сайлау бірден-бір аяқ бұйымдары аяқталған жоқ.  | таңдау                      | қолдаушылар    | шешім              | өлкіне                  | Негізгі дауыс беру          |   |                                    |
| сайлау    | verb           | homonym     | сайлаудың немесе сайлаудың ресми және ұйымдасқан процесі  | Қалаған үлгілерді сайлау  | дауыс беру                  | қолдау         | ойластыру          |                         |                             |   |                                    |
| сайлау    | adjective      | homonym     | адамдай, сайлаушы   | Тәрбағатай сипемдері сайлау, шұқмақ ойпаттарына ең қымбаттысын жасырып тастағандай (П.Бөкеев, Өз аталары... 17)   | әзірлеу                     | дәрілеу        | жабынсау           |                         |                             |   |                                    |
| тораға    | noun           | unambiguous | Мәңгілікті, министрлікті басқарушы ресми тұлға.   | Палаталарды мемлекеттік тілді қорғау мәселесіндегі депутаттарының арасынан (Палаталар депутаттары жалпы санының көпшілігі даусымен жасырып дауыс беру арқылы Сметпен Мәңгілікті сайлаған торағалар басқарды (ҚР Конституциясы). | ресми тұлға                 | басқарушы      | бастық             | директор                | басшы                       | ең бірінші басшы                                    | министрлікті басқарушы ресми тұлға |
| делегация | noun           | unambiguous | қандай да бір ұйымның немесе мемлекеттің форум, съезд, конференция, халықаралық кеңестерде және т.б. нұсқасын қорғайтын тұлғалар тобы (делегаттар). | Бүгінгі экономикалық кеңістікті құрушы төрт мемлекет делегациясы өкілдерінің қатысуымен жетпісін рет Астанада өтті (ЖИАС-Алматы).   | сайлаған топ                | сайлаған құрам | сайлаған орта      | көрсетілген ердің ұйымы | делегаттар тобы             | Мемлекет я ұйым атқаны сайлаған ресми адамдар тобы. |                                    |
| облыс     | noun           | unambiguous | Кез келген республика құрамындағы ең өзгешелік-құрылымдық бөлік, аймақ.   | Меншік байлық шетімен Тәрбағатай баурайында өткен болатын.  | өңірлік-территориялық бөлік | аймақ          | белгісіз бір бөлік | ағырап                  | өңірлік-территориялық бөлік | аймақ   |                                    |

Figure 2 – A fragment of the synonymizer of standard samples of synonymous series of words in the texts of socio-political discourse and public speech

the volume of dictionary entries of the synomizer obtained in the project is 1000 dictionary entries [Bekmanova, 2023, 191–205] and 2,000 more dictionary entries on election topics will be expanded: «Election advertising», «Speech of political candidates», «Election debates».

This table has the following columns:

- **Word:** This column indicates the specific word for which the synonymy analysis is performed.
- **Part of Speech:** This column shows the part of speech to which a given word belongs (for example, noun, verb, adjective, etc.).
- **Status (homonym, unambiguous, polysemous):** This column indicates the status of the word in the context of synonymy. It can be a homonym (have several different meanings), unambiguous (have only one meaning) or multi-valued (have several meanings, but not a homonym).
- **Meaning:** This column describes the meaning or meanings of the word that are considered in this table.
- **Example:** Here is an example of the use of a word in the context of socio-political discourse or public speech.
- **Synonym:** This column contains information about synonyms for this word, if any. Synonyms are words with a similar or similar meaning.
- **Periphrasis:** This column indicates the periphrases for this word. Periphrases are expressions that are used to convey synonymous meaning, but using other words or phrases.

This table is intended for the analysis and comparison of synonyms in texts related to socio-political discourse and public speech, which can be useful for studying and analyzing the language used in this area.

## Conclusion

As a result of the research work at the moment, sources for the text corpus have been found and a text corpus has been created.

But since the goal of the project is to develop methods for analyzing political discourse in social networks in the Kazakh language in order to identify official and unofficial information sources of political discourse, as well as to determine the mood of discussion in these sources, we plan to work harder and further.

To achieve this goal, the following tasks will be solved in the future: creation of ontological models on election topics: “Election advertising”, “Speech of political candidates”, “Election debates”, cre-

ation of knowledge bases with semantic features, formal recording of logical rules for inference from knowledge bases, creation of a processor for processing official and unofficial information sources of political discourse, creation of a sentiment analyzer (software application) of official and unofficial information sources of political discourse based on the analysis of the sentiments of texts in the Kazakh language.

#### LIST OF REFERENCES

1. Methodology of political discourse research: Actual problems of meaningful analysis of socio-political texts. Issue 1 / Under the general editorship of I.F. Ukhvanova-Shmygova. – Mn.: Belgosuniversitet, 1998. – 283 p.
2. Sheigal, E. I. (2004). Semiotics of political discourse. Moscow: Gnosis.
3. Yelibayeva, G., Orynbay, L., Bekmanova, G., & Sairanbekova, A. (2023). Proper names knowledge base for intelligent mobile application. *Journal of Theoretical and Applied Information Technology*, 101(11).
4. Bekmanova, G., Yergesh, B., Ukenova, A., Omarbekova, A., Mukanova, A., & Ongarbayev, Y. (2023, June). Sentiment Processing of Socio-political Discourse and Public Speeches. In *International Conference on Computational Science and Its Applications* (pp. 191–205). Cham: Springer Nature Switzerland.
5. Yergesh B., Bekmanova G., Sharipbay A. Sentiment analysis of kazakh text and their polarity. *Web Intelligence*, 2019. 17(1), pp. 9–15. doi:10.3233/WEB-190396
6. Bekmanova G., Yergesh B., Sharipbay A. Sentiment analysis model based on the word structural representation 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings. Pp 170–178. doi: 10.1007/978-3-030-86993-9 16.
7. Bekmanova G., Yergesh B., Sharipbay A., Mukanova A. Emotional speech recognition method based on word transcription. *Sensors*, 22(5) doi:10.3390/s22051937.
8. Zhetkenbay L., Sharipbay A., Bekmanova G., Kamanur U. Ontological modeling of morphological rules for the adjectives in kazakh and turkish languages. *Journal of Theoretical and Applied Information Technology*, 91(2). 2016. Pp. 257–263.
9. Bekmanova G., Sharipbay A., Altenbek G., Adali E., Zhetkenbay L., Kamanur U., Zulkhazhav A. A uniform morphological analyzer for the kazakh and turkish languages. Paper presented at the CEUR Workshop Proceedings. 2017. Pp. 20–30.
10. Bekmanova, G., Yelibayeva, G., Aubakirova, S., Dyussupova, N., Sharipbay, A., & Nyazova, R. (2019). Methods for analyzing polarity of

the Kazakh texts related to the terrorist threats. In Computational Science and Its Applications–ICCSA 2019: 19th International Conference, Saint Petersburg, Russia, July 1–4, 2019, Proceedings, Part I 19 (pp. 717–730). Springer International Publishing.

11. <https://informburo.kz/>
12. <https://tengrinews.kz/>
13. <https://kazpravda.kz/>
14. <https://sputnik.kz/>
15. <https://baq.kz/>
16. [www.inform.kz](http://www.inform.kz)
17. <https://egemen.kz/>
18. <https://kznews.kz/>
19. <https://el.kz/>
20. <https://www.minber.kz/>
21. <https://qazaqtimes.com/>
22. <https://amanatpartiasy.kz/>
23. <https://halykpartiyasy.kz/>
24. <https://akzhol.kz/ru>
25. <https://auyl.kz/>
26. BAYTAQ
27. <http://kalamendala.kz/>
28. <https://arka-azhary.kz/>
29. <https://dknews.kz/>
30. <https://almaty-akshamy.kz/>
31. <https://ser-per.kz/>
32. [maqat-tyynysy.kz](http://maqat-tyynysy.kz)
33. <https://aqtobegazeti.kz/>
34. <https://kazgazeta.kz/>
35. <https://atpress.kz/>
36. <https://alashainasy.kz/>
37. <https://www.azattyq.org/>
38. «Qazaqstan»
39. [Almaty.tv](http://Almaty.tv)

УДК 81'33:811.512.145

**РУССКО-ТАТАРСКИЙ МАШИННЫЙ ПЕРЕВОДЧИК:  
ПОДГОТОВКА ДАННЫХ ДЛЯ ЗАПОЛНЕНИЯ  
БД РУССКО-ТАТАРСКИХ ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ**

*Мадехур Аюпов*

*Академия наук Республики Татарстан,  
Казанский федеральный университет, Казань, Россия,  
madehur@mail.ru*

В статье описываются разные методы сбора и обработки данных для заполнения БД русско-татарских параллельных текстов. Полученные данные используются для обучения русско-татарского машинного переводчика.

**Ключевые слова:** русско-татарский машинный переводчик, база данных, татарский язык.

**RUSSIAN-TATAR MACHINE TRANSLATOR:  
PREPARATION OF DATA FOR FILLING THE DB  
OF RUSSIAN-TATAR PARALLEL TEXTS**

*Madehur Ayupov*

*Tatarstan Academy of Sciences,  
Kazan Federal University, Kazan, Russia,  
madehur@mail.ru*

The article describes different methods of collecting and processing data to fill the database of Russian-Tatar parallel texts. The obtained data is used to train a Russian-Tatar machine translator.

**Keywords:** Russian-Tatar machine translator, data base, Tatar language.

## 1. Введение

Русско-татарский машинный переводчик «Татсофт» (translate.tatar) [Khusainov, 2020] является одним из первых нейросетевых машинных переводчиков, где перевод осуществляется с «пониманием» всего предложения, а не отдельных слов или фраз. В настоящее время «Татсофт» является лучшим по качеству перевода среди своих аналогов (Яндекс и Google). Разработка такой системы требует наличия большого объема параллельных предложений (десятки, сотни миллионов пар предложений). В отличие от крупных распространённых мировых языков таких, как англий-

ский, китайский, татарский язык относится к малоресурсным языкам и в связи с этим при создании русско-татарского машинного переводчика важной и самой трудоемкой частью является задача заполнения базы данных русско-татарских параллельных текстов. Для решения данной задачи нами выработаны различные методы и технологии расширения, сбора и обработки данных.

Накопление базы данных русско-татарских параллельных текстов для обучения машинного переводчика включает в себя следующие этапы:

- сбор текстов на русском и татарском языках,
- разбиение текстов на предложения и выравнивание параллельных предложений,
- проверка на правильность перевода параллельных предложений,
- нахождение и исправление пунктуационных, грамматических, синтаксических и других ошибок в предложениях.

## **2. Сбор текстов на русском и татарском языках**

На первом этапе реализации данной работы было необходимо собрать электронные тексты на одном языке и их перевод на другом языке. Одним из способов является оцифровка книг, журналов и иной печатной продукции. Анализ существующей ситуации показал, что у большинства печатной продукции отсутствует электронная версия, а если она и существует, то электронная версия является неполной. Или, если произведение имеет электронную версию на одном языке, то перевод на другой язык имеет только бумажную форму. Поэтому возникла необходимость работы с бумажными источниками. Далее рассмотрим этапы этой работы.

На первом шаге необходимо было найти отсканированную версию нужного произведения или, если нет версии с хорошим качеством, отсканировать произведение.

На втором шаге произведение распознается с помощью специальных программ распознавания.

На третьем шаге в полуавтоматическом режиме удаляется ненужная в дальнейшем информация, например, рисунки, номера страниц.

В рамках данной работы оцифровано 283926 страниц текста.

Второй способ пополнения электронных параллельных текстов – это использование имеющихся интернет ресурсов. При

обращении к Интернету, следует помнить, что материал, представленный в нем, очень неоднороден. Перед тем как анализировать те или иные источники из Интернета, следует убедиться в их надежности, аутентичности. Для выполнения этой работы необходимо было разработать программное обеспечение для автоматического обхода Интернет-ресурсов, содержащих тексты с одинаковым содержимым на татарском и русском языках. Программное обеспечение должно выполнять следующие задачи:

- получение текстов на татарском, русском языках,
- установление между текстами связей типа «является переводом на татарском языке» и «является переводом на русском языке» соответственно.

В рамках данной работы собрано более 617 754 связанных пар страниц на русском и татарском языках.

Третий способ пополнения электронных текстов – это искусственное увеличение исходного набора татарских предложений. Для этого модернизировался межтюркский морфопереводчик с максимальным учетом и поддержкой всех языковых явлений татарского языка и доработкой для башкирского и крымскотатарского языков. В итоге разработанный башкирско-татарский, татарско-башкирский, татарско-крымскотатарский и крымскотатарско-татарский машинный переводчик, который работает на правилах, создает возможность увеличения электронных корпусов на тюркских языках. Полученные дополнительные электронные корпуса используются для обучения машинного переводчика, что помогает улучшить качество татарско-русского и русско-татарского машинного перевода.

## **2. Разбиение текстов на предложения и выравнивание параллельных предложений**

Большинство русско-татарских параллельных текстов не дают возможности непосредственного извлечения информации о переводных соответствиях. Во-первых, между переводом и текстом оригинала не существует однозначного соответствия на уровне слов, имеются различия грамматической структуры, лексическая неоднозначность. Во-вторых, одному предложению оригинала может соответствовать несколько предложений перевода, и наоборот. Наконец, имеются неточности перевода, среди которых наиболее существенными являются пропуски [Хакимов, 2021].



Поэтому разбиение текстов на предложения представляет собой не столь тривиальную задачу. В наиболее простом подходе для деления текста на предложения используются синтаксические признаки конца и начала, в более сложных подходах применяются методы, использующие знания о лексике конкретного языка – списки сокращений и др.

Во время выравнивания параллельных предложений, распределение слов в предложениях используется как основной источник информации при установлении лексических соответствий. От точности выравнивания предложений зависит успех работы машинного переводчика.

Для выравнивания параллельных текстов на первом шаге использовался инструмент АВУУ Aligner 2.0 – программа, которая находит соответствующие друг другу предложения в текстах на разных языках, сопоставляет их между собой и автоматически создает выравненные сегменты. После автоматического выравнивания, на втором шаге, производится ручное редактирование выравненного текста для исправления возможных ошибок выравнивания и улучшения качества результата.

В рамках данной работы всего выравнено 954829 параллельных предложений.

### **3. Проверка на правильность перевода параллельных предложений**

Качественный перевод должен в точности передавать смысл оригинала. Кроме стандартной проверки, для получения хорошего качества необходимо учитывать следующие моменты:

- термины должны иметь одинаковый перевод по всем текстам,
- если существует несколько вариантов написания слов, они должны быть приведены к единому написанию,
- необходимо соблюдать правила употребления неразрывных пробелов в нужных местах,
- написание чисел должно соответствовать правилам конкретного языка и др.

Так же необходимо устранить пунктуационные, грамматические и орфографические ошибки, чтобы убедиться, что в параллельных предложениях нет ошибок.

Все эти работы проводились в ручном режиме.

#### **4. Проблемы, возникающие при подготовке данных**

Во время накопления русско-татарских параллельных текстов возникает множество проблем. Рассмотрим некоторые из них.

При распознавании отсканированных текстов с помощью специальных программ распознавания возникают орфографические ошибки, если встречаются неуверенно распознанные слова и слова, отсутствующие в словаре программы распознавания. Отсутствие пробелов между словами – также распространенный случай при работе с распознанным текстом. Такие ошибки автоматически исправлять не получается, их выявлять и исправлять можно только во время ручного просмотра распознанного текста. Такое редактирование необходимо осуществить до выравнивания параллельных предложений с помощью инструмента АBBYU Aligner 2.0, так как вышеперечисленные ошибки влияют на качество выравнивания.

Для обучения машинного переводчика нужны качественные параллельные тексты, но как показывает анализ ситуации, требование «качественные» не всегда удовлетворяется хорошо, особенно это касается информации взятой из Интернета. Иногда попадаются такие пары предложений, у которых мало или ничего общего. Поэтому, для создания качественного машинного переводчика, приходится базу данных русско-татарских параллельных текстов чистить от таких плохих пар.

#### **5. Заключение**

Выработанные нами различные методы и технологии расширения, сбора и обработки данных, позволили получить достаточно объемную и качественную базу данных русско-татарских параллельных текстов, что позволило русско-татарскому машинному переводчику «Татсофт» на сегодняшний день стать лучшим по качеству общедоступным переводчиком в русско-татарской языковой паре среди своих аналогов (Google, Яндекс).

В дальнейшем планируется продолжить работу по увеличению объема базы данных русско-татарских параллельных текстов, что в итоге даст прирост в качестве работы русско-татарского машинного переводчика.

## ЛИТЕРАТУРА

1. Khusainov A., Suleymanov D., Gilmullin R. (2020) The Influence of Different Methods on the Quality of the Russian-Tatar Neural Machine Translation. In: Kuznetsov S.O., Panov A.I., Yakovlev K.S. (eds) Artificial Intelligence. RCAI 2020. Lecture Notes in Computer Science, vol 12412. Springer, Cham. Pp. 251–261. [https://doi.org/10.1007/978-3-030-59535-7\\_18](https://doi.org/10.1007/978-3-030-59535-7_18).
2. Хакимов Б.Э., Шаехов М.Р. Проблема эквивалентности параллельных предложений в тестовом корпусе для русско-татарского машинного переводчика // Proceedings of the 9th International Conference on Turkic Languages Processing (TURKLANG-2021). (Tyva, September 21-23, 2021). – Tyva, 2021.

УДК 81'322.2

ТЕЗАУРУС ПО МАТЕМАТИКЕ СРЕДНИХ ШКОЛ  
НА КАЗАХСКОМ ЯЗЫКЕ*А. А. Шарипбаев, А. К. Альжанов, С. А. Нариман,  
Г. Ж. Ахметова**Евразийский национальный университет имени Л.Н.Гумилева,  
Астана, Казахстан*sharalt@mail.ru, alzhanov\_ak@mail.ru, saniya\_khairova@mail.ru,  
gulya\_kish@mail.ru

В данной работе рассматриваются вопросы применения интеллектуальных технологий в системе образования, как один из наиболее эффективных путей развития сферы образования. Внедрение интеллектуальных технологий в образовательный процесс приводит к достижению качественно новых образовательных результатов. Представлена модель тезауруса по математике средних школ на казахском языке. Работа выполнена в рамках проекта AP19678613 «Разработка технологии создания интеллектуальных учебников, способных осуществлять интерактивное обучение, консультирование и оценку знаний по предметам, преподаваемым на казахском языке».

**Ключевые слова.** Интеллектуальные технологии, тезаурус, математика.

THESAURUS OF MATHEMATICS IN SECONDARY SCHOOLS IN  
THE KAZAKH LANGUAGE*Altynbek Sharipbay, Aitugan Alzhanov, Sania Nariman,  
Gulzhan Akhmetova**L. N. Gumilyov Eurasian national university  
Astana, Kazakhstan*sharalt@mail.ru, alzhanov\_ak@mail.ru, saniya\_khairova@mail.ru,  
gulya\_kish@mail.ru

This paper discusses the use of intelligent technologies in the education system, as one of the most effective ways to develop the education sector. The introduction of intelligent technologies into the educational process leads to the achievement of qualitatively new educational results. A model of thesaurus for secondary school mathematics in the Kazakh language is presented. The work was carried out within the framework of the project AP19678613 “Development of technology for creating intelligent textbooks capable of interactive learning, counseling and assessment of knowledge in subjects taught in the Kazakh language.”

**Keywords.** Intellectual technologies, thesaurus, mathematics.

**Введение.** С течением времени мир меняется, и с ним меняются и предпочтения молодежи. Интеллектуальные технологии становятся все более важными в образовании, включая обучение школьных предметов на казахском языке.

Обучение предмету на казахском языке, а именно математики становится вызовом при обучении представителей других национальностей.

1. Применение интеллектуальных технологий в обучении школьного предмета на казахском языке способствует:

- совершенствованию практических навыков и умений, позволяет эффективнее организовать самостоятельную работу и индивидуальный процесс обучения школьников, повышает интерес к занятиям; активизирует познавательную деятельность студентов, развивает интеллектуальные творческие способности.

- использование достижений в области интеллектуальных технологий позволяет облегчить работу учителя;

- интенсифицировать образовательный процесс, повысить у учащихся мотивацию к обучению;

- осуществить индивидуальный подход к обучению;

- повысить эффективность и качество образования на казахском языке

В настоящее время в образовании активно используются Интеллектуальные технологии, такие как интерактивные доски, телевидение и интеллектуальные учебники. Важно следовать тенденциям и внедрять инновационные методы обучения математике, чтобы обеспечить школьников качественными знаниями [McLaren, 2008; McLaren, 2016; McLaren].

**2. Описание работы.** Основные задачи интеллектуальных технологий обучения математике на казахском языке включают:

- Особенности внедряемых технологий: эффективность инноваций зависит от характеристик самих технологий, их соответствия образовательным целям и потребностям обучающихся.

- Потенциал новаторов: роль учителей и других инициаторов инноваций важна. Их знания, умения и мотивация способствуют успешному внедрению инноваций.

- Пути внедрения инноваций: важным аспектом является выбор и организация методов внедрения инноваций в учебный процесс, включая обучение педагогов и стимулирование изменений в образовательных практиках.

Интеллектуальные технологии обучения – это особый вид профессиональной деятельности, требующий использования боль-

шого педагогического опыта. Они требуют от педагогов творческого подхода и личностного отношения к работе. Каждый педагог, осваивая Интеллектуальные технологии, также занимается саморазвитием и самосовершенствованием. В современном мире подготовка будущих специалистов и формирование их профессиональных навыков является важным требованием общества, не терпящим отлагательств. Поэтому мы разработали модель формирования у педагогов-математиков умений применения интеллектуальных технологий обучения математике [McLaren, 2011].

Формирование навыков применения интеллектуальных технологий обучения математике для будущих специалистов охватывает следующие сферы:

1. Подготовка к использованию новых образовательных технологий, включая изучение современных методик и подходов.

2. Развитие творческого мышления и способности к инновациям, чтобы смело экспериментировать и применять новые идеи.

3. Укрепление педагогической компетентности и умений адаптировать инновационные методы к конкретным образовательным ситуациям.

4. Развитие навыков цифровой и информационной грамотности для эффективного использования современных технологий в образовании.

5. Создание условий для саморефлексии и самоанализа, чтобы педагоги могли постоянно улучшать свою деятельность.

6. Развитие коммуникационных навыков, так как современные технологии обучения часто требуют коллаборации и взаимодействия.

7. Привитие чувства ответственности и готовности к постоянному профессиональному росту.

Это лишь некоторые аспекты формирования навыков применения интеллектуальных технологий. Развитие таких навыков является ключевым для эффективной работы педагогов в современном обществе [Schiaffino, 2008].

Инновации в учебно-воспитательном процессе можно разделить на три вида: модификационные, комбинаторные и радикальные.

1. Модификационные инновации основаны на улучшении и изменении уже существующей формы обучения. Это могут быть различные усовершенствования, модернизация методов и техник преподавания. Примером модификационной инновации может

быть разработка опорных конспектов по математике, которые применяются учителями и позволяют улучшить процесс обучения.

2. Комбинаторные инновации заключаются в создании новых сочетаний известных методических элементов. Это означает использование комбинации различных методов и подходов в обучении, которые ранее не применялись вместе. Например, современные методики преподавания часто комбинируют традиционные методы, интерактивные подходы, использование информационных технологий и другие элементы.

3. Радикальные инновации связаны с глубокими изменениями в системе образования, такими как внедрение государственных стандартов образования. Государственные стандарты определяют параметры и показатели уровня и качества обучения, обновляют содержание образования и ориентируют систему на мировые тренды. Эти изменения вносят основательные перемены в учебно-воспитательный процесс [Cheung, 2003].

В современном образовательном пространстве развитие интеллектуальных возможностей, способностей и талантов обучающихся становится одной из главных задач. В этом контексте использование интеллектуальных педагогических систем и технологий, включающие мультимедийные занятия, играет важную роль.

Занятия с использованием интеллектуальных технологий представляют собой форму обучения, в которой используются различные средства коммуникации, такие как текст, изображения, аудио, видео и анимация. Цель таких занятий заключается в использовании компьютерных технологий для развития навыков самостоятельного обучения, освоения материала и познавательной деятельности, а также для стимулирования самостоятельной работы и развития творческих способностей школьников [McLaren, 2015; Adams].

Применение интеллектуальных технологий по математике имеет несколько преимуществ.

Во-первых, они способствуют повышению интереса обучающихся к предмету математика и активизации их вовлеченности в учебный процесс. Во-вторых, мультимедийные занятия могут визуализировать сложные концепции и идеи, делая их более понятными и доступными для школьников. В-третьих, они позволяют обучающимся развивать навыки работы с информацией из разных источников и медиа-ресурсов.

Однако внедрение таких занятий требует квалифицированных преподавателей, которые могут адаптировать содержание урока к

данным средствам обучения и эффективно использовать их в процессе обучения. Также важно обеспечить доступность необходимого оборудования и программного обеспечения для проведения занятий с использованием интеллектуальных технологий.

В целом, применение интеллектуальных технологий по математике представляют собой инновационный подход к обучению, который может способствовать более качественному и интересному учебному процессу, развитию навыков самостоятельности и творческого мышления обучающихся.

Когда применяются интеллектуальные системы обучения математике, выполнение описанных выше условий может быть улучшено и дополнено.

Во-первых, данные занятия способствуют личностной подготовке обучающихся путем стимулирования их самостоятельной работы, поиска необходимой информации и развития навыков самообучения.

Во-вторых, использование интеллектуальных технологий может способствовать совершенствованию интеллектуальных способностей школьников. Мультимедийные презентации, интерактивные задания и визуализация сложных концепций могут помочь студентам лучше понять и усвоить учебный материал.

Третье условие – эстетическое воспитание, также может быть достигнуто с помощью таких занятий. Создание презентаций, видеоматериалов или других мультимедийных проектов позволяет школьникам развивать свою творческую активность и эстетический вкус.

Четвертое условие – формирование объективного подхода, может быть поддержано с помощью занятий с применением интеллектуальных технологий, которые позволяют студентам исследовать различные источники информации, делать сравнительные анализы и формировать собственное мнение на основе разнообразных данных.

Наконец, углубление междисциплинарных связей может быть достигнуто с помощью мультимедийных проектов, которые объединяют разные аспекты знаний и позволяют школьникам исследовать тему с разных точек зрения [McLaren, 2012; Lew, 2020].

Таким образом, применение интеллектуальных технологий на занятиях по математике может способствовать выполнению всех описанных условий, обогащая образовательный процесс и развивая навыки школьников.



Математика является одним из фундаментальных предметов в школьной программе. Ее изучение позволяет развивать логическое мышление, абстрактное мышление и умение решать разнообразные задачи.

Одним из видов интеллектуальных систем по математике на казахском языке может выступать тезаурус по математике средних школ на казахском языке.

Тезаурус (от греческого «θησαυρος») (thesauros) – сокровище) – в обобщенном смысле это словарь ключевых терминов (понятий), описывающих некоторую предметную область, с указанием семантических отношений (связей) между терминами (понятиями).

В данной статье мы рассмотрим тезаурус по математике средних школ, охватывающий основные понятия, принципы и методы.

Данный тезаурус по математике на казахском языке представляет собой средство организации и структурирования математических понятий, терминов и их взаимосвязей. Он помогает упростить изучение и понимание математических концепций, а также обеспечивает точность и ясность в обмене информацией в мире математики. Тезаурус включает в себя систему схожих терминов, синонимов, антонимов и связанных понятий, что упрощает поиск и обмен знаниями в этой области науки [Ергеш. 2018; Жеткенбай, 2019].

Результаты работы по разработке тезауруса по математике на казахском языке для средних школ будет отражено в отчете по проекту AP19678613.

Тезаурус в математике может быть полезным инструментом для учеников и учителей, помогая им лучше понимать и использовать математические термины. Он способствует систематизации знаний и улучшает коммуникацию в сообществе математиков.

Такой тезаурус включает термины, связанные с разделами – арифметикой, геометрией, алгеброй.

Арифметика – это раздел математики, который изучает числа, операции над ними (сложение, вычитание, умножение, деление) и их свойства. Она помогает нам решать задачи, связанные с подсчетом, измерением и оценкой количества. Арифметика является основой для более сложных математических дисциплин, таких как алгебра, геометрия и тригонометрия.

Геометрия – это раздел математики, изучающий формы, размеры, свойства и отношения объектов в пространстве. Геометрия исследует различные виды фигур, а также способы измерения расстояний, углов, объемов и площадей. Она имеет множество при-

ложений в науке, инженерии, архитектуре и других областях, где важны вопросы о форме и структуре объектов. Геометрия также играет ключевую роль в образовании, помогая учащимся развивать логическое мышление и пространственное воображение.

Алгебра – это раздел математики, который изучает символичные выражения и операции над ними. В алгебре рассматриваются числа, переменные и математические выражения, а также методы их анализа и преобразования. Она включает в себя концепции, такие как уравнения, неравенства, многочлены, функции и многие другие алгебраические структуры. Алгебра играет важную роль в различных областях математики, физики, инженерии и других науках, а также имеет широкое практическое применение в решении проблем из реального мира.

Тезаурус по математике средних школ на казахском языке был основан на терминологии школьных учебников (Таблицы 1, 2).

**Таблица 1. Авторы школьных учебников по арифметике**

| Класс | Авторы учебника «Арифметика»         |
|-------|--------------------------------------|
| 1     | Акпаева А.Б., Жакупова Г.Ш.          |
| 2     | Акпаева А.Б., Оспанов Т.К.           |
| 3     | Акпаева А.Б.                         |
| 4     | Акпаева А.Б.                         |
| 5     | Алдамуратова Т.А., Әбілқасымова А.Е. |
| 6     | Алдамуратова Т.А., Әбілқасымова А.Е. |

**Таблица 2. Авторы школьных учебников по алгебре и геометрии**

| Класс | Авторы учебника «Геометрия»              | Авторы учебника «Алгебра»                        |
|-------|--|--|
| 7     | Шыныбеков А. Н., Смиров В. А.,           | Абылкасымова А. Е., Шыныбеков А. Н., Забара Л.М. |
| 8     | Шыныбеков А. Н., Смиров В. А., Солтан Г. | Абылкасымова А. Е., Шыныбеков А. Н., Солтан Г.   |
| 9     | Шыныбеков А. Н., Смиров В. А., Солтан Г. | Абылкасымова А. Е., Шыныбеков А. Н., Солтан Г.   |
| 10    | Шыныбеков А. Н., Смиров В. А., Солтан Г. | Абылкасымова А. Е., Шыныбеков А. Н., Пак О.      |
| 11    | Шыныбеков А. Н., Смиров В. А., Солтан Г. | Абылкасымова А. Е., Шыныбеков А. Н.              |

Тезаурус по математике средних школ на казахском языке будет содержать все термины и их определения по всем разделам школьного курса математики, а также антонимы, синонимы, гиперонимы, гипонимы, меронимы и холонимы этих определений. Также будут включены содержательные вопросы по всем разделам математики для самоконтроля обучающихся.

### 3. Заключение

Таким образом, использование интеллектуальных технологий в образовании на казахском языке может приносить значительные преимущества. Эти преимущества включают:

- Совершенствование практических навыков и умений: Технологии позволяют студентам более активно взаимодействовать с учебным материалом, выполнять задания и учиться на практике.

- Организация самостоятельной работы: Системы электронного обучения и онлайн-ресурсы позволяют студентам изучать материалы в своем темпе и вне учебных аудиторий.

- Повышение интереса к занятиям: Интерактивные методы обучения, мультимедийные материалы и обучающие игры могут сделать учебу более увлекательной и интересной.

- Активизация познавательной деятельности: Технологии могут помочь стимулировать аналитическое и критическое мышление студентов.

- Индивидуальный подход: Адаптивные образовательные платформы могут предоставлять персонализированный контент, учитывая потребности и способности каждого студента.

- Повышение эффективности и качества образования: Современные технологии могут помочь учителям более эффективно преподавать и оценивать знания студентов.

- Развитие интеллектуальных творческих способностей: Работа с инновационными технологиями может способствовать развитию креативности и умения решать сложные задачи.

Внедрение таких систем и технологий в образовательный процесс может действительно улучшить качество обучения и подготовку обучающихся к современным вызовам.

### СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. McLaren, B.M., Lim, S., Gagnon, F., Yaron, D., & Koedinger, K.R. Studying the effects of personalized language and worked examples in the

context of a web-based intelligent tutor. In M. Ikeda, K.D. Ashley, & T-W. Chan, Proc. of the 8th Inter.Conf. on Intelligent Tutoring Systems, Lecture Notes in Computer Science, 4053 (pp. 318–328). Berlin: Springer.

2. McLaren, B.M., Lim, S., & Koedinger, K.R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conf. of the Cognitive Science Society (pp. 2176-2181). Austin, TX: Cognitive Science Society.

3. McLaren, B.M., van Gog, T., Ganoë, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in classroom experiments. *Computers in Human Behavior*, 55, 87–99.

4. McLaren, B.M., DeLeeuw, K.E., & Mayer, R.E. (2011). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human Computer Studies*, 69(1-2), 70–79. doi:10.1016/j.ijhcs.2010.09.001.

5. Schiaffino, S., Garcia, P., & Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students. *Computers & Education* 51, 1744-1754.

6. Cheung, B.; Hui, L.; Zhang, J.; Yiu, S. M. (2003). «SmartTutor: An intelligent tutoring system in web-based adult education». *Journal of Systems and Software*. 68: 11–25. doi:10.1016/s0164-1212(02)00133-4.

7. McLaren, B. M., Adams, D. M., & Mayer, R.E. (2015). Delayed learning effects with erroneous examples: A study of learning decimals with a web-based tutor. *International Journal of Artificial Intelligence in Education*, 25(4), 520–542.

8. Adams, D., McLaren, B.M., Mayer, R.E., Gogvadze, G., & Isotani, S. Erroneous examples as desirable difficulty. In Lane, H.C., Yacef, K., Mostow, J., & Pavlik, P. Proc. of the 16th Inter.Conf. on Art. Intelligence in Education. LNCS 7926 (pp. 803–806). Springer, Berlin.

9. McLaren, B.M., Adams, D., Durkin, K., Gogvadze, G. Mayer, R.E., Rittle-Johnson, B., Sosnovsky, S., Isotani, S., & Van Velsen, M.. To err is human, to explain and correct is divine: A study of interactive erroneous examples with middle school math students. In A. Ravenscroft, S. Lindstaedt, C. Delgado Kloos, & D. Hernández-Leo, Proc. of EC-TEL 2012: Seventh Euro Conference on Technology Enhanced Learning, LNCS 7562 (pp. 222-235). Springer, Berlin.

10. Lew, Hc. Developing smart math textbook in Korea. *Afr. Mat.*31, 143–153 (2020). <https://doi.org/10.1007/s13370-019-00732-w>.

11. Ергеш Б.Ж., Шарипбай А.А., Гатиатуллин А.Р. Тезаурус: обзор моделей, применение // Вестник КазНУТУ. – Алматы, 2018. – №4. – С. 208–211.

12. Жеткенбай Л., Шәріпбай А.Ә. Елибаева Г.К., Муқанова А.С., Ергеш Ж.Қазақ және түрік тілдерінің затесімінің онтологиялық моделі // ҚазҰТЗУ Хабаршысы, №3. – 2019. – б. 439–445.

УДК 81'322.2

РЕСУРСЫ СИСТЕМ АВТОМАТИЧЕСКОГО АНАЛИЗА  
ТЕКСТА*М. А. Абжалова<sup>1</sup>, М. А. Адилова<sup>2</sup>**<sup>1</sup>Ташкентский университет информационных технологий  
имени Аль-Хорезми**<sup>2</sup>Ташкентский государственный университет узбекского языка  
и литературы*

abjalova.manzura@gmail.com, adilovamunojot@gmail.com

Анализ текста является одной из наиболее важных проблем обработки естественного языка (NLP). Анализ текста выполняется в рабочем процессе любого программного обеспечения, работающего с текстом. Поэтому очень важно анализировать устные или письменные тексты на естественном языке на основе языковых норм: орфографических, лексических, семантических, грамматических и даже прагматических. В данной статье рассматриваются лингвистические ресурсы, необходимые для лингвистического обеспечения программ для узбекского языка, предназначенных для обработки текста на основе компьютерных технологий.

**Ключевые слова:** обработка естественного языка (NLP), лингвистическое обеспечение, филологические программы, обработка текста, источник.

## RESOURCES OF AUTOMATIC TEXT ANALYSIS SYSTEMS

*Abjalova Manzura Abdurashetovna<sup>1</sup>,  
Adilova Munojot Asrorovna<sup>2</sup>**<sup>1</sup>Tashkent University of Information Technologies  
named after Al-Khorazmi, Tashkent, Uzbekistan**<sup>2</sup>Alisher Navo'i Tashkent State University of Uzbek  
Language and Literature, Tashkent, Uzbekistan*

abjalova.manzura@gmail.com, adilovamunojot@gmail.com

Text analysis is one of the most important issues in natural language processing (NLP). Text analysis is performed in the workflow of any software that works with text. Therefore, it is very important to analyze spoken or written texts in natural language based on language norms: orthographic, lexical, semantic, grammatical even pragmatic. This article deals with the linguistic resources that are necessary for the linguistic support of programs designed for text processing based on computer technologies.

**Keywords:** natural language processing (NLP), linguistic support, philological programs, text processing, source.

## INTRODUCTION

To reproduce texts in any language, in general, to create philological programs, it is necessary to have a certain linguistic resource (LR) in the computer memory. Linguistic provisions and norms form the basis of the provision. Also, the philological vocabulary of a particular language can be learned from LR. Therefore, when creating a linguistic processor, it is necessary to have sufficient knowledge about the natural language being processed. This, in turn, ensures that the linguistic processor is perfect.

When creating a linguistic processor for automatic editing and analysis of texts in the Uzbek language, a linguist needs a collection of linguistic dictionaries and grammatical rules of the Uzbek literary language.

The creation of linguistic modules for automatic text analysis programs lays the foundation for the development of promising and ideal programs that serve for the competent preparation of texts in the Uzbek literary language. Thus, the linguistic base of the automatic text analysis program has been replenished with many sources. One of the main problems in natural language processing (NLP) is the issue of automatic text analysis.

## DISCUSSION

A relatively common type of linguistic dictionaries in foreign Computational Linguistics is a morphological dictionary. In such a dictionary, which word group the lexeme belongs to, the list of its grammatical forms, if the lexeme belongs to an inflected language, the cases of change (for example: “*писать – пишу*”, “*водитъ – возжу*”; “*child – children*”, “*man – men*”) are reflected, and the dictionary helps in the morphological analysis of texts. The creator of the linguistic processor can add grammatical affixes to the dictionary according to his working style.

A morphological dictionary is a lexicographic resource that reflects the type of word form, that is, the type of nouns and the declension of verbs through a special system of conditional marking [Abjalova, 2020; Abjalova, 2021]. The dictionary is arranged alphabetically. In inversion order, the last letter of the word is taken into account. At the beginning of the dictionary, “Grammatical information” is given, which shows the phenomenon of inflection and declension in the form of a sample. Therefore, each word has a grammatical symbol and an index that refers to “grammatical information”. This feature helps the

user to identify the occurrence of a specific word change. In order to speed up and make it easier to search for a sample, at the top or footer of each page of the dictionary, the number of indexes found on the page and information about the grammatical information represented by the indexes is given which page of the dictionary. In order to clearly show the scope and importance of creating a morphological dictionary, the famous Russian lexicographer A.A. A fragment from Zaliznyak's book "Грамматический словарь русского языка" ("Grammar dictionary of the Russian language") was given:

ж 8а —47 | св (нсв) 2 —92

ВАТЬ

|                   |   |                |                |
|-------------------|---|----------------|----------------|
| отретушивать      | св 2а   | стасовать      | св 2а          |
| варьировать       | нсв 2а  | растасовать    | св 2а ⊗I       |
| интерьюировать    | св-нсв 2а   | фасовать       | нсв 2а         |
| проинтерьюировать | св 2а   | расфасовать    | св 2а ⊗I       |
| кровать           | ж 8а  | всовать        | св 2б ⊗I (см.) |
| диван-кровать     | м, склоняются обе части (ков. формы и оп. деления избегаются) | подсовать      | св 2б ⊗I (см.) |
| крёсло-кровать    | с, склоняются обе части (ков. формы и оп. деления избегаются) | колесовать     | св-нсв 2а      |
| соборовать        | св-нсв 2а   | адресовать     | св-нсв 2а      |
| особоровать       | св 2а   | переадресовать | св 2а ⊗I       |
| воровать          | нсв 2а  | пересовать     | св 2б          |
| наворовать        | св 2а ⊗I  | интересовать   | нсв 2а         |
| обворовать        | св 2а ⊗I  | заинтересовать | св 2а ⊗I       |
| разворовать       | св 2а ⊗I  | рисовать       | нсв 2а         |
| поворовать        | св 2а   | зарисовать     | св 2а ⊗I       |
| своровать         | св 2а   | нарисовать     | св 2а          |
| уворовать         | св 2а ⊗I  | обрисовать     | св 2а ⊗I       |
| озоровать         | нсв нп 2а   | подрисовать    | св 2а ⊗I       |
| созоровать        | св нп 2а  | перерисовать   | св 2а ⊗I       |
| селитровать       | св-нсв 2а   | разрисовать    | св 2а ⊗I       |
| нитровать         | св-нсв 2а   | изрисовать     | св 2а ⊗I       |
| титровать         | нсв 2а  | пририсовать    | св 2а ⊗I       |
| оттитровать       | св 2а   | дорисовать     | св 2а ⊗I       |
| секвестровать     | св-нсв 2а   | порисовать     | св 2а          |
| оркестровать      | св-нсв 2а   | прорисовать    | св 2а ⊗I       |
| муштровать        | нсв 2а  | срисовать      | св 2а ⊗I       |
| промуштровать     | св 2а   | вырисовать     | св 2а ⊗I       |
| вымуштровать      | св 2а   | киковать       | нсв нп 2а      |
| фильтровать       | нсв 2а  | скиковать      | св нп 2а       |
| профильтровать    | св 2а ⊗I  | коковать       | нсв 2а         |
| отфильтровать     | св 2а ⊗I  | накоковать     | св 2а          |
|                   |   | буксовать      | нсв нп 2а      |
|                   |   | забуксовать    | св нп 2а       |
|                   |   | силовать       | нсв 2а         |
|                   |   | засиловать     | св 2а          |

A morphological dictionary is also very important in creating modules for the morphological analysis stage of the automatic text analysis program of Uzbek texts. However, due to the lack of such a dictionary, morpheme dictionaries were used as the main source. Such a dictionary was compiled for the first time by A.Gulomov, A.N.Tikhonov, R.K.Kongurov and published by the "O'qituvchi" ("Teacher") publishing house in 1977 [Mengliyev, 2007; Samad, 2007; Tikhonov, 1977].

"An explanatory dictionary of the Uzbek language" [6] is an important auxiliary source in determining which word group the lexemes in the spelling dictionaries belong to. "The morpheme dictionary of

the Uzbek language” was used to determine what affixes a lexeme can attach to itself, and to create combinations of affixes.

It is known that the writer’s language skills are visible, first of all, in the use of synonymous lines. Due to the fact that selecting a suitable, necessary unit from the synonymous line is the most accurate way to accurately express expressiveness and subjective evaluation, it is considered a necessary linguistic tool in composing a text. In addition, there is an opportunity to set a stylistic criterion for the presence of synonymous variants in the language. A. Hojiyev’s “Explanatory Dictionary of synonyms of the Uzbek Language” [Hojiyev, 1974] provides practical help in differentiating these synonymous options.

In the dictionary of borrowed lexemes, lexemes introduced into the Uzbek language from another language are explained. A dictionary in this language is a form of an explanatory dictionary, and they are called differently depending on the language from which they interpret the derived lexeme [Kurbanova, 2021; Usmon, 1965]. Due to the renewal and changes in the technological age, many diminutive words are appearing in the terminology, and the range of scientific and official terms is increasing. Therefore, it is time to release a new edition of the dictionary of diminutive words.

## RESULTS

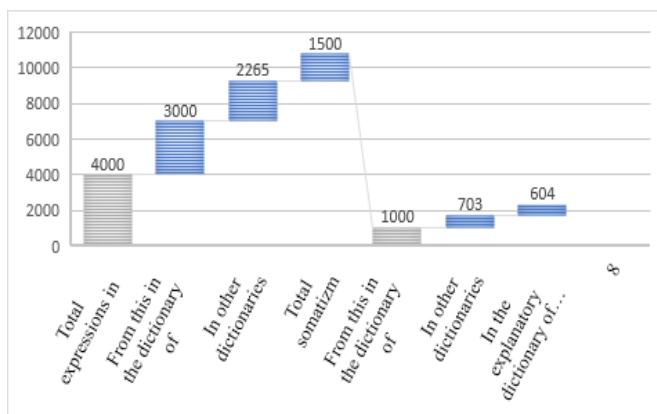
Phraseologisms in Uzbek Language, in essence, are mainly the product of tuning and artistic speech. Limitations specific to other styles appear after a certain period of time. For example, the emergence of a synonymous line of about a hundred euphemisms that mean the meaning of a single pronoun lexeme creates the possibility of their functional limitation. As an example, these words are used in the oral style, “to die”, “*olamdan o‘tmoq*”, “*dunyodan o‘tmoq*”, “*omonatini topshirmoq*”, “*qulog‘i ostida qolmoq*”, “*jon bermoq*” etc, however, “*vafot etmoq*”, “*hayotdan ko‘z yummoq*”, “*dunyodan ko‘z yummoq*”, “*hayot bilan vidolashmoq*” are used in journalistic and official styles. Such things as “*Alloh rahmatiga yo‘l tutmoq*”, “*shahodat sharobini ichmoq*”, “*dorilfanodan dorilbaqoga rixlat bo‘lmoq*” belong to the artistic text. It can be seen that the dictionary of phraseology [Mengliyev, 2007; Rakhmatullayev, 2001] is one of the sources that ensure the perfection of the program.

Although phrases are considered stable word combinations, the syntactic derivation between their components is dynamic in nature. This condition affects only the formal aspect of expressions, and its seman-



tics preserve the integrity of meaning and figurativeness. The syntactic relationship of the components of the combination is primary in the introduction of the phraseological combination into the syntactic relationship in the sentence. For example: a phrase with a predicative sign,  $N(N_0 + N_{e.a.} \Rightarrow \text{noun compound}) + V$ : “*Mening sabr kosam to ‘ldi*”; “*Sabr kosang to ‘lgandir*”; “*Uning sabr kosasi ham to ‘ldi*”; “*Sabr kosamiz to ‘lib bormoqda*”. Therefore, a base of 1393 phrases was created in LR, a phrase was taken as a unit equivalent to a word [**phrase=word**], and the place of change of grammatical forms was determined in them for the morphological analysis module [Абжалова, 2020].

The number of phrases in the Uzbek language Sh.Rakhmatullayev [Rakhmatullayev, 2001], the explanatory dictionary of the Uzbek language (EDUzL) [6] and other dictionaries has the following numbers: total PhU in the Uzbek language - about 4000: of which Sh.Rakhmatullayev’s dictionary has about 3000; in other dictionaries it is 2265. The total number of somatic phrases is about 1,500: of which Sh. Rakhmatullayev’s dictionary contains about 1,000; in other dictionaries it is 703, and in EDUzL it is 604 (Figure 1).



Every language has somatic phrases involving human body parts, and now they are a special object of research in Uzbek linguistics. The number of somatic expressions in the Uzbek language is more than a thousand. The total number of somatic phrases in Sh.Rakhmatullayev’s dictionary is 604, with the following statistics by category: verb (*fe’l*) – 489, noun(*ot*) – 22, adjective (*sifat*) – 60, adverb (*ravish*) – 33 (Figure 1).

In the process of writing the text, there are also cases where lexemes whose pronunciation is close to each other are mistakenly replaced or written according to pronunciation. For example, “*tambur*” instead of “*tanbur*”, “*adl*” instead of “*adil*”, “*tire*” instead of “*teri*”, “*yonilg ‘i*” instead of “*yoqilg ‘i*”, “*sudxo ‘r*” instead of “*sutxo ‘r*”. Most people do not care that the abbreviation “AyoQSh” is an extension of an “*avtomobil yonilg ‘i quyish shoxobchasi*”. Therefore, most people think that as “*avtomobil yoqilg ‘i quyish (yoxud qo ‘yish) shaxobchasi*”. However, “*yonilg ‘i*” is a set of flammable liquids such as gas, kerosene, gasoline; “*yoqilg ‘i*” – a set of wood, coal, etc. [Abjalova, 2022]. With this in mind, in order to distinguish the meanings of such lexemes, we refer to M. Abjalova’s “A Dictionary of paronymous words in the Uzbek Language”. When the person entering the text doubts that the lexeme has been entered correctly, he refers to the dictionary of paronyms placed in the program processor and, depending on the meaning of the lexeme, leaves it in its place.

One of the linguistic problems in creating an automatic text analysis program is the automatic detection of homonymy. This issue is somewhat relevant, and in finding its solution, of course, special modules are important [Abjalova, 2020; Abjalova, 2021]. Linguistic modules are designed for existing homonyms in the program base. Therefore, a total of 1,638 homonymous units, appeared in Sh. Rakhmatullayev’s book “Explanatory Dictionary of Homonyms of the Uzbek Language” [Rakhmatullayev, 1984], which appeared later in the speech, was added to the database.

Thesauruses and ontologies are important types of linguistic resources. Thesaurus (from Greek *thesaurós* – treasure, wealth) is a unique special view of the dictionary, which shows the semantic relationship between linguistic units (synonyms, antonyms, paronyms, hyponyms, hyperonyms, etc.). Therefore, these dictionaries are important in the semantic analysis stage of the automatic text analysis system [Abjalova, 2021].

Unlike an explanatory dictionary, a thesaurus not only explains the meaning of a word but also shows the connection (semantic relationship) of a particular word with other words. This shows that thesaurus dictionaries can be used to increase the knowledge base of the artificial intelligence system. Popularization of such dictionaries provides a solution to the task of informational search [Лукашевич, 1996].

The concept of thesaurus is closely related to the concept of ontology [Abjalova, 2021]. Ontology (Greek *ontós* – existence, *logos* – doc-

trine) is a philosophical doctrine of existence and is a general lexical database of language. The reason that ontological dictionaries are considered linguistic is that they are built on the basis of the vocabulary of a certain language [Abjalova, 2021].

It is important to fill the ontology with extra-linguistic activities such as speech, the spirit of the language owner, knowledge of the world, intellectual system-social thinking, and national-cultural views. Only then will linguistics be of perfect use. After all, lexemes in the language as a product of human cognition serve as devices that hold knowledge about the world of language.

The WordNet [22] system is such a linguistic ontology, and it is a large lexical resource. This system contains lexemes related to English noun, adjective, verb, and adverb groups and their semantic connections. The words in each of the categories shown in the system are divided into groups of synonyms (*synsets*) that relate to antonyms, hyponyms, hyperonyms (*genus – species*), holonyms and meronyms (*part – whole*). The system resource covers about 25,000 words. The number of words in the hierarchical relationship for the genus-species relationship is 15,000 (7–8). The highest level of the hierarchy creates a general ontology, a system of basic concepts about the world. The lexical resources of other Western languages are compiled according to the English WordNet scheme. Typologically, the belonging of languages to different families, the fact that languages are of a flexive and agglutinative nature, the basis of lexical data of the English language – the meaning relationship in the WordNet system is not exactly consistent. When marking the tag of word categories, the English-language adverb category may correspond to the word translation in the adjectival category in Uzbek [Abjalova, 2021].

A special linguistic resource of all linguistic programs that not only automatic text analysis but also text processing is the grammar of natural language. First of all, grammar is important with its normative rules. The number of grammatical rules depends on the morphology and syntax model.

## CONCLUSIONS

In conclusion, it is worth saying that grammar dictionaries are important in creating the framework of linguistic programs. Grammatical dictionaries are dictionaries that contain information about the morphological and syntactic properties of lexemes, in which lexemes are

written in alphabetical or inversion order. The amount of information about the lexeme and the principle of selection of lexemes differ depending on the purpose of the creator of each grammar dictionary [10]. A. Polatov states that the computer dictionary is the most effective dictionary according to the requirements of the present time and cites the parts of the computer dictionary and the requirements for it [Polatov, 2011]. In her research, E.I. Bolshakova states that the linguistic processor is based on the linguistic data model, that is, the linguistic resources of the software are: computer dictionaries, natural language grammar, phrase base, thesaurus and ontologies, text sets, and corpus [Большакова, 2011].

Compilation of computer thesauruses and computer grammar dictionaries is a more extensive and laborious task than the development of linguistic modules of the program. Therefore, one of the important tasks of CL is the creation of automated linguistic resources [Hearst, 1998] [Figure 2].

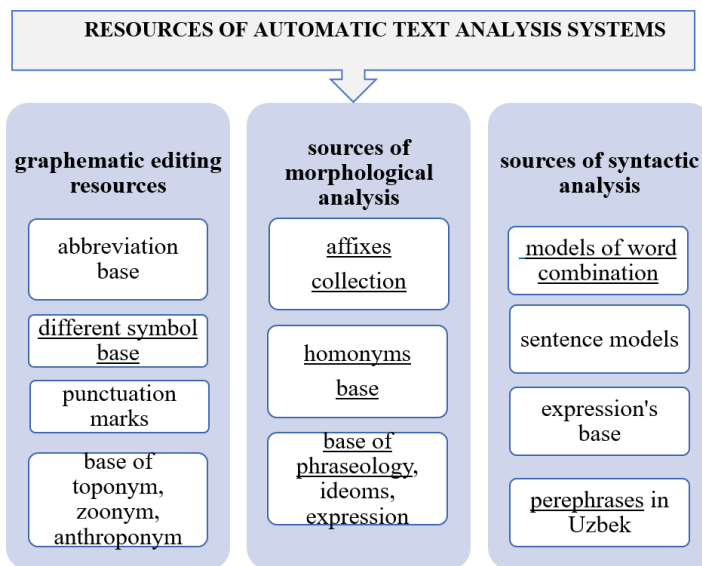


Figure 2. Automatic editing and analysis sources.

The program of automatic editing and analysis of texts can be noted as a form of innovative learning, and it can be shown that it has the following capabilities:

- 1) quickly edits texts of various sizes;
- 2) improves the typist's written speech skills;
- 3) helps the user to independently write texts correctly and literately;
- 4) acts as a linguistic resource in studying the lexical and grammatical norms of the Uzbek language. Therefore, it is considered important to place in the Microsoft Office Word program, perfectly creating the base of the text analysis package of the Uzbek language.

## REFERENCES:

1. Abjalova M. Linguistic modules of editing and analysis programs. [Text]: monograph / M.A. Abjalova. – Tashkent: Nodirabegim, 2020. – 176 p.
2. Abjalova M.A. Ontology of the Uzbek language: technology and concept of creation. [Text]: monograph / M.A. Abjalova. – Tashkent: Nodirabegim, 2021. – 215 p. ISBN 978-9943-7804-5-3
3. Abjalova M. Dictionary of paronyms of the Uzbek language [text]: educational-methodological dictionary. – Termez, 2022. – 64 p.
4. Abjalova M., Iskandarov O. Methods of Tagging Part of Speech of Uzbek Language. // IEEE – UBMK – 2021: 6<sup>th</sup> International Conference on Computer Science and Engineering. 15-16-17 September 2021. Ankara – Turkey. DOI: 10.1109/UBMK52708.2021.9558900. – pp. 82–85. Impact Factor 5.5
5. Abjalova M., Yuldashov A. Methods for Determining Homonyms In Homonymy And Linguistic Systems. ACADEMICIA: An International Multidisciplinary Research Journal. Vol. 11, Issue 2, February 2021. Impact Factor: SJIF 2021 = 7.492 (<https://saarj.com>). ISSN: 2249-7137.
6. Explanatory Dictionary of the Uzbek language: more than 80,000 words and combinations of words (ed. under A. Madvaliev). In 5 volumes. – Tashkent: National Encyclopedia of Uzbekistan, 2006.
7. Hearst, M.A. Automated Discovery of WordNet Relations. In: Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database. MIT Press, Cambridge, 1998, – P. 131–151.
8. Hirst, G. Ontology and the Lexicon. In.: Handbook on Ontologies in Niformation Systems. Berlin, Springer, 2003.
9. Hojiyev A. Explanatory Dictionary of synonyms of the Uzbek language. – Tashkent: Okituvchi, 1974. – 308 p.
10. [http://gramota.ru/slovari/types/17\\_5](http://gramota.ru/slovari/types/17_5)
11. Kurbanova M., Abjalova M. Stressed Dictionary of borrowed words of Uzbek Language. [Text]: educational-methodological dictionary / M.Kurbanova, M.Abjalova, N.Akhmedova, R.Tolaboyeva. – Tashkent: Nodirabegim, 2021. – 988 p. ISBN 978-9943-6940-9-5

12. Matsumoto Y. Lexical Knowledge Acquisition. In: The Oxford Handbook of Computational Linguistics. Mitkov R. (ed.). Oxford University Press, 2003, – P. 395–413.
13. Mengliyev B. And others. Educational Explanatory Dictionary of Uzbek language phrases. – Tashkent: Yangi asr avlodi, 2007.
14. Mengliyev B., Bahriddinova B. Word composition of the Uzbek language educational dictionary. – Tashkent: Yangi asr avlodi, 2007.
15. Polatov A. Computational linguistics. – Tashkent: Akademnashr, 2011. – pp. 213.
16. Rakhmatullayev Sh. Explanatory Dictionary of the homonyms of the Uzbek language. – Tashkent: Okituvchi, 1984. – 215 p.
17. Rakhmatullayev Sh. Explanatory phraseological Dictionary of the Uzbek language. – Tashkent: Okituvchi, 2001. – 407 p.
18. Rakhmatullayev Sh., Mamatov N., Shukurov R. Explanatory Dictionary of the antonyms of the Uzbek language. – Tashkent: Okituvchi, 1980. – 232 p.
19. Samad A. spelling dictionary of words starting with “X” and “H”. – Tashkent, 2007. – 346 p.
20. Tikhonov A., Gulomov A. and others. Morpheme Dictionary of the Uzbek language. – Tashkent. 1977. – 463 p.
21. Usmon O., Doniyorov R. Explanatory Dictionary of Russian-International words. – Tashkent, 1965.
22. Word Net: an Electronic Lexical Database / Edit. by Christiane Fellbaum. Cambridge, MIT Press, 1998.
23. Абжалова М. Автоматический анализ фразеологических единиц в лингвистических программах // VIII Международная конференция по компьютерной обработке тюркских языков «TurkLang-2020». (Труды конференции). Уфа: ИИЯЛ УФИТС РАН, 2020. – С. 76–80.
24. Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие. – Москва: МИЭМ, 2011. – С. 97–99.
25. Лукашевич Н.В., Салий А.Д. Тезаурус для автоматического индексирования и рубрицирования: разработка, структура, ведение // НТИ, Сер. 2, №1, 1996. – С. 1–6.

УДК: 004.891.2

**ПРОЕКТИРОВАНИЕ ЭКСПЕРТНОЙ СИСТЕМЫ  
«DIALECTEXPERT» С ИСПОЛЬЗОВАНИЕМ  
ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ*****Р. А. Бурнашев, М. Р. Галимов****Институт прикладной семиотики Академии Наук Республики Та-  
тарстан, Казань, Россия*

r.burnashev@inbox.ru, magl.galimov@gmail.com

В статье представлены результаты проектирования экспертной системы «DialectExpert». Созданный в рамках исследования прототип экспертной системы предназначен для специалистов работающих в области диалектологии, лингвистики, типологии и компаративных исследований. Экспертная система собирать, обрабатывать и анализировать пространственные данные о языковых единицах различных языковых уровней, представленными в различной форме (графовые и реляционные).

В процессе проектирования экспертной системы, а также визуализации пространственных данных были использованы геоинформационные технологии.

Для повышения уровня поиска информации по базе знаний были использованы современные программные библиотеки с элементами нечеткой логики.

Для создания веб-сервиса была использована программная библиотека (фреймворк) Django языка программирования Python. Для хранения и обработки знаний была использована графовая база данных Neo4j. Визуализация полученных и обработанных пространственных данных на карте выполнялась с помощью современных программных библиотек Python.

**Ключевые слова:** база знаний, экспертная система, нечёткая логика, геоинформационные технологии

**DESIGNING OF THE EXPERT SYSTEM «DIALECTEXPERT»  
USING GEOINFORMATION TECHNOLOGIES*****R.A. Burnashev, M.R. Galimov****Institute of Applied Semiotics of the Academy of Sciences of Tatarstan  
Republic Kazan, Russia*

r.burnashev@inbox.ru, magl.galimov@gmail.com

The paper presents the results of the design of the expert system «DialectExpert». The prototype of the expert system created as part of the research is intended for specialists working in the field of dialectology, linguistics, typology and comparative studies. The expert system collects, processes and analyzes spatial data on language units of various language levels, presented in various forms (graph and relational).

Geoinformation technologies were used in the process of designing the expert system, as well as visualization of spatial data.

Modern software libraries with elements of fuzzy logic were used to increase the level of information search in the knowledge base.

To create a web service, the Django software library (framework) of the Python programming language was used. The Neo4j graph database was used for storing and processing knowledge. Visualization of the received and processed spatial data on the map was performed using modern Python software libraries.

**Keywords:** knowledge base, expert system, fuzzy logic, geoinformation technologies

## 1. Введение

Из большого количества разнородной информации, поступающей к нам сегодня из разных источников, порой бывает трудно выделить главное и сделать правильное решение. Часто у специалиста в области диалектологии возникает сложность сбора и обработки входной информации. Использование в научных исследованиях современных программных средств по считыванию и обработке различий диалектов народов с последующим картографированием является актуальностью задачей. Графическая информация воспринимается человеком в несколько раз быстрее, нежели текстовая.

Для разработки экспертной системы [Burnashev, 2020] были использованы библиотеки языка программирования Python: Django, Pandas, Neo4j Driver, os, NumPy, Neomodel и др.

В ходе исследования мы провели эксперименты по интеграции графовой базы данных Neo4j в экспертную систему.

Для работы с интерактивными картами была использована библиотека Folium.

Folium позволяет легко визуализировать данные, которые были обработаны в Python на интерактивной карте Leaflet. Он позволяет картографировать данные, а также передавать векторные/растровые/HTML элементы в качестве маркеров на карте.

## 2. Обработка и анализ геопространственных данных

В процессе разработки пользовательских интерфейсов были использованы следующие данные:

1. Личная информация об исследователе и респонденте:
  - ФИО исследователя и респондента
  - Образование и профессия исследователя
  - Образование и профессия респондента



– Место проживания респондента и др

2. Информация о языке и диалекте, который изучается:

– Название языка

– Название диалекта

– Географическое расположение диалекта и др.

3. Базы знаний с описанием фонетических (произношение звуков, интонация), морфологических (склонение и спряжение слов, образование словоформ) и лексических (уникальные слова и выражения) особенностей диалекта.

4. Примеры употребления конкретных слов и фраз на диалекте (идёт запись звука или видео). В процессе употребления слов могут быть использованы устройства записи (телефон, диктофон и др.)

Эти вопросы при анкетировании могут быть дополнены или изменены в зависимости от конкретной задачи исследования.

Ниже приведена главная страница созданной экспертной системы «DialectExpert» с использованием геоинформационных технологий (Рис. 1.):

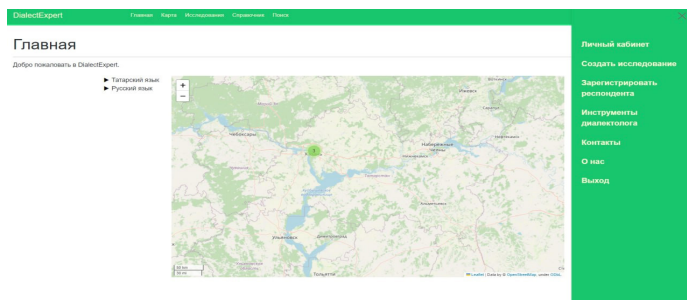


Рис. 1. Интерфейс главной страницы экспертной системы «DialectExpert»

Интерфейс веб-сервиса состоит из следующих вкладок:

- Личный кабинет;
- Создать исследование;
- Зарегистрировать респондента;
- Инструменты диалектолога;
- Контакты;
- О нас.

Для работы с графиками и отчётами была использована программная библиотека Matplotlib. Matplotlib – библиотека на языке

программирования Python для визуализации данных двумерной (2D) графикой. Обработанные графики могут быть использованы в качестве иллюстраций при формировании отчёта (статистики).

Использование данных программных инструментов позволило осуществлять поиск корреляций между языковыми параметрами и географическим распределением языков.

Ниже представлен интерфейс с инструментами для исследователя (Рис. 2.). Интерфейс включает в себя набор справочников, который в ходе исследования заполняется в ручном и автоматическом режиме.

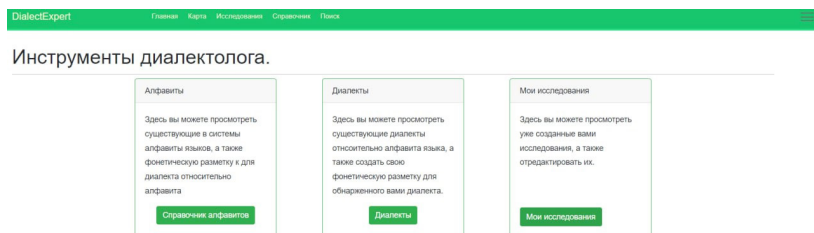


Рис. 2. Инструменты диалектолога

Созданные в рамках исследования личные кабинеты пользователей (респондент, диалектолог и администратор) (Рис. 3, 4) позволяют регистрировать исследователей и респондентов для последующей визуализации пространственных данных на карте.

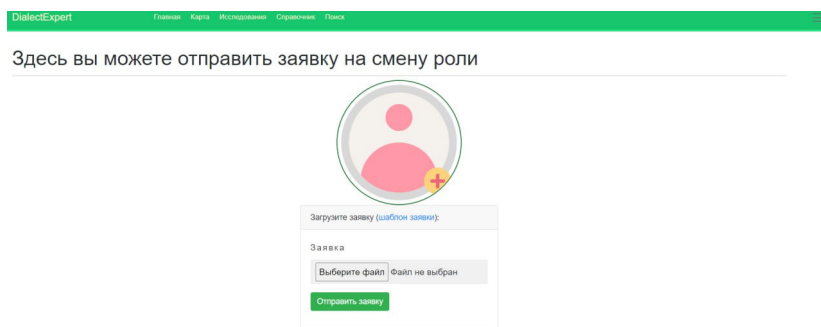


Рис. 3. Примеры личного кабинета и регистрация роли пользователей

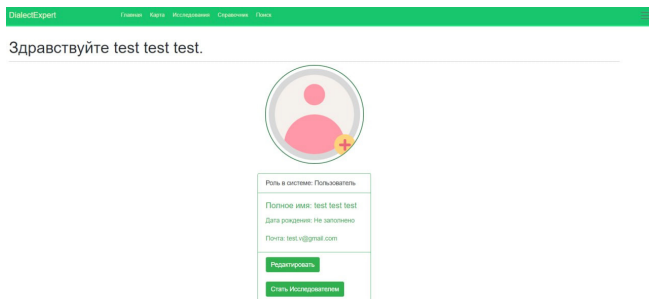


Рис. 4. Интерфейс с личной информацией о пользователе

### 3. Заключение

Разработанный прототип экспертной системы «DialectExpert» предназначен для специалистов в области диалектологии. Система позволяет собирать, обрабатывать и анализировать пространственные данные о языковых различиях в разных регионах.

Созданную экспертную систему с элементами нечёткой логики и геоинформационных технологий в виде единого веб-сервиса планируется интегрировать в портал [Gatiatullin., 2022; Sulevmanov, 2020] на базе Института прикладной семиотики Академии наук Республики Татарстан. Программные инструменты этого портала позволяют вам описывать как языки, так и диалекты на разных лингвистических уровнях и в разных проявлениях.

### ЛИТЕРАТУРА

1. A. Gatiatullin, L. Kubedinova, N. Prokopyev and A. Ibraim, “Toolset of “Turkic Morpheme” Portal for Creation of Electronic Corpora of Turkic Languages in a Unified Conceptual Space,” 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 408–412, doi: 10.1109/UBMK55850.2022.9919449.

2. D. Sulevmanov, A. Gatiatullin, N. Prokopyev and N. Abdurakhmonova, “Turkic Morpheme Web Portal as a Platform for Turkology Research,” 2020 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2020, pp. 1–5, doi: 10.1109/ICISCT50599.2020.9351500.

3. R. A. Burnashev, I. A. Enikeev and A. I. Enikeev, “Design and Implementation of Integrated Development Environment for Building Rule-Based Expert Systems,” 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, Russia, 2020, pp. 1–4, doi: 10.1109/FarEastCon50210.2020.9271143.

УДК 004.8:81'33

**ПОРТАЛ ИТ-РЕСУРСОВ ПО РАСШИРЕНИЮ ФУНКЦИЙ  
И ПОВЫШЕНИЮ КУЛЬТУРЫ КАЗАХСКОГО ЯЗЫКА**

*А. А. Шарипбай<sup>1</sup>, Г. Т. Бекманова<sup>1</sup>, Б. Ж. Ергеш<sup>1</sup>, Алтынбек  
Зулхажав<sup>1</sup>, А. С. Омарбекова<sup>1</sup>, А. С. Муканова<sup>2</sup>*

*<sup>1</sup>Евразийский национальный университет им. Л. Н. Гумилева  
Астана, Казахстан*

*<sup>2</sup>Международный университет Астана, Казахстан  
sharalt@mail.ru, gulmira-r@yandex.kz, b.yergesh@gmail.com,  
altinbekpin@gmail.com, omarbekova\_as@mail.ru, asel\_ms@bk.ru*

В данной статье описаны результаты исследования по разработке научно-лингвистических основ и ИТ-ресурсов по расширению функций и повышению культуры казахского языка. В рамках данного исследования построены формальные описания грамматики казахского языка, созданы базы данных синонимических слов, собственных имен, терминов для соответствующих предметных области (общественно-политическая и публичная жизнь, грамматика казахского языка и школьные учебники) с учетом их истории, использования и возможных толкований, формированы аудио и текстовый корпуса для синтеза казахской речи, а также созданы базы знаний по научному наследию Ахмета Байтурсынулы и по всем структурным ярусам языка в свете его учений. Результаты реализованы в портале ИТ-ресурсов по расширению функций и повышению культуры казахского языка.

**Ключевые слова:** казахский язык, научно-лингвистические основы, синонимайзер, формальные правила, терминология школьных учебников, научное наследие А. Байтурсынулы

**PORTAL OF IT RESOURCES FOR EXPANDING  
THE FUNCTIONS AND IMPROVING THE CULTURE OF THE  
KAZAKH LANGUAGE**

*Altynbek Sharipbay<sup>1</sup>, Gulmira Bekmanova<sup>1</sup>, Banu Yergesh<sup>1</sup>, Altan-  
bek Zulkhazav<sup>1</sup>, Assel Omarbekova<sup>1</sup>, Assel Mukanova<sup>2</sup>*

*<sup>1</sup>L. N. Gumilyov Eurasian National University, Astana, Kazakhstan*

*<sup>2</sup>Astana International University,  
Higher School of Information Technology and Engineering  
Astana, Kazakhstan*

*sharalt@mail.ru sharalt@mail.ru, gulmira-r@yandex.kz,  
b.yergesh@gmail.com, altinbekpin@gmail.com,  
omarbekova\_as@mail.ru, asel\_ms@bk.ru*

The article describes the results of research on the development of scientific-linguistic foundations and IT resources for expanding functions and enhancing the culture of the Kazakh language. Within this research, formal descriptions of the Kazakh language grammar were constructed, databases of synonymous words, proper nouns, and terms for relevant subject areas (socio-political and public life, Kazakh language grammar, and school textbooks) were created, taking into account their history, usage, and possible interpretations. Audio and text corpora were formed for Kazakh speech synthesis, and knowledge bases were established regarding the scientific legacy of Ahmet Baitursynuly and all structural levels of the language in the light of his teachings. The results have been implemented in the IT resource portal for expanding functions and enhancing the culture of the Kazakh language.

**Keywords:** Kazakh language, scientific and linguistic foundations, synonymizer, formal rules, terminology of school textbooks, scientific heritage of A. Baitursynuly

## Введение

Для развития современного казахского языка, его доступности к изучению в сети Интернет, распространению цифровых сервисов, повышению его культуры существует потребность в цифровых лингвистических ресурсах. Всестороннее изучение и комплексное решение данной междисциплинарной проблемы позволит получить необходимые результаты.

Разработка научно-лингвистических основ и IT-ресурсов по расширению функций и повышению культуры казахского языка, как языка межэтнического общения в цифровом формате является актуальной и важной задачей в Казахстане.

Казахский язык – один из тюркских языков с древней историей, формировавшейся на протяжении многих веков, как и языки родственных тюркских племен, проживающих на территории современного Казахстана [1].

Решение данной задачи осуществляется на основе: анализа научных, методологических и нормативных основ грамматики казахского языка и компьютерной лингвистики; исследования моделей и методов синтеза речи; построения формального описания грамматики казахского языка, технологии искусственного интеллекта для разработки интеллектуального синонимайзера, электронного справочника, мобильного приложения «Увлекательная ономастика», электронного словаря терминологии школьных учебников, синтезатора казахской речи и интеллектуальной системы «Ахметтану».

Данная задача была реализована в рамках программы BR11765535 «Разработка научно-лингвистических основ и IT-ресурсов по расширению функций и повышению культуры казахского языка». Решением данных проблем занимаются группа видных ученых Республики Казахстан и Евразийского национального университета им. Л.Н. Гумилева в области искусственного интеллекта и компьютерной лингвистики, имеющих большой научный задел по проблеме исследования.

## Модули портала IT-ресурсов

Использованы интеллектуальные технологий по компьютерной обработке (анализу и синтезу) устных и письменных текстов (данных) на казахском языке.

В результате разработан портал [<https://kazlangres.enu.kz/#/>] включающие следующие подсистемы:

– интеллектуальное приложение – синонимайзер стандартных образцов синонимического ряда слов в текстах общественно-политического дискурса и публичной речи, состоящее из онтологической модели, исходного кода и базы данных синонимических слов общественно-политического дискурса и публичной речи, пользовательского интерфейса на казахском языке с функцией корректировки имеющихся и добавления новых слов-синонимов и исходного кода (Рисунок 1). Синонимайзер позволяет осуществлять поиск слов в казахском языке, с дополнительной информацией, часть речи, однозначность или многозначность слово, толкование слова, пример, а также синонимы и перифразы слова; возможность добавления новых синонимов и перифраз к заданному слову в базу [Vekmanova, 2023a; Vekmanova, 2023b].

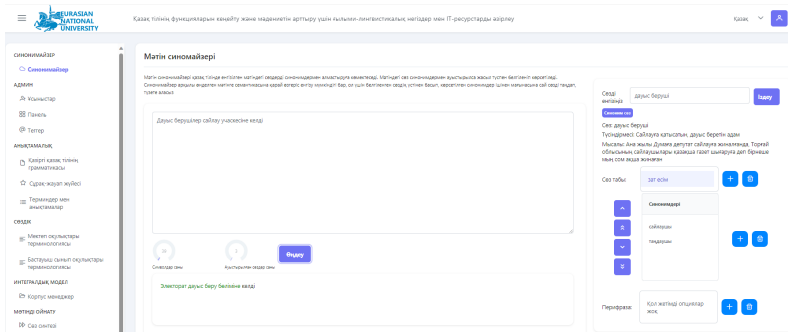


Рис. 1. Интеллектуальное приложение – синонимайзер

– *грамматический электронный справочник грамматики современного казахского языка*, состоящий из формального описания грамматики и синтаксиса казахского языка [Муканова, 2022; Yelibayeva, 2022; Жеткенбай, 2023], исходного кода и базы данных концептов и примеров, пользовательского интерфейса на казахском языке с функцией корректировки имеющихся и добавления новых концептов и примеров (рисунок 2).

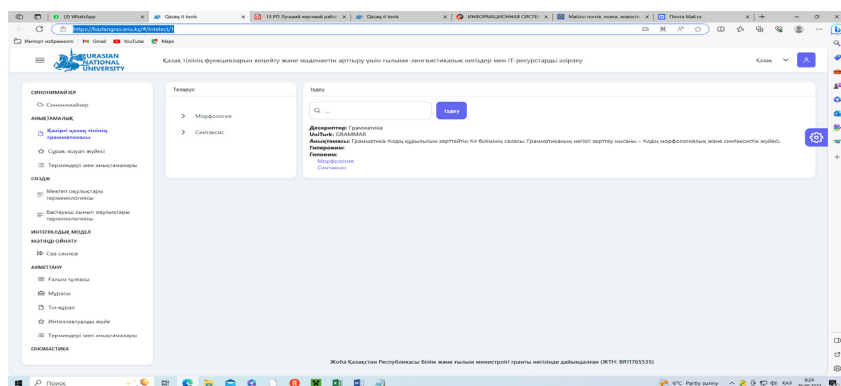


Рис. 2. Грамматический электронный справочник грамматики современного казахского языка

– *мобильное приложение «Увлекательная ономастика»*, состоящее из исходного кода, базы данных собственных имен и примеров, пользовательского интерфейса на казахском языке с функцией корректировки имеющихся и добавления новых собственных имен и примеров. В мобильном приложении «Увлекательная ономастика» учитываются семантические признаки казахских имен собственных, в том числе по области антропониимики [7, 8].

– *электронный словарь терминологии школьных учебников*, состоящий из исходного кода, базы данных терминов школьных учебников, пользовательского интерфейса на казахском языке с функцией корректировки имеющихся и добавления новых терминов [9, 10].

При разработке терминологического словаря особое внимание уделяется возрастным особенностям пользователей, разработан индивидуальный интерфейс для учащихся 1–4 классов, интерфейсные элементы больше, возвращается определение каждого термина только на 1 уровне. Кроме того, поскольку у учащихся

1-2 классов навыки чтения еще не сформированы, предусмотрена передача термина звуком, интерфейс снабжен анимационными элементами для детей (Рисунок 3).

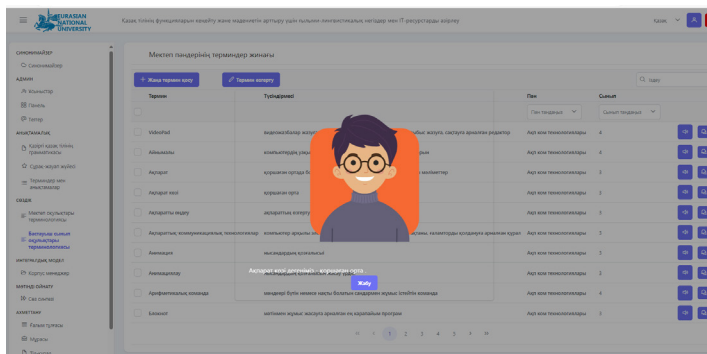


Рис. 3. Электронный словарь терминологии школьных учебников

– лингвистические основы синтеза казахской речи: *информационная система для синтеза речи казахского языка*, состоящая из исходного кода, аудио и текстовый корпуса для синтеза казахской речи, пользовательского интерфейса на казахском языке с функцией корректировки имеющихся и добавления новых слов и предложений (Рисунок 4).

В рамках данной работы был проведен весь спектр научно-исследовательских работ для разработки системы синтеза казахской речи – от сбора данных до разработки моделей синтеза речи [11, 12, 13].

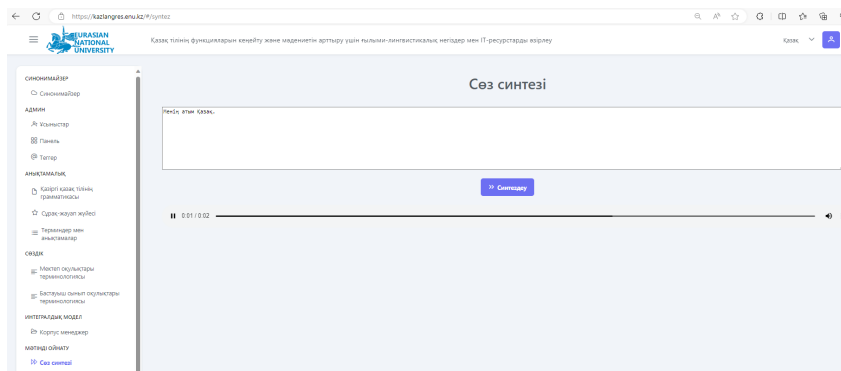


Рис. 4. Информационная система для синтеза речи казахского языка





3. Bekmanova, G., Omarbekova, A., Mukanova, A., Zulkhazhav A., Zakirova A., Ongarbayev, Y. Development of an ontological model of words in public political discourse. The 7th International Conference on Education and Multimedia Technology (ICEMT 2023) August 29–31, 2023, Tokyo, Japan.

4. A. Mukanova, B. Yergesh, G. Yelibayeva, G. Bekmanova. Applying the Ontological Approach to Electronic Guide Development [Электрондық гидтерді әзірлеуге онтологиялық тәсілді қолдану] // 2022 International Conference on Engineering & MIS (ICEMIS), Istanbul, Turkey, 4-6 July, 2022. – pp. 1–4, doi: 10.1109/ICEMIS56295.2022.9914299.

5. G. Yelibayeva, B. Yergesh, G. Bekmanova, B. Razakhova, A. Sharipbay, A. Mukanova. Modelling of Verb Phrases of the Kazakh Language // 2022 International Conference on Engineering & MIS (ICEMIS), 2022, pp. 1–3, doi: 10.1109/ICEMIS56295.2022.99140157

6. Л. Жеткенбай\*, А.Ә. Шәріпбай, Б.Ш. Разахова, А.Б. Барлыбаев. Формализация притяжательных и личных окончаний казахского языка с использованием логических правил вывода [Қазақ тіліндегі тәуелдік, жіктік жалғауларын логикалық шығарым ережелері арқылы формалдау] // Вестник Алматинского университета энергетики и связи, – No 3(62). – 2023, [https://doi.org/10.51775/2790-0886\\_2023\\_62\\_3\\_118](https://doi.org/10.51775/2790-0886_2023_62_3_118)

7. Gulmira Bekmanova, Gaziza Yelibayeva, Banu Yergesh, Laura Orynbay, Ayaulym Sairanbekova, and Zulfiya Kaderkeyeva. The Emotional Coloring of Kazakh Names of People in the Semantic Knowledge Database of the Mobile Application “Fascinating Onomastics” // The 21st IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT’22, ACER-EMORE2022), Niagara Falls, Canada, 17-20 November, 2022.

8. Орынбай Л.О., Сайранбекова А.Д., Елибаева Г.К., Бекманова Г.Т. Қазақ есімдерінің семантикалық базасының құрылымын анықтау жолдары // «TURKLANG 2022» “Түркі тілдерін компьютерлік өңдеу” атты X халықаралық конференция еңбектері. – Нұр-Сұлтан, 2022. – 302–309 б.

9. Шарипбай А.А., Омарбекова А.С. Структура базы данных терминов школьных предметов и виды запросов к ним // КОМПЬЮТЕРНАЯ ОБРАБОТКА ТЮРКСКИХ ЯЗЫКОВ. X международная конференция: Труды.. – Нұр-Сұлтан, 2022. – б. 280–285.

10. G. Bekmanova, N. Amangeldy, A. Nazyrova, A. Sharipbay, S. Kudubayeva. A new approach to developing a terminological dictionary of school subjects in the Kazakh language [Қазақ тіліндегі мектеп пәндерінің терминологиялық сөздігін жасауға жаңа көзқарас] // Proceedings of 7th International Conference on Computer Science and Engineering (UBMK). – 2022. – P. 568–574. (will be indexed in Scopus).

11. Bekmanova, G.; Yergesh, B.; Sharipbay, A.; Mukanova, A. Emotional Speech Recognition Method Based on Word Transcription [Сөзді транскрипциялауға негізделген эмоциялық сөйлеуді тану әдісі] // *Sensors*, 2022, 22(5):1937. <https://doi.org/10.3390/s22051937>.

12. Bekmanova, G., Yergesh, B., Sharipbay, A., Omarbekova, A., Zakirova, A. (2022). Linguistic Foundations of Low-Resource Languages for Speech Synthesis on the Example of the Kazakh Language [Қазақ тілі мысалында ресурсы аз сөйлеу синтезі тілдерінің лингвистикалық негіздері] // In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C. (eds) *Computational Science and Its Applications – ICCSA 2022 Workshops. ICCSA 2022. Lecture Notes in Computer Science*, vol 13379. Springer, Cham. [https://doi.org/10.1007/978-3-031-10545-6\\_1](https://doi.org/10.1007/978-3-031-10545-6_1).

13. Kozhirkbayev Z., Yessenbayev Z., Islamgozhayev T. Sharipbay A., Preliminary tasks of unsupervised speech recognition based on unaligned audio and text data [Біркелкі емес дыбыс және мәтін деректеріне негізделген бақылаусыз сөйлеуді танудың алдын ала тапсырмалары] // 2022 International Conference on Engineering & MIS (ICEMIS), Istanbul, Turkey, 2022, pp. 1–3, doi: 10.1109/ICEMIS56295.2022.9914249.

14. Сыздықова Г.О., Жұмағұлова А.А. Ахмет Байтұрсынұлы – ұлт ұстазы “Ұлт ұстазы Ахмет Байтұрсынұлының 150 жылдық мерейтойына арналған «Ахмет Байтұрсынұлының мұрасы және ұлттық құндылықтар» атты халықаралық ғылыми-прикладикалық конференция материалдары. Том 1. – Алматы: Абай атындағы ҚазҰПУ, «Ұлағат» баспасы, 2022. – 139–142 б.

15. Ергеш, Б., Бекманова, Г., Сыздықова, Г., & Жумагулова, А. (2023). Онтологическое моделирование звуковой системы по наследию Ахмета Байтұрсынұлы. *Вестник Евразийского национального университета имени Л.Н.Гумилева серия: технические науки и технологии*, 144(3), 52–59.

## УДК 811.512.15

СОЗДАНИЕ БАЗЫ ДАННЫХ БУДДИЙСКИХ ТЕКСТОВ  
ХРАМА «ЦЕЧЕНЛИНГ»<sup>1</sup>*А. Я. Салчак, А. Б. Хертек**Тувинский государственный университет,  
Кызыл, Тува, Россия  
aelita\_74@mail.ru, khertek.ab@yandex.ru*

В статье рассматриваются итоги выполнения проекта «Духовное наследие храма «Цеченлинг». Проект направлен на сохранение и цифровизацию духовного наследия храма Цеченлинг. В рамках проекта проведена научно-исследовательская работа по созданию электронного корпуса переводных традиционных буддийских текстов, осуществлены переводы на тувинский язык традиционных буддийских текстов Идэгэл (Принятие Прибежища Трех Драгоценностей), Ногоон Дарь Эх (восхваления и благопожелания Зеленой Таре), Цагаан Дарь Эх (восхваления и благопожелания Белой Таре) с тибетского и старомонгольского на тувинский язык.

**Ключевые слова:** храм Цеченлинг, буддизм, традиционные буддийские тексты, базы данных переводных буддийских текстов, тувинский язык, старописьменный монгольский язык, тибетский язык.

CREATION OF A DATABASE OF BUDDHIST TEXTS  
OF THE TEMPLE «TSECHENLING»*Aelita Salchak, Arzhaana Khertek**Tuvan State University,  
Kyzyl, Tuva, Russia  
aelita\_74@mail.ru, khertek.ab@yandex.ru*

The article discusses the results of the project “Spiritual Heritage of the Tsechenling Temple”. The project aims to preserve and digitalize the spiritual heritage of Tsechenling Temple. The research work on creation of an electronic corpus of translations of traditional Buddhist texts from Tibetan and Old Mongolian into Tuvan was carried out. Such traditional Buddhist texts as Idegel (Taking Refuge of the Three Jewels), Nagoon Dar Ekh (praises and well wishes to the Green Tara), Tsagaan Dar Ekh (praises and good wishes to White Tara) were translated into Tuvan.

**Keywords:** Tsechenling Temple, Buddhism, traditional Buddhist texts, databases of translated Buddhist texts, Tuvan language, old written Mongolian language, Tibetan language.

---

<sup>1</sup> Статья подготовлена при поддержке Фонда содействия буддийскому образованию и исследованиям.

Российский буддизм является интегральной частью российского духовного и культурного пространства. Различные буддийские общины начали развиваться в России в XVIII в. и стали неотъемлемой частью российской культурной и гражданской идентичности.

В настоящее время вопросы развития российской цивилизации и достижения ее ценностных целей особенно актуальны в свете современных вызовов. В этой связи важно содействие деятельности буддийских организаций, направленное на консолидацию общества и укрепление государства. Буддизм как философия, религия и культура вносит огромный вклад в выработку и определение универсальных идеалов, ценностей, сформулировав их в том числе в понятиях, объединяющих Россию и мир в общем стремлении к светлому и гармоничному будущему для новых поколений.

Тува наряду с Бурятией и Калмыкией является центром развития традиционных буддийских общин. Большая часть населения Тувы исповедует буддизм более 70% населения.

В Стратегии научно-технологического развития Российской Федерации, утвержденной Указом Президента Российской Федерации от 1 декабря 2016 г., № 642, отмечается, что в ближайшие 10-15 лет приоритетами научно-технологического развития Российской Федерации следует считать те направления, которые позволят получить научные и научно-технические результаты и создать технологии, являющиеся основой инновационного развития внутреннего рынка продуктов и услуг, устойчивого положения России на внешнем рынке. В качестве приоритетных направлений указываются также «противодействие техногенным, биогенным, социокультурным угрозам, терроризму и идеологическому экстремизму, а также киберугрозам и иным источникам опасности для общества, экономики и государства; возможность эффективного ответа российского общества на большие вызовы с учетом взаимодействия человека и природы, человека и технологий, социальных институтов на современном этапе глобального развития, в том числе применяя методы гуманитарных и социальных наук» [Стратегия научно-технологического развития РФ 2016].

В этой связи важно содействие деятельности буддийских организаций, на консолидацию общества и укрепление государства. Буддизм как философия, религия и культура вносит огромный вклад в выработку и определение универсальных идеалов, ценностей, сформулировав их в том числе в понятиях, объединяющих

Россию и мир в общем стремлении к светлому и гармоничному будущему для новых поколений.

В настоящее время важной задачей является сохранение духовного наследия буддийских храмов и центров. Одним из эффективных методов сохранения духовного наследия и содействия его развитию является цифровизация традиционных буддийских текстов.

Буддийский храм «Цеченлинг» является одним из крупнейших духовных центров Тувы, где проводится просветительская работа в области буддизма. Название храма Цеченлинг в переводе означает «Обитель безграничного сострадания». В 1998 г. президентом Тувы был заложен первый камень, а уже в следующем году готовое здание храма освятил Его Святейшество Богдо-Гегээн IX.

Храм «Цеченлинг» является объектом культурного наследия регионального значения, поставлен на государственную охрану (Постановление Правительства Республики Тыва от 28 октября 2004 г. № 968 «О внесении дополнений в Государственный список памятников истории и культуры Республики Тыва»).

На территории храма Цеченлинг построены восемь ступ, каждая из которых соответствует событиям в жизни Будды: рождение, просветление, поворот колеса учения, чудесные деяния, сошествие Тушита на землю с небес, восстановление согласия в Сангхе, совершенная победа и уход в нирвану. Храм Цеченлинг сочетает в себе старинные традиции буддийской архитектуры и современные строительные технологии. В храме находится резиденция Камбы-Ламы Республики Тыва.

При храме «Цеченлинг» проходят молебны, ритуалы и обряды, прием граждан, а также ламы проводят занятия по тибетскому языку, медитации и философии буддизма для духовно-нравственного, патриотического развития граждан.

В храме Цеченлинг имеется архив, содержащий собрания буддийских текстов на тибетском и старомонгольском языках: Идэгэл (Принятие Прибежища Трех Драгоценностей), НогоонДарь Эх (восхваления и благопожелания Зеленой Таре), ЦагаанДарь Эх (восхваления и благопожелания Белой Таре), Дуккар (восхваления и благопожелания Белозонтичной Таре), Шэрнинг («Праздникапарамита») (Сутра Запредельной Мудрости) и т.д.

В рамках проекта «Духовное наследие храма «Цэчэнлинг», поддержанного Фондом содействия буддийскому образованию и исследованиям, исполнителями проекта начата работа по соз-

данию электронного корпуса переводных традиционных буддийских текстов храма Цеченлинг.

Коллективом исполнителей проводится научно-исследовательская работа, направленная на сохранение и цифровизацию духовного наследия храма Цеченлинг.

Выполнены следующие виды работ:

1. Составлены базы данных традиционных буддийских текстов, имеющихся в архиве храма Цеченлинг;

2. Перевод традиционных буддийских текстов Идэгэл (Принятие Прибежища Трех Драгоценностей), НогоонДарь Эх (восхваления и благопожелания Зеленой Таре), ЦагаанДарь Эх (восхваления и благопожелания Белой Таре) с тибетского и старомонгольского на тувинский язык: подготовка подстрочных переводов и литературные переводы;

3. Анализ переведенных текстов.

Руководителем проекта является Хомушка Ольга Матпаевна – доктор философских наук, профессор, Заслуженный деятель науки Республики Тыва, российский религиовед, государственный деятель, ректор Тувинского государственного университета. В состав рабочей группы входят преподаватели ТувГУ, имеющие опыт руководства и участия в научно-исследовательских проектах: Салчак Аэлига Яковлевна, кандидат филологических наук, доцент кафедры тувинской филологии и общего языкознания, ХертекАржаана Борисовна, кандидат филологических наук, доцент кафедры тувинской филологии и общего языкознания, Уламсурен Цецегдарь, кандидат филологических наук, доцент кафедры тувинской филологии и общего языкознания, Ондар Валентина Сувановна, кандидат филологических наук, доцент кафедры русского языка и литературы, Санчай Чойган Херел-ооловна, кандидат культурологии, старший преподаватель кафедры философии, Сарыглар Кара-кыс Александровна, старший преподаватель кафедры иностранных языков.

В рамках проведения работ по переводу составлен перечень традиционных буддийских текстов, имеющихся в архиве храма Цеченлинг; составлены базы данных переводов с тибетского и старомонгольского языков на тувинский язык традиционных буддийских текстов Идэгэл (Принятие Прибежища Трех Драгоценностей), Ногоон Дарь Эх (восхваления и благопожелания Зеленой Таре), Цагаан Дарь Эх (восхваления и благопожелания Белой Таре). Рукописные тексты переведены в электронный формат,

сопровожаются переводами на тувинском языке, базы данных составлены в формате Excel. Параллельные тексты выравнены по смысловым частям.

Пример параллельного выравнивания буддийского текста «Восхваления и благопожелания Белой Таре»

| ЦагаанДарь Эх (монгольский язык)   | Ак –Дарийги (тувинский язык)  |
|--|---|
| <p>Ум хутагт гэтэлгэгч дар эхэд мөргөмүй.<br/>Орчлонгоос гэтэлгэгч дар эх, дудари-гээр 8 аюулнаас гэтэлгэгч, дуурий-гээр өвчнүүдээс тонилгогч гэтэлгэгч эхэд мөргөн магтмуй.</p> | <p>Ыдыктыг Дарийгимге тейлеп тур мен!<br/>Ом Таре – октаргайдан хосталгаш<br/>Туутгаре – коргуушкуннардан камгалап тур сен<br/>Туре -аарыг-аржыктан хостап тур сен.</p>   |
| <p>Цагаан лянхуагийн дунд агсан саран өнгөт дэвсгээрийн дээр, очырын завилалыг зохиогч эх дээдийг өгөгч эхдээ өргөмүй.</p>   | <p>Ак Бадма чечектиң ортузунда саадап олулар<br/>Ай олут кырында саадап олулар<br/>Очур ышкаш аспактанып олуруп алган<br/>Үш эртинени көргүзүп авыралды хайырлаан силерлерге тейледим<br/>Дээди ыдыктыгны өргээниинге тейлеп тур мен.</p> |

Переведенные тексты были проанализированы. Были определены наиболее частотные переводные клише, стандарты и способы переводов.

Собранный и проанализированный материал размещен на сайте <https://tuvan-buddha-texts.ru/>. Также в рамках проекта начата работа по созданию тибетско-монгольско-тувинско-русского словаря буддийских терминов. Работа по созданию электронного корпуса переведенных на тувинский язык традиционных буддийских текстов будет продолжена, результаты будут размещены на сайте <https://tuvan-buddha-texts.ru/>.



УДК 81'33

**ЛИНГВИСТИК АТАМАЛАРНИ ЭНЦИКЛОПЕДИК  
ЛУҒАТДА ИФОДАЛАШ ХУСУСИДА*****D. Rustamova****Andijon davlat universiteti**dilrabo@list.ru*

**Аннотация.** Мазкур мақолада лингвистик энциклопедияда умумий лингвистик терминларни кодировка қилиш масаласи тавсифланади. Бундай луғатларда умумий тилшунослик терминлари изоҳини, луҳат мақоласини шакллантириш методикаси таҳлил қилинади ҳамда умумий тилшунослик терминларини тавсифлаш бўйича тавсиялар берилади. Луғат сўзлигини тузиш энциклопедик луғат тузишнинг муҳим босқичлардан бири, луғат макро-структураси лойиҳасидан кейин луғатнинг асосий қисми – луғат мақоласини ёзиш учун луғатда изоҳланадиган сўзлар рўйхати шакллантирилиши талаб этилади. Мазкур мақолада умумий лингвистик терминлар сўзлигини шакллантириш ва уларни изоҳлаш ҳақида фикр юритилади.

**Калит сўзлар.** Энциклопедик луғат, лингвистик термин, умумий лингвистик термин, хусусий лингвистик термин, тиллар оиласи, трансформ.

**О ВЫРАЖЕНИИ ЯЗЫКОВЫХ ТЕРМИНОВ  
В ЭНЦИКЛОПЕДИЧЕСКОМ СЛОВАРЕ*****Д. Рустамова****Андижанский государственный университет,**Андижан, Узбекистан**dilrabo@list.ru*

**Абстракт.** В данной статье рассмотрен вопрос кодификации общеупотребительных лингвистических терминов в лингвистической энциклопедии. Так же анализируется объяснение общелингвистических терминов, способ формирования словарной статьи, даются рекомендации по описанию общелингвистических терминов. Создание словаря – один из важных этапов создания энциклопедического словаря, после проектирования макроструктуры словаря необходимо создать список слов, подлежащих пояснению в словаре, чтобы написать основную часть словаря – словарную статью. В данной статье рассматривается создание словаря общелингвистических терминов и их толкование.

**Ключевые слова.** Энциклопедический словарь, лингвистический термин, общелингвистический термин, видовой лингвистический термин, языковая семья, трансформация.

## ON THE EXPRESSION OF LINGUISTIC TERMS IN AN ENCYCLOPEDIA DICTIONARY

*Dilrabo Han Rustamova*  
*Andijan State University, Andijan,*  
*dilrabo@list.ru*

This article describes the issue of codification of common linguistic terms in the linguistic encyclopedia. In such dictionaries, the explanation of general linguistic terms, the method of forming a dictionary article is analyzed, and recommendations are given for the description of general linguistic terms. Creating a dictionary is one of the important stages of creating an encyclopedic dictionary, after the project of the macrostructure of the dictionary, it is necessary to create a list of words to be explained in the dictionary in order to write the main part of the dictionary - the dictionary article. This article discusses the creation of a dictionary of general linguistic terms and their interpretation.

**Keywords.** Encyclopedic dictionary, linguistic term, general linguistic term, specific linguistic term, language family, transform.

### КИРИШ

Энциклопедик луғат тузишда муҳим босқичлардан бири луғат сўзлигини тузишдир. Луғат макроструктураси лойиҳалангандан кейин луғатнинг асосий қисми – луғат мақоласини ёзиш учун дастлаб унинг бош сўзи, яъни луғатда изоҳланадиган сўзлар рўйхати шакллантирилиши лозим. Сўзлик луғатнинг тури, фойдаланувчилари доирасига қараб танланади. Мазкур мақолада ўзбек тили лингвистик атамалар энциклопедияси учун сўзлик йиғиш хусусида фикр юритамиз. Демак, дастлабки эътибор луғатнинг соҳа энциклопедияси эканлигига қаратилади. Бунинг учун бошқа тилларда яратилган лингвистик энциклопедияларнинг сўзлиги таркибини шакллантириш тажрибаси ўрганилди, жумладан, рус тилидаги бир неча электрон лингвистик луғат сўзликлари қиёсий таҳлил қилинди.

### АСОСИЙ ҚИСМ

Умумий тилшуномликка оид терминларни энциклопедик луғатда кодировка қилиш, турли энциклопедия ва дарсликлардаги маълумотларни умумлаштириш бўйича тавсияларни ишлаб чиқиш учун бир неча терминни танлаб оламиз.

Мазкур мавзуй гуруҳга мансуб “трансформ” термини А.Ҳожиёв луғатида шундай изоҳланади:

ТРАНСФОРМ. Трансформация натижасида ҳосил бўлган тузилма. Мас., аниқ тузилманинг ўзгариши билан ҳосил бўлган мажхул тузилма. Қ. Трансформация<sup>1</sup>.

В. Жеребило мазкур терминни шундай тавсифлайди:

ТРАНСФОРМ [< трансформация]. В лингвистике: языковая единица, полученная в результате трансформации; напр. *Ветер снес крышу*. – *Ветром снесло крышу*<sup>2</sup>.

О.Ахманова луғатида бу термин мавжуд эмас. Бошқа бир манбада шундай берилади:

#### ТРАНСФОРМ

(от лат. transformare – преобразовывать). Преобразованная языковая форма, структура; см. трансформация. Вопросительный трансформ повествовательного предложения.

Юқоридаги маълумотлар асосида мазкур луғат мақоласини куйидагича шакллантириш мумкин:

#### Transform

Transformatsiya natijasida hosil boʻlgan tuzilma. Masalan., aniq tuzilmaning oʻzgarishi bilan hosil boʻlgan majhul tuzilma. Q. Transformatsiya.

Умумий лингвистик терминлар сирасида “тил шахобчаси” термини мавжуд бўлиб, бу атама ҳамма луғат сўзлигида ҳам учрайвермайди. А.Ҳожиев луғатида “ТИЛЛАРНИНГ ШАЖАРАВИЙ ТАСНИФИ”, “ТИЛЛАРНИНГ ҚАРИНДОШЛИГИ”, “ТИЛЛАР ОИЛАСИ” каби терминлар келтирилган, аммо “тил шахобчаси” термини сўзликда ҳам, кўрсаткичда ҳам мавжуд эмас. А.Ҳожиев мазкур атамаларни шундай изоҳлайди:

ТИЛЛАРНИНГ ШАЖАРАВИЙ ТАСНИФИ. Тилларни келиб чиқиш манбаининг бирлиги, умумийлигига кўра гуруҳларга ажратиш. Бундай умумий манбадан келиб чиққан тилларнинг катта гуруҳлари қариндош тиллар оиласини ташкил этади. Мас., ҳинд-европа тиллари оиласи, мўғул тиллари оиласи, туркий тиллар оиласи ва б<sup>3</sup>.

<sup>1</sup> Ҳожиев А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 111.

<sup>2</sup> Жеребило Т.В. Словарь лингвистических терминов. Изд. 5-е, испр. и доп. – Назрань: ООО «Пилигрим», 2010. – 486 с. – С. 418.

<sup>3</sup> Ҳожиев А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 105.

ТИЛЛАРНИНГ ҚАРИНДОШЛИГИ. Тилларнинг бир асос тилдан келиб чиққанлиги ва уларнинг фонетик, лексик ва грамматик қурилишида бир асос тилдан келиб чиққанлигини кўрсатувчи умумайлик, изчил мосликларнинг мавжудлиги<sup>4</sup>.

ТИЛЛАР ОИЛАСИ. Ўзаро ўхшашликлари келиб чиқиш асосининг умумийлиги билан изоҳланадиган тиллар гуруҳи (тил шохобчаси): Туркий тиллар оиласи<sup>5</sup>.

Кўринадики, мазкур атама алоҳида луғат мақоласи сифатида келтирилмаса-да, “ТИЛЛАР ОИЛАСИ” луғат мақоласида “тиллар гуруҳи”нинг дублети сифатида берилган. В.Жеребило луғатида ҳам алоҳида луғат мақоласи сифатида берилган:

ВЕТВЬ ЯЗЫКОВ (группа языков). Группировка внутри семьи языков. Например, внутри индоевропейской семьи выделяются индоиранская, славянская и др. ветви. Кавказская семья включает картвельскую, абхазо-адыгскую и нахско-дагестанскую ветви<sup>6</sup>.

Ушбу термин инглиз тилидаги луғатларда ҳам алоҳида термин сифатида берилмаган, бошқа луғат мақоласида қуйидагича таърифланади:

A language family is a group of languages related through descent from a common ancestral language or parental language, called the proto-language of that family. The term “family” reflects the tree model of language origination in historical linguistics, which makes use of a metaphor comparing languages to people in a biological family tree, or in a subsequent modification, to species in a phylogenetic tree of evolutionary taxonomy. Linguists therefore describe the daughter languages within a language family as being genetically related. The divergence of a proto-language into daughter languages typically occurs through geographical separation, with different regional dialects of the proto-language spoken by different speech communities undergoing different language changes and thus becoming distinct languages from each other.

<sup>4</sup> Ҳожиёв А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 106.

<sup>5</sup> Ҳожиёв А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 106.

<sup>6</sup> Жеребило Т.В. Словарь лингвистических терминов. Изд. 5-е, испр. и доп. – Назрань: ООО «Пилигрим», 2010. – 486 с. – С. 56.

A language family is a group of languages related through descent from a common ancestral language or parental language, called the proto-language of that family. The term “family” reflects the tree model of language origination in historical linguistics, which makes use of a metaphor comparing languages to people in a biological family tree, or in a subsequent modification, to species in a phylogenetic tree of evolutionary taxonomy. Linguists therefore describe the daughter languages within a language family as being genetically related. The divergence of a proto-language into daughter languages typically occurs through geographical separation, with different regional dialects of the proto-language spoken by different speech communities undergoing different language changes and thus becoming distinct languages from each other.

The language families with the most speakers are the Indo-European family, which includes many widely spoken languages native to Europe (such as English and Spanish) and South Asia (such as Hindi, Urdu and Bengali); and the Sino-Tibetan family, mainly due to the many speakers of Mandarin Chinese in China. A language family may contain any number of languages: Some families, such as the Austronesian and Niger-Congo families, contain hundreds of different languages; while some languages, termed isolates, are not known to be related to any other languages and therefore constitute a family consisting of only one language.

Membership of languages in a language family is established by research in comparative linguistics. Genealogically related languages can be identified by their shared retentions; that is, they share systematic similarities that cannot be explained as due to chance, or to effects of language contact (such as borrowing or convergence), and therefore must be features inherited from their shared common ancestor. However, some sets of languages may in fact be derived from a common ancestor but have diverged enough from each other that their relationship is no longer detectable; and some languages have not been studied in enough detail to be classified, and therefore their family membership is unknown.

Estimates of the number of language families in the world may vary widely. According to Ethnologue there are 7,151 living human languages distributed in 142 different language families. Lyle Campbell (2019) identifies a total of 406 independent language families, including isolates<sup>7</sup>.

Кўринадики, бу мақолада ҳам мазкур атама алоҳида берилмаган, аммо келтирилган луғат мақоласида бу ҳақида маълумот келтирилган.

<sup>7</sup> [https://en.wikipedia.org/wiki/Language\\_family](https://en.wikipedia.org/wiki/Language_family)

Шундан келиб чиқиб мазкур терминнинг кодировкасини қуйидагича кенгайтириш ва алоҳида луғат мақоласи сифатида бериш мақсадга мувофиқ:

### Til shohobchalari

Fonetik jihatdan yaqinlik belgilariga ega bo'lgan til oilalarining ichki guruhlari, shu oila ichidagi til guruhlari. Mas., hind-yevropa tillari oilasiga kiruvchi hind tili, slavyan tili shohobchalari.

Qadimgi hind tilshunoslari grammatika, morfologiya sohasida ham ancha ishlarni amalga oshirdilar. Ular bu yo'nalishda ham grek tilshunoslaridan ancha ilgari ketdilar. Aniqrog'i, hind tilshunosi Guru morfologiyaning uch bo'limdan tashkil topishini aniq ko'rsatib beradi va unga quyidagilarni kiritadi: 1. So'zlar tasnifi (so'z turkumlari). 2. So'z yasalishi. 3. So'z o'zgarishi. Hindlar to'rtta so'z turkumini farqlaganlar: ot, fe'l, old ko'makchi va yuklama. Hind tilshunosligida ot predmet ifodalovchi, fe'l esa harakat, holat ifodalovchi so'z sifatida beriladi. Old ko'makchilar esa otlarning, asosan, fe'llarning ma'nosini belgilaydi. Yuklamalar esa ma'nolariga ko'ra 1) bog'lovchi va 2) qiyoslovchi kabi turlarga ajratiladi. Olmosh va ravishlar esa ot va fe'l turkumlariga qo'shib yuborilgan, alohida ajratilmagan.

Yunonlardan farqli holda hindlar so'z turkumlarini gap bo'laklaridan farqlaganlar, ya'ni ular bilan qorishtirmaganlar, adashtirmaganlar. Shunga ko'ra hindlar, yuqorida aytilganidek, otlarni predmet, fe'llarni harakat ifodalovchi so'z sifatida «baholaganlar».

Qadimgi hindlar so'zlarni tahlil qilish, tarkibini o'rganish jarayonida ularni quyidagi bo'laklarga ajratganlar: 1) o'zak, 2) suffiks, 3) qo'shimcha (turlovchi qo'shimcha). Shuningdek, so'z yasovchi va so'z o'zgartiruvchi morfemalar farqlangan.

Yevropa olimlari hind tilshunoslarining ishlari bilan yaqindan tanishib, so'zlardan o'zak, so'z yasovchi va so'z o'zgartiruvchi morfemalarni ajratishga «kirishganlar».

Hindlar otlarda yettita kelishikni qayd etganlar: 1) bosh kelishik, 2) qaratqich kelishigi, 3) jo'nalish kelishigi, 4) tushum kelishigi, 5) qurol kelishigi, 6) chiqish (ablativ) kelishigi, 7) o'rin kelishigi. Hind tilshunoslari qo'shma so'zlarning o'ttizga yaqin turini farqlaganlar. Ular qo'shma so'zlarning tuzilishida komponentlar orasidagi munosabatlarga e'tibor berganlar. Masalan: ot+ot//fe'l; sifat// sifatdosh //ravish+ot// sifat//fe'l; son+ot va boshqalar. Hind grammatikachilari fe'l turkumining morfologik kategoriyalarini mukammal ishlagan edilar. Ular fe'lning uch zamonga birlashadigan yetti xil zamon formasini: hozirgi zamon, o'tgan zamonning tugallangan, tugallanmagan, uzoq o'tgan zamon turlarini, kelasi zamon, odatdagi kelasi zamon va juda kam qo'llaniladigan shart fe'li shaklini ajratganlar.

Rasulov R. Umumiy tilshunoslik. – Toshkent, 2013. Ирискулов М. Тилшуносликка кириш. – Т.: Ўқитувчи, 1992. Жамолхонов Х, Ҳозирги ўзбек адабий тили. –Т., 2005. Abduazizov A. O‘zbek tili fonologiyasi va morfonologiyasi. – Т.,1992. Sayfullayeva R., Mengliyev B., Boqiyeva G., Qurbonova M., Yunusova Z., Abuzalova M. Hozirgi o‘zbek adabiy tili.–Т., 2009. Ҳожиев А. Тилшунослик терминларининг изоҳли луғати.–Тошкент: Фан, 2002. Ўзбек тилининг изоҳли луғати. –Тошкент: «Ўзбекистон миллий энциклопедияси», 2008. Ўзбек тилининг изоҳли луғати. –Тошкент: Ўзбекистон миллий энциклопедияси, 2007. Jamolxonov H. O‘zbek tilining nazariy fonetikasi. pdf. –Т.:Fan, 2009. Tursunov U., Muxtorov A., Rahmatullayev Sh. “Hozirgi o‘zbek adabiy tili”. Darslik. –Т.: Fan,1992. Нурмонов А. Танланган асарлар. – Тошкент, 2012. www.ziyouz.com kutubxonasi

**Bog‘lanuvchi so‘zlar manzarasi: Til oilalari, hind-yevropa tillari.**

Умумий фонетикага тегишли “танглай товуши” терминининг луғатларда тавсифланишини таҳлил қиламиз. Н. Маҳкамов, И.Эрматовнинг “Тилшунослик терминлари луғати”да бу термин берилмаган. А.Ҳожиев луғатида “Орқа танглай ундошлари” бирикмаси ва “Танглай товуши” сифатида берилган:

ОРҚА ТАНГЛАЙ УНДОШЛАРИ. Тил орқа қисмининг юмшоқ танглай томон кўтарилиши натижасида ҳосил бўлувчи ундошлар, Мас., *х, з* ва *б. Қ.* Тил орқа ундошлари<sup>8</sup>.

Шунингдек, ушбу атама билан боғлиқ “танглай-тиш ундошлари” термини ҳам келтирилади:

ТАНГЛАЙ-ТИШ УНДОШЛАРИ. Ҳаво оқими оғиз бўшлиғида икки тўсиқдан (тил билан милк орасидан ва тил орқа қисми билан юмшоқ танглай орасидаги тўсиқдан) сирғалиб ўтиши билан ҳосил бўладиган тил олди ундошлари. Мас., *ж, ш* ундошлари<sup>9</sup>.

“Танглай товуши” луғат мақоласи жуда қисқа бўлиб, бошқа терминга ҳавола берилган:

<sup>8</sup> Ҳожиев А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 77.

<sup>9</sup> Ҳожиев А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 79.

ТАНГЛАЙ ТОВУШИ – қ. Палатал товуш<sup>10</sup>.

Ҳавола қилинган луғат мақоласи эса қуйидаги мазмунга эга:

ПАЛАТАЛ ТОВУШ. Тил орқа қисминииг қаттиқ танглай томон кўтарилиши билан ҳосил бўлувчи товуш: **й** ундоши.<sup>11</sup>

Шу луғат кўрсаткичида қуйидагича таржима қилинган:

Танглай-тиш ундошлари – небно-зубные согласные  
Танглай товуши – небный звук (палатальный звук)<sup>12</sup>.

Кўринадики, мазкур термин билан боғлиқ бир неча кичик луғат мақолалари мавжуд. Энциклопедик луғатда шуларни бирлаштириб, батафсил мақола шакллантириш мумкин. Рус тилидаги лингвистик луғатларда *небный звук* (*палатальный звук*) терминлари мавжуд эмас. Бу масала билан боғлиқ терминлар ва уларнинг тавсифи қуйидагича:

ПАЛАТАЛИЗАЦИЯ ЗАДНЕЯЗЫЧНЫХ СОГЛАСНЫХ ВТОРАЯ. Изменение заднеязычных (заднеязычных) согласных в свистящие перед \*e, \*i дифтонгического характера: \*d, \*k, \*ch – z', ц', с' (друг – друзья и др.).

ПАЛАТАЛИЗАЦИЯ ЗАДНЕЯЗЫЧНЫХ СОГЛАСНЫХ ПЕРВАЯ. Изменение задне-язычных (заднеязычных) согласных в шипящие перед j и гласными переднего ряда в прасла-вянском языке, отраженное в современном русском языке в виде чередований к/ч, г/ж, х/ш: дух – душа и др.

<sup>10</sup> Ҳожиёв А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 102.

<sup>11</sup> Ҳожиёв А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 79.

<sup>12</sup> Ҳожиёв А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 152.



**ПАЛАТАЛИЗАЦИЯ ЗАДНЕЯЗЫЧНЫХ СОГЛАСНЫХ ТРЕТЬЯ.** Палатализация, осуществившаяся в поздний праславянский период и по своим результатам совпавшая со второй палатализацией. Заднеязычные г, к, х изменились в мягкие свистящие: з', ц', с'. Изменения наблюдались: 1) после гласных переднего ряда е,, , î, . и после слогаобразующего рь: мръцание

**ПАЛАТАЛИЗАЦИЯ СОГЛАСНЫХ** [< лат. palatum нёбу]. Смягчение согласных путем добавочного участия в артикуляции средней части языка (поднятия ее к нёбу). Например: нь, ль. Палатализация – один из видов дополнительной артикуляции. Накладываясь на основную артикуляцию, она и создает дополнительную окраску смягченного (палатализованного) согласного.

**ПАЛАТАЛИЗАЦИЯ СОГЛАСНЫХ ЗВУКОВ ЙОТОВАЯ.** Смягчение согласных в сочетании с [j] в праславянском языке было связано с действием закона открытого слога. Йотовой палатализации подверглись всеогласные звуки праславянского языка: ноша, ложа, вожжи, ключидр<sup>13</sup>.

Юкоридаги материаллардан фойдаланиб, мазкур терминни куйидагича кодировка қиламиз:

#### **Tanglay tovushi**

Tanglay tovushi. Til oldi – tanglay undoshlari – til oldi qismining qattiq tanglay tomon ko‘tarilishidan hosil bo‘lgan to‘siqda yuzaga keladigan undoshlar. **Q.** Palatal tovush

Хулоса сифатида айтиш жоизки, лингвистик энциклопедия учун кодировка қилишда турли манбалардаги маълумотларни таҳлил қилиш, уларни бирлаштириш ва ўзбек тили ходисаларига мос равишда луғат мақоласи сифатида шакллантириш мақсадга мувофиқ.

#### **АДАБИЁТ**

1. [https://en.wikipedia.org/wiki/Language\\_family](https://en.wikipedia.org/wiki/Language_family)
2. Жеребило Т.В. Словарь лингвистических терминов. Изд. 5-е, испр. и доп. – Назрань: ООО «Пилигрим», 2010. – 486 с. – С. 418.
3. Ҳожиёв А. Тилшунослик терминларининг изоҳли луғати. – Тошкент: «Ўзбекистон Миллий Энциклопедияси» давлат илмий нашриёти, 1997. – 164 б. – Б. 111.

<sup>13</sup> Жеребило Т.В. Словарь лингвистических терминов. Изд. 5-е, испр. и доп. – Назрань: ООО «Пилигрим», 2010. – 486 с. – С. 249–250.

---

## МАШИННОЕ ОБУЧЕНИЕ

УДК 81'322.4

### THE DEVELOPMENT OF NEURAL MACHINE TRANSLATION MODELS FOR THE KAZAKH-RUSSIAN PAIR OF LANGUAGES

*Vladislav Karyukin<sup>1</sup>, Nilufar Abdurakhmonova<sup>2</sup>*

*<sup>1</sup>Al-Farabi Kazakh National University, Institute of Information and Computational Technologies, Almaty, Kazakhstan*

*<sup>2</sup>National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan*

*vladislav.karyukin@gmail.com, n.abduraxmonova@nuu.uz*

Applied intelligent systems have made remarkable strides in addressing a multitude of challenges. Among these advancements, machine translation of natural languages is a pivotal direction. Automated translation plays a significant role in assisting professional translators. The development of effective machine translation systems hinges on the availability of substantial parallel corpora for various resource-rich language pairs. For many European languages, extensive corpora are readily accessible through various platforms, while it does not apply to the Kazakh language. It lies in the resource-low group of languages. This paper explores the preparation of the Kazakh-Russian machine translation models, taking the following steps: the preparation of the parallel corpora, its preprocessing, training with the neural machine translation architectures, such as Seq2Seq based on RNN and Transformer, and evaluating with popular machine translation metrics: BLEU, WER, and TER. The trained models have shown effective results with metrics scores above 0.39.

**Keywords:** Neural machine translation, parallel corpora, forward translation, low-resource languages, Kazakh, Russian

### РАЗРАБОТКА МОДЕЛЕЙ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА ДЛЯ КАЗАХСКО-РУССКОЙ ПАРЫ ЯЗЫКОВ

*В. И. Карюкин<sup>1</sup>, Н. З. Абдурахмонова<sup>2</sup>*

*<sup>1</sup>Казахский национальный университет им. Аль-Фараби Алматы, Казахстан*

*<sup>2</sup>Национальный университет имени Мирзо Улугбека Ташкент, Узбекистан*

*vladislav.karyukin@gmail.com, n.abduraxmonova@nuu.uz*

Прикладные интеллектуальные системы добились значительных успехов в решении множества проблем. Среди этих достижений ключевым направлением является машинный перевод естественных языков. Автоматический

перевод играет важную роль в помощи профессиональным переводчикам. Разработка эффективных систем машинного перевода зависит от наличия больших параллельных корпусов многоресурсных языковых пар. Для многих европейских языков большие корпуса доступны на различных платформах, но это не относится к казахскому языку. Он относится к группе малоресурсных языков. В данной статье рассматривается подготовка казахско-русских моделей машинного перевода, включающая следующие шаги: подготовка параллельных корпусов, их предварительная обработка, обучение с помощью архитектур нейронного машинного перевода, таких как Seq2Seq на основе RNN и Transformer, а также оценка значимыми метриками машинного перевода: BLEU, WER и TER. Обученные модели показали эффективные результаты со значениями выше 0.39.

**Ключевые слова:** нейронный машинный перевод, параллельные корпуса, прямой перевод, малоресурсные языки, казахский, русский

## 1. Introduction

Today, applied intelligent systems that solve many problems have become widely developed. One of them is the direction of machine translation (MT) of natural languages. MT allows us to eliminate various language barriers between people. It also allows you to automate translation processes for various areas of society, including business, medicine, economics, etc. Automated translation is often used to help professional translators to speed up and simplify their work. After automatic translation, all that remains is to edit the stylistic and grammatical errors in the text.

The development of MT systems [Babulgaonkar, 2017] requires the availability of a large volume of parallel corpora for different language pairs. Large corpora are available for most European languages, such as English, Spanish, Portuguese, German, and Czech. They can be found on the websites of OPUS, WMT, CLARIN, and others. More than 1 billion parallel sentences are available for the English-Spanish language pair, more than 400 million for the English-Portuguese language pair, and more than 250 million for the English-Czech language pair.

For many other languages, the situation with the availability of language resources is significantly different. Not only are parallel corpora difficult to find for some, but quality monolingual texts are also a challenge. This especially applies to the Kazakh language, which is an agglutinative language with complex morphological and syntactic structures. Nevertheless, the need for translation of Kazakh texts is growing every year due to the emergence of a large number of legal, educational, and technical texts. Existing MT systems [Singh,

2019] can fully translate texts, but they still contain a large number of errors. When translating from Russian into Kazakh, various errors may occur since the Kazakh language differs from other languages and has special characteristics: the proximity of the lexical structure, the law of synchronism, agglutination (a series of affixes), a special word order (in the Kazakh language the word order in phrases and types of connections between words are strictly defined), etc.

The collection of parallel corpora for the experiment was carried out using both a parser and a web crawler that analyzed various multilingual websites and using existing MT platforms such as Google Translate and Promt. The open-source application Bitextor was used for parsing, which uses a set of rules to identify the same texts in two different languages. Bitextor was applied to the sites of the Kazakh National University (<http://www.kaznu.kz>), Bolashak International Scholarship (<http://www.bolashak.gov.kz>), Eurasian National University (<http://www.enu.kz>), Kazakh mail (<http://www.kazpost.kz>), news portals (<http://inform.kz>, <http://tengrinews.kz>), etc.

After collecting the data, they are processed, during which unnecessary elements, signs, symbols, etc. are removed. After processing, monolingual texts are translated into the target language, forming parallel corpora [Laskar, 2023] together with the source language. Several architectures, such as Seq2Seq (RNN, BRNN) and Transformer, were used to train the model. Model training results were verified, tested, and evaluated using BLEU, TER, and WER metrics.

## **2. Related works**

The problems of MT are covered in many research papers. The paper [Mondal, 2023] focuses on statistical and neural MT research. Specifically, it covers the history of their development and transformation. In addition, it highlights the significant roles of enhancing MT techniques. [Datta, 2023] describes the intensive development of NMT technologies for low-resource languages. The problem of scarcity of resources for some languages, called low-resource languages, is a serious barrier to building high-quality models.

In [Heryadi, 2023], the large experimental research of building statistical machine translation (SMT) and NMT models for English-Bangla and English-Hindi pairs was done. Both Bengali and Hindi belong to the Indian groups of languages, and they are considered to be poor resources. NMT models with Byte pair encoding (BPE) and Transform-

ers architecture for English-Hindi and Hindi-English pairs received BLEU scores from 28.8 to 39.9, while Attention-based NMT for the Bengali-Hindi pair reached a score of 20.41. [Tukeyev, 2020] analyzes the performance of the Long Short-term Memory (LSTM) NMT model involving Indonesian and Sundanese languages. Both trained models achieved accuracy scores of 0.92 for training and 0.88 for testing.

The Kazakh language is also a low-resource language, and a number of works are devoted to building NMT models for various pairs of languages that include Kazakh. The study [Liu, 2023] covered the morphological segmentation for the Kazakh language with the complete set of endings (CSE). The Tensorflow Seq2Seq model was used in the experimental part to train the parallel corpora of the Kazakh-English 109,772 sentences, resulting in the BLEU score values from 0.18 to 0.25. The work [Karyukin, 2023] used back translation data enhancement methods to increase the Kazakh-Chinese corpora and train NMT models. The experimental results in this research showed that this approach could increase the BLEU scores for Chinese-Kazakh and Kazakh-Chinese translations by 4.47 and 5.97, respectively.

### 3. Methodology

Due to the insufficient volume of parallel corpora for the Russian-Kazakh language pair, the approach of their formation using direct translation (Forward Translation) by the Google Translate and Prompt systems was used. In this method, the source language corpora are translated into the target language by a translation system. The generated synthetic data is then added to existing parallel corpora to increase their size. Forward Translation [Wang, 2022] is also used to check the quality of parallel corpora. Translated texts are also checked by comparing them with the original version. This can help identify errors, ambiguity, or confusion that may arise in the translation.

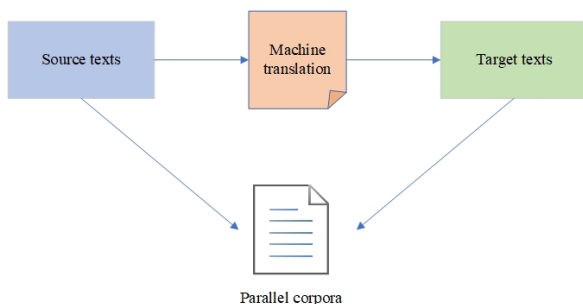
This approach is implemented in the following sequence of steps:

There are source texts  $X = \{x\}_{i=1}^N$ , where  $X$  is a source set of texts;  $S_t$  is a translation system.

Source texts  $X$  are translated by the translation system  $S_t$  for getting target texts  $Y = \{y\}_{i=1}^N$ .

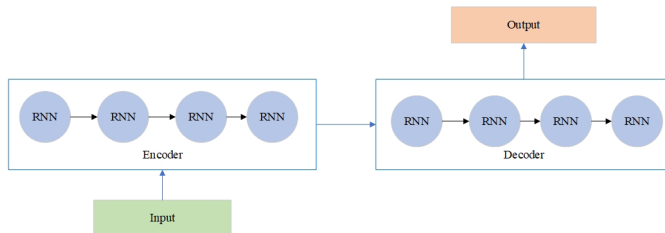
The resulting source and target texts  $X$  and  $Y$  are gathered to obtain a new parallel corpus.

A scheme of the Forward Translation approach is shown in Figure 1.



**Figure 1.** Forward Translation approach [9]

There were 380 thousand Russian-Kazakh parallel sentences generated from a text corpus. NMT (Neural Machine Translation) architectures, which have become widespread in this field in recent years, are actively used to train MT models. One of the most famous NMT architectures is Seq2Seq [HaiLong, 2023], designed to translate text from one language to another. This deep learning model can take sequences of elements, be they words, letters, time series, etc., and generate other sequences of elements. The structure of this model consists of two main components: an encoder and a decoder. The encoder processes the input data and transforms it into a common vector called context. This vector is then passed to the decoder, which generates the output sequence. Since this task involves serial data, forms of recurrent neural networks are usually used for the encoder and decoder, such as LSTM (Long short-term memory), GRU (Gated recurrent unit), etc. The latent state vector can have different sizes, although its size is often chosen as a power of two, and its length is usually 128, 256, 512, or 1024. The structure of the Seq2Seq architecture is shown in Figure 2.



**Figure 2.** Seq2Seq architecture [9]

The so-called Transformer model was the next architecture that became particularly successful in MT tasks. The Transformer model

[Saxena, 2021] is a neural network architecture that has become the basis for many modern MT systems and has significantly advanced natural language processing. RNN layers are replaced by attention layers. As a result, Transformer does not process input text sentences in sequential order. Instead, the internal attention mechanism identifies the context that gives meaning to the input sequence. Hence, it achieves high parallelization and reduces training time.

In this architecture, the input data is the source text. It is divided into tokens: individual words and subwords; each token is encoded using embedding that takes into account the order of words in the sentence. The attention mechanism allows the model to focus on different parts of the input text and observe the context, and feedforward neural network layers process the information after the attention mechanism. Each decoding layer includes an attention mechanism to process the output from the encoder and predict the next token in the target text. The decoder also uses normalization, adding connections and positional embeddings. The architecture of the Transformer is shown in Figure 3.

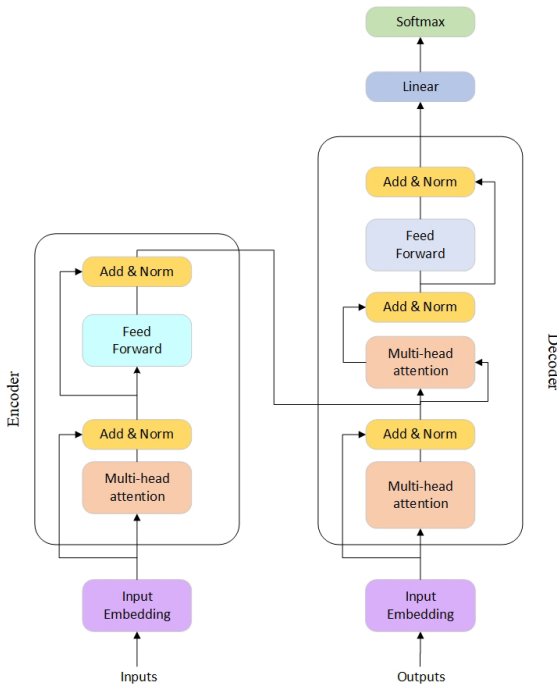


Figure 3. Transformer architecture [9]

When the model is trained, it is necessary to evaluate MT quality with special metrics called BLEU [Mouratidis, 2020], WER, and TER [Abdul Rauf, 2011], which show a high correlation with human quality ratings. These metrics' values range from 0 to 1 (from 0% to 100% in percentage).

BLEU is a very popular metric for evaluating the quality of MT. It generally measures the similarity between the machine-generated translation and the reference translation. The formula of the BLEU metric is

$$BLEU = BP * e^{\sum_{i=1}^n P_i}, \quad (1)$$

where  $BP$  is a factor for penalizing short translation;  $P_i$  is the precision of  $n$ -grams of order  $i$  in MT compared to the reference translation;  $n$  is the maximum  $n$ -gram order.

TER is a metric that estimates the number of different edits (insertion, deletion, and substitution) required for transforming MT into the reference translation. The formula of the TER metric is

$$TER = \frac{S + D + E}{N}, \quad (2)$$

where  $S$  is the number of substitutions;  $D$  is the number of deletions;  $E$  is the number of shifts;  $N$  is the total number of words in the reference translation.

#### 4. The experiments

The parallel Kazakh-Russian corpus in the amount of 380 thousand sentences was used to train the MT model. The whole corpus was split into the training part – 90%, the validation part – 5%, and the testing part – 5%. The experiment was run on a computer with the following characteristics: Core i7 4790K CPU, 32GB RAM, 1TB SSD, and RTX 2070 Super and GTX 1080 GPUs.

The open-source OpenNMT framework built on top of the popular deep learning framework PyTorch was implemented for training the models. One of the most significant advantages of this framework is that it supports various NMT architectures, including Seq2Seq and Transformer. It also provides instruments for tokenization and data preparation, significantly simplifying MT model preparation. The framework also allows configuring neural network parameters, such as the number of layers, hidden dimensions, and dropout rates.



Once the corpus was prepared for training, the NMT architectures were configured and utilized for training the models. Then, the quality of translation was evaluated with BLEU, TER, and WER metrics. The experimental results are presented in Table 1.

**Table 1 – Evaluation of NMT translation models for the Kazakh-Russian corpora**

| Architecture | BLEU | WER  | TER  |
|--------------|------|------|------|
| RNN          | 0.42 | 0.55 | 0.48 |
| Transformer  | 0.39 | 0.62 | 0.55 |

The trained models showed good performance results for the Kazakh–Russian corpora, comparable with the high-resource trained models. All three metrics scores proved that the proposed NMT architectures have been useful in the preparation of high-quality NMT models.

## 5. Conclusions

This paper is devoted to the development of NMT models for the low-resource languages, specifically for the Kazakh-Russian language pair. One of the significant problems of low-resource languages is the absence of necessary resources for training high-quality MT models. In this way, the parallel corpora in a size of 380 thousand sentences were generated by the Forward Translation approach with the use of the Google Translate and Prompt systems. Then, the data is processed, and the OpenNMT machine translation framework is utilized for training the models. The Seq2Seq based on RNN and Transformer models are configured in the framework. One of the most significant advantages of this framework is that it supports various NMT architectures, including Seq2Seq and Transformer. The translation quality of the models was evaluated with BLEU, TER, and WER metrics. The results shown by the trained models were effective, with the metrics scores above 0.39. Nevertheless, there is much work to be done to increase the size of the parallel corpora, not only for the Kazakh-Russian pair of languages but for the ones, too. It is planned to implement this research in the future.

## REFERENCES

1. A.R. Babhulgaonkar, S.V. Bharad. Statistical machine translation. In 1st International Conference on Intelligent Systems and Information

Management (ICISIM), Aurangabad, India, pp. 62-67, 2017. <https://doi.org/10.1109/ICISIM.2017.8122149>.

2. S.P. Singh et al. 2019. Overview of Neural Machine Translation for English-Hindi. International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 27–28 September 2019, pp. 1–4. <https://doi.org/10.1109/ICICT46931.2019.8977715>.

3. S.R. Laskar, B. Paul, P. Dadure, R. Manna, P. Pakray, S. Bandyopadhyay. English–Assamese neural machine translation using prior alignment and pre-trained language model. *Computer Speech & Language*, vol. 82, 101524, 2023. <https://doi.org/10.1016/j.csl.2023.101524>.

4. S.K. Mondal, H. Zhang, H.M.D. Kabir, et al. Machine translation and its evaluation: a study. *Artificial Intelligence Review*, 56, 10137–10226, 2023. <https://doi.org/10.1007/s10462-023-10423-5>.

5. G. Datta, N. Joshi, K. Gupta. Performance Comparison of Statistical vs. Neural-Based Translation System on Low-Resource Languages. *International Journal on Smart Sensing and Intelligent Systems*, 16(1), 2023. <https://doi.org/10.2478/ijssis-2023-0007>.

6. Y. Heryadi, B. D. Wijanarko, D. F. Murad, C. Tho, K. Hashimoto. Neural Machine Translation Approach for Low-resource Languages using Long Short-term Memory Model. *International Conference on Computer Science, Information Technology and Engineering (IC-CoSITE)*, Jakarta, Indonesia, pp. 939-944, 2023. <https://doi.org/10.1109/ICCoSITE57641.2023.10127724>.

7. U. Tukeyev, A. Karibayeva, Zh. Zhumanov. Morphological segmentation method for Turkic language neural machine translation. *Cogent Engineering*, 7:1. <https://doi.org/10.1080/23311916.2020.1856500>.

8. C. Liu, W. Silamu, Y. Li. A Chinese–Kazakh Translation Method That Combines Data Augmentation and R-Drop Regularization. *Applied Sciences*, 13, 10589, 2023. <https://doi.org/10.3390/app131910589>.

9. V. Karyukin, D. Rakhimova, A. Karibayeva, A. Turganbayeva, A. Turarbek. The neural machine translation models for the low-resource Kazakh–English language pair. *PeerJ Computer Science*, 9:e1224, 2023. <https://doi.org/10.7717/peerj-cs.1224>.

10. X. Wang. Analysis of Machine Translation and Computer Aided Techniques in English Translation. In: Abawajy, J.H., Xu, Z., Atiquzzaman, M., Zhang, X. (eds) Tenth International Conference on Applications and Techniques in Cyber Intelligence (ICATCI 2022). ICATCI 2022. Lecture Notes on Data Engineering and Communications Technologies, vol. 170. Springer, Cham, 2023. [https://doi.org/10.1007/978-3-031-29097-8\\_91](https://doi.org/10.1007/978-3-031-29097-8_91).

11. Y. HaiLong, S. Wei, L. Lei, Zh. Jing, C. Chuan, X. Cunlu Xu. Pre-training model for low-resource Chinese–Braille translation. *Displays*, vol. 79, 102506, 2023. <https://doi.org/10.1016/j.displa.2023.102506>.

12. S. Saxena, S. Chauhan, P. Daniel. Analysis of Unsupervised Statistical Machine Translation Using Cross-Lingual Word Embedding for English–Hindi. *Topical Drifts in Intelligent Computing. ICCTA 2021. Lecture Notes in Networks and Systems*, vol. 426, Springer, Singapore, 2022. [https://doi.org/10.1007/978-981-19-0745-6\\_7](https://doi.org/10.1007/978-981-19-0745-6_7).

13. D. Mouratidis, K.L. Kermanidis, V. Sosoni. Innovative Deep Neural Network Fusion for Pairwise Translation Evaluation. *Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology*, vol 584. Springer, Cham, 2020. [https://doi.org/10.1007/978-3-030-49186-4\\_7](https://doi.org/10.1007/978-3-030-49186-4_7).

14. S. Abdul Rauf, H. Schwenk. Parallel sentence generation from comparable corpora for improved SMT. *Machine Translation*, 25, pp. 341–375, 2011. <https://doi.org/10.1007/s10590-011-9114-9>.

УДК 81'322.4

**ПРИМЕНЕНИЕ МЕТОДА TRANSFER LEARNING  
К ЗАДАЧЕ МАШИННОГО ПЕРЕВОДА ДЛЯ ПАРЫ  
РУССКИЙ-ХАКАССКИЙ*****А. Ю. Лебедева***

*Национальный исследовательский университет  
«Высшая школа экономики»  
Санкт-Петербург, Россия  
annlebedeva.spb@gmail.com*

Статья посвящена применению метода transfer learning (передачи обучения) к задаче перевода для языковой пары русский-хакасский. В исследовании рассматривается процесс выбора языковой пары для предварительного обучения, сбора и предварительной обработки данных, а также описывается один из возможных способов токенизации, подразумевающий деление текста на подслово, эффективный для языков с богатой морфологией. Подробно описывается процесс обучения модели, несколько вариантов экспериментов, таких как настройка пропорции языков в словаре, аугментация при помощи технологии дропаута, подбор параметра максимальной длины последовательности. Результаты модели сравниваются с базовой моделью, обученной без transfer learning, а также с существующим русско-хакасским переводчиком и результатами других исследований по переводу малоресурсных языков. Исследование показывает, что предварительное обучение на паре русский-чувашский может значительно улучшить производительность модели для пары русский-хакасский.

**Ключевые слова:** *нейронный машинный перевод, малоресурсные языки, передача обучения.*

**APPLICATION OF THE TRANSFER LEARNING APPROACH TO  
TRAINING MACHINE TRANSLATION MODEL  
FOR THE RUSSIAN-KHAKAS LANGUAGE PAIR*****Lebedeva A.***

*National Research University Higher School of Economics  
Saint-Petersburg, Russia  
annlebedeva.spb@gmail.com*

The article is devoted to the application of the transfer learning method to the translation problem for the Russian-Khakas pair. The study examines the process of selecting a language pair for pre-training, data collection and pre-processing. One of the possible tokenization methods is described, which involves dividing text into subwords, which proved effective for languages with rich morphology. The process of training the model, several experimental options, such as adjusting the

proportion of languages in the dictionary, augmentation using dropout technology, and selecting the maximum sequence length parameter are described in detail. The model's results are compared with a baseline model trained without transfer learning, as well as with an existing Russian-Khakassian translator and the results of other studies on the translation of low-resource languages. The study shows that pretraining on the Russian-Chuvash pair can significantly improve the model performance for the Russian-Khakas pair.

**Keywords:** neural machine translation, low-resource languages, transfer learning.

## 1. Введение

Мы описываем результаты применения метода transfer learning к задаче перевода русско-хакасской языковой пары. Задача включала обучение базовой модели, выбор языковой пары для предварительного обучения, предварительную обработку, корректировку параметров модели, обучение и оценку модели. Модель обучалась в обоих направлениях перевода. Для русский-хакасской пары языков удалось достичь state-of-the-art результатов, и идеи для будущего улучшения будут представлены в конце исследования.

## 2. Обзор существующих исследований

Эффективность моделей нейросетевого машинного перевода существенно снижается, когда объем данных для обучения уменьшается [Knowles, 2020, p. 1112–1122]. Один из предложенных способов улучшения точности перевода – это передача знаний, полученных от пары языков с более богатыми ресурсами. В исследовании [Zoph, 2016, p. 1568–1575] предлагают метод transfer learning, который заключается в обучении рекуррентной модели на большом параллельном корпусе богатой языковой пары, а затем использование весов этой модели как начальной конфигурации для языковой пары с малыми ресурсами, при этом сохраняя эмбединги той стороны, где язык остается прежним. Это уменьшает необходимость в больших объемах специализированных размеченных данных, которые могут быть дорогостоящими и затратными по времени для получения. Transfer learning предоставляет хорошую исходную точку для модели, так как значимые признаки уже были выделены во время предварительного обучения. Эта инициализация помогает моделям быстрее сходиться во время дообучения на специализированных

данных, сокращая время обучения и вычислительные ресурсы. В этом случае предобученная модель часто называется родительской моделью, а дообученная – дочерней.

Валеев и др. [Валеев, 2019] также использовали метод передачи обучения [Zoph, 2016, p. 1568–1575]. Они использовали русско-казахскую пару для обучения родительской модели, а затем дообучали ее на русско-татарских данных. При обучении использовался общий словарь для всех трех языков. Таблица 1 сравнивает объемы данных, которые разные исследователи использовали для родительских и дочерних моделей соответственно при применении метода передачи обучения.

**Таблица 1. Размеры корпусов, используемых разными авторами для transfer learning**  
**Table 1. Dimensions of corpora used by different authors for transfer learning**

| Статья  | Родительский, пар предложений | Дочерний, пар предложений |
|---|-------------------------------|---------------------------|
| Валеев и др., 2019                              | 5,000,000                     | 324,000                   |
| Knowles и др., 2020, p. 1112–1122               | 22,000,000                    | 60,000                    |
| Zoph и др., 2016, p. 1568–1575                  | 53,000,000                    | 2,500,000                 |
| Космі и Војаг, 2018, p. 244–252, пара enFI-enET | 2,800,000                     | 800,000                   |
| Космі и Војаг, 2018, p. 244–252, пара enET-enCS | 40,100,000                    | 800,000                   |
| Нижон и др., 2023                               | 1,000,000                     | 36,601                    |

### 3. Данные

#### 3.1 Данные для дочерней модели

**Источник данных.** Для параллельных данных русско-хакасского языка мы использовали корпус TPL параллельных предложений [Mirzakhlov, 2021, p. 5876–5890], который состоит из 60,295 пар в обучающем наборе и 1,000 пар в наборах для валидации и тестирования соответственно.

**Предварительная обработка.** Изначально этот набор данных имел пробелы с обеих сторон знаков пунктуации и т.н. “мусор-

ные” символы, такие как ‘\*’, ‘& # 91’ или случайные числа. Мы удалили лишние пробелы перед знаками пунктуации для получения более корректного вывода при подаче модели новых предложений с правильной пунктуацией. Мы также очистили данные от мусорных символов и случайных чисел. Почти все хакасские предложения в корпусе TIL не имеют пунктуации, но некоторые предложения относятся к литературным произведениям, предоставленным нам Электронным корпусом хакасского языка<sup>1</sup>, поэтому нам удалось восстановить пунктуацию для этих предложений, что составило примерно половину набора данных.

### 3.2 Данные для родительской модели

Идея выбора языка для предварительного обучения модели включала в себя следующие факторы: а) он также должен быть на кириллице, так как мы используем общий словарь между родительскими и дочерними моделями, б) предпочтительно также тюркский, чтобы иметь схожую морфологию и синтаксис с хакасским, в) набор данных должен быть относительно большим, чтобы модель могла хорошо обучаться. Исходя из размера корпуса хакасского языка и работы [Нижон, 2023], представляющей хорошие результаты для дочернего набора схожего размера (36,601 пар) и имея 1,000,000 пар для предварительного обучения, мы решили выбрать пару русский-чувашский для предварительного обучения. Источником данных является Двухязычный корпус чувашского языка<sup>2</sup>. Мы дополнительно перетасовали предложения и получили в итоге 1,000,013 пар в обучающем наборе и 2,000 пар в наборе для валидации.

## 4. Методы

**Кодирование с помощью метода Byte-pair encoding.** Модели нейронного машинного перевода (NMT) обычно работают с фиксированным словарем, но для перевода периодически требуется использование неизвестных и редких слов. Это особенно актуально для агглютинативных языков, где слова формируются путем комбинирования меньших морфем, поэтому модели перевода требуют механизмов, работающих ниже уровня слова. Чтобы справиться с этим, предыдущие методы используют словари

<sup>1</sup> <https://khakas.altai.ca.ru/>

<sup>2</sup> <https://ru.corpus.chv.su/>

для неизвестных слов. Такой метод кодирования как Byte-pair encoding, который мы используем в этой работе, представлен в работе [Sennrich, 2016]. Он реализует кодирование слов как последовательностей подслов. Этот инструмент использует алгоритм Byte-pair encoding для задачи сегментации слов. Желаемое количество операций слияния может быть установлено как параметр.

**Transfer learning.** В этом исследовании мы используем метод передачи обучения, предложенный в работе [Kosmi, 2018, p. 244–252]: процесс обучения начинается с первоначального обучения на родительской языковой паре в течение определенного числа итераций, после чего происходит переход к дочерней языковой паре без сброса каких-либо (гипер)параметров. Этот подход схож с методом передачи обучения, представленным в работе [Zoph, 2016, p. 1568–1575], но использует общий словарь, как предложено в работе [Nguyen, 2017].

**Метрики оценки.** Bilingual Evaluation Understudy (BLEU) по-прежнему широко используется как автоматическая метрика оценки машинного перевода, несмотря на то, что считается, что у нее низкая корреляция с человеческой оценкой [Macháček, 2014, p. 293–301]. Она измеряет сходство между результатом машинного перевода и одним или несколькими эталонными переводами на основе точности n-грамм. Значение BLEU в процентах варьируется от 0 до 100, причем более высокий балл указывает на лучший перевод. Мы будем использовать BLEU для оценки наряду с ChrF (Character F-Score) [Popović, 2015, p. 392–395]: методом оценки, который измеряет перекрытие символьных n-грамм вместо словесных n-грамм, как в метрике BLEU. ChrF использует F-меру, которая сочетает в себе точность символьных n-грамм (ChrP) и полноту символьных n-грамм (ChrR). Нам не удалось использовать более современные фреймворки, такие как метрика оценки COMET, потому что для неё нет предварительно обученной модели для русско-хакасской языковой пары, а обучение метрики COMET для оценки этих языков требует размеченных человеком данных, которых также нет.

## 5. Эксперименты.

**Настройки обучения.** Все исследования были проведены с использованием инструментария Sockeye [Hieber, 2022]. Sockeye –



это открытый инструментарий для нейронного машинного перевода по принципу seq2seq. Текущая версия Sockeye – Sockeye 3 – основана на Transformer, типичной архитектуре кодировщик-декодировщик с механизмами внимания [Vaswani, 2017]. Мы использовали классическую конфигурацию из 6 слоев кодировщика и 6 слоев декодировщика с 8 головами внимания как на стороне кодировщика, так и на стороне декодировщика. Что касается настройки гиперпараметров, мы установили только следующие: *max-num-checkpoint-not-improved 30, bucket-width 10, keep-last-params 1, batch-size 2048, max-num-epochs 100, weight-tying none, optimized-metric bleu, cache-metric bleu, source-factors-combine concat*. Эти параметры были одинаковыми для базовой, родительской и дочерней моделей.

Модели обучались на одном графическом процессоре Nvidia Tesla V100 с 32 ГБ памяти на суперкомпьютере Университета Высшей школы экономики [Kostenetskiy, 2021].

Перед проведением экспериментов мы предварительно очистили данные, как описано в подразделе 3.1. Помимо предобучения модели на чувашском языке были проведены следующие эксперименты: изменение количества данных на хакасском языке, на основе которых создавались словари (колонка “Словарь” в таблице 2), аугментация данных при помощи технологии VPE-dropout (колонка “Dropout” в таблице 2), изменение параметра максимальной длины последовательности (колонка “Длина” в таблице 2).

**Бейзлайн.** Создание бейзлайна заключалось в обучении модели Sockeye с вышеуказанными параметрами только на русско-хакасских данных без какого-либо предварительного обучения. Результаты обучения этой модели приведены в таблице 2 под номером 1.

**Увеличение размера доли дочернего корпуса при формировании словаря.** Так как мы использовали общий словарь для языков, которые меняются от предобучения к дообучению, и учитывая размеры этих корпусов, мы первоначально продублировали данные на хакасском языке 16 раз, чтобы сделать этот корпус сопоставимым по размеру с данными на чувашском языке, для того чтобы токены на хакасском языке были равнозначно представлены в общем словаре, как предложено в работе [Knowles, 2020, p. 1112–1122]. Далее доля предложений на русском-хакасском была увеличена, и оказалось, что увеличение объема данных на хакасском языке – в нашем случае, дублирование корпуса 26

раз – улучшает результаты окончательной модели (модели 2 и 3 в таблице 2).

**Настройка WPE-dropout.** Мы применили WPE-dropout, как описано в работе [Provilkov, 2019]. Идея WPE-dropout заключается в том, что во время WPE-токенизации некоторые слияния отбрасываются, делая полученные последовательности более устойчивыми. В наших экспериментах мы применили dropout на стороне источника, как предложено в работе [Knowles и др., 2020, p. 1112–1122], и для некоторых экспериментов мы также выполнили операцию dropout 5 раз на одних и тех же данных, объединяя результаты, чтобы сделать наш набор данных больше и исходные последовательности более разнообразными, как это также было предложено в работе [Knowles и др., 2020, p. 1112–1122]. Это позволило нам улучшить результаты как в плане метрик BLEU, так и ChrF в обоих направлениях перевода (модели 3 и 4 в таблице 2).

**Максимальная длина последовательности.** Основная причина экспериментирования с максимальной длиной последовательности заключается в том, что, хотя 99% токенизированных последовательностей в данных русско-чувашского корпуса имеют длину менее 100 токенов, 99% последовательностей русско-хакасского корпуса имеют длину менее 75 токенов. Это дало нам идею, что корректировка этого параметра может улучшить результаты модели. По результатам эксперимента, это действительно улучшило показатель ChrF для направления русский-хакасский и показатели BLEU и ChrF для направления Kh-Ru (Модели 4 и 5 в таблице 2).

**Таблица 2. Результаты модели в различных экспериментах**  
**Table 2. Model results in various experiments**

| N | Словарь | Длина | Дропаут | Русский-чувашский |      | Чувашский-русский |      | Русский-хакасский |             | Хакасский-русский |             |
|---|---------|-------|---------|-------------------|------|-------------------|------|-------------------|-------------|-------------------|-------------|
|   |         |       |         | BLEU              | ChrF | BLEU              | ChrF | BLEU              | ChrF        | BLEU              | ChrF        |
| 1 | –       | 75    | 1       | –                 | –    | –                 | –    | 19                | 41.1        | 11.6              | 37.3        |
| 2 | 16      | 100   | 1       | 12.8              | 53.2 | 15.8              | 46.1 | 25.5              | 55.5        | 12.7              | 43.0        |
| 3 | 26      | 100   | 1       | 13                | 53.4 | 16                | 46.4 | 25.9              | 56          | 16.6              | 45.8        |
| 4 | 26      | 100   | 5       | 13                | 53.4 | 20.3              | 49.4 | <b>29.6</b>       | 55.5        | 17                | 46.8        |
| 5 | 26      | 75    | 5       | 12.4              | 53.8 | 17.1              | 49.5 | 28.5              | <b>57.6</b> | <b>17.2</b>       | <b>47.1</b> |

## 6. Результаты

### 6.1 Метрики

Результаты применения метода transfer learning к задаче машинного перевода хакасского языка показали результаты, сравнимые с единственным русско-хакасским переводчиком, доступным в Интернете<sup>1</sup>. Помимо transfer learning, автор модели также использовал метод back-translation для улучшения производительности модели. Данный подход описан в работе [Sennrich и др., 2016] и подразумевает перевод моноязычного корпуса существующей моделью перевода с дальнейшим прибавлением полученных данных к основному корпусу. Мы не использовали back-translation в нашем исследовании, но это перспектива для будущих улучшений. Тем не менее, даже без back-translation, наша модель превосходит существующую модель в направлении русский-хакасский на 7.26 пунктов BLEU. К сожалению, веса модели для сравнения не доступны публично, поэтому нет возможности проверить метрики на одних и тех же тестовых данных. Результат сравнения приведен в таблице 3.

**Таблица 3. Результаты работы нашей модели в сравнении с существующей**  
**Table 3. Results of our model in comparison with the existing model**

| Модель                | русский-хакасский | хакасский-русский |
|-----------------------|-------------------|-------------------|
| github/adeshkin TL    | 17.2              | 17.78             |
| github/adeshkin TL+BT | 20.24             | 18.73             |
| Наша модель           | <b>28.5</b>       | 17.2              |

### 6.2 Анализ удовлетворительных результатов и ошибок

**Анализ удовлетворительных результатов.** Визуальный анализ переводов показал, что модель иногда способна правильно переводить полные предложения, хотя в основном не с точными словами, как в эталонном переводе, но схожими по смыслу. Вот некоторые примеры (S = Исходное предложение, R = Эталонное предложение, T = Переведенное предложение):

S: Хайзы харындас-пичелер, туған-чағыннар, чыылызып, колхозха кірер-кірбестенер алай аалдаң пасха чирзер тизердеңер чөптес турғаннар.

<sup>1</sup> [https://github.com/adeshkin/khakas\\_russian\\_translator](https://github.com/adeshkin/khakas_russian_translator)

R: Некоторые братья и сёстры, родственники, собравшись, обсуждали, нужно ли входить в колхоз или необходимо уехать из села в другое место.

T: Кто-то из них родственники и сестёр, собравшись, уговаривали о том, как или не входить в колхоз, или переправляться из деревни в другое место.

S: Сах андох тигір чалынни түскен, күгүрт күзүреен паза табыстар истіле түскеннер. Чир угаа тың тітіреп сыххан. Кізілер чир ұстүнде чуртааннаң сығара, чир пір дее іди тітіребеен полған. Угаа тың тітіреп сыххан! Угаа илбек тітірес!

R: Тогда засверкали молнии, зазвучали голоса, загремел гром, и произошло сильное землетрясение. Такого мощного землетрясения не было за все время существования людей на земле! Столь великим было это землетрясение!

T: И произошли молнии, громы и голоса, и сделалось великое землетрясение, какого не бывало с тех пор, как люди на земле. Такое землетрясение! Так великое!

S: Солдаттар пазох сых килділер

R: Солдаты снова вышли.

T: Солдаты опять вышли.

S: Худай Пабабысха хачан даа сабланыс ползын. Аминь.

R: Богу же и Отцу нашему слава во веки веков! Аминь.

T: Хвала Богу и Отцу во веки веков. Аминь.

**Анализ ошибок.** Чаще всего модель пропускает некоторые слова или добавляет дополнительные слова. Примеры:

S: – Тооза тартын полбаза чаалазарға килем

R: – Если не сможет до конца натянуть, воевать приду.

T: – Если не натянет – воевать приду.

S: Пір хатап ұс харындас аңнап партырлар

R: Один раз три брата на охоту пошли, говорят.

T: Однажды три брата на охоту поехали.

S: Ана іди міні үгредерге ыспааннар.

R: Вот так меня не отправили на учёбу.

T: Вот так меня не послали.

В основном модель переводит некоторые слова правильно, но общий смысл предложения теряется.

S: Хайдағ андағ ниме мағаа анда таап алды полар, хайди ла хатығлирға итче полар?

R: Интересно, что она там для меня выискала, какое наказание готовит?

Т: Что это за вещь там нашла меня, наверно, как только хочет наказать?

С: Хаçан ізік хыринда прайзы апсыр сыхханнарындох, Алиса кем килгенін сизін салған.

Р: и, соответственно, Алиса догадалась, кого увидит, как только у дверей все зачихали.

Т: Когда у дверей все появились в молодости, Алиса заметила, кто пришел.

## 7. Заключение и дальнейшие исследования

Применение метода transfer learning показало хорошие результаты в улучшении производительности модели перевода на русско-хакасских данных. Будущие исследования могут включать:

1. Корректировку количества операций слияния ВРЕ таким образом, чтобы токенизация была максимально морфологической, то есть слова делились бы на части, максимально близкие к их морфологической структуре. Вопрос о том, должна ли токенизация повторять морфологию, в литературе не полностью обсуждается и является одним из направлений будущих исследований.

2. Другой способ расширения исследования будет заключаться в изучении влияния увеличения размера дочернего корпуса. К сожалению, параллельные данные для русско-хакасской пары языков очень ограничены. Но, кроме корпуса TIL, существует около 7 литературных произведений, переведенных с русского на хакасский, также предоставленных Электронным корпусом хакасского языка, но они не выровнены, поэтому они еще не могут составлять параллельный корпус. Использование инструментов выравнивания, таких как `lingtrain aligner`<sup>1</sup>, может помочь сопоставить параллельные строки таких книг, и этот инструмент показал перспективные результаты, когда мы начали его использовать. Есть возможность, что мы расширим параллельный корпус этим способом и проверим, насколько это влияет на производительность модели.

3. Также возможным улучшениям может способствовать применение метода back-translation, описанного в разделе 6.1. Этот метод можно использовать в нашем случае, так как существуют книги на хакасском языке, доступные в Электронном корпусе хакасского языка, а также возможно использование данных газет после веб-скрейпинга.

<sup>1</sup> <https://github.com/averkij/lingtrain-aligner>

4. Другой способ экспериментов может включать в себя попытку использовать другой тюркский язык для родительской модели, например, казахский.

5. Использование ансамбля описанных подходов может помочь достичь лучшего качества результирующей модели.

#### СПИСОК ЛИТЕРАТУРЫ

1. Hieber, F., Denkowski, M., Domhan, T., Barros, B.D., Ye, C.D., Niu, X., Hoang, C., Tran, K., Hsu, B., Nadejde, M., et al.: Sockeye 3: Fast neural machine translation with pytorch. arXiv preprint arXiv:2207.05851 (2022).

2. Hujon, A.V., Singh, T.D., Amitab, K.: Transfer learning based neural machine translation of english-khasi on low-resource settings. *Procedia Computer Science* 218, 1–8 (2023).

3. Knowles, R., Larkin, S., Stewart, D., Littell, P.: Nrc systems for low resource german-upper sorbian machine translation 2020: Transfer learning with lexical modifications. In: *Proceedings of the Fifth Conference on Machine Translation*. pp. 1112–1122 (2020).

4. Kocmi, T., Bojar, O.: Trivial transfer learning for low-resource neural machine translation. In: *Proceedings of the Third Conference on Machine Translation: Research Papers*. pp. 244–252. Association for Computational Linguistics, Brussels, Belgium (Oct 2018). <https://doi.org/10.18653/v1/W18-6325>, <https://aclanthology.org/W18-6325>.

5. Koehn, P., Knowles, R.: Six challenges for neural machine translation (2017).

6. Kostenetskiy, P., Chulkevich, R., Kozyrev, V.: Hpc resources of the higher school of economics. In: *Journal of Physics: Conference Series*. vol. 1740, p. 012050. IOP Publishing (2021).

7. Mach'áček, M., Bojar, O.: Results of the wmt14 metrics shared task. In: *Proceedings of the Ninth Workshop on Statistical Machine Translation*. pp. 293–301 (2014).

8. Mirzakhlov, J., Babu, A., Ataman, D., Kariev, S., Tyers, F., Abduraufov, O., Hajili, M., Ivanova, S., Khaytbaev, A., Laverghetta Jr, A., et al.: A large-scale study of machine translation in turkic languages. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. pp. 5876–5890 (2021).

9. Nguyen, T.Q., Chiang, D.: Transfer learning across low-resource, related languages for neural machine translation. arXiv preprint arXiv:1708.09803 (2017).

10. Popovi'c, M.: chrF: character n-gram f-score for automatic mt evaluation. In: *Proceedings of the tenth workshop on statistical machine translation*. pp. 392–395 (2015).

11. Provilkov, I., Emelianenko, D., Voita, E.: Bpe-dropout: Simple and effective subword regularization. arXiv preprint arXiv:1910.13267 (2019).
12. Sennrich, R., Haddow, B., Birch, A.: Improving neural machine translation models with monolingual data (2016).
13. Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units (2016).
14. Valeev, A., Gibadullin, I., Khusainova, A., Khan, A.: Application of low-resource machine translation techniques to russian-tatar language pair (2019).
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* 30 (2017).
16. Zoph, B., Yuret, D., May, J., Knight, K.: Transfer learning for low-resource neural machine translation. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. pp. 1568–1575. Association for Computational Linguistics, Austin, Texas (Nov 2016). <https://doi.org/10.18653/v1/D16-1163>, <https://aclanthology.org/D16-1163>.

УДК 81'33, 81.512.1

**РАСПОЗНАВАНИЕ ИМЕНОВАННЫХ ОБЪЕКТОВ  
НА ОСНОВЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ  
ТАТАРСКОГО ЯЗЫКА*****В. Р. Гафарова<sup>1</sup>, Ф. М. Гафаров<sup>2</sup>****<sup>1</sup>Институт прикладной семиотики Академии наук РТ  
Казань, Татарстан, Россия**<sup>2</sup>Казанский федеральный университет  
Казань, Татарстан, Россия*

79046639045@yandex.ru, fgafarov@yandex.ru

В этой работе мы представили результаты использования методов глубокого обучения для решения задачи распознавания именованных сущностей (NER) для татарского языка. Для решения этой задачи мы использовали два типа моделей нейронных сетей. В качестве первой модели использована система, основанная на сверточной нейронной сети (CNN) и двунаправленной рекуррентной LSTM нейронной сети (Bi-LSTM) с условным случайным полем (CRF). Вторая модель – это модель, основанная на предварительно обученной нейронной сети BERT. Для обучения и тестирования моделей мы использовали корпус из 5333 предложений. Экспериментальные результаты показывают, что модель на основе BERT работает лучше, чем Bi-LSTM, CNN и CRF, и достигает 86,7% по показателю F1.

**Ключевые слова:** Распознавание именованных сущностей, глубокое обучение, сверточная нейронная сеть, рекуррентная нейронная сеть, BERT

**NAMED ENTITY RECOGNITION BASED ON DEEP LEARNING  
METHODS FOR TATAR LANGUAGE*****Gafarova V. R.<sup>1</sup>, Gafarov F. M.<sup>2</sup>****<sup>1</sup>Institute of Applied Semiotics, Tatarstan Academy of Sciences,  
Kazan, Russian Federation**<sup>2</sup>Kazan Federal University  
Kazan, Russian Federation*

79046639045@yandex.ru, fgafarov@yandex.ru

Named Entity Recognition (NER) is a fundamental task in natural language processing. NER for English texts has been extensively researched in recent years, but only limited research has focused on Tatar NER due to the lack of resources for Tatar named entities. In this work, we used a recently created labeled NER dataset for the Tatar language. Neural networks are a powerful tool for learning representations of data with multiple levels of abstraction. Therefore the named entity recognition system has been developed by using neural network-based approaches, based on researches conducted in other languages [Richa,2020] and



by using the latest methods in NLP, as the vector representation of words. We used two types of models, and compared their prediction accuracy. The first type of models based on convolutional neural network (CNN) and bidirectional long short-term memory neural network (Bi-LSTM), with conditional random field (CRF) layers. For Bi-LSTM, CNN and based models initially, we used GloVe model to represent words in semantic vectors as input vector of neural networks.

Another problem in this research is the small amount of training data, with only a few thousand training examples per dataset. To address this problem, we used pre-trained BERT model. BERT is designed to pre-train deep bidirectional representations from unannotated text by jointly processing both left and right contexts at all levels based on Transformer architecture. Transformers have generally outperformed convolutional neural networks, recurrent neural networks and long-term memory networks (LSTMs). Therefore, in this work we used also a pre-trained BERT, and trained the model to perform named entity recognition. Experimental results show that the trained model performance for Bi-LSTM, CNN and CRF reached 71%, for BERT based model 86.7% on F1 measure on train dataset.

**Keywords:** Named entity recognition, deep learning, convolutional neural network, recurrent neural network, BERT

## Introduction

Named Entity Recognition (NER) is a fundamental task in natural language processing. NER for English texts has been extensively researched in recent years, but only limited research has focused on Tatar NER [Nevzorova, 2018] due to the lack of resources for Tatar named entities. In this work, we used a recently created labeled NER dataset for the Tatar language, presented in [Khakimov, 2022]. Neural networks are a powerful tool for learning representations of data with multiple levels of abstraction. Therefore, the named entity recognition system has been developed by using neural network-based approaches, based on researches conducted in other languages [Richa,2020] and by using the latest methods in NLP, as the vector representation of words. We used two types of models, and compared their prediction accuracy.

The first type of models based on convolutional neural network (CNN) and bidirectional long short-term memory neural network (Bi-LSTM), with conditional random field (CRF) layers. For Bi-LSTM, CNN and based models initially, we represented words as semantic vectors by GloVe model, to use then as an input vectors of neural networks.

Another problem in this research is the small amount of training data, with only a few thousand training examples per dataset. To address this problem, we used pre-trained BERT model [Kotei, 2023].

BERT is designed to pre-train deep bidirectional representations from unannotated text by jointly processing both left and right contexts at all levels, based on Transformer architecture. Transformers have generally outperformed convolutional neural networks, recurrent neural networks and long-term memory networks (LSTMs). Therefore, in this work we used also a pre-trained BERT, and trained the BERT-based model to perform named entity recognition. Experimental results show that the trained model performance for Bi-LSTM, CNN and CRF reached 71%, for BERT based model 86.7% on F1 measure on train dataset.

### Related work

Named entity recognition is both a text analysis task and a natural language processing task that aims to extract predefined named entities from a given text. A named entity can be an entity that can be encountered in all domains in general, such as “PERSON”, “LOCATION” or “ORGANIZATION” [Li, 2022]. On the other hand, there may be subject-oriented named entities that are widely used in some domains such as medicine, genetics, finance, business, government entities, agriculture, history, archaeology, environment and so on [Zhang, 2020, Raza, 2022, Repke, 2021]. Recognizing named entities is necessary for various high-level semantic applications such as knowledge graph construction. Named Entity Recognition is an essential NLP task that is used in many applications such as information retrieval, question answering, and machine translation [Li, 2022]. Currently, deep learning methods have been widely used for this task and have achieved remarkable results.

Paper [Sheping, 2022]. presents a model BERT-Conv-RBiGRU (BCRB), based on the combination of BERT and neural network for Chinese named entity recognition task on MSRA and OntoNotes dataset. The experimental results show that the model achieves very good results on with F1 values of 94.95% (MSRA) and 77.74% (OntoNotes). By using the WeiboNER data set and the People-Daily2004 data set for English–Chinese cross-lingual named entity recognition task authors obtained 53.22% F1 value of the optimal model [Wang, 2023].

A complex named entity recognition model for Hindi language, based on convolutional neural network (CNN), bidirectional long short-term memory (Bi-LSTM) neural network and conditional random field (CRF) is presented in [Richa, 2020]. Results obtained by authors show that the proposed architecture performs better as compared to the others

based on recurrent neural network, long short-term memory and Bi-LSTM individually. In paper [Ajees, 2018] authors propose a neural network based NER system for Malayalam. With a small number of features (authors used a corpus of 20615 sentences), the system was able to obtain the state-of-the-art performance in NER for Malayalam. In other work [Litake, 2023], by using different variations of BERT like base-BERT, RoBERTa, and ALBERT by publicly on the basis of available Hindi and Marathi NER datasets, authors provide an exhaustive comparison of different monolingual and multilingual transformer-based models. Authors shows that the monolingual MahaRoBERTa model performs the best for Marathi NER whereas the multilingual XLM-RoBERTa performs the best for Hindi NER.

In paper [Zali, 2018], on the basis of the study conducted in other languages a Persian named entity recognition system has been developed based on neural network. In other work a Named Entity Recognition of Arabic (NERA) is also designed based on neural network approach [Mohammed, 2012]. The results showed that the neural network approach achieved better than decision tree, the maximal accuracy of the system reached 92 %. Two RNN-based models by fine-tuning the pretrained BERT language model for NER task on Classical Arabic NER dataset (CANERCorpus) [Alsaaran, 2021]. Authors obtained that; their model outperformed the other models by achieving an F-measure of 94.76%. The deep neural network model [Attia, 2018], which combines word and character-based representations in convolutional and recurrent networks with a CRF layer, showed f-score macro of 70.09% on Modern Standard Arabic and Egyptian dialectal Arabic. A neural network model for Turkish named entity recognition is presented in [Güngör, 2018]. The authors developed a model (using a conditional random field (CRF)) that creates a context vector for each position in a sentence, processing words in the forward and backward directions.

Named entity recognition has also an important application of multimodal learning in the field of natural language processing [Zhai, 2023; Shenyi, 2023]. Currently, multimodal approaches are widely used in basic image recognition and natural language processing tasks. For example, paper [Sun, 2023] proposes named entity recognition model, in which lexical features, word boundary features and pinyin features are fused by using a multi-headed attention mechanism for the analysis of government texts in Chinese. In work [Hanming, 2023] authors presented a network architecture based on Transformer for multimodal feature fusion. In the model image and text encoding was performed

separately. Existing multimodal named entity recognition techniques for social media pre-render in review paper [Qian, 2023].

Named entity recognition (NER) of medical text is a basic task in electronic medical text processing for recognition entities such as drug, protein, disease, symptom, detection of missing relationships between biomedical entities such as diseases and chemicals, and determination of drug interactions and side effects in biomedical texts [Zhang, 2020]. The experimental study shows that the most successful method for extracting diseases and symptoms from biomedical texts is BioBERT, with an F1 score of 0.72 [Çelikten, 2023]. A BERT-based model developed to identify biomedical named entities on biomedical Arabic biomedical text dataset reached 85% F1-score [Boudjellal, 2012]. Authors propose a novel neural network architecture for NER that detects word features automatically without feature engineering.

### **Data preprocessing**

The dataset containing 5333 sentences was labeled by NER tags defining it as one of the entities from the set containing the following: CARDINAL, DATE, FACILITY, LOCATION, MONEY, ORGANIZATION, PERSON, QUANTITY, TIME, TITLE, OTHER. The NER tag follows a special format commonly used in the NER literature, called the IOB format:

- O: This tag means that the word is not part of an object
- B: this tag means that the word is either a single word object name or the first word in a multiword object name.
- I: This tag means that the word is part of a multi-word object, but is not the first word in the full object name.

First of all, we convert the data into required format for model training. The whole dataset is pre-separated into sentences and labels. All sentences are divided into training 70% (4265 sentences) and test 30% (1077 sentences). Initially all sentences were embedded into vector space by using Global Vectors (GloVe). GloVe is a distributed word representation model that is obtained by using unsupervised learning algorithm to produce vector representations of words. GloVe vectors were pre-trained on a corpus of Tatar texts of 332 MB.

### **Convolutional and recurrent neural networks-based models**

We used a neural network models based on recurrent (BiLSTM) and convolutional neural networks (CNN). The structure of BiLSTM-CRF

neural network is mainly consists of three layers: input layer of GloVe word vectors, BiLSTM layer and CRF layer. A sequence of word vectors used as inputs. Feature extraction is performed by BiLSTM layer to get the probability of each word in each tag. Finally, a CRF layer is used to restrict the different combinations and get the optimal tagging sequence. BiLSTM can only detect the relationship between the text sequence and the tag, and the relationship between tags can be calculated by using the CRF transition matrix. In BiLSTM word vector matrix, the output result of the hidden layer of BiLSTM is obtained by combining forward LSTM and backward LSTM, and then combining the output word vectors.

For the convolutional neural network (CNN) based model, each word in the input sequence is converted into an N-dimensional vector. Then a convolutional layer is used to create local features around each word, and the output size of convolutional layers depends on the number of words in the sentence. The global feature vector is constructed by combining the local feature vectors extracted using convolution layers. The size of the global feature vector is fixed, independent of the sentence length, for applying subsequent standard layers. Two approaches are widely used to extract global features: a maximum operation or averaging over the position (i.e., “time” step) in the sentence. Finally, these fixed-size global features are passed to a tag decoder to compute distributional estimates for all possible tags for words in the neural network input data.

The recurrent neural network model is implemented as a layer of bidirectional LSTM (Long short-term memory) neurons. The convolutional neural network (CNN) is based on 4 consecutive convolutional neural layers with ReLU activation function, with Dropout layer and batch normalization layer. The output layer of both models is linear, with the number of neurons equal to the number of tags. Both models have added layers for character embedding and Conditional Random Field (CRF) coding.

### **BERT based neural network**

BERT – refers to a type of neural networks called transformers. Transformers do not contain any recurrent mechanisms and convolutions, but they are still the most successful architectures for processing sequences. Transformers have gained popularity in natural language processing and have surpassed convolutional neural networks (CNNs),

long-term memory networks (LSTMs), and recurrent neural networks (RNNs) in general. Transformer architecture has become one of the most popular neural network-based architectures. Transformer are trained on all words in a sequence simultaneously and adds positional coding to each word to prioritize each word in the sequence [Kotei, 2023].

In this project, we use bert-base-multilingual-cased (pre-trained model on 104 major languages with the largest Wikipedia) from the Huggingface (<https://huggingface.co/>) PyTorch library. This model is typically used for either masked language modeling or predicting the next sentence, but it is mainly intended to be fine-tuned for use in a variety of tasks.

Before fine-tuning BERT, we need to prepare the dataset. We have to tokenize all sentences using the BERT tokenizer. It splits the tokens into subword tokens. These tokens are special tokens that are added at the beginning and end of the input text sentence. These tokens are then converted into identifiers by using BertTokenizer. We then added an attention mask for completion and trimmed the sentences to the maximum length. The input identifiers, which include sentence identifiers, attention mask identifier and label identifiers, are then converted into Pytorch tensors.

A model from the Huggingface library called BertForTokenClassification in pytorch is used for training, which is a regular BERT model with the addition of one linear layer, which is used to classify each token entity class, to perform token-level predictions. This model is a fine-tuning model that wraps the BERT model and adds a tag-level classifier on top of the BERT model. The top-level classifier is a linear layer that takes the last hidden state of a sequence as input. When we feed the input data, the pre-trained Bert model and an additional untrained classification layer are trained on our specific task -NER.

## Results

In all experiments, three evaluation metrics are used: precision, recall and F1 score, which are widely used in NER evaluation criteria. The above three evaluation metrics are obtained from the conjugacy matrix of classification results. Precision is the proportion of the number of correctly named objects recognized by the model to the number of all identified named objects. Recall is the proportion of the number of correctly named entities identified by the model to the total number

of named entities in the sample. In real experimental results, there may be conflicts between precision and recall. In this case, the F1 metric is also computed to comprehensively evaluate the quality of model training.

Training process of convolutional neural network-based model during training of neural networks are presented in Figure 1, and in Figure 2 training process for recurrent neural network-based model is presented. An early learning stopping mechanism is used if accuracy does not change seriously within 10 training steps. On the training dataset, F1 accuracy reaches almost 100% rather quickly, but on the test dataset the maximum value is 71%, and a slight overfitting of the models is also observed during further training.

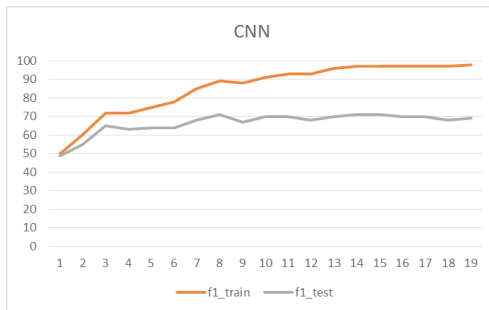


Figure 1. Plots of the F1 metric dynamics during training for a convolutional neural network, on the training and test datasets

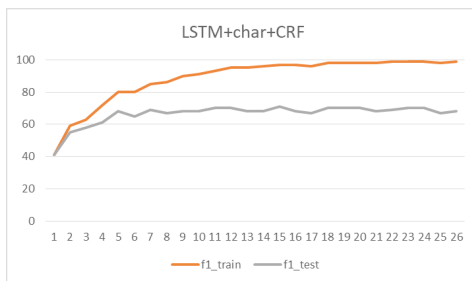


Figure 2. Plots of the F1 metric dynamics during training for the recurrent neural network, on the training and test datasets

Comparison of the learning processes for LSTM and CNN showed that these types of neural networks are trained almost similarly.

**Table 1 shows the maximum values of the F1 metric for all possible model configurations on the test dataset.**

**Table 1. Maximum F1 values on the test dataset**

| Тип модели | char | crf | F1 |
|------------|------|-----|----|
| CNN        | –    | –   | 68 |
| CNN        | +    | –   | 71 |
| CNN        | –    | +   | 65 |
| CNN        | +    | +   | 70 |
| LSTM       | –    | –   | 68 |
| LSTM       | +    | –   | 71 |
| LSTM       | –    | +   | 67 |
| LSTM       | +    | +   | 71 |

**Table 2 and Table 3 show the values of measures characterizing the quality of the trained LSTM model (Table 5) and CNN (Table 6) containing character embedding and CRF based encoding on the test dataset separately for each of the tags. The highest model quality is observed for the tags TIME – 87%, MONEY – 75%, QUANTITY – 76%, CARDINAL-76% and the worst for FACILITY – 17%, OTHER – 36%.**

**Table 2: Values of performance metrics of the LSTM + char\_CRF model on the test dataset for individual tags**

| Tag          | precision | recall | F1    | support |
|--------------|-----------|--------|-------|---------|
| CARDINAL     | 76.04%    | 75.08% | 75.56 | 313     |
| DATE         | 83.39%    | 87.83% | 85.56 | 277     |
| FACILITY     | 21.62%    | 13.79% | 16.84 | 37      |
| LOCATION     | 67.48%    | 64.91% | 66.17 | 329     |
| MONEY        | 82.35%    | 70.00% | 75.68 | 34      |
| ORGANIZATION | 52.60%    | 61.69% | 56.79 | 346     |
| OTHER        | 34.02%    | 37.90% | 35.85 | 244     |
| PERSON       | 67.39%    | 72.09% | 69.66 | 460     |
| QUANTITY     | 80.43%    | 72.55% | 76.29 | 46      |
| TIME         | 87.14%    | 87.14% | 87.14 | 70      |
| TITLE        | 57.46%    | 51.74% | 54.45 | 181     |



**Table 3: Values of performance metrics of the CNN + char\_CRF model on the test dataset for individual tags**

|              | precision | recall | F1    | support |
|--------------|-----------|--------|-------|---------|
| CARDINAL     | 77.74%    | 80.44% | 79.07 | 313     |
| DATE         | 81.94%    | 89.73% | 85.66 | 277     |
| FACILITY     | 24.00%    | 20.69% | 22.22 | 37      |
| LOCATION     | 61.73%    | 66.96% | 64.24 | 329     |
| MONEY        | 78.05%    | 80.00% | 79.01 | 34      |
| ORGANIZATION | 53.69%    | 54.24% | 53.96 | 346     |
| OTHER        | 40.11%    | 32.42% | 35.86 | 244     |
| PERSON       | 73.54%    | 73.02% | 73.28 | 460     |
| QUANTITY     | 81.25%    | 76.47% | 78.79 | 46      |
| TIME         | 88.24%    | 85.71% | 86.96 | 70      |
| TITLE        | 55.67%    | 56.22% | 55.94 | 181     |

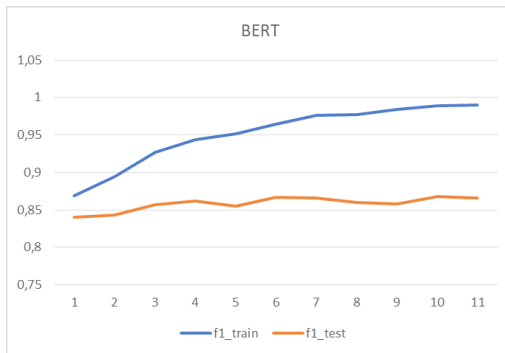


Figure 4. Plots of the F1 metric dynamics during training for BERT based neural network model, on the training and test datasets

Figure 4 shows the graphs of training the neural network based on Bert for the parameter `batch_size=32`. As can be seen from the graphs, after the first training step the model reaches a fairly high accuracy value. The training was carried out for 11 steps, later it was stopped, because the accuracy on the test dataset stopped increasing, and on the training dataset it reached almost 100%. The maximum F1 value on the test set reaches 86.7%.

**Table 4: Values of BERT-based model quality measures on the test dataset for individual tags**

|                | precision | recall | f1-score | support |
|----------------|-----------|--------|----------|---------|
| I-ORGANIZATION | 0.85      | 0.78   | 0.81     | 1061    |
| I-LOCATION     | 0.69      | 0.69   | 0.69     | 233     |
| B-FACILITY     | 0.39      | 0.45   | 0.42     | 75      |
| I-MONEY        | 0.84      | 0.81   | 0.83     | 80      |
| B-CARDINAL     | 0.76      | 0.79   | 0.78     | 443     |
| B-ORGANIZATION | 0.68      | 0.66   | 0.67     | 528     |
| B-MONEY        | 0.67      | 0.67   | 0.67     | 48      |
| I-DATE         | 0.83      | 0.87   | 0.85     | 345     |
| B-PERSON       | 0.69      | 0.76   | 0.72     | 624     |
| I-TITLE        | 0.74      | 0.74   | 0.74     | 304     |
| I-FACILITY     | 0.48      | 0.53   | 0.51     | 136     |
| B-TIME         | 0.73      | 0.63   | 0.68     | 30      |
| I-OTHER        | 0.62      | 0.45   | 0.52     | 506     |
| O              | 0.93      | 0.94   | 0.94     | 13382   |
| B-QUANTITY     | 0.67      | 0.68   | 0.67     | 84      |
| B-OTHER        | 0.56      | 0.42   | 0.48     | 311     |
| B-TITLE        | 0.68      | 0.71   | 0.70     | 302     |
| I-QUANTITY     | 0.78      | 0.66   | 0.72     | 105     |
| I-PERSON       | 0.78      | 0.83   | 0.80     | 499     |
| B-DATE         | 0.83      | 0.85   | 0.84     | 324     |
| I-TIME         | 0.84      | 0.91   | 0.87     | 23      |
| I-CARDINAL     | 0.77      | 0.81   | 0.79     | 42      |
| B-LOCATION     | 0.75      | 0.70   | 0.73     | 564     |
| accuracy       |           |        | 0.87     | 20049   |
| macro avg      | 0.72      | 0.71   | 0.71     | 20049   |
| weighted avg   | 0.86      | 0.87   | 0.86     | 20049   |

Tables 4 summarize the values of the quality measures of the BERT-based model on the test dataset for individual tags. The model achieves the highest prediction accuracy on the tags ORGANIZATION, CARDINAL, TIME.

## Conclusions

In this work we presented the result of using a neural network models based on recurrent (BiLSTM) convolutional neural networks (CNN), and BERT on NER task for Tatar language. The quality of trained models NER was accessed by three evaluation metrics: precision, recall and F1 score. On the training dataset, F1 accuracy reaches almost 100% rather quickly, but on the test dataset the maximum value is 71% (for BiLSTM and CNN), and a slight overfitting of the models is also observed during further training. The highest model quality is observed for the tags TIME- 87%, MONEY- 75%, QUANTITY- 76%, CARDINAL-76% and the worst for FACILITY-17%, OTHER-36%. For BERT based model the accuracy on the training dataset it reached almost 100%. The maximum F1 value on the test set reaches 86.7%. This model achieves the highest prediction accuracy on the tags ORGANIZATION, CARDINAL, TIME. This study was conducted on a fairly small dataset (only 5333 labeled sentences in total), and further increases in the dataset size may lead to higher accuracy values.

## REFERENCES

1. Ajees A, Sumam I. (2018). A Named Entity Recognition System for Malayalam using Neural Networks. *Procedia Computer Science*. 143. 962–969. 10.1016/j.procs.2018.10.338.
2. Alsaaran N., Alrabiah M. (2021) “Classical Arabic Named Entity Recognition Using Variant Deep Neural Network Architectures and BERT,” in *IEEE Access*, vol. 9, pp. 91537-91547, 2021, doi: 10.1109/ACCESS.2021.3092261.
3. Attia M., Samih Y., Maier W. (2018). GHHT at CALCS 2018: Named Entity Recognition for Dialectal Arabic Using Neural Networks. In *Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching*, pages 98–102, Melbourne, Australia. Association for Computational Linguistics.
4. Boudjellal, N., Zhang, H., Khan, A., Ahmad, A., Naseem, R., Shang, J., Dai, L. (2021). ABioNER: A BERT-Based Model for Arabic Biomedical Named-Entity Recognition. *Complexity*. 2021. 1–6. 10.1155/2021/6633213.
5. Çelikten, A., Onan, A., Bulut, H. (2022). Investigation of Biomedical Named Entity Recognition Methods. In *The International Conference on Ar-*

tificial Intelligence and Applied Mathematics in Engineering (pp. 218-229). Cham: Springer International Publishing.

6. Güngör O., Üsküdarlı S., Güngör T., (2018) “Recurrent neural networks for Turkish named entity recognition,” 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404788.

7. Hanming Z., Xiaojun L. Zhiwen H., Xin T., Fanliang B. (2023) ML-Net: a multi-level multimodal named entity recognition architecture. *Frontiers in Neurorobotics*. 17. 10.3389/fnbot.2023.1181143.

8. Khakimov B, Gafarova V, Размеченный корпус именованных существностей на татарском языке, 2022 <https://publications.hse.ru/pubs/share/direct/859123164.pdf#page=90>

9. Kotei E, Thirunavukarasu R. A Systematic Review of Transformer-Based Pre-Trained Language Models through Self-Supervised Learning. *Information*. 2023; 14(3):187. <https://doi.org/10.3390/info14030187>

10. Li J., Sun A., Han J. and Li C. (2022) A Survey on Deep Learning for Named Entity Recognition, in *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50–70, doi: 10.1109/TKDE.2020.2981314.

11. Litake, O., Sabane, M., Patil, P., Ranade, A., Joshi, R. (2023). Mono Versus Multilingual BERT: A Case Study in Hindi and Marathi Named Entity Recognition. In: Gunjan, V.K., Zurada, J.M. (eds) *Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications. Lecture Notes in Networks and Systems*, vol 540. Springer, Singapore. [https://doi.org/10.1007/978-981-19-6088-8\\_56](https://doi.org/10.1007/978-981-19-6088-8_56)

12. Mohammed, N. F., Omar, N. (2012). Arabic Named Entity Recognition Using Artificial Neural Network. *Journal of Computer Science*, 8(8), 1285-1293. <https://doi.org/10.3844/jcssp.2012.1285.1293>

13. Nevzorova O., Mukhamedshin D., Galieva A. Named Entity Recognition in Tatar: Corpus Based Algorithm // *Proceedings of Computational Models in Language and Speech Workshop (CMLS 2018)*. CEUR, vol. 2303. – P. 58–68.

14. Qian S., Jin W., Chen Y., Ma J., Qiao Y., Lu J. (2023) A Survey on Multimodal Named Entity Recognition. In *Advanced Intelligent Computing Technology and Applications: 19th International Conference, ICIC 2023, Zhengzhou, China, August 10–13, 2023, Proceedings, Part IV*. Springer-Verlag, Berlin, Heidelberg, 609–622. [https://doi.org/10.1007/978-981-99-4752-2\\_50](https://doi.org/10.1007/978-981-99-4752-2_50).

15. Raza S, Reji DJ, Shajan F, Bashir SR (2022) Large-scale application of named entity recognition to biomedicine and epidemiology. *PLOS Digital Health* 1(12): e0000152. <https://doi.org/10.1371/journal.pdig.0000152>

16. Repke, T., Krestel, R. (2021). Extraction and Representation of Financial Entities from Text. In: Consoli, S., Reforgiato Recupero, D.,

Saisana, M. (eds) *Data Science for Economics and Finance*. Springer, Cham. [https://doi.org/10.1007/978-3-030-66891-4\\_11](https://doi.org/10.1007/978-3-030-66891-4_11).

17. Richa S., Sudha M., Basant A, Ramesh C., Shoeb K. (2020) A deep neural network-based model for named entity recognition for Hindi language. *Neural Computing and Applications*. 32. 10.1007/s00521-020-04881-z.

18. Sheping Z. Dan G., Huizhen W., Yun C. (2022) Chinese Named Entity Recognition Based on BERT and Neural Network. 10.1007/978-3-030-89698-0\_137.

19. Sun Z., Sun R., Liang Z., Su Z., Yu Y., Wu S. (2023) Chinese Named Entity Recognition Based on Multi-feature Fusion. In *Advanced Intelligent Computing Technology and Applications: 19th International Conference, ICIC 2023, Zhengzhou, China, August 10–13, 2023, Proceedings, Part IV*. Springer-Verlag, Berlin, Heidelberg, 670–681. [https://doi.org/10.1007/978-981-99-4752-2\\_55](https://doi.org/10.1007/978-981-99-4752-2_55).

20. Wang, H.; Zhou, L.; Duan, J.; He, L. (2023) Cross-Lingual Named Entity Recognition Based on Attention and Adversarial Training. *Appl. Sci.*, 13, 2548. <https://doi.org/10.3390/app13042548>.

21. Zhang, R, Gao, Y., Yu R., Wang, R., Lu, W. (2020). Medical Named Entity Recognition Based on Overlapping Neural Networks. *Procedia Computer Science*. 174. 27-31. 10.1016/j.procs.2020.06.052.

22. Zali, M. & Firoozbakht M. (2018). Named entities recognition and classification system for Persian texts based on neural network. *Iranian Journal of Information Processing Management*. 34. 473–486.

УДК 004.021

**ПОСТРОЕНИЕ АЛГОРИТМА ПОЛУЧЕНИЯ  
СИНОНИМА ИЗ МЕДИЦИНСКИХ ТЕКСТОВ  
ДЛЯ КАЗАХСКОГО ЯЗЫКА*****Д. Р. Рахимова, А. С. Карибаяева, Е. Р. Сулейменов****Институт информационных и вычислительных технологий,  
Алматы, Казахстан*di.diva@mail.ru, a.s.karibayeva@gmail.com,  
erken.suleimenov@gmail.com

В статье представлен алгоритм, направленный на разработку инструмента, способного автоматически идентифицировать и извлекать синонимы из медицинских текстов. Улучшение понимания и использования синонимов может улучшить обмен информацией в здравоохранении. С точки зрения технической реализации статья включает в себя использование подходов обработки естественного языка (NLP) и семантического анализа. Библиотеки NLTK и WordNet используются для обработки текста и создания синонимов на одном языке. Векторные модели слов, такие как Word2Vec, GloVe или BERT, используются для создания векторных представлений слов, которые затем используются для определения семантического сходства между словами. Слова с высокой семантической близостью рассматриваются как синонимы.

**Ключевые слова:** казахский язык, малоресурсный язык, синонимы, медицинские тексты, алгоритм.

**A SYNONYM OBTAINING ALGORITHM CONSTRUCTION FROM  
MEDICAL TEXTS FOR THE KAZAKH LANGUAGE*****Diana Rakhimova, Aidana Karibayeva, Yerkin Suleimenov****Institute of Information and Computing Technologies,  
Almaty, Kazakhstan*di.diva@mail.ru, a.s.karibayeva@gmail.com,  
erken.suleimenov@gmail.com

The article presents an algorithm aimed at developing a tool capable of automatically identifying and extracting synonyms from medical texts. Improving the understanding and use of synonyms can improve communication in healthcare. In terms of technical implementation, the paper includes the use of natural language processing (NLP) and semantic analysis approaches. NLTK and WordNet libraries are used to process text and create synonyms in a single language. Word vector models such as Word2Vec, GloVe or BERT are used to create vector representations of words, which are then used to determine semantic similarity between words. Words with high semantic similarity are considered synonyms.

**Key words:** Kazakh language, low-resource language, synonyms, medical texts, algorithm.

## 1. Введение

В медицине существует огромное количество терминов и синонимов, и они постоянно обновляются и уточняются. Потребность упрощении в текстах возрастает с увеличением объема информации, а государственные органы нуждаются в эффективных способах упрощения текстов для увеличения охвата цифровыми технологиями. Это, в свою очередь, мотивирует необходимость автоматизации процесса упрощения текстов. Одна часть проблемы в этом процессе – разработать метод замены трудных для понимания слов более простыми, то есть синонимами. Имея систему, которая автоматически извлекает синонимы из медицинских текстов, можно упростить поиск и анализ медицинской информации.

Разные страны и культуры могут использовать разные термины для обозначения одного и того же медицинского состояния или процедуры. Система, которая автоматически определяет синонимы, может помочь в международном обмене медицинскими знаниями и улучшении качества здравоохранения. В современном мире информации становится все больше, и поиск нужной информации становится все более сложной задачей.

Автоматическое извлечение синонимов может упростить и ускорить процесс поиска информации. Таким образом, эта тема является актуальной и имеет большой потенциал для дальнейшего развития и применения в медицине и других областях.

## 2. Обзор литературы

Словарь синонимов был использован для поиска синонимов-заменителей в текстах четырех разных жанров. Одна из вещей, которую авторы обнаружили, заключается в том, что замена синонимов не так проста, как кажется. Выбор слов часто зависит от контекста, поэтому просто слова замена не всегда может помочь [Jurafsky, 2017].

Основываясь на методах замены слов и системной оценки, авторы работали над заменой синонимов в медицинском контексте [Abrahamsson, 2014].

Двумя обычно используемыми мерами для оценки читабельности являются индекс читабельности, и индекс вариаций слов.

Значение индекса читабельности учитывает количество слов и предложений, а также количество длинных слов. Чем длиннее предложение и чем длиннее слова в тексте, тем менее читаемым оно является. Индекс вариаций слов, с другой стороны, использует количество уникальных слов в тексте для измерения лексической вариативности. Чтение текстов с уникальными словами считается более сложным. Результаты показывают, что замена синонимов в медицинском тексте снижает читаемость по индекс читабельности и повышает читаемость по индекс вариаций слов.

Более сложные греческие и латинские слова были заменены более длинными, но простыми словами, что привело к более высокому значению индекса читабельности. В трети предложений, где были изменены слова, изменился смысл предложений. Это потенциальная проблема в текущей работе, и ее следует учитывать при анализе результатов. Оба исследования, упомянутые выше, использовали показатели читабельности для оценки своих систем. Однако настоящая работа направлена только на выявление синонимов, поэтому такие оценки невозможны.

Различные комбинации случайной индексации и случайной перестановки могут использоваться для идентификации синонимов в медицинских текстах [Benoit, 2013]. Случайное индексирование – это метод, который генерирует матрицу совпадений путем присвоения каждому слову вектора распределения, состоящего из заранее определенного количества нулей размерности. Кроме того, каждое слово получает индексный вектор, состоящий из случайно распределенных ненулевых элементов ( $-1$  и  $1$ ).

Если слово находится рядом с другим словом, его вектор индекса добавляется к вектору распределения соседа. Случайная перестановка очень похожа на случайное индексирование, но также учитывает порядок появления слов. Помимо поиска синонимов, они рассмотрели пары сжатия-расширения в обоих направлениях. Было использовано пять различных экспериментальных установок: случайная индексация, случайная перестановка, случайная индексация с фильтрацией случайной перестановки, случайная перестановка с фильтрацией случайной перестановки и случайная индексация со случайной перестановкой.

Через два года после вышеупомянутой статьи те же авторы опубликовали доклад на ту же тему. И снова комбинация случайной индексации и случайной перестановки оказалась весьма эффективной при выполнении задачи. Модель случайного ин-



дексирования использовала 1000 измерений, а векторы индексов состояли из восьми ненулевых элементов. В этом исследовании они были ограничены включением только слов, которые встречались более 50 раз. Это ограничение возникает из-за того, что чем меньше слово встречается в корпусе, тем сложнее смоделировать его значение. Результаты также показали, что использование различных кейсов помогло улучшить производительность. Были использованы два разных корпуса: один состоял из шведских медицинских записей, а другой – из выпусков шведских медицинских журналов. Причина, по которой комбинация случайной индексации и случайной перестановки хороша, заключается в том, что они различаются способом моделирования семантических отношений [Baroni, 2014].

Авторы создали шведский словарь синонимов [Benoit, 2018]. С помощью краудсорсинга. Собрав список возможных синонимов, они позволяют людям определить, являются ли два слова синонимами. Система реализована в онлайн-сервисе Lexin on-line, представляющем собой шведский лексикон.

Не было необходимости определять смысловые отношения, создающие синонимию. Скорее, важно было восприятие синонима людьми. Всякий раз, когда человек ищет что-то на веб-сайте Lexin, отображаются два слова, которые просят пользователя оценить синоним по шкале от нуля до пяти, где ноль означает отсутствие синонима, а пять – полный синоним. Также был вариант «не знаю». Когда достаточное количество людей оценит синоним, система может принять решение удалить или сохранить синоним на основе средней оценки. В этом случае краудсорсинг оказался полезным инструментом. То есть за пять месяцев система получила более двух миллионов оценок. Один из уроков, извлеченных из этого проекта, заключается в том, что представление плохих (далеко не синонимичных) пар слов может раздражать пользователей [Snomed CT., 2018]. Эту проблему качества следует учитывать при составлении списка возможных синонимов.

Некоторые методы, используемые для извлечения семантически связанных слов, основаны на гипотезе распределения. То есть слова, которые появляются в схожих контекстах, часто имеют схожие значения. Эта группа моделей известна как распределительные семантические модели. В то время как первоначальные распределительные семантические модели создавали векторы на основе значений, полученных из частот событий, позже был раз-

работан другой тип распределительных семантических моделей, решающий векторную оценку как контролируруемую задачу с целью предсказать термин с учетом контекста или контекст с учетом термина. Одной из групп таких распределительных семантических модели является word2vec с двумя подходами CBOW и Skip-gram [Mikolov, 2017]. Как и другие распределительные семантические модели, модели word2vec не идеальны, когда дело доходит до различения сходств распределения

Прежде чем создавать систему автоматического обнаружения синонимов, возможно, будет полезно узнать, что характерно для медицинского языка и чем он отличается от обычного языка. Этот аспект был изучен в шведском медицинском тексте [Henriksson, 2014]. Их результаты показывают, что сокращения и технические термины чаще встречаются в медицинских текстах, чем в общем языке. Их результаты показывают, что существуют различия между текстами в разных частях системы здравоохранения.

В нескольких исследованиях были представлены методы, которые можно использовать для извлечения синонимов или выбора синонима к слову с учетом нескольких вариантов. В работе по лексическому упрощению шведских медицинских текстов представлены методы замены синонимов [Alam, 2020].

Авторы с помощью словаря синонимов казахского языка заменили неизвестное слово словом, близким по смыслу в задаче неизвестных слов в нейронном машинном переводе. Словарь синонимов использовались для поиска слов, близких по значению к неизвестным словам, которые были определены. Кроме того, найденные синонимы проверяются на наличие в словаре обученной модели. После этого выполнялся перевод отредактированного предложения исходного языка. Собрана база слов-синонимов из обычных текстов казахского языка [Turganbayeva, 2020].

### 3. Алгоритм создания словаря синонимов

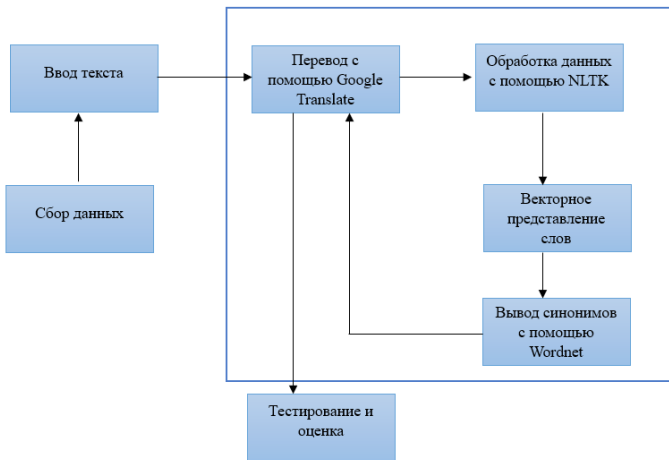
Проанализировав работы Алгоритм создания словаря синонимов состоит из следующих шагов (Рис.1):

*Сбор данных.* На этом этапе будет собран большой корпус медицинских текстов, которые будут использоваться в качестве исходных данных для проекта. Медицинские статьи, исследовательские работы, клинические отчеты и т. д. б. может быть использованы.

*Манипулирование данными.* Следующим шагом является использование Natural Language Toolkit (NLTK) для предварительной обработки текстовых данных. Это может включать токенизацию (разбиение текста на отдельные слова или фразы), лемматизацию (сокращение слова до его корня) и исключение стоп-слов (общие слова, которые обычно удаляются из процессов NLP).

*Создание векторного представления слов.* На этом этапе модели обработки естественного языка, такие как Word2Vec или BERT, используются для создания векторных представлений слов. Эти векторы представляют слова в многомерном пространстве, где слова со схожим значением расположены близко друг к другу.

*Извлечение синонимов с помощью WordNet.* WordNet, крупная английская семантическая сеть слов, затем используется для поиска синонимов. Семантические векторные представления слов также используются для определения семантической близости и возможных синонимов.



**Рис. 1.** Схема работы разработанного алгоритма

*Использование Google Translate для поиска межъязыковых синонимов.* Между тем API Google Translate используется для перевода медицинских текстов на разные языки. Затем процесс синонимизации этих переведенных текстов повторяется для получения межъязыковых синонимов.

*Оценка и проверка.* После получения синонимов они оцениваются и проверяются на точность. Это может включать сравнение с существующими базами данных синонимов или оценку экспертами в области медицинской терминологии.

*Усовершенствование.* На основании отзывов методы получения синонимов будут усовершенствованы и доработаны.

*Развертывание и мониторинг.* После разработки и тестирования система развертывается.

Создание системы автоматического извлечения синонимов из медицинских текстов – важная задача, значительно упрощающая работу медицинских работников и исследователей. Этот алгоритм, основанный на методах обработки естественного языка (NLP), таких как Nltk, WordNet и Google Translate, предлагает системный подход к этой задаче. Алгоритм начинается со сбора данных и их обработки с помощью NLTK, а затем использует методы представления векторов слов для создания семантических моделей слов. В WordNet алгоритм извлекает синонимы, а Google Translate позволяет распространить этот процесс на несколько языков. После этого полученные синонимы оцениваются и проверяются, процесс повторяется для улучшения результатов и, наконец, система развертывается и контролируется. В целом этот алгоритм является многообещающим способом автоматизации извлечения синонимов из медицинских текстов. В будущем он может быть усовершенствован и адаптирован для работы с различными медицинскими текстами и базами данных.

В таблице приведены примеры слов синонимов, полученные разработанным алгоритмом (Таблица 1):

**Таблица 1.** Примеры слов синонимов, полученные разработанным алгоритмом

| Слово        | Синоним                         |
|--------------|---------------------------------|
| гипертония   | қан қысымы, жоғары қан қысымы   |
| фарингит     | тамақ ауруы, жұтқыншақ қабынуы  |
| астма        | бронх демікпесі, астма ұстамасы |
| жара         | ойық, тесік                     |
| қатерлі ісік | рак, онкология                  |

## Заключение

Одной из основных задач работы с медицинскими данными является создание медицинских словарей для облегчения понимания и использования медицинских терминов в научной работе, исследованиях, практике и преподавании должен быть словарь, содержащий термины, определения, синонимы и связи между ними. В статье показан алгоритм создания словаря синонимов.

Основной задачей исследования является разработка модели на основе современных информационных технологий, учитывающей особенности медицинской терминологии, а также алгоритма создания медицинского словаря на основе этой модели. В статье представлены результаты экспериментов, проведенных для проверки эффективности и точности предложенного алгоритма.

В ходе работы была успешно создана система автоматического извлечения синонимов из медицинских текстов. Эта система обрабатывает медицинские тексты с использованием методов обработки естественного языка (NLP) и идентифицирует синонимы с помощью WordNet.

Алгоритм успешно извлекает синонимы из медицинских текстов и помогает улучшить понимание и анализ этих текстов. Использование библиотек NLTK и WordNet показало, что эти инструменты могут быть очень полезны для обработки естественного языка и извлечения синонимов. Интеграция Google Translate позволила системе работать с многоязычными текстами и переводить синонимы на разные языки.

Эта система вносит значительный вклад в область медицинской информатики, помогая людям лучше понимать и анализировать медицинские тексты.

### Благодарности

Данное исследование выполнено и профинансировано грантовым проектом ИРН АР 09259556 «Разработка методов и систем комплексного обучения и обработки естественного языка на основе технологий искусственного интеллекта». Министерства науки и высшего образования Республики Казахстан.

## ЛИТЕРАТУРА

1. Abrahamsson, E., Forni, T., Skeppstedt, M., Kvist, Medical text simplification using synonym replacement: Adapting assessment of word difficulty to a compounding language. Proceedings of The 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations; 2014, pp 57–65.

2. Alam, F., Afzal, M., Malik, K. M. Comparative Analysis of Semantic Similarity Techniques for Medical Text. International Conference on Information Networking; 2020, pp. 106–109.
3. Baroni M, Dinu G, Kruszewski G. Don't count, predict! a systematic comparison of contextcounting vs. context-predicting semantic vectors. In ACL (1); 2014, pp 238–247.
4. Benoit K., Nulty P., Oben A., Wang H., Müller S., Lauderdale B., Lowe W. Package 'quanteda', 2018.
5. Henriksson A., Moen H., Skeppstedt M., Daudaravičius V., Duneld, M. Synonym extraction and abbreviation expansion with ensembles of semantic spaces. Journal of Biomedical Semantics; 5(1), 2014, pp 1–25.
6. Jurafsky, D., Martin, J. H. Speech and Language Processing. Cambridge: Pearson, 2017, 1031 p.
7. Mikolov T, Chen K, Corrado G., Dean J. Efficient estimation of word representations in vector space. CoRR; abs/1301.3781, 2013, pp. 1–12.
8. Socialstyrelsen. (n.d.). Snomed CT. 2018, <https://www.socialstyrelsen.se/nationellehalsa/snomed-ct>.
9. Turganbayeva A., Tukeyev U. The solution of the problem of unknown words under neural machine translation of the Kazakh language. Journal of Information and Telecommunication; 5. 1–12, 2020, pp. 1–12.

---

## ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

УДК 372.881.1

### ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ ДЛЯ ОБУЧЕНИЯ АЛТАЙСКОМУ ЯЗЫКУ

**Ч. П. Сабина**

*Издательский Дом «Алтай Бичик»  
Горно-Алтайск, Алтай, Россия  
sabcheynesh@gmail.com*

Образовательная среда становится ведущей сферой общения на алтайском языке, так как родной язык перестаёт быть средством внутрисемейного общения. Сегодня сохранить родной язык могут помочь современные технологии. Мы живем в мировом информационном изобилии. Дети погружены в мир информации, виртуальное пространство все больше становится для них удобным и интересным. Поэтому разработка и внедрение современных методик и информационных технологий преподавания алтайского языка в общеобразовательных организациях Республики Алтай является одним из важнейших направлений научно – методической работы, которая обеспечит освоение родного языка в современной социокультурной ситуации.

**Ключевые слова.** Алтайский язык, образовательная среда, методическая работа, государственный язык, государственная программа, искусственный интеллект, мультфильмы на алтайском языке, чат-бот на алтайском языке, мобильное приложение, образовательный ресурс, аудиосказки на алтайском языке.

### INTELLIGENT SYSTEMS AND TECHNOLOGIES FOR TEACHING THE ALTAI LANGUAGE

**Cheynesh Sabina**

*Publishing House “Altai Bichik”  
Gorno-Altai, Altai, Russia  
sabcheynesh@gmail.com*

The educational environment is becoming the leading sphere of communication in the Altai language, as the native language ceases to be a means of intra-family communication. Today, modern technologies can help to preserve the native language. We live in a world of information abundance. Children are immersed in the world of information, virtual space is becoming more and more convenient and interesting for them. Therefore, the development and implementation of modern methods and information technologies for teaching the Altai language in

educational institutions of the Altai Republic is one of the most important areas of scientific and methodological work that will ensure the development of the native language in the modern socio-cultural situation.

**Keywords.** Altai language, educational environment, methodical work, state language, state program, artificial intelligence, cartoons in the Altai language, chatbot in the Altai language, mobile application, educational resource, audio stories in the Altai language.

В Республике Алтай функционируют два государственных языка: русский и алтайский. Ведётся работа по поддержке и продвижению государственных языков республики, по сохранению национальной идентичности, культурного разнообразия как неперменного и обязательного условия устойчивого развития общества, мирного и уважительного сосуществования народов.

Государственным языком Российской Федерации на всей её территории является русский язык. Русский язык является одной из важных основ российской государственности и способствует взаимопониманию, укреплению межнациональных связей народов Российской Федерации в едином многонациональном государстве. Обеспечение развития русского языка на территории Республики Алтай отражено в государственной программе Республики Алтай «Реализация государственной национальной политики».

Государственный алтайский язык функционирует в сфере культуры, науки, образования, средств массовой информации. На двух государственных языках в Республике Алтай проводятся научно-практические конференции, семинары, транслируются театральные постановки, проводятся конкурсы, лектории. На государственном алтайском языке осуществляются исследования в области языкознания, литературоведения, фольклора, культуры, истории и национальной культуры, издаются научные монографии, различного рода словари, художественная, учебная, методическая литература; ведётся работа по подготовке цифровых ресурсов. Однако современное пространство требует усиления роли родного языка в цифровой и виртуальной среде.

Образовательная среда становится ведущей сферой общения на алтайском языке, так как родной язык перестаёт быть средством внутрисемейного общения. Сегодня сохранить родной язык могут помочь современные технологии. Мы живем в мировом информационном изобилии. Дети погружены в мир информации, виртуальное пространство все больше становится для них



удобным и интересным. Поэтому разработка и внедрение современных методик и информационных технологий преподавания алтайского языка в общеобразовательных организациях Республики Алтай является одним из важнейших направлений научно – методической работы, которая обеспечит освоение родного языка в современной социокультурной ситуации.

В связи с этим Правительство Республики Алтай утвердило государственную программу Республики Алтай «Сохранение и развитие алтайского языка», которое вступило в силу с 1 января 2022–2027 годов.

#### Развитие алтайского языка в сфере образования

В вопросах сохранения и развития алтайского языка особая роль принадлежит системе образования. Современная система преподавания алтайского языка и литературы в Республике Алтай включает все уровни общего образования (дошкольное, начальное, основное, среднее), а также среднее профессиональное образование и высшее образование. На региональном уровне создана нормативная правовая база, обеспечивающая условия преподавания алтайского языка. В 50% или в 83 дошкольных образовательных организациях региона организованы занятия и кружки по алтайскому языку. Доля детей, изучающих алтайский язык в дошкольном образовании в 2021–2022 учебном году, составляет 26% или 3584 ребёнка. Алтайский язык преподаётся в 72% или 129 школах региона. Доля обучающихся в общеобразовательных организациях, изучающих алтайский язык в 2021–2022 году, составляет 30,5% или 11 961 человек, от общего числа обучающихся в общеобразовательных организациях. Одним из важных направлений современного образования является использование цифровых ресурсов. В настоящее время начальное, основное и среднее общее образование обеспечено 19 электронными формами учебников по алтайскому языку и алтайской литературе; электронные образовательные ресурсы составлены для учебников начального общего образования и для 5–6 классов.

Отдельная работа проводится по пропаганде и популяризации алтайского языка, повышению мотивации у детей к его изучению посредством проведения олимпиады школьников по алтайскому языку, открытых лекториев по истории, языку и культуре алтайского народа, лингвистических онлайн-смен по углублённому изучению алтайского языка. Подтвердили свою эффективность такие конкурсы регионального уровня, как интеллектуальный кон-

курс «Тийинеш» («Белочка») по алтайскому языку и литературе для обучающихся 5-10 классов, республиканский конкурс чтецов «Тирү классика алтай тилле» («Живая классика на алтайском языке») для обучающихся и студентов, республиканский слет юных поэтов и писателей «Амаду» («Мечта»).

Начата разработка и внедрение интерактивной образовательной продукции и мультимедийного контента обучения алтайскому языку. Сегодня успешно функционируют мобильные приложения «Говорящий алфавит» пока только в формате арк файла для системы Android, «Алтайские сказки детям» приложение аудиосказок на алтайском языке с тестовыми заданиями и текстом на двух языках на платформе Android, электронный журнал для детей «Косуленок», электронный образовательный ресурс «Изучаем алтайский язык» где есть видеоуроки как выучить алтайский язык, словари с аудио и аудиорассказы на родном алтайском языке и в него встроены чат-бот переводчик «Ырысту» «Счастливчик».

В 2021 пробовали создавать голосовой переводчик AiA, который распознавал слово на русском языке находил перевод слова в введенной базе данных, озвучивал это слово голосом робота и находил картинку, основа Искусственный интеллект.

Научная новизна предлагаемых в проекте решений: Голосовая платформа должна представлять собой программное обеспечение, в котором будет осуществляться мгновенный голосовой перевод слов и языковой пары «русский - алтайский» с возможностью понимания и обработки естественного языка NLU. Должна произойти интеграция алтайского языка и его правил в платформу для грамотного и правильного перевода.

В ходе тестовых проб, мы решили приостановить проект, в связи с тем, что проект не имел базы данных.

Дубляж мультсериала Тиг и Лео на алтайский язык.

Созданы 2 д мультфильмы на алтайском языке «Ырысту» и «Түлкү ле Турна».

Активно введутся работы по работе перевода статей в Википедии.

Есть раскладка клавиатур на Android, iOS, Windows имеется необходимость принятия мер:

1) обновление учебно-методической базы с учетом современных технологий образования;

2) издание и выпуск современных аудиопособий, мультимедийных пособий, теле-и радиопередач, детских книг, журналов, учебной литературы; 3) материально-техническое оснащение кабинетов алтайского языка и литературы общеобразовательных организаций Республики Алтай;

4) обновление методики и технологии обучения родному языку.

#### ЛИТЕРАТУРА

1. <http://publication.pravo.gov.ru/Document/View/0400202111300002>
2. [https://nbra.ru/index.php?option=com\\_attachments&task=download&id=322](https://nbra.ru/index.php?option=com_attachments&task=download&id=322).

УДК 372.881.1

**МОБИЛЬНОЕ ПРИЛОЖЕНИЕ «ТЫВАЛАП ЧУГААЛАЖЫЫЛ»  
‘ПОГОВОРИМ ПО-ТУВИНСКИ’*****С. А. Мылдыргыновна****Тувинский государственный университет,  
Кызыл, Республика Тыва, Россия  
soyan-a@mail.ru*

В данной работе рассматриваются функции мобильного приложения «Тывалап чугаалажыыл» ‘Поговорим по- тувински’ и его роль при изучении тувинского языка. В приложении, созданном на платформе Android, содержатся тувинский алфавит со звуковым сопровождением, имена числительные, тематический словарь на тувинском и русском языках, стихотворения, детские песни, тувинские игры, мультфильмы.

Мобильное приложение даёт возможность изучить тувинский язык вне зависимости от места, времени и направлено на сохранение и развитие тувинского языка.

**Ключевые слова:** мобильное приложение, тувинский язык, лексика, алфавит, тематический словарь.

**THE MOBILE APPLICATION «TYVALAP CHUGAALAZHYYL»  
‘LET’S TALK IN TUVAN’*****Aylanmaa Soyana****Tuvan State University,  
Kyzyl, Republic of Tuva, Russia  
soyan-a@mail.ru*

**Abstract.** This paper discusses the functions of the mobile application «Tyvalap Chugaalazhyyl» ‘Let’s Talk in Tuvan’ and its role in learning the Tuvan language. The application, created on the Android platform, contains the Tuvan alphabet with sound accompaniment, numeral names, thematic vocabulary in Tuvan and Russian, poems, children’s songs, Tuvan games, and cartoons.

The mobile application gives an opportunity to learn the Tuvan language regardless of place, time and is aimed at preserving and developing the Tuvan language.

**Keywords:** mobile application, Tuvan language, vocabulary alphabet, thematic dictionary.

Мобильное приложение «Тывалап чугаалажыыл» «Поговорим по-тувински» создано в результате коллаборации студентов филологического (Донгак Ангыр-Чечек, Элбек Олча) и физико-математического (Монгуш Ям-Суюн) факультетов Тувинского

государственного университета под руководством кандидатов филологических наук, доцентов Елены Куулар, Айланмаа Соян и старшего преподавателя Чойганы Ооржак. Оно предназначено для развития устной связной речи и для изучения тувинского языка детьми дошкольного возраста.

Приложение на платформе Android представлено на двух языках – тувинском и русском – и содержит тувинский алфавит со звуковым сопровождением, имена числительные, тематический словарь, стихотворения, детские песни, тувинские игры, мультфильмы.

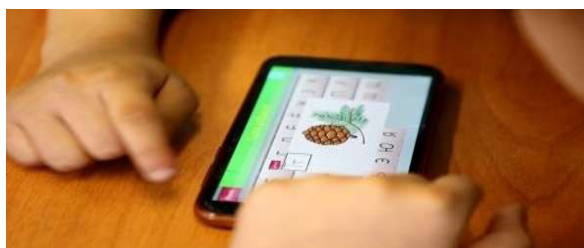


Рис. 1. Тувинский алфавит Fig.

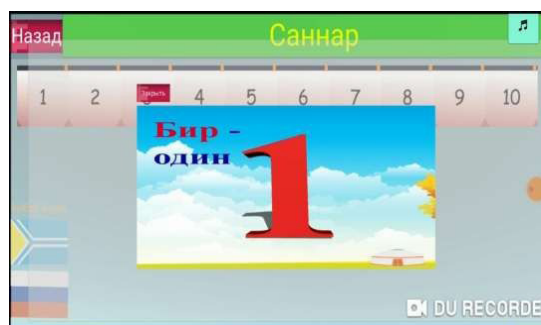


Рис. 2. Имена числительные

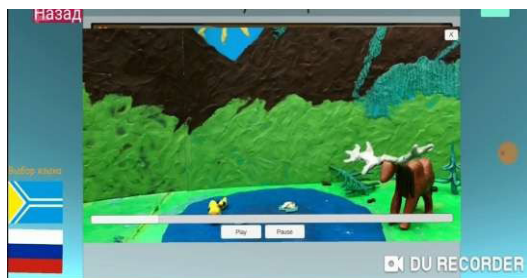


Рис. 3. Мультфильмы

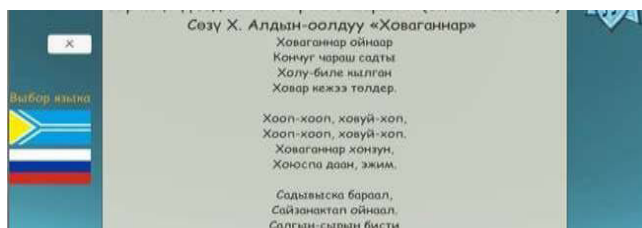


Рис. 4. Песни

В тематическом словаре представлены названия тувинских блюд (*быштак* ‘сыр’, *пиң* ‘лепёшка с топленным маслом’), ягод (*чыжыргана* ‘облепиха’, *честек-кат* ‘земляника’), наименования родственных отношений (*авай* ‘мама’, *ачай* ‘папа’, *кырган-авай* ‘бабушка’), домашних и диких животных (*анай* ‘козленок’, *адыг* ‘медведь’), птиц (*торга* ‘дятел’), слова, обозначающие времена года и природные явления (*кыш* ‘зима’, *шуурган* ‘ураган’), базовая лексика тувинского языка.



Рис. 5. Тематический словарь

В настоящее время в Республике Тыва нарастает тенденция незнания детьми дошкольного и младшего школьного возраста своего родного (тувинского) языка. Поэтому создание данного приложения очень актуально. Известно, что незнание родного языка приводит к утрате культуры и исчезновению народа.

Приложение направлено на расширение активного словарного запаса детей, не владеющих и плохо владеющим родным тувинским языком, а также позволяет детям быстро и весело научиться понимать значение слов базовой лексики тувинского языка и правильно их произносить. Ведь родной язык – это начало всех начал.

Роль родного языка в развитии ребенка раннего огромна и неоспорима. В раннем возрасте ребенок должен знать базовую лексику на родном языке: числительные, основные глаголы, наименования родственных отношений, названия диких и домашних животных, названия частей тела, слова-цветообозначения и т.д.

Потенциальными потребителями данного мобильного приложения являются дошкольные учреждения Республики Тыва, а также желающие попрактиковаться в общении на тувинском языке.

Для того, чтобы выявить роль специализированных программных обеспечений (далее – ПО) для повышения эффективности изучения других языков, в марте 2023 г. были проведены статистические исследования среди студентов филологического факультета ФГБОУ ВО «Тувинский государственный университет». Результаты показали, что более 70% студентов-языковедов в процессе обучения используют ПО. При этом более 80% респондентов из числа пользователей ПО отметили, что использование программного обеспечения позволяет им повысить эффективность освоения иностранного языка.

Мобильное приложение подобного рода можно использовать при изучении других тюркских языков. При составлении концептуальных основ предлагаемой системы предполагается утилизировать высокий уровень схожести языков тюркской языковой семьи, что позволит качественно повысить удобство и результативность эксплуатации программного обеспечения со стороны специального пользователя. По нашему мнению, разработка специализированного ПО, учитывающего особенности языков тюркской языковой группы, будет способствовать популяризации изучения тюркских языков, а также позволит облегчить про-

цесс их изучения, в особенности для действующих носителей тюркских языков. В частности, одним из наиболее перспективных методов изучения языков тюркской группы с использованием специализированного программного обеспечения представляется сопоставление языковых массивов для выведения идентичных и схожих лексических единиц. Например, произношение и написание слова солнце очень схоже во многих тюркских языках: тув. *хун*, алт. *кун*, хак. *кун*, каз. *кун*.

При этом стоит учитывать использование языками тюркской группы различных алфавитов. В связи с этим одним из наиболее рациональных решений представляется создание нескольких массивов для отдельного языка, содержащих транслитерацию его лексических единиц из одной алфавитной системы в другую. Таким образом, посредством сопоставления различных вариантов транслитерации возможно повысить точность осуществления поиска и перевода лексических единиц, а также обеспечить удобство эксплуатации для различных категорий пользователей.

Применение метода аналогии на основании уже имеющихся у потенциального обучающегося знаний родного языка способно в значительной степени повысить степень усвояемости языкового материала.

Таким образом, ожидаемые эффекты от внедрения мобильного приложения «Тывалап чугаалажыыл» ‘Поговорим по-тувински’ заключаются в следующем:

- возможность изучения тувинского языка посредством мобильного приложения вне зависимости от места и времени;
- самостоятельная и групповая формы обучения;
- использование разных видов речевой деятельности в обучении языку;
- визуальное восприятие информации.

Мобильное приложение «Тывалап чугаалажыыл» «Поговорим по-тувински» подготовлено на фичеринг и направлено на изучение и сохранение тувинского языка, разработано с помощью среды разработки Unity3d с использованием базы данных SQLite. Язык программирования – C#.



УДК 372.881.1

**ПРИМЕНЕНИЕ STEAM-ОБРАЗОВАНИЯ В ПРОЦЕССЕ  
ОБУЧЕНИЯ БУДУЩИХ УЧИТЕЛЕЙ ТУВИНСКОГО ЯЗЫКА***А. К. Тарыма,**А. М. Чулдум**Тувинский государственный университет**Кызыл, Россия*

taryma\_ak@mail.ru, ajkys.chuldum@bk.ru

В данной статье рассматривается подготовка будущих учителей тувинского языка к работе в условиях изменения современной школы. А именно рассматривается основное понятие и роль STEAM-технологий в процессе формирования у студентов педагогического направления профессиональных компетенций, необходимых для работы в школе. В работе выделены ключевые тенденции в системе высшего педагогического образования, актуализированные с помощью STEAM-технологий. Предмет исследования: применение STEAM-образования в процессе обучения будущих учителей тувинского языка. STEAM – это аббревиатура от английских слов Science (естественные науки), Technology (технологии), Engineering (инженерия, проектирование), Mathematics (математика). Иными словами, STEAM-образование предлагает систему межпредметных знаний для проектной работы в области научных и инженерных технологий. Сегодня, идет активное продвижение от STEAM к STEAM образованию, с активным дополнением творческих и художественных предметов и дисциплин, например, дизайн, искусство, архитектура, индустриальная эстетика и других, объединенных общим термином Arts. Под STEAM также понимают ряд или последовательность курсов или программ обучения, готовит учеников к успешному трудоустройству, к образованию после школы или для того и другого, требует различных и более технически сложных навыков, в частности с применением математических знаний и научных понятий.

**Ключевые слова:** STEAM – образование, STEAM – технология, подготовка будущих учителей, ИКТ, моделирование.

**APPLICATION OF STEAM TECHNOLOGIES IN TRAINING  
TEACHERS OF THE TUVAN LANGUAGE***Aldynsai Taryma, Ai-kys Chuldum**Tuvan State University**Kyzyl, Tuva, Russia*

taryma\_ak@mail.ru, ajkys.chuldum@bk.ru

This article discusses the preparation of future teachers of the Tuvan language to work in a changing modern school. Namely, the main concept and the role of STEAM technologies in the process of formation of students of the

pedagogical direction of professional competencies necessary for work at school are considered. The paper highlights the key trends in the system of higher pedagogical education, updated with the help of STEAM technologies. Subject of research: the use of STEAM education in the process of teaching future teachers of the Tuvan language. STEM is an abbreviation of the English words Science (natural sciences), Technology (technology), Engineering (engineering, design), Mathematics (mathematics). In other words, STEM education offers a system of interdisciplinary knowledge for project work in the field of scientific and engineering technologies. Today, there is an active promotion from STEM to STEAM education, with the active addition of creative and artistic subjects and disciplines, for example, design, art, architecture, industrial aesthetics and others, united by the common term Arts. STEAM is also understood as a series or sequence of courses or training programs, prepares students for successful employment, for education after school or for both, requires various and more technically complex skills, in particular with the use of mathematical knowledge and scientific concepts.

**Keywords:** STEAM – education, STEAM – technology, training of future teachers, ICT, modeling.

В настоящее время Республика Тыва находится на стадии постепенного перехода к четвертой технологической революции, связанной с активным внедрением системы автоматизации, искусственного интеллекта и био- и нанотехнологий во все сферы человеческой деятельности, что требует постоянного обновления знаний и умений, необходимых для успешного овладения новыми технологиями. Применение высоких технологий позволяет в полной мере реализовывать цели и задачи обеспечения современного качества образования, напрямую связанного с реализацией федерального проекта «Кадры для цифровой экономики» [Григорьев, 2018].

Процесс внедрения цифровых технологий в сферу образования обуславливает новые требования к подготовке будущих учителей разных специальностей, в том числе и гуманитарных. Учителя гуманитарных предметов играют важную роль в реализации культурообразующей функции школы. При этом в федеральных стратегических документах: Указ Президента Российской Федерации от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017 – 2030 годы» [2]; Распоряжение Правительства Российской Федерации от 28.07.2017 № 1632-р «Об утверждении программы «Цифровая экономика Российской Федерации» (раздел 2 – «Кадры и образование») [2] и т.д. указано, что педагогические кадры нового поколения должны

владеть конвергентными или междисциплинарными знаниями из разных образовательных областей естественных наук, инженерии и технологии, умением свободно ориентироваться в общемировом потоке информации, квалифицированно находить и обрабатывать нужные данные и затем на их основе принимать решения, использовать новейшие технологии в профессиональной деятельности.

Однако, несмотря на общее ускорение процесса глобализации образования, многие учителя гуманитарного профиля, в том числе учителя родного (тувинского) языка, отмечают недостаточное владение компетенциями, необходимыми для нового века цифровых технологий, связанных с использованием междисциплинарных и прикладных знаний для обработки информации в области культурного наследия в целях сохранения, использования в научно-исследовательской и образовательной деятельности.

Одним из путей решения обозначенной проблемы может стать использование технологий STEAM (Science, Technology, Engineering, Arts, Mathematics) в процессе обучения информатике будущих учителей тувинского языка. Особенностью ее является интеграция не только инженерных и естественнонаучных STEM-предметов, но и гуманитарных и творческих дисциплин, способствующих развитию креативного мышления обучающихся, позволяя им гармонично сочетать в проектной деятельности научную строгость и творческую свободу [Блинов, 2019].

Один из путей решения данной проблемы является – внедрение в учебный план дисциплины «Цифровые методы в сохранении и презентации культурного наследия» направления подготовки 44.04.01 Педагогическое образование по программе «Цифровые технологии в гуманитарном образовании», которая демонстрирует гибридизацию цифровых технологий и тувинского филологического образования в рамках современного динамично развивающегося междисциплинарного направления науки – цифровой гуманитаристики (Digital Humanities).

Цель этой дисциплины заключается в том, чтобы обучить студентов использованию современных цифровых методов и технологий для сохранения и представления культурного наследия. Это включает в себя не только создание точных копий предметов культурного наследия, но и создание виртуальных музеев и галерей, в которых можно изучать и исследовать культурное наследие, не выходя из дома.

По учебному плану данную дисциплину изучают во втором семестре. Общее время, отводимое на ее изучение – 72 часа (10 ч. – лекции; 10 ч. – практические работы; 52 ч. – самостоятельная работа).

Студенты, изучающие эту дисциплину, могут узнать о том, как использовать цифровые технологии для создания цифровых копий культурных объектов, их сохранения и доступности для широкой аудитории. Они также могут изучать техники виртуализации и создания виртуальных выставок, а также обзорные экскурсии по музеям и памятникам, которые могут использоваться для популяризации и обучения культурному наследию.

Рабочая программа дисциплины «Цифровые методы в сохранении и презентации культурного наследия» для направления подготовки «44.04.01 Педагогическое образование по программе «Цифровые технологии в гуманитарном образовании» может включать, например, следующие разделы и темы:

- Web-программирование и дизайн;
- Цифровые методы в сохранении и презентации культурного наследия;
- Программирование на языке Python;
- Введение в анализ данных;
- Основы графического дизайна и 3D-моделирования;
- Технологии создания цифровых культурных объектов;
- Информационные системы и базы данных в гуманитарных областях;
- Основы SQL и работы с базами данных.

Таким образом, внедрение STEAM-образования в процессе подготовки будущих учителей тувинского языка позволит создать условия для непрерывного обучения, построения карьеры и профессионального роста педагога в соответствии с требованиями образования будущего.

## ЛИТЕРАТУРА

1. Григорьев С.Г., Курносенко М.В. Внедрение элементов STEM-образования в подготовку педагогов по профилю «Информатика и технологии» // Известия института педагогики и психологии образования. – 2018. – № 2. – С. 5–13.

2. Стратегия развития информационного общества в Российской Федерации на 2017–2030 годы [Электрон. ресурс] // Судебные и нормативные акты РФ – URL: <https://sudact.ru/law/ukaz-prezidenta->

rf-ot-09052017-n-203/strategiia-razvitiia-informatsionnogo-obshchestva-v/?ysclid=lpgae0itl898261834 (Дата обращения: 19.10.2023)

3. Блинов В.И., Дулинов М.В., Есенина Е.Ю., Сергеев И.С. Проект дидактической концепции цифрового профессионального образования и обучения. – М.: Издательство «Перо», 2019. – 72 с.

4. Тарыма А.К. ИКТ-компетентность учителей тувинского языка / А.К., Тарыма // Сборник трудов участников конференции. – Саратов: Изд-во ГОУ ДПО «СарИПКипРО», 2009. – 368 с. – С. 298–301.

---

# ЛИНГВИСТИЧЕСКИЕ ТЕХНОЛОГИИ

УДК 81'33

## МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСЧЕТА НОРМ ЯЗЫКА ОРТАТЮРК КАК МОДЕЛЬ ЯЗЫКОВОГО РАЗВИТИЯ СИСТЕМЫ ТЮРКСКИХ ЯЗЫКОВ

**Б. Р. Каримов<sup>1</sup>, Ш. Ш. Муталов<sup>2</sup>**

*<sup>1</sup>Институт востоковедения Академии наук  
Республики Узбекистан*

*<sup>2</sup>Ташкент, Узбекистан*

karimov.bahtiyor@yahoo.com, Shahahmad.mutal@gmail.com

Предлагается создать усредненные языки методами математической и компьютерной лингвистики на основе усреднения лексических, фонетических и грамматических норм генеалогически родственных языков. Сущность предлагаемого метода состоит в том, чтобы с помощью специальной математической процедуры определить общий фонд для групп родственных языков и кодифицировать общий фонд как эталон, т.е. как единицы соответствующих уровней усредненного языка. Единицы каждого из уровней языка имеют варианты: разные значения слова, позиционные варианты фонем и морфем, разные способы выражения синтаксических отношений. Процедура усреднения состоит в том, что единицам языковых уровней сопоставляются векторы и на этой основе производится вычисление лингвистических норм усредненного языка. Предлагается идея создания усредненного мирового языка путем усреднения усредненных и изолированных языков. Исследованы социальные, культурные, лингвистические, информационные, коммуникативные проблемы усредненных языков, пути использования этой математической модели в области тюркских языков и создания языка ОртаТюрк.

**Ключевые слова.** Усредненные языки; языковая идентичность; этнолингвопанизм; математическая и компьютерная лингвистика, тюркские языки, ортаТюрк

## MATHEMATICAL MODEL FOR CALCULATING THE NORMS OF THE ORTATURK LANGUAGE AS A MODEL OF LINGUISTIC DEVELOPMENT OF THE SYSTEM OF TURKIC LANGUAGES

***Bakhtiyor Karimov<sup>1</sup>, Shoakhmad Mutalov<sup>2</sup>***

*<sup>1</sup>Institute of Oriental Studies of the Academy of Sciences  
of the Republic of Uzbekistan, Tashkent, Uzbekistan*

*<sup>2</sup>scientific researcher, Tashkent, Uzbekistan*

karimov.bahtiyor@yahoo.com, Shahahmad.mutal@gmail.com

Creation averaged languages by using the methods of mathematical and computer linguistics based on averaging vocabularies and grammar rules and norms of genealogically cognate languages has been proposed. The essence of the proposed method is to use a special mathematical procedure to determine the common fund for groups of related languages and codify the common fund as a standard, i.e. as units of the corresponding levels of averaged language. Units at each level of language have variants: different word meanings, positional variants of phonemes and morphemes, different ways of expressing syntactic relationships. The averaging procedure consists of comparing vectors to units of linguistic levels and, on this basis, calculating the linguistic norms of the averaged language. The idea of creating averaged world language is proposed by means of averaging the averaged and isolated languages. Social, cultural, linguistic, information, communication problems of averaged languages, ways of using this mathematical model in the area of Turkic languages and creating of language Ortaturk have been investigated.

**Keywords** – averaged languages; language identity; ethnolingvopanism; mathematical and computer linguistics, Turkic languages, Ortaturk

## **Introduction**

While exchanging information, the parties must understand the language that the information is encoded in. The language of social communication, i.e. oral language (sound encoding of information), written language (text encoding of information), as well as partially sign language, the language developed for the community of deaf people, make the basis of information exchange in the modern epoch. The real situation in the system of the world civilisation radically differs from this idealised model that, thus, happens to be applicable to the social group knowing certain presumed common language. For to be closer to adequate comprehension of the reality, it is necessary to go beyond this idealised model and establish a wider model embracing the former as a particular case.

The language situation has been studied in many research works that show complexity, controversy, conflict, dramatic, tragic nature of relations of language groups in the world civilisation. Although the world civilisation recognises that each language is shared value of the humankind, the research shows that one of living languages vanishes, turning into extinct language. The language identity is closely related to the national and ethnic identity, therefore, the language processes are connected with national and ethnic processes.

The today's course of the world civilisation language development evolves on the basis of establishment of nation-states, where the

language of so called state forming nations is recognised as the state language, and the languages of other nations in the territory, in the majority of cases, acquires a second grade status. As the majority of nations fail to establish their nation-state, the languages of such nations find themselves in the second grade status. Such nations make majority of nations in the world and their languages make the majority of the world languages.

## 1. Statement of the problem

The paper proposes different way of addressing this fundamental problem of social communication in the modern information civilisation based on the rational scientific regulation of information exchange and language processes based on the methods of mathematical and computer linguistics.

Despite some peculiarities of nation formation process the existing nations have developed of related tribes and all-nation language, literary language of a nation, has developed on the base of tribe dialects as a result of their interference. From the point of view of genesis it seems quite natural that germs of nations were among related tribes because related tribes inhabited neighbouring territories and had easy communications. Besides, the closer were their languages the easier was the communication.

At present, when unity of mankind is necessary for its survival, it is expedient to consciously seek for the ways of formation and preserving such unity. Consolidation of close nations into federations, confederations, or unitary states based on their will is one of such ways.

A regional mediating language must be a language that is equally close to the people's languages of the region where it is supposed to be a mediating language. The world mediating language must meet similar requirements. In our opinion, regional mediating languages should be constructed on the basis of group of close languages according to a special method providing its neutrality as well as optimum degree of its intelligibility to as many people as possible among native speakers of these languages.

Acceleration of the social processes, emergence of the humankind global problems system brings the necessity to move to noosphere, to simulation and control of information and language processes, in particular. The mediating language creation method must conform and imitatively simulate objective mechanism of *koine* formation process



in the course of communication of the related languages native speakers. It means the more is number (1) of the related languages where and (2) the individual native speakers of these languages with whom certain language form (or norm) is observed the larger the probability for this very form (or norm) to be included into the *koine* forms or norms system.

The above is not merely theoretical speculation. Mechanism, or method, of mediating language creation is important because when persons that consider related languages as their vernacular realize their ethnic affiliation the closeness degree of the related languages, realization of the persons' languages by them as independent languages or dialects of one language play important, though not decisive, role. In the systems of dialects of various languages in the world the dialects closeness measure that results in realization of affiliation with the same language by bearers of a dialect varies essentially from one language to another.

For the vocabulary it is necessary to select an upper part of a certain length from the frequency dictionary of one of the languages. This language serves as a basic one. The meanings of polysemantic words that are marked in the dictionary entry of the basic language by figures are considered as separate lexemes. The chosen words in the list are to be numerated. For every lexeme of ordered list the equivalents should be found in the languages or dialects the averaged language must be constructed on.

## 2. The concept of the problem solution

The essence of the offered method is to define a common fund for groups of dialects or languages by means of a special mathematic procedure, and to codify the common fund as a standard, i.e. as units of corresponding levels of the language being compiled. The procedure is as follows: units of language levels are connected to vectors. As a rule, the level units have versions: different meanings of a word, positional variants of phonemes and morphemes, different ways of expressing syntactic relations. A vector space is associated with the unit system of a level and this space is constructed in such a way that when the main meaning of a unit is bound with magnitude 1, other variants are correlated with numbers less than 1 according to the formula:

### 3. Realization of the concept

The essence of the proposed method [Karimov B.R., Mutalov Sh.Sh., 1992] is identification of the common stock for a group of dialects or languages by means of special mathematical procedure and codification of this common stock as a standard, i.e. as a unit of respective level of the language we create. The procedure is as follows: the units of the language levels are related to vectors. As a rule, the level units have versions: different meanings of the words, positional versions of phonemes and morphemes, different options for expressing syntactic relations. The vector space related to the level units is built in a way that the main meaning is attached the value of 1, other versions correspond to values smaller than 1, according to the formula

$$X_{n,\alpha}^i = 1 - \frac{i-1}{2s}, \quad (1)$$

where “n” stands for the number of units in the list of units; “α” stands for a concrete language; “i” stands for the variant number of the level unit with number “n”; “s” is the number of variants of the level units with number “n”. These numbers form a set of the vector components. We named the method averaging method and the languages built by the method averaged languages.

The next step is to order the roots in accordance with the values of the function.

Then tables of simple lists are to be transformed into other ones, where the words of same roots are located in the same line.

A function  $F_n^j(\hat{x})$  is introduced for the words of same root with the formula:

$$F_n^j(\hat{x}) = \sum_{\alpha=1}^A \left( 1 + \frac{K_\alpha}{K} \right) X_{n,\alpha}^j \quad (2),$$

where A – total number of the languages under averaging,  $K_\alpha$  – number of native speakers of the language α,  $\bar{K}$  – the arithmetic mean of the native speakers of one group of kin languages that is calculated as fraction of dividing the total number of native speakers of all the languages of a group by number of languages in the group, j means the ordinal number of the words of same root in the ordered tables.

The factor  $1 + \frac{K_\alpha}{K}$  serves for taking into account relative native speakers’ number of a language from a group of languages the averaged language is being built for.

The first roots in the list are included into vocabulary of the averaged language as the main meanings of the words. Other words make a kind of synonyms stock for enrichment of averaged language vocabulary.

All the work may be accomplished by means of computers. The advantages of the suggested solution of world language problem are as follows:

1. as averaging will be done on the basis of ethnic languages (dialects), native speakers of those languages (or dialects), averaged language will be built on, are sure to understand it to some extent without special learning;

2. belonging to none ethnic community the averaged language gives no privilege to either of them, so it will not promote national discord based on the language policy;

3. the averaged language eliminates some arbitrariness in choosing one of the local languages as official state language as well as interethnic conflicts;

4. the averaged language eliminates the introducing of languages of former colonizers as the only state official language, weakens the dependence on the former parent state in the spheres of culture and education.

5. many nations speaking the same language are divided by state borders. Thus an averaged language constructed by the offered method might play the role of macromediator.

One more advantage of averaged languages may be brought to the light here – averaged language may be a rescue for endangered languages many of which are cognate minor languages or dialects.

In the process of formation of universal information civilization, it is expedient to develop computer soft ware for each language that transforms text in one alphabet into the text in another alphabet. It is expedient to develop soft ware for translation from one language into another. In this process, translation of the texts into averaged language of a certain group on language could be the main stage of translation into other genealogically kin languages. It is necessary to expand information and multimedia resources in the Internet in the national and averaged languages. In particular, it is expedient to develop Wikipedia in the national and averaged languages. Implementation of these proposals would foster development both of each local civilization and the system of world information civilization as a whole. [Karimov B.R. Mutalov Sh.Sh., 1993, 2008, 2019; Karimov B.R., 2003]

## Conclusion

Mathematical model for calculating the norms of the Ortaturk language correspond to a model of linguistic development of the system of Turkic languages. The language Ortaturk would be a closely related language to most Turkic languages. It can be studied in addition to studying the native Turkic language. Each Turkic nation will be able to choose for itself the forms, methods and level of studying the Ortaturk language, based on its national and state interests. The Ortaturk language will be a voluntarily used language of interethnic communication, a language of accumulation of information of general Turkic and global significance in the community of Turkic nations.

Each national Turkic language will have the status of a state language and will fully develop in its own national state, on its own ethnic territory.

The functioning of the Ortaturk language will serve to strengthen the attention of representatives of Turkic languages to their own languages, giving grounds for hope and faith in the usefulness and promise of their native languages for familiarization with world culture. The democratic nature of the procedures for the creation and functioning of the language Ortaturk, their compliance with universal human norms of linguistic relations leads to the voluntary study of this language by the majority of Turkic nations and Turkic-speaking individuals along with their native national Turkic languages. The use of the language Ortaturk does not lead to waste, but to savings, to optimization of the use of vital forces of both each of the Turkic nations and the entire system of Turkic nations as a whole.

To achieve the goals of understanding and explanation, the Ortaturk language will not require special study for many Turkic peoples, since it will be very close to them. Speakers of a number of other Turkic languages will require some additional training for these purposes, primarily in terms of mastering the changed frequency of word use and differences in grammatical forms.

To do this, after the creation of the Ortaturk language, it will be necessary to train a group of enthusiasts to a high level of practical knowledge of the normative system of this language. Members of this group could specialize in performing a number of functional duties in the media: announcers of common Turkic and national television and radio broadcasting in the language Ortaturk; editors of newspapers, magazines and books published in that language; teachers of the lan-

guage Ortaturk teaching via radio, television, and computer information systems. Frequently repeated auditory and visual perception of information transmitted in the Ortaturk language through the media will lead to passive acquisition of this language by the majority of speakers of Turkic languages and an increase in the degree of mutual understanding of speakers of various Turkic languages, even if they actively use only their native Turkic languages. After some time, a significant part of the speakers of Turkic languages, who often find themselves in situations of inter-Turkic linguistic communication, will develop the skills of some transformation of their idiolects in the direction of a number of basic norms of the Ortaturk language in the corresponding language situations.

#### REFERENCES

1. Karimov B.R. The oikumenic concept of the nation and development of languages. *Oykumenicheskaya kontsepsiya natsii i razvitiye yazykov*. Qarshi, 2003. 160 p.
2. Karimov B.R., Mutalov Sh.Sh. *Ortaturk tili*. Toshkent, 1992. 32 p.
3. Karimov B.R., Mutalov Sh.Sh. *Averaged languages: an attempt to solve the world language problem*. Tashkent: Fan, 1993. (the second revised editions in 2008, the third – in 2019). 60 p
4. Karimov B.R., Mutalov Sh.Sh. *Usrednyonnye yaziki: popytka resheniya mirovoy yazykovoy problemy*. Tashkent: Fan, 2008. (the second revised editions in 2019). 64 p.

УДК 81'322.4

**УЗБЕКСКИЙ НЕЙРО МАШИННЫЙ ПЕРЕВОДЧИК  
НА БАЗЕ BART: ИСПОЛЬЗОВАНИЕ  
МНОГОИСТОЧНИКОВЫХ ДАННЫХ*****А. И. Зохиоров<sup>1</sup>, Н. З. Абдурахмонова<sup>2</sup>, А. М. Нарзуллаев<sup>2</sup>***<sup>1</sup>*Mohirdev, Ташкент, Узбекистан*<sup>2</sup>*Национальный университет имени Мирзо Улугбека  
Ташкент, Узбекистан*adham@mohirdev.uz, n.abduraxmonova@nuu.uz,  
anvar@mohirdev.uz

В этой статье представлено новое исследование нейронного машинного перевода (NMT) на узбекский язык, язык с низким уровнем ресурсов, с использованием архитектуры BART (двунаправленные и авторегрессивные преобразователи). Признавая проблемы, с которыми сталкиваются языки с ограниченными ресурсами в NMT, в исследовании особое внимание уделяется разработке и внедрению индивидуальных решений для таких языков на примере узбекского языка.

Исследование основано на комплексной стратегии сбора данных с использованием разнообразных наборов данных из различных источников, включая онлайн-платформы и чат-бота MohirAI. Этот подход решает проблему нехватки параллельных корпусов в языках с ограниченными ресурсами, что является критическим барьером в NMT. В документе также рассматривается эволюция NMT, прослеживается его путь от интеграции с традиционными статистическими системами машинного перевода до появления моделей на основе преобразователей, подчеркиваются ключевые разработки, такие как сверточные модели и модели последовательности-последовательности, а также ключевая роль механизмов внимания.

**Ключевые слова:** естественный язык; искусственный интеллект; технологии машинного перевода; машинный перевод

**BART-BASED UZBEK NMT: LEVERAGING  
MULTISOURCE DATA*****Adkham Zokhirov<sup>1</sup>, Nilufar Abdurakhmonova<sup>2</sup>, Anvar Narzullaev<sup>2</sup>***<sup>1</sup>*NLP Team Lead at Mohirdev*<sup>2</sup>*National University of Uzbekistan named after Mirzo Ulugbek,  
Tashkent, Uzbekistan*adham@mohirdev.uz, n.abduraxmonova@nuu.uz,  
anvar@mohirdev.uz

In recent years, Neural Machine Translation (NMT) has achieved remarkable progress, primarily due to the advent of transformer-based architectures such as

BART (Bidirectional and Auto-Regressive Transformers). Despite the significant strides in high-resource language pairs, low-resource languages, such as Uzbek, still face substantial challenges in achieving accurate and contextually appropriate translations. This paper delves into the domain of Neural Machine Translation using BART specifically tailored for Uzbek, a low-resource language with limited parallel corpora.

A significant contribution of this research lies in our comprehensive data collection strategy. We gathered diverse datasets from a multitude of sources, including websites, blogs, and communication with our MohirAI chatbot.

Our findings not only contribute to the growing body of research in low-resource NMT but also offer practical insights for improving translation systems for other underrepresented languages. As the demand for accurate machine translation services continues to rise globally, our work provides valuable advancements in ensuring linguistic inclusivity and accessibility for speakers of low-resource languages like Uzbek.

**Keywords:** Neural machine translation, parallel corpora, transformer-based architectures such as BART, low-resource languages, Uzbek, English

Neural network training proceeds for several epochs, i.e., full iterations over the training data. Through track training progress, it could be seen that the error on the training set continuously decreases. However, at some point over-fitting sets in, where the training data is memorized and not sufficiently generalized. [Philipp Koehn, 2017]

The modern resurrection of neural methods in machine translation started with the integration of neural language models into traditional statistical machine translation systems. The pioneering work by Schwenk (2007) showed large improvements in public evaluation campaigns. However, these ideas were only slowly adopted, mainly due to computational concerns. The use of GPUs for training also posed a challenge for many research groups that simply lacked such hardware or the experience to exploit it. More ambitious efforts aimed at pure neural machine translation, abandoning existing statistical approaches completely. Early steps were the use of convolutional models (Kalchbrenner and Blunsom, 2013) and sequence-to-sequence models (Sutskever et al., 2014; Cho et al., 2014). These were able to produce reasonable translations for short sentences, but fell apart with increasing sentence length. The addition of the attention mechanism finally yielded competitive results (Bahdanau et al., 2015; Jean et al., 2015b). With a few more refinements, such as byte pair encoding and back-translation of target-side monolingual data, neural machine translation became the new state of the art.

## Methodology

### Data Collection:

In our pursuit of building a proficient Uzbek-English and English-Uzbek Neural Machine Translation system, we undertook a comprehensive data collection effort. Scrutinizing bilingual content from a variety of sources, including websites such as [it-park.uz](http://it-park.uz), [president.uz](http://president.uz), [egov.uz](http://egov.uz), and and communication with our MohirAI chatbot, to ensure a wide variety of language styles, topics, and contexts. Crucially, our dataset consisted of over 500,000 meticulously curated sentence pairs in both Uzbek and English. To create a comprehensive dataset for training and evaluating our model, we employed a multi-faceted approach to data collection. Data was gathered from diverse sources, including websites, blogs.

To ensure the highest quality of translations, we implemented a rigorous verification process. Each sentence pair was manually inspected and verified for accuracy and fluency by linguists proficient in both languages. This meticulous manual verification step played a pivotal role in enhancing the reliability and trustworthiness of our dataset.

#### *Websites and Blogs:*

We systematically crawled Uzbek websites and blogs, extracting text content in both formal and informal contexts. This method allowed us to capture written language from various domains such as news, literature, and everyday conversations.

#### *Communication with MohirAI Chatbot:*

To simulate real-world conversational data, we leveraged interactions with MohirAI, our chatbot designed to engage users in Uzbek language conversations. These interactions provided us with valuable informal language data, including slang, colloquialisms, and idiomatic expressions.

### Multilingual Website Crawling and Parsing

In our pursuit of creating a rich and diverse dataset for Uzbek-English and English-Uzbek Neural Machine Translation, we adopted a meticulous approach to data collection. Leveraging crawling techniques, we targeted multilingual websites to ensure a wide array of language styles, topics, and contexts. One of our innovative strategies involved the utilization of scripts tailored to handle multilingual content on websites, exemplified by links such as:

English Version: <https://president.uz/en/lists/view/6653>

Uzbek Version: <https://president.uz/oz/lists/view/6653>



To extract bilingual data from such multilingual websites, we engineered a sophisticated crawling script. This script was meticulously designed to navigate the intricacies of websites offering content in multiple languages. By identifying corresponding pages in both Uzbek and English versions, we ensured the extraction of parallel sentence pairs essential for our translation task.

### **Handling Language Variants**

Multilingual websites often present challenges due to variations in language structure and content placement across different language versions. Our script was engineered to adeptly handle these challenges, enabling the extraction of aligned sentence pairs even in the presence of divergent webpage structures.

### **Data Parsing and Structuring**

Upon successful crawling, the extracted data underwent parsing and structuring processes. Our parsing algorithms were tailored to extract text while preserving the original sentence alignments between the Uzbek and English versions. Special attention was given to preserving the semantic meaning and context during the parsing phase.

### **Enriching the Dataset with Multilingual Content**

By employing this innovative approach of crawling and parsing content from multilingual websites, we significantly enriched our dataset. The inclusion of diverse, real-world multilingual content played a pivotal role in enhancing the robustness of our training data, enabling our models to capture the intricacies of translation in both Uzbek to English and English to Uzbek directions.

### **Data Preprocessing and Alignment:**

The collected dataset, comprising the verified 500,000+ sentence pairs, underwent thorough preprocessing and alignment. Employing advanced algorithms, we meticulously aligned corresponding sentences between Uzbek and English, maintaining the integrity of the bilingual context. The collected data underwent rigorous preprocessing to ensure uniformity and consistency. This preprocessing included tokenization, lowercasing, removing special characters, and cleaning noisy data from web sources. Additionally, we performed language-specific preprocessing techniques to handle Uzbek-specific challenges, such as agglutination and complex morphological structures.

### Data Splitting:

| Dataset Split   | Description   | Percentage of Total Dataset |
|-----------------|---|-----------------------------|
| Training Data   | Sentence pairs used for training the BART model                               | 80%                         |
| Validation Data | Subset used for fine-tuning hyperparameters and optimizing model performance. | 10%                         |
| Test Data       | Independent subset for comprehensive evaluation of translation quality.       | 10%                         |

*\*Total Dataset Size: 500,000+ sentence pairs (bilingual Uzbek-English and English-Uzbek)*

To facilitate effective training, validation, and testing of our Neural Machine Translation model, we split the preprocessed dataset into the following segments:

**Training Data:** This subset was used for training the BART model. It comprised 80% of the cleaned dataset and served as the foundation for the model's learning process.

**Validation Data:** 10% of the dataset was set aside for validation purposes. This subset allowed us to fine-tune hyperparameters, assess the model's performance during training, and prevent overfitting.

**Testing Data:** The remaining 10% of the dataset was used as a separate test set. It was crucial for evaluating the model's generalization to unseen data and determining its overall translation quality.

## Modeling and Fine-Tuning

### Model Selection:

For our Uzbek-English and English-Uzbek Neural Machine Translation task, we opted for the BART (Bidirectional and Auto-Regressive Transformers) model. BART, an extension of the Transformer architecture, has demonstrated exceptional capabilities in various natural language processing tasks, including machine translation. Its bidirectional and auto-regressive nature makes it ideal for our bidirectional translation requirements.

### Pretraining and Initialization:

We initialized our BART model with pre-trained weights to harness the power of transfer learning. These weights were pre-trained on large-scale multilingual corpora, enabling the model to learn universal language representations. This initialization jump-started our training process, allowing the model to leverage its knowledge of multiple lan-

guages, which is especially crucial for low-resource languages like Uzbek.

### **Training Configuration:**

The training process was conducted over an intensive three-day period, utilizing two Nvidia 3090 24GB GPUs. This high-performance hardware configuration significantly accelerated the training speed and allowed us to handle the extensive dataset efficiently.

During training, we employed mixed-precision training techniques, utilizing 16-bit floating-point precision, which effectively reduced memory usage and accelerated the computations. This approach maximized the utilization of the available GPU memory, ensuring efficient processing of the large-scale dataset.

### **Optimization and Fine-Tuning:**

Optimizing the model's performance was a multifaceted process. We employed the Adam optimizer with a carefully tuned learning rate schedule. The learning rate was adjusted dynamically during training, ensuring the model's convergence to an optimal solution. Gradient clipping techniques were applied to prevent exploding gradients, enhancing the stability of the training process.

To enhance training stability and facilitate convergence, we utilized teacher forcing, a technique where the model is trained using the ground truth (reference) translations during training. This approach helped the model learn correct translations and mitigate the risk of error accumulation during the training process.

### **Training Strategies and Challenges:**

Training a Neural Machine Translation model for low-resource languages presents unique challenges. The scarcity of labeled data necessitates innovative training strategies. To address this, we explored techniques such as back-translation and data augmentation. Back-translation involves generating synthetic source-target pairs from monolingual data, effectively augmenting our dataset and providing additional training signal for the model.

### **Iterative Fine-Tuning and Model Evaluation:**

Our training approach was iterative and guided by continuous evaluation. We regularly evaluated the model's performance on the validation set, adjusting hyperparameters and training strategies accordingly. This iterative fine-tuning process was instrumental in refining the model's translation quality and ensuring its adaptability to diverse linguistic patterns and contexts.

By leveraging the computational power of two Nvidia 3090 24GB GPUs, coupled with advanced training strategies, we endeavored to create a Neural Machine Translation system that excels in translating between Uzbek and English. The iterative fine-tuning, coupled with the robust evaluation process, stands as a testament to our commitment to delivering accurate, fluent, and contextually appropriate translations in both translation directions.

## Results

### Quantitative Evaluation Metrics

Our Uzbek-English and English-Uzbek Neural Machine Translation systems were rigorously evaluated using established metrics to assess the quality and fluency of translations. The following key metrics were employed for the quantitative evaluation:

- **BLEU (Bilingual Evaluation Understudy):** BLEU scores were calculated to measure the overlap between machine-generated translations and human reference translations. Higher BLEU scores indicate better translation quality.

- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** ROUGE metrics were used to evaluate the overlap between machine-generated translations and reference summaries. ROUGE focuses on recall and provides insights into the quality of generated translations, especially for tasks such as text summarization.

text

Poytaxt aholisi soni ik bor 1960 yilda 1 million, 1985 yilda 2 million, hozirgi kunga kelib 3 millondan oshdi.  
2023 yil 1 oktyabr holatiga Toshkent shahrida yashovchi doimiy aholi soni 3 mln 20.5 ming kishini tashkil etgan.  
Statistika agentligi ma'lumotlariga ko'ra, shundan erkaklar - 1,4 mln kishini, ayollar - 1,5 mln kishini tashkil etgan.  
Poytaxt aholisi soni ik bor 1960 yilda 1 million, 1985 yilda 2 million, hozirgi kunga kelib 3 millondan oshdi.  
Bundan oldin Toshkentda aholisi soni eng ko'p tuman ma'lum qilingandi.  
Shahar hududlari kesimida tabiiy shuni ko'rsatganlik, ene ko'p aholi soni Olmazor tumaniida bo'lib, 395 ming kishini tashkil etgan.

Choose a language  
Choose a language pair

uz-en

Choose a model version  
Choose a model version

v2

Clear Submit

output

The population of the capital was 1 million in 1960, 2 million in 1985, and 3 million today. As of October 1, 2023, the total population of permanent residents living in the city of Tashkent was 3 million 20.5 thousand. According to the statistics agency, there were 1.4 million men and 1.5 million women. The population of the capital was 1 million in 1960, 2 million in 1985, and 3 million today. It was previously reported that the population of the most districts in Tashkent is the most. This was shown by the analysis of the population between the regions of the city, the largest population was in Olmazor district, which was 395,000 people.

Flag

## Experimental Results

The BART model, fine-tuned on our extensive dataset of over 500,000 manually verified sentence pairs, exhibited exceptional performance in both Uzbek-English and English-Uzbek translation tasks. The results demonstrated the effectiveness of our meticulous data pre-processing, extensive dataset, and advanced training techniques. Here are the summarized results:

| Translation Direction        | BLEU Score | ROUGE Score |
|------------------------------|------------|-------------|
| Uzbek to English Translation | 0.62       | 0.41        |
| English to Uzbek Translation | 0.64       | 0.43        |

## REFERENCES

1. N. Abdurakhmonova, I. Alisher and G. Toirova, “Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing,” 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 73-75, doi: 10.1109/UBMK55850.2022.9919521.
2. N. Abdurakhmonova, I. Alisher and R. Sayfulleyeva, “MorphUz: Morphological Analyzer for the Uzbek Language,” 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 61-66, doi: 10.1109/UBMK55850.2022.9919579.
3. Aripov, M., Sharipbay, A., Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology – Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018)
4. Abduraxmonova, N. Z. “Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef.” (2018).
5. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. *Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta*. 2016;2 (38):12-7.
6. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
7. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. *Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL)*. 2019;6(1-2019):131-7.

8. Abdurakhmonov N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation. Journal of Social Sciences and Humanities Research. 2017;5(03):89-100.

9. Kubedinova L. Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N. First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020) .2020/11: 90-101

10. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.

УДК 81'322.3

**МОДЕЛИРОВАНИЕ КАРАКАЛПАКСКОЙ ГЛАГОЛЬНОЙ  
ГРУППЫ ДЛЯ ЭТАПА МОРФОАНАЛИЗА*****А. З. Отемисов, Ж. Шарбаев****Каракалпакский государственный университет  
им. Бердака Нукус, Каракалпакстан, Узбекистан  
utemisov.aziz@mail.ru, sharbaevj77@gmail.com*

**Аннотация.** В статье говорится о моделировании каракалпакской глагольной группы слов для этапа морфоанализатора. Для этапа морфоанализатора использованы работы родственных тюркских народов и других зарубежных ученых по моделированию глагольной группы слов. В статье рассматриваются такие вопросы, как создание базового словаря глаголов каракалпакского языка, включение их в словари с аффиксами и без них, образование глаголов, суффиксов множественного числа и суффиксов личного числа. Ещё рассматриваются такие вопросы, как индексирование в словаре подобном формате и пометки записей раскрываются с айди.

**Ключевые слова:** морфология, морфоанализатор, аффикс, ID, глагол, составной глагол, существительное действия, переходный глагол, непереходный глагол, вспомогательный глагол, наречие, символ.

**MODELING THE KARAKALPAK VERB GROUP FOR THE  
MORPHOANALYSIS STAGE*****Aziz Otemisov, Jaras Sharbaev****Berdak Atyndagy Karakalpak State University  
Nukus, Karakalpakstan, Uzbekistan  
utemisov.aziz@mail.ru, sharbaevj77@gmail.com*

**Abstract.** The article deals with the modeling of the Karakalpak verbs that supposed to be as part of speech for the morphoanalyzer stage. For the morphoanalyzer stage, the works of the sisterly Turkic peoples and other foreign scientists on the modeling of the verbs were used. The issues discussed in the article include creating a basic dictionary of verbs of the Karakalpak language, including them in dictionaries with and without affixes, forming verbs, plural suffixes, and personal-number suffixes. Issues such as indexing in a dictionary-like format and tagging entries with IDs are revealed.

**Key words:** morphology, morphoanalyzer, affix, ID, verb, compound verb, action noun, transitive verb, non-transitive verb, auxiliary verb, adverb, symbol

## MORFOANALIZATOR BOSQICHI UCHUN QORAQALPOQ TILIDAGI FE'L SO'Z TURKUMINI MODELLASHTIRISH

*Ötemisov A. Z., Sharbaev J.*

*Berdax nomidagi Qoraqalpoq Davlat Universiteti*

*utemisov.aziz@mail.ru, sharbaevj77@gmail.com*

Maqolada qoraqalpoq tilidagi fe'l so'z turkumining morfoanalizator bosqichi uchun modellashtirish haqida so'z boradi. Morfoanalizator bosqichi uchun fe'l so'z turkumini modellashtirish bo'yicha qardosh turkiy xalqlar va boshqa xorijiy olimlarning asarlaridan foydalanildi. Maqolada so'z yuritilgan muammolarga qoraqalpoq tilining fe'l so'z turkumining bazaviy lug'atini yaratish, ularni affiksli va affikssiz shaklda lug'atlarga kiritish, fe'l so'zlarni yashovchi, ko'plik qo'shimchalari, shaxs-son affikslarini ham lug'at tarzida bazaga joylashtirish, affikslarga IDlar orqali kod kiritishga o'xshagan muammolar ochib beriladi.

**Kalit so'zlar:** morfologiya, morfoanalizator, affiks, ID, fe'l, qo'shma fe'l, harakat nomi, o'timli fel, o'timsiz fel, yordamchi fe'l, ravishdosh, simvol

**Kirish.** Morfologik analizator muayyan tokenning morfologik tarkibini tahlil qilish uchun mas'ul bo'lgan dastur sifatida izohlanadi. Morfologik analizator berilgan tokenni tahlil qiladi va turkum, turli grammatik ma'nolar kabi ma'lumotlarni shakllantiradi [Elov, 2022: 76]. Morfoanalizatorning ko'p vazifa va funksiyalari mavjud. Ulardan matnlarni tokenlarga ajratish, asos va qo'shimchalarni ajratish, so'zlarni shakllariga qarab turkumlarga ajratish, grammatik ma'nolarni keltirib chiqarish, matndagi so'z shakllarini tahlil qilish va boshqalar.

**Asosiy qism.** Morfologik tahlilda asosan matn imlosini tekshirish, so'zlarning grammatik shakllari va ularning tahlili nazarda tutilgan. Ushbu maqsadga erishish uchun quyidagi vazifalar amalga oshirilishi kerak.

1) tilning lug'at tarkibini komputerga kiritish, ya'ni elektron komputer lug'atini tuzish;

2) lug'atdagi adabiy tilga oid so'zlarni ajratish;

3) ajratib olingan so'zlarni qo'llanish xususiyatiga (ilmiy, badiiy, rasmiy va so'zlashuv) ko'ra guruhlariga taqsimlash;

4) adabiy so'zlarni turkumlarga ajratish;

5) so'zlarni so'z turkumi doirasida semantik guruhlar (masalan, otlarni shaxs otlari, o'simlik nomlari, yer-suv nomlari kabi guruhlariga ajratish).

6) guruhlariga ajratilgan so'zlarning qo'shimchalar kombinatsiyasini tuzish. Bunda qo'shimchalar kombinatsiyasini amalda adabiy til doirasida qo'llanishi qamrab olinadi. Kombinatsiyalarning qatorini tuzishda qo'shimchalar ketma-ketligiga e'tibor qaratiladi.



7) tuzilgan qo‘shimchalar kombinatsiyasi so‘zlarga birlashtiriladi.

8) lingvist tomonidan amalga oshirilgan yuqoridagi ishlar dasturchi tomonidan dasturga kiritiladi. [Abjalova, 2019: 64–65]

Komputer lingvistikasining rivojlanishi davomida matnni qayta ishlash dasturlarining algoritmlarda tilda mavjud bo‘lgan har bir lingvistik tushunchaga ramz berilgan bo‘lib, bugungi kunga kelib bunday ramzlar umumiy qabul qilingan belgilar sifatida foydalaniladi.

Tahlil dasturi bazasida xalqaro iste‘molga kiritilgan lingvistik tushunchalarning maxsus belgilarga ega shakllaridan foydalaniladi. Quyida mazkur ramzlarning umumiy qo‘llanishda bo‘lgan shakllari berildi, ish davomida ular yakkalanadi.

1) Ot – N (noun), ko‘plik shaklidagi ot – N<sub>s</sub>, turdosh ot – N<sub>com</sub>, atoqli ot – N<sub>prop</sub>, ot birliklar;

2) Sifat – Adj (adjective), sifat birliklar – AdjP;

3) Son – Num (numeral);

4) Fe‘l – V (verb), o‘timli fe‘l - V<sub>p</sub>, o‘timsiz fe‘l – V<sub>i</sub>, fe‘lning predikativ (shaxs-son shakli – finite form) shakli – Vf, fe‘lning nopredikativ shakli (shaxs, sonsiz shakli – nonfinite form) – Vnf, Vh – harakat nomi, Vs– sifatdosh, Vr – ravishdosh, birikkan va birikmali qo‘shma fe‘l – VP, yordamchi fe‘llar– Vaux;

5) ravishlar – Adv (adverb);

6) olmoshlar – Pron (pronoun);

7) ko‘makchilar – PostP (postposition);

8) bog‘lovchi – Conj (conjunction);

9) yuklama – Part (particle);

10) modal so‘zlar – Mod (modal);

11) undov so‘zlar – Interj (interjection);

12) taqlid so‘zlar – Mim. [Abjalova, 2019:16-17]. [Xolmanova, 2019: 62]

Yuqorida keltirilgan misollardan ilmiy ishimizda o‘rganiladigan obyektimiz fe‘l bo‘lib, dastlab fe‘llarni morfoanalizatorni dasturlashda fe‘lning bazasini yaratib olish lozim.

Fe‘l so‘z turkumi boshqa so‘z turkumlariga nisbatan murakkab grammatik kategoriyalarga, shakllar tizimiga ega. Usbu ishimizda umumiy nazariy adabiyotlarga tayangan holda fe‘llarni formallashtirishga e‘tibor qaratildi.

Adabiyotlarda fe‘llar kategorial va funksional shakllarga ajratiladi. Kategorial shakllar fe‘lning muayyan bir grammatik kategoriyasiga xos ma‘nolarni anglatadi. Ularga mayl, shaxs-son, zamon kiradi.

Funksional shakllarga fe'ning turli gap bo'laklari, yetakchi fe'l, sifatdosh, ravishdosh shakllari kiradi.

Fe'llar – harakat va holat ma'nolarini bildiradigan so'z turkumi.

Fe'ning boshqa so'z turkumlaridan o'ziga xosligi shundaki, leksik-semantik jihatdan harakat ma'nosiga, leksik-grammatik jihatdan bo'lishli va bo'lishsiz, o'timli va o'timsiz fe'l, nisbat, mayl, zamon, shaxs-son kategoriyalariga egaligi bilan ajralib turadi.

Fe'llar ko'pincha otlar bilan bog'liq qo'llaniladi, ularning leksik-semantik belgisi, sintaktik vazifasi otlarga nisbatan aniqlanadi. Otlar leksik-semantik jihatdan predmetlarning, tabiat hodisalarining nomlarini bildiruvchi so'z turkumi bo'lsa, fe'llar shu predmetlarning, tabiat hodisalarining ish-harakat jarayonini, belgisini ifodalaydi.

Fe'llar leksik-semantik va grammatik xususiyatlariga ko'ra ikki: yetakchi va ko'makchi fe'llar guruhiga ajraladi.

Ma'noli fe'llar to'liq leksik ma'noga ega bo'lib, ish-harakat va holat belgilarini bildiradi, gapda gap bo'lagi vazifasini bajaradi. Ular ikkinchi ma'noli so'zlarga birikib, so'z birikmasini hosil qiladi va uning boshqaruvchi komponenti bo'lib keladi.

Ko'makchi fe'llar bunday belgilarga ega bo'la olmaydi. Ular boshqa ma'noli so'zlar bilan qo'llanilib, ularga yordamchi grammatik ma'no yuklaydi. [Dáwletov, 2010: 135]

Morfologik tahlilni amalga oshirish uchun undagi qoidalar formal holatda kiritiladi. Morfologiya yoki sintaksisga komputer orqali munosabatda bo'lish bu tabiiy tilni modellashtirish, algoritm, tahlil yoki parsing orqali amalga oshiriladi. [Abduraxmanova, 2021: 101].

Fe'l so'z turkumini modellashtirishda yuqorida kiritilgan shartli belgilardan foydalanishimiz mumkin. Masalan, ism asosli qo'shma fe'llarni, fe'l asosli qo'shcha fe'llarni modellashtirishni quyidagi namuna orqali isbotlash mumkin:

*VP-apar, áper, ákel, baratır, kórip shıq, oqıp boldı, jazıp al, názer sal, sóyley-sóyley, bara-bara, ayta-ayta, kóre-bil;*

**o'timli fe'llarni esa:**  $V_t$ -*kúldır, túsındır, keltir, ótkiz, kes, shap, sazla, bawla, súyre, sana, ter, kór;*

**o'timsiz fe'llarni modellashtirishda :**  $V_1$ -*kiyindi, tarandı, juwındı, bezendi, kúldı, quwandı, qayğırdı, juladı, júrdı, sharshadı, jilamsıradı, esnedi;*

**harakat nomini modellashtirish:**  $V_h$ -*juwirıw, julqılaw, iytermelw, almaw, xabarlaw, jigerleniw, oqıw, jazıw;*

**sifatdoshni modellashtirish:** *Vs* - juwılmağan kóylek, oqıǵan , súrilgen jer, qaynaǵan suw, kirip kelgen ,shıǵıp ketken, alıp kiyatırǵan, jasaǵan.

**ravishdoshni modellashtirish:** *Vr*- súrinip, jıǵılıp, qulap, taranıp, barmay, sóylemey, kelmey, kórmegenshe.

**Fe'ning predikativ (shaxs-son shaklli) shaklini modellashtirish:** – *Vp*

| <i>I shaxs birlik/ko'plik</i> | <i>II shaxs birlik/ko'plik</i> | <i>III shaxs birlik/ko'plik</i> |
|-------------------------------|--------------------------------|---------------------------------|
| <i>Vp – bardım/ bardıq</i>    | <i>Vp – bardıń/ bardıńız</i>   | <i>Vp – bardı</i>               |
| <i>Vp – jazdım/ jazdıq</i>    | <i>Vp – jazdıń/ jazdıńız</i>   | <i>Vp – jazdı</i>               |
| <i>Vp – kórdım/ kórdik</i>    | <i>Vp – kórdıń/ kórdıńız</i>   | <i>Vp – kórdi</i>               |
| <i>Vp – oqıdım/ oqıdıq</i>    | <i>Vp – oqıdıń/ oqıdıńız</i>   | <i>Vp – oqıdı</i>               |

**Ko'makchi fe'llarni modellashtirish:** *Vaux* - edi, eken, emes, ediń, dese, desti, bolar, bolmas.

Har qanday so'z turkumining tarkibiga kiruvchi so'zlar, agar ular yasama so'z bo'lsa, morfemalardan iborat bo'lishi ma'lum. Morfemalar esa ikki xil bo'ladi:

I. Asos morfemalar – asos morfemalar alohida holda ham, o'ziga qo'shimcha morfema qo'shib ham qo'llanishi mumkin va alohida holda ma'no anglatadi, qo'shimcha morfemalarda esa bunday xususiyat yo'q.

II. Affiks morfemalar – alohida holda ma'no anglatmaydigan va alohida qo'llana olmaydigan, asos morfemalarga qo'shib, yangi ma'no hosil qilishda yoki shakl hosil qilishda qo'llaniladigan morfemalar.

Abjalova Manzura muallifligidagi “Tahrir va tahlil dasturlarining lingvistik modullari» monografiyada qo'shimchalarni IDlar orqali boshqarishning quyidagi namunasi berib o'tilgan. [Abjalova, 2019: 16-17]

Biz ham ishimizda qo'shimchalarni IDlar orqali boshqarishda natijaga erishish uchun ushbu namunadan foydalanish mumkin.

Shuningdek, fe'l so'z turkumini modellashtirishda asosga qo'shiluvchi affiksalar quyidagi jadvaldagidek belgilanadi.

Matnni tahlil qiluvchi dasturning algoritmini tuzishda har bir affiksga maxsus belgi (dastur affiksni tanib olishi uchun) qo'yish talab qilinadi:

| Ai  | Belgining izohi            | ID     | Lemma                      |
|-----|----------------------------|--------|----------------------------|
| A1  | oʻzlik nisbat affiksi      | ó_d_a  | <b>ism asosli shakllar</b> |
| A2  | birgalik nisbat affiksi    | sh_d_a |                            |
| A3  | oʻzlik nisbat affiksi      | óz_d_a |                            |
| A4  | belgisizlik nisbat affiksi | b_d_a  |                            |
| A5  | buyruq mayli affiksi       | b_m_a  |                            |
| A6  | istak mayli affiksi        | t_m_a  |                            |
| A7  | boʻlishsizlik affiksi      | b_a    |                            |
| A8  | koʻplik affiksi            | k_a    |                            |
| A9  | shaxs-son affiksi          | b_s_a  |                            |
| A10 | shart mayli affiksi        | sh_m_a |                            |
| A11 | maqsad mayli affiksi       | m_m_a  |                            |

Quyidagi jadvalda esa bu qoʻshimchalarning soʻz yasash namunalari berilgan:

| Asos                | Qoʻshimcha  | Yasama soʻz         |
|---------------------|-------------|---------------------|
| bas, bazar          | -la/-le     | basla, bazarla      |
| kók, taza, jańa     | -ar/-er, -r | kóger, tazar, jańar |
| tar, keń, kóp       | -ay/-ey, -y | taray, keńey, kóbey |
| ot, tún, oyin       | -a/-e       | ota, túne, oyna     |
| bas, es,            | -qar/-ker   | basqar, esker       |
| suw, ań             | -ǵar/-ger   | suwǵar, ańǵar       |
| qarıw, kewil        | -lan/-len   | qarıwlan, kewillen  |
| boz, tot            | -ıq/-ik     | boziq, totıq        |
| bala, júrek         | -sı/-si     | balasın, júreksin   |
| kóp, awır           | -sın/-sin   | kórsin, awırsin     |
| menmen              | -sıra/-sire | menmensire          |
| siyrek, kem         | -sit/-sit   | siyreksit, kemsit   |
| tumaw, quw          | -ra/-re     | tumawra, quwra      |
| gúrkirew, jiltıldaw | -ıra/-ire   | gúrkire, jiltıra    |

| Asos        | Qo‘shimcha | Yasama so‘z   |
|-------------|------------|---------------|
| búk, saba   | -la/-le    | búkle, sabala |
| qis, qımtaw | -ta/te     | qısta, qımta  |
| buraw, ótew | -a/-e      | bura, óte     |

Yuqorida keltirilgan misollar ishning tugallanganligini bildirmaydi. Bundan keyingi bosqichda fe‘l so‘z turkumini formallashtirish masalasi turadi.

**Xulosa.** Morfologik analizator bo‘yicha ishlarni olib borishda e‘tibor qaratish lozim bo‘lgan ko‘p masalalar mavjud. Birinchi navbatda tilning elektron lug‘at bazasini yaratish, qo‘shimchalarni alohida bazaga joylashtirish, tilning o‘ziga xosligiga, so‘z turkumlarining o‘ziga xos xususiyatlariga qarab ularni tartibli ketma-ketlikda formallashtirish modellarini algoritmlar doirasida tuzish lozim. Qoraqalpoq tilining morfoanalizatorini yaratish davr talabi, bunday tadqiqotlarning miqdor va sifat jihatdan ortib borishi tilning rivojlanishiga ham o‘zining ijobiy ta‘sirini o‘tkazadi.

#### ADABIYOT

1. Abduraxmanova N. Kompyuter lingvistikasi. Toshkent, “Nodirabegim” nashriyoti, 2021.
2. Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. Tashkent, 2019.
3. Xolmanova Z. Kompyuter lingvistikasi. Tashkent. 2019.
4. A.Dáwletov, M.Dáwletov, M.Qudaybergenov. Házirgi qaraqalpaq ádebiy tili (morfemika, morfonologiya, sóz jasalıw, morfologiya). Nókis. 2010.
5. Elov B., Hamroyeva Sh., Elova D. Morfologik analizatorni yaratish usullari. O‘zbekiston: til va madaniyat. Amaliy filologiya masalalari. Toshkent. 2022. 67-86-b. [www.researchgate.net](http://www.researchgate.net)

УДК 81'322.2

**ЛЕКСИКО-СЕМАНТИЧЕСКАЯ РЕАЛИЗАЦИЯ КОНЦЕПТА  
“КРАСОТА” В УЗБЕКСКИХ И АНГЛИЙСКИХ  
ЭЛЕКТРОННЫХ КОРПУСАХ**

**Т. Р. Яндашева**

*Национальный университет имени Мирзо Улугбека*

*Ташкент, Узбекистан*

yandashova92@gmail.com

В статье проведено сравнительное исследование узбекского и английского национальных корпусов понятия красоты и репрезентирующих его лексико-семантических средств. Благодаря этому в словаре на двух языках найдены эти лексемы, отражающие человека и его внутреннюю, внешнюю красоту, с целью определения места в лексикографии, дискурсе и поэтическом тексте.

**Ключевые слово:** красота, слово-ассоциация, лексический синоним, внутренняя красота, внешняя красота, национально-ментальное мировоззрение.

**LEXICO-SEMANTIC IMPLEMENTATION OF THE CONCEPT  
“BEAUTY” IN UZBEK AND ENGLISH ELECTRONIC CORPORA**

**Yandasheva Tursunoy**

yandashova92@gmail.com

**Annotation.** The article performed a comparative study in Uzbek and English national corpuses of beauty concept and lexical-semantic tools representing it. Through this, in two languages dictionary found of these lexemes reflecting man and its internal, external beauty is aimed at setting a place in a lexicography, discourse and poetic text.

**Keywords:** beauty, association word, lexic synonym, internal beauty, external beauty, national-mental worldview.

**O‘ZBEK VA INGLIZ TILLARI ELEKTRON KORPUSLARIDA  
GO‘ZALLIK KONSEPTINING LEKSIK-SEMANTIK VOQELANISHI**

**Yandashova Tursunoy Rustam qizi**

*Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti*

*Toshkent, O‘zbekiston*

yandashova92@gmail.com

**Annotatsiya.** Maqolada go‘zallik konsepti va uni ifodalovchi leksik-semantik vositalarining o‘zbek va ingliz tillari elektron korpuslaridagi qiyosiy tadqiqi amalga

oshirilgan. Bu orqali inson va uning ichki, tashqi go‘zalligini aks ettiruvchi mazkur leksemalarning ikki til elektron lug‘at fondida, spontan nutqda hamda badiiy matndagi o‘rmini belgilash maqsad qilingan.

**Kalit so‘zlar:** go‘zallik, assotsiativ so‘z, lug‘aviy ma‘nodoshlik, ichki go‘zallik, tashqi go‘zallik, milliy-mental dunyoqarash.

Go‘zallikning izohi sifatida “O‘zbek tilining izohli lug‘ati”da chiroy va husn so‘zlari keltiriladi. Ko‘pchilik shaxslarda ham go‘zallik deganda birinchi bo‘lib assotsiyalanadigan tushuncha sifatida “chiroy” va “chiroyli” tushunchalarini e‘tirof etish mumkin. Ikkinchi o‘rinda esa “tashqi ko‘rinish, shakl” leksik birliklari turadi.

Go‘zallik konsepti qaysi til aspekti doirasida voqe‘lanishidan qat‘iy nazar leksik-semantik, morfologik, sintaktik, frazeologik vositalar orqali ifodalanadi. Go‘zallik konseptining leksik-semantik tadqiqida xuddi shu konsept atrofida to‘planuvchi jamiki leksemalarning kelib chiqishi, lug‘aviy ma‘nosi bilan bog‘liq ma‘lumotlar to‘plab o‘rganiladi. Bu go‘zallik konsepti bilan bog‘liq assotsiativ til fondini yaratish uchun juda samarali tadqiqot bo‘lib xizmat qilishi mumkin. Chunonchi, birgina “go‘zal” so‘zining o‘zini leksik-semantik tahlilga tortsak, bu so‘z o‘zbek tilining besh jildli izohli lug‘atida “juda ham chiroyli, husndor, xushro‘y” kabi sinonimlar va “kishi ko‘zini quvontiradigan, kishini maftun etadigan, zavqlantiradigan”, “zavq-shavqqa boy, serzavq, unutilmas (davr, vaqt haqida)” kabi sifatlar bilan izohlangani, shuningdek, u muloqot jamiyatida xotin-qizlar ismi vazifasini bajarishi ham aytib o‘tilgan. Bunda e‘tibor berishimiz kerak bo‘lgan asosiy nuqta ham aynan shu – oxirgi izohdir. Nega aynan xotin-qizlar ismi? Chunki go‘zal so‘zi tilmizda hamisha ayollarga, tabiatga va ba‘zi bir voqea-hodisalarga nisbatan qo‘llaniluvchi leksema hisoblanadi. Va bu bevosita mazkur leksemaning gender ahamiyatga ega so‘z ekanligiga ishora qiladi. Fikrimizni asoslash uchun tilimizda mavjud “go‘zal” so‘zining sinonimlariga murojaat qilamiz. “Go‘zal” leksemasi o‘zbek tilining jonli muloqot sharoitida hech qachon erkaklarga nisbatan qo‘llanilmaydi. Yoinki, “nozli, ishvali” so‘zlari go‘zallik parametrlarida ayollarga nisbatan ijobiy semantik voqelikni gavdalantirsa, bil‘aks, erkaklarga nisbatan salbiy semantik bo‘yoqdorlik kasb etadi. Go‘zal so‘zining mutlaq gender ahamiyatga ega ekanligini “O‘zbek tilining etimologik lug‘ati”da keltirilgan quyidagi izoh ham yaqqol isbotlaydi. “Go‘zal - xushro‘y, husndor. Bu sifat asli qadimgi turkiy

<sup>1</sup> O‘zbek tilining izohli lug‘ati. – B.530.

tildagi “nazar sol” ma’nosini anglatgan ‘ko’s’ fe’liga –al qo’shim-chasi qo’shilishi bilan yasalgan. Keyinchalik so’z boshidagi k undoshi g undoshiga, unlilar oralig’idagi s undoshi z ga almashgan; o’zbek tilida o’ unlisining yumshoqlik belgisi yo’qolgan. Bu sifat asli “o’ziga e’tiborni tortadigan” ma’nosini anglatgan bo’lib, “xushro’y” ma’nosi shu ma’no asosida o’sib chiqqan<sup>1</sup>”.

Leksemaning etimologik ma’nosi ham mazkur sifatning ko’rish, his qilish ongli faoliyat turlari bilan bog’liq estetik xususiyat ekanligini anglatmoqda. Shu ikki lug’atning o’zidayoq “go’zal” so’zining ham gender (lingvokulturologik), ham estetik jihatga ega ikki eng muhim tomonini ta’kidlab ko’rsatish mumkin. Shu jihatdan, go’zallik konseptining assotsiativ til fondini leksik-semantik tadqiq etish qimmatli ahamiyatga ega. Bundan tashqari, go’zal so’zi semasiologik nuqtai-nazardan, tilimizda darajalanish xususiyatiga ega leksema hisoblanadi. Uning ma’no taraqqiyotida o’ziga yondosh leksemalar orasida sezilarli ma’no nozikliklari bilan ajralib turuvchi ba’zi jihatlarni ilg’ab olish qiyin emas. Chunonchi, “go’zal” so’zi “istarali” so’zi bilan qiyoslanganda, mazkur so’zning semantik tarkibi “go’zal” so’zi ifodalagan mukammallik etaloniga nisbatan biroz qashshoq, biroz quyiroq darajani ifodalagani bois, garchi ikkisi ham bir xil nutq sharoitida qo’llanilsa-da, har doim ham kontekstda teng munosabatni ifodalamaydi. Shu boisdan ham, tilda, ayniqsa, o’zbek tilida so’zning ma’no qirralari, leksik-semantik tadrijiy taraqqiyoti bilan bog’liq hosilalar muhim ahamiyatga ega hisoblanadi.

Inglizcha “beauty” so’zining etimologiyasiga nazar tashlaydigan bo’lsak, bu so’z dastlab lotincha “yaxshi, kelishgan” ma’nosida qo’llaniluvchi “bellus”, “bellitalem” sifat so’zidan “bealte”, “beaute” shaklida yevrofransuz tiliga o’zlashgan. 1325-yilga qadar mustaqil so’z sifatida iste’molda bo’lgan. Keyinchalik mazkur tildan ingliz tiliga o’zlashib, hozirgi “chiroy”, “go’zallik” ma’nosida qo’llana boshlagan. Har ikkala tilda ham bu so’zlarning sinonimik qatori juda uzun hisoblanadi. Chunonchi, o’zbek tilida go’zallik konseptini ifodalovchi leksik-semantik birliklarga quyidagilarni misol keltirishimiz mumkin: *go’zal, nozik, nafis, latif, yoqimli, chiroyli, sohibjamol, hurliqo, asal, shirin, shakar, jonon, ko’rkam, xushbichim, kelishgan, alpqomat, xushsurat, xushro’y, jozibali, maftunkor, latofatli, barkamol, mukammal, ko’hlik, ajoyib, suqsur, yaxshi, zebo, dono, xushqad, pahlavon, malo-*

<sup>1</sup> O’zbek tilining izohli etimologik lug’ati. – B.88.



*hatli, bokira, sharmli, hayoli, iboli, suluv, vafoli, tabiiy, betakror, sodda, haqiqiy, pok, mard, jo'mard, komil, soliha, oqila, aqlli, nazokatli, shirinso'z, shirinsuxan, mehribon, rahmdil, farosatli, fahmli, fasohatli, nozanin, diyonatli, zamonaviy, samimiy, ideal, istarali, nozli, ishvali, zehnli, sabrli, zukko...* Bu qatorni cheksiz davom ettirish mumkin. Boisi o'zbek milliy mentalitetida ichki go'zallik tashqi go'zallikdan yuqori turadi. Va bu tanlov o'zbek tilining lug'at fondida ham o'z aksini topgan.

Ushbu sifat so'zlarning yasalishi til rivoji jarayonida to'xtab qolmaydi. Chunonchi, *-li*, *xush*, *-kor* kabi unumdor sifat yasovchi qo'shimchalar yordamida yangi yangi so'zlarni yasash, iste'molga kiritish mumkin. masalan,

*-li* qo'shimchasi bilan yasaladigan go'zallik konsepti sifat so'zlari: *chiroyli, yoqimli, jozibali, latofatli, malohatli, sharmli, hayoli, iboli, vafoli, aqlli, nazokatli, fahmli, farosatli, diyonatli, muomalali, ishvali, nozli, vijdonli, istarali, ko'rkli, zehnli, sabrli...* va b. ot + li = sifat qolipi asosida yasaluvchi ushbu sifatlar muayyan mavhum yoki aniq tushuncha (predmet, xususiyatlarga nisbatan)ga egalikni, xoslikni bildiradi.

*-kor* qo'shimchasi bilan yasaladigan go'zallik konsepti sifat so'zlari: *maftunkor, jilokor, ishvakor, diyonatkor.*

*-dor* qo'shimchasi bilan yasaladigan go'zallik konsepti sifat so'zlari: *jozibador, vafodor, mazmundor.*

*Xush* old qo'shimchasi bilan yasaladigan go'zallik konsepti sifat so'zlari: *xushqad, xushro'y, xushqomat, xushbichim, xushsurat.* Umuman, tashqi va ichki go'zallikni ifodalovchi sifat so'zlar va sifat qo'shimchalar o'zbek tilida boy va rang-barangdir.

Endi ingliz lingvomadaniyatiga xos go'zallik konsepti sifatlarini tahlilga tortamiz. Oldingi boblarimizda aytilganidek, inglizlar tashqi go'zallikka ko'proq ahamiyat qaratishadi, va bu tanlov ham ularning til birliklarida yaqqol o'z aksini topgan. Ingliz tilida: *beautiful, smart, handsome, attractive, fine, good, nice, beauteous, cute, fair, good-looking, gorgeous, sheen, hot (slang), lovely, nice-looking, pretty, shapely, fit (slang), clear, pleasant, excellent, exceptional, great, marvelous, perfect, stylish, wonderful, sunny, alluring, dazzling, fascinating, graceful, magnificent, appealing, charming, delicate, delightful, elegant, exquisite, grand, pleasing, splendid, stunning, superb, well-formed, taking, symmetrical, sublime, statuesque, slightly, resplendent, refined, ravishing, radiant, pulchritudinous, ideal, foxy,*

*enticing, divine, comely, classy, bewitching, angelic, admirable*<sup>1</sup> etc. Bu soʻzlar qatorini yana davom ettirish mumkin. Chunki ingliz tili ham Shekspir ijodi, dunyo tamadduni markazi sifatida juda katta lugʻat fondiga ega. Ayniqsa, mazkur tilda yaratilgan tezaurus lugʻatlar bu tilni oʻrganish va oʻrgatishda katta ahamiyatga ega. Mazkur tilda ham yasali holatlari koʻp uchraydi. Birgina *beauty* soʻzidan *beautiful, beautifler, beautify, beautifully, beautifulness* kabi soʻzlar yasalgan. Oʻzbek tilidan farqli ravishda bu soʻzlar turli morfologik turkumga mansub. *Beautifully* – ravish, *beautify* – feʼl, *beautiful* – sifat, *beautifulness* – ot. Har ikki tilda mavjud mazkur leksemalar emotsional ekspressivligiga koʻra, nutq jarayonidagi ishtirokida bir-biridan farqlanishi mumkin. Masalan, *dilbar, goʻzal, barno, lobar; beauteous, gorgeous, sheen, marvelous, appealing* soʻzlari badiiy uslubga; *chiroyli, aqli, yaxshi, farosatli; good, nice, fine, smart, pretty* – neytral uslubga; *yoqimli, shirin, shakar, asal, sodda, tabiiy; cute, fair, hot, fit, attractive* soʻzlari soʻzlashuv uslubiga xos leksemalar hisoblanadi.

Bunday xoslangan soʻzlar milliy til madaniyatini ochishda kalit vazifasini bajarishi mumkin. Axloqshunoslikdagi “yaxshi odam” tushunchasi hammaga – ayolga ham, erkakka ham, yosh-u qariga ham tegishli boʻlishi mumkin. Estetikada esa “goʻzal odam” tushunchasi yoʻq; yo “goʻzal yigit”, yo “goʻzal qiz” degan tushunchalargina mavjud. Chunki, erkak kishidagi chiroyli moʻylov faqat erkakning yuzida, ayol kishidagi husnlardan biri – uzun soch faqat ayol kishi vujudida goʻzallikka ega. Endi moʻylov burab soʻzlayotgan ayolni-yu, soch-popuk taqib yurgan erkakni tasavvur qiling! Boyagi goʻzalliklar xunuklikka aylanadi-qoladi. Shuningdek, goʻzallik bir vujudda ham faqat oʻz oʻrnini talab qiladigan “oʻta injiqlik” xususiyatiga ega. Shu joyda olmon nafosatshunosi Fexner qoʻllagan misolni keltirish oʻrinlidir. Mutafakkirlarning fikricha, qiz bola yuzidagi qizillik uning goʻzalligidan dalolat beradi. Biroq, qizillik uning burni ustiga koʻchsa – xunuklikka aylanadi. Demak, axloq uchun – umumiylik, nafosat uchun esa – muayyanlik mavjudlik sharti hisoblanadi.

Bundan tashqari goʻzallik bilan bogʻliq har bir leksemaning “goʻzal” soʻzi bilan muayyan darajada bogʻliqliklari boʻlgani kabi, baʼzi semantik farqlari ham mavjud. Bu, ayniqsa, kontekstda yorqinroq koʻrinadi. Yuqorida biz ularning uslubiy maʼno nozikliklariga eʼtibor qaratgandik. Quyida esa “beautiful” soʻzining sinonimlari orasidagi baʼzi bir maʼno qirralariga eʼtiborimizni qaratamiz. merriam-webster.

<sup>1</sup> <https://www.thesaurus.com/browse/beautiful>

com saytida bu haqida quyidagi ma'lumotlarga duch keldik: "Some common synonyms of *beautiful* are *comely*, *fair*, *handsome*, *lovely*, and *pretty*. While all these words mean "exciting sensuous or aesthetic pleasure", *beautiful* applies to whatever excites the keenest of pleasure to the senses and stirs emotion through the senses. For example: *beautiful mountain scenery*<sup>1</sup>". Ingliz tili so'zlashuv nutqida "*beautiful*" so'zining *comely*, *fair*, *handsome*, *lovely* va *pretty* kabi sinonimlari eng ko'p qo'llanilishga ega.

*Comely* ajoyib degan ma'noda, kishiga beixtiyor o'zgacha emotional his uyg'ota oladigan xatti-harakat yoki holatga nisbatan ishlatiladi. Masalan, raqqosning ajoyib harakati. *Fair* go'zallikning eng ko'p qadrlanuvchi sifati: soflik, beg'uborlik, poklikka nisbatan qo'llaniladi. Kishilardagi yuz go'zalligini ta'riflashda eng ko'p asqotadi, beg'ubor, tiniq yuz ma'nosida. Proporsion, simmetrik mutanosiblik inglizlarda *handsome* so'zi bilan ifodalanadi. Bizda bu so'zning eng muqobil variantlari: *kelishgan*, *xushqomat* kabi so'zlardir. *Pretty* va *beautiful* so'zlari ko'pincha kontekstda bir xil ma'noda keladi. *Pretty* asosan tashqi go'zallikka nisbatan faolroq ishlatilsa, *beautiful* ham tashqi, ham ichki go'zallikni ifodalash uchun birday munosib so'z sanaladi.

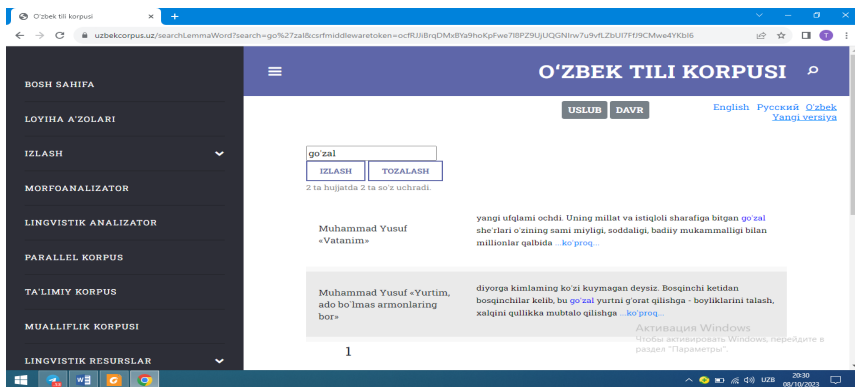
Butun dunyo xalqlari lingvomadaniyatining go'zallik tavsifida eng avval ayol siymosi gavdalanishi tabiiy hol. Ayniqsa, Sharq xalqlari adabiyotida ayol go'zalligi, latofati uning sharm-u hayosi, odob-axloqi bilan nihoyatda katta qadr-qimmatga ega bo'lgan. Bu haqda avvalgi boblarimizda keltirilgan tajribamiz tavsifida ham to'xtalgan edik. O'zbek milliy mentalitetida go'zallik konsepti yadro maydonida go'zal, go'zallik, tashqi go'zallik, ichki go'zallik, chiroyli, betakror, kelishgan, ayol, tabiat tushunchalari jonlansa, periferik maydonda vafo, sadoqat, farosat kabi ichki go'zallik asoslari aks etadi. Bu haqda mashhur o'zbek shoiri A.Oripovning "Ayol" she'ridan olingan quyidagi misralarda Sharq xalqlari idealidagi go'zal ayol qiyofasi juda ta'sirchan ravishda ochib berilgan:

...*Shu cho'lpon ko'zlarning buyuk hurmati,*  
*Shu aqiq lablarning rost so'zi deya,*  
*So'ylangchi, vafoning nadir qimmati –*  
*Siz ham kutganmisiz biror soniya?!*

She'rda keltirilgan *cho'lpon ko'z*, *aqiq lab*, *olovli nafas*, *parishon sochlar*, *vafo shevasi* kabi o'xshatishlar ayol go'zalligining mukammal tavsifi desak, mubolag'aga yo'l qo'ymagan bo'lardik.

<sup>1</sup> <https://www.merriam-webster.com/thesaurus/beautiful>

Biz go‘zallik konseptini elektron korpuslar bo‘yicha qiyosini amalga oshirganda o‘zbek tili korpusidagi manbalar juda oz ekanligiga guvoh bo‘ldik. Har holda bu borada qilinajak ishlarimiz ko‘p ekan. DSc N.Abdurahmonova rahbarligidagi O‘zbek tili korpusi<sup>1</sup> manbalarida “go‘zal” so‘zi bilan bog‘liq quyidagi ma’lumotlarga duch keldik.



Saytning yangi versiyasida esa “go‘zal” so‘zi bilan bog‘liq 9 ta matn turi ochib berildi. Sayt ancha to‘ldirilgan<sup>2</sup>. Boyitilgan. Korpus test rejimida ishlaganligi bo‘smi, tezaurus lug‘at bizga go‘zallik konsepti bilan bog‘liq so‘zlarni topib berolmadi. Loyihaning yanada kengayib mukammallashishiga umid bog‘laymiz. Ingliz tili tezaurus lug‘ati birgina Marriam webster tezaurusi misolida ochib berildi. Shundan kelib chiqib ham xulosa qilishimiz mumkinki, o‘zbek tili elektron korpuslari, tezaurus lug‘atlari ko‘payishiga va ma’lumotlar hajmining kengayishiga juda katta ehtiyojimiz bor.

## ADABIYOT

1. Ощепкова В. В. Культурологические, этнографические и типологические аспекты лингострановедения: Автореф. ... док. филол. наук. – М., 1995.

2. Раҳматуллаев Ш. Ўзбек тилининг этимологик луғати. – Тошкент: Университет, 2009. – 284 б.

<sup>1</sup> <https://uzbekcorpus.uz/search>

<sup>2</sup> <https://uzbekcorpus.uz/searchLemmaNew?noun=None&verb=None&adj=None&numeric=None&pron=None&adv=None&search=go%27zal&page=1>

3. Рустамов Д. Лексемалар миллий-маданий хосланган семемасининг лингвомаданий тадқиқи: Филол. фан. бўй. фалс. фан. докт. дисс. автореф. – Фарғона, 2018. – 49 б.
4. Сафаров Ш., Боймирзаева С. Гендер тилшунослиги ва матн тадқиқи // Хорижий филология, 2006. №4. – Б. 33.
5. Юсупов Ў.Қ. Маъно, тушунча, концепт ва лингвокультурега атамалари хусусида // Стилистика тилшуносликнинг замонавий йўналишларида: Илмий амалий конференция материаллари. – Тошкент, 2011. – Б.49.
6. Cambridge Advanced Learner's Dictionary Electronic resource. – Cambridge University Press, 2004. <http://www.dictionary.cambridge.org> (24.06.2005).
7. Hornby A.S. Oxford Advanced Learners Dictionary of Current English, Oxford University Press, 1974. – P. 176.
8. Ўзбек тилининг изоҳли луғати. 5 жилдлик. – Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти. 2006–2008.
9. O‘zbek tilining izohli etimologik lug‘ati. – B.88.
10. <https://uzbekcorpus.uz/>
11. <https://www.thesaurus.com/browse/beautiful>
12. <https://www.merriam-webster.com/thesaurus/beautiful>

УДК. 809:494.3

## КОМБИНАТОРНЫЕ СВОЙСТВА ЛЕКСИЧЕСКИХ ЕДИНИЦ

**Б. А. Юнусова**

*Самаркандский государственный университет  
имени Шарофа Рашидова  
Самарканд, Узбекистан  
yunuovabakhora@gmail.com*

В статье рассматривается и анализируется отличие слова от лексемы, их сходство, виды лексического сочетания в процессе употребления слова в контексте текста, комбинаторная лексикология.

**Ключевые слова:** лексема, синтагматика, лексическая единица, парадигма, лексическое сочетание, комбинатор.

## COMBINATORIAL PROPERTIES OF LEXICAL UNITS

***Yunusova Bahora Akhtamzhonovna***

*Sharof Rashidov Samarkand State University,  
Samarkand, Uzbekistan  
yunuovabakhora@gmail.com*

The article examines and analyzes the difference between a word and a lexeme, their similarity, types of lexical combinations in the process of using a word in the context of a text, combinatorial lexicology.

**Keywords:** lexeme, syntagmatics, lexical unit, paradigm, lexical combination, combinator.

Известно, что все слова, существующие в языке, называются словарным составом или лексикой. При этом изучаются проблема слова, являющегося основной единицей языка, построение словарного состава, применение, обогащение, развитие словарных единиц и другие аспекты. Несмотря на многолетнее изучение слова и связанных с ним явлений языка и речи, в настоящее время оно остается основным источником исследования в языкознании. Основной причиной этого является постоянное изменение и обновление слова и связанных с ним явлений и тот факт, что слово, связанные с ним понятия занимают важное место в качестве средства общения в обществе. Поэтому проблема слова является основным источником изучения лексикологии. Мы постараемся осветить тему, опираясь на суждения ученых о различии, общих чертах слова и лексемы. Лексема реализуется в речи в слове.

Наряду с тем, что лексема является готовой, общей и обязательной для всех членов общества, она обладает также следующими иными свойствами: 1. Член общества не создает лексему, принимает ее в готовом виде. 2. В сознании члена общества лексема «живет» в одном ряду со схожими лексемами (в парадигмах). Например: [daftar] ~ [bloknot]; [daftar] ~ [oynoma] ~ [ro'znoma]; [daftar] ~ [qissa] ~ [roman]; [daftar] ~ [miqova] ~ [varaq] ~ [bet] ~ [bob] (*тетрадь* ~ *блокнот*; *тетрадь* ~ *журнал* ~ *газета*; *тетрадь* ~ *повесть* ~ *роман*; *тетрадь* ~ *обложка* ~ *лист* ~ *страница* ~ *глава*). Слово *daftar* на основе этих отношений имеет несколько смыслов. 3. Лексемы в сознании человека «живут» также в соседских (синтагматических) отношениях. Например: [*тетрадь*] ~ [*пиши*] ~ [*возьми*] ~ [*качественный*] ~ [*математика*] ~ [*родной язык*]; [*тетрадь*] ~ [*числовые дополнения*] ~ [*притяжательные аффиксы*] ~ [*надежные дополнения*]... Эти сходственные и соседские отношения, возможности смысла и задач проясняются, уточняются в речи. Следовательно, лексемы являются также совокупностью речевых возможностей, реализованных и реализуемых в сознании носителей языка [1]. В процессе применения слова в окружении текста существует 2 вида лексических комбинации: *внутренняя комбинация* и *внешняя комбинация*. *Внутренняя комбинация* – имеет целостный смысл, состоит из стабильных отношений двух и более слов до процесса речи, привносится в речь в готовом виде, образуя переносный смысл, реализуется посредством фразеологической или лексической единицы. При этом ярко проявляются, в основном, в *описательном выражении* (*перифразах*), *фразах* (*фразеологизмах*), *паремнологических единицах* (*поговорах* и *поговорках*), *мудрых словах* (*афоризмах*).

Смысл слова «подняться» от слова «высокий» скрыт в значении «отличаться, побеждать», «превзойти друг друга, превзойти друг друга, не опуститься ниже». Он пил воду из высокого корыта. Прийти с высоты 1) высокомерно говорить, высокомерно поступать; 2) установить большую, высокую цену, зависить цену. Не поднимайся с такой высоты, спускайся. Хочешь продать, возмем (в Торге). Его нос (или клюв) высокий. димог. Высокомерный Очень самоуверенный, высокомерный. Рука высоко 1) повезло, крупный бизнес, повезло. Не забывай, моя дорогая, мы будем рады, если в этом году получим хороший урожай. Ш. Рашидов, Сильнее бури; 2) победитель, победитель. Но в то время, когда сугдийцы были в приподнятом настроении, я получил дополнительную поддержку из Мароканда в размере одного округа.

Из приведенных примеров видно, что через взаимодействие действия и его результата, отношение действия и его исполнителя, взаимодействие материала и сделанной из него вещи в словах может обретаться новый смысл. выражать вещи. Чтобы назвать вещи и предметы в человеческом существовании, необходимо выявить их важные признаки, знать изменение отношения к этим вещам и предметам в повседневной жизни, а также понимать, что одно слово может сочетаться с другим словом или сочетанием. То есть, анализируя качественные свойства предмета или события, используя его для выделения и описания важного признака, воспринимать, замечать, воспринимать, понимать, знать и в уме воспринимать набор признаков, принятый группой говорящих. должен уметь воплотить в жизнь описываемую в его воображении вещь или событие.

Поэтому говорящий, опираясь на свои знания, основанные на языковых и жизненных обобщениях, замечает в характеристиках определенной вещи или человека некоторые общие черты между другой вещью или человеком, характер связи между ними, т. е. путем их соединения находят общий доминирующий характер и назвать его на своем языке. В результате у названия первой вещи (первичного референта) появляется новый смысл и на его основе появляется новое имя у второй вещи (вторичный референт). При создании нового имени говорящий должен обладать высоким уровнем мышления, то есть человеческий разум должен быть способен выносить суждения и выводы. Потому что, если слушатель или читатель не сможет вынести суждение о наличии сходства между признаками предмета, если он не увидит общности признаков, он не поймет смысла. В этом процессе важно, чтобы слова говорящего выступали в разных значениях в потоке речи, сочетании слов и положении слова в возникновении таких значений.

В именовании выделяют три аспекта: именуемый объект, субъект именования и элементы выбираемого языка. Объектом для наименования может быть отдельное понятие, предмет, знак (красота, книга, скажем, зелень), предмет с конкретными признаками (зеленое дерево) или целое событие (Весна! Птицы полетели). Содержание символа, выбранного в процессе именовании в качестве основы для наименования, является основой для образования внутренней комбинации. Итак, один и тот же объект может называться по-разному в зависимости от его разных знаков. В общий словарный запас обычно включаются имена, соот-



ветствующие законам внутреннего развития языка и способные удовлетворить потребности представителей того или иного языка.

*Внешняя комбинация* – реализуется посредством лексических единиц, образованных на основе переноса в процессе речи в качестве имени слов в прямом смысле и означающих прямой смысл из предметов и явлений в другие предметы и явления. Это ярко проявляется в *словосочетаниях и метафорах*. Паремиологические единицы и афоризмы, составляющие структуру, рассматриваются как объект изучения литературы в целом, а поскольку содержание и цель обучения в этих единицах занимают первостепенное место, то давая определения и пояснения лингвистически и принося их в классификации как лингвистическое явление сбивает с толку и может привести к ложным выводам [Абузалова, 2008: 23-31].

Комбинаторная лексикология изучает взаимодействие слов в потоке языка и речи [Кадирова, 2022: 13]. Комбинаторная лексикология – это оптимальная интерпретация сочетания лексических единиц, которая осуществляется в словарях. Разработка двух язычных словарей осуществляется посредством изучения комбинаторных свойств уровней языка. Комбинаторное языкознание является особой областью узбекского языкознания, изучающее отношения между различными единицами языка в качестве системы признаков языка. Комбинаторное языкознание в соответствии с предметом изучения и в качестве самостоятельного направления охватывает масштабные проблемы соответствия (сочетания) единиц языка. С этой точки зрения возникает проблема существования определенного метаязыка и возможность описания с его помощью. Результат исследования показывает, что в XX веке возникло множество понятий соответствия, которые используются в качестве синонимов. Приведение к определенной норме этих понятий в терминологической системе и выбор приемлемого варианта зависит от развития комбинаторного языкознания.

#### ЛИТЕРАТУРА

1. Влавацкая М.В. Комбинаторная лингвистика в структуре науке о языке // Вестник Ленинградского государственного университета имени А.С.Пушкина. – Ленинград, 2010. – С. 23–31.
2. Абузалова М., Назарова С. Систем тилшунослик асослари. – Бухоро, 2008. – Б. 20.
3. Кадирова З.З. Алишер Навоийнинг насрий асарларида перифразлар: Фил. фан. бўйича фалс. докт. (PhD). – Термиз, 2022. – Б. 13.

УДК 81'322.2

**МОРФОЛОГИЧЕСКИЕ СОЧЕТАНИЯ ЛЕММ  
ДЛЯ ЯЗЫКОВЫХ КОРПУСОВ И ИХ ПРАКТИЧЕСКИЕ  
РЕЗУЛЬТАТЫ**

*С. А. Каримов, С. М. Умирова, Б. Ф. Холмухамедов,  
Дж. У. Тиркашев*

*Самаркандский государственный университет  
Самарканд, Узбекистан*

*suyun1950@rambler.ru, umirova.s.m06@mail.ru,  
bxolmuxamedov@mail.ru, jtirkashev@mail.ru*

Одной из основных особенностей корпусов национального языка является морфологическая разметка слов, этапы ее реализации поясняются на примере существительных. На основе примеров анализируется сочетание имен-лемм с различными суффиксами.

**Ключевые слова:** язык, языковой корпус, лемма, морфологическая классификация, существительное, суффикс множественного числа, притяжательный суффикс, суффикс согласия.

**MORPHOLOGICAL COMBINATIONS OF LEMMAS FOR  
LANGUAGE CORPS AND THEIR  
PRACTICAL RESULTS**

*Suyun Karimov, Svetlana Umirova,  
Bakhtier Kholmukhamedov, Zhasurzhon Tirkashev*

*Samarkand State University  
Samarkand, Uzbekistan*

*suyun1950@rambler.ru, umirova.s.m06@mail.ru,  
bxolmuxamedov@mail.ru, jtirkashev@mail.ru*

One of the main features of national language corpora is the morphological marking of words, the stages of its implementation are explained using the example of nouns. Based on examples, the combination of lemma names with various suffixes is analyzed.

**Key words:** language, language corpus, lemma, morphological classification, noun, plural suffix, possessive suffix, agreement suffix.

TIL KORPUSLARI UCHUN OT-LEMMALARNING  
MORFOLOGIK KOMBINATSIYALARI VA AMALIYOTDAGI  
NATIJALARI

*Karimov S. A., Umirova S. M.,  
Xolmuxamedov B. F., Tirkashev J. U.,  
Samarqand davlat universiteti  
Samarqand, Uzbekistan*

suyun1950@rambler.ru, umirova.s.m06@mail.ru,  
bxolmuxamedov@mail.ru, jtirkashev@mail.ru

**Annotatsiya.** Milliy til korpuslarining asosiy xususiyatlaridan biri soʻzlarni morfologik teglash boʻlib, uni amalga oshirish uchun bajariladigan amallar ot soʻz turkumi misolida tushuntirilgan. Ot-lemmalarning turli qoʻshimchalarni olib kombinatsiyalashuvi misollar asosida tahlil etilgan.

**Kalit soʻzlar:** til, til korpuslari, lemma, morfologik razmetka, ot, koʻplik qoʻshimchasi, egalik qoʻshimchasi, kelishik qoʻshimchasi.

Dunyo tillari korpuslarini oʻrganish jarayonida shu narsa maʼlum boʻldiki, ularning barchasida til materiallari morfologik tomondan tavsiflangan. Toʻgʻri, tillarning morfologik tavsifida birxillik kuzatilmaydi, chunki tillarning geneologik va morfologik xususiyatlari bunga yoʻl qoʻymaydi. Sintaktik tomondan esa turlicha yondashuvlarni koʻrish mumkin. Korpusga kiritilgan soʻz va soʻz-shakllarning har biriga morfologik xususiyatlarni biriktirish morfologik teglash hisoblanadi hamda bu jarayon dastlab asosan qoʻl mehnati orqali amalga oshirilgani bois koʻp vaqt va mehnat talab qiladi. Shu sababli soʻzning morfologik xususiyatlarini avtomatik ravishda aniqlaydigan maxsus morfologik analizator dasturlar yaratilishi lozim. Soʻzlarning grammatik xususiyatlarini avtomatik tarzda belgilash natijalarida tabiiy tildagi matnlarni tahlil qilish, turli lugʻatlar tuzish, mashinali tarjima tizimlarini yaratish va tabiiy tillarni qayta ishlash vazifalari tez va oson bajariladi. Oʻzbek tili milliy korpusi ishlab chiqilar ekan, u uchun ham tabiiy tilni turli tomonlama qayta ishlash, soʻzlarning morfologik xususiyatlarini avtomatik biriktirish uchun maxsus dasturlar yaratish, komputer yordamida tilni tadqiq etish bugunning dolzarb vazifalaridan sanaladi.

Jahon tilshunosligida korpus lingvistikasi haqidagi ilk fikrlar R.G.Piatrovskiy tomonidan aytilgan [Кутузов]. Bu sohadagi asosiy tadqiqotlar 40-yillarda Blumfeld, Frays va Bondjerslar tomonidan amalga oshirilgan [Блумфилд, 1968; Fries, 1969; Bongers, 1947]. N.Frensis va G.Kuchera tomonidan korpus tuzish tamoyillari ishlab chiqilgan [Френсис, 1967]. Dastlabki korpusni esa ingliz tili misolida J.Sin-

kler yaratgan [Синклер]. Rus tilshunosligida V.P.Zaxarov, A.B.Kutuzov [Кутузов], E.V.Nedoshivina [Недошвина, 2006], V.V.Rikov, V.Plungyan, O.V.Kukushkina, A.A.Polikarpov, E.V.Surovsevalar tadqiqotlari muhim sanaladi. Ular korpus, korpus lingvistikasi, uning turlari, o'ziga xos xususiyati, ahamiyati, korpus tuzish tamoyillari borasidagi ishlarni amalga oshirishgan.

O'zbek tilshunosligida A.Po'latov, S.Muhamedov, M.Ayimbetov, S.Muhamedova, S.Karimov, G.Jumanazarova, A.Babanarov, D.O'rinboyevalar kompyuter dasturlari yordamida matnga leksikografik, lingvostastistik ishlov berishgan, zamonaviy kompyuter usullarini tavsiya etishgan, ammo korpus lingvistikasi sohasi bilan shug'ullanishmagan. Korpus lingvistikasi sohasida Sh.Xamroyevaning "O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari" [Хамроева, 2018] nomli tadqiqotini dastlabki ishlardan biri sifatida baholash mumkin. Keyinchalik bu sohaga oid ilmiy tadqiqotlar, ilmiy maqola va tezislari yaratildi. Jumladan, B.Mengliyev [Mengliyev, 2020; Менглиев], G.Toirova [Тоирова, 2020], A.Eshmo'minov [Эшмуминов, 2019], D.Ahmedova [Ахмедова, 2020], N.Ataboyevlarning [Ataboyev, 2019] ilmiy izlanishlari natijalarini misol sifatida keltirish mumkin.

Morfologik teglash jarayoni turli xil algoritmlar, jumladan, qoidalarga asoslangan usullar va neyron tarmoqqa asoslangan modellar yordamida amalga oshirilishi mumkin. Qoidalarga asoslangan ma'lum bir so'zning matnga qarab qanday teganishini boshqaradi, oldindan belgilangan qoidalar to'plamiga asoslanadi. Boshqa tomondan modellar ma'lum bir so'z uchun eng ehtimoliy tegni avtomatik ravishda biriktiradi.

Morfologik teglash ko'plab tabiiy til ilovalarida, jumladan, mashina tarjimasida, matndan nutqqa sintez va ma'lumotlarni qidirishda qo'llaniladi. Darhaqiqat, bu tabiiy tilni qayta ishlashning eng asosiy vazifalaridan biridir, chunki u sintaktik tahlil va semantik tahlil kabi murakkabroq tilni qayta ishlash vazifalari uchun asos yaratadi.

Morfologik teglash kompyuterlashtirilgan tilni qayta ishlashning muhim tarkibiy qismi bo'lib, tabiiy til ilovalarida tobora ko'proq foydalanilmoqda. Belgilash jarayonining to'g'riligi ishlatiladigan resurslarning sifatiga, shuningdek, qo'llaniladigan algoritmlarga bog'liq. Bu soha doimiy ravishda rivojlanib bormoqda, morfologik teglashning samaradorligini oshirish uchun yangi usullar ishlab chiqilmoqda.

Biz ushbu maqolamizda A.Qahorning "Bemor" hikoyasi matnidagi ot so'z turkumiga oid so'zlarni avtomatik aniqlashga yordam beradigan kombinatsiyalar, ularning bosqichlari va natijalari xususida fikr yuritimiz. Amaliyot uzbekcorpora.uz saytida sinovdan o'tkazilgan.

Til o'rganish metodikasidan bizga ma'lumki, dastlab ot so'z turkumi va uning grammatik xususiyatlari o'rgatiladi. Shunday ekan, otga qo'shiladigan lug'aviy va sintaktik qo'shimchalarning qo'shilish tartibi va ular anglatgan ma'nolar muhimdir. Jumladan, otlarning son kategoriyasi uning asosiy ma'nolaridan biri sanaladi.

*Ot-lemmalarga son shaklini qo'shish bosqichi.* Otlarda birlik va ko'plik son shakli mavjud, birlik shaklining morfologik ko'rsatkichi mavjud emas. Ko'plikning morfologik belgisi -lar qo'shimchasi bo'lib, u har doim ko'plik ma'nosini ifodalamaydi, balki tildagi boshqa uslubiy vazifalarni ham bajaradi.

O'zbek tilida yaratilgan lug'atlar asosida to'plangan ot so'z turkumi bazasiga o'ttiz besh mingdan ortiq lemma jamlandi. Bu ot-lemmalar asosan turdosh ot bo'lib, atoqli otlar bazaga kiritilmadi. Ot-lemmalarga turli morfologik qo'shimchalar qo'shib, 72 xil kombinatsiyaga kiritish natijasida ularning soni 2.5 milliondan oshdi. Dasturda otlarning son kategoriyasi birlik [birl.s.] va ko'plik [ko'p.s.] teglari ostida razmetkalab chiqildi. Ularning dasturiy kodi esa birlik [a] va ko'plik [b] belgisi ostida yozildi.

*Ot-lemmalarga egalik shakllarini qo'shish bosqichi.* Ismlardagi egalik shakllari grammatikaning muhim qismidir. Ular narsa yoki obyektga egalik yoki egalik huquqini ko'rsatish uchun ishlatiladi. Ismlarda egalik shakllarini shakllantirishning mos usulini bilish juda muhim, chunki bu har qanday yozma hujjat yoki og'zaki muloqotning ravshanligi va to'g'riligiga sezilarli ta'sir ko'rsatadi. Egalik shakllaridagi noaniqliklar tushunmovchiliklarga, chalkashliklarga va oxir-oqibatda yomon muloqotga olib kelishi mumkin.

Otlarda egalik shakllarining to'g'ri shakllanishi o'zbek tilida ham samarali muloqot qilishda muhim ahamiyatga ega. Har qanday noaniqlik yoki noto'g'ri talqinni oldini olish uchun qoidalarni to'g'ri bilish va qo'llash juda muhimdir. Otlardagi egalik shakllari qoidalarini o'zlashtirib, aniq va ravshan lug'atini tuzish lozim.

Unli bilan tugovchi otlarga undosh bilan boshlanuvchi egalik qo'shimchalari, undosh bilan tugovchi otlarga unli bilan boshlanuvchi egalik qo'shimchalari qo'shilishi barchamizga ma'lum. Shuning uchun ot-lemmalarning unli bilan tugaganlari ajratib olindi va ular o'n bir mingdan ortiq lemmalarni tashkil qildi. Undosh bilan boshlanadigan qo'shimchalar alohida kodlashtirildi, o'rtasiga undosh tovush orttirib yoziladiganlari alohida kodlashtirildi.

Undosh bilan tugagan ot-lemmalar 23 mingdan ortiq bo'lib ularning hammasi ham egalik qo'shimchasi qo'shilganda asos shaklini to'liq saqlab qolmaganligi sababli tovush tushadiganlari,

orttiriladiganlari va almashadiganlari ajratib olindi. Jumladan, *q* va *k* bilan tugagan lemmalarga egalik qo‘shimchalari qo‘shilganda hosil bo‘ladigan tovush almashinish hisobga olinib,  $q \rightarrow g$  va  $k \rightarrow g$  harflariga almashtirib chiqildi.

Shuningdek, birinchi shaxs ko‘plikdagi egalik qo‘shimchasi biz kishilik olmoshi bilan bog‘langanda belgisiz qo‘llanilishi [*bizning uy(imiz)*], [*bizning maktab(imiz)*]; *-(lar)i*, *-(s)i* egalik qo‘shimchalari ayrim ravishlar tarkibida yaxlitlanib qolganligi [*kechasi*], [*kunduzi*], [*kechalari*]; viloyat, shahar, tuman, korxonona, muassasa nomlariga qo‘shilgan egalik qo‘shimchasi egalikni emas, xoslik, umumdan ajratilganlik ma‘nolarini bildirishi [*Orol dengizi (umumdan ajratilganlik)*], [*o‘qish kitobi (xoslik)*], [*Toshkent shahri (umumdan ajratilganlik)*] ham inobatga olindi.

Egalik qo‘shimchalari birlikdagi ot-lemmalarga (35000 ta lemma + *-(i)m*, *-(i)miz*, *-(i)ng*, *-(i)ngiz*, *-i*, *-si* = *kitobim*, *kitobimiz*, *kitobing*, *kitobingiz*, *kitobi*), va ko‘plikdagi ot-lemmalarga (35000 ta lemma + *-lar* = *-(i)m*, *-(i)miz*, *-(i)ng*, *-(i)ngiz*, *-i*, *kitoblarim*, *kitoblarimiz*, *kitoblaring*, *kitoblaringiz*, *kitoblari*) kombinatsiyalangan holda qo‘shib chiqildi. Shundan so‘ng ot-lemmalar bazasi uch yuz ellik mingtani tashkil qildi.

Dasturda otlarning egalik kategoriyasidan I shaxs, birlik [I\_sh.b., II shaxs, birlik [II\_sh.b., III shaxs, birlik [III\_sh.b.], I shaxs, ko‘plik [I\_sh.k.], II shaxs, ko‘plik [II\_sh.k.] va III shaxs, ko‘plik [III\_sh.k.] teglari ostida razmetkalab chiqildi. Ularning dasturiy kodi esa I shaxs, birlik [a], II shaxs, birlik [b], III shaxs, birlik [c], I shaxs, ko‘plik [d], II shaxs, ko‘plik [e] va III shaxs, ko‘plik [f] belgisi ostida yozildi.

*Ot-lemmalarga kelishik shakllarini qo‘shilish bosqichi.* Otlarning kelishik kategoriyasi o‘zbek tilida grammatikaning asosiy elementlaridan biridir. Bu otning gapdagi grammatik rolini aks ettirish uchun shaklini o‘zgartirish jarayonidir. Kelishik kategoriyasi tilning muhim jihattir, chunki u ma‘noni yetkazish va sintaksisni to‘g‘ri shakllantirishga yordam beradi.

Ot-lemmalarga kelishik shakllari qo‘shilishida bo‘ladigan tovush o‘zgarishlari ham inobatga olindi va ular ham alohida kodlashtirildi.

Kelishiklar otlarning lemmasiga (35000 ta lemma + **-ni**, **-ning**, **-ga**, **-da**, **-dan** = *kitobni*, *kitobning*, *kitobga*, *kitobda*, *kitobdan*), ko‘plik shaklidan so‘ng (35000 ta lemma + *-lar* + **-ni**, **-ning**, **-ga**, **-da**, **-dan** = *kitoblarni*, *kitoblarning*, *kitoblarga*, *kitoblarda*, *kitoblardan*), otlarning yuqorida aytilgan olti xil egalik qo‘shimchasini olgan shakllaridan so‘ng (35000 ta lemma + *-(i)m*, *-(i)ng*, *-i*, *-si*, *-(i)miz*, *-(i)ngiz* + **-ni**, **-ning**, **-ga**, **-da**, **-dan** = *kitobimni*, *kitobimning*, *kitobimga*, *kitobimda*,

kitobimdan, kitobimizni, kitobimizning, kitobimizga, kitobimizda, kitobimizdan, kitobingni, kitobingning, kitobingga, kitobingda, kitobingdan, kitobingizni, kitobingizning, kitobingizga, kitobingizda, kitobingizdan, kitobini, kitobining, kitobiga, kitobida, kitobidan), ko‘plik qo‘shimchasini olib, olti xil egalik qo‘shimchasini olgan shakllaridan so‘ng (35000 ta lemma + *-lar* + *-(i)m, -(i)ng, -i, -si, -(i)miz, -(i)ngiz* + **-ni, -ning, -ga, -da, -dan** = kitoblarimni, kitoblarimning, kitoblarimga, kitoblarimda, kitoblarimdan, kitoblarimizni, kitoblarimizning, kitoblarimizga, kitoblarimizda, kitoblarimizdan, kitoblaringni, kitoblaringning, kitoblaringga, kitoblaringda, kitoblaringdan, kitoblaringizni, kitoblaringizning, kitoblaringizga, kitoblaringizda, kitoblaringizdan, kitoblarini, kitoblarining, kitoblariga, kitoblarida, kitoblaridan) kombinatsiyalangan holda qo‘shib chiqildi.

Dasturda otlarning kelishik kategoriyasidan bosh kelishik [b.k.], qaratqich kelishik [qar.k.], tushum kelishigi [tush.k.], jo‘nalish kelishigi [j.k.], o‘rin-payt kelishigi [o‘.p.k.] va chiqish kelishigi [ch.k.] teglari ostida razmetkalab chiqildi. Ularning dasturiy kodi esa bosh kelishik [a], qaratqich kelishik [b], tushum kelishigi [c], jo‘nalish kelishigi [d], o‘rin-payt kelishigi [e] va chiqish kelishigi [f] belgisi ostida yozildi.

Ot-lemmalarning teglanishi va ularning dasturga joylanishi natijasi uzbekcorpora.uz saytida sinovdan o‘tkazildi.

### **Bemor**

*Osmon yiroq, yer qattiq.  
Maqol*

Sotiboldining xotini og‘rib qoldi. Sotiboldi kasalni o‘qitdi – bo‘lmadi, **tabibga**

ko‘rsatdi. **Tabib qon** oldi. Betobning **ko‘zi** tinib, **boshi** aylanadigan bo‘lib qoldi. **Baxshi** o‘qidi. Allaqanday bir **xotin** kelib **tolning xipchini** bilan savaladi, **tovuq** so‘yib qonladi... Bularning hammasi, albatta, **pul** bilan bo‘ladi. Bunday vaqtlarda yo‘g‘on cho‘ziladi, ingichka uziladi.

**Shaharda** bitta **doktorxona** bor. Bu **doktorxona** to‘g‘risida Sotiboldining bilgani shu: salqin, tinch **parkda**, **daraxtlar** ichiga ko‘milgan baland va chiroyli oq imorat; **shisha** qabzali kulrang **eshigida qo‘ng‘iroq tugmasi** bor. **Chigit po‘choq** va **kunjara** bilan **savdo** qiladigan **xo‘jayimi Abdug‘aniboy omborda** qulab ketgan **qoplar** ostida qolib o‘ladigan bo‘lganida bu **doktorxonaga** bormay **Simga** ketgan edi. **Doktorxona** deganda Sotiboldining ko‘z oldiga izvosh va oq **podshoning surati** solingan 25 **so‘mlik pul** kelar edi...

Natija kutilganidek bo'ldi. Unda lug'atga kiritilgan ot turkumiga doir so'z shakllarining 97%i qamrab olindi. Ish uchun dasturda nazarda tutilmagan, ammo matnda uchraydigan bog'lamali otlarni dastur tanimadi. Bu uchun yana bog'lamali otlar bazasi shakllantirildi. Shuningdek, omonim shakllar masalasi ham uning kamchiligi sifatida ko'rinib turibdi.

Yuqoridagi matn faqat bosh yoki ikkinchi darajali admin uchun vizual shakldir. Uning foydalanuvchi uchun otlarning ko'rinishi quyidagicha teglangan holda bo'ladi:

**osmon** <**osmon**> [ot], [tur.ot], [o'.j.o.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**yer** <**yer**> [ot], [tur.ot], [o'.j.o.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**maqol** <**maqol**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**sotiboldining** <**sotiboldi**> [ot], [at.ot.], [shaxs\_n], [ya.ot.], [qo'sh.ot.], [yas.o.], [komp.], [birl.], [qar.k.]

**xotini** <**xotin**> [ot], [tur.ot], [sh.ot.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.], [III\_sh.b.]

**sotiboldi** <**sotiboldi**> [ot], [at.ot.], [shaxs\_n], [ya.ot.], [qo'sh.ot.], [yas.o.], [komp.], [birl.], [b.k.]

**tabibga** <**tabib**> [ot], [tur.ot], [sh.ot.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [j.k.]

**tabib** <**tabib**> [ot], [tur.ot], [sh.ot.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**qon** <**qon**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**baxshi** <**baxshi**> [ot], [tur.ot], [sh.ot.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**xotin** <**xotin**> [ot], [tur.ot], [sh.ot.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**tolning** <**tol**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [qar.k.]

**xipchini** <**xipchin**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [tush.k.], [III\_sh.b.]

**tovuq** <**tovuq**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**pul** <**pul**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**vaqtlarda** <**vaqt**> [ot], [tur.ot], [nar\_n], [mav.o.], [ya.ot.], [s.ot.], [t.o.], [ko'p.s.], [o'.p.k.]



- shaharda** <**shahar**> [ot], [tur.ot], [o' .j.o.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [o' .p.k.]
- doktorxona** <**doktorxona**> [ot], [tur.ot], [o' .j.o.], [an.o.], [ya.ot.], [s.ot.], [yas.o.], [aff.], [birl.], [b.k.]
- doktorxona** <**doktorxona**> [ot], [tur.ot], [o' .j.o.], [an.o.], [ya.ot.], [s.ot.], [yas.o.], [aff.], [birl.], [b.k.]
- sotiboldining** <**sotiboldi**> [ot], [at.ot.], [shaxs\_n], [ya.ot.], [qo' sh. ot.], [yas.o.], [komp.], [birl.], [qar.k.]
- parkda** <**park**> [ot], [tur.ot], [o' .j.o.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [o' .p.k.]
- daraxtlar** <**daraxt**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [ko' p.s.], [b.k.]
- imorat** <**imorat**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]
- shisha** <**shisha**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]
- eshigida** <**eshik**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [o' .p.k.]
- qo'ng'iroq** <**qo'ng'iroq**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]
- tugmasi** <**tugma**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [yas.o.], [aff.], [birl.], [b.k.], [III\_sh.b.]
- chigit** <**chigit**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]
- po'choq** <**po'choq**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]
- kunjara** <**kunjara**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]
- xo'jayini** <**xo'jayin**> [ot], [tur.ot], [sh.ot.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [tush.k.], [III\_sh.b.]
- abdug'aniboy** <**abdug'aniboy**> [ot], [at.ot.], [shaxs\_n], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]
- omborda** <**ombor**> [ot], [tur.ot], [o' .j.n.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [o' .p.k.]
- qoplar** <**qop**> [ot], [tur.ot], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [ko' p.s.], [b.k.]
- ostida** <**ost**> [ot], [tur.ot], [o' .j.o.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [o' .p.k.]
- doktorxonaga** <**doktorxona**> [ot], [tur.ot], [o' .j.o.], [an.o.], [ya.ot.], [s.ot.], [yas.o.], [aff.], [birl.], [j.k.]

**simga** <sim> [ot], [at.ot.], [geog.n.], [ya.ot.], [s.ot.], [t.o.], [birl.], [j.k.]

**doktorxona** <doktorxona> [ot], [tur.ot.], [o'.j.o.], [an.o.], [ya.ot.], [s.ot.], [yas.o.], [aff.], [birl.], [b.k.]

**sotiboldining** <sotiboldi> [ot], [at.ot.], [shaxs\_n], [ya.ot.], [qo'sh.ot.], [yas.o.], [komp.], [birl.], [qar.k.]

**ko'z** <ko'z> [ot], [tur.ot.], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**oldiga** <old> [ot], [tur.ot.], [o'.j.o.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [j.k.]

**izvosh** <izvosh> [ot], [tur.ot.], [sh.ot.], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

**oqpodshoning** <oqpodsho> [ot], [at.ot.], [shaxs\_n], [ya.ot.], [qo'sh.ot.], [yas.o.], [komp.], [birl.], [qar.k.]

**surati** <surat> [ot], [tur.ot.], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.], [III\_sh.b.]

**pul** <pul> [ot], [tur.ot.], [nar\_n], [an.o.], [ya.ot.], [s.ot.], [t.o.], [birl.], [b.k.]

Natijadan kelib chiqib shuni ayta olamizki, bu lug'at bazasi matni samarali segmentatsiyalash, mashina tarjimasini, ma'lumotni qidirish va obyektini tanib olish kabi ko'plab ilovalar uchun juda muhimdir. Bu ma'lum matn hujjatidagi alohida so'z yoki iboralarni tegishli toifalari asosida aniqlash va tasniflashni o'z ichiga oladi. Bu toifalar nutqning bir qismi, semantik tarkib yoki boshqa belgilar bo'lishi mumkin.

Matn izohi bir necha shaklda bo'lishi mumkin, jumladan, odamlar tomonidan qo'lda izohlash, avtomatik izohlash yoki ikkalasining kombinatsiyasi. Matni segmentatsiyalash va belgilash samaradorligi annotatsiya jarayonining to'g'riligiga va to'plangan tegishli ma'lumotlarning miqdoriga bog'liq.

Qo'lda matn izohida har bir so'z yoki ibora inson annotatori tomonidan oldindan belgilangan toifalar to'plamiga muvofiq etiketlanadi. Bu usul ko'p vaqt talab qiladi va xatolarga moyil bo'lishi tabiiy, chunki odamlar so'z va iboralarning ma'nolarini turlicha talqin qilishlari mumkin. Biroq bunday izohlar mashinaning o'rganish algoritmlarini ishlab chiqish va avtomatik izohlash uchun oltin standartlarni o'rnatishda foydali bo'ladi. Bu standartlar yuqorida keltirilgan adminlar yoki foydalanuvchi uchun ko'rinadigan shaklda bo'lmaydi. Ular qanday dasturiy kod sifatida avtomatlashtirilgan bo'lsa, xuddi shunday taniydi. Avval ta'kidlaganimizdek, ikkinchi darajali admin va foydalanuvchi uchun tushunarsizroq bo'lgan dasturning orqa qismidan foydalanib, quyidagicha ko'rinishdagi lug'at bazasi shakllantirildi:

| 1<br>soʻz<br>shakllarining<br>navbatga<br>asoslangan<br>dasturiy kodi | 2<br>lugʻat<br>bazasidagi soʻz<br>shakllarining<br>dasturiy roʻyxati | 3<br>lugʻat<br>bazasidagi soʻz<br>shakllarining<br>lemmalari<br>roʻyxati | 4<br>soʻz shakllarining<br>individual morfologik<br>dasturiy kodi | 5<br>dasturda<br>lugʻatlarni,<br>matnlarni<br>kirituvchi<br>adminga<br>yoʻnaltiruvchi<br>kod |
|---|--|--|---|--|
| 4593  | osmon  | osmon  | b%caaaa%aa%%  | 6  |
| 4594  | yer  | yer  | b%caaaa%aa%%  | 6  |
| 4595  | maqol  | maqol  | b%baaaa%aa%%  | 6  |
| 4596  | sotiboldining  | sotiboldi  | ab%aabbab%%   | 6  |
| 4597  | xotini   | xotin  | b%aaaa%aac%   | 6  |
| 4598  | sotiboldi  | sotiboldi  | ab%aabbbaa%%  | 6  |
| 4599  | tabibga  | tabib  | b%aaaa%ad%%   | 6  |
| 4600  | tabib  | tabib  | b%aaaa%aa%%   | 6  |
| 4601  | qon  | qon  | b%baaaa%aa%%  | 6  |
| 4602  | baxshi   | baxshi   | b%aaaa%aa%%   | 6  |
| 4603  | xotin  | xotin  | b%aaaa%aa%%   | 6  |
| 4604  | tolning  | tol  | b%baaaa%ab%%  | 6  |
| 4605  | xipchini   | xipchin  | b%baaaa%acc%  | 6  |
| 4606  | tovuq  | tovuq  | b%baaaa%aa%%  | 6  |
| 4607  | pul  | pul  | b%baaaa%aa%%  | 6  |
| 4608  | vaqtlarda  | vaqt   | b%bbaaa%be%%  | 6  |
| 4609  | shaharda   | shahar   | b%caaaa%ae%%  | 6  |
| 4610  | doktorxona   | doktorxona   | b%caabaabaa%%   | 6  |
| 4611  | doktorxona   | doktorxona   | b%caabaabaa%%   | 6  |
| 4612  | sotiboldining  | sotiboldi  | ab%aabbab%%   | 6  |
| 4613  | parkda   | park   | b%caaaa%ae%%  | 6  |
| 4614  | daraxtlar  | daraxt   | b%baaaa%ba%%  | 6  |
| 4615  | imorat   | imorat   | b%baaaa%aa%%  | 6  |
| 4616  | shisha   | shisha   | b%baaaa%aa%%  | 6  |
| 4617  | eshigida   | eshik  | b%baaaa%ae%%  | 6  |
| 4618  | qoʻngʻiroq   | qoʻngʻiroq   | b%baaaa%aa%%  | 6  |

| 1<br>soʻz<br>shakllarining<br>navbatga<br>asoslangan<br>dasturiy kodi | 2<br>lugʻat<br>bazasidagi soʻz<br>shakllarining<br>dasturiy roʻyxati | 3<br>lugʻat<br>bazasidagi soʻz<br>shakllarining<br>lemmalari<br>roʻyxati | 4<br>soʻz shakllarining<br>individual morfologik<br>dasturiy kodi | 5<br>dasturda<br>lugʻatlarni,<br>matnlarni<br>kirituvchi<br>adminga<br>yoʻnaltiruvchi<br>kod |
|---|--|--|---|--|
| 4619  | tugmasi  | tugma  | b%baaabaaac%  | 6  |
| 4620  | chigit   | chigit   | b%baaaa%aa%%  | 6  |
| 4621  | poʻchoq  | poʻchoq  | b%baaaa%aa%%  | 6  |
| 4622  | kunjara  | kunjara  | b%baaaa%aa%%  | 6  |
| 4623  | xoʻjayini  | xoʻjayin   | b%aaaa%acc%   | 6  |
| 4624  | abdugʻaniboy   | abdugʻaniboy   | ab%aaaa%aa%%  | 6  |
| 4625  | omborda  | ombor  | b%caaaa%ae%%  | 6  |
| 4626  | qoplar   | qop  | b%baaaa%ba%%  | 6  |
| 4627  | ostida   | ost  | b%caaaa%ae%%  | 6  |
| 4628  | doktorxonaga   | doktorxona   | b%caaabaad%%  | 6  |
| 4629  | simga  | sim  | ab%aaaa%ad%%  | 6  |
| 4630  | doktorxona   | doktorxona   | b%caaabaaa%%  | 6  |
| 4631  | sotiboldining  | sotiboldi  | ab%aabbbab%%  | 6  |
| 4632  | koʻz   | koʻz   | b%baaaa%aa%%  | 6  |
| 4633  | oldiga   | old  | b%caaaa%ad%%  | 6  |
| 4634  | izvosh   | izvosh   | b%aaaa%aa%%   | 6  |
| 4635  | oqpodshoning   | oqpodsho   | ab%aabbbab%%  | 6  |
| 4636  | surati   | surat  | b%baaaa%aac%  | 6  |
| 4637  | pul  | pul  | b%baaaa%aa%%  | 6  |

Koʻrinib turibdiki, avtomatik matn izohi matn maʼlumotlarini tasniflash va belgilash uchun mashinaning oʻrganish algoritmlaridan foydalanishni oʻz ichiga oladi. Bu jarayon algoritmgga soʻzlar va iboralarni aniq tasniflashni oʻrganish imkonini beruvchi izohli matn maʼlumotlarining katta korpusida mashinaning oʻrganish modelini oʻrgatishni oʻz ichiga oladi. Avtomatik izoh qoʻlda izohlashdan koʻra tezroq va kengaytirilishi mumkin, ammo maʼlumotlar sifati va miqdori bilan cheklangan.

Matndagi soʻzlar va iboralarni muayyan toifalarga guruhlash va teglash orqali berilgan matn hujjatiga qimmatli maʼlumotlarni beradi. Matnni segmentlash va belgilash uchun tanlangan usul ilovaning oʻziga xos ehtiyojlariga, aniqlik va vaqt talablariga va mavjud boʻlgan tegishli maʼlumotlar miqdoriga bogʻliq. Tabiiy tilni qayta ishlash va hisoblash tilshunosligi sohasi rivojlanishda davom etar ekan, matnni izohlash jarayoni tadqiqot va ishlanmalarni rivojlantirishda muhim rol oʻynaydi.

#### ADABIYOT

1. Ataboyev N. Compiling dictionaries by using corpus analysis and its advantages // International Journal of Progressive Sciences and Technologies.
2. Ataboyev N. Problematic issues of corpus analysis and its shortcomings // ISJ Theoretical & Applied Science, 10 (78), 2019.
3. Bongers H. The history and principles of Vocabulary control. – Woerden: WOCOPI, 1947.
4. Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – L., 1969.
5. Mengliyev B. Oʻzbek tili taraqqiyoti va rivojlanish omillari // “Oʻzbek tilini dunyo miqyosida keng targʻib qilish boʻyicha hamkorlik istiqbollari” mavzusidagi xalqaro ilmiy-amaliy anjuman materiallari. – Toshkent, 2020.
6. Ахмедова Д. Атов бирликларини ўзбек тили корпуслари учун лексик-семантик теглашнинг лингвистик асос ва моделлари: Филол. фан. бўйича фалсафа доктори (PhD) дисс. – Бухоро, 2020. – 247 б.
7. Блумфилд Л. Язык. – М.: «Прогресс», 1968. – 608 с.
8. Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) – //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf.
9. Менглиев Б., Бобожонов С., Хамроева Ш. Ўзбек тилининг миллий корпуси. <http://marifat.uz/marifat/ruknlar/fan/1241.htm>
10. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург. – 2006. 26 с.
11. Плунгян В. Зачем мы делаем Национальный корпус русского языка? «Отечественные записки» 2005, №2. [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html)
12. Рыков В.В. Курс лекций по корпусной лингвистике. URL: <http://rykov-cl.narod.ru/c.html>
13. Синклер Д. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка»/ Д. Синклер [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-infro.html>

14. Тоирова Г. Миллий корпус яратишнинг технологик жараёни хусусида // Ўзбекистонда хорижий тиллар. Электрон илмий-методик журнал. – Тошкент. 2020, № 2 (31), – Б. 57–64. <https://journal.fledu.uz/uz/2-31-2020>

15. Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – М., 1967.

16. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Қарши, 2018.

17. Эшмўминов А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол. фан. бўйича фалсафа доктори (PhD) дисс. – Қарши, 2019. – 137 б.

УДК 81'322.2

**ОСНОВАННЫЙ НА ЗНАНИЯХ WSD ПОДХОД  
В УЗБЕКСКОМ ЯЗЫКЕ****Н. З. Абдурахмонова<sup>1</sup>, Ж. Б. Исроилов<sup>2</sup>**<sup>1</sup>*Национальный университет имени Мирзо Улугбека  
Ташкент, Узбекистан,*<sup>2</sup>*Наманганский государственный университет  
Наманган, Узбекистан*

В этой статье проведен сравнительный анализ НЛП и подхода WSD, основанного на знаниях. Даны рекомендации по решению проблемы WSD Узбекского корпуса

**Ключевые слова:** Обработка естественного языка, ОСНОВАННЫЙ НА ЗНАНИЯХ, подход, WSD, Сенсеваль

**KNOWLEDGE-BASED WSD APPROACH IN UZBEK LANGUAGE*****Abdurakhmonova Nilufar<sup>1</sup>, Isroilov Jasur<sup>2</sup>***<sup>1</sup>*National University of Uzbekistan  
Tashkent, Uzbekistan*<sup>2</sup>*Namangan State University  
Namangan, Uzbekistan, 160107*

n.abduraxmonova@nuu.uz, jasurisroilov@namdu.uz

**Abstract:** In this paper, a comparative analysis of NLP and Knowledge-based WSD approach was performed. Recommendations for solving the WSD problem of the Uzbek Corpus were given.

**Keywords:** NLP, Knowledge-based, approach, WSD, Senseval

Natural language processing (NLP) is a field of computer science that gives computers the ability to understand, interpret, and process human language. It is a broad field that encompasses a wide range of tasks, such as:

**Machine translation:** translating text from one language to another.

**Text summarization:** extracting the main points of a text and presenting them in a concise form.

**Question answering:** answering questions posed in natural language.

**Sentiment analysis:** determining the sentiment of a piece of text, such as whether it is positive, negative, or neutral.

**Named entity recognition:** identifying and classifying named entities in text, such as people, organizations, and locations.

NLP is a rapidly growing field, and there are many new research directions being explored. Some of the most promising research directions include:

Developing new methods for understanding and processing natural language that are more accurate and efficient.

Developing new applications for NLP, such as chatbots and virtual assistants.

Making NLP more accessible to a wider range of users, such as non-technical people and people with disabilities.

NLP has the potential to revolutionize the way we interact with computers. As NLP technology continues to improve, it will become increasingly common for computers to understand and process human language in a natural way. This will make it easier for people to use computers to accomplish a wide range of tasks, from finding information to communicating with others.

Today, there are many unsolved problems of the Uzbek language NLP. One such problem is Word Sense Disambiguation (WSD), which has not been adequately studied by Uzbek linguists.

WSD is a subfield of NLP that deals with the problem of determining the correct sense of a word in a given context. This is a challenging task because words can have multiple senses, and the meaning of a sentence can change depending on the sense of a word that is used.

Knowledge-based WSD is a subfield of NLP that deals with the problem of determining the correct sense of a word in a given context. This is a challenging task because words can have multiple senses, and the meaning of a sentence can change depending on the sense of a word that is used.

The history of knowledge-based WSD can be traced back to the early 1980s. In 1983, Steven L. Small and his colleagues published a paper titled "Parsing and Knowledge-based Inference for Machine Translation". This paper introduced the first knowledge-based WSD system, called MEAD. MEAD used a knowledge base of word senses to disambiguate words in a corpus of text.

In the following years, there was a growing interest in knowledge-based WSD. In 1993, the Senseval evaluation campaign was held. Senseval was a competition between different WSD systems. The goal of Senseval was to compare the performance of different WSD systems and to identify the best WSD system.



The first Senseval evaluation campaign was a success, and it led to a number of improvements in knowledge-based WSD systems. In the following years, there were several more Senseval evaluation campaigns, and the performance of knowledge-based WSD systems continued to improve.

Today, knowledge-based WSD is a well-established field of NLP. There are a number of different knowledge-based WSD systems available, and these systems are used in a variety of applications, such as machine translation and text understanding.

In 1983 Steven L. Small and his colleagues publish the paper “Parsing and Knowledge-based Inference for Machine Translation”, which introduces the first knowledge-based WSD system, called MEAD. After then in 1993 the Senseval evaluation campaign is held by Jarowski. Senseval is a competition between different WSD systems. In 1998 the second Senseval, in 2004 the third Senseval, in 2010 the fourth Senseval, in 2014 the fifth Senseval evaluation campaign is held.

Knowledge-based WSD is a rapidly evolving field, and there are many new research directions being explored. Some of the most promising research directions include:

- Developing new methods for disambiguating words in new contexts, even if the word has not been seen before in a corpus of text.
- Developing methods for disambiguating words that are ambiguous due to their morphology or syntax.
- Developing methods for disambiguating words in noisy or informal text.
- Developing methods for disambiguating words in different languages.

As WSD technology continues to improve, it will enable NLP applications to perform more accurately and effectively.

Having studied several WSD approaches, I set myself the task of creating a WSD tool for the corpus of the Uzbek language using the Knowledge-based WSD approach in my future research work.

## REFERENCES

1. Ranjan Pal, Diganta Saha, WSD: A Survey. International Journal of Control Theory and Computer Modeling (IJCTCM) Vol.5, No.3, July 2015.
2. Loïc Vial, Benjamin Lecouteux, Didier Schwab. Sense Embeddings in Knowledge-Based WSD. 12th International Conference on Computational Semantics (IWCS), 2017.

3. Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. An Enhanced Lesk WSD Algorithm through a Distributional Semantic Model. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 1591–1600, Dublin, Ireland.
4. N.Abdurakhmonova, Dependency Parsing Based On Uzbek Corpus. Language technology for all (LT4all). 2019.
5. N.Abdurakhmonova, J.Isroilov,. Personal names spell-checking—a study related to Uzbek. Journal of Social Sciences and Humanities Research. Volume 6, 2018
6. Hwee Tou Ng. Exemplar-Based WSD Some Recent Improvements. In Second Conference on Empirical Methods in Natural Language Processing. 1997.

УДК 81'322.2

**РАЗРАБОТКА УЗБЕКСКО-АНГЛИЙСКОЙ ДВУЯЗЫЧНОЙ  
ПРОГРАММЫ НА ОСНОВЕ МОДЕЛИРОВАНИЯ  
ГРАММАТИЧЕСКИХ КАТЕГОРИЙ ГЛАГОЛОВ  
И МОРФОЛОГИЧЕСКОГО АНАЛИЗА ДЛЯ МАШИННОГО  
ПЕРЕВОДА**

*Э. Ш. Назирова<sup>1</sup>, Н. З. Абдурахмонова<sup>2</sup>, Усмонова Камола<sup>1</sup>*

*<sup>1</sup>Ташкентский университет информационных технологий  
имени Мухаммада аль-Хоразми, Ташкент, Узбекистан*

*<sup>2</sup>Национальный университет имени Мирзо Улугбека  
Ташкент, Узбекистан*

*elmira\_nazirova@mail.ru, n.abduraxmonova@nuu.uz,  
kamolausmonova1234@gmail.com*

В данной статье раскрывается подход к разработке морфемы и моделированию грамматических категорий глаголов в узбекском языке.

Более того, в статье выделены аспекты, которые особенно необходимо учитывать при их отражении в базе данных.

Обозначены виды морфологического анализа грамматических категорий глагола, а также намечены общие парадигмы и различия в происхождении и переводе языка. Дифференцируются аналитические особенности и принципы образования глаголов. Эти особенности глаголов, как и других частей речи, определяют необходимость специальных лингвистических исследований сложных, вспомогательных глаголов, всех видов словосочетаний, введения в созданную автором программную базу модулей для идеального машинного перевода.

А также указано строение, связи, состав, определения сочетаний глагольных аффиксов при моделировании грамматических категорий. Глаголы узбекского языка характеризуются аналитическим характером и особой формой образования в виде совместного, вспомогательного сочетания глаголов. Классификация этих типов требует большого лингвистического анализа в компьютерной морфологии. Обогащение словарного запаса узбекского языка и развитие науки и техники, внедрение неологизмов и введение заимствованных слов также требуют отдельного исследования и анализа.

Основной акцент в статье сделан на построении предложений при переводе с узбекского языка на английский и наоборот. Данная статья может послужить хорошей основой для новых исследований и научных работ, так как требует более глубокого изучения внутренних языковых возможностей узбекской слово семейства и их лексико-семантических моделей.

**Ключевые слова** комплексная разработка, узбекский язык, английский язык, двуязычие, программа, базис, модель, глагольная категория, морфологический анализ, машинный перевод, компьютерная лингвистика, автома-

тическая морфология, грамматическая категория, аналитические глаголы, словосочетание, морфологический анализ, моделирование, база данных.

## DEVELOPMENT OF UZBEK-ENGLISH BILINGUAL PROGRAM SOFTWARE ON THE BASIS OF MODELLING OF GRAMMATICAL CATEGORIES OF VERBS AND MORPHOLOGICAL ANALYSIS FOR MACHINE TRANSLATION

*Elmira Nazirova<sup>1</sup>, Nilufar Abdurakhmonova<sup>2</sup>, Kamola Usmonova<sup>1</sup>*

<sup>1</sup>Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan

<sup>2</sup>National University of Uzbekistan named after Mirzo Ulugbek Tashkent, Uzbekistan

elmira\_nazirova@mail.ru, n.abduraxmonova@nuu.uz,  
kamolausmonova1234@gmail.com

This article reveals the approach of development of the morphemic and modelling of grammatical categories of verbs in Uzbek language.

Moreover, the article highlights the aspects that are specifically necessary to be considered while reflecting them in the database.

The types of morphological analysis of grammatical categories of the verb are indicated, as well as common paradigms and differences in the origin and the translation of the language are outlined. Analytical features and principles of verb formation are differentiated. These features of verbs as well as other parts of speech determine the need for special linguistic research on complex, auxiliary verbs, on all types of word combinations, for the introduction of modules in the program base created by the author for perfect machine translation.

As well as the structure, connections, composition, definitions of combinations of verb affixes in the modelling of grammatical categories are indicated. Verbs of Uzbek language are characterized by their analytical character and special form of formation in the form of joint, auxiliary combination of verbs. Classification of these types requires a large linguistic analysis in computer morphology. Enrichment of vocabulary in Uzbek language and development of science and technology, implementing neologisms and introducing borrowed words also require separate research and analysis.

The main emphasis in the article is made on the construction of sentences when translating from Uzbek into English and vice versa. This thesis can serve as a good basis for new research and scientific works, as it requires a deeper study of the internal linguistic capacities of the Uzbek word family and their lexicosemantic models.

**Keywords:** complex development, Uzbek language, English language, bilingual, program, basis, model, verb category, morphological analysis, machine translation, computer linguistics, automatic morphology, grammatical category, analytic verbs, word combination, morphological analysis, modelling, database.

## **Introduction**

Nowadays it is hard to imagine any sphere of activity without ICT. Information technologies are crucial for the development of science and have a strong impact on the optimization of infrastructure in the field of knowledge accumulation, mainly on Internet Technologies in the world as well as in Uzbekistan. Since independence of our country, the infrastructure has undergone crucial changes. These changes had mostly affected the sphere of education. Widespread use of computing technologies and the Internet has made it possible to solve serious scientific problems. In the fields with where computer technologies are used new scientific directions have been introduced.

Today we can witness not only a few examples of this. There are doctors, teachers, and bankers who widely use information technologies both in daily basis and in scientific reports and research, projects and trainings.

In this way, laboratories, IRC Universities, and program classes have been equipped and furnished since the 2000s. One of the main ideologies of Uzbek computer linguistics is the most correct machine translation based on mathematical modelling and containing all the features of the languages being translated. After all, translation is the key to progress and rapid exchange of experience and innovations.

It is worthy to note that the plenty of scientists from our republic have conducted the first experiments and laboratory surveys: PhD M. Khakimov has conducted a lot of surveys on mathematic modelling in the field of machine translation. Under the guidance of M. Khakimov there were presented the following handbooks and training manuals: “Computer Linguistics” [A. Pulatov, 2011], “Fundamentals of Computer Linguistics” [A. Rakhimov, 2011], “Linguistic Fundamentals of Machine Translation” [N. Abdurakhmonova, 2012] and others. Most of these works had theoretical form and contents, yet there were no program realisations based on linguistic databases. There was no cooperation work between linguists and programmers. Currently, computer linguistics as a science is being taught in several state universities in Uzbekistan: in Tashkent, Andijan, Namangan, Fergana, Khorezm, Samarkand, and Bukhara.

Based on the Decrees of the first President and clearly defined strategies of the current President Sh. Mirziyoyev the following essential issues have been emphasized that “...ensuring the appropriate place of our native language in the world information network Internet. Furthermore, its computer support and availability of scientific and

methodological aids relate to machine translation and electronic dictionaries, preparation of recommendations for wide application of the results in practice”.

Uzbek language – the language of great philosophers, scientists, poets and writers recognised on the international level, such as Ulugbek, Ibn Sina, Alisher Navoi, Al Khwarizmiy and others.

Uzbek language belongs to Turkic languages, that have an ancient history and constantly changing conditions due to variety of reasons. We can observe its distinguishing features from other languages at each of the language levels.

For example, the preserved vowel harmony in Turkish in words such as “üzüm, velâyet”. However, there are more loanwords in Uzbek than in Turkish.

For instance, “insulin, management, budget, test” from English, “стол, поезд, бухгалтер” from Russian, “agronomiya, allergiya, nargis” from Greek, “vazir, maktab, maorif” from Arabic.

On the other hand, one of the important tasks of computer linguistics in Uzbekistan is to create grammar analysers for Uzbek language corresponding to Latin script. In neighbour-countries such as Kyrgyzstan and Kazakhstan “...an initial version of morphological analyser system in the sphere of Embarcadero RAD Studio has been developed, taking into account morphology. Of course, it requires further development and research and development in this direction is underway”. Similarly, in Uzbekistan a number of young scientists are striving to develop and implement their own Uzbek-English bilingual software system based on modelling of grammatical categories of verbs and morphological analysis in machine translation.

## **TECHNOLOGIES, MODELS and SYSTEMS**

1. Annotation Uzbek Grammar The grammar consists of two parts: morphology and syntax.

### **Speech parts of Uzbek language**

| Notional parts of speech                                   | Functional parts of speech                      | Individual word groups                          |
|--|---|---|
| Noun<br>Adverb<br>Adjective<br>Pronoun<br>Verb<br>Numerals | Conjunction<br>Interjections<br>Auxiliary words | Service words<br>Modal words<br>Imitation words |

Grammatical meanings, derivations, word rules and formal models in morphology are considered as linguistic processes. Formal morphological patterns are the result of the use of word combinations and the relations between them in a text. Formal patterns always exist in a syntagma.

A syntagma is a semantic-syntactic unit that expresses some unified words as a meaningful part of a sentence. A linguistic database includes a grammar and a dictionary.

Any syntactic analysis consists of three main components:

- 1) Parts of speech;
- 2) Parts of sentence;
- 3) Types of sentences.

Uzbek is a morphologically rich language with nouns, adjectives, and verbs changing in case, number, and other word forms. This property requires adding morphological information to machine translation systems to eliminate the lack of multiple inflectional forms. So also parallel to it alternative information from English. For machine translation it is important to create a formal grammar of Uzbek language in all its forms and reflections. Uzbek language has agglutinative morphology with productive inflectional and derivational suffixes. Suffixes can be added sequentially and a single word can contain many parameters such as possessive, plural/singular, case, modality, etc.

Case modification is a common linguistic category present in many languages of the world. In the literature on formal syntax, there are two main approaches to distinguishing cases. The first approach is mainly associated with the work of Noam Chomsky, who views case as a syntactic phenomenon known in NLP; the second approach, proposed in the work of Alec Marantz, views case as a presyntactic, purely morphological phenomenon.

As Nilufar Abdurakhmonova mentioned that “There are the following derivational models of the Uzbek language”:

W+A=>nok+zor  
A+W=>be+foyda  
W+W=>tez+yurar  
W-W=>ota-ona  
W W=>sotib olmoq  
W-u/yu W=>Erta-yu kech

Due to the lack of grammatical information for natural language processing, language description for linguistic database is carried out.

Modelling of grammatical categories for machine translation in Uzbek language is done in comparison with English. English and Uzbek languages belong to different language groups. Therefore, highlighting the unique features and differences of both languages is considered important for morphological analysis". If we consider this process on the example of Uzbek verbs, we can see that the translation process is a difficult job because of the mental and conceptual differences that exist in different language families, societies and cultures. Both linguistic (ambiguity, synonyms, paronyms, homonyms) and extra-linguistic (psychological) factors as well as culture and mentality influence the quality of translation. Even human translators face the same problems in the translation process that machine translation systems face.

When the translation is done between related languages, it is easier to translate, but when it is contexts from unrelated languages, it makes machine translation inefficient, because word roots, endings, prepositions can have different meanings. The translator's work is also complicated by the order in which sentences are constructed and the way they are expressed.

In the process of comparing two languages, we can observe that both languages have the identical forms of verbs:

1. Imperative.
2. Perfect verbs are used for the past tense in Uzbek.
3. Imperfect verbs are used for future tense in English, but are used to express different tenses in Uzbek (past, present and future) in combination with different inflections and particles.
4. In Uzbek, active and passive participles are used to a lesser extent than in English.

The verbs have the following grammatical categories:

If we look at agglutinative languages such as Finnish, we find that morphosyntactic features are systematically encoded by individual morphemes, which are arranged in linear order.

Word-forming and syntactic affixes are particularly difficult to translate. If in Uzbek language there are more than 6000 dictionary words and more than 206 types of affixes and their variations of parts of speech, 130 of which are verbal, it is necessary to do a lot of work to enter all variants in English into machine translation. And if we also take into account the synonymic series of both one and two languages, we can realize that in this direction it is necessary to carry out complex work and research. This fact became the main incentive for the author of the article, who speaks both English and Uzbek languages, acquires



the basics of information and technical skills, experience of translator to give start to this fundamental work along with scientific and practical research.

For comparison we can refer to the verb “uchmoq”, so we could see some examples of different models of verb structures:

- 1) Simple verb – uchmoq (fly)
- 2) Compound verb – uchib ketmoq (fly away)
- 3) Collocation – raketa uchirmoq (fly the rocket)
- 4) Collocation with verb – varrak uchirib bermoq (fly the kite to smb.)
- 5) Combination with a modal word – uchirish kerak (must fly)
- 6) Idiom - kapalagim uchib ketdi (be afraid)

While entering text, the morphological analyser must correctly analyse each segment of the text. Otherwise, homonymy problems arise when translating text units. For example, the word combination “qo’yib berdi” is used in many functions as contextual homonymy, as in the following examples:

U hujjatni stolga qo’yib berdi-> He gave document as putting on the table.

U bolani hovlida o’yinab olishiga qo’yib berdi-> He let the boy play in the yard.

Direktor ko’rsatilgan hujjatlarga darhol imzo qo’yib berdi-> The director signed abruptly brought documents.

U bolalar o’yinab olsin deb, sho’x ashula qo’yib berdi-> He played music so that to dance the children.

The purpose of study at this stage of research is to create a database of phrasal verbs as analytical models in English-Uzbek translation.

In order to build a system of machine translation from English into Uzbek, the size of the vocabulary stored in the database should be specified.

English and Uzbek have very large databases including all linguistic levels, and they are very different. The verb category in English is phrasal verb. Anyway, phrasal verbs in English as a verb phrase in Uzbek have their own peculiarities. It is a problem for the structural components of the sentence. Phrasal verbs are considered a very important and frequent feature of the English language. Firstly, they are so common in everyday conversation, and foreigners who want to seem natural when speaking English have to study grammar to know how to pronounce them correctly. Second, the habit of inventing phrasal verbs has been a source of great enrichment for the language. Phrasal verbs

are used to describe the greatest variety of human actions and relationships. And this means that English verb constructions are very difficult to analyse and coherently describe in synchronic terms.

In conclusion, it should be noted that:

– linguistic models and semantic relations of each linguistic unit play an important role in creating databases for machine translation systems;

– because of globalization processes, everything is subject to rapid change;

– there are no obstacles to prevent the unification of cultural and social relations between people.

Therefore, the understanding of foreign languages is significant, so we cannot disregard this fact while compiling a completely new, including all the constructions of the two languages, and the most convenient software. Today the result of machine translation, which was introduced in the last half of the 20th century, plays a huge role in the development of sciences and exchange of information.

#### REFERENCES

1. Abdurakhmonova. The bases of automatic morphological analysis for machine translation / Известия Кыргызский государственный технический университет им. И. Раззокова теоритической и прикладной научно-технический журнал, № 2 (38), 2016 – p. 12–17.

2. Abdurakhmonova. Modeling grammatical categories of verb in Uzbek as stage of morphological analysis in machine translation «В международной конференции по компьютерной обработке тюркских языков «turklang »», 2017, p. 155.

3. Lyutikova. Formal Modeling of case variation: a parametric approach // Computer Linguistics and Intellectual Technologies Proceedings of the Annual International Conference, «dialog» Edition 15, 2016, p. 461.

4. Olteanu. A holistic approach to phrasal verbs, “Editura Sfântul Ierarh Nicolae”, 2012, p. 16.

5. Roark & Sproat. Computational Approaches to Morphology and Syntax. Oxford University Press Inc., New York, 2007, p. 63.

УДК 81'322.3

**МОРФОЛОГИЧЕСКИЙ АНАЛИЗ ПРИЛАГАТЕЛЬНЫХ  
В УЗБЕКСКО-АНГЛИЙСКИХ ЯЗЫКАХ ДЛЯ ПРОГРАММНОГО  
ОБЕСПЕЧЕНИЯ «ALIGNER»*****Ш. М. Хамроева<sup>1</sup>, Н. Ш. Матъякубова<sup>1</sup>, А. Ю. Даулетов<sup>2</sup>****<sup>1</sup>Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои, Ташкент, Узбекистан.**<sup>2</sup>Университет Альфраганус, Ташкент, Узбекистан.  
shaxlo.xamrayeva@navoiy-uni.uz, nailya89mm@mail.ru,  
davletov--odilbek@mail.ru*

Морфологический анализ является одним из основных этапов процесса выравнивания параллельных текстов, он помогает сохранить лингвистическую и грамматическую информацию, повысить качество перевода, облегчить исследование разных языков. Он играет важную роль в обеспечении правильного выравнивания параллельных текстов, что необходимо для реализации различных задач обработки естественного языка и разработки приложений. Реализация такого взгляда на анализ играет важную роль, особенно при работе с языками, богатыми флективно и имеющими словообразовательную морфологию. Узбекский и английский – совершенно разные языки по морфологическому строению, в том числе это касается прилагательных и их сравнительных форм. В данной статье анализируются морфологические различия в образовании прилагательных и их форм в узбекском и английском языках.

**Ключевые слова:** Параллельный текст, морфологическая форма, токенизация, флективные языки, морфема.

**MORPHOLOGICAL ANALYSIS OF ADJECTIVES IN THE  
UZBEK-ENGLISH LANGUAGES FOR “ALIGNER” SOFTWARE*****Shahlo Hamroyeva<sup>1</sup>, Noila Matyakubova<sup>1</sup>, Adilbek Dauletov<sup>2</sup>****<sup>1</sup>Tashkent State University of Uzbek  
Language and Literature named after Alisher Navoi,  
Tashkent, Uzbekistan**<sup>2</sup>Alfraganus University, Tashkent, Uzbekistan.  
shaxlo.xamrayeva@navoiy-uni.uz, nailya89mm@mail.ru,  
davletov--odilbek@mail.ru*

Morphological analysis is one of the main steps in the process of aligning parallel texts, it helps to preserve linguistic and grammatical information, improve the quality of translation, and facilitate research on different languages. It plays an important role in ensuring the correct alignment between parallel texts, which

is necessary for the implementation of various natural language processing tasks and application development. Implementation of this view of the analysis plays an important role, especially when working with languages that are inflectionally rich and have derivational morphology. Uzbek and English are completely different languages in terms of morphological structure, and this also applies to adjectives and their comparative forms. This article analyzes the morphological differences in the formation of adjectives and their forms in Uzbek and English.

**Key words:** Parallel text, morphological form, tokenization, inflectional languages, morpheme.

### “ALIGNER” DASTURIY VOSITASI UCHUN O‘ZBEK-INGLIZ TILIDA SIFATNING MORFOLOGIK TAHLILI

*Hamroyeva Sh. M.<sup>1</sup>, Matyakubova N. Sh.<sup>1</sup>, Dauletov A. Y.<sup>2</sup>*

<sup>1</sup>*Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili  
va adabiyoti universiteti Toshkent, O‘zbekiston*

<sup>2</sup>*Alfraganus universiteti, Toshkent, O‘zbekiston*  
shaxlo.xamrayeva@navoiy-uni.uz, nailya89mm@mail.ru,  
davletov--odilbek@mail.ru

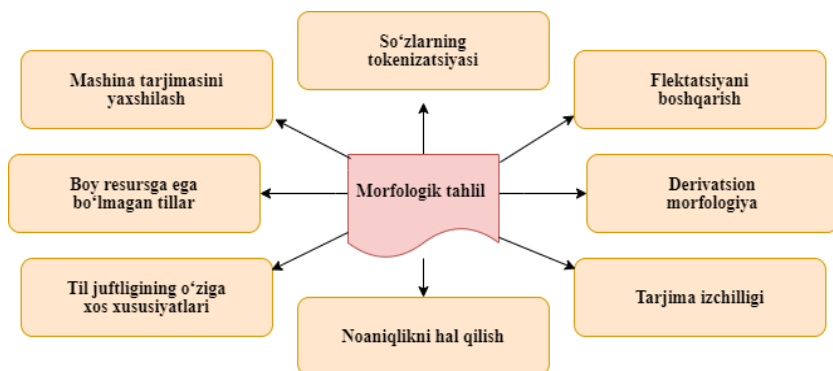
Morfologik tahlil parallel matnlarni moslashtirish jarayonidagi asosiy qadamlardan biri bo‘lib, lingvistik va grammatik ma‘lumotlarni saqlash, tarjima sifatini oshirish, turli tillar bo‘yicha tadqiqotlarni osonlashtirishga yordam beradi. Bular tabiiy tilni qayta ishlashning turli vazifalarini amalga oshirish va ilovalarni ishlab chiqishda zarur bo‘lgan parallel matnlar o‘rtasida to‘g‘ri moslashtirishni ta‘minlashda muhim ro‘l o‘ynaydi. Tahlilning bu ko‘rinishini amalga oshirish, ayniqsa, flektiv jihatdan boy va derivativ morfologiyaga ega bo‘lgan tillar bilan ishlashda muhim ro‘l o‘ynaydi. O‘zbek va ingliz tillari morfologik tuzilishi jihatidan bir-biridan mutlaqo farq qiladigan tillar bo‘lib, bu sifatlar va ularning qiyosiy shakllariga ham taluqli. Mazkur maqolada o‘zbek va ingliz tilida sifat va uning shakllari yasashigi morfologik farqlar tahlil qilinadi.

**Kalit so‘zlar:** Parallel matn, morfologik shakl, tokenizatsiya, flektiv tillar, morfema.

**Kirish.** Parallel matnlarni moslashtirish jarayonining samarali va sifatli ishlashini taminlash uchun eng muhim omillardan biri moslashtirish jarayonidagi matnlar tahlili hisoblanadi. Matnlarni sintaktik, semantik, grammatik va morfologik tahlil qilish ikki tildagi matnlaridagi sintaktik butunliklar – gaplarni o‘zaro moslashtirish jarayonining aniqliligini ta‘minlashga yordam beradi. Tahlil qilish orqali matnlarning mazmuni va tuzilishini tushunib, moslashtirish algoritmlari aslyat matnning qaysi qismlari tarjima matnning qaysi qismlariga mos kelishi haqida ko‘proq asosli qaror qabul qilish imkonini beradi.

Chunki tillarning sintaktik strukturasi turli xil; soʻz tartibi ham oʻziga xos boʻladi. Matnни tahlil qilish ushbu struktur farqlarini aniqlash va hisobga olishga yordam beradi, bu esa moslashtirilgan matnlarning tarjima tilida maʼnoli tarjimaga ega boʻlishini taʼminlaydi. Bundan tashqari moslashtirish vositalari, matnlarni tahlil qilish orqali mashina tarjimasi, koʻp tilli ilovalar, lingvistik tadqiqotlar hamda tilni qayta ishlash vositalarida sifat nazorati harakatlarini shakllantirishda muhim vosita hisoblanadi.

**1. Parallel matnlarni moslashtirish jarayonida morfologik tahlilning oʻrni.** Takidlanganidek, matnlarni bir necha usulda tahlil qilish mumkin boʻlib, ularga boʻlgan ehtiyoj asliyat va tarjima tillarining tuzilish va qoʻllanilishiga bogʻliq. Morfologik tahlil parallel matnlarni moslashtirish jarayonida, ayniqsa, flektiv jihatdan boy va derivativ morfologiyaga ega boʻlgan tillar bilan ishlashda hal qiluvchi roʻl oʻynaydi.



1- Rasm. Moslashtirish jarayonida morfologik tahlilning ahamiyati

Quyida parallel matnlarni moslashtirishda morfologik tahlilning muhimligining bir necha asosiy sabablarini koʻrib chiqamiz:

**1. Soʻzlarning tokenizatsiyasi.** Morfologik tahlil soʻzlarni prefiks, qoʻshimcha va oʻzak kabi morfemalarga ajratishga yordam beradi. Bu, ayniqsa, flektiv – grammatik shakllar chegarasi aniq belgilanmagan tillarda juda muhim. Tokenizatsiyasi parallel matnlarni moslashning birinchi bosqichidir.

**2. Flektatsiyani boshqarish.** Koʻp tillarda zamon, jins, son va hol kabi grammatik maʼnolarni ifodalashda flektiv morfemalardan foydalanadi. Morfologik tahlil bu jihatlarni aniqlashga va parallel matn-

lar bo‘ylab mos keladigan shakllarni moslashtirishga yordam beradi, grammatik istisnolar saqlanishini ta’minlaydi.

**3. Derivatsion morfologiya.** Tillar ko‘pincha so‘zlarda o‘zakdan oldin (prefiks) yoki keyin qo‘shimchalar qo‘shish kabi derivativ jarayonlar orqali yangi so‘zlarni yaratadi. Morfologik tahlil bu hosilaviy so‘zlarni tanib olishga imkon beradi; tillararo o‘zaro bog‘liq so‘zlar va ularning hosilalarini moslashtirishga yordam beradi.

**4. Tarjima izchilligi.** Morfologik tahlil tarjimalarda izchillikni saqlashga yordam beradi. Birlikning morfologik strukturasi aniqlash orqali parallel matnlar bo‘ylab ekvivalent morfemalarning izchil tarjima qilinishi ta’minlanib, umumiy tarjima sifati yaxshilanadi.

**5. Noaniqlikni hal qilish.** Morfologik tahlil omograflarning (imloli bir xil, ammo ma’nolari har xil bo‘lgan so‘zlar) morfologik xususiyatlarini aniqlashga yordam beradi. Bu, ayniqsa, homografiya darajasi yuqori bo‘lgan tillarda juda ahamiyatli.

**6. Til juftligining o‘ziga xos xususiyatlari.** Turli tillarda o‘ziga xos morfologik struktura va qoidalar mavjud. Morfologik tahlil moslashtirish algoritmlariga tilga xos xususiyatlarni hisobga olish imkonini beradi, bu esa muayyan til juftlari uchun moslash aniqligini oshiradi.

**7. Mashina tarjimasini yaxshilash.** Mashina tarjimasini kontekstida morfologik tahlil asliyat matn haqida ko‘proq ma’lumot berish orqali tarjima sifatini yaxshilashi mumkin. Bu tarjima modeliga kontekstga mos tarjimalarni yaratishga yordam beradi.

**8. Boy resursga ega bo‘lmagan tillar.** Morfologik tahlil, ayniqsa, parallel korpus va lingvistik resurslari cheklangan, resurslari kam bo‘lgan tillarda amaliy ahamiyat kasb etadi. Bu moslashtirish jarayonini avtomatlashtirishga yordam beradi va til bilan bog‘liq vazifalar uchun qimmatli parallel ma’lumotlarni yaratishga zamin hozirlaydi.

Morfologik tahlil parallel matnlarni moslashtirish jarayonida asosiy qadam bo‘lib, lingvistik va grammatik ma’lumotlarni saqlash, tarjima sifatini oshirish va turli tillar bo‘yicha tadqiqotlarni osonlashtirishga yordam beradi. Bu tabiiy tillarni qayta ishlashda turli vazifalarni bajarish va ilovalari uchun zarur bo‘lgan parallel matnlar o‘rtasida to‘g‘ri moslashtirishni ta’minlashda muhim.

Quyida o‘zbek va ingliz tillari parallel matnlardagi sifat va uning darajalarining morfologik tahlilini ko‘rib chiqamiz.

**II. “Aligner” dasturiy vositasi uchun o‘zbek va ingliz tilida sifatlarning morfologik tahlili**

Sifatlar tuzilishiga ko‘ra tub va yasama bo‘ladi. Yasovchi qo‘shimchasi mavjud bo‘lmagan belgi bildiruvchi so‘zlar tub sifat sanaladi. Masalan, *xunuk*, *go‘zal*, *oq*, *qora* kabi. Yasovchi qo‘shimchalar yordamida boshqa so‘z turkumlaridan hosil qilingan sifatlar yasama sifat hisoblanadi. Tub sifatlar va ayrim yasama sifatlarni lug‘atdan topish oson, ammo hamma yasama sifatning tarjimasini o‘zbek-ingliz tili lug‘atlarida mavjud emas. Bunday hollarda moslashtiruvchi dasturiy vositarda so‘zlarning orasidagi moslikni aniqlashda bir muncha qiyinchiliklar yuzaga keladi. Sifat morfologik va sintaktik usulda yasaladi. Ayni shu jihatdan ularning morfologik va sintaktik tahlili bu muammoni hal qilishga yordam beradi. Morfologik usulda so‘z o‘zak, negiziga maxsus qo‘shimchalar qo‘shish orqali sifat yasaladi [1]. Ingliz tili flektiv tillardan bo‘lganligi tufayli bitta o‘zakka turli qo‘shimchalar qo‘shish orqali uning shakli va ma‘nosini o‘zgartirish ko‘p uchraydi. Ingliz tilida maxsus sifat yasovchi qo‘shimchalar bo‘lib, otdan, fe‘ldan sifat yasaydi va ular o‘zbek tilidan farqli ravishda gapning istalgan qismida turli gap bo‘laklarini ifodalab kelishi mumkin. Bunday qo‘shimchalar o‘zbek tilida ham ko‘plab uchraydi: *-ma*, *-chi*, *-li*, *-dor*, *-ch*, *-i(sh)*, *-ar*, *-ik*, *-ish*, *-qin*, *-choq*, *-a(y)*, *-siz* kabi qo‘shimchalarni bunga misol qilib keltirish mumkin [1]. Ammo tarjima yoki parallel matnlarni moslashtirish jarayonida sifatlarni ravishdan farqlash bir muncha qiyin bo‘ladi, chunki o‘zbek tilida sifatlar otlarni aniqlab kelsa, ingliz tilida sifatlar otlardan tashqari holat fe‘llari bilan ham qo‘llaniladi. Ayni murakablikni hal qilishda sifatlarni morfologik tahlil qilish eng yaxshi yechimdir. Masalan, *We have not done a dangerous task yet* gapida *dangerous* yasama sifat bo‘lib **danger** so‘ziga – **ous** qo‘shimchasini qo‘shish orqali yasalgan. Bu misolda *dangerous* so‘zini sifat ekanligini topish oson, chunki u gapda otdan oldin uning belgisini ifodalab kelgan, ammo *The situation is getting much more dangerous* misolida u fe‘l bilan birikib kelgan. Ushbu holatda gapda qo‘llanilgan *dangerous* so‘zining sifat ekanligini ikkita tahlil orqali aniqlash mumkin:

1) morfologik tahlil orqali, ya‘ni *danger* o‘zagiga sifat yasovchi qo‘shimcha qo‘shilganligi va ravish yasovchi – *ly* qo‘shimchasi yo‘qligi orqali;

2) grammatik tahlil orqali, ya‘ni ingliz tilida holat fe‘llari o‘zidan keyin ravish emas, sifat oladi degan qoidaga asoslanib.

## 1-jadval. Ingliz tilida sifat yasovchi qo'shimchalar

|        |                    |       |                     |
|--------|--------------------|-------|---------------------|
| -able  | <i>comfortable</i> | -ical | <i>historical</i>   |
| -al    | <i>accidental</i>  | -ious | <i>victorious</i>   |
| -ant   | <i>reluctant</i>   | -ish  | <i>childish</i>     |
| -ar    | <i>circular</i>    | -ist  | <i>racist</i>       |
| -ary   | <i>imaginary</i>   | -ive  | <i>attractive</i>   |
| -ate   | <i>passionate</i>  | -less | <i>careless</i>     |
| -some  | <i>wholesome</i>   | -like | <i>businesslike</i> |
| -ent   | <i>dependent</i>   | -ly   | <i>friendly</i>     |
| -esque | <i>picturesque</i> | -ory  | <i>compulsory</i>   |
| -ful   | <i>careful</i>     | -ous  | <i>dangerous</i>    |
| -ian   | <i>Italian</i>     | -y    | <i>lucky</i>        |
| -ible  | <i>horrible</i>    |       |                     |
| -ic    | <i>historic</i>    |       |                     |

Bundan tashqari ingliz tilida qo'shma sifatlarning yasalishi ham o'zbek tilidagi qo'shma sifatlardan birmuncha farq qilib, mutloqo sifat qatnashmasdan ham yasalishi mumkin. Misol uchun, ***It is impossible to drive a broken-down car*** misolida **broken-down** (buzilgan) so'zi *break* fe'lining o'tgan tugallangan (Verb3) shakliga predlog qo'shilishi orqali yasalgan bo'lib, ikkita mustaqil ma'noga ega so'zlar hisoblanadi, ammo o'zbek tiliga tarjima qilingan bitta "**buzilgan**" so'ziga to'g'ri keladi. Bu holat moslashtirish jarayonida so'zlar miqdoridagi tafovut hosil bo'layotganini va ushbu holatga asosiy sabab ingliz tilidagi qo'shma sifatlarning yasalishida mustaqil so'z turkumlarini birga qo'llab sifat birikmasini hosil qilinishidir. Buni Aligner dasturida quyidagicha ko'rsatish mumkin:

| Ingliz tilida                                      | O'zbek tilida                                     |
|--|---|
| It is impossible to drive a <b>broken-down</b> car | <b>Buzilgan</b> mashinani haydashning imkoni yo'q |

Quyida ingliz tilidagi qo'shma sifatlarning yasalish strukturasi ko'rib chiqamiz.

$$N+V_{ing} \quad (1)$$

Bu yerda *N*- ot, *V<sub>ing</sub>* – fe'ning birinchi shakliga **-ing** qo'shimchasini qo'shish orqali yasalgan bo'lib, ot va fe'ning birikishi orqali qo'shma sifat hosil bo'ladi. Masalan, **a time-consuming task (vaqt talab qiluvchi vazifa)**. Ushbu shaklda yasalgan qo'shma sifatlarda so'z birikmalari orasida " - " bo'lishi shart, aks holda, gapning ma'no-



sida keskin o‘zgarish yuzaga keladi. Buni quyidagi misoldan aniqlash mumkin:

| Ingliz tilida                              | O‘zbek tilida   |
|--|---|
| <i>I saw a <b>man-eating</b> alligator</i> | <i>Men <b>odamxo‘r</b> timsohni ko‘rdim</i>                   |
| <i>I saw a <b>man eating</b> alligator</i> | <i>Men <b>bir odamning timsohni yeyayotganini</b> ko‘rdim</i> |

Birinchi misoldagi **a man-eating alligator** timsohning belgisini ifodalab, *odamxo‘r* degan ma’noda kelgan, ikkinchi misoldagi **a man eating alligator** esa *qandaydir bir shaxs timsohni istemol qilayotganini* ifodalagan. B yerda *N* gapda sof ot bo‘lib,  $V_{ing}$  esa gerund bo‘lib kelgan: ish-harakatni ifodalab to‘ldiruvchi vazifasida kelgan.

$$A_{dv} + V_{ing} \quad (2)$$

Ushbu shaklda  $A_{dv}$  ravish,  $V_{ing}$  – fe’ning birinchi shakliga **-ing** qo‘shimchasini qo‘shish orqali yasalgan bo‘lib, ravish va fe’ning birikishi orqali qo‘shma sifat hosil bo‘ladi. Bunda **-ing** qo‘shimchasi sifat holatining davomiyligini ifodalaydi. Misol uchun: *a never-ending story* (nihoyasiz hikoya), *ever-lasting love* (boqiy muhabbat).

$$A_{dv} + V_3^{(ed)} \quad (3)$$

Ushbu shaklda  $A_{dv}$  ravish,  $V_3$  - fe’ning o‘tgan tugallangan shakli, ravish va fe’ning birikishi orqali qo‘shma sifat hosil bo‘ladi. Bunda **-ed** qo‘shimchasi sifat holatining tugallanganligini ifodalaydi. Misol uchun: *a well-known writer* (mashhur yozuvchi), *a poorly-built house* (sifatsiz uy).

$$V_3^{(ed)} + P \quad (4)$$

Ushbu shaklda  $V_3$  – fe’ning o‘tgan tugallangan shakli, **P** – esa predlog, fe’l va predlog birikishi orqali qo‘shma sifat hosil qilingan. Fe’ning o‘tgan tugallangan shakli ham ikki xil ko‘rinishda yasaladi:

1) o‘zakdan o‘zgarish orqali;

2) fe’lga **-ed** qo‘shimchasini qo‘shish orqali, har ikkala holatda ham fe’l va predlog o‘rtasida “-” bo‘lishi lozim.

Masalan, *worn-out shoes* (eskirgan poyabzal), *a broken-down car* (buzulgan mashina).

$$CN + N_{(s)} \quad (5)$$

Ushbu shaklda CN – son,  $N_{(s)}$  - ot bo‘lib, hech qanday qo‘shimchasiz birlik shaklda yasalishi lozim. Masalan, **“a three-day holiday”** (**uch**

*kunlik ta'til*)da uch soni ko'plik ma'nosini bersa ham, sifat vazifasini bajarganligi uchun unga birikib kelgan ot ham birlikda yasaladi va qo'shma sifatni hosil qiladi.

### 2-jadval. O'zbek tilidagi qo'shma sifatlarning yasalishi

|               |  |
|---------------|--|
| Sifat+ ot     | yapoloqyuz, xomkalla, sho`r peshona, kaltafahm, ochofat, maydagap, shikastahol, shirinsuxan, sovuqqon;   |
| Ot+sifat:     | boshqarong`u, yoqavayron, jig`ibiyron, tepakal, xonavayron, xudobexabar, otabezori, dilxasta, nonko`r;   |
| Ot+ot:        | bodomqovoq, sheryurak, darveshsifat, devqomat, dilozor, dilrom, dilpora, izzattalab, kafangado, otashnafas, sohibjamol, jigar rang, havo rang; |
| Ravish+ot:    | hozirjavob, kamgap, kamsuxan, kamsuqum, kamqon, kamxarj;   |
| Fe'l+fe'l:    | yebto`ymas;  |
| Ravish +fe'l: | tezpishar, kechpishar, cho`rtkesar;  |
| Olmosh+ot:    | o`zboshimcha   |
| Ot+fe'l:      | tilyog`lama, gadoytopmas, tinchliksevar, jonkuyar;   |
| Olmosh+sifat  | o`zbilarmon;   |
| Son+ot:       | ikkiyuzlamachi, qirqyamoq, sakkizoyoq;   |

Ingliz tilidan farqli ravishda o'zbek tilida qo'shma sifatlarning hech qanday punktuatsion belgilar yordamida ajratilmaydi, aksincha, qo'shib yoziladi [2]. Misol uchun: **man-eating = odamxo`r**

**Xulosa.** O'zbek va ingliz tillarida sifatlarning morfologik tahlili bir necha sabablarga ko'ra tabiiy tillarni qayta ishlash va mashinani o'rganish dasturlarida moslashish jarayoni uchun zarur. Chunki ikki tildagi sifatning o'ziga xos maxsus morfologik belgisining yo'qligi, sifat yasovchi affikslarning nihoyatda turli-tumanligi, ba'zan sifatlarni aniqlashda ma'lum qiyinchilik tug'diradi. Asosan, yasama sifatlarning lug'atlarda to'liq shaklda mavjud emasligi tarjima va moslashtirish jarayonida ko'plab qiyinchiliklarga sabab bo'ladi. Shu tufayli sifat so'z turkumining algoritmini tuzishda uchrashi mumkin bo'lgan barcha holatlarning lisoniy qolipi aniqlanishi, omonimlik keltiruvchi qo'shimchalar va so'zlar bazasi shakllantirilishi kerak.

## ADABIYOT

1. Abdullayeva. O.X., Nizomova F.B. “O‘zbek tilida sifat so‘z turkumini modellashtirishda so‘z yasalishi va omonimiya masalasi”, international scientific-theoretical conference on the topic: «Problems of research and education of the Uzbek language» [www.myscience.uz](http://www.myscience.uz), 2023.

2. Muqimova Gulnora Rashidovna, “Hozirgi o‘zbek adabiy tilida sifatlarning semantikfunksional xususiyatlar” dissertatsiya ishi. 2018.

3. Stowell.T. “The alignment of arguments in adjective phrases”, Brill, 1991.

4. Hang.S. Hungston. S. “Adjective complementation patterns and judgement: Aligning lexical-grammatical and discourse-semantic approaches in appraisal research”, Text & Talk, 27 May 2019.

5. Ambarita.E. “Morphological Analysis Of Adjective Reduplications In Toba Batak Language”. Jurnal Penelitian Pendidikan Bahasa dan Sastra. 2020.

6. Brinton, Laurel J. “The Structure of Modern English: A Linguistic Introduction”. Amsterdam: John Benjamin’s Publishing Company, 2000.

7. <https://osf.io/3w2g4/>.

8. [https://www.academia.edu/94404615/Morphological\\_and\\_Syntactical\\_Features\\_of\\_Adjectives\\_in](https://www.academia.edu/94404615/Morphological_and_Syntactical_Features_of_Adjectives_in).

9. <https://oxfordre.com/linguistics/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-558>

10. [https://www.grammar-monster.com/lessons/adjectives\\_compound\\_adjectives.htm](https://www.grammar-monster.com/lessons/adjectives_compound_adjectives.htm).

УДК 81'322.2

**КЛАССИФИКАЦИЯ ФЕЙК-НОВОСТЕЙ С ИСПОЛЬЗОВАНИЕМ  
МОРФОЛОГИЧЕСКИХ ТЕГОВ И N-ГРАММ****Б. Б. Элов, Н. У. Худайбергенов, З. Ю. Хусаинова***Ташкентский государственный университет узбекского языка  
и литературы им. Алишера Навои**Ташкент, Узбекистан*e-lov@navoiy-uni.uz, nizomaddin@navoiy-uni.uz,  
xusainovazilola@navoiy-uni.uz

Научиться эффективно выявлять фейковые новости в социальных сетях в настоящее время очень важная и актуальная задача. Методы выявления изучаются во многих областях исследований, включая морфологический анализ. Исследователи данных методов утверждают, что простых n-грамм, связанных с контентом, и POS-тегов недостаточно для классификации фейковых новостей. Однако за последнее десятилетие не получено результатов каких-либо эмпирических исследований, которые могли бы экспериментально подтвердить эти утверждения. Основной целью статьи является описание технологий экспериментальной оценки возможностей совместного использования **n-грамм** и **POS-тегов** для решения задач корректной классификации фейковых и реальных новостей. Нами были идентифицированы и дополнительно проанализированы n-граммы и POS-теги в текстах корпуса. Три метода, основанные на POS-метках различных групп n-грамм, были предложены и применены на этапе предварительной обработки обнаружения фейковых новостей. Для этого сначала проверялся размер n-грамма. На основе обнаруженных n-грамм определялась оптимальная глубина деревьев решений для достаточного обобщения. Наконец, производительность моделей, основанных на предложенных методах, сравнивалась со стандартизированными значениями TF-IDF. Показатели **эффективности модели, такие как точность, полнота и f1-оценка**, проверялись несколько раз. Также подробно исследован вопрос о том, можно ли улучшить метод TF-IDF с помощью POS-меток. Результаты исследования показали, что новый предложенный метод дает более точные результаты по сравнению с традиционным методом TF-IDF. В заключение можно сказать, что морфологический анализ может улучшить базовый TF-IDF метод.

**Ключевые слова:** обнаружение фейковых новостей, интеллектуальный анализ текста, обработка естественного языка, морфологический разметка, морфологический анализ.

## FAKE NEWS CLASSIFICATION USING MORPHOLOGICAL TAGS AND N-GRAMS

**Botir Elov, Nizomaddin Xudayberganov, Zilola Xusainova**

*Alisher Navoi' Tashkent State University*

*of the Uzbek Language and Literature*

*Tashkent, Uzbekistan*

e-elov@navoiy-uni.uz, nizomaddin@navoiy-uni.uz,

xusainovazilola@navoiy-uni.uz

Today, learning how to effectively identify fake news in social networks is a very important and urgent task. These methods are studied in many research areas, including morphological analysis. Some NLP researchers argue that simple content-related n-grams and POS tagging are insufficient to classify fake news. However, they have not received any empirical research results that could experimentally confirm these statements in the last decade. Considering this contradiction, the main goal of the paper is to experimentally evaluate the possibilities of general use of n-grams and POS tagging for correct classification of fake and real news. The n-grams of the POS tags of the corpus texts were identified and further analyzed. Three methods based on POS tagging of different groups of n-grams were proposed and applied in the preprocessing stage of fake news detection. For this purpose, the size of n-gram was checked first. Based on the detected n-grams, the optimal depth of the decision trees was determined for sufficient generalization. Finally, the performance of the models based on the proposed methods was compared with the standardized TF-IDF values. Performance indicators of the model, such as precision, recall and f1-score, were checked several times. Also, the question of whether the TF-IDF method can be improved using POS tagging was investigated in detail. The results of the study showed that the newly proposed method recorded more accurate results compared to the traditional TF-IDF technique. In conclusion, it can be said that morphological analysis can improve the basic TF-IDF method.

**Keywords:** Fake news detection, intelligent text analysis, natural language processing, POS tagging, morphological analysis.

## MORFOLOGIK TEG VA N-GRAMMLAR VOSITASIDA SOXTA YANGILIKLARNI TASNIFLASH

**Elov Botir Boltayevich, Xudayberganov Nizomiddin Uktambay o'g'li,  
Xusainova Zilola Yuldashevna**

*Alisher Navoiy nomidagi Toshkent davlat o'zbek tili*

*va adabiyoti universiteti, Toshkent, O'zbekiston*

e-elov@navoiy-uni.uz, nizomaddin@navoiy-uni.uz,

xusainovazilola@navoiy-uni.uz

Bugungi kunda ijtimoiy tarmoqlardagi soxta yangiliklarni samarali aniqlash usullarini o'rganish juda muhim va dolzarb vazifa hisoblanadi. Ushbu usullar

ko'plab tadqiqot sohalarida, jumladan morfologik tahlilda o'rganiladi. Ba'zi NLP tadqiqotchilarning ta'kidlashicha, oddiy kontent bilan bog'liq n-gramlar va POS teglash orqali soxta yangiliklarni tasniflash uchun etarli emas. Biroq, ular so'nggi o'n yillikda bu bayonotlarni eksperimental ravishda tasdiqlashi mumkin bo'lgan hech qanday empirik tadqiqot natijalarini olmaganlar. Ushbu qarama-qarshilikni hisobga olgan holda, maqolaning asosiy maqsadi soxta va haqiqiy yangiliklarni to'g'ri tasniflash uchun n-gramlar va POS teglashdan umumiy foydalanish imkoniyatlarini eksperimental baholashdan iborat. Korpus matnlarini POS teglarning n-grammlari aniqlandi va keyinchalik tahlil qilindi. Soxta yangiliklarni aniqlashning dastlabki ishlov berish bosqichida n-grammlarning turli guruhlariga POS teglash asoslangan uchta usul taklif qilindi va qo'llanildi. Shu maqsadda n-gramm o'lchami birinchi bo'lib tekshirildi. Aniqlangan n-grammlar asosida yetarli darajada umumlashtirish uchun qaror daraxtlarining eng mos chuqurligi aniqlandi. Nihoyat, tavsiya etilgan usullarga asoslangan modellarning ishlash ko'rsatkichlari standartlashtirilgan TF-IDF qiymatlari bilan taqqoslandi. Aniqlik (precision), recall va f1-score kabi modelning samaradorlik ko'rsatkichlari bir necha marta tekshirildi. Shuningdek, TF-IDF usulini POS teglash yordamida yaxshilash mumkinmi degan savol batafsil o'rganildi. Tadqiqot natijalari shuni ko'rsatdiki, yangi taklif qilingan metod an'anaviy TF-IDF texnikasi bilan solishtirilganda aniqroq ko'rsatkichlarni qayd etdi. Xulosa sifatida morfologik tahlilning asosiy TF-IDF metodini yaxshilashi mumkinligini aytish mumkin.

**Kalit so'zlar:** Soxta yangiliklarni aniqlash, matnni intellektual tahlil qilish, tabiiy tilni qayta ishlash, POS teglash, morfologik analiz.

## Kirish

Soxta yangiliklar hozirda rivojlangan dunyoning eng katta muammolaridan biri hisoblanadi [JANG, 2018]. Shaxsiy yoki siyosiy manfaatlar uchun yolg'on ma'lumot yoki yolg'on xabarlarini tarqatish, albatta, yangilik bo'lmasa-da, ijtimoiy media kabi hozirgi tendentsiyalar har bir shaxsga har qachongidan ham oson yolg'on ma'lumot yaratish imkonini beradi [Allcott, 2017]. Maqolada morfologik tahlildan foydalangan holda o'zbek tilidagi soxta va haqiqiy yangiliklarni tasniflash uchun to'rtta taklif qilingan modelni baholash haqida so'z boradi.

Morfologik analiz tabiiy tilni qayta ishlash tadqiqotining asosiy vositalaridan biridir. U kontekstdagi so'zning morfologik xususiyatlari sifatida POS teglari bilan bog'liq bo'lib, ilmiy tadqiqotlarda **uslubga asoslangan usul** sifatida ta'riflangan [Zafarani, 2019]. Lingvistik funksiyalar orqali matn tarkibidan *belgilar*, *so'zlar*, *gaplar* va *hujjatlar* kabi turli darajadagi strukturlangan ma'lumotlar aniqlanadi. Gap darajasidagi funksiyalar gaplar miqyosiga asoslangan barcha muhim atributlarni aniqlaydi. Bu turdagi funksiyalarga *POS teglar*, *gapning o'rtacha uzunligi*, *tvit/postning o'rtacha uzunligi*, *tinish belgilarining chastotasi*, *gapdagi manoga ega so'z birikmalari* va *iboralar*, *gapning*

*o'rtacha qutbliligi (ijobiy, neytral yoki salbiy)*, yoki *gapning murakkabligini aniqlash* kabilarni misol sifatida keltirish mumkin [Khan, 2021].

Mavjud ilmiy tadqiqot ishlarida asosan, noto'g'ri (yolg'on) ma'lumotlarning ichki xususiyatlarini aniqlash maqsadida standart lingvistik xususiyatlarni, jumladan **leksik**, **sintaktik**, **semantik** va **diskurs** xususiyatlarini o'rganadi. Sintaktik xususiyatlarni *POS teglar*, *tinish belgilar* va *chuqur sintaktik chastotalar* kabi guruhlariga ajratish mumkin [5;172]. Ushbu maqolada n-grammga asoslangan POS teglash orqali o'zbek tilidagi matnning morfologik analiz qilish orqali soxta yangiliklarni tasniflash masalasi ko'rib chiqiladi.

**N-gramm** – N ta token (so'z)lardan iborat ketma-ketligidir. Matndagi N-grammlar *ko'p so'zli iboralar* yoki *leksik birliklar* sifatida aniqlanadi. Quyidagi so'z birikmalari mos ravishda 2- va 3-grammni ifodalaydi: “*Amir Temur*”, “*Katta Buxoro kanali*”. Ko'p hollarda matndagi alohida so'zlarni (tokenlarni) tahlil qilishdan ko'ra **N-grammlarni tahlil qilish** samarali natijalarni qaytaradi. Ba'zi ilmiy tadqiqot ishlarida oddiy kontent bilan bog'liq n-grammlar va POS teglash usulining tasniflash vazifasi uchun yetarli emasligi isbotlangan [Feng, 2012]. Biroq, bu asosan mualliflarning fikrini aks ettiradi xolos. Chunki ular so'nggi o'n yillikda ushbu bayonotlarni tasdiqlovchi hech qanday empirik tadqiqot natijalarini tushunmagan yoki nashr etmagan.

Ushbu qarama-qarshilikni hisobga olgan holda, maqolaning asosiy maqsadi soxta va haqiqiy yangiliklarni to'g'ri tasniflash uchun n-grammlar va POS teglashdan foydalanish imkoniyatlarini eksperimental baholashdan iborat. Shu sababli, POS teglashning (n-gramm) berilgan namunasidan **n** ta elementning uzluksiz ketma-ketligi tahlil qilindi. Ushbu maqsadga erishish uchun POS teglariga asoslangan usullar taklif qilingan va ishlatilgan. Keyingi qadamlarda, ushbu usullar matn xususiyatlarini baholash uchun standartlashtirilgan mos yozuvlar TF-IDF metodi bilan taqqoslandi. Shuningdek, TF-IDF metodi natijasi samaradorligini va aniqligini POS teglash usuli yordamida yaxshilash mumkinligi batafsil o'rganiladi [Conroy, 2015]. Maqolada keltiriladigan barcha usullarni matnga dastlabki ishlov berish bosqichida qo'llash mumkin. Olingan natijalar to'plami qarorlar daraxti tasniflagichlari yordamida tahlil qilinadi.

Maqola tanlangan klassifikatorning kirish vektorlarini oldindan qayta ishlash uchun tavsiya etilgan usullarni taqdim etish va baholashga qaratilgan. Ushbu usullar matn POS teglaridan foydalangan holda n-grammlarni shakllantirishga asoslangan.

Barcha tavsiya etilgan usullar n-grammning turli darajalariga qo‘llanilgan bo‘lib, ushbu usullarning qarorlar daraxti tasniflagichining kirish vektorlari sifatida ishlatilgan. POS teglashning n-grammlarga asoslangan taklif qilingan yondashuvning muvofiqligini baholash uchun quyidagi metodologiya qo‘llanilgan:

– *Tahlil qilingan ma‘lumotlar to‘plamida POS teglarni aniqlash;*  
 – *POS teglaridan foydalanib N-gramm (1 gramm, 2 gramm, 3 gramm, 4 gramm) larni aniqlash. N-gramm POS teglar ketma-ketligini ifodalaydi.*

– *Hujjatlarda n-grammlar chastotasini hisoblash. Ya‘ni, tekshirilgan soxta va haqiqiy yangiliklarda n-grammning nisbiy chastotasi hisoblanadi.*

– *POS teglash va boshqariladigan TF-IDF metodi uchun tavsiya etilgan uchta usuldan foydalangan holda tasniflagichlarning kirish vektorlarini aniqlash;*

– *Qarorlar daraxti tasniflagichlarini qo‘llash. N-grammlarning turli chuqurliklari va uzunligi bo‘yicha parametrlarni sozlash;*

– *Qaror daraxtlarining xususiyatlarini, asosan daraxtlarning aniqligi, chuqurligi va vaqt ko‘rsatkichlarini aniqlash va taqqoslash.*

## POS teglar

Ma‘lumotlar to‘plamidagi yangiliklardagi barcha so‘zlariga TreeTagger nomli vosita yordamida morfologik teglar tayinlangan. Schmid 1994 yilda English Penn Treebank deb nomlangan teglar to‘plamini ishlab chiqqan. Ingliz tilidagi to‘liqroq Penn Treebank teglar to‘plami 35 ta morfologik tegni o‘z ichiga oladi [10;8,11;56]. Biroq, tadqiqot maqsadini hisobga olgan holda, ba‘zi teglar paydo bo‘lishining past chastotasi yoki nomuvofiqligi sababli keyingi tahlillarga kiritilmagan. Shuning uchun tahlilda foydalanilgan morfologik teglarning yakuniy ro‘yxati quyidagi 1-jadvalda keltirilgan [11;53,12;43,13;58]:

### *1-jadval. Yangiliklarni tasniflashda foydalaniladigan morfologik teglar*

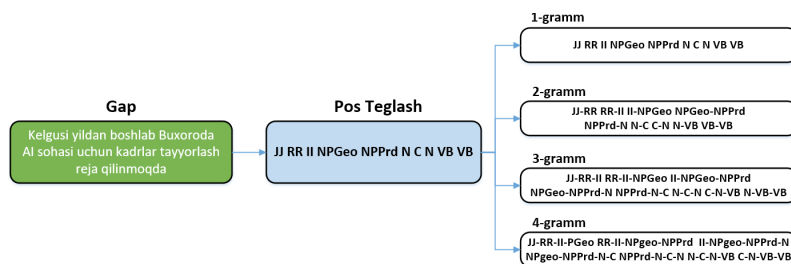
| GTAG    | Pos teglar                                      |
|---------|---|
| C guruh | C (bog‘lovchi), NUM (sanoq son)                 |
| E guruh | EX (biror narsani tasdiqlash uchun ishlatiladi) |
| F guruh | NP (Neologizm)                                  |
| I guruh | II (ko‘makchi so‘zlar)                          |



| GTAG    | Pos teglar  |
|---------|---|
| J guruh | JJ (sifat), JJR (sifat, Qiyosiy daraja), JJT (sifat, orttirma daraja)   |
| M guruh | MD (modal)  |
| N guruh | N (ot, birlik), N (ko‘plikdagi ot), NP (atoqli ot, nomlar), NP (atoqli ot, ko‘plik)   |
| P guruh | PInd (belgilash olmoshi), PossP (egalik kategoriyasi), PP (kishilik olmoshi)  |
| R guruh | RR (ravish), RRR (ravish, qiyosiy), RRT (ravish, orttirma), Prt (yuklama)   |
| T guruh | (fe’lning lug‘at shakli “moq”)  |
| U guruh | UH (undov so‘zlar)  |
| V guruh | VB (fe’l), TPast (fe’l, o‘tgan zamon), VBS (fe’l, sifatdosh) Prs3s (fe’l, birlik, hozirgi zamon), VBZ (fe’l, 3-shaxs birlik. Hozirgi zamon) |
| W guruh | PQues (aniqlovchi so‘zlar, so‘roq olmoshi)  |

## POS teglardan N-grammni aniqlash

Ushbu bosqichda ma’lumotlarni boshlang‘ich qayta ishlash POS teglaridan N-grammlar olingan. Natijada, POS teglarining berilgan namunasiidan n-gramm ketma-ketligi shakllantirildi. Quyidagi 1-rasmda bu jarayon 2019-yilda Facebookda baham ko‘rilgan eng ko‘p ko‘rilgan o‘ninchi soxta yangiliklardan olingan gaplardan foydalangan holda ko‘rsatilgan.



1-rasm. Soxta yangilikni POS teglash va N-grammlarga ajratish

Yuqorida keltirilgan “Kelgusi yildan boshlab Buxoroda AI sohasi uchun kadrlar tayyorlash reja qilinmoqda” gapiga mos aniqlangan POS teglar quyidagicha:

- **JJ** (Sifat),
- **RR** (Ravish),
- **II** (Ko‘makchi),
- **NPGeo** (Geografik nom),
- **NPPrd** (Mahsulot nomi),
- **N** (Ot),
- **C** (Bog‘lovchi),
- **VB** (Fe‘l).

1-gramm va aniqlangan POS teglar bir xil bo‘lganligi sababli, keyingi tadqiqotlarda ishlatiladigan 1-grammli kirish fayli aniqlangan POS teglari bilan bir xil bo‘ladi. TF-IDF usuli uchun n-grammlar xuddi shu tarzda shakllantirilgan. Ammo shuni ta’kidlash kerakki, ushbu usulda so‘zning **lemmalari** yoki **stemlarini** ifodalovchi terminlar ishlatilgan.

#### *Kirish vektorlarini boshlang‘ich qayta ishlash usullari*

Tanlangan klassifikator uchun kirish vektorlarini boshlang‘ich qayta ishlash uchun quyidagi to‘rtta usul qo‘llanilgan.

**Term frequency - inverse document frequency** (TF-IDF) usulidan tokenlarning korpus hujjatlaridagi ahamiyatini baholash uchun foydalaniladi [7]. TF-IDF yondashuvi odatda “*shovqin*” sifatida identifikatsiya qilinadigan ma’lum bir domen bilan yuqori darajada bog‘liq bo‘lgan ko‘p ishlatiladigan terminlarani aniqlash uchun ishlatiladi. An’anaviy TF-IDF usuli katta hajmdagi ma’lumot (yangilik) larni qayta ishlash uchun qo‘llanilmaydi. Odatda, TF-IDF og‘irligi ikki elementdan iborat: birinchisi terminning normalangan chastotasi (Term Frequency, TF), ikkinchisi esa teskari hujjat chastotasi (Inverse Document FrequencyIDF). Quyidagi belgilanishlarni aniqlab olamiz:

- **t** – termin/so‘z;
- **d** – hujjat;
- **w** – hujjatdagi istalgan termin.

**d** hujjatdagi **t** termin/so‘zning chastotasi quyidagi formula orqali hisoblanadi:

$$tf(t, d) = \frac{f(t, d)}{f(w, d)}$$

Bu yerda **f(t, d)** – **d** hujjatdagi termin/so‘zlar soni va **f(w, d)** – hujjatidagi barcha terminlar soni. Shuningdek, TF-IDF qiymatni hisoblashda ma’lum bir termin/so‘z sodir bo‘lgan barcha hujjatlar soni ham hisobga olinadi. Bu qiymat **idf(t, D)** kabi belgilanadi va uning qiymati quyidagi formula orqali aniqlanadi:

$$\text{idf}(t, D) = \ln \frac{N}{\sum(d \in D: t \in d) + 1}$$

Bu yerda,  $D$  – hujjatlar korpusi va  $N$  – korpusdagi hujjatlar soni. **Tf-Idf** qiymati quyidagi formula orqali aniqlanadi:

$$\mathbf{Tf - Idf}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D)$$

**Tf** formulasi turli xil variantlarga ega bo‘lib, masalan

$$\mathbf{log}(\text{tf}(t, d)) \text{ yoki } \mathbf{log}(\text{tf}(t, d)) + 1$$

Xuddi shunday, **idf** qiymatni hisoblash mumkin bo‘lgan bir nechta variantlari mavjud. Yuqorida keltirilgan formulalarga mos **Tf-Idf** qiymatni hisoblash uchun Pythondagi **scikit-learn**<sup>1</sup> kutubxonasidan foydalanish mumkin. Tajribada qo‘llaniladigan TF-IDF usuli orqali tavsiflangan tanlangan xususiyatlarini taqqoslash uchun ishlatiladi. Xuddi shu ma’lumotlar to‘plami keyingi qadamlar uchun kirish sifatida ishlatilgan. Biroq, bu holatda nomuhim so‘zlari olib tashlangan.

#### *POS chastotasi (PosF) usuli*

Ushbu usul Term Frequency usuliga o‘xshashidir. Biroq, u POS teglar chastotasi bilan hisoblab chiqadi. Quyidagi belgilanishlarni aniqlab olamiz:

- **pos** – identifikatsiyalangan POS teg;
- **d** – hujjat;
- **w** – hujjatda aniqlangan har qanday POS teg.

Bu holda **d** hujjatidagi **POS** teglar chastotasini quyidagicha hisoblash mumkin:

$$\mathbf{PosF}(pos, d) = \frac{f(pos, d)}{f(w, d)} \quad (3)$$

Bu yerda  $f(pos, d)$  – **d** hujjatdagi POS teglar soni va  $f(w, d)$  – hujjatidagi barcha POS teglar soni. Yuqorida keltirilgan 3-formula orqali hujjatda aniqlangan POS teglarining tahlil qilingan ro‘yxati doirasidagi har bir POS tegining nisbiy chastotasini ifodalaydi.

#### *PosF-IDF usuli*

Ushbu usul TF-IDF usulining analogidir. Yuqorida keltirilgan PosF usuliga o‘xshab, u alohida so‘zlar va gaplar asosida tahlil qilingan ma’lumotlar to‘plamidagi har bir hujjatda aniqlangan POS teglarini

<sup>1</sup> <https://scikit-learn.org>

ko‘rib chiqadi. Faqat identifikatsiyalangan POS teglardan iborat hujjatlar PosF-IDFni hisoblash uchun kerakli ma‘lumotlar hisoblanadi. Hujjatdagi POS teglarining nisbiy chastotasidan tashqari, ma‘lum bir POS tegi aniqlangan barcha hujjatlar soni ham hisobga olinadi.

#### *TF-IDF va PosF usullarini birlashtirish*

Ushbu usul an‘anaviy TF-IDF usulini POS teglash yordamida yaxshilash mumkinmi yoki yo‘qligini tasdiqlash uchun ishlab chiqilgan. Shu maqsadda har bir hujjat uchun quyidagi vektorlar shakllantirilgan:

– *Tf-Idf* vektori;

– *Hujjatdagi POS teglarining nisbiy chastotasini ifodalovchi PosF vektori.*

Yuqorida keltirilgan  $\overline{Tf - Idf(d)}$  va  $\overline{PosF(d)}$  vektorlarni birlashtirilgan natijasida yangi ***merge(d)*** vektori hosil qilinadi ( $m \leq n$ ).

$$\begin{aligned}\overline{Tf - Idf(d)} &= (t_1, t_2, \dots, t_n), \\ \overline{PosF(d)} &= (p_1, p_2, \dots, p_m), \\ \overline{merge(d)} &= (t_1, t_2, \dots, t_n, p_1, p_2, \dots, p_m),\end{aligned}$$

Ma‘lumotlarni tasniflashdagi kirish vektorlarini boshlang‘ich qayta ishlash uchun bir qator usullar ishlab chiqildi. Ushbu usullarni avvalgi TF-IDF usulining modifikatsiyasi deb hisoblash mumkin bo‘lib, bunda POS teglari asl terminlarga qo‘shimcha ravishda hisobga olinadi. Natijada, yuqorida tavsiflangan to‘rtta usul tipik o‘zgarishlarni ifodalash orqali terminlar va POS teglar asosida hosil qilingan gibrid usul orqali asosiy xususiyatlarini taqqoslash va tahlil qilish imkonini beradi.

#### **Qaror daraxtlari vositasida modellashtirish**

Bugungi kunda ma‘lumotlarni tasniflash uchun quyidagi tasniflagichlardan foydalanish mumkin:

– *qarorlar daraxti tasniflagichlari (tree classifiers);*

– *Bayes klassifikatorlari (Bayesian classifiers);*

– *eng yaqin k-qo‘shni tasniflagichlari (k-nearest-neighbour classifiers);*

– *holatlarga asoslangan fikrlash (case-based reasoning);*

– *genetik algoritmlar (genetic algorithms);*

– *qo‘pol to‘plamlar (rough sets);*

– *oshkormas mantiq usullari (fuzzy logic techniques).*

Kirish vektorlarini hisoblash uchun tavsiya etilgan usullarning mosligini baholash va ularning xususiyatlarini tahlil qilish uchun *qaror daraxtlari (decision trees)* usulini ko'rib chiqamiz. Qaror daraxtlari nafaqat ishlarni oddiy tasniflash imkonini beradi, balki ular bir vaqtning o'zida oson izohlanadigan va tushunarli tasniflash qoidalarini yaratadi. Xuddi shu yondashuv Kapusta, Benko va Munk tadqiqot ishlarida qisman qo'llanilgan.

Qarorlar daraxti yaratilayotganda qo'llaniladigan *ma'lumot olish, Jini indeksi* kabi atributlarni tanlash ko'rsatkichlari muhim omili hisoblanadi. Qaror daraxtini ishlab chiqishning har bir bosqichida eng yaxshi funksiya har doim tanlanadi. Ushbu funksiya kirish atributlari soniga bog'liq emas. Bu esa tanlangan klassifikatorning kiritilishida ko'proq atributlar (kirish vektorining elementlari) berilgan bo'lsa ham, aniqlik o'zgarishini anglatadi.

#### *K-o'Ichovli tekshiruv*

Amalga oshirilgan tajribada shakllantirilgan qaror daraxtlarini taqqoslash uchun qaror daraxtlarining *tugunlari* yoki *barglari* soni kabi muhim xususiyatlardan foydalaniladi. Bu xususiyatlar daraxtning o'lchamini anglatib, ularni mos ravishda kamaytirish kerak.

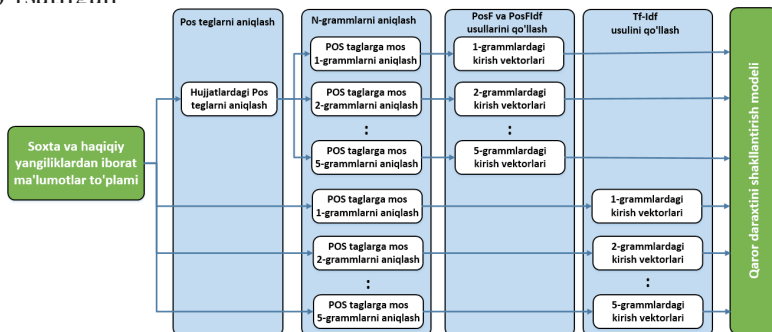
Bir vaqtning o'zida *precision, recall* va *f1-score* kabi modelning ishlash ko'rsatkichlari bir necha (10) marotaba tekshirish orqali sinovdan o'tkaziladi. Modellarni baholash uchun *K-o'Ichovli tekshiruv*dan foydalanilgan. Bu, odatda, boshqa usullarga nisbatan kamroq noxolis modelga olib keladi, chunki u asl ma'lumotlar to'plamidagi har bir kuzatuv mashg'ulot va test to'plamida paydo bo'lish imkoniyatiga ega bo'lishini ta'minlaydi.

#### *n-gramm uzunligini o'rnatish*

Yuqorida keltirilgan kirish vektorlarini boshlang'ich qayta ishlash usullarida umumiy shartlar talab etilgan. Shu sababli birinchi qadam sifatida n-grammdagi eng yuqori qiymatlar aniqlanadi. Ko'pgina NLP vazifalarida odatda  $n = \{1,2,3\}$  oraliqdagi qiymatlardan foydalaniladi.  $n$  ning yuqori qiymati (4 gramm, 5 gramm va boshqalar) apparat va dasturiy ta'minotga, hisoblash vaqtiga va umumiy ishlashga sezilarli murakkabliklarni yuzaga keltiradi. Boshqa tomondan, yaratilgan modellarning aniqligini oshirishda yuqori n-grammlarning potensial hissasi cheklangan. Yuqoridagi fikrlarni tasdiqlash uchun bir nechta qarorlar daraxti modellari ishlab chiqildi. *Tokenlar/so'zlar* va *POS teglar* uchun N-grammlar (1 gramm, 2 gramm, ..., 5 gramm) tayyorlandi. Keyinchalik, TF-IDF usuli n-gramm tokenlar/so'zlar

uchun qo‘llanildi. Shu bilan birga, n-gramm POS teglarida **PosF** va **PosfIdf** usuli qo‘llanildi.

Natijada, kirish vektorlaridan iborat **15** ta fayl yaratildi (*1-5 gramm × 3 usul*). Quyidagi 2-rasmda ushbu jarayonning individual bosqichlari ko‘rsatilgan



2-rasm. Kirish vektorlari tajriba jarayoni

Har bir boshlang‘ich qayta ishlangan 15 ta fayl ustida 10 marta o‘tkazilgan sinov natijasida o‘nta qaror daraxti modeli shakllantirildi. Barcha holatlarda **aniqlik darajasi** sifatida modelning samaradorlik o‘lchovi hisoblandi.

Amalga oshirilgan sinov natijalari shuni ko‘rsatadiki, aniqlik darajasi n-gramm uzunligi bilan, asosan, TF-IDF usulini qo‘llashda pasayadi. Cheklangan vaqt va hisoblash jarayonining murakkabligi tufayli kattaroq o‘lchamli n-grammlarni (6 gramm, 7 gramm va boshqalar) qayta ishlash imkonini mavjud emas. Shu sababli tadqiqotda **n = 5** maksimal chegara sifatida qabul qilingan.

Qarorlar daraxti modelini ishlab chiqish jarayonida n-grammlarni bitta kirish fayliga birlashtirish eng yuqori aniqlikni ta‘minlaganini qayd etish lozim. Natijada, keyingi qadamdagi barcha tajribalar 1-gramm, 2-gramm, 3-gramm va 4-gramm (1,4) dan iborat fayl ustida amallar bajariladi.

Amalga oshirilgan natijalariga ko‘ra ijtimoiy tarmoqlardagi soxta yangiliklarni samarali aniqlash uchun quyidagi ketma-ketlikda amallari bajarish lozim:

1. *Ma'lumotlar to'plamidagi POS teglarni aniqlash.*
2. *Aniqlangan POS teglar uchun PosF va PosfIdf kirish vektorlarini qo'llash.*
3. *Kirish vektorlarini aniqlash uchun TF-IDF usulini qo'llash.*

Ushbu usul orqali stemming algoritmi tomonidan o'zak so'zlar aniqlanadi. Shuningdek, ushbu qadamda nomuhim so'zlari olib tashlanadi.

4. PosF va TF-IDF metodlari orqali vektorlarni birlashtirish.

5. Quyida keltirilgan amallarni maksimal chuqurlikning turli qiymatlari bilan takrorlash (1...n):

– PosF, PosFidf, TfIdf va Merge metodlarining kirish vektorlarini 10 martalik o'zaro tekshirish talablariga muvofiq o'qitish va to'plam ostilariga tasodifiy taqsimlash.

– Berilgan maksimal chuqurlikka ega bo'lgan har bir mashg'ulot to'plami uchun qarorlar daraxtini hisoblash.

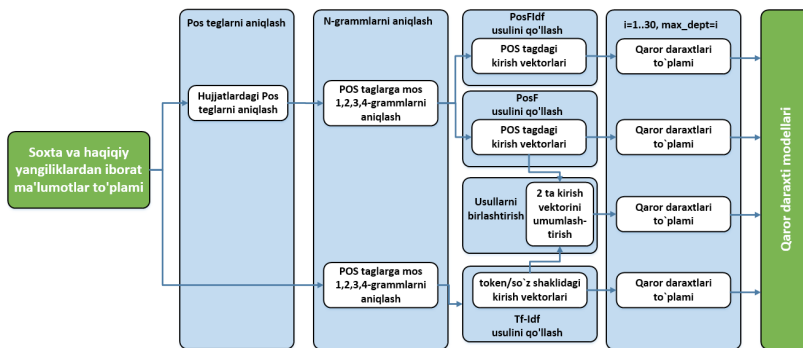
Sinov kichik hajmli ma'lumotlar to'plamida model bashoratlarining sifatini sinab ko'rish. Quyidagi xarakteristikalar o'rnatiladi:

- *prec\_fake* (soxta guruh uchun aniqlik);
- *prec\_real* (haqiqiy guruh uchun aniqlik);
- *rec\_fake* (soxta guruh yozuvi);
- *rec\_real* (haqiqiy guruh yozuvi);
- *f1-ball/baho*;
- *har bir iteratsiyaga sarflangan vaqt*.

– Natijalarni tahlil qilish (modellarni baholash).

1–4 bosqichlar natijasi yuqorida aytib o'tilgan to'rtta kirish vektoridir. Taklif etilayotgan metodologiyaning beshinchi bosqichi ushbu to'rtta tekshirilgan qadamni/natijani baholashga qaratilgan.

Taklif etilayotgan metodologiyani 10 marta o'zaro tekshirish bilan qo'llash natijasida 1200 ta turli xil qarorlar daraxtlari yaratildi. Quyidagi 3-rasmda tajriba metodologiyasining alohida bosqichlari tasvirlangan.



3-rasm. Tavsiya etilgan usul bosqichlari

## Xulosa

Ushbu maqolada n-grammalar orqali til korpusidan yaratilgan POS teglashga asoslangan ijtimoiy tarmoqlardagi soxta yangiliklarni ishonchli tasniflash usullari tahlil qilindi. Maqolada POS teglashga asoslangan ikkita usul taklif qilindi va TF-IDF usuli asosida o'zaro solishtirildi. Olingan natijalar PosF va TF-IDF usuli o'rtasidagi statistik jihatdan ahamiyatsiz farqlarni ko'rsatdi. Bu farqlar barcha kuzatilgan ishlash ko'rsatkichlarida, jumladan, *accuracy*, *precision*, *recall* va *f1-score* asosida solishtirildi. Shu sababli, morfologik tahlilni soxta yangiliklar tasnifiga qo'llash mumkin degan xulosaga kelish mumkin. Shuningdek, tavsiflovchi statistik jadvallar TF-IDF usuli statistik jihatdan ahamiyatsiz bo'lsa-da, yaxshiroq natijalarga erishishini ko'rsatdi. Morfologik tahlilga asoslangan usullar zamonaviy ma'lumotlar to'plamida, shu jumladan o'zbek tili korpusidagi 1100 ta haqiqiy va yolg'on yangiliklarda sinovdan o'tkazildi va samarali natija qaytardi.

## FOYDALANILGAN ADABIYOTLAR RO'YXATI

1. Jang, S. M., Geng, T., Queenie Li, J. Y., Xia, R., Huang, C. T., Kim, H., & Tang, J. (2018). A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior*, 84. <https://doi.org/10.1016/j.chb.2018.02.032>
2. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. In *Journal of Economic Perspectives* (Vol. 31, Issue 2). <https://doi.org/10.1257/jep.31.2.211>
3. Zafarani et al. (2019) Zafarani R, Zhou X, Shu K, Liu H. Fake news research: theories, detection strategies, and open problems. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19; New York, NY, USA: Association for Computing Machinery; 2019. pp. 3207–3208.*
4. Khan, J. Y., Khondaker, Md. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4. <https://doi.org/10.1016/j.mlwa.2021.100032>
5. Feng, S., Banerjee, R., & Choi, Y. (2012). Syntactic stylometry for deception detection. *50th Annual Meeting of the Association for Computational Linguistics, ACL 2012 – Proceedings of the Conference*, 2.
6. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1).
7. <https://doi.org/10.1002/pra2.2015.145052010082>



9. B.Elov, Z.Xusainova, N.Xudayberganov. O'zbek tili korpusi matnlari uchun TF-IDF statistik ko'rsatkichni hisoblash. *SCIENCE AND INNOVATION INTERNATIONAL SCIENTIFIC JOURNAL VOLUME 1 ISSUE 8 UIF-2022: 8.2 ISSN: 2181-3337*

10. [https://www.academia.edu/105829396/OZBEK\\_TILI\\_KORPUSI\\_MATNLARI\\_UCHUN\\_TF\\_IDF\\_STATISTIK\\_KORSATKICHNI\\_HISOBLASH](https://www.academia.edu/105829396/OZBEK_TILI_KORPUSI_MATNLARI_UCHUN_TF_IDF_STATISTIK_KORSATKICHNI_HISOBLASH)

11. B.Elov, Sh.Khamroeva, Z.Xusainova (2023). The pipeline processing of NLP. *E3S Web of Conferences 413, 03011, INTERAGROMASH 2023*. <https://doi.org/10.1051/e3sconf/202341303011>

12. Boltayevich, E. B., Mirdjonovna, H. S., & Ilxomovna, A. X. (2023). Methods for Creating a Morphological Analyzer. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13741 LNCS*. [https://doi.org/10.1007/978-3-031-27199-1\\_4](https://doi.org/10.1007/978-3-031-27199-1_4)

13. Kaing, H., Ding, C., Utiyama, M., Sumita, E., Sam, S., Seng, S., Sudoh, K., & Nakamura, S. (2021). Towards Tokenization and Part-of-Speech Tagging for Khmer: Data and Discussion. *ACM Transactions on Asian and Low-Resource Language Information Processing, 20(6)*. <https://doi.org/10.1145/3464378>

14. B.Elov, Sh.Hamroyeva, O.Abdullayeva, M.Uzoqova. O'zbek tilida pos tegging masalasi: muammo va takliflar. *O'zbekiston: til va madaniyat (Amaliy filologiya), 2022, 5(4)*.

15. B.Elov, Sh.Hamroyeva, O.Abdullayeva, Z.Xusainova, N.Xudayberganov. O'zbek, turk va uyg'ur tillarida pos teglash va stemming. *O'zbekiston: til va madaniyat (Kompyuter lingvistikasi), 2023, 1(6)*.

16. B.Elov, E.Adali, Sh.Khamroeva, O.Abdullayeva, Z.Xusainova, N.Xudayberganov. The Problem of Pos Tagging and Stemming for Agglutinative Languages. *8 th International Conference on Computer Science and Engineering UBMK 2023, Mehmet Akif Ersoy University, Burdur – Turkey*.

УДК 81'322.2

**ОБРАБОТКА КОРПУСНЫХ ТЕКСТОВ УЗБЕКСКОГО ЯЗЫКА  
МЕТОДАМИ WORD2VEC, GLOVE, ELMO, BERT**

**Б. Б. Элов, Р. Х. Алаев, З. Ю. Хусайнова, А. У. Юлдашев**  
*Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои  
Ташкент, Узбекистан*

e-elov@navoiy-uni.uz, mr.ruhillo@gmail.com,  
xusainovazilola@navoiy-uni.uz, yuldashevaziz@navoiy-uni.uz

В данной статье описаны методы машинного обучения Word2Vec, GloVe, ELMO, BERT для обработки текстовых данных и особенности их применения к обработке текстов на узбекском языке. Благодаря методам дискретного представления текста каждое слово в корпусе считается уникальным и преобразуется в числовую форму на основе одного из приведенных методов. В статье дается анализ этих методов для разработки современных NLP приложений на основе CNN и LSTM методов.

**Ключевые слова:** корпус узбекского языка, обработка текста, Word2Vec, GloVe, ELMO, BERT.

**METHODS OF PROCESSING UZBEK LANGUAGE CORPUS TEXTS  
WORD2VEC, GLOVE, ELMO, BERT**

**Botir Elov, Ruhillo Alayev, Zilola Xusainova, Aziz Yuldashev**  
*Alisher Navoi' Tashkent State University of the Uzbek Language  
and Literature, Tashkent, Uzbekistan*

e-elov@navoiy-uni.uz, mr.ruhillo@gmail.com,  
xusainovazilola@navoiy-uni.uz, yuldashevaziz@navoiy-uni.uz

**Abstract.** This article describes how to process data in the form of symbols, words, and text, as well as the application of Word2Vec, GloVe, ELMO, BERT methods to the Uzbek language from the methods of teaching a computer to process natural language. Through Discrete Text Representation methods, each word in the corpus is considered unique and converted into a numerical form based on the various methods discussed above. The article presents several advantages and disadvantages of the different methods. Currently, these methods are used in the development of modern NLP applications based on CNNs and LSTMs.

**Keywords:** Uzbek language corpus, text processing, Word2Vec, GloVe, ELMO, BERT.

**O‘ZBEK TILI KORPUSI MATNLARINI QAYTA ISHLASH  
WORD2VEC, GLOVE, ELMO, BERT USULLARI**

*Elov B. B., Alayev R. H., Xusainova Z. Y., Yuldashev A. U.*

*Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va  
adabiyoti universiteti, Toshkent, O‘zbekiston*

*e-elov@navoiy-uni.uz, mr.ruhillo@gmail.com,*

*xusainovazilola@navoiy-uni.uz, yuldashevaziz@navoiy-uni.uz*

Mazkur maqolada ma’lumotlar belgilar, soʻzlar va matnli shaklda boʻlganda ularni qanday qayta ishlash, kompyuterni tabiiy tilni qayta ishlashga oʻrgatish usullaridan Word2Vec, GloVe, ELMO, BERT usullarining oʻzbek tiliga qoʻllanishi tavsiflanadi. Matnni diskret koʻrinishlari usullari orqali korpusdagi har bir soʻz unikal deb hisoblanadi va yuqorida muhokama qilgan turli usullarga asoslanib, sonli shaklga aylantiriladi. Maqolada keltirilgan turli xil usullarning bir-biriga oʻxshash bir nechta afzalliklari va kamchiliklari keltirildi. Hozirgi vaqtda ushbu usullar CNN va LSTMlarga asoslangan zamonaviy NLP ilovalarini ishlab chiqishda qoʻllanilmoqda.

**Kalit soʻzlar:** Oʻzbek tili korpusi, matnlarini qayta ishlash, Word2Vec, GloVe, ELMO, BERT.

*Word2Vec usuli*

Word2Vec – bu soʻzlarni joylashtirish uchun mashhur algoritm. Ushbu algoritim 2013-yilda Tomas Mikalov tomonidan “Vektor fazosida soʻzlarning ifodalanishini samarali baholash” tadqiqoti ostida ishlab chiqilgan [Cahyani, 2021]. Metod soʻzlarni ifodalashning bashorat qilishga asoslangan.

**Soʻzlar bogʻliqligi (Word embeddings)** – bu soʻzning vektor koʻrinishi boʻlib, har bir soʻzning boshqa soʻzlar bilan semantik va sintaktik aloqasini hisobga olgan holda, belgilangan vektor oʻlchami bilan ifodalanadi. Word2vec arxitekturasi bitta yashirin qatlamli tarmoqdir. Yashirin qatlamning ogʻirligi *soʻzning yoʻqotish funksiyasi (normal backprop)* orqali aniqlanadi.

Ushbu arxitektura avtokoderga oʻxshab, bu yerda kodlovchi va dekoder qatlami mavjud boʻlib, oʻrta qism oʻlchamlarni kamaytirish yoki anomaliyalarni aniqlashda foydalanish uchun ishlatilishi mumkin boʻlgan kirishning siqilgan koʻrinishidir. **Word2vec** usuli orqali korpus tasviri 2 xil usulda amalga oshiriladi [Xiong, 2019; Jang, 2019]:

**CBOW** – atrofda soʻz konteksti asosida oraliq soʻzni taxmin qilishga asoslangan. CBOW usulida kontekst/atrofda soʻzlarni hisobga olgan holda qaysi soʻz koʻproq mos kelishi haqida boʻsh joylarni toʻldirishga harakat qilinadi. Ushbu usul kichikroq maʼlumotlar toʻplamlari bilan samarali natijani beradi.

**Skip-Gram** – maqsadli soʻzdan (CBOWga teskari) atrofidagi kontekst soʻzlarini taxmin qilishga harakat qiladi. Kattaroq hajmdagi maʼlumotlar toʻplamida yaxshiroq natija beradi. Biroq oʻquv maʼlumotlar toʻplamini qayta ishlashga koʻp vaqt sarflanadi.

Word2vec usuli orqali vektor arifmetikasidan foydalangan holda soʻzlar oʻrtasidagi oʻxshashlik darajasi aniqlanadi. **“Man is to woman as king is to queen”** kabi shablondan arifmetik amallar orqali **“king”** - **“man”** + **“woman”** = **“queen”** kabi natijaga ega boʻlish mumkin. Shuningdek, **“queen”** soʻzi hozirgi va oʻtgan zamon kabi sintaktik va semantik munosabatlarni ifodalaydi.

**Gensim** paketi yordamida word2vec usulini koʻrib chiqamiz:

```
from gensim.models import Word2Vec
sentences = ['Men NLP bilan ishlayman', 'NLP juda ajoyib',
             'NLP - bu mashinalarga tabiiy tilni qayta ishlashga imkon berish-
             dir',
             'bu misol nlp texnikasiga namuna']
# gapni oldindan qayta ishlash Word2Vec uchun zarur formatga
aylantirish
sentence_list=[]
for i in sentences:
    li = list(i.split(" "))
    sentence_list.append(li)
model = Word2Vec(sentence_list, min_count=1,
                 workers=4, sg=1, window=4)
model.wv['nlp']
model.wv.most_similar(positive=['nlp'])
```

```
[('imkon', 0.24666069447994232),
 ('Men', 0.11936754733324051),
 ('ajoyib', 0.11928389966487885),
 ('ishlashga', 0.11663015931844711),
 ('texnikasiga', 0.09614861011505127),
 ('bu', 0.08543577790260315),
 ('ishlayman', 0.07172605395317078),
 ('tilni', 0.05970853567123413),
 ('mashinalarga', 0.04119439423084259),
 ('-', 0.012471411377191544)]
```

Yuqoridagi dastur kodining bir necha satrida biz nafaqat soʻzlar-ni vektor sifatida oʻrgatish va koʻrsatish imkoniyatiga egamiz, balki oʻxshash va turli soʻzlarni aniqlashimiz mumkin. Vektorlar oʻrtasidagi oʻxshashlikni aniqlashning ikki yoʻli mavjud:

**1. Normallashtirilgan:** vektorlar orasidagi skalyar ko'paytmani hisoblab, ularni o'xshashligini aniqlash mumkin.

**2. Normallashtirilmagan:** vektorlar orasidagi kosinus o'xshashligini quyidagi formuladan foydalanib hisoblash mumkin:

$$\text{cosine similarity} = 1 - \text{cosine distance} = \frac{u * v}{||u|| * ||v||}$$

Korpus asosida raqamli vektorlarni aniqlash, korpusni mashinali o'qitish va so'zlar o'rtasidagi munosabatlarni aniqlash algoritmlari keyingi ilmiy nashrlarda ko'rib chiqiladi. Word2Vec usulining afzalliklari va kamchiliklari quyidagi jadvalda keltirilgan:

| Afzalliklari   | Kamchiliklari   |
|--|---|
| Turli so'zlar o'rtasidagi sintaktik va semantik munosabatlarni aniqlashga imkon beradi | Lug'atdan tashqari so'zlar (OOV)ni qayta ishlab bo'lmaydi.                            |
| So'zga mos raqamli vektorining o'lchami kichik va moslashuvchan.                       | So'zning semantik ifodasi faqat qo'shnilariga asoslanadi.                             |
| Korpusni o'qitish jarayoni inson omiliga bog'liq emas.                                 | Word2Vec usulini yangi tabiiy tilga qo'llash uchun juda ko'p amallarni bajarish lozim |
|  | Ma'lumotlar aniqligi osjishi uchun kattaroq korpus talab qilinadi.                    |

### *GloVe usuli*

**Global Vectors yoki qisqacha GloVe usuli** so'zlarni raqamli ifodalashning zamonaviy NLP usulidir. Ushbu usul Jefferi Pennington, Richard Socher va Kristofer Manning tomonidan 2014-yilda ishlab chiqilgan va joriy etilgan [5]. Yuqorida keltirilgan Word2vec usulidan farqli ravishda ushbu usulda so'zning lokal va global statistikasi o'rganiladi hamda so'zlarni ifodalash uchun **gibrid yondashuv** deb ataladi. GloVe usulida quyidagi belgilashlardan foydalaniladi:

$$v_i^T v_j = \log P(i|j)$$

yoki,

$$v_i^T v_j = \log P(X_{ij}) - \log P(X_i)$$

Shunday qilib,  $P(i|j)$ ga mos tarzda  $V_i$  va  $V_j$  so'z vektorlar qiymatlari hisonlanadi. Ushbu vektorlar birgalikda joylashish matritsada global statistik ma'lumotni ifodalaydi. O'zbek tilidagi matnlarni korpusda GloVe usulida qayta ishlash maqsad va vazifalari alohida tadqiqotni talab qiladi. Katta hajmdagi matnlarni oldindan o'rgatilgan

modellardan asosida GloVe vektorlarini shakllantirish va ulardan foydalanish quyida keltirilgan:

```
import gensim.downloader as api
# 2 milliard tvitning 25 o'Ichamli GloVe tasvirini yuklab olish
twitter_glove = api.load("glove-twitter-25")
# O'xshash so'zlarni topish
print(twitter_glove.most_similar("book",topn=10))
# 25D vektorlarni olish
print(twitter_glove['book'])
print(twitter_glove.similarity("book", "school"))
```

```
[('books', 0.94181889295578), ('project', 0.9214614033699036),
 ('review', 0.9140495657920837), ('script', 0.9069417119026184),
 ('new', 0.9069172143936157), ('feature', 0.8995184302330017),
 ('guest', 0.897861659526825), ('read', 0.8931056261062622), ('post',
 0.8916701674461365), ('art', 0.8880472183227539)]
```

```
[ 0.21621  0.056781  0.82955 -0.1424  0.82832 -0.87341  1.699
 -0.25702  0.65303 -0.82435  0.26496  0.4612  -4.0463
 -0.044556
  0.15648 -0.083655  0.72399  0.20802 -0.27561 -0.024987
 -0.83992
 -0.92536 -0.95454  0.42348 -0.14709 ]
```

0.7545484

GloVe usulining afzallikalri va kamchiliklari quyidagi jadvalda keltirilgan:

| Afzalliklari   | Kamchiliklari  |
|--|--|
| Word2vec usuliga qaraganda yaxshiroq ishlaydi  | Birgalikda yuzaga keladigan matritsa va global ma'lumotdan foydalanganligi sababli, GloVe usulida word2vec usuliga qaraganda ancha ko'p xotira talab etiladi |
| Vektorlarni qurishda so'z juftligi va so'z juftligi munosabatini inobatga oladi                                | Word2vec usuliga o'xshab, u ko'p ma'noli so'zlar muammosini hal qilmaydi   |
| Word2Vec bilan solishtirganda GloVe usulini parallellashtirish osonroq, shuning uchun o'rgatish vaqti qisqaroq |  |

## Zamonaviy yondashuvlar

### ELMO usuli

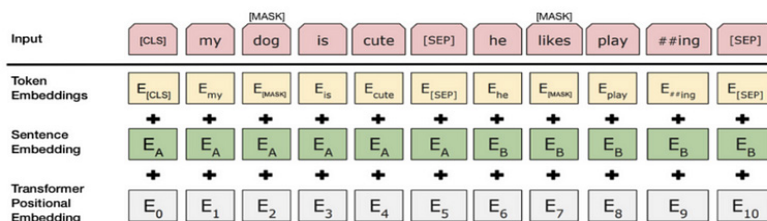
2018-yilda Metyu Peters va boshqalar chuqur kontekstli soʻz koʻrinishlari nomidagi maqolasi taqdim etildi [6]. U taklif qilgan usulda Word2vec va GloVe usullari kamchiliklarini vektor koʻrinishi va u ifodalovchi soʻz oʻrtasida koʻpdan-bir munosabatga ega boʻlish orqali hal qilishga harakat qilinadi. ELMO usulida kontekstni inobatga olinib, soʻzning vektor koʻrinishini mos ravishda oʻzgartiriladi.

ELMO usulida soʻzlarni boshlangʻich soʻz vektorlariga aylantirish uchun belgilar darajasidagi CNNlardan foydalaniladi. Oʻqitish jarayonida qoʻshimcha ravishda ikki tomonlama LSTMlardan foydalaniladi. Usul orqali yoʻnaltirish va orqaga iteratsiya kombinatsiyasi mos ravishda soʻzdan oldin va keyingi kontekst maʼlumotlarini ifodalovchi oraliq soʻz vektorlarini yaratiladi. Boshlangʻich soʻz vektori va 2 ta oraliq soʻz vektorining vaznli yigʻindisi yakuniy qiymatni beradi.

### BERT usuli

**BERT** – 2019-yildagi **Google AI** jamoasining “Pre-training of Deep Bidirectional Transformers for Language Understanding” nomli maqolasida tavsiflangan tilni tushunish uchun chuqur ikki yoʻnalishli **transformerlar (Transformers)**ni oldindan oʻrgatish usuli hisoblanadi [7].

Bu usulda transformerlarni oldindan oʻrgatish uchun yangi oʻz-oʻzini nazorat qiladigan mashinali oʻrganish vazifasidir.



Source: BERT [Devlin et al., 2018], with modifications

### 1-rasm. BERT usuli arxitekturasi

BERT usulida til modelining ikki tomonlama kontekstidan foydalanadi. Bunda bashorat qilish vazifalari uchun ishlatiladigan oraliq tokenlarni yaratish uchun *chapdan oʻngga* va *oʻngdan chapga* “niqoblash”ga harakat qilinadi.

BERT modeliga kiritilgan maʼlumotlar tokenlarni joylashtirish, segmentatsiyalashdan iborat boʻlib, kontekstda soʻzni toʻgʻri bashorat

qilish uchun model uchun niqoblash strategiyasiga amal qiladi. BERT usuli orqali soʻzlar orasidagi kontekstual munosabatni oʻrganadigan va NER va savol-javob tizimlari kabi boshqa vazifalarni bajarish uchun sozlangan transformer tarmogʻidan foydalanadi.

*Matnni sonli koʻrinishi ilovalari*

Ushbu maqolda keltirilgan matnning sonli koʻrinishdagi modellarini quyidagi NLP vazifalariga qoʻllash mumkin:

**Matn tasnifi (Text Classification):** Matnni tasniflash vazifasida matnga boshlangʻich ishlov berish uchun matnni vektor koʻrinishida shakllantirish muhimdir.

**Mavzuni modellashtirish (Topic Modelling):** Mavzuni model-lashtirish vazifasida matnni turli mavzularda modellashtirish uchun toʻgʻri formatda taqdim etilishi talab qilinadi.

**Avtomatik tuzatish modeli (Autocorrect Model):** Avtomatik tuzatish modeli orqali matndagi imlo xatolari tuzatiladi. Avtomatik tuzatish modeli vositasida berilgan matnni talab qilingan sonli formatda taqdim etish kerak.

**Yangi matn generatsiya qilish (Text Generation):** Matnni generatsiya qilish uchun ehtimollarga asoslangan sonli matn formati talab qilinadi.

Mashinali oʻrganish modelini oʻrgatishdan oldin matnni maʼlum bir formatda ifodalash juda muhim. Format qanchalik murakkab boʻlsa, modelning aniqligi va natijalari shunchalik yaxshi boʻladi. Matnli maʼlumotlarini oʻz ichiga olgan har bir NLP ilovasi yaxshi matn koʻrinishini talab qiladi.

## XULOSA

Matnni diskret koʻrinishlari usullari orqali korpusdagi har bir soʻz unikal deb hisoblanadi va yuqorida muhokama qilgan turli usullarga asoslanib, sonli shaklga aylantiriladi. Maqolada keltirilgan turli xil usullarning bir-biriga oʻxshash bir nechta afzalliklari va kamchiliklari keltirildi. NLPdagi murakkab vazifalarini *taqsimlangan matn koʻrinishlari algoritmlari* vositasida hal qilish mumkin. Taqsimlangan matn koʻrinishlaridan til korpusni tushunish va oʻrganishda foydalanish mumkin. Korpus ichidagi soʻzlarni va ularning bir-biri bilan qanday bogʻlanishini oʻrganishni bunga misol sifatida keltirish mumkin. Bugungi kunda *Savol-javob tizimlari, hujjatlar tasnifi, chatbot, NER obyektini tanib olish* kabi murakkab NLP vazifalarini hal qilish uchun nazorat ostidagi oʻrganish modellarini ishlab chiqishda taqsimlangan



matn ko‘rinishlaridan keng miqyosida foydalanilmoqda. Hozirgi vaqtda ushbu usullar CNN va LSTMlarga asoslangan zamonaviy NLP ilovalarini ishlab chiqishda qo‘llanilmoqda.

#### ADABIYOT

1. Cahyani D.E., Patasik, I. (2021). Performance comparison of tf-idf and word2vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, 10(5). <https://doi.org/10.11591/eei.v10i5.3157>
2. Method N.W., Goldberg Y., Levy, O., Mikolov, T., Sutskever, I., Chen, K., Corrado G., Dean J. (2014). word2vec Explained : Deriving Mikolov et al. *ArXiv:1402.3722 [Cs, Stat]*, 2.
3. Xiong Z., Shen, Q., Xiong Y., Wang Y., Li W. (2019). New generation model of word vector representation based on CBOW or skip-gram. *Computers, Materials and Continua*, 60(1). <https://doi.org/10.32604/cmc.2019.05155>
4. Jang B., Kim I., Kim J.W. (2019). Word2vec convolutional neural networks for classification of news articles and tweets. *PLoS ONE*, 14(8). <https://doi.org/10.1371/journal.pone.0220976>
5. Pennington J., Socher R., Manning C.D. (2014). GloVe: Global vectors for word representation. *EMNLP 2014–2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. <https://doi.org/10.3115/v1/d14-1162>
6. Kutuzov A., Kuzmenko E. (2021). Representing ELMo embeddings as two-dimensional text online. *EACL 2021 – 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the System Demonstrations*. <https://doi.org/10.18653/v1/2021.eacl-demos.18>
7. Joshi M., Levy O., Weld D.S., Zettlemoyer L. (2019). BERT for coreference resolution: Baselines and analysis. *EMNLP-IJCNLP 2019–2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*. <https://doi.org/10.18653/v1/d19-1588>

УДК 81'322.2

## ТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ SVD И NMF

**Б. Б. Элов, А. У. Юлдашев, Н. Р. Алоев**

*Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои*

*Ташкент, Узбекистан*

elov@navoiy-uni.uz, yuldashevaziz@navoiy-uni.uz,  
vip.alayev@gmail.com

Целью тематического моделирования в области обработки естественного языка (NLP) является – идентификация абстрактных тем из набора документов. При тематическом моделировании на основе текстового корпуса из множества документов выявляются абстрактные темы текста. Тематическое моделирование – неконтролируемая задача для машинного обучения (Machine Learning) (ML). В данной статье рассматривается вопрос тематического моделирования текстов языкового корпуса методами SVD и NMF.

**Ключевые слова:** Термин-документ, тематическое моделирование, Обработка естественного языка, Усеченная сингулярная декомпозиция.

## TOPIC MODELING USING SVD AND NMF

**Botir Elov, Aziz Yuldashev, Narzillo Aloyev**

*Alisher Navoi' Tashkent State University of the Uzbek Language  
and Literature, Tashkent, Uzbekistan*

elov@navoiy-uni.uz, yuldashevaziz@navoiy-uni.uz,  
vip.alayev@gmail.com

The goal of topic modeling in the field of natural language processing (NLP) is to extract abstractly considered topics from a set of documents. Topic modeling uses a text corpus to identify abstract topics in a text from multiple documents. Topic modeling is an unsupervised task for Machine Learning (ML). This article discusses topic modeling of texts in a language corpus using SVD and NMF methods.

**Keywords:** NLP, SVD, NMF, term-document, topic modeling

## SVD VA NMF METODLARI ORQALI TEMATIK MODELLASHTIRISH

**Botir Elov, Aloyev Narzillo, Aziz Yuldashev**

*Alisher Navoiy nomidagi Toshkent davlat o'zbek tili  
va adabiyoti universiteti, Toshkent, O'zbekiston*

elov@navoiy-uni.uz, yuldashevaziz@navoiy-uni.uz,  
vip.alayev@gmail.com

Tabiiy tilni qayta ishlash (NLP) sohasida tematik modellashtirish nazoratsiz o'rganish vazifasi bo'lib, uning maqsadi hujjatlar to'plamidan mavhum hisoblangan mavzularni aniqlashdan iborat. Tematik modellashtirishda ko'p hujjatlarning matn korpusini hisobga olgan holda, matn haqidagi mavhum mavzular aniqlanadi. Tematik modellashtirish – Machine Learning (ML) uchun nazorat qilib bo'lmaydigan vazifa hisoblanadi. Ushbu maqolada til korpuysi matnlarini SVD va NMF metodlari orqali tematik modellashtirish masalasi ko'rib chiqiladi.

**Kalit so'zlar:** svd, nmf, tematik modellashtirish, hujjat-atama, nlp.

## Kirish

Bir nechta hujjatlardan iborat katta hajmdagi til korpusi berilgan bo'lsin. Har bir hujjatni o'qib chiqmasdan (tahlil qilmasdan) turib, berilgan hujjatlar to'plamidagi asosiy mavzularni aniqlash lozim bo'lsin. **Tematik modellashtirish** orqali til korpusidagi ma'lumotlarni ma'lum miqdordagi *mavzularga ajratiladi* [Vayansky, 2020; Churchill, 2022]. **Mavzular** – bu kontekstga o'xshash va hujjatlar to'plamidagi ma'lumotlarni ifodalaydigan *so'zlar guruhi* hisoblanadi. **M** ta hujjat va **N** ta termin (atama)dan iborat til korpusi uchun **“hujjat - atama”** matritsasi (Document-Term Matrix, DTM)ning umumiy tuzilishi quyida ko'rsatilgan [Pradana, 2020]:

|           |    | Terminlar |     |     |         |
|-----------|----|-----------|-----|-----|---------|
|           |    | T1        | T2  | T3  | TN      |
| Hujjatlar | D1 | w11       | w12 | w13 | ... w1N |
|           | D2 | w21       | w22 | w23 | ... w2N |
|           | D3 | w31       | w32 | w33 | ... w3N |
|           | .  | .         | .   | .   | .       |
|           | DM | wM1       | wM2 | wM3 | ... wMN |

**1-rasm.** M ta hujjat va N ta termin (atama)dan iborat jadval (matritsa)

Berilgan matritsani tahlil qilamiz:

–  $D_1, D_2, \dots, D_M$  – M hujjatlar;

–  $T_1, T_2, \dots, T_N$  – N atamalar.

“Hujjat-atama” matritsasini to'ldirish uchun keng qo'llaniladigan

**TF-IDF** usulidan foydalanamiz.

*TF-IDF baholash formulasi*

TF-IDF baholash quyidagi tenglama orqali aniqlanadi [Elov, 2022;

5]:

$$w_{ij} = TF_{ij} * \log\left(\frac{M}{df_j}\right)$$

Bu verda,

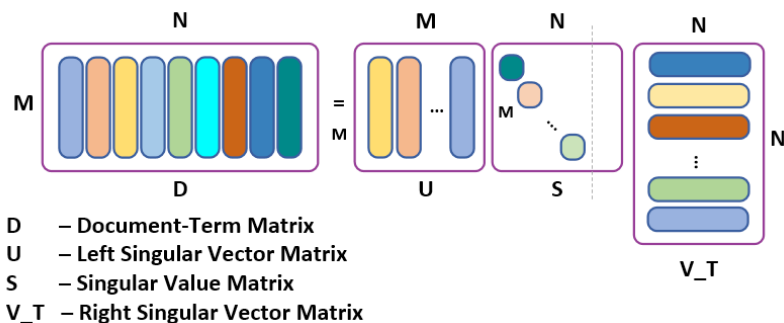
–  $\mathbf{TF}_{ij} - \mathbf{D}_i$  hujjatda  $\mathbf{T}_i$  atamasining uchrash soni;

–  $\mathbf{df}_j - \mathbf{T}_j$  atamasini o'z ichiga olgan hujjatlar soni.

Muayyan hujjatda ko'p ishlatilgan, biroq til korpusda kamdan-kam uchraydigan atama yuqori IDF bahosiga ega bo'ladi. Keyingi qadamda matritsalarini faktorizatsiya qilish usullarini ko'rib chiqiladi.

*Singular qiymatlarni ajratish (Singular Value Decomposition, SVD) yordamida tematik modellashtirish*

Tematik modellashtirishda SVDning ishlatilishi quyidagi 2-rasmda ko'rsatilgan[6]:



2-rasm. Singular qiymatlarni ajratish (SVD) yordamida tematik modellashtirish

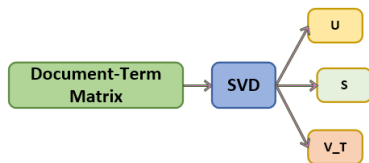
**D** “hujjat - atama” matritsasiidan **SVD** metodi yordamida quyidagi 3 ta matritsa hosil qilinadi:

– **U** – *chap singular vektor matritsasi*. Bu matritsa  $\mathbf{D} \cdot \mathbf{D}^T$  Gram matritsasining o'ziga xos bo'linishi orqali hosil qilinadi. Ko'p hollarda ushbu matritsa *hujjatlarning o'xshashlik matritsasi* deb ham nomlanadi. O'xshashlik matritsasining  $i, j$ -chi yozuvi  $i$  hujjat  $j$  hujjatiga qanchalik o'xshashligini anglatadi.

– **S** – *Singular qiymat matritsasi*. Ushbu matritsa mavzularning nisbiy ahamiyatini ifodalaydi.

– **V\_T** – *o'ng singulyar vektor matritsasi*. Shuningdek, ushbu matritsa *mavzu matritsasi* deb ham ataladi. Matndagi mavzular ushbu matritsaning satrlari bo'ylab joylashtiriladi.

**SVD** metodi orqali **D** “hujjat - atama” matritsasiidan 3 ta matritsa (**U**, **S** va **V\_T**) hosil qilinadi. Natijada hosil qilingan **V\_T** matritsasi qatorlarida mavzular joylashtiriladi.

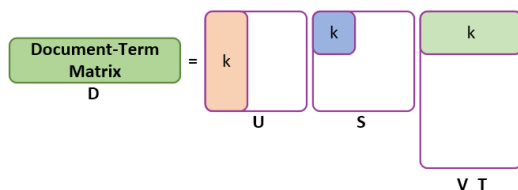


3-rasm. SVD metodi vositasida mavzuni modellashtiruvchidir

SVD metodi baz'i hollarda navbatida **Latent Semantic Indexing (LSI)** deb ham nomlanadi.

*Qisqartirilgan SVD yoki k-SVD*

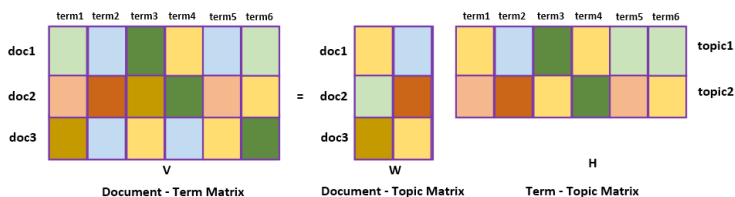
Aytaylik, **150** ta hujjatdan iborat til korpusi mavjud bo'lsin. Til korpusini tavsiflovchi **150** xil hujjatni yoki korpus mazmunini ifodalaydigan **10** ta mavzuni o'qish masalasini ko'rib chiqamiz. Matn mazmunini yaxshi yetqazib bera oladigan oz sonli mavzularni belgilab olish ko'pincha foydali hisoblanadi. **k-SVD** metodi vositasida ushbu vazifani bajarish mumkin. Katta o'lchamli matritsalarini o'zaro ko'paytirish katta sondagi murakkab amallarni talab qilganligi sababli, eng katta **k singular qiymatlarni** va ularga mos keladigan **mavzularni** tanlash afzaldir. **k-SVD** metodining ishlashi quyidagi 4-rasmda ko'rsatilgan [ELov, 2023; Ke, 2022; 8]:



4-rasm. k-SVD - eng yaxshi k-darajali approximatsiya

*Negativ bo'lmagan matritsali faktorizatsiya (Non-Negative Matrix Factorization, NMF) yordamida tematik modellashtirish*

NMF metodining ishlash prinsipi quyidagi 5-rasmda ko'rsatilgan [Alghamdi, 2015; Qiang, 2022; Kherwa, 2020]:



5-rasm. Negativ bo'lmagan matritsali faktorizatsiya

Bu yerda,

– **W** – **hujjat-mavzu (document-topic matrix)** matritsasi. Ushbu matritsa mavzularning korpusi hujjatlari bo'yicha taqsimotini ifodalaydi.

– **H** – **atama-mavzu (term-topic matrix)** matritsasi. Ushbu matritsa mavzular bo'yicha terminlarning qiymatini ifodalaydi.

NMF metodidagi **W** va **H** matritsalarining barcha elementlari manfiy emasligi sababli, korpusga qo'llash birmuncha soddaroq. Shu sababli, NMF metodi orqali natijaning aniqligi biroz yuqori.

NMF – *aniq bo'lmagan matritsalarini faktorizatsiya qilish (non-exact matrix factorization technique)* usuli bo'lib, **W** va **H** matritsalar ko'paytmasi orqali boshlang'ich **V** matritsani aniqlab bo'lmaydi.

Birinchi qadamda **W** va **H** matritsalar tasodifiy tarzda shakllantiriladi. NMF algoritmidagi qadamlar iterativ ravishda bajarilishi natijasida ushbu matritsa qiymatlari yangilanadi va *cost function (CF)* deb nomlanuvchi funksiya qiymatini minimallashtiradi. *CF* funksiyasi quyida ko'rsatilganidek, **V-W.H** matritsasining *Frobenius normasini* ifodalaydi:

$$\text{minimize } ||V - WH||_F$$

bu yerda,

– **V** – (*Document - Term Matrix*);

– **W** – (*Document - Topic Matrix*);

– **H** – (*Term - Topic Matrix*).

**MxN** o'lchovli **A** matritsaning Frobenius normasi quyidagi tenglama bilan aniqlanadi:

$$||A||_F = \sqrt{\sum_{i=1}^M \sum_{j=1}^N |a_{ij}|^2}$$

*SVD usuli orqali tematik modellashtirish bosqichlari*

SVD usuli orqali tematik modellashtirish uchun quyidagi qadamlarni amalga oshirish lozim.

**1-qadam.** Mavzularni aniqlashda SVD usulidan foydalanish uchun birinchi qadamda **matn korpusini aniqlab olish** lozim. Quyidagi kod katakchasi kompyuter dasturlash bo'yicha matn bo'lagini o'z ichiga oladi.

text=[“Computer programming is the process of designing and building an executable computer program to accomplish a specific computing result or to perform a specific task.”,

“Programming involves tasks such as: analysis, generating algorithms, profiling algorithms’ accuracy and resource consumption, and the implementation of algorithms in a chosen programming language (commonly referred to as coding).”

“The source program is written in one or more languages that are intelligible to programmers, rather than machine code, which is directly executed by the central processing unit.”

“The purpose of programming is to find a sequence of instructions that will automate the performance of a task (which can be as complex as an operating system) on a computer, often for solving a given problem.”

“Proficient programming thus often requires expertise in several different subjects, including knowledge of the application domain, specialized algorithms, and formal logic.”

“Tasks accompanying and related to programming include: testing, debugging, source code maintenance, implementation of build systems, and management of derived artifacts, such as the machine code of computer programs.”

“These might be considered part of the programming process, but often the term software development is used for this larger process with the term programming, implementation, or coding reserved for the actual writing of code.”

“Software engineering combines engineering techniques with software development practices.”

“Reverse engineering is a related process used by designers, analysts and programmers to understand and re-create/re-implement”]

**2-qadam.** Matn ma’lumotlari uchun **scikit-learn** paketidan **TfidfVectorizer** sinfini import qilish lozim:

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

Matn korpusi uchun TF-IDF ball (qiymat)lari bilan to’ldirilgan **V** matritsaga ega bo’lish uchun **TfidfVectorizer** sinfidan foydalaniladi:

**3-qadam.** Yuqorida muhokama qilingan Truncated SVD (k-SVD) dan foydalanish uchun **scikit-learn** paketinida **TruncatedSVD** sinfini import qilishingiz kerak:

```
from sklearn.decomposition import TruncatedSVD
```

Barcha zarur modullar import qilganidan so’ng, matndagi mavzularni qidirishga o’tish mumkin.

**4-qadam.** Ushbu bosqichda **Tfidfvectorizer** obyektini yaratish lozim:

```
vectorizer = TfidfVectorizer(stop_words='english', smooth_idf=True)
```

# kichik harflar, maxsus belgilarni, nomuhim so'zlarini olib tashlash

```
input_matrix = vectorizer.fit_transform(text).todense()
```

1–4 qadamlar natijasida quyidagi amallar bajarildi:

– matnlar to'plami jamlandi;

– zarur modullarni import qilindi;

– Document-Term Matrix matritsasi aniqlandi.

**5-qadam.** 3-qadamda import qilingan **TruncatedSVD** sinfidan foydalanamiz:

```
svd_modeling= TruncatedSVD(n_components=4, algorithm='randomized', n_iter=100, random_state=122)
```

```
svd_modeling.fit(input_matrix)
```

```
components=svd_modeling.components_
```

```
vocab = vectorizer.get_feature_names()
```

**6-qadam.** Korpus hujjatlariga mos mavzularni aniqlash:

```
topic_word_list = []
```

```
def get_topics(components):
```

```
    for i, comp in enumerate(components):
```

```
        terms_comp = zip(vocab, comp)
```

```
        sorted_terms = sorted(terms_comp, key= lambda x:x[1], reverse=True)[:7]
```

```
        topic=""
```

```
        for t in sorted_terms:
```

```
            topic= topic + ' ' + t[0]
```

```
        topic_word_list.append(topic)
```

```
        print(topic_word_list)
```

```
    return topic_word_list
```

```
get_topics(components)
```

**7-qadam.** Aniqlangan mavzularni va ularni mantiqiy to'g'ri shakllantirilganligini tahlil qilish. SVDdan olingan komponentlarni **get\_topics()** funksiyasiga parameter sifatida uzatib, *mavzular ro'yxatini* va ushbu mavzularning har biridagi *ommabop so'zlarni* aniqlanash:

Topic 1:

code programming process software term computer engineering



Topic 2:  
engineering software development combines practices techniques used

Topic 3:  
code machine source central directly executed intelligible

Topic 4:  
computer specific task automate **complex** given instructions

Yuqorida amalga oshirilgan 7 ta qadam natijasida 4 ta hujjatdagi mavzular aniqlandi. Aniqlangan mavzularni vizualizatsiya qilish uchun **word cloud** usulidan foydalanish mumkin. Ushbu usul orqali aniqlangan mavzular nisbiy ahamiyatiga ko'ra ko'rsatiladi. Har bir hujjatdagi eng muhim so'z eng katta shrift bilan ajratilgan.

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt
for i in range(4):
    wc = WordCloud(width=1000, height=600, margin=3, prefer_horizontal=0.7, scale=1, background_color='black', relative_scaling=0).generate(topic_word_list[i])
    plt.imshow(wc)
    plt.title(f"Topic {i+1}")
    plt.axis("off")
    plt.show()
```

Scikit-learn paketidagi NMF sinfi metodlari orqali korpus matnlarini **tematik modellashtirish** mumkin:

```
from sklearn.decomposition import NMF
NMF_model = NMF(n_components=4, random_state=1)
W = NMF_model.fit_transform(input_matrix)
H = NMF_model.components_
```

Keyingi qadamda **get\_topics()** metodi orqali **H** matritsasidagi mavzular ro'yxatini aniqlash mumkin:

Topic 1:  
code machine source central directly executed intelligible

Topic 2:  
engineering software process development used term combines

Topic 3:  
algorithms programming application different domain expertise formal

**Topic 4:**

computer specific task programming automate **complex** given

Berilgan korpus matnlari uchun SVD va NMF usullari o'xshash mavzular ro'yxatini qaytarishini ko'rish mumkin.

*SVD va NMF usullari farqlari*

Til korpusi matnlarini tematik modellashtirish uchun ushbu ikkita matritsani faktorizatsiya qilish usullari o'rtasidagi farqlarni keltiramiz:

– *SVD matritsalarini real faktorizatsiya qilish usulidir. SVD usuli natijasida olingan matritsalaridan kirish DTMni qayta hosil qilish mumkin;*

– *Agar korpusga k-SVD usuli qo'llangan bo'lsa, bu holda DTM kirishiga eng yaxshi k-darajali yaqinlashuv amalga oshiriladi;*

– *NMF usuli SVD usuliga qaraganda mavzularni aniqlash natijasi yuqori.*

**Xulosa**

Ushbu maqolada til korpusi hujjatlarini SVD va NMF metodlari orqali tematik modellashtirish masalasi ko'rib chiqildi va Python tilidagi sklearn paketi vositalari orqali dasturiy ta'minotni ishlab chiqish ketma-ket qadamlar namoyish etildi. Maqolada M ta hujjat va N ta termin (atama)dan iborat til korpusi uchun DTM (Document-Term Matrix) matritsasini shakllantirish uchun TF-IDF usulidan foydalanildi. Shunigdek, DTM matritsalarini faktorizatsiya qilishniung singular qiymatlarni ajratish (Singular Value Decomposition, SVD) va Negativ bo'lmagan matritsali faktorizatsiya (Non-Negative Matrix Factorization, NMF) usullari ko'rib chiqilgan va ularning yutuq va kamchiliklari keltirilgan.

**ADABIYOT**

1. Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, 94. <https://doi.org/10.1016/j.is.2020.101582>
2. Churchill, R., & Singh, L. (2022). The Evolution of Topic Modeling. *ACM Computing Surveys*, 54(10). <https://doi.org/10.1145/3507900>
3. Musthofa Galih Pradana. (2020). Penggunaan Fitur Wordcloud dan Document Term Matrix dalam Text Mining. *Jurnal Ilmiah Informatika*, 8(1).
4. B.Elov, Z.Xusainova, N.Xudayberganov. O'zbek tili korpusi matnlari uchun TF-IDF statistik ko'rsatkichni hisoblash. *SCIENCE AND INNOVA-*

TION INTERNATIONAL SCIENTIFIC JOURNAL VOLUME 1 ISSUE 8  
UIF-2022: 8.2 ISSN: 2181-3337

5. [https://www.academia.edu/105829396/OZBEK\\_TILI\\_KORPUSI\\_MATNLARI\\_UCHUN\\_TF\\_IDF\\_STATISTIK\\_KORSATKICHNI\\_HISOBLASH](https://www.academia.edu/105829396/OZBEK_TILI_KORPUSI_MATNLARI_UCHUN_TF_IDF_STATISTIK_KORSATKICHNI_HISOBLASH)

6. B.ELov, Sh.Khamroeva, Z.Xusainova (2023). The pipeline processing of NLP. *E3S Web of Conferences 413, 03011, INTERAGROMASH 2023*. <https://doi.org/10.1051/e3sconf/202341303011>

7. Ke, Z. T., & Wang, M. (2022). Using SVD for Topic Modeling. *Journal of the American Statistical Association*.

8. <https://doi.org/10.1080/01621459.2022.2123813>

9. Alghamdi, R., & Alfalqi, K. (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*, 6(1). <https://doi.org/10.14569/ijacsa.2015.060121>

10. Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X. (2022). Short Text Topic Modeling Techniques, Applications, and Performance: A Survey. In *IEEE Transactions on Knowledge and Data Engineering* (Vol. 34, Issue 3). <https://doi.org/10.1109/TKDE.2020.2992485>

11. Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24). <https://doi.org/10.4108/eai.13-7-2018.159623>

12. Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, 2. <https://doi.org/10.1016/j.caeai.2021.100016>

13. Wang, J., & Zhang, X. L. (2023). Deep NMF topic modeling. *Neurocomputing*, 515. <https://doi.org/10.1016/j.neucom.2022.10.002>

УДК 81'322.4

**ПРАКТИКА МАШИННОГО ПЕРЕВОДА  
В СОВРЕМЕННОМ МИРЕ: ОБЗОР*****М. М. Кадирова****Ташкентский государственный университет узбекского языка  
и литературы им. Алишера Навои  
Ташкент, Узбекистан  
begonammm@gmail.com*

Машинный перевод сегодня является активно развивающейся сферой. В связи с этим в мировой науке каждый день создаются новые исследования и новые идеи. В этой статье сравниваются и глубоко анализируются работы, проделанные международными учеными в области машинного перевода. Следует отметить, что в последние годы большие исследования проводились в области статистического перевода и перевода на основе нейронных систем. Также в статье приведены сведения о системах перевода, которые разрабатываются по результатам этих исследований, в них изучаются и анализируются результаты узбекско-английского межъязыкового перевода.

**Ключевые слова:** лог-линейное моделирование, n-грамма, апостериорная вероятность, энтропийный классификатор, иерархические лексические модели, схема модальности/отрицания (MN), метод «дерево-дерево».

**THE PRACTICE OF MACHINE TRANSLATION  
IN THE MODERN WORLD: SURVEY*****Madinabonu Kadirova****Tashkent State University of Uzbek Language and Literature name  
after Aisher Navai.  
begonammm@gmail.com*

**Abstract:** Machine translation is an actively developing field today. In this regard, new researches and new ideas are being created every day in world science. In this article, the work done by international scientists in the field of machine translation is compared and deeply analyzed. It should be noted that in recent years, great researches have been carried out in the areas of statistical translation and translation based on neural systems. The article also provides information about the translation systems that are being developed based on the results of these studies, and the results of Uzbek-English interlanguage translation are studied and analyzed in them.

**Keywords:** log-linear modeling, n-gram, posterior probability, entropy classifier, hierarchical lexical models, modality/denial (MN) scheme, tree-to-tree method.

Today, when globalization and active integration are being observed, it is natural that relations between countries and nations will increase. These processes are expected to become more intense in the following years. In such conditions, removing inter-linguistic barriers is considered one of the urgent tasks of modern science. In this regard, scientists from all over the world, especially from Europe, Asia, and America, are conducting important research. It should be mentioned here that the international journal “Computer Linguistics”, which has been published since 1988 and has been open access since 2009, has been widely promoted and introduced to the production of these studies. This journal is the primary archival forum for research in computational linguistics and natural language processing.

Also, large corporate organizations such as the University of Northern California in the USA, IBM Research Center in Germany, Aachen University in Germany, Google, Unbabel, and Alibaba are making a great contribution to the development of machine translation.

It has been confirmed that the rule-based machine translation (RBMT) approach, which was born almost simultaneously with the ideas of machine translation, is not able to provide perfect translation. Currently, the field of modern machine translation places high hopes on statistical machine translation (SMT), and especially on the approach of machine translation based on neural systems, which has been introduced to the translation system since 2016. After all, we can see that researches in the direction of automatic translation in the following years are mainly focused on these areas.

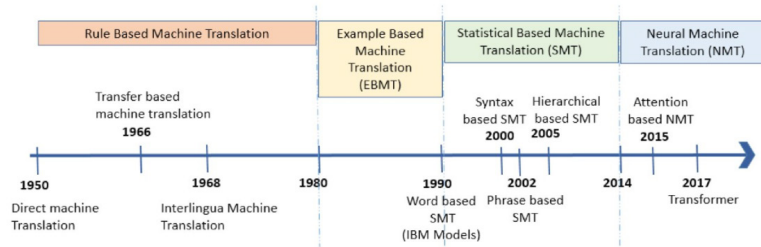


Figure 1. Development stages of machine translation

Most of the research from 2000-2016 is based on the SMT approach. In particular, Franz Josef Och and Hermann Ney, scientists of the Aachen University of Technology in Germany, “A systematic template approach to statistical machine translation” [Och, Ney, 2004, p.

418] published their articles. According to him, the context of words is taken into account in the translation model, and local changes in word order can be transferred from the source language to the target language with high accuracy. The model is described using the log-linear modeling method. According to scientists, it is expected to be easier to expand the model using this method than classical statistical machine translation systems.

Another group of scientists from the Aachen University of Technology proposed the use of dynamic programming technologies in statistical translation. K. Tillman and H. Neylar in his "Word Reordering and Dynamic Programming Ray Search Algorithm for Statistical Machine Translation" [Tillmann, Ney, 2003, p. 97] expressed their ideas in the article. This paper describes an efficient beam search algorithm for statistical machine translation based on dynamic programming. They present a new method of reordering applicable words between the source and target languages to achieve an efficient search algorithm. The direction of translation from German to English was chosen as the object of research. Accordingly, four sets of parameters are introduced to control the reordering of words, which can then be easily transferred to new translation directions. Hermann Ney's research continues even later. Together with Sonya Niessen, he published "Statistical Machine Translation with Few Resources Using Morpho-Syntactic Information" [Nießen, Ney, 2004, p. 182] announces his research. In this research work, he tries to completely separate translation based on traditional rules from statistical translation. According to him, in statistical machine translation, correspondence between source and target language words is checked on the basis of parallel corpora, and often no linguistic knowledge is used to build the underlying models. In particular, existing statistical systems for machine translation often treat different inflectional forms of the same lemma as independent words. They also propose to build hierarchical lexical models based on equivalence classes of words. In addition, it is proposed to create a software system trained to learn word order in connected sentences.

Hermann Ney later elaborated on these points in "Word-level confidence estimation for machine translation" [Ney, Weffing, 2007, p. 10] continues in his research. It provides several methods for estimating confidence between word levels in machine translation. They allow you to mark each word in the automatic translation as correct or incorrect. According to the article, confidence estimation approaches are based on posterior probabilities of the word. Different concepts of

word posterior probability and different methods of their calculation are compared. They can be divided into two categories: system-based methods that study the translation made by the program, and direct methods that are independent of the translation system. Posterior probability is determined by summing the probabilities of sentences in the translation hypothesis space containing words in the target language.

Aachen University researchers have also taken a special approach to evaluating the results of machine translation and analyzing errors. In particular, in 2014, M. Popovic and H. Ney published an article entitled "Evaluation of machine translation results and error analysis". [Popovic, Ney, 2014, p.657] It deals with the extraction of true false words using error-in-word (WER) and error-in-position (PER) detection algorithms. This study was a first step towards the development of automatic evaluation measures that provide more accurate information about specific translation problems. The proposed approach makes it possible to use different types of linguistic knowledge to classify translation errors in different ways. This work focuses on five categories of errors: inflectional errors, word order violations, word omissions, overuse, and incorrect word choice. They also compared the results of automatic error analysis with the results of human error analysis and found favorable results.

American scientists have created a unique school of statistical machine translation. The cities of Northern California and Cambridge served as research centers. In particular, scientists from the University of Northern California, Alexander Fraser and Daniel Marcu, dedicated to determining the quality level of the process of matching words in two languages in statistical machine translation "Measuring the quality of word MATCHING for statistical machine translation" [Fraser, Marcu, 2007, p. 293] published the article. In this research institute, Daniel Marcu is conducting major research on statistical translation. Together with the scientist D.S. Munteanu in 2006, "Improving the efficiency of machine translation by using non-parallel corpora" [Munteanu, Marcu, 2006, p. 477] published his research. The paper presents a new method for discovering parallel sentences in comparable, non-parallel corpora. They use a maximum entropy classifier. According to it, given a pair of sentences, the program can reliably determine whether they are translations of each other. They also show that it is possible to build a good quality MT system from scratch, starting with a very small parallel corpus (100,000 words) and using a large, non-parallel corpus. Thus, this method can be used among language pairs with very few sources.

Researchers from the world-famous company Google conducted research on query rewriting using monolingual statistical machine translation. [ Riezler, Liu, 2010, p.569] According to them, in the information technologies that work on the basis of the conjunctive method, queries entered some time ago and rarely used may not be stored in the system memory. Therefore, they propose to solve the problem by rewriting other terms with similar syntactic properties. They argue that the best results can be achieved by adopting the perspective of bridging the “lexical gap” between queries and documents by translating user queries from the source language into the target language of web documents.

Libin Shen and Jinxi Hu, researchers at Rayton BBN University of Technology, located in Massachusetts, USA, put forward the principle of text-to-dependency in statistical machine translation. [Shen, Hu, Weischedel, 2010] They propose a sentence dependency algorithm. This algorithm uses the target association language Munteanu model for decoding to use word relationships that cannot be modeled by the traditional n-gram model.

Also in 2010, researchers from the Kevin Knight Institute for Information Sciences and the Daniel Marcu Institute for Information Sciences published the paper “Restructuring, Redefinition, and Readaptation for Syntax-Based Machine Translation.” [Wang, May, Knight, Marcu, 2010]

This paper demonstrates that standard parsing and matching tools are not optimal for syntax-based statistical machine translation systems. And instead of them, three modification programs are presented to increase the accuracy of the MT system based on modern syntax: restructuring – changes the syntactic structure of the text to ensure the reuse of substructures; recoding and re-matching words from two different language materials – removing mistranslated words and combining words across sentences for correct structures. The most important thing is that all this work is done by EHM.

In 2009, Johns Hopkins University’s Center for Human Language Technologies established a summer camp for the study of applied languages. Within the SCALE-2009 program, which lasted for 8 weeks, scientists engaged in the creation of resources and systems for semantically informed machine translation (SIMT). They publicly demonstrated a new modality/negation (MN) annotation scheme and two automated MN taggers using a lexicon for using modality and negation in semantically informed syntactic MT. The annotation scheme



distinguishes three components of modality and negation: the trigger (the word denoting the modality or negation), the target (the action associated with the modality or negation), and the base (the experience of the modality). They state that the structure-based MN tagger leads to an accuracy of about 86% (depending on the genre) for tagging the standard LDC data set [Baker, Bloodgood, Dorr, Burch, Filardo, Piatko, Levin, Miller, 2010, p. 411].

And Israeli scientists proposed to adapt translation models to the translation language in order to improve the quality of machine translation [Lembersky, Ordan, Wintner, 2013, p. 999]. Translation models used for statistical machine translation are constructed from parallel corpora that are translated manually. Many studies in the field of translation studies show that the direction of translation is important, but it is natural that the language of translation (the target language) has unique characteristics. By adapting the translation model to the special features of the translation, we use the information about the direction of the translation to construct the phrase tables. And they also propose to create a mixed model by interpolating the tables of phrases in the translated texts in the “correct” and “incorrect” directions. They predict that this will lead to a consistent, statistically significant improvement in translation quality.

Scientists from America, Europe, and Eastern countries have proposed many models for statistical translation. After all, Kevin Gimpel, a scientist at the Toyota Institute of Technology located in Chicago, and A. Smith, a researcher at Carnegie Mellon University, put forward the idea of machine translation based on phrases with “tree-to-tree” quasichronic features [Gimpel, Smith, 2014, p. 349]. In 2016, Daniel Ortiz-Martinez, a teacher at the Polytechnic University of Valencia, Spain, presented an online course on statistical machine translation [Martinez, 2016]. In 2016, Dutch scientists Arianna Bisazza and Marcello Federico published a study titled “A Study of Word Reordering in Statistical Machine Translation: Computational Models and Linguistic Phenomena.” [Bisazza, Federico, 2016]

Considerable work has been done in this regard in the Asian region as well. For example, Japanese scientists such as G. Neubing and Taro Watanabe presented several methods of optimizing statistical machine translation [Neubig, Watanabe, 2016].

Source language adaptation approaches for machine translation in a resource-poor environment were studied by Pidong Wang, a scientist at the Qatar Computing Research Center, and Hwe Toular, a researcher at

the Singapore National Research Center [Wang, Nakov, TouNg, 2016, p. 111]. Edinburgh researchers have also worked on optimizing low-resource machine translation. After all, in 2016, B. Haddov, R. Bavden and A. Michel published a major research work entitled “Learning of low-resource machine translation”. [Haddow, Bawden., Miceli, Birch, 2022]

Chinese scholars Dun Deng and Nianwen Xue conducted empirical research on the topic of Chinese-English translation differences in machine translation. [Deng, Xue, 2017] A number of Qatari scientists have created speech structure evaluation methods in machine translation. created a model of the sequence of operations by combining the statistical translation based [Durrani, Schmid, Fraser, Koehn, Schutze, 2015]. Indian scientist Payal Khullar Do ellipses matter for Machine Translation in 2021? He published an article entitled [Khullar, 2021] Ellipsis is a linguistic phenomenon in which parts of a sentence are omitted and must be taken from a discourse or real-world context. In his research work, he proved the importance of ellipses and in some cases their impossibility. Canadian National Research Council scientist Saif M. Muhammad presented the automatic emotion detection technology. [Muhammad, 2022].

The years 2000–2016 can literally be called the heyday of statistical machine translation. Many works, monographs, research works, scientific articles were published in this field. “Learning Machine Translation” by K. Gutte, N. Canseda, M. Dimetman and George Foster in 2009 “Statistical Machine Translation” by Philip Koch in 2010, P. Williams, R. Senrich, F. Koch in 2016 Dozens of papers have been published, including Syntax-Based Statistical Machine Translation, 2017 by L.Spesy, K.Skarton, and G.Henriks, “Assessing the Quality of Machine Translation.”

In 2016, the emergence of ideas about the introduction of neural networks to machine translation led to radical changes in this field. In this regard, Google took the initiative to translate 20 languages based on neural networks. Almost all research in the following years focused on neural systems. Japanese scientists are showing great interest in this. For example, Microsoft researcher Akiko Eriguchi of Japanese origin, Yoshimasa Tsuruoka and Kazuma Hashimoto, professor of Tokyo University of Information and Communications Economics, worked on introducing source-side phrase structures into neural machine translation [Eriguchi, Tsuruoka, Hashimoto, 2019, p. 267]. Neural Machine Translation (NMT) has gained great success as a new alternative to the

traditional statistical machine translation model in several languages. Early NMT models were based on sequence learning, which encodes a sequence of source words into a vector space and generates another sequence of target words from the vector. In these NMT models, sentences are treated as sequences of words with no internal structure. These scholars have focused on the syntactic structure of sentences and propose a new syntactic NMT model called the NMT tree sequence model. Their proposed model has a mechanism that allows the decoder to generate a translated word and easily match it with phrases. Experimental results show that using a syntactic structure can be useful when the data set is small, but is not as efficient as using a bidirectional encoder. As the size of the data set increases, the advantages of using a syntax tree diminish.

Testing of sequence level for non-autoregressive neural machine translation is being conducted by the Intelligent Information Processing Laboratory of the Institute of Computing Technology, Chinese Academy of Sciences [Shao, Feng, Zhang, Meng, 2021, p. 861]. Scientists propose a new model for sequence level optimization based on several new amplification algorithms adapted for NMT. This model is superior to the traditional method in the estimation of gradients by the property of reducing their difference.

Research conducted by the Chinese corporation Alibaba Group focuses on the problems of working with short texts [Wan, Yang, Wong, Chao, Zhang, 2021, p. 321].

Short texts are available in a variety of formats, including query, dialog, and message formats. Much of the research in Neural Machine Translation (NMT) has focused on solving the open problems of long sentences rather than short sentences. This is because in humans, relatively short sequences of learning and processing are generally regarded as easy examples. Researchers prove that this is not the case with the help of experiments. They argue that the longer a sentence is, the more its content determines and reduces inter-sentence options, the lack of contextual information causes NMT to be biased against information on short sentences, and therefore the NMT model is flawed. leads to translation.

Despite the recent success of deep neural networks in natural language processing and other areas of artificial intelligence, they remain difficult to interpret. Scientists from different countries of the world have analyzed the images studied by neural machine translation (NMT) models at different levels and are trying to evaluate their quali-

ty through the relevant external features. In particular, they are looking for answers to the following questions:

(I) How well is word structure preserved within expressions, which is an important aspect in translating morphologically rich languages?

(II) Do translational reflection mechanisms capture long-range dependencies and efficiently handle syntactically divergent languages?

(III) Do reflective mechanisms capture lexical semantics? In particular, Yonatan Belinkov and James Glans, researchers of the Massachusetts Institute of Computer Technologies and Artificial Intelligence Laboratory, Nadir Durrani and Fahim Dalvi, scientists of the Qatar Computer Research Institute. Professor John F. Paulson of Harvard University conducted research on the linguistic representation power of neural machine translation models and conducted a detailed investigation in the following areas [Belinkov, Paulson, Durrani, Dalvi, Sajjad, Glans, 2020] :

(I) which layers of translation architecture cover each of these linguistic phenomena;

(II) How does the choice of translation unit (word, symbol or sub-word unit) affect the linguistic properties captured by the underlying images?

(III) Do the encoder and decoder learn the translation object independently?

(IV) Do representations learned by multilingual NMT models capture the same amount of linguistic information as their bilingual counterparts?

As a result of these studies, hundreds of translation programs have been created today, which are based on some method of translation. Below, we will focus on the features of online translator programs that have the highest ratings in the world and the analysis of Uzbek-English translations. That is, with the help of the translator program, we translate the sentence in Table 1 into English, and then translate the translation given in the same translator program back into Uzbek:

| № | Machine Translator | Tebranib yonayotgan sham zo'r mo'jizaday hammaning diqqatini jalb qilgan edi. |  |
|---|--------------------|---|--|
|   |                    | In English  | In Uzbek   |
| 1 | Google             | <i>The flickering candle attracted everyone's attention like a miracle</i>    | <i>Miltillovchi sham hammaning e'tiborini mo'jizadek o'ziga tortdi</i> |

| № | Machine Translator               | Tebranib yonayotgan sham zo'r mo'jizaday hammaning diqqatini jalb qilgan edi.                     |  |
|---|----------------------------------|---|--|
|   |                                  | In English  | In Uzbek   |
| 2 | <b>Yandex translator</b>         | <i>The vibrating burning candle was the perfect miracle that attracted everyone's attention</i>   | <i>Vibratsiyali yonayotgan sham barchaning e'tiborini tortgan mukammal mo'jiza edi.</i>        |
| 3 | <b>Prompt. One</b>               | <i>Lightning candle attracted the attention of everyone</i>                                       | <i>Cho'qqon shami barchaning e'tiborini o'g'irladi</i>   |
| 4 | <b>Bing Microsoft Translator</b> | <i>The vibrating burning candle had attracted the attention of everyone like a great miracle.</i> | <i>Tebranayotgan yonayotgan sham barchaning e'tiborini buyuk mo'jizadek o'ziga tortgan edi</i> |

**Table 1. Analysis of translation of modern machine translations between Uzbek and English languages**

**Google** is an excellent online translator based on neural systems. The interface is simple and supports translation into more than a hundred languages. By the number of dictionaries and available functions, it is considered the most functional and versatile service to date. It can be used to translate articles, documents, web pages, as well as PPT, PDF and other file formats.

It is possible to save translated sentences and articles in Google memory. If you start translating individual words, the service will automatically switch to online dictionary mode. Alternative options are suggested for each word with a brief description. Transcription and transliteration are supported. It also has the ability to work with voice: voice-to-text, text-to-voice technology, and the ability to return the translation result in voice format. However, the amount of text input for translation is limited to 5000 characters, and there are many meaningful misunderstandings in the direction of English-Uzbek translation.

The translation of Google translator between Uzbek and English languages is considered in Table 1. According to him, the word “multilamoq” cannot translate the word “tebranib yonmoq” but the program did not stray too far from the meaning of the sentence.

**Yandex translator** is a high-quality online translation system based on a hybrid translation model created by mixing statistical and neural translation systems. Translates between more than 90 language pairs and automatically detects the input language. The translation base

develops itself based on the statistical approach. Based on the online translator system, the user can create his own dictionary database. The program performs a hold translation based on his personal terminology.

We reviewed the Uzbek-English translation status of the translator program above. As we can see in Table 1, changes were also observed in the structure and content of sentences.

**Prompt.One** is a new hybrid technology that combines a neural network approach and a rule-based translation system, that is, Rule-Based Translation (RBMT), and provides automatic translation services mainly for European and Asian languages. This machine translation service is one of the most popular online translators created in Russian computer linguistics. The service is based on its own linguistic technologies that guarantee high-quality translation. PROMT's neural algorithms pre-analyze the text and decide which technology is best suited to translate a particular piece of text. The system supports 22 languages. We can see that it has the following advantages:

- checks the spelling of words;
- there is an opportunity to evaluate the proposed translation;
- has the function of automatic detection of the input language;
- Developers have the option to send their own version of the translation.
- has the ability to determine the topic of the text;
- a perfect system with a dictionary system that includes word usage options.

However, we can see in Table 1 that abbreviations are observed in the sentence structure when translated using the translator program.

Bing Microsoft translator. Since 2000, Microsoft has been providing translation services based on statistical machine translation. SMT uses advanced statistical analysis to estimate the best translations of a word given the context of multiple words. Since the beginning of 2010, the introduction of new artificial intelligence technology, deep neural networks, has made it possible to improve the quality of translation. The steps of Microsoft's neural network translation algorithm are as follows:

- Each word, or rather the symbols that represent it, pass through the first layer of 500-dimensional vector "neurons" that encode it into a 1000-dimensional vector (b) that represents the word in the context of other words in the sentence.

- after all the words have been encoded once into these 1000-dimensional vectors, the process is repeated several times, allowing each layer to fine-tune the 1000-dimensional representation of the word.

- The final output matrix is used by a software algorithm to determine which word should be translated next from the original sentence of previously translated words. Also, in this process, unnecessary words in the target language are removed.

- The decoder layer converts the selected word (or, more precisely, the 1000-dimensional vector representing that word in the context of a full sentence) into the most appropriate target language equivalent. The output of this last layer (c) is then fed back to the verification layer to calculate which word should be translated next from the original sentence.

- When this decoder reaches the translation level, the appropriate translation is transmitted.

Thanks to this approach, the end result is in most cases smoother and closer to human translation than SMT-based translation.

We can see the translation quality from Table 1. According to that, there was no significant change in the content or structure of the sentence. Therefore, the most perfect system for the Uzbek-English translation direction in the translation of artistic texts is the Microsoft Bing translator program.

Today, there are hundreds of other world-famous translation programs. For example, SYSTRAN, Babylon, DeepL, Translatedict, Translate.net, and Reverso are known for their excellent translation services and the wide range of options they provide to users. However, it is impossible to use these systems as they do not include the Uzbek language. In general, machine translation systems are improving day by day. This is determined by the increasing demand for them day by day. Facilitating the work of the translator in the translation process is the main task of today's machine translation industry.

Machine translation is becoming a necessity in many aspects of life in today's globalization era. First of all, in this era of increasing volume of business and scientific texts in the world, publishing books and articles in the scientific and business domains in foreign languages is becoming a necessity. It is natural that there is a need to translate international relations and cooperation, legal agreements, technical documents, instructions, announcements, advertisements and other texts.

Second, every word in a foreign language is a crossroad of cultures, a practice of intercultural communication, because they reflect a for-

eign world and a foreign culture: behind every word there is a world idea conditioned by national consciousness. Translation, with its versatility, is important for the correct perception of the world and intercultural communication. But there are still language barriers or translation difficulties in communication between experts from different countries.

There are several advantages to applying machine translation using computer science and artificial intelligence. One of them is the **speed** of the machine. In particular, a sentence can be translated immediately by copying the text on a computer or smartphone. Of course, the result may not be 100% correct, but if used in standard sentences based on specific facts, it can fully satisfy the relevance. For example, shopping sites, weather information. Also, texts with a lot of strict sentences, such as travel, administrative documents and government documents, can be quickly translated using machine translation.

The second is multilingual **compatibility**. Unlike electronic dictionaries with a limited number of languages, most modern automatic translation applications can work with multiple languages. In particular, Google Translate has the ability to translate from one language to many languages and quickly change the output language. Electronic dictionaries are limited to certain languages.

The third is **comfort in space and time**. In the age of technology, the most important thing for users is convenience. Machine translation can be easily used in any situation, even without a sophisticated machine, using a smartphone. Its speed and multilingual interface make it more efficient than pocket dictionaries. Also, most of the modern machine translation programs have a speech recognition and text conversion system, which allows them to be used in any place and time.

Machine translation is one of the most important areas of computational linguistics and covers all the problems of speech processing at all language levels. Reflecting on the advantages of machine translation, scientists note that they provide a high translation speed with the possibility of processing a large amount of data and the general “**neutrality**” of the output texts.

#### REFERENCES

1. Baker K., Bloodgood M., Dorr B. J., Burch Ch., Filardo N. W., Piatko Ch., Levin L., Miller S. Use of Modality and Negation in Semantically-Informed Syntactic MT // Johns Hopkins University, Moulton Street, Cambridge. 2012.
2. Belinkov Y., Paulson J., Durrani N., Dalvi F., Sajjad H. Glans J. On the Linguistic Representational Power of Neural Machine Translation



Models // Qatar Computing Research Institute HBKU Research Complex. – Doha, 2020.

3. Bisazza A., Federico M. A Survey of Word Reordering in Statistical Machine Translation: Computational Models and Language Phenomena, Amsterdam, 2016.

4. Deng D., Xue N. Translation Divergences in Chinese–English Machine Translation: An Empirical Investigation // Department of Chinese Languages and Literature, Tsinghua University, Beijing, China, 2017.

5. Durrani N., Schmid H., Fraser A., Koehn P., Schutze H. The Operation Sequence Model – Combining N-Gram-Based and Phrase-Based Statistical Machine Translation // Ludwig Maximilian University – Munich, 2015.

6. Khullar P. Are Ellipses Important for Machine Translation? IIIT Hyderabad, 2021.

7. Eriguchi A., Hashimoto K., Tsuruoka Y. Incorporating Source-Side Phrase Structures into Neural Machine Translation // The University of Tokyo Department of Information and Communication Engineering – Tokyo, 2019.

8. Fraser A., Marcu D. Measuring Word Alignment Quality for Statistical Machine Translation. University of Southern California. USA, 2007.

9. Gimpel K., Smith N. A. Phrase Dependency Machine Translation with Quasi-Synchronous Tree-to-Tree Features // Carnegie Mellon University. - Germany, 2014.

10. Goutte C., Cancedda N., Dymetman M., Foster G. Learning Machine Translation Institute for Information Technology, National Research Council// MA: The MIT Press, Cambridge, 2009

11. Haddow B., Bawden R., Miceli A. V., Birch A. Survey of Low-Resource Machine Translation // University of Edinburgh School of Informatics - Edinburgh, 2022.

12. Joty Sh., Guzman F. Marquez L., Nakov P. Discourse Structure in Machine Translation Evaluation // School of Computer Science and Engineering Nanyang Technological University – Qatar Foundation, 2017.

13. Koehn P. Statistical Machine Translation (book). Cambridge University Press// Edinburg, 2010.

14. Lembersky G., Ordan N. Wintner S. Improving Statistical Machine Translation by Adapting Translation Models to Translationese //\*\*University of Haifa, Israel, 2013.

15. Martinez D.O. Online Learning for Statistical Machine Translation// Universitat Politècnica de Valencia, 2016.

16. Muhammad S. M. Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis. National Research Council Canada, 2022.

17. Munteanu D.S., Marcu D. Improving Machine Translation Performance by Exploiting Non-Parallel Corpora. Information Sciences Institute University of Southern California. – USA, 2006.
18. Neubig G., Watanabe T. Optimization for Statistical Machine Translation: A Survey // Tokyo, Japan, 2016.
19. Nießen S., Ney H. Statistical Machine Translation with Scarce Resources Using Morpho-syntactic Information // Computer Science Department, RWTH Aachen, Germany, 2004.
20. Och F. J. Ney H. The Alignment Template Approach to Statistical Machine Translation. Computer Science Department, RWTH Aachen–University of Technology, Ahornstr. 55, 52056 Aachen, Germany. – 2004.
21. Popovic M., Ney H. Towards Automatic Error Analysis of Machine Translation Output// RWTH Aachen University – German Research Centre for Artificial Intelligence, Alt-Moabit Berlin, Germany. 2011.
22. Riezler S., Yi Liu Query Rewriting Using Monolingual Statistical Machine Translation. Zurich, Switzerland, 2010.
23. Shao Ch. Feng Y., Zhang J., Meng F. Incorporating Source-Side Phrase Structures into Neural Machine Translation // Laboratory of Intelligent Information Processing Institute of Computing Technology Chinese Academy of Sciences – China, 2021.
24. Shen L., Xu J. Weischedel R. String-to-Dependency Statistical Machine Translation Raytheon BBN Technologies. Moulton Street, Cambridge, 2010.
25. Specia L., Scarton K., Paetzold G.H. Quality Estimation for Machine Translation University of Sheffield and Federal University of Technology, Parana – 2017.
26. Tillmann Ch., Ney H. Word Reordering and a Dynamic Programming Beam Search Algorithm for Statistical Machine Translation. Computer Science Department, RWTH Aachen, Germany. – 2003.
27. Tillmann Ch., Ney H. Word Reordering and a Dynamic Programming Beam Search Algorithm for Statistical Machine Translation. Computer Science Department, RWTH Aachen, Germany. - 2003.
28. Ueffing N., Ney H. Word-Level Confidence Estimation for Machine Translation RWTH Aachen University, Germany, 2007.
29. Wang P., Nakov P., TouNg H. Source Language Adaptation Approaches for Resource-Poor Machine Translation // National University of Singapore, 2016.
30. Wang W., May J., Knight K., Marcu D. Re-structuring, Re-labeling, and Re-aligning for Syntax-Based Machine Translation Los Angeles, USA. – 2010.
31. Wan Y., Yang B., Wong D. F., Chao L. S., Zhang H. On the Linguistic Representational Power of Neural Machine Translation Models // Alibaba DAMO Academy, 2021.

УДК 81'322.2

**ВАЖНОСТЬ ПРОГРАММЫ ALIGNER  
ДЛЯ ПАРАЛЛЕЛЬНОГО КОРПУСА*****И. А. Холмонова****Ташкентский государственный университет узбекского языка  
и литературы им. Алишера Навои**Ташкент, Узбекистан**iqbolabintualisher@gmail.com*

В этой статье обсуждается важность программного обеспечения для выравнивания в параллельный корпус. Проанализированы исследования, проведенные по программе Aligner и разработанным алгоритмам, а также предложенные модели создания программы. При этом можно изучить взаимосвязь параллельного корпуса и программы выравнивания.

**Ключевые слова:** параллельный корпус, программное обеспечение для выравнивания, модели, выравнивание текста, модели IBM, фильтрация корпуса.

**IMPORTANCE OF ALIGNER PROGRAM  
FOR PARALLEL CORP*****Iqbola Xolmonova****Tashkent State University of the Uzbek language and literature  
named after Alisher Navoi**Tashkent, Uzbekistan**iqbolabintualisher@gmail.com*

This article discusses the importance of an aligner program for a parallel case. You can get detailed information about why the Aligner program is needed and about the research conducted on this program and the developed algorithms and the proposed models for creating the program. At the same time, it is possible to study the connection between the parallel case and the aligner program.

**Keywords:** parallel corpus, aligner software, models, text alignment, IBM models, corpus filtering.

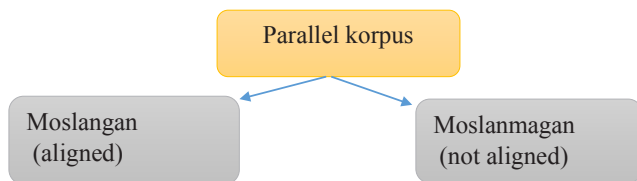
**PARALLEL KORPUS UCHUN ALIGNER DASTURINING  
AHAMIYATI ORA*****Xolmonova Iqbola Alisher qizi****Alisher Navoiy nomidagi Toshkent davlat**o'zbek tili va adabiyoti universiteti**Toshkent, O'zbekiston**iqbolabintualisher@gmail.com*

**Annotatsiya.** Ushbu maqolada parallel korpus uchun aligner dasturining ahamiyati haqida soʻz boradi. Aligner dasturi ustida olib borilgan tadqiqotlar va ishlab chiqilgan algoritmlar hamda dastur tuzish uchun taklif qilingan modellar tahlil qilinadi. Shu bilan birga parallel korpus va aligner dasturi bogʻliqligini oʻrganish mumkin.

**Kalit soʻzlar:** *parallel korpus, aligner dasturi, modellar, matnni moslashtirish, IBM modellari, korpusni filtrlash.*

**Kirish.** Parallel korpus deganda ikki yoki undan ortiq tildagi matnlar toʻplami tushuniladi. Ular jumlar yoki iboralar darajasida moslashtiriladi. Ushbu korpuslar turli xil tabiiy tillarni qayta ishlash vazifalarida, masalan, mashina tarjimasida, tillararo maʼlumot qidirish va ikki tilli lugʻat yaratishda qoʻllaniladi.

Parallel korpus – bir mavzuga oid ikki tildagi matnli korpus. Birinchi tipdagi korpus “parallel korpus” (parallel corpora) deb nomlanib, maʼlum bir tarjimaning turli aspektini oʻrganish uchun qoʻllaniladi. Parallel korpus, oʻz navbatida, yana 2 turga boʻlinadi. Bu moslangan (aligned) va moslanmagan (not aligned) korpusdir.



“Moslangan” atamasi korpusda tarjima birliklari orasida bir-birini taqozo etuvchi aniq aloqa mavjudligini bildiradi. Bunday korpusdan u yoki bu gapni qanday tarjima qilinganini oʻrganish mumkin. Bu turdagi korpus tarjimon uchun ahamiyatli, chunki unda noyob resurs – “tarjima xotira” (translation memory) mavjud [Xamroeva, 2019: 48].

Demak, bundan kelib chiqadiki, parallel korpus yaratish qiyin, chunki u jummalarni yoki iboralarni toʻgʻri moslashni talab qiladi. Bu jarayon, odatda, qoʻlda yoki avtomatik moslashtirish algoritmlari orqali amalga oshiriladi. Moslashtirgandan soʻng parallel korpuslar statistik mashina tarjimasida modellarini oʻrgatish yoki ikki tilli leksikonlarni olish uchun ishlatilishi mumkin [Zong, 2019].

**Asosiy qism.** Parallel korpusga turli tillardagi jumlar yoki iboralarni toʻgʻri moslashtirish uchun “aligner” dasturi kerak. Ushbu moslashtirish turli maqsadlarda, jumladan, tarjima, tilni taqqoslash va mashina tarjimasida modellarini oʻrgatish uchun juda muhim.

Hozirgi kunda parallel korpuslar uchun aligner dasturiga qaratilgan tadqiqotlar ko'p emas. Masalan, mashina tarjimasini va moslashtirish bo'yicha taniqli tadqiqotchi Filipp Koen GIZA++ va fast\_align kabi moslashtirish vositalarini ishlab chiqishga hissa qo'shgan. Ushbu vositalar ingliz, fransuz, nemis, ispan, italyan, golland, rus, xitoy, yapon, arab tillaridagi parallel korpuslarni moslashtirish uchun keng qo'llaniladi [Xie, 2023].

Mashina tarjimasini tadqiqotchisi Kris Kallison-Burch GIZA++ va uning variantlari kabi moslashtirish vositalarini ishlab chiqishga ham hissa qo'shgan. U parallel korpuslar uchun moslashtirish texnikasini takomillashtirish ustida ishlagan [Callison-Burch, 2004].

Kanada Milliy Tadqiqot Kengashi (NRC) Bitextor deb nomlangan ochiq manbali moslashtirish vositasini ishlab chiqdi, u parallel korpuslarni moslashtirish uchun maxsus mo'ljallangan bo'lib har xil turdagi parallel ma'lumotlar bilan ishlashga qaratilgan [Bañón, 2020].

Moses SMT Toolkit (Muson) keng qo'llaniladigan statistik mashina tarjimasini manbalar to'plami GIZA++ va fast\_align kabi moslashtirish vositalarini o'z ichiga oladi. Musoning ishlab chiquvchilari va mualliflari parallel korpuslar uchun aligner dasturiga harakat qilishdi [Koehn, 2013: 121-191].

Avtomatik jumalarni moslashtirishning birinchi yondashuvlari uzunlikka asoslangan edi. Gale va Church (1991) "belgilardagi paragraf uzunligi va uning tarjimasini uzunligi o'rtasidagi bog'liqlik juda yuqori ekanligini" aniqladilar. Bunga asoslanib, ular belgilar uzunligining oddiy statistik modeliga asoslangan jumalarni moslash usulini tasvirladilar. Brovn va boshqalar (1991) ham uzunlikka asoslangan usulni tavsiflaydi, lekin belgilar o'rniga tokenlardan foydalanganlar. Bunga qo'shimcha ravishda, ular korpusni kichikroq bo'laklarga bo'lish uchun asosiy nuqtalari sifatida belgilashdagi signallardan foydalanadilar [8]. Papageorgi va boshqalar (1994) PoS-teglar asosida optimal moslashni hisoblash orqali, odatda, tarjimada saqlanadigan nutq qismidan foydalanganlar. Tschorn va Ludeling (2003) lug'atga asoslangan masofa o'lchovini yaxshilash uchun morfologik analizatoridan foydalanadi, Ma esa (2006) kamroq tarjima qilingan so'zlarga katta vaznlarni belgilash orqali leksikonga asoslangan moslashgichning mustahkamligini oshiradi. Tompson va Koen (2019) ikki tilli jumalarni joylashtirishga asoslangan usulni tasvirlab beradi, bunda joylashtirishlar orasidagi o'xshashlik moslashish uchun ball funksiyasi sifatida ishlatiladi.

So‘nggi paytlarda neyron tarmoqlar asosiy nuqtalarni topish va noto‘g‘ri moslanishlarni aniqlash uchun ishlatilgan. Ushbu usullarning ko‘pchiligi manba va maqsad jumlarlar parallel yoki yo‘qligini aniqlash uchun tasniflagichlarni o‘rgatish orqali taqqoslanadigan korpusdan parallel jumlalarni ajratib olish uchun ishlab chiqilgan. Avvalgi ishlar so‘zlarni moslash uchun IBM modellaridan foydalanishni o‘z ichiga oladi (Brovn va boshq., 1993). Khadivi va Ney (2005) korpusning ortiqcha belgilar qismini IBM 1 va 4 modellari va uzunlikka asoslangan modellari asosida filtrlaydi va ularning chiziqli kombinatsiyasi bo‘yicha moslanishlarni baholaydi. Sarikaya va boshqalar (2009) jumlarar juftligi qamrovini kengaytirish uchun kontekst ekstrapolyatsiyasidan foydalanadi, jumalarning bog‘lanish nuqtasidan masofasi bir xil yoki yo‘qligini tekshiradi va jumlarar belgilangan chegaradan past bo‘lishiga qaramay, oyna ichidagi boshqa juftliklar bilan solishtirganda eng yuqori o‘xshashlik balliga egami yoki yo‘qligini tekshiradi.

Zipporah (Xu va Koen, 2017) jumlarar juftlarini tasniflash uchun o‘qitilgan logistik regressiya modelidan foydalanadi. BiCleaner (Sanchez-Kartagena va boshq., 2018) nuqsonli jumlararni aniqlash uchun qo‘lda ishlangan qoidalar to‘plamidan foydalanadi, keyin leksik tarjimalar va tegishli uzunlik, mos keladigan raqamlar va tinish belgilari kabi bir nechta sayoz xususiyatlarga asoslangan tasodifiy moslashgichdan foydalanishga kirishadi [9; 182–190].

Aligner dasturini yaratish bir necha bosqichlarni o‘z ichiga oladi.

**1. Ma’lumotlarni tayyorlash.** Moslashtirilishi talab qilinuvchi parallel matnlardan iborat bo‘lgan korpus tayyorlanadi. Ushbu korpus asl manba va uning tarjimai mavjud bo‘lgan matnlardan iborat bo‘lishi talab etiladi.

**2. Oldindan ishlov berish.** Matnlarni moslashtirish jarayoniga xalaqit beradigan keraksiz belgilar va tinish belgilaridan tozalash lozim. Formatlash va taqdim etishda izchillikni ta’minlash uchun matnlarni normallashtirish lozim [10].

**3. Algoritm tanlash.** Tadqiqotchi o‘zining talablaridan hamda ma’lumotlarning tabiatidan kelib chiqib algoritm tanlashi mumkin. Smit-Waterman algoritmi, Needleman-Wunsch algoritmi va mashina tarjimai uchun IBM modellari kabi turli xil algoritmlar mavjud.

**4. Amalga oshirish.** Dasturchi istagan dasturlash tilida tanlangan moslashtirish algoritmi yordamida kod yozishi lozim. Bu o‘xshashlik ko‘rsatkichlari, masofalarni tahrirlash yoki ehtimollik modellari asosida jumalarni yoki so‘zlarni solishtirish va moslashtirish uchun kod yozishni o‘z ichiga oladi.

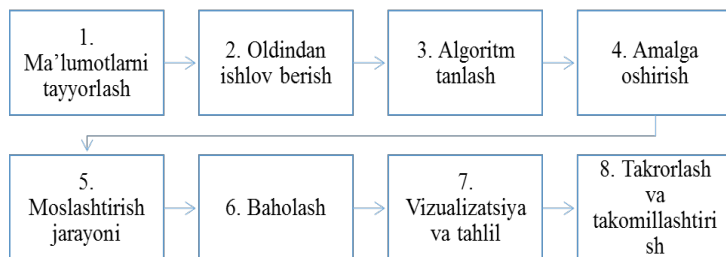
**5. Moslashtirish jarayoni.** Parallel korpusga moslashtirish algoritmini qo'llash kerak. Ikkala tildagi jumlarlar yoki so'zlarni takrorlab, ularni solishtirib va tanlangan algoritm asosida ularning mos kelishini aniqlash zarur. Moslashtirishlarni matritsa yoki lug'at kabi ma'lumotlar tuzilmasida saqlash talab etiladi [Zariņa. 2015].

**6. Baholash.** Tegishli ko'rsatkichlar, masalan, aniqlik, eslab qolish yoki F1 balli ko'rsatkich (F1 reytingi tasniflash modelining ish faoliyatini baholash uchun ishlatiladigan ko'rsatkichdir. Matnni tasniflash, his-tuyg'ularni tahlil qilish va spamni aniqlash kabi vazifalarda modellarning samaradorligini baholash uchun odatda ma'lumot olish, tabiiy tilni qayta ishlash va mashinali o'rganish kabi sohalarda qo'llaniladi. U aniqlik va eslab qolishni yagona ballga birlashtirib, model aniqligining muvozanatli o'lchovini ta'minlaydi) yordamida moslashtirish sifatini baholash mumkin [Maučec, 2019].

**7. Vizualizatsiya va tahlil.** Lisoniy struktura, tarjima aniqligi yoki lingvistik hodisalar haqida tushunchaga ega bo'lish uchun moslashtirilgan ma'lumotlarni vizualizatsiya qilish va tahlil qilish lozim. Ushbu bosqich moslashtirish vizualizatsiyasini yaratish, statistik tahlil yoki lingvistik tadqiqotlar uchun maxsus vositalardan foydalanishni o'z ichiga oladi [13].

**8. Takrorlash va takomillashtirish.** Fikr-mulohaza, baholash natijalari va domenga xos talablar asosida aligner dasturini doimiy ravishda takomillashtirib borish kerak bo'ladi. Buning uchun dasturga moslashtirish aniqligi va qamrovini oshirish maqsadida qo'shimcha funktsiyalar yoki usullarni qo'shib borish kerak.

Bu bosqichlarni quyidagicha umumlashtirish mumkin:



Aligner dasturini yaratish tabiiy tillarni qayta ishlash, algoritmlar va dasturlash ko'nikmalarini yaxshi tushunishni talab qiladi. Bundan tashqari, moslashtirish texnikasi va eng yaxshi amaliyotlar haqida tushunchaga ega bo'lish uchun mavjud aligner dasturiy ta'minoti va tadqiqot hujjatlarini o'rganish foydali.

Aligner dasturi turli amallarni bajarishda qo‘l keladi. Masalan, Aligner dasturi parallel korpusdagi jumlarlar yoki so‘zlarni moslashtirishga yordam beradi. Parallel matnlarni moslashtirish turli xil tabiiy tillarni qayta ishlash vazifalari, masalan, mashina tarjimasi, tillararo ma’lumot olish va ikki tilli lug‘at yaratish uchun zarur. Aligner dasturi ko‘pincha statistik mashina tarjimasi modellarini o‘rgatish uchun ham ishlatiladi. Ushbu modellar tillar orasidagi tarjima matnlarni moslashtirish asosida o‘rganadi va tarjimaning aniqligini oshiradi [Dragos, 2005]. Bu dasturni lingvistik tadqiqotlarda ham keng qo‘llash mumkin. Dastur tilshunoslarga parallel matnlarni moslashtirish orqali til tuzilmalari va hodisalarini o‘rganishda yordam beradi. Bu tadqiqotchilarga turli tillarning tegishli qismlarini solishtirish va tahlil qilish imkonini beradi, bu esa tillararo tadqiqotlarni osonlashtiradi [Adamou, 2019]. Aligner dasturi til o‘rganish va o‘qitish maqsadlarida ikki tilli jumlarlar juftlarini yaratish uchun ishlatilishi mumkin. Parallel korpusdagi jumlarlarni moslashtirish orqali o‘quvchi tarjimalarni osongina solishtiriradi va tillar orasidagi farqni tushunadi. Shu bilan birga Aligner dasturi parallel matnlardan atamalarni ajratib olish, nomli obyektlarni aniqlash yoki ikki tilli matnlardan ma’lumotlarni olish uchun ishlatilishi mumkin. Umuman olganda, aligner dasturiy ta’minoti parallel korpuslarni moslashtirishda hal qiluvchi rol o‘ynaydi, tabiiy tillarni qayta ishlash, mashina tarjimasi, lingvistik tadqiqotlar, til o‘rganish va ma’lumot olishda turli xil ilovalarga imkon beradi.



Parallel korpus va aligner dasturlari tabiiy tilni qayta ishlash va mashina tarjimasi kontekstida bog‘lanadi. Parallel korpuslar jumla yoki ibora darajasida birlashtirilgan ikki yoki undan ortiq tildagi matnlar to‘plamiga ishora qiladi. Parallel korpuslar turli tillardagi jumla/ iboralarni taqqoslash va moslashtirish orqali kerakli ma’lumotlarni taqdim etadi. Aligner dasturi esa, jumlarlar yoki iboralarni parallel korpusda avtomatik ravishda moslashtirish uchun ishlatiladigan vosita



yoki dasturdir. Bu turli tillardagi matnning mos segmentlarini aniqlashga yordam beradi hamda ikki yoki ko'p tilli moslashuvlarni yaratishga imkon beradi. Aligner dasturi jumlar yoki iboralar orasidagi eng yaxshi moslashuvni aniqlash uchun statistik modellar yoki lingvistik qoidalar kabi turli usullardan foydalanadi.

Parallel korpus va aligner dasturiy ta'minoti o'rtasidagi munosabat shundan iboratki, dasturiy ta'minot parallel korpusdagi matnlarni qayta ishlash va moslashtirish uchun ishlatiladi. Aligner dasturi jumla/iboralarni qo'lda moslashtirish bo'yicha ko'p vaqt talab qiladigan vazifani avtomatlashtiradi, bu esa yuqori sifatli parallel korpuslarni yaratishni oson va samaraliroq qiladi.

Aligner dasturi mos segmentlarni topish uchun turli tillardagi matnlarning lingvistik jihatlari va strukturasi tahlil qiladi. U so'zlarni, iboralarni va sintaktik strukturalarni solishtiradi va o'xshashlikni aniqlaydi va ularni moslashtiradi.

Xulosa. Aligner dasturi parallel korpusdagi matnlarni qayta ishlab, turli tillarda tegishli segmentlarni ko'rsatadigan moslashtirilgan ma'lumotlarni ishlab chiqaradi. Keyinchalik bu moslashtirilgan ma'lumotlar mashina tarjimasiz tizimlari uchun o'quv ma'lumotlari sifatida ishlatiladi [16]. Aligner dasturi korpus ichidagi matnlarni qayta ishlash va moslashtiruvchi sifatida parallel korpusning bir qismi hisoblanadi. Dastur ikki yoki undan ortiq tildagi matnlarni oladi va matnning tegishli segmentlarini jumla yoki ibora darajasida avtomatik ravishda aniqlaydi. U jumlar yoki iboralar orasidagi eng yaxshi moslikni aniqlash uchun algoritmlar, statistik modellar yoki lingvistik qoidalardan foydalanadi.

## ADABIYOT

1. Xamroeva Sh. Korpus lingvistikasi atamalarining qisqacha izohli lug'ati. Terminologik lug'at, 2020 – B. 48.
2. Zhaorong Zong, Changchun Hong. Research on Alignment in the Construction of Parallel Corpus. School of Foreign Languages, Huangshan University, 2019.
3. Ieva Zariņa, Pēteris Nīkiforovs, and Raivis Skadiņš. Word Alignment Based Parallel Corpora Evaluation and Cleaning Using Machine Learning Techniques. In Proceedings of the 18th Annual Conference of the European Association for Machine Translation, Antalya, Turkey. 2015 – P. 185–192.
4. Ruoyu Xie, Antonios Anastasopoulos. Noisy Parallel Data Alignment. Department of Computer Science, George Mason University. 2023.
5. Chris Callison-Burch, David Talbot, Miles Osborne. Statistical Machine Translation with Word- and Sentence-Aligned Parallel Corpora. Janu-

ary 2004. Conference: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, 21–26 July, 2004, Barcelona, Spain.

6. Marta Bañón, Pinzhen Chen, Barry Haddow, Kenneth Heafield. ParaCrawl: Web-Scale Acquisition of Parallel Corpora. January 2020. Conference: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.

7. Philipp Koehn. MOSES Statistical Machine Translation System User Manual and Code Guide. University of Edinburgh. October 19, 2013 – P. 121–191.

8. Steinþór Steingrímsson, Hrafn Loftsson, and Andy Way. 2020. Effectively Aligning and Filtering Parallel Corpora under Sparse Data Conditions. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 182–190, Online. Association for Computational Linguistics.

9. Corpora under Sparse Data Conditions. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 182–190, Online. Association for Computational Linguistics.

10. Effectively Aligning and Filtering Parallel Corpora under Sparse Data Conditions Steinþór Steingrímsson Department of Computer Science Reykjavik University Iceland steinhor18@ru.is Hrafn Loftsson Department of Computer Science Reykjavik University Iceland hrafn@ru.is Andy Way School of Computing ADAPT Centre Dublin City University Ireland andy.way@adaptcentre.ie

11. Ieva Zariņa, Pēteris Ņikiforovs, Tilde Raivis. Word Alignment Based Parallel Corpora Evaluation and Cleaning Using Machine Learning Techniques. Conference: Volume: Proceedings of the 18th Annual Conference of the European Association for Machine Translation (EAMT 2015).

12. Mirjam Sepesy Maučec, Gregor Donaj. Machine Translation and the Evaluation of Its Quality. September 2019. In book: Natural Language Processing - New Approaches and Recent Applications [Working Title]

13. De Sutter, G., Cappelle, B., De Clercq, O., Loock, R., & Plevoets, K.. Towards a corpus-based, statistical approach of translation quality. *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 16, XX–YY. 2017.

14. Dragos Stefan Munteanu, Daniel Marcu. Improving Machine Translation Performance by Exploiting Non-Parallel Corpora. December 2005 *Computational Linguistics* 31(4):477-504

15. Adamou E. (2019). Corpus linguistic methods. J. Darquennes, J. Salmons & W. Vandebussche (eds). *Language contact*. Boston & Berlin: Mouton de Gruyter.

16. Contrastive analysis and learner language: A corpus-based approach Stig Johansson, University of Oslo 2008.

УДК 81'322.2

**МЕТОДЫ ONE-HOT КОДИРОВАНИЯ И МЕШКА СЛОВ ПРИ  
ОБРАБОТКЕ КОРПУСА ТЕКСТОВ УЗБЕКСКОГО ЯЗЫКА*****Б. Б. Элов, Ш. М. Хамроева, Н. Ш. Матякубова,  
У. С. Йодгоров****Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои**Ташкент, Узбекистан*e-elov@navoiy-uni.uz, shaxlo.xamrayeva@navoiy-uni.uz,  
nailya89mm@mail.ru, yodgorov@navoiy-uni.uz

Компьютеры предназначены для обработки информации в числовой форме, но данные не всегда представляются в числовой форме. В данной статье описывается обработка данных в виде символов, слов и текста узбекского языка с применением методов ONE-HOT ENCODING и BAG-OF-WORDS. Как Alexa, Google Home и многие другие «умные» помощники понимают и реагируют на нашу речь сегодня? В этой статье представлены подходы к обработке текстового корпуса узбекского языка с помощью таких методов обработки текста, как Bag-of-words (BOW), кодирование ONE-HOT в области искусственного интеллекта и обработки естественного языка.

**Ключевые слова:** корпус узбекского языка, обработка текста, мешок слов (BOW), ONE-HOT кодирование.

**ONE-HOT ENCODING AND BAG-OF-WORDS METHODS IN  
PROCESSING THE UZBEK LANGUAGE CORPUS TEXTS*****Elov B. B., Hamroyeva Sh. M., Matyakubova N. Sh., Yodgorov U. S.****Tashkent State University of Uzbek Language**and Literature named after Alisher Navoi**Tashkent, Uzbekistan*e-elov@navoiy-uni.uz, shaxlo.xamrayeva@navoiy-uni.uz,  
nailya89mm@mail.ru, yodgorov@navoiy-uni.uz

Computers are designed to process information in digital or numerical form. But data is not always in numerical form. This article describes how to process data in the form of characters, words, and text, as well as the application of ONE-HOT ENCODING and BAG-OF-WORDS methods to the Uzbek language, among the methods of teaching a computer to process natural language. How do Alexa, Google Home, and many other “smart” assistants understand and respond to our speech today? This article presents the approaches of text processing of the Uzbek language corpus through text processing methods such as Bag-of-words (BOW), ONE-HOT encoding in the field of artificial intelligence called natural language processing.

**Keywords:** Uzbek language corpus, text processing, Bag-of-words (BOW), ONE-HOT encoding.

**Introduction.** Natural language processing is a subfield of artificial intelligence that helps machines understand and process human language. For most natural language processing (NLP) tasks, the most basic step is to convert words into numbers to understand and decode patterns in natural language. In NLP, this stage is called text representation [Naseem, 2021; Chai, 2023; Probiez, 2023].

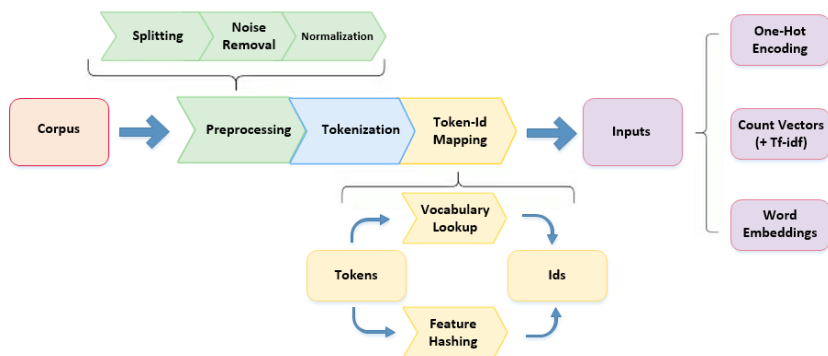
The “raw” text in the language corpus is pre-processed and converted into a suitable format for the machine learning model. Data is processed through tokenization, de-wording, punctuation removal, stemming, lemmatization, and a number of other primary processing NLP tasks (Figure 1). In this process, existing “noise” in the data is cleaned [ELov, 2023; ELov, 2023; ELov, 2022]. This cleaned data is presented in various forms (templates) according to the input requirements of the NLP application and machine learning model. Common terms used in text processing in NLP are:

**Corpus (Corpus, C):** a collection of data or multiple textual data together interpreted as a corpus.

**Vocabulary (V):** collection of all unique words in the corpus.

**Document (D):** A single text record of a dataset.

**Word(Word, W):** words in the dictionary.



**Figure 1.** Stages of initial processing of language corpus texts

Figure 1 shows the process of converting the corpus matrix to different input formats for the ML model. Starting from the left, a cor-

pus goes through several steps before obtaining tokens, a set of text building blocks, i.e. words, characters, etc. Since ML models are based on numerical value processing only, the tokens in the sentence are replaced by the corresponding numerical values. In the next step, they are converted to the various input formats shown on the right. Each of these formats has its pros and cons and should be chosen strategically based on the specifics of a given NLP task.

### **Types of text processing**

Although the process of text processing is iterative, it plays an important role for a machine learning model/algorithm. Text views can be divided into two parts [Siebers, 2022; Elov, 2022]:

1. Discrete text representations;
2. Distributed/Continuous text representations.

This article focuses on discrete text representations and introduces text processing methods using the Python package Sklearn.

### **Discrete views of text**

In the discrete representation of corpus texts, words in the corpus are represented independently of each other. In this approach, words are represented by indexes corresponding to their position in the vocabulary of the corpus(s). Methods belonging to this category are listed below [Naseem, 2021; Probiez, 2023; Siebers, 2022]:

- One-Hot encoding;
- Bag-of-words (BOW);
- CountVectorizer;
- TF-IDF
- Ngram.

### **One-Hot encoding method**

In the One-Hot encoding method, a vector consisting of 0 and 1 is assigned to each word in the corpus [Elov, 2022]. In the coding of this method, only one element of the vector is assigned – 1, and all other elements – 0. This value represents the element category. The resulting digital vectors are called hot vectors in NLP, and a unique hot vector is assigned to each word in the corpus. This action allows the machine learning model to recognize each word individually by its vector. One-Hot encoding method can be useful when there is a categorical feature in the data set. For example: The vector values corresponding to the sentence I like to read are expressed corresponding to each word in the sentence as follows:

Men → [1 0 0 0], o‘qishni → [0 1 0 0], yaxshi → [0 0 1 0], ko‘raman → [0 0 0 1] or,

$$\begin{array}{l} \text{Men:} \\ \text{o‘qishni:} \\ \text{yaxshi:} \\ \text{ko‘raman:} \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In this case, the sentence is expressed numerically as follows:

$$\text{sentence} = [ [1,0,0,0], [0,1,0,0], [0,0,1,0], [0,0,0,1] ]$$

In One-Hot encoding, each bit represents a possible category, and if a given variable does not belong to more than one category, one bit is sufficient to represent it. By this method, the words “Men” and “men” are matched with different vectors. By applying lowercase to all words in word processing, it is possible to match the same vector to uppercase and lowercase letters. In this method, the size of the one-dimensional vector is equal to the size of the dictionary.

When a corpus is encoded using the One-Hot encoding method, each word or token in the dictionary is converted into a digital vector. So, sentences in the corpus, in turn, become a matrix of size (p, q). In this,

- “p” is the number of tokens in the sentence;
- “q” is the size of the dictionary.

The size of the digital vector corresponding to the word in the One-Hot encoding method is directly proportional to the dictionary size of the corpus. So, with the increase in the size of the case, the size of the vector also increases. This method is not useful for large corpora, which may contain up to 100,000 or more unique words. We implement the One-Hot encoding method using the Sklearn package:

```
from sklearn.preprocessing import OneHotEncoder
import itertools
# 4 ta namunaviy hujjat
docs = ['Men NLP bilan ishlayman', 'NLP juda ajoyib texnologiya',
'Tabiiy tilni qayta ishlash', 'Zamonaviy texnologiyalar bilan ishlash']
# hujjatlarni tokenlarga ajratish
tokens_docs = [doc.split(" ") for doc in docs]
# tokenlar ro'yxatini umumlashtirish va so'zni identifikatoriga moslashtiradigan lug'atni yaratish
```

```

all_tokens = itertools.chain.from_iterable(tokens_docs)
word_to_id = {token: idx for idx, token in enumerate(set(all_to-
kens))}
# tokenlar ro'yxatini token-id ro'yxatlariga aylantirish
token_ids = [[word_to_id[token] for token in tokens_doc] for to-
kens_doc in tokens_docs]
# token-id ro'yxatlarini umumlashtirish
vec = OneHotEncoder(categories="auto")
X = vec.fit_transform(token_ids)
print(X.toarray())

[[0. 0. 0. 1. 0. 0. 1. 0. 0. 1. 0. 0. 1. 0.]
 [0. 0. 1. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 1. 0. 0.]
 [0. 1. 0. 0. 1. 0. 0. 0. 0. 1. 0. 1. 0. 0.]]

```

We show the advantages and disadvantages of this method in the table below:

| Advantages                       | Disadvantages   |
|----------------------------------|---|
| Easy to understand and implement | if the number of categories is very large, a large amount of memory is required                                     |
|                                  | the vector representation of words is orthogonal, and the relationship between different words cannot be determined |
|                                  | the meaning of the word in the sentence cannot be determined  |
|                                  | a large number of computations are required to represent a high-dimensional sparse matrix                           |

### Bag-of-words method

In the bag-of-words method, words from the corpus are placed in a “bag of words” and the frequency of each word is calculated. In this method, word order or lexical information is not taken into account to represent the text. In algorithms based on the BOW method, documents with similar words are returned as similar regardless of word placement.

The BOW method converts a text fragment into vectors of fixed length. Word frequency detection helps to compare documents. The BOW method can be used in a variety of NLP applications, such as

thematic modeling, document classification, and email spam detection. Below is the BOW vector corresponding to 2 Uzbek sentences.

| 1-sentence  | 2-sentence                              |
|---|---|
| “Adirlar ham bahorda lola bilan go‘zal, chunki lola – bahorning erka guli”. | “Lola ham shifokorlik kasbini tanladi”. |

|       | Adirlar | bahorda | lola | go‘zal | bahorning | erka | guli | shifokorlik | kasbini | tanladi |
|-------|---------|---------|------|--------|-----------|------|------|-------------|---------|---------|
| 1-gap | 1       | 1       | 2    | 1      | 1         | 1    | 1    | 0           | 0       | 0       |
| 2-gap | 0       | 0       | 1    | 0      | 0         | 0    | 0    | 1           | 1       | 1       |

The article “Using bag of words algorithm in natural language processing” written by B.Elov, N.Khudaiberganov and Z.Khusainova presents methods of converting Uzbek texts into digital form using the BoW algorithm [Fu, 2020].

**Conclusion.** Through Discrete Text Representation methods, each word in the corpus is considered unique and converted into a numerical form based on the various methods discussed above. The article presents several advantages and disadvantages of the different methods. We summarize them as a whole. Methods that generate discrete numerical values of text are easy to understand, implement, and interpret. Discrete representations of text are widely used in classical machine learning techniques and deep learning applications to solve NLP tasks such as document similarity, sentiment classification, spam classification, and topic modeling.

## REFERENCES

1. Naseem, U., Razzak, I., Khan, S. K., & Prasad, M. (2021). A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(5). <https://doi.org/10.1145/3434237>
2. Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3). <https://doi.org/10.1017/S1351324922000213>
3. Probiez, B., Hrabia, A., & Kozak, J. (2023). A New Method for Graph-Based Representation of Text in Natural Language Processing. *Electronics*, 12(13). <https://doi.org/10.3390/electronics12132846>



4. B.Elov, E.Adali, Sh.Khamroeva, O.Abdullayeva, Z.Xusainova, N.Xudayberganov (2023). The Problem of Pos Tagging and Stemming for Agglutinative Languages. *8 th International Conference on Computer Science and Engineering UBMK 2023, Mehmet Akif Ersoy University, Burdur – Turkey.*

5. B.Elov, Sh.Khamroeva, Z.Xusainova (2023). The pipeline processing of NLP. *E3S Web of Conferences 413, 03011, INTERAGROMASH 2023.* <https://doi.org/10.1051/e3sconf/202341303011>

6. B.Elov, Sh.Hamroyeva, X.Axmedova. Methods for creating a morphological analyzer. *14th International Conference on Intelligent Human Computer Interaction, IHCI 2022, 19-23 October 2022, Tashkent.* [https://dx.doi.org/10.1007/978-3-031-27199-1\\_4](https://dx.doi.org/10.1007/978-3-031-27199-1_4)

7. Siebers, P., Janiesch, C., & Zschech, P. (2022). A Survey of Text Representation Methods and Their Genealogy. *IEEE Access, 10.* <https://doi.org/10.1109/ACCESS.2022.3205719>

8. B.Elov, Z.Xusainova, N.Xudayberganov. Tabiiy tilni qayta ishlashda Bag of Words algoritmidan foydalanish. *O‘zbekiston: til va madaniyat (Amaliy filologiya), 2022, 5(4).* <http://aphil.tsuull.uz/index.php/language-and-culture/article/download/32/29>

9. B.Elov, Z.Xusainova, N.Xudayberganov. O‘zbek tili korpusi matnlari uchun TF-IDF statistik ko‘rsatkichni hisoblash. *SCIENCE AND INNOVATION INTERNATIONAL SCIENTIFIC JOURNAL VOLUME 1 ISSUE 8 UIF-2022: 8.2 | ISSN: 2181-3337*

[https://www.academia.edu/105829396/OZBEK\\_TILI\\_KORPUSI\\_MATNLARI\\_UCHUN\\_TF\\_IDF\\_STATISTIK\\_KORSATKICHNI\\_HISOBLASH](https://www.academia.edu/105829396/OZBEK_TILI_KORPUSI_MATNLARI_UCHUN_TF_IDF_STATISTIK_KORSATKICHNI_HISOBLASH)

10. Fu, Y., & Yu, Y. (2020). Research on text representation method based on improved TF-IDF. *Journal of Physics: Conference Series, 1486(7).* <https://doi.org/10.1088/1742-6596/1486/7/072032>

УДК 81'322.2

**ОПРЕДЕЛЕНИЕ ОМОНИМОВ В УЗБЕКСКОМ ЯЗЫКЕ  
С ПОМОЩЬЮ АЛГОРИТМА ЛЕСКА****Б. Б. Элов, Х. И. Ахмедова***Ташкентский государственный университет узбекского языка и  
литературы им. Алишера Навои**Ташкент, Узбекистан*

e-elov@navoiy-uni.uz, xolisa9029@mail.ru

Решение задачи лексической многозначности в настоящее время является одной из наиболее актуальных задач компьютерной лингвистики. Особенно актуально это для тюркских языков, которые относятся к малоресурсным языкам и не обладают такими семантическими ресурсами, как WordNet, VerbNet, FrameNet и др. В языках обладающих подобными электронными лингвистическими ресурсами они используются во многих прикладных задачах семантической обработки текстов.

Одним из известных способов решения лексической многозначности является алгоритм Леска. В данной статье предлагается реализация алгоритма Леска с помощью известного ресурса WordNet. Таким образом, эта статья еще раз подтверждает актуальность создания онтологических графов знаний для тюркских языков, в число которых входит и узбекский язык.

**Ключевые слова:** семантический анализатор, омонимия, метод на основе правил, статистический метод, алгоритм Леска, весовое слово, WordNet.

**DETERMINING HOMONYMS IN THE UZBEK LANGUAGE USING  
THE LESK ALGORITHM****Botir Elov, Xolisa Axmedova***Alisher Navoi' Tashkent State University of the Uzbek Language  
and Literature**Tashkent, Uzbekistan*

e-elov@navoiy-uni.uz, xolisa9029@mail.ru

Solving the problem of lexical ambiguity is currently one of the most pressing problems in computer linguistics. This is especially true for Turkic languages, which are low-resource languages and do not have such semantic resources as WordNet, VerbNet, FrameNet, etc. In languages that have such electronic linguistic resources, they are used in many applied tasks of semantic text processing.

One of the well-known methods for solving lexical ambiguity is the Lesk algorithm. This article proposes an implementation of the Lesk algorithm using the well-known WordNet resource. Thus, this article once again confirms the

relevance of creating ontological knowledge graphs for Turkic languages, which include the Uzbek language.

**Keywords** semantic analyzer, homonymy, rule-based method, statistical method, Lesk's algorithm, weight word, WordNet.

## LESK ALGORITMI YORDAMIDA O'ZBEK TILIDAGI OMONIM SO'ZLARNI ANIQLASH

*Elov B. B., Axmedova X. I.*

*Alisher Navoiy nomidagi Toshkent davlat  
o'zbek tili va adabiyoti universiteti, Toshkent, O'zbekiston  
e-[elov@navoiy-uni.uz](mailto:elov@navoiy-uni.uz), [xolisa9029@mail.ru](mailto:xolisa9029@mail.ru)*

Leksik noaniqlik muammosini hal qilish hozirgi vaqtda kompyuter tilshunosligining eng dolzarb muammolaridan biridir. Bu, ayniqsa, kam resursli va WordNet, VerbNet, FrameNet va boshqalar kabi semantik resurslarga ega bo'lmagan turkiy tillar uchun to'g'ri keladi. Bunday elektron lingvistik resurslarga ega bo'lgan tillarda semantik matnni qayta ishlash kabi ko'plab amaliy vazifalarda qo'llaniladi.

Leksik noaniqlikni yechishning mashhur usullaridan biri Lesk algoritmidir. Ushbu maqola taniqli WordNet resursidan foydalangan holda Lesk algoritmini amalga oshirishni taklif qiladi. Shunday qilib, ushbu maqola turkiy tillar, jumladan o'zbek tilini ham o'z ichiga olgan ontologik bilim grafiklarini yaratishning dolzarbligini yana bir bor tasdiqlaydi.

**Kalit so'zlar.** semantik analizator, omonimiya, qoidalarga asoslangan usul, statistik usul, Lesk algoritmi, so'z vazni, WordNet.

### I. Kirish

Tabiiy tilni avtomatik qayta ishlash muammosi yarim asrdan ko'proq vaqt davomida dolzarb bo'lib qolmoqda. Muammoning murakkabligi va aniq g'oyaning yo'qligi uni hal qilish yo'llarining qiyinligini ko'rsatadi. Matn, nutq va paralingvistik vositalarni tanib olish uchun barcha yangi tizimlar ishlab chiqilmoqda. Matnni qayta ishlash ushbu sohadagi eng qadimgi va eng muhim tadqiqotlardan biridir. Matnni avtomatik qayta ishlash bo'yicha birinchi tadqiqotlar XX asrning 50-yillariga to'g'ri keladi. Matnni avtomatik qayta ishlash bir necha bosqichlarga bo'linadi, ulardan biri morfologik tasnifdir. Bu bosqichda har bir so'z uchun morfologik tavsiflar (jins, son, hol, tuslanish, tur va hokazo) va lemma deb ataladigan so'zning boshlang'ich shakli aniqlanadi. Morfologik tasnifni omonimiya hodisasi murakkablashtiradi.

Ayrim flektiv tillardagi matnlar uchun ehtimollik modellaridan foydalanishga asoslangan omonimiyani aniqlash usullari juda keng tarqalgan bo'lsa-da, ular juda yuqori aniqlikni ta'minlaydi. Yashirin Markov modeli ruscha matnlardagi omonimlarni aniqlash uchun yaxshiroq ishlashi isbotlangan.

Semantik qidiruv semantik tahlil orqali amalga oshiriladi. U qanchalik yaxshi ishlab chiqilgan bo'lsa, qidiruv shunchalik samarali bo'ladi. Semantik tahlilni amalga oshirish bevosita lingvistik resurslarga bog'liq. Leksik manbalarga lug'atlar, tezauruslar va ontologiyalar kiradi. Semantik tahlil ham alohida o'rganishni talab qiladigan elementlarga ega. Ushbu maqolada homonimiyani aniqlash muammosini hal qilish haqida so'z boradi. Omonimiya semantik tahlilning muhim elementlaridan biridir. Omonimiyani aniqlash turli tabiiy tillarda turlicha talqin qilinadi. Jahon kompyuter tilshunosligida gaplarni semantik tahlil qilishda asosan 3 ta usuldan foydalaniladi:

**Qoidaga asoslangan usul** - bu tabiiy tilning grammatik xususiyatlariga asoslanib, oldindan belgilangan til qoidalariga asoslangan omonimiyani aniqlash.

**Statistik ma'lumotlarga asoslangan usulni** til korpusi ma'lumotlari asosida qaror qabul qilish usuli deb ham atash mumkin. Ya'ni, statistik ma'lumotlar til korpusidagi ma'lumotlar orasida olib borilgan kuzatishlar asosida olinadi. Olingan statistik ma'lumotlar asosida yangi omonim baholanadi. Statistik usullar yordamida omonimiyani aniqlash muammosi gaplarni POS (Past of Speech) yorlig'i muammosini hal qilish jarayonida o'z yechimini topadi. POS teglash - bu yangi kiritilgan matndagi har bir so'zni otlar, fe'llar, sifatlar va boshqalar kabi tegishli POS teglari bilan bog'lash jarayonidir. Bu vazifa NLP (tabiiy til jarayoni)dagi ma'noni aniqlash vazifalaridan biridir. Buning sababi shundaki, tildagi ko'p so'zlar bir necha xil ma'noga ega bo'lishi mumkin.

Shuning uchun ham turli gap bo'laklari orasidagi omonimlarni aniqlashda statistik usullardan foydalanish samarali natijalar beradi.

**Mashinani o'rganishga asoslangan usul** nafaqat statistik ma'lumotlardan foydalanadi, balki bevosita neyron tarmoqlarga ham tegishli usuldir. Mashinani o'rganishga asoslangan yondashuv, o'z navbatida, nazorat qilinadigan va nazoratsiz algoritmlarga bo'linadi. Omonimiyani aniqlashda ushbu yondashuvni qo'llash orqali ham yaxshi natijalarga erishish mumkin. Ushbu usullar yordamida so'zning ma'nosini aniqlash mumkin. Tabiiy tildagi so'zlarning ma'nosini aniqlash murakkab va muhim vazifadir. So'zning ma'nosini aniqlash mashina tarjimasining aniqlik darajasini oshirishga imkon beradi va matn, asarlar, ilmiy maqolalar va dissertatsiyalar uchun avtomatik annotat-

siya yaratadi. Soʻz maʼnosini aniqlashdagi eng muhim vazifalardan biri omonim soʻzlarni semantik jihatdan farqlashdir. Gapdagi omonim soʻzning maʼnosini aniqlash.

## II. Material va metodlar

Xorijiy tajribani chuqur oʻrganib, oʻzbek omonimlarini farqlashda qoidaga asoslangan, stokastik, mashina oʻrganish va neyron tarmoq usullaridan foydalanamiz. Oʻzbek tilida omonimlarni farqlashda ularni gap boʻlaklari ichida kelishiga koʻra bir boʻlak ichidagi omonimlar, ikki boʻlak boʻlaklari, uch boʻlaklari, toʻrt boʻlaklari kabi guruhlarga ajratdik. Grammatik jihatdan bir-biriga oʻxshamaydigan soʻz turkumlari ichida omonimiyani aniqlash uchun qoidaga asoslangan usuldan foydalandik. Bu haqda [Boltayevich, 2022: 278-283-b; Elov, 2022: 150-162-b; Axmedova: 393-400-b] keltirilgan ilmiy maqolalarda aytib oʻtgan edik.

Turli soʻz turkumlari oʻrtasidagi omonimiyani aniqlash muammosi gaplarni POS (Part of speaking tagging) teglash jarayonida hal qilinadi. Soʻz turkumlarini aniqlashda Hidden Markov modelidan foydalanish boʻyicha koʻplab ilmiy maqolalarni topish mumkin. Markov modelidan foydalanish jarayonida Viterbi algoritmidan foydalangan holda natijalar aniqroq boʻlishini koʻrish mumkin[12; 57-62-b][13;105-112-b].

Turli xil tabiiy tillarda bir xil soʻz turkumida uchraydigan omonimlar ham mavjud.

**1-jadval: Turli soʻz turkumlari orasidagi omonim soʻzlar**

| Soʻz | Soʻz turkumlari | Maʼnosi                                     |
|------|-----------------|---|
| Oʻt  | Feʼl            | Oʻtmoq feʼli                                |
|      | Ot              | Maysa, oʻt-oʻlan                            |
|      | Ot              | Inson oʻrgani                               |
|      | Ot              | Olov  |
| Oz   | Ravish          | Kam, miqdori nisbatan koʻp boʻlmagan        |
|      | Feʼl            | Oriqlamoq, etidan yoʻqotmoq                 |
|      | Feʼl            | Noxush boʻlmoq, kuchsizlanmoq, holsizlanmoq |
|      | Feʼl            | Adashmoq, toʻgʻri yoʻldan chetga chiqmoq    |
| Boʻy | Ot              | Uzunlik oʻlchovi                            |
|      | Ot              | Hid, is                                     |
| ...  | ...             | ...   |

1-jadvalda ko'rsatilganidek, nutqning bir qismida yoki hatto turli so'z turkumlari orasida omonim hosil qila oladigan so'zlar bo'lishi mumkin. Omonimlarning so'z turkumini aniqlashda Frequentist, Naïve Bayes, Hidden Markov modeli kabi usullardan foydalanish mumkin [11,44-54-b.].

Lekin bir gap bo'lagi doirasidagi omonimlarni semantik farqlashda bu usullardan foydalanish samarali emas. Bu so'zlar jumlada aniq nimani anglatishini aniqlash muhimdir. Bu masalani hal qilish mashina tarjimasining aniqligini oshirishga, jumlaning qisqacha mazmunini, jumla-matn va katta hajmdagi asarlarning matn-matn konspektini aniqlashga yordam beradi. So'z turkumidagi omonim so'zlarning ma'nosini aniqlash masalasi xorijiy manbalarda so'z ma'nosini ajratish (WSD- word sense disambiguation) deb ataladi.

So'z ma'nosini aniqlash (WSD) usullari kirishning semantik talqinini talab qiladigan ko'plab NLP vazifalari uchun foydalidir. Bundan tashqari, bunday usullar leksikografik tadqiqotlar va til o'rganish resurslari uchun muhim bo'lgan turli korpuslardagi so'zlarning turli ma'nolarining chastotasini baholashga yordam beradi. Rus tilidagi polisemantik fe'llarning ma'nosiga oid oldingi tadqiqotlar muhim va qiziqarli natijalar bergan bo'lsa-da, u asosan noaniqlikni kamaytirish yoki tez-tez uchraydigan ma'nolarni aniqlashga qaratilgan. maqsadli edi, lekin so'z ma'nosini aniqlashning to'g'riligini baholashga emas. Ma'lumotlarga ko'ra, ruscha fe'llar uchun yarim nazoratli so'z ma'nosini tanib olishni amalga oshiradigan har tomonlama baholangan usul mavjud emas. Usulning turli xil variantlarini solishtirish va uning cheklovlarini tahlil qilish mumkin. Leksik-semantik noaniqlik har qanday tabiiy tilga xos xususiyatdir, shuning uchun so'z ma'nosini ajratish ko'plab tabiiy tillarni qayta ishlash vazifalarining muhim qismidir. SemEval sessiyalarida (Pradhan va boshq. 2007) va WSD so'rovlarida (Ide and Véronis 1998; Navigli 2009; Mihalcea 2011) turli WSD algoritmlari muhokama qilingan. Eng zamonaviy va istiqbolli yondashuvlar - bu allaqachon mavjud resurslardan foydalanadigan va inson ishtirokini talab qilmaydigan yondashuvlar. Qoidalarga asoslangan yondashuvlar tezaurusdan foydalanadi. Nazorat qilinmagan korpusga asoslangan yondashuvlar odatda ma'nolar inventariga aniq havola qilmasdan korpusda klasterlashni amalga oshiradi. Korpus ma'lumotlari asosida so'zlarning ma'nosini aniqlash uchun LESK algoritmidan foydalanish jarayonini ko'rib chiqamiz.

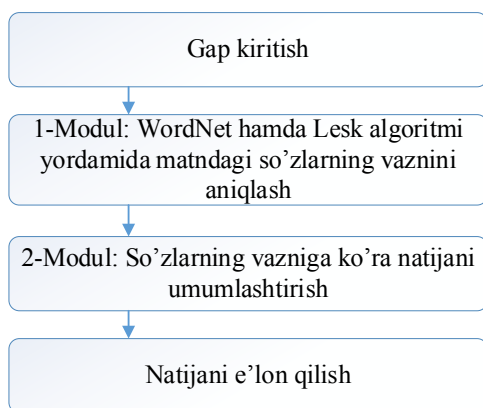
Lesk algoritmi 1986 yilda E. Maykl Lesk tomonidan kiritilgan va klassik WSD algoritmi hisoblanadi. Lesk algoritmi faqat matnning ma'lum bir qismidagi so'zlar o'xshash ma'noga ega degan fikrga aso-

slanadi. Soddalashtirilgan Lesk algoritmidagi har bir soʻz kontekstining toʻgʻri maʼnosi berilgan kontekst va uning lugʻat maʼnosi oʻrtasidagi eng oʻxshash maʼnoni topish orqali topiladi. Ushbu algoritmi hind tilidagi soʻzning maʼnosini aniqlash uchun ishlatilgan. U savol-javob tizimlari va hissiyotlarni tahlil qilish tizimlarini ishlab chiqishda ishlatilgan [Tripathi, 2021: 939-954-b]. Zouaghi, A., Merhbene, L. va boshqalar tomonidan arabcha soʻzlarning maʼnosini aniqlash uchun foydalanilgan. Oʻtkazilgan tadqiqot davomida 73% aniqlikka erishildi [Zouaghi, 2012: 257-269-b]. Basuki, S., Xolimi, A. S. va boshqalar. Indoneziya omograflarining semantik identifikatsiyasidan foydalanilgan. Natijada bir soʻz turkumidagi soʻzlarni aniqlashda 78,6%, ikki soʻz turkumidagi omonim soʻzlarni aniqlashda 62,5% aniqlikka erishildi [Basuki, 2019: 8-15-b]. Lesk algoritmi tabiiy tilning WordNet-ga asoslangan qoidaga asoslangan usuldir. Lesk algoritmi yordamida omonimiyani aniqlash uchun quyidagi maʼlumotlar talab qilinadi:

**Maʼnolar toʻplami:** koʻp maʼnoli soʻzning mavjud maʼnolari yigʻindisi.

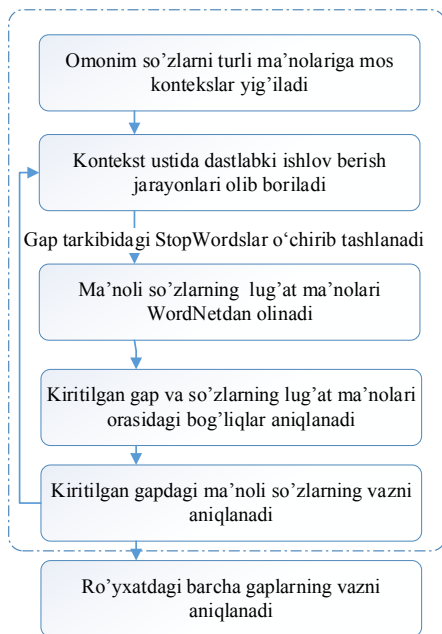
**Kontekstlar toʻplami:** soʻzning har bir maʼnosi uchun kontekstlar toʻplami.

Lesk algoritmi yordamida omonimiyani aniqlash jarayoni ikki moduldan iborat.



1-rasm: Lesk algoritmidagi modullar

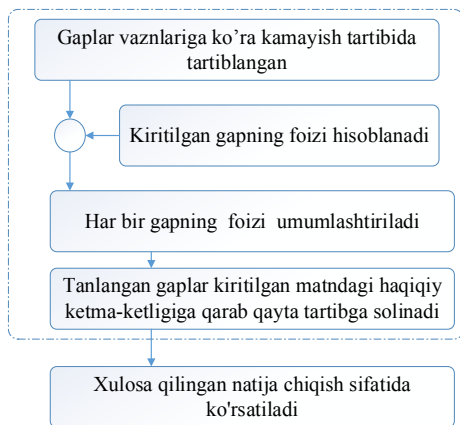
1-modul: Keyingi qadam bu soʻzlarning har bir maʼnosi uchun kontekst yaratishdir. Buning uchun korpus maʼlumotlari ishlatiladi. Dastlab tarkibida omonimlar boʻlgan kontekstlar ajratiladi va ular omonim soʻzining maʼnosiga qarab ajratiladi. Oldindan ishlov berish ajratilgan kontekstlarda amalga oshiriladi. (2-rasm).



2-rasm: 1-modulning vazifalari

1-modul: bu algoritm  $n$  ta gap uchun  $O(n^3)$  vaqt oladi va gaplardagi o'xshash so'zlar sonini aniqlash uchun  $O(n^2)$  amallarni bajaradi.

2-modulda 1-modul natijalari umumlashtiriladi va umumlashtiriladi.



3-rasm: 2-modulning vazifalari



Quyidagi gap yordamida Lesk algoritmi yordamida noaniq soʻzlarni semantik farqlash ketma-ketligini koʻrib chiqamiz. “*Bu fe’ling bilan hammani qon qilasan-ku*”

Berilgan gapda “fe’l” va “qon” omonimlari mavjud. Bu soʻzlarning maʼnolari toʻplami maʼlumotlar bazasiga kiritilgan (2-jadval).

## 2-jadval: gapdagi omonimlar va ularning maʼnolari

| Soʻz | Soʻz turkumlari | Maʼnosi   |
|------|-----------------|---|
| Feʼl | Ot              | Harakat, xulq-atvor, xarakter   |
|      |                 | Grammatik termin  |
| qon  | Ot              | Organizm tomiridan yurak faoliyati bilan harakatlanuvibi taʼminlovchi qizil rangli suyuqlik |
|      | Feʼl            | Toʻymoq, qoniqmoq   |

Soʻzlarning vaznini aniqlashda har bir maʼno uchun soʻzlar sonini sanash orqali amalga oshiriladi. Kontekstlar tomonidan olingan maʼlumotlar maʼlumotlar bazasida saqlanadi. Yuqorida

### “Bu fe’ling bilan hammani qon qilasan-ku”

gapdagi omonimlarning maʼnosini Lesk algoritmidan foydalanib aniqlaymiz. Kiritilgan gapda quyidagi harakatlar bajariladi.

*Tokenizatsiya;*

*Lemmatizatsiya;*

*StopWords-ni oʻchirish;*

*POST belgisi;*

Natijada

### Feʼl, bilan, qon, qilmoq

soʻzlar qoladi. Keyingi bosqichda kiritilgan gapdagi omonim soʻz aniqlanadi va uning birikmalari ajratiladi. Ajratilgan birikmalarning ogʻirligi maʼlumotlar bazasidan aniqlanadi. Masalan, “fe’l” konjugatsiyalari va ularning ogʻirliklaridan iborat maʼlumotlar toʻplami 3-jadvalda keltirilgan.

**3-jadval: ‘fe’l’ so‘zining semantik birikmalari va ularning vazni**

| BIRIKUVCHILAR | 1-MA’NO | 2-MA’NO |
|---------------|---------|---------|
| <b>YOMON</b>  | 10      | 2       |
| <b>YAXSHI</b> | 15      | 1       |
| <b>OT</b>     | 0       | 25      |
| <b>QURMOQ</b> | 35      | 0       |
| <b>TOR</b>    | 18      | 8       |
| <b>QON</b>    | 20      | 12      |
| <b>QILMOQ</b> | 14      | 18      |

Xuddi shu tarzda “qon” omonim so‘zining birikmalari va ularning vaznlaridan iborat ma’lumotlar to‘plami olinadi..

**4-jadval: “qon” so‘zining semantik birikmalari va ularning vazni**

| BIRIKUVCHILAR    | 1-MA’NO | 2-MA’NO |
|------------------|---------|---------|
| <b>SUV</b>       | 10      | 35      |
| <b>TOMIR</b>     | 123     | 0       |
| <b>KENGAYMOQ</b> | 45      | 5       |
| <b>FE’L</b>      | 47      | 26      |
| <b>QILMOQ</b>    | 31      | 11      |
| ...              | ...     | ...     |

Berilgan ma’lumotlardan gapdagi omonim so‘zning birikmalari va ularning vazni aniqlanadi. Aniqlangan ma’lumotlardan eng yuqori og‘irlikdagi birikmaning ma’nosi tanlanadi.

Gapning vazni jumlada mavjud bo‘lgan birikmalarning har birining vazniga qarab belgilanadi. Bu voqea qo‘shma hodisa bo‘lgani uchun gapning vazni vaznlarning ko‘paytmasiga teng bo‘ladi

$$p = \prod_{i=1}^n p_i$$

Bu yerda  $n$  – har bir gapdagi so‘zlar soni,  $p_i$  esa gapdagi  $i$ - so‘zining bir ma’noni anglatish ehtimoli. Bu ehtimollik shartli ehtimollik deb

ham ataladi. Bu shartli ehtimol gapdagi har bir soʻz uchun hisoblab chiqiladi va gapning ehtimolligi umumlashtirish yoʻli bilan aniqlanadi.

### III. XULOSA

Lesk algoritmi yordamida omonimiyani aniqlash ham qoidaga asoslangan usulning tarkibiy qismidir. Ushbu algoritmdan foydalanish uchun tabiiy tilning WordNet tizimi mavjud boʻlishi kerak. Aniqroq aytganda, Lesk algoritmi WordNet asosida ishlaydi. Lesk algoritmi yordamida omonimlikni aniqlash uchun omonim soʻzlar har bir maʼnosiga koʻra kontekstlarda inson omili yordamida semantik belgilar qoʻyiladi. Ushbu algoritmnning vazifasi semantik teglangan kontekstdagi noyob soʻzlar sonini aniqlash va xulosa qilishdir. Lesk algoritmi qoidaga asoslangan usulning bir qismi boʻlib, uning yordamida turli soʻz turkumlaridagi omonimlarni emas, balki bir xil soʻz turkumidagi omonimlarni ham semantik jihatdan farqlash mumkin. Ushbu algoritmdan foydalanib, omonimli jumalarning katta toʻplami kerak boʻladi. Ushbu algoritmi har bir omonim uchun 200-2000 ta jumalar toʻplami bilan ishlatilishi mumkin.

### ADABIYOT

1. Boltayevich, E. B., & Ilxomovna, A. X. (2022). Business Process Modeling That Distinguishes Homonymy Within Three Parts of Speeches in The Uzbek Language. In Proceedings - 7th International Conference on Computer Science and Engineering, UBMK 2022 (pp. 278–283). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/UBMK55850.2022.9919453>
2. Elov B.B., Axmedova X.I. Uchta soʻz turkumi doirasidagi omonimiyani farqlovchi biznes jarayonni modellashtirish// Oʻzbekiston respublikasi innovatsion rivojlanish vazirligining, Ilm-fan va m innovasion rivojlanish ilmiy jurnal 2022 / 1, 150-162-b.
3. Axmedova X. I. Turli soʻz turkumlari orasidagi omonimiyani aniqlovchi matematik modellar// Science and innovation international scientific journal volume 1 issue 7 uif-2022: 8.2 | issn: 2181-3337. <https://doi.org/10.5281/zenodo.7238546>
4. Axmedova X.I. Chastotali usul yordamida omonimiyani aniqlash// “OʻZBEK AMALIY FILOLOGIYASI ISTIQBOLLARI” Respublika ilmiy-amaliy konferensiyasi Toshkent: 2022. – 164–170 b.
5. Elov B.B., Axmedova X.I. Determining homonymy using statistical methods.// “Hisoblash modellari va texnologiyalari (HMT 2022)”

O‘zbekiston-Malayziya ikkinchi xalqaro konferensiyasi materiallari-Toshkent, 2022 16-17 sentabr,-106 b.

6. Uri Roll, Ricardo A. Correia, Oded Berger-Tal// Using machine learning to disentangle homonyms in large text corpora- Conservation Biology 31 October 2017 <https://doi.org/10.1111/cobi.13044>

7. Tripathi, P., Mukherjee, P., Hendre, M., Godse, M., & Chakraborty, B. (2021). Word Sense Disambiguation in Hindi Language Using Score Based Modified Lesk Algorithm. International Journal of Computing and Digital Systems, 10(1), 939–954. <https://doi.org/10.12785/IJCDS/100185>

8. Zouaghi, A., Merhbene, L., & Zrigui, M. (2012). Combination of information retrieval methods with LESK algorithm for Arabic word sense disambiguation. Artificial Intelligence Review, 38(4), 257–269. <https://doi.org/10.1007/s10462-011-9249-3>

9. Basuki, S., Kholimi, A. S., Minarno, A. E., Sumadi, F. D. S., & Effendy, M. R. A. (2019). Word Sense Disambiguation (WSD) for Indonesian homograph word meaning determination by LESK Algorithm Application. In Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019 (pp. 8–15). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICTS.2019.8850957>

10. J.Y. Park, Shin, H.J.; Lee, J.S. Word Sense Disambiguation Using Clustered Sense Labels. Appl. Sci. 2022, 12, 1857. <https://doi.org/10.3390/app12041857>

11. B.B. Elov, X.I. Axmedova “So‘z ma’nosini aniqlashda naive bayes algoritmidan foydalanish”, илм-фан ва инновацион ривожланиш ilmiy jurnali, Toshkent, 3/2023,44-54-b.

12. Elov Botir Boltayevich, Sirojiddinov Shuhrat Samariddinovich, Khamroeva Shahlo Mirdjonovna, Eʃref Adalı, Xusainova Zilola Yuldashevna. “Pos Taging of Uzbek Text Using Hidden Markov Model”, 8 th International Conference on Computer Science, 13-14-15-September 2023, Burdur-Turkey, 57–62-pp.

13. Elov B., Hamroyeva Sh., Xusainova Z., Xudayberganov N., Yodgorov U., Yuldashev A. “POS TAGGING OF UZBEK TEXTS USING HIDDEN MARKOV MODELS (HMM) AND VITERBI ALGORITHM”. “O‘zbekiston Milliy universitetining ilm-fan rivoji va jamiyat taraqqiyotida tutgan o‘rni” mavzusidagi xalqaro ilmiy-amaliy konferensiya, 2023 yil, 12 may, 104-115-b

УДК 81'322.2

## ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕКСТОВЫХ ДАННЫХ

*М. Х. Примова**Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои**Ташкент, Узбекистан*

primovamastura@navoiy-uni.uz

Интеллектуальный анализ текстовых данных – один из важнейших способов анализа и обработки неструктурированных данных, на долю которого приходится около 80% мировых данных. Сегодня большинство организаций и учреждений собирают и хранят большие объемы данных в хранилищах данных и облачных платформах. Каждую минуту в информационную систему поступают новые данные из множества источников, и она продолжает расти в геометрической прогрессии. Сегодня обработка больших объемов текстовых данных является актуальной задачей NLP. В данной статье представлены приложения для анализа больших объемов текстовых данных.

**Ключевые слова:** Text Mining, Big Data, information extraction, NLP, information retrieval.

## INTELLECTUAL ANALYSIS OF TEXT MINING

*Mastura Primova**Tashkent State University of Uzbek Language and Literature  
name after Aisher Navai.**Tashkent, Uzbekistan*

primovamastura@navoiy-uni.uz

Text mining is one of the most important methods for analyzing and processing unstructured data, accounting for about 80% of the world's data. Today, most organizations and institutions collect and store large amounts of data in data warehouses and cloud platforms. Every minute, new data from many sources enters the information system, and it continues to grow exponentially. Today, processing large volumes of text data is one of the most important NLP tasks. This article presents applications for analyzing large volumes of text data mining.

**Keywords:** Text Mining, Big Data, information extraction, NLP, information retrieval.

## MATNLI MA'LUMOTLARINI INTELLEKTUAL TAHLIL QILISH

*Primova Mastura Hakim qizi**Alisher Navoiy nomidagi Toshkent davlat  
o'zbek tili va adabiyoti universiteti**Toshkent, O'zbekiston*

primovamastura@navoiy-uni.uz

Matnli ma'lumotlarni intellektual tahlil qilish – strukturalanmagan ma'lumotlarni tahlil qilish va qayta ishlashning eng muhim usullaridan biri bo'lib, dunyodagi ma'lumotlarining qariyb 80 foizini tashkil qiladi. Bugungi kunda ko'pchilik tashkilot va muassasalar ma'lumotlar omborlari va bulutli platformalarda katta hajmdagi ma'lumotlarni to'playdi, saqlaydi. Har daqiqada bir nechta manbalardan yangi ma'lumotlar axborot tizimiga qabul qilinadi va ular eksponent ravishda o'sishda davom etadi. Bugungi kunda katta hajmdagi matnli ma'lumotlarni qayta ishlash NLPning dolzarb vazifasi hisoblanadi. Ushbu maqolada katta hajmdagi matnli ma'lumotlarni tahlil qilish ilovalarikeltililadi.

**Kalit so'zlar:** Text Mining, Big Data, information extraction, NLP, information retrieval.

## Kirish

**Text Mining** – matnli ma'lumotlarni intellektual tahlili bo'lib, tabiiy tilni qayta ishlash (NLP) vazifalaridan biri hisoblanadi. Ushbu jarayonda strukturalanmagan matn qayta ishlanib strukturalangan va ma'noga ega formatga keltiriladi [Aggarwal, 2013]. *Naïve Bayes*, *Support Vector Machines (SVM)* va boshqa chuqur o'rganish algoritmlarini qo'llash natijasida katta hajmdagi strukturalanmagan ma'lumotlardagi yashirin munosabatlar, qonuniyatlar aniqlanadi va tadqiq qilinadi [Kao, 2005]. Elektron pochta xabarlar, hujjatlar, matnli xabarlar fayllar orqali yaratilgan barcha ma'lumotlar umumiy matnda jamlanadi.

### Matnli intellektual tahlil qilish ilovalari

So'ngi yillarda matnli Text Mining sohasi jadal suratda rivojlandi. *Elektron tijorat tizimlari veb-saytlari, ijtimoiy media platformalari, nashr etilgan maqolalar va qidiruv tizimlaridagi so'rovlar* katta hajmdagi (BigData) strukturalanmagan ma'lumotlarni hosil qiladi [Hassani, 2020; Salloum, 2017]. Bu turdagi ma'lumotlarni qo'lda qayta ishlash jurakkab amallarni talab etadi. Shu sababali Text Mining metodlari vositasida strukturalanmagan ma'lumotlarni qayta ishlash bugungi kunda muhim ahamiyat kasb etadi. Matnli intellektual tahlil qilish uchun quyida ilovalar mavjud:



1-rasm. Text Mining ilovalar

### • Ochiq-yopiq soʻrov javoblarni tahlil qilish

Foydalanuvchi tomonidan oʻrganayotgan mavzular doirasida soʻrovlar koʻpincha turli xil ochiq savollarni berishadi. Bu esa, foydalanuvchilar oʻz “fikri”ni bildirishadi. Bundan tashqari, fikrlar maʼlum oʻlchamlar yoki maʼlum bir javob formati bilan cheklanmaydi.

### • Elektron pochta xabarleri va boshqa xabarleri avtomatik qayta ishlash

Yana bir keng tarqalgan dastur matnni avtomatik tasniflashga yordam beradi. Ushbu dastur orqali keraksiz elektron pochta xabarlerini avtomatik ravishda “filtirlash”, kiruvchi elektron pochta xabarini aniqlash va oʻchirilish, elektron pochta xabarleri mos boʻlimga yoʻnaltirish, elektron pochta xabarleri “nomaqbul” xabarlar va avtomatik ravishda joʻnatuvchilarga “xafa” qiluvchi soʻzlarga tekshirish.

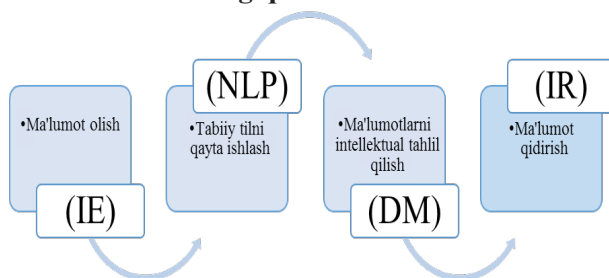
### • Kafolat yoki sugʻurta daʼvolarini, tibbiy diagnostik natijalar va boshqalarni tahlil qilish

Katta hajmdagi maʼlumotlar bazasida ayrim sohalariidagi maʼlumotlarning katta qismi ochiq holda toʻplanadi. Kafolat daʼvolari yoki tibbiy diagnostik natijalari elektron shaklda umumlashtirib toʻplanadi. Bu turdagi maʼlumotlardan foydalanish koʻpgina qulayliklarni beradi. Masalan, tibbiyot sohasida bemorlarning oʻz belgilarini tavsifini ochadi va tibbiy tashxis uchun foydali maslahatlar beradi.

### • Raqobatchilarni veb-saytlarini skanerlash orqali oʻrganish

Boshqa turdagi ilovalar maʼlum bir domendagi veb-sahifalar tarkibini qayta ishlash uchun moʻljallangan. Veb-sahifani oʻrnatgandan soʻng barcha veb-sahifalar qayta ishlashi uchun u yerda joylashgan havolalarni “skanerlashni” boshlashingiz mumkin. Ushbu saytda mavjud boʻlgan shartlar va hujjatlar roʻyxatini yuklab olib, tavsiflangan eng muhim atamalar va xususiyatlar aniqlanadi.

## Big Data da Text Mining qoʻllanish sohalari



2-rasm. Big Data da Text Mining qoʻllanish sohalari

Matnli ma'lumotlarni tahlil qilish quyidagi masalalarni hal qilishda qo'llanilmoqda:

– **Ma'lumot olish (Information Extraction, IE)** – strukturlanmagan matndan o'zaro bog'liq ob'ektlarni va ob'ekt atributlarini avtomatik aniqlash. Ko'pgina hollarda, bu faoliyat NLP yordamida inson tilidagi matnlarni qayta ishlashni o'z ichiga oladi;

– **Tabiiy tilni qayta ishlash (Natural Language Processing, NLP)** – sun'iy intellekt (AI) sohasining tarkibiy qismi hisoblanadi. Kompyuterlar odatda insonlar muloqot qiladigan tabiiy tillarni tushunishlari uchun dasturchilar tomonidan *maxsus algoritmlar, metodlar va vositalar* ishlab chiqiladi. Bunday turdagi dasturiy ta'minotlar orqali mashina tabiiy tildagi yozma va og'zaki nutqni tanib oladi va qayta ishlaydi. Biroq tabiiy til o'ziga xos xususiyatlarga ega ekanligi sababli har bir tabiiy til uchun individual yondashuv talab etiladi. Shungdek, inson nutqida *jargon, ijtimoiy kontekst va mintaqaviy dialektlar* kabi istisnoli elementlar mavjudligi, ularni qayta ishlash murakkab amallarni talab qiladi.

– **Ma'lumotlarni intellektual tahlil qilish (Data Mining, DM)** – BigData matnidagi foydali, yashirin shablonlar va qonuniyatlarni aniqlashdan iborat. Data Mining orqali kelajakdagi tendensialarni bashorat qilishi mumkin. Bu esa tashkilotlarga mavjud ma'lumotlar asosida aniq qaror qabul qilish imkonini beradi. Data Mining vositalari an'anaviy ravishda juda ko'p vaqt talab qiladigan ko'plab biznes muammolarni hal qilish uchun ishlatilishi mumkin.

– **Ma'lumot qidirish (Information Retrieval, IR)** – Katta hajmdagi ma'lumotlar bazasida saqlanadigan ma'lumotni qidiradi va ma'lumotlardan foydali ma'lumotlarni olib ularni birlashtirib kengi bosqichga o'tadi. Bu esa foydalanuvchi tomonidan berilgan mos adekvat natijalarni taqdim etish vazifasiga javob beradi. Bugungi kunda ko'plab axborot tizimlarida IR komponentlari mavjud bo'lib, ma'lum bir muammoga tegishli hujjatlar sonini kamaytirish orqali tahlilni sezilarli darajada tezlashtirish imkonini beradi.

### **Text Mining yondashuvlari**

Bugungi kunda aktual bo'lgan Text Mining yondashuvlari quyida keltirilgan:

#### **1. Kalit so'zlar asosida assotsiatsiyalarni tahlil qilish**

Ushbu yondashuv asosida matndagi birga uchraydigan kalit so'zlar yoki atamalar to'plamini va ular orasidagi assotsiatsiya munosabatlarni aniqlanadi. Bunda matnli ma'lumotlardagi so'zlarning o'zagi, stemi,



lemmasi aniqlanadi va nomuhim soʻzlar olib tashlanadi. Matn ustida boshlangʻich amallar bajarilgach, oʻzaro bogʻlanishlarni aniqlash jarayoni ishga tushiriladi. Bu jarayon toʻliq avtomatik tarzda, inson ishtirokisiz amalga oshirilishi sababli, bajarilish vaqti kamayadi [Zanini, 2015].

## 2. Hujjatlarni tasniflash tahlili

Internetdagi veb-sahifalar, elektron pochta xabarlari va katta hajmdagi matnli hujjatlarni avtomatik ravishda tasniflash lozim boʻladi. Matnli hujjatlar tasnifi relyatsion maʼlumotlar tasnifidan farq qilib, hujjat maʼlumotlar bazalari atribut qiymatlari juftligiga koʻra tashkil etilmaydi.

### Matnni raqamlashtirish

Matnni raqamlashtirish bosqichida quyidagi amallar bajariladi:

– *Oʻzak (stem)larni aniqlash algoritmlari*. Hujjatlarni qayta ishlashning ilk qayta ishlash bosqichi – gapdagi **soʻzlar oʻzagini aniqlash (stemming)** dan boshlanadi. “Stemming” jarayonini soʻzlarni oʻz ildiz (oʻzak)lariga olib kelish deb taʼriflash mumkin. Bunda, soʻzlarning turli grammatik shakllari uchun umumiy oʻzak aniqlanadi. Stemmingning asosiy maqsadi matnni intellektual tahlil qilish dasturiga oʻxshash soʻzni taqdim etishdir.

– *Turli tillarni qoʻllab-quvvatlash*: Tabiiy til xususiyatidan kelib chiqqan holda *oʻzakni, sinonimlarni va harflarni aniqlash* kabi amallar bajariladi.

– *Muayyan belgilarni istisno qilish: Raqamlar, maʼlum belgilar yoki belgilar qatori va maʼlum miqdordagi harflardan qisqaroq yoki uzunroq* soʻzlarni istisno qilish berilgan hujjatlarini qayta ishlashdan oldin amalga oshirilishi lozim.

– *Roʻyxatlarni qoʻshish, roʻyxatlarni (nomuhim soʻzlarni) chiqarib tashlash*: Berilgan hujjatlardagi soʻzlarning chastotasiga qarab toifalarga ajratish mumkin. Shuningdek, homuhim soʻzlarni berilgan NLP masalasi shartiga koʻra chiqarib tashlanishi lozim. Koʻp hollarda til korpusi asosida nomuhim soʻzlar roʻyxati shakllantiriladi.

### Xulosa

Bugungi kunda matnli maʼlumotlar hajmining ortib borayotganligi sababli, maʼlumotlarni tahlil qilish va undan zarur/foydali maʼlumotlarni olish uchun samarali usullarni qoʻllash kerak. Elektron pochta xabarlari, hujjatlar, matnli xabarlar, ijtimoiy tarmoq postlari, bloglardagi maʼlumotlarni operativ tarzda qayta ishlash uchun NLP vositalari va usullaridan foydalanish maqsadga muvofiqdir. Text Mining

kompaniyalarga o`z faoliyatlarini yanada samaraliroq tashkil qilishga, o`z mijozlarini yaxshiroq tushunishga va operativ ma`lumotlarga asoslangan qarorlar qabul qilish uchun tushunchalardan foydalanishga yordam beradi. Ko`p vaqt talab qiluvchi va takrorlanuvchi vazifalarni Text Mining usullari orqali tez va yuqori aniqlikda bajarish mumkin. Katta ma`lumotlar to`plamini tahlil qilish va hissiyotlarni tahlil qilish, mavzuni belgilash yoki kalit so`zlarni aniqlash kabi turli usullardan foydalanish imkoniyati mijozlarning mahsulot haqida o`ylashlari va his-tuyg`ulari haqida zarur bilimlarni shakllantirishga xizmat qiladi. Text Mining texnologiyalarida nafaqat dasturlash ko`nikmalariga ega bo`lganlar, balki marketing, savdo, mijozlarga xizmat ko`rsatish va ishlab chiqarish sohalarida ishlayotganlar uchun barcha sohalardagi odamlar foydalanishi mumkin.

#### ADABIYOT

1. Aggarwal, C. C., & Zhai, C. X. (2013). An introduction to text mining. In *Mining Text Data* (Vol. 9781461432234). [https://doi.org/10.1007/978-1-4614-3223-4\\_1](https://doi.org/10.1007/978-1-4614-3223-4_1)
2. Kao, A., & Poteet, S. (2005). Text mining and natural language processing: introduction for the special issue. *SIGKDD Explor Newsl*, 7(1).
3. Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1). <https://doi.org/10.3390/bdcc4010001>
4. Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: Facebook and Twitter perspectives. *Advances in Science, Technology and Engineering Systems*, 2(1). <https://doi.org/10.25046/aj020115>
5. Zanini, N., & Dhawan, V. (2015). Text Mining: An introduction to theory and some applications. *Research Matters: A Cambridge Assessment Publication*, 19.

УДК 81'322.2

CREATION OF A LINGUISTIC DATABASE FOR ALISHER NAVOI  
AUTHORSHIP CORPUS*Abjalova M. A.<sup>1</sup>, Gulomova N. S.<sup>2</sup>**<sup>1</sup>Tashkent University of Information Technologies  
named after Al-Khorazmi, Tashkent, Uzbekistan**<sup>2</sup>Navoi Innovation University  
Navoi, Uzbekistan*

abjalova.manzura@gmail.com, gulomovamoi@mail.ru

The article is devoted to the current issues of the development of corpus linguistics. In order to present the scientific and creative heritage of Alisher Navoi, recognized as a great thinker, the sultan of words property to the general public in a modern and convenient way, the creation of Alisher Navoi's author corpus, the scientific and practical importance of the corpus, the vocabulary used in the author's work, information about the scope of expressions, the stages of describing the meanings of lexemes by explanation, the processes of creating a database of semantic tags of lexemes used by the thinker are presented.

**Keywords:** language corpus, author corpus, representation, identification, tagging, frequency, code.

СОЗДАНИЕ ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ  
ДЛЯ АВТОРСКОГО КОРПУСА АЛИШЕРА НАВОИ*М. А. Абжалова<sup>1</sup>, Н. С. Гуломова<sup>2</sup>**<sup>1</sup>Ташкентский университет информационных технологий  
имени Мухаммада аль-Хорезми, Ташкент, Узбекистан**<sup>2</sup>Навийский инновационный университет  
Навои, Узбекистан*

abjalova.manzura@gmail.com, gulomovamoi@mail.ru

**Аннотация.** Статья посвящена актуальным вопросам развития корпусной лингвистики. В целях современного и удобного представления широкой публике научного и творческого наследия Алишера Навои, признанного великим мыслителем, султаном слова, создается авторский корпус Алишера Навои, придается научное и практическое значение корпусе представлена лексика, используемая в творчестве автора, сведения об объеме выражений, этапах описания значений лексем путем объяснения, процессы создания базы данных семантических тегов лексем, используемых мыслителем.

**Ключевые слова:** языковой корпус, авторский корпус, репрезентация, идентификация, маркировка, частота, код.

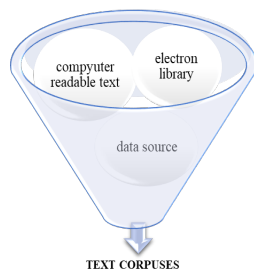
## INTRODUCTION

The need to create convenient methods of finding, storing, processing, editing and delivering information to the general public with the help of computer technologies is increasing in the field of linguistics, as in all areas of world civilization. This created a huge task for the field of computer technology. At the moment, textual information is directly related to the field of corpus linguistics. As a result of the growing interest and need in this field, many scientific studies have been conducted, therefore, a number of definitions and relationships have been given to the language corpus [Abjalova, 2022, p. 8]. For example, when Edward Finegan defines: “A corpus is a representative set of texts, you can familiarize yourself with the types of texts that are usually read by a machine, or you can search for a specific word or phrase” [Finegan, 2004, p.24], McEnery and A. Wilson say as follows defines: “Corpus linguistics is a branch of computational linguistics that develops general principles for the creation and use of linguistic corpus (text corpus).” [Zaharov, 2004, p. 11],

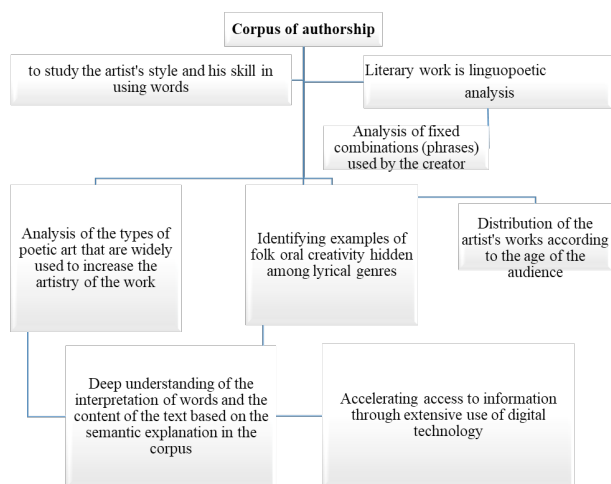
### Analysis

New Renaissance – In today’s historical period, when a solid foundation of the Third Renaissance is being created, to achieve wide use of our national heritage in the modern information and communication system, for this purpose, to promote the works of our ancestors among young people, authorship that presents examples of classic literature in a readable and understandable way creating corpora has become an important task of computational linguistics [Abjalova, 2023]. The scientific and applied research on the creation of authorship corpora has been proven by world experience that corpora are important not only for representatives of the field of linguistics, but also for the development of the nation and the development of the language. This type of corpora is not only a tool for speeding up the technical process, but it is also an innovative resource that can answer questions and include various forms of the language of a certain author in the information system. For this purpose, it is important to improve the corpus of Alisher Navoi’s authorship and for this purpose, to place all his ghazals in the collection of “Khazayin ul-maoni” into the corpus by semantic tagging. It is a laborious process to make modern large-scale housings look like a whole. [Zaharov, 2004, c. 34],

Among the different types of text corpora created in the field of computational linguistics, author corpora are distinguished by several advantages:



The corpus of works belonging to a certain writer will have the following convenient features:



The explanatory words found in 650 ghazals from the “Badoye’ ul-vasat” library of the Navoi “Khazayin ul-maoni” college were semantically tagged and placed in the corpus base using dozens of explanatory dictionaries created to clarify the language of the thinker’s works in a comprehensible way. [2, 3, 4, 5]. Summarizing Navoi’s words, sorting them, and commenting on them requires special knowledge, hard work and strong talent. Discovering Navoi’s vocabulary related to eternity, studying the linguistic features of his works is one

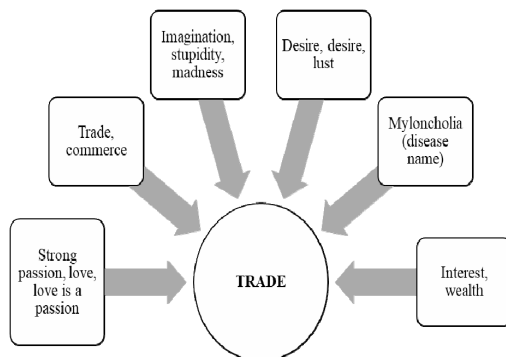
of the scientific researches that have interested hundreds of researchers and are in the center of attention. The famous scientist of the Renaissance, Justus Scaliger, said that this process is very laborious: "If someone is condemned to hard labor and suffering, the blacksmiths and miners do not let him see the hardships and give him order to make a dictionary. This work itself is the most difficult of all efforts". he says, referring to the responsible and complex scientific process of creating a dictionary.

Annotated dictionaries compiled for the study of the language of Alisher Navoi are considered an important source for understanding the leading idea of the thinker's works, and are an incomparable treasure that provides information on the cultural, social and political life of the 15th century. As a result of comparative study of annotated dictionaries, it can be said that the process of creating special annotated dictionaries to explain the language of Navoi's inimitable creativity and perfect works in a complete and integrated state should still be continued. The reason is that in the world of linguistics, a perfect explanatory dictionary that fully covers Alisher Navoi's vocabulary has not yet been created. As proof of our opinion, we can say that in the "Annotated Dictionary of the Language of Alisher Navoi's Works" published under the editorship of E. Fazilov, which is currently considered the most perfect in this field, *behrak*, *nafyi*, *manfi*, *do'loshib*, *rag'im*, *uqor*, *lavn* are used in ghazals. [Fozilov, 1983, p. 265]. Words like In the annotated dictionary "Language of Navoi's works" compiled by P. Shamsiyev and S. Ibrohimov: "*behrak*" – better (appears in 1 ghazal); "*nafyi*" – to reject, distance, drive away (occurs in 9 ghazals); "*do'loshib*" – to wrap around, to spin (occurs in 1 ghazal); "*rag'im*" – doing the opposite of something, doing something against it (occurs in 9 ghazals); "*uqor*" – crane (found in 1 ghazal); we find explanations of words such as "*lavn*" – color, paint (found in 8 ghazals). [Shamsiev P., 1983, p. 102]. Also, the explanation of some words can be found in other dictionaries.

### Research Methodology

In the process of entering the ghazals into the base of Alisher Navoi's authorship corpus with semantic tags, the semantic meanings of the explanatory words, which are difficult for the user to understand, are found from several existing explanatory dictionaries and typed one by one by hand on the computer. This painstaking process requires a lot of scientific and creative work, extreme care, strong attention and a

lot of time. The reason is that many of the words used by Navoi have the characteristic of homonymy, and some of them have dozens of subtle semantic differences. In particular, in the 2600 ghazals of the “Khazayin ul-maoni” collection, the lexeme “**trade**” appears in 73 places with the following different meanings:



The analysis of explanatory words in the verse is as follows: “*xurshidi raxshon*” – bright sun, “*zulf-hair*”, “*savdo*” – trade, “*chirmash*” – wrap. By day the bright sun brings tears to my eyes, and at night my hair is wrapped in dreams.

If we enter the explanation of a certain word in the ghazals into the computer’s memory, it will interpret this word in the same way in all verses. This leads to the fact that the verses are not analyzed correctly, that is, the original meaning of the ghazal is not covered. At the very beginning of the process of semantic analysis of ghazals, tables are created in MS Word, and ghazals and their stanzas and verses are entered based on the order in the book. How many explanatory words there are in each verse, are copied back to a separate table. For example, the initial process of semantically tagging the explanatory words in the 210 th ghazal of the book “*Garayib us-sigar*” and placing them in Alisher Navoi’s author corpus is carried out in the form of the following table:

| №   | Word form | Explanation           | Ghazal verse  |
|-----|-----------|-----------------------|---|
| 210 | aro       | aro – ichida          | Vasl aro ashkim qilur tug‘yon<br>o‘shul yuz tobidin |
| 210 | ashkim    | ashk – ko‘z<br>yoshim | Vasl aro ashkim qilur tug‘yon<br>o‘shul yuz tobidin |

| №   | Word form | Explanation                | Ghazal verse                                     |
|-----|-----------|----------------------------|--|
| 210 | tug‘yon   | tug‘yon – hayajonga kelish | Vasl aro ashkim qilur tug‘yon o‘shul yuz tobidin |
| 210 | o‘shul    | o‘shul – o‘sha             | Vasl aro ashkim qilur tug‘yon o‘shul yuz tobidin |
| 210 | tobidin   | tobi – issiqlik            | Vasl aro ashkim qilur tug‘yon o‘shul yuz tobidin |
| 210 | o‘lsa     | o‘lsa – yo‘q bo‘lsa        | Yoz faslida quyosh tez o‘lsa bo‘lur saylxez      |
| 210 | saylxez   | saylxez – kuchli sel       | Yoz faslida quyosh tez o‘lsa bo‘lur saylxez      |

In order to analyze ghazals, interpret “ancient” (archaic, historical) words that are difficult for a modern reader to understand, to understand the thoughts and ideas expressed in verses, the field of linguistics is used as a system - a whole object, i.e. dynamic and requires validation based on static linguistics. Dynamic linguistics studies the language in its real existence, “development”, and change, while static linguistics separates and describes a specific period (section) of language activity that is completely synchronized. This branch of linguistics studies a certain “old”, that is, outdated layer of the language, which is completely separated from the process of development and change in the language, related to the present – modern period. Static corpora contain texts created over a period of time. Author corpora – a collection of writers’ texts belong to this type of corpus. [Zaharov, 2020, p. 60].

After the process of semantic tagging of ghazals is completed, the annotated words are coded by ID. Code is a means (method) of recording information. [4]. There are 20.003 (twenty thousand three) verses and 40.006 (forty thousand six) verses in the 2.600 ghazals collected in the “Khazayin ul-maoni” collection. In the process of improving Alisher Navoi’s author’s corpus, 1850 ghazals, 15.002 verses, 30.004 verses were collected in “Garayib us-sigar”, “Navodir ush-shabab”, “Favoid ul-kibar” in addition to “Badoye’ ul-vasat” separate ID codes are obtained for and IDs are manually assigned to each annotated word found in the lines:

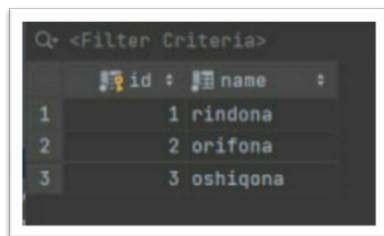
According to information theory, it is important to parallelize each word that is being semantically tagged with an ID code, as noted above. Because the code appears as an identification that allows recording in-



|     | A  | B           | C                                    | D      | E    |
|-----|----|-------------|--------------------------------------|--------|------|
| 585 | 90 | tasbih      | tasbih – tasbeh                      | g'azal | 2608 |
| 586 | 90 | rido        | rido – to'n                          | g'azal | 2608 |
| 587 | 90 | shayx       | shayx – tasavvuf yo'lining boshlig'i | g'azal | 2608 |
| 588 | 90 | dayr piri   | dayr piri – mayxonachi               | g'azal | 2909 |
| 589 | 90 | ilgidin     | ilg – qo'l                           | g'azal | 2609 |
| 590 | 90 | sog'ar      | sog'ar – qadah                       | g'azal | 2609 |
| 591 | 90 | fano        | fano – yo'q bo'lish                  | g'azal | 2610 |
| 592 | 90 | bahrig'a    | bahr – dengiz                        | g'azal | 2610 |
| 593 | 90 | cho'mmoqdin | cho'mmoq – sho'ng'imoq               | g'azal | 2610 |
| 594 | 90 | g'araz      | g'araz – yomon niyat                 | g'azal | 2610 |
| 595 | 90 | tolibqa     | tolib – xohlovchi                    | g'azal | 2610 |
| 596 | 90 | vasl        | vasl – uchrashuv                     | g'azal | 2610 |
| 597 | 90 | g'avvos     | g'avvos – suv tagiga tushuvchi       | g'azal | 2611 |
| 598 | 90 | za'fig'a    | za'f – zaiflik                       | g'azal | 2612 |
| 599 | 90 | bedilliq    | bedil – oshiq                        | g'azal | 2612 |
| 600 | 90 | bedillig'in | bedil – oshiq                        | g'azal | 2613 |
| 601 | 91 | hadis       | hadis – so'z                         | g'azal | 2614 |

formation, distinguishing or identifying one word from other words. In the table above, 6 words are identified with the ID number 2610 which means that there are so many annotated words in one sentence, and the semantic meaning of these coded words can be seen in the corpus interface. If a different ID code is assigned to a word in the verse, the original meaning of the word will not be visible. Therefore, every word in the ghazal must be marked with an ID number very carefully. This process is considered complicated for a computer, and a person can establish such associative relations relying on his intelligence.

In the analysis of ghazals, it is of particular importance to determine its content (romantic, orifona, rindona). When entering ghazal types into the corpus database, the information search language, which is considered a formal language, is used, and special ID numbers are blocked:



The words are identified in the database by means of a special numbering technique when including the young people of the target audience based on the nature of the text of the ghazals:

- 1 ID the age of the audience- 18;
- 2 ID the age of the audience- 17;
- 3 ID the age of the audience- 16;
- 4 ID the age of the audience- 15.

So, in the semantic tagging of explanatory verb forms found in ghazals, the columns of the database were formed in the following order: in the first column of the table, the sequence number of the ghazal, in the second row, the explanatory verb form of the interpreted words in the ghazals, and in the third row, the second row the bases of the word forms are given with an explanation, the genre of the poem is shown in the fourth line, and the stanza in which the word takes part is noted in the fifth line. A verse was written as many times as there are several explanatory units in one verse, because a certain word is used in many and different meanings in the divan, and their contextual meaning is introduced to the program using the verse, in the sixth line, the text type, the seventh line shows the age of the audience.

Various statistical data related to the poet's lyrics are also important in improving the corpus of Alisher Navoi's authorship. In the statistical analysis of the ghazal text, Zif's law of computer linguistics can be used appropriately. Zif's law is used to calculate how often each word is repeated (frequency) in large texts and their rate of repetition. [Plat, 1965, p.184.]. With the help of this law, it will be possible to determine the number of words in the text and the amount of their repetition. For example, American writer Mark Twain's "The Adventures of Tom Sawyer" was analyzed based on Zif's law. The work contains a total of 71.370 characters and 8.018 non-repeated word types. The average frequency of repetition of words used in the text is 8.9, that is, the words in the text are repeated approximately 9 times. [Rakhimov, 2011, p. 50]. All the words in the text of the novel do not have the same degree of use in the text. Some words are repeated 700 times, while some words occur only once. These words are called hapas legomena (Greek for "read only once"), and they make up about half of the work.

## Results

Using this Zif law, the number of words in the kulliyat divans was determined. The ghazals in "Garayib us-sigar" divan have a total of 4975 verses - 9950 verses (75 thousand 350. 66 954 words in ghazals), in the "Navodir ush-shabab" divan 4998 verses - 9996 verses (71 thousand 527 words, 66 596 word forms in ghazals), 5001 verses - 10002 verses (71 thousand 580 words, 66 539 word forms in ghazals) in "Ba-

doye' ul-vasat" divan, "Favoyid ul-kibar" divan contains 5029 verses – 10058 verses (75 911 words, 66 722 words in ghazals), and 20003 verses – 40006 verses in 2600 ghazals in "Khazayin ul-maoni" (a total of 294368 words in ghazals) 266811 word forms) are available. In Kulliyat ghazals, some words are used more than 1000 times, while some words are mentioned only once. According to Zif's law, the most frequently used word in the text is defined as  $r = 1$ , the less frequently used word is defined as  $r = 2$ , and the next word is defined as  $r = 3$ . Another important aspect of Zif's law is that with its help, the range of words found in ghazals is determined, and each creator's unique way of using words is known. From the 2600 ghazals in the "Khazayin ul-maoni" collection, the frequency of the words listed in the following table was determined based on the Zif law:

| Word   | Frequency | Determining the amount of words according to Zif's law |
|--------|-----------|--|
| Ko'z   | 2000      | $r = 1$  |
| Ishq   | 1988      | $r = 1$  |
| La'l   | 727       | $r = 1$  |
| Sarv   | 529       | $r = 1$  |
| Dahr   | 276       | $r = 2$  |
| Soqiy  | 262       | $r = 2$  |
| Jonon  | 187       | $r = 2$  |
| Ko'zgu | 169       | $r = 2$  |
| Miqroz | 4         | $r = 3$  |
| Ig'moz | 1         | $r = 3$  |

Any research carried out in society is important only if it serves to improve human life and its future. In the creation of Alisher Navoi's corpus, special attention was paid to increasing its educational value.

Verses using poetic arts such as talmeh, tanosub, tazad, irsoli masal from each divan in "Khazain ul-maoni" collection were analyzed separately and included in the corpus base. This increases the need to use corpora in the subjects of history, mother tongue, literature, and geography in secondary schools. In Navoi's ghazals, the names of historical figures, historical places: cities, villages, rivers, ponds, and lakes serve to increase their educational value. When the art of talmeh

is used, the poet refers to the legend, the heroes of historical events, and the persons whose names are shown, of course, must be familiar to a wide readership. [Hojiahmedov, 1998, p. 7]. Only then the real meaning of the text of the ghazal will be clear to the reader. It is known that Navoi skillfully used the poetic art of *talmeh* in ghazals in order to prove his opinion, increase the artistry of verses, and ensure the uniqueness of ghazals. Verses in which *Talmeh*'s poetic art took part are also significant in that they are studied in the teaching of other subjects in addition to literature classes of general education schools. Knowledge of historical, literary, and mythical figures is acquired in geography and economics training sessions - geographical and ethnic place names, and in lesson sessions organized on the basis of the curriculum of history. As a result, students develop the ability to learn subjects by connecting them with each other, and strengthen their knowledge of historical figures, names of historical and ethnic places. How clearly the poet imagined the earth can be seen from the fact that the basic terms and concepts of geography - road, sea, river, wind, sky, etc. are expressed in several alternative versions or the names of the country, region, city, village we can see that he skillfully used it in poetry. Geographical names indicate that the ghazals were created on the basis of artistic excellence and realism, and that Navoi imagined the world scientifically.

### **Conclusion**

In conclusion, it should be said that increasing the opportunity to use the centuries-old rich national heritage of the language, preserving it and making it a complete electronic corpus in order to fully deliver it to future generations is an urgent issue not only in Uzbek linguistics, but also in the whole world. It is an honorable task of our researchers to improve the authorship corpus of our great grandfather Alisher Navoi at the level of world standards, to translate it into many languages of the world, and to create parallel corpora is our duty and author corpora are considered the most convenient source of information for researchers and have become an important part of theoretical and practical fields in modern linguistics. Because author corpora are not only a means of speeding up the technical process, various forms of the language of a particular author are included in the information system, and unexpected questions about the author is an innovative system that can respond.

After all, author corpora are not only a means of speeding up the technical process, but also an innovative system in which various forms of the author's language are included in the information system and can answer unexpected questions about the author.

## REFERENCES:

1. Abjalova M. (2021). The importance of language corpus in the construction of lexicographic sources. *Current Research Journal Of Philological Sciences* (2767-3758), 2(12), 161–166. <https://doi.org/10.37547/philological-crjps-02-12-31>. <https://masterjournals.com/index.php/crjps>
2. Abjalova M. *Corpus Linguistics. [Text]: methodological manual / M.A. Abjalova. – Tashkent: Nodirabegim, 2022. – 110 p.*
3. Abjalova M., Gulomova N. Alisher Navoi and The Third Renaissance Period. // *Procedia of Theoretical and Applied Sciences. Vol. 4* (2023). 28.02.2023. – pp. 111–115. ALISHER NAVOI AND THE THIRD RENAISSANCE PERIOD | *Procedia of Theoretical and Applied Sciences*
4. Abjalova M., Gulomova N. Author's Corpus of Alisher Navoi and its Semantic Database. // *IEEE – UBМК – 2022: 7th International Conference on Computer Science and Engineering. 24–26 September 2022. – Diyarbakir, Turkey. – pp. 182–187. Impact Factor 5.5. DOI: 10.1109/UBMK55850.2022.9919546*
5. Abjalova M., Gulomova N., Sadullaeva Sh. *Author corpus of Alisher Navoi. Certificate No. DGU 18544. – Tashkent, 2022. (authorship certificate).*
6. *Alisher Navoi. Garayib us-sigar. – Tashkent. Civilization, 2011. – 570 p.*
7. Finegan E. *Language: its structure and use. New York: Harcourt Brace College Publishers, 2004. – P. 24.*
8. Hojjahmedov A. *Poetic arts and classical rhyme. – Tashkent. East, 1998. – 160 p.*
9. Plat U. *Matematicheskaja lingvistika // Novoe v lingvistike. – M. Progress, 1965. Vyp.IV. – 204 s.*
10. Rahimov A. *Fundamentals of computer linguistics. Andijan. – T.: Akademnashr, 2011. – 160 p.*
11. Shamsiev P., Ibrohimov S. *Dictionary of Alisher Navoi's works. – Tashkent: Gafur Ghulam, 1972. – 784 p.*
12. Zaharov V. P., Bogdanova S. Ju. *Korpusnaja lingvistika. – SPb.: Izdvo S. – Peterb. 2020. – 234 s.*

УДК: 004.934.5

## ФОРМИРОВАНИЕ РЕЧЕВОЙ БАЗЫ УЗБЕКСКОГО ЯЗЫКА

*С. Н. Ибрагимова<sup>1</sup>, М. И. Абдуллаева<sup>2</sup>*

*<sup>1</sup>Научно-исследовательский институт Развития цифровых технологий и искусственного интеллекта, Ташкент, Узбекистан,*

*<sup>2</sup>Ташкентский университет информационных технологий имени Мухаммада аль-Хоразми, Ташкент, Узбекистан  
snibragimova@mail.ru, malika.ilkhmovna@gmail.com*

Научная работа посвящена формированию речевой базы узбекского языка для дальнейшего создания TTS системы для узбекского языка. Узбекский язык является одним из самых распространенных и важных тюркских языков, используемых в Узбекистане и других сопредельных регионах. Несмотря на это, существует недостаток в речевых базах и ресурсах для разработки и исследования речевых технологий на узбекском языке. Синтез речи по тексту на узбекском языке – предполагает создание фонетико-акустической базы данных. Для формирования такой базы необходимо определить принципы создания и обработки текстового и речевого корпуса для узбекского языка и особенности формирования на их основе речевой базы. Результаты исследования позволили разработать эффективный алгоритм формирования речевой базы узбекского языка, который может быть использован для различных задач, связанных с распознаванием речи, синтезом речи, автоматическим переводом и другими речевыми технологиями на узбекском языке.

**Ключевые слова:** речевой сигнал, обработка сигнала, речевая база, нормализация текста, синтез, TTS система.

## FORMATION OF THE SPEECH BASE OF THE UZBEK LANGUAGE

*Ibragimova S. N.<sup>1</sup>, Abdullaeva M. I.<sup>2</sup>*

*Research Institute for the Development of Digital Technologies and Artificial Intelligence, Tashkent, Uzbekistan,  
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan  
snibragimova@mail.ru, malika.ilkhmovna@gmail.com*

The scientific work is devoted to the formation of the speech base of the Uzbek language for the further creation of a TTS system for the Uzbek language. The Uzbek language is one of the most widely spoken and important Turkic languages used in Uzbekistan and other adjacent regions. Despite this, there is a lack of speech databases and resources for the development and research of speech technologies in the Uzbek language. Synthesis of speech from the text in

the Uzbek language - involves the creation of a phonetic-acoustic database. To form such a database, it is necessary to determine the principles of creating and processing the text and speech corpus for the Uzbek language and the features of the formation of a speech base on their basis. The key task in the formation of the speech base is to achieve high quality and naturalness of synthesized speech. To do this, it is necessary to ensure high clarity and clarity of audio recordings, minimize noise and distortion. In addition, synthesis models should take into account the intonation, rhythm, accents and other prosodic features of the Uzbek language in order to create the most natural and understandable result. This work is devoted to the solution of these issues. The results of the study made it possible to develop an effective algorithm for the formation of the speech base of the Uzbek language, which can be used for various tasks related to speech recognition, speech synthesis, automatic translation and other speech technologies in the Uzbek language. This algorithm can be the basis for further research and development in the field of Uzbek speech processing and related technologies.

**Keywords:** speech signal, signal processing, speech base, text normalization, synthesis, TTS system.

Speech synthesis is a field of artificial intelligence that deals with the creation of artificial speech using computer systems. It allows you to generate human-like speech from textual data. Speech synthesis is widely used in various fields, including technologies to help people with speech disorders, automated voice assistants, audiobooks, announcements and broadcasts, educational applications and much more.

In recent years, speech synthesis has evolved significantly due to advances in deep learning and neural networks. Modern TTS systems have a high degree of naturalness and allow you to create speech with different intonations, accents and emotional expressions. They are able to reproduce not only individual words and phrases, but also convey complex melodic and rhythmic aspects of speech, making it more natural and understandable for the listener.

The main component of modern high-quality TTS systems is a speech base with a large volume. The formation of a speech base for the Uzbek language is a complex and multifaceted process that requires considerable effort and resources. It is necessary to collect an extensive set of audio recordings in the Uzbek language. This may include recordings of different genres of speech, accents, dialects, and intonations. However, finding and collecting enough quality recordings can be time-consuming and require collaboration with native speakers and local communities. The speech base should be quite diverse and cover different styles and topics.

The speech database of TTS systems is a database consisting of a set of audio data and corresponding text files [Dargis, 2018, pp. 26-29].

Audio files consist of samples of speech elements (sounds, syllables, words, sentences), and text files contain transcriptions corresponding to these speech elements<sup>3</sup> [Jindrich, 2001, p. 45].

To date, there are a number of speech databases for world languages in the public domain. Unfortunately, for the Uzbek language there is no such speech base in the public domain. Below are the highest quality and most popular of them [Panayotov, 2015, pp. 5206–5210; Muller, 2000, pp. 259–264; Radova, 2000, pp. 732–735; Cooper, 2001, p. 1134].

**Table 1. List and characteristics of open access speech databases**  
**Таблица 1. Список и характеристики речевых баз открытого доступа**

| Name of the speech base | Number of announcers | Tongue       | The volume of the speech base (hours) |
|-------------------------|----------------------|--------------|---------------------------------------|
| LJ Speech               | 1                    | English      | 24                                    |
| Libri-TTS               | Multi-announcer      | English      | 585                                   |
| RUSLAN                  | 1 (male)             | Russian      | 29                                    |
| NATASHA                 | 1 (female)           | Russian      | 13                                    |
| M-AILABS                | Multi-announcer      | Multilingual | 1000                                  |

In practice, the quality of synthetic speech depends on the quality of the speech base [Acero, 2012, p. 173; Prahallad, 2010, p. 128]. This is especially confirmed for speech synthesis based on the concatenative method and neural network architectures.

However, it is important to keep in mind that learning a neural network for speech synthesis usually requires a significant computational resource. Training a model can take a long time, especially when using deep neural network architectures. This may require the use of powerful computing systems or cloud platforms. After the model is trained, optimization and tuning of parameters are required to achieve the best quality of synthesized speech. This process may require a lot of experimentation and analysis of the results.

Common methods for the formation of a speech base for TTS systems are:

- recording of the announcer reading a pre-prepared text, material;
- recording of an announcer delivering spontaneous speech, narratives, etc.

Both methods are costly due to the need to involve additional specialists and speakers for the pre-processing of textual information and



post-processing of transcriptions and related audio data [6 Chalamandaris, 2014, p. 52. Nevertheless, the first method has the advantage in terms of the ability to adapt the TTS system being developed to a specific area, including it in the speech base terminology and suggestions from this area.

When creating a speech base, attention was paid to a number of features listed below, affecting its quality [Ochilov, 2022, p. 1–145]:

**1. The phonetic structure of textual information, which makes up the speech base and the variety of vocabulary.** The first step to the formation of a competent speech base is the collection and preparation of a variety of textual information. First of all, it is necessary to take into account the area for which the TTS system is being developed, thereby contributing to the predisposition of the system for the selected area. This is achieved by including sentences, terms and keywords of this direction in the text. A wealth of phonetic and prosodic coverage of the vocal corpus is also needed.

**2. Professionalism and literacy of the announcer.** The recording of the prepared text should be carried out by a native speaker with a good, clear pronunciation that meets the established language standards. The speaker's speech should be without unnecessary breaks, non-lexical vocabulary, false beginnings and fillers, such as «uh», «uh». For a competent recording of one hour of audio speech base, the announcer spends an average of two or more hours of time. When recording audio data based on the prepared text information, the announcer is also required to take it into account position relative to the microphone. The speech corpus for TTS systems should consist of audio data that differs little from each other in terms of intonation, voice volume, pronunciation speed, etc.

**3. The state of the audio recording environment.** It is very important to set the exact boundaries and tasks for the system, which will serve as a speech base. Text-to-speech systems require a studio environment, without extraneous noise, conversations, or music.

**4. The volume of the speech base.** The total duration of the audio data also varies depending on the task at hand. The minimum volume of the speech base for TTS systems that synthesize intelligible and understandable speech is 25 hours of audio data with their transcription. Audio recordings enter the speech database after strict multi-stage filtering by specialists.

In fig.1 shows the form of creating a competent speech base, consisting of four main stages, according to which the speech base for the Uzbek language was formed [Abdullaeva, 2023, pp. 1–12].

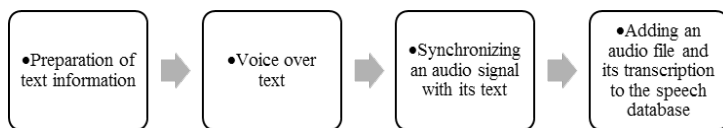


Fig. 1. Stages of creating a speech base

**Stage 1. Preparation of textual information.** The speaker is provided with pre-prepared normalized material in the form of text. Each text document goes through a series of processing stages to standardize it. The text may consist of words whose pronunciation is not usually found in dictionaries or lexicons, such as «BMT», «UzHDP», «TATU», etc. Such words are called non-standard words.

Non-standard words have several categories:

- numbers whose pronunciation changes depending on whether they refer to currency, time, phone numbers, postal codes;
- abbreviations, abbreviations, acronyms;
- Punctuation;
- dates, times, units, and URL links.

Many non-standard words are also homographs, i.e., words with the same written form but different pronunciations:

- IV, which may sound different: four (to 'rt), the fourth (to 'rtinchi);
- Three- or four-digit numbers, which can be dates and regular numbers (e.g., 2040-year, 2040-ton).

If the first component of textual information processing is normalization, then the second is prosodic analysis. At this stage, the text is analyzed in terms of stress and intonation and the prosodic markup of the normalized text is performed.

**Stage 2. Voice-over of the text.** The stage is the processing of recorded audio files for their standardization. The announcer records an audio file with expressive pronunciation. The audio file format is defined as .wav. Other main parameters of the audio file are set as 16-bit mono files, with a sampling rate of 44100 Hz. In some cases, when recording, these parameters may not be observed by the announcer, for this reason it is necessary to standardize all audio files. In addition to the above settings, each audio file must be checked in terms of the correctness of the spoken speech and the removal of speech zones in case the speech does not correspond to the text or is pronounced incorrectly. All audio data should be carefully filtered to avoid redundancy of audio information. It is for this reason that zones of silence are removed.

**Stage 3. Synchronization of the audio signal with its text.** This step is the most time-consuming and important for creating a speech base. This step is performed by an expert who carefully synchronizes audio and text files.

Synchronization of audio and its text consists of determining the pronunciation interval of segmented text in an audio file and marking it. This step requires a special approach and creates the need to create special algorithms and consists of 4 steps:

1. Voice-over of the text;
2. Calculation of informative coefficients of original and synthesized speech;
3. Calculation of the optimal coincidence of informative coefficients;
4. Temporary definition of the current text in the original audio file.

At the file verification step, there is an audio file and its text transcription. File verification is a mechanical process when an expert, listening to each audio file, checks the accuracy of the coincidence of the sounds spoken in it with the transcription of the current text.

**Stage 4. Adding an audio file and its transcription to the speech database.** Audio files checked for matching with their transcription are added to the speech database.

**Experiments and results.** To form the speech base of the Uzbek language, the software environment for recording Audacity with a wired, condenser microphone Hyper X QUADCAST S was used and was recorded by one female speaker. Hardware components of the audio recording system:

1. Monitor №1. Monitor LG 19M38A-B - designed to display text to the speaker
2. Monitor №2. Monitor LG 19M38A-B – the main monitor on which the audio data is recorded
3. SONY WH 1000 XM4 headphones for listening to recorded audio data
4. System Unit -Dell Optiplex 3080 Micro for recording, processing and storing audio data
5. Hyper X QUADCAST S microphone for voice recording
6. Logitech MX KEYS keyboard for entering and modifying information
7. Microsoft Sculpt Mouse to control.

The main devices of the recording system are a monitor with the optimal size for displaying text, a condenser studio microphone

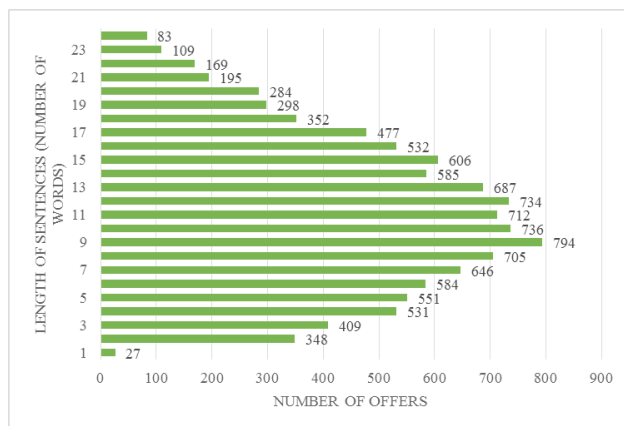
for high-quality speech recording, and a computer that supports the continuous operation of sound recording programs such as Audacity, Adobe Audition.

In total, the formed speech base of the Uzbek language for the speech synthesis system was  $\sim 30$  hours, which was formed by one female speaker. In total, more than 11 thousand sentences were used in the texts provided for reading. The sentences used  $\sim 151500$  words in the Uzbek language, 22721 of which are non-repeating words.

**Table 2. Parameters of the formed speech base**

| Category                                 | Teaching (Training) | Parameter settings (Validation) | General |
|--|---------------------|---------------------------------|---------|
| The volume of the speech base (in hours) | 27,5                | 3                               | 30,5    |
| Expression                               | 11107               | 584                             | 11692   |
| Words                                    | 143902              | 7574                            | 151476  |
| Non-repetitive words                     | 21585               | 1136                            | 22721   |

The statistical parameters of the generated many-hour speech database of the Uzbek language are shown in Fig. 2–3. According to statistics, it is known that sentences with a duration of 7–8 seconds and a length of 9 words are most common [Khuzhayarov, 2021, pp. 1–145; Musaev, 2022, p. 1].



**Fig. 2. Distribution of sentences found in the speech base by length**

The formed speech base of the Uzbek language has been introduced into the TTS system for the Uzbek language and is an integral part of the speech synthesis process. Synthesized Uzbek speech to assess the quality of the transmitted voice was rated on a MOS scale from 1 to 5 and received a score of 4.34, while the natural speech reproduced by a native speaker received a score 4.45.

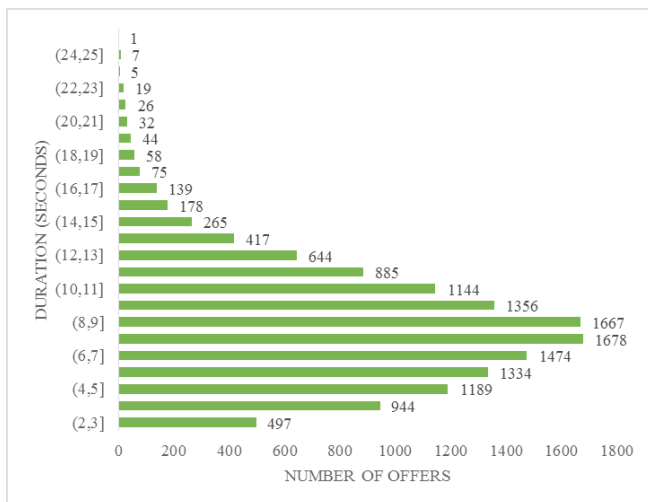


Fig. 3. Distribution of sentences found in the speech base by sound time

#### REFERENCES:

1. A. Acero, Acoustical and environmental robustness in automatic speech recognition. Springer Science & Business Media, 2012, vol. 201, 173 p.
2. A. Chalamandaris, P. Tsiakoulis, S. Karabetsos, and S. Raptis, Using audio books for training a text-to-speech system, Pro-ceedings of the 9th International Conference on Language Resources and Evaluation, 2014, 5p.
3. Dargis, R. and Auzina, I., Towards a Modern Text-to-Speech System for Latvian, In Human Language Technologies – The Baltic Perspective, Frontiers in Artificial Intelligence and Applications vol. 307, 2018, pp. 26–29.
4. Erica Cooper, Emily Li, Julia Hirschberg, Characteristics of Text-to-Speech and Other Corpora Columbia University, USA, 1p.
5. Jindrich Matousek, Josef Psutka, Jiri Kruta, Design of Speech Corpus for Text-to-Speech Synthesis, Eurospeech 2001 – Scandinavia, 2001, 4p.

6. Kishore Prahallad Automatic Building of Synthetic Voices from Audio Books CMU-LTI-10-XXX July 26, 2010, 128 p.

7. Khuzhayarov I.Sh. Integrallashgan neuron tarmoklar asosida uzbek tili nutkini tanish algorithmi va dasturiy vositalari. The technique of fanlar buyicha falsafa doctor (PhD) dissertation autorefrati. Toshkent – 2021. B. 145.

8. M.I.Abdullaeva, D.B.Juraev, M.M.Ochilov, M.F.Rakhimov. Uzbek Speech Synthesis Using Deep Learning Algorithms. IHCI 2022, LNCS 13741, pp. 1–12, 2023. [https://doi.org/10.1007/978-3-031-27199-1\\_5](https://doi.org/10.1007/978-3-031-27199-1_5).

9. Musaev M.M., Abdullayeva M.I., Ochilov M.M., Raximov M.F., Jurayev D.B. O'zbek tili nutqini sintezlovchi "Matn-nutq" dasturi, № DGU 17273. 01.07.2022.

10. Muller, L., Psutka, J., Smidl, L., Design of Speech Recognition Engine, Proceedings of TSD2000, Springer Verlag, Berlin, 2000, pp. 259–264.

11. Ochilov M.M. Uzbek tilidagi uzluksiz nutkni tanish texnologiyasi, algorithmi va dasturiy majmuasi. Tekhnika fanlar bo'icha falsafa doktori (phd) dissertatsiyasi avtorefrati. Toshkent – 2022. 145B.

12. Radova, V., UWB S01 Corpus – A Czech Read-Speech Corpus, Proceedings of ICSLP2000, vol. IV, Beijing, 2000, pp. 732–735.

13. V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, Librispeech: an ASR corpus based on public domain audio books, 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2015, pp. 5206–5210.

УДК 81'322.2

**МЕТОД СЕНТИМЕНТ АНАЛИЗА, ОСНОВАННЫЙ НА ЛЕКСИКЕ*****С. Ю. Алланазарова****Ташкентский государственный университет узбекского языка  
и литературы имени Алишера Навои**Ташкент, Узбекистан*

allanazarovasabohat@gmail.com

Анализ тональность – это метод НЛП, который помогает идентифицировать настроения в тексте. В данном анализе объектом исследования являются комментарии, оставленные пользователями, которые классифицируются по эмоциональной выразительности. Хотя в богатых ресурсами языках уже созданы лингвистические ресурсы для анализа, для узбекского языка таких ресурсов нет. Это тормозит развитие многих направлений. В связи с тем, что анализ настроений используется в процессе принятия решений. В данной статье анализируются исследования и современные подходы в этой области.

**Ключевые слова:** Анализ тональность, НЛП, лингвистическая база, отзывы, социальные сети, общественное мнение.

**LEXICON-BASED METHOD SENTIMENT ANALYSIS*****Sabohat Allanazarova****Alisher Navoi' Tashkent State University of the Uzbek Language  
and Literature**Tashkent, Uzbekistan*

allanazarovasabohat@gmail.com

Sentiment analysis is a natural language processing (NLP) technique that helps identify sentiments in text. In this analysis comments left by users are the object of research and they are classified in terms of emotional expressiveness. Although resource-rich languages already have built of linguistic resources for analysis, but there are no such resources for Uzbek language. This is a hindrance to the development of many areas. Owing to the fact that sentiment analysis is used in the decision-making process. This article analyzes the research and modern approaches in this field.

**Keywords:** Sentiment analysis, NLP, linguistic database, reviews, social network, public opinion.

## SENTIMENT TAHLILI UCHUN LEKSIKAGA ASOSLANGAN YONDASHUV

*Allanazarova Sabohat Yusupboyevna*

*Alisher Navoiy nomidagi Toshkent Davlat O'zbek Tili va Adabiyoti  
Universiteti*

*Toshkent, O'zbekiston*

allanazarovasabohat@gmail.com

Sentiment tahlili – matndagi his-tuyg'ularni aniqlashga yordam beradigan NLP sohasi hisoblanadi. Bu tahlilda foydalanuvchilar qoldirgan sharhlar tadqiqot obyektini bo'lib, ular emotsional-ekspressivlik jihatdan tasniflanadi. Resurslarga boy tillarda tahlil uchun allaqachon o'nlab lingvistik ta'minot manbalari mavjud bo'lsa-da, o'zbek tili uchun bunday resurslar mavjud emas. Bu esa ko'plab sohalarning rivoji uchun to'stinlik qiladi. Sababi sentiment tahlili qaror qabul qilish jarayonida qo'llaniladi. Ushbu maqolada mazkur sohada amalga oshirilgan tadqiqotlar va zamonaviy yondashuvlar tahlil qilinadi.

**Kalit so'zlar:** Sentiment tahlili, NLP, lingvistik ta'minot, sharhlar, ijtimoiy tarmoq, jamoatchilik fikri.

So'nggi yillarda matnni hissiy tahlil qilishga ko'proq e'tibor qaratilib, asta-sekin axborot olish, tabiiy tilni qayta ishlash (NLP) va boshqa sohalarning tadqiqot nuqtasiga aylandi [Pei, 2022]. So'nggi yillarda his-tuyg'ularni tahlil qilish bo'yicha katta ishlar amalga oshirildi [Rosa, 2015; Riedl, 2012]. ST bo'yicha dunyo miqyosidagi olimlar turli tadqiqotlar bilan shug'ullanib kelganlar va hozirda ham shug'ullanmoqdalar. Ingliz tilidagi matnlarni fikrlar asosida tahlil qilish juda mashhur va yaxshi o'rganilgan mavzu hisoblanadi. Ingliz tili uchun ko'plab tasniflash manbalari mavjud. Masalan: SentiWordNet [Riedl, 2012], SenticNet [Hoffman, 2010] va NRC Emotion Lecikon va boshqalar [Hofmann, 1999]. Hissiy jihatdan qutblanish manbalari yaratilmagan tillar uchun eng oddiy yechim ingliz tilidagi manbalarni tarjima qilish bo'lishi mumkin. Ammo agglutinativ xarakteriga ega o'zbek tili uchun bu to'g'ri yechim emas.

O'zbek tili uchun Sentiment tahlilida qo'llaniladigan lingvistik ta'minot ishlab chiqish bo'yicha yetarlicha ma'lumotlar mavjud emas va to'laqonli o'rganilmagan. Shunga qaramasdan o'zbekcha matnlar ustida bir nechta tadqiqotlar o'tkazilgan. Xususan, S.Matlatipov va boshqalar tomonidan o'zbek tilidagi matnlarning sentiment tahlili uchun annotatsiyali korpus qilingan [Matlatipov, 2019]. Shuningdek, tabiiy tillarni qayta ishlash jarayonida so'z shakllarini morfologik tahlil qilish model va algoritmlari asosida dasturiy model va vositalari ishlab chiqilgan. Korpus va morfologik dasturiy moduliga asoslanib, o'zbek tilida



bildirilgan taklif va fikrlarni klassik mashinaviy o'qitish va zamonaviy neyron to'rlar algoritmlaridan foydalangan holda sentiment tahlil qilish modeli qurilgan. Ushbu qurilgan ikkala modelning natijasi mos ravishda 88.89% va 89.56% foiz aniqlik ko'rsatgan [Kuriyozov, 2019]. I. Rabbimov boshchiligidagi tadqiqotchilar guruhi filmlariga qoldirilgan sharhlarni emojilar asosida o'rganishdi [8, 436]. Sababi ko'pchilik fikr bildirishdan ko'ra emoji ni afzal bilishadi. Tadqiqot ishi uchun YouTube videoxostingdagi o'zbek filmlariga qoldirilgan sharhlar obyekt qilib olingan. Tadqiqot ishi 85.25% ga teng aniqlikda natija ko'rsatib, boshqa tillarda o'tkazilgan avvalgi tadqiqotlar bilan mos kelgan.

Turkiy tillar oilasiga mansub bo'lgan qo'shni Turk va Qozoq tillari bu sohada ancha taraqqiy etdi. Masalan: Turk tilidagi kino sharhlariga o'tkazilgan hissiy tahlil qilish tizimi 79.06% aniqlik ko'rsatdi [Lee, 2011]. Qozoq tilidagi terrorizm tahdidlari bo'lgan matnlar ustida ishlab, morfologik, sintaktik va hissiy jihatdan tahlil qilish usullari ishlab chiqilgan [Blei, 2003]. Ushbu ishda morfologik qoidalar va ontologik modelga asoslanib, qozoq tilidagi matnlarni hissiy tahlil qilish uchun emotsional so'zlar lug'ati yordamida qoidalarga asoslangan usul 83% aniqlik ko'rsatgan.

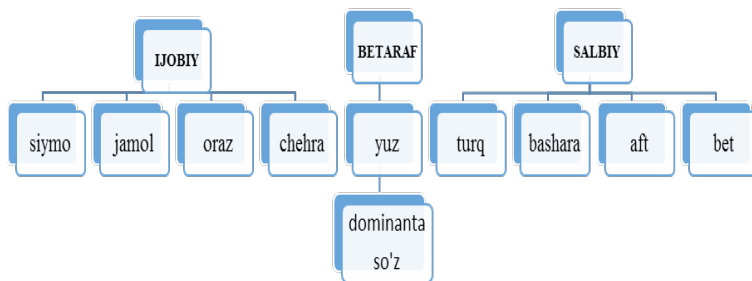
Xu va Liu [Ding, 2008] o'z tadqiqotlarini mijozlar sharhlarini tasniflashga qaratdilar, ya'ni ular hissiyotlarni o'z ichiga olgan mahsulot xususiyatlarini ajratib oldilar, so'ngra ushbu xususiyatlar asosida fikrlarni tasnifladilar va natijada mahsulot sharhlarining qisqacha mazmuni yaratildi. Misol uchun, agar kamera haqidagi sharhlar ko'riloyatgan bo'lsa, mualliflar tasvir sifati va kamera o'lchami kabi xususiyatlarni ajratib olishadi va ushbu xususiyatlar asosida ijobiy va salbiy kamera sharhlarini tasniflaydilar. Jumlagi ijobiy yoki salbiy belgi qo'yish uchun tadqiqotchilar birinchi navbatda har bir sharhdan qutbli so'z (sifat)larni ajratib olishgan. Tasniflash sifatning sinonimlari bilan bir xil qutbliligi va antonimlarining qarama-qarshi qutbliligiga asoslangan. Emotsional-ekspressiv bo'yoqdor so'zlar WordNetda ma'lum yo'nalishga ega bo'lgan sinonim va antonimlarni qidirish uchun ishlatilgan. Shu sababli, sharhda topilgan qutbli so'zlarning yo'nalishi aniqlandi va o'rtacha aniqlik 84% ni tashkil etdi. Shuning uchun bu usul sifatlarning semantik yo'nalishini va gaplarning qutbliligini tahlil qilishda samarali bo'lishi mumkin.

Kim va Hovy [Kim, 2004] berilgan mavzudagi matn va uning ma'ruzachisining kayfiyatini o'rganib chiqdilar. Tadqiqot mualliflari bir nechta tasniflagichlardan foydalanganlar. Birinchi klassifikator jumladagi har bir so'zning qutbligini olish uchun qo'llanildi. Ikkinchi tasniflagich fikr tashuvchisi tomonidan ifodalangan butun jumlaning qut-

biligini aniqladi. Bundan tashqari, mualliflar boshlang'ich so'zlarning kichik boshlang'ich ro'yxatidan foydalanishni sifat va fe'llar bilan bir xil tarzda kiritdilar. Bu oxirigisi WordNetda mos keladigan sinonim va antonimlarni qidirish orqali kengaytirildi. Mualliflarning ta'kidlashicha, ba'zi sinonimlar/antonimlar betaraf yoki qarama-qarshi yo'nalishga ega, bu ularni ishlatish uchun nomaqbul hisoblanadi. Bundan tashqari, tadqiqotchilar so'zlarning ijobiylik va inkorlik kuchini aniqlash zarurligini ta'kidladilar, bu esa polisemantik so'zlarni yo'q qiladi. Kim va Hovy jumlada fikr egasiga yaqin bo'lgan va his-tuyg'ularni o'z ichiga olishi mumkin bo'lgan to'rt xil sohani aniqladilar. Taklifning yo'nalishini aniqlash uchun mualliflar uchta modelni ishlab chiqdilar. Birinchi model "negativlar bir-birini bekor qiladi" degan taxminga asoslangan edi. Ikkinchi va uchinchi modellar mos ravishda ma'lum bir mintaqada hissiyot kuchining harmonik va geometrik o'rtacha qiymatini ifodalaydi. Tajribalar o'tkazilgandan so'ng eng yaxshi natijalar birinchi model va fikr egasidan boshlab jumlaning oxirigacha bo'lgan maydon yordamida olinadi, degan xulosaga kelingan.

Garchi his-tuyg'ularni tahlil qilish juda faol tadqiqot bo'lgan bugungi kunda ushbu sohaning bir qator murakkab muammolari hali ham qolmoqda. Birinchidan, sarkazm muammosi; Sarkazm - nutq birliklarining murakkab shakli bo'lib, unda sharh muallifi yoki yozuvchi o'zi nazarda tutgan narsaning aksini aytadi yoki yozadi. Tuyg'ularni tahlil qilishda istehzoli jummalarni tahlil qilish juda qiyin, chunki fikr aniq va to'g'ridan-to'g'ri ifodalangan.

Tilning leksikasi eng nozik ma'noga va stilistik bo'yoqdorlik bilan axborot berishda juda katta imkoniyatlarga ega. Uslubiyatda tilning lug'at boyligidagi birliklarning stilistik xususiyatlari o'rganiladi. Ular ichida, ayniqsa, sinonimlar, antonimlar stilistik imkoniyatga boy vositalar hisoblanadi. Stilistik bo'yoqli so'zlar, ayniqsa, sinonimik qatorda o'z ifodasini topadi.



**1-rasm.** O'zbek tilidagi ko'p qatorli sinonimik sinset na'munasi.

Yuz, aft, bashara, turq, bet, chehra, siymo, oraz, diydor. Bu sinonimik guruhdan **yuz** - bosh, dominant soʻz boʻlib, betaraf bahoni ifodalaydi va nutq uslublarini tanlamaydi. **Aft, turq, bashara, bet** soʻzlari esa salbiy boʻyoqdor soʻzlardir. **Chehra, diydor, oraz, siymo** kabi soʻzlar esa ijobiy boʻyoqdorlikka ega. Bu sinonimik qatordagi har bir soʻz turli nutq vaziyatlarida tanlab ishlatiladi.

**Xulosa.** Ushbu maqolada Sentiment tahlili boʻyicha yondashuvlar va shu kungacha amalga oshirilgan tadqiqot ishlari koʻrib chiqildi. Ingliz, fransuz, ispan, hind, xitoy, rus, turk kabi tillardagi boy resurslarning mavjudligi oʻzbek tili uchun lingvistik taʼminot ishlab chiqish va turli tillarda bajarilgan ishlarni oʻrganish boʻyicha qilinadigan tadqiqotlarga sezilarli darajada ragʻbatlantiradi.

#### ADABIYOT

1. Pei, Y.; Chen, S.; Ke, Z.; Silamu, W.; Guo, Q. AB-LaBSE: Uyghur Sentiment Analysis via the Pre-Training Model with BiLSTM. Appl. Sci. 2022, 12, 1182.

2. La Rosa M., Fiannaca A., Rizzo R., Urso A. Probabilistic topic modeling for the analysis and classification of genomic sequences // BMC Bioinformatics. – 2015. – Vol. 16, no. Suppl 6. – P. S2.

3. Riedl M., Biemann C. TopicTiling: A text segmentation algorithm based on LDA // Proceedings of ACL 2012 Student Research Workshop. – ACL '12. – Stroudsburg, PA, USA: Association for Computational Linguistics, 2012. – Pp. 37–42.

4. Hoffman M. D., Blei D. M., Bach F. R. Online learning for latent Dirichlet allocation // NIPS. – Curran Associates, Inc., 2010. – Pp. 856–864.

5. Hofmann T. Probabilistic latent semantic indexing // Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. – New York, NY, USA: ACM, 1999. – Pp. 5

6. Matlatipov S., Kuryozov E., Miguel A. A., Corlos-Rodriguez. Deep learning vs. classic models on a new uzbek sentiment analysis dataset. Conference: 9th language & technology conference: Human language technologies as a challenge for computer science and linguistics Poznan, – Poland–2019. – P. 258–262

7. Kuriyozov E., Matlatipov S. Building a new Sentiment Analysis Dataset for Uzbek Language and Creating Baseline Models.// Multidisciplinary Digital Publishing Institute Proceedings. 2019.– №1. Pages 37.

8. Rabbimov I., Mporas I., Kobilov S. Investigating the effect of emoji in opinion classification of uzbek movie review comments. International Conference on Speech and Computer Science. – 2020. – P. 435-445.

9. Lee S. S., Chung T., McLeod D. Dynamic item recommendation by topic modeling for social networks // Information Technology: New Generations (ITNG), 2011 Eighth International Conference on. – IEEE, 2011. – Pp. 884–889.

10. Blei D. M., Jordan M. I. Modeling annotated data // Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. – New York, NY, USA: ACM, 2003. – Pp. 127–134. [26] Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet allocation // Journal of Machine Learning Research. – 2003. – Vol. 3. – Pp. 993–1022.

11. Ding X., Liu, B., & Yu, P. S. (2008, February). A holistic lexicon-based approach to opinion mining. In Proceedings of the 2008 international conference on web search and data mining (pp. 231–240). ACM.

12. Kim, S. M., & Hovy, E. (2004, August). Determining the sentiment of opinions. In Proceedings of the 20th international conference on Computational Linguistics (p. 1367). Association for Computational Linguistics.

УДК 81'322.2

**NER: МЕТОДЫ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ  
ТОПОНИМОВ В ТЕКСТАХ НА УЗБЕКСКОМ ЯЗЫКЕ*****Б. Б. Элов, М. Т. Саматбоева****Ташкентский государственный университет узбекского языка  
и литературы им. Алишера Навои**Ташкент, Узбекистан*

e-elov@navoiy-uni.uz, samatboyevamadina@navoiy-uni.uz

В данной статье рассмотрены методы автоматического распознавания топонимов с помощью NER объектов (Named Entity Recognition) технологий. Топонимы не только обозначают название места, но и отражают историю, прошлое и языковые особенности этого места. По этой причине при изучении и моделировании топонимов в первую очередь использовались методы идентификации названных объектов. В статье исследованы NER методы идентификации объектов на основе правил и словаря, были представлены сведения о проводимых научных исследованиях по топонимам. Также представлены основные характеристики топонимов, такие как характеристика имени собственного и его отличие от аналогов.

**Ключевые слова:** NER, имя, именованный объект, относительное существительное, наименование, имя собственное, топоним, топонимия, ономастика, лингвистика.

**NER: METHODS FOR AUTOMATIC DETECTION OF  
TOPONYMS IN TEXTS IN THE UZBEK LANGUAGE*****Botir Elov, Madina Samatboyeva****Alisher Navoi' Tashkent State University of the Uzbek Language  
and Literature, Tashkent, Uzbekistan*

e-elov@navoiy-uni.uz, samatboyevamadina@navoiy-uni.uz

This article discussed methods of automatic identification of toponyms (place names) from NER (Named Entity Recognition) objects. Toponyms do not only indicate the name of a place, but also reflect the history, past, and language characteristics of this place. For this reason, in the study and modeling of toponyms, first of all, the methods of identifying the named objects were used. In the article, rule-based and dictionary-based methods were studied for identifying NER objects. In this research work, information about the scientific research work carried out on toponyms was presented. Also, one of the main characteristics of toponyms - the characteristic of a proper name and its difference from a similar name, important terms within the toponym were presented.

**Key words:** NER, name, named object, relative noun, appellative, proper noun, place name, toponymy, onomastics, linguistics.

## NER: O‘ZBEK TILIDAGI MATNLARDA TOPONIM(LAR)NI AVTOMATIK ANIQLASH METODLARI

**Elov B. B., Samatboyeva M. T.**

*Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va  
adabiyoti universiteti, Toshkent, O‘zbekiston*

e-elov@navoiy-uni.uz, samatboyevamadina@navoiy-uni.uz

Ushbu maqolada NER(Named Entity Recognition) obyektlaridan – toponimlar (joy nomlari)ni avtomatik aniqlash metodlari haqida fikr yuritildi. Toponimlar – faqat joy nomini bildiribgina qolmay, ushbu joy tarixi, o‘tmishi, til xususiyatlarini ham o‘zida aks etadi. Shu sababli toponimlarni o‘rganishda va modellashtirishda, avvalo, nomlangan obyektlarni aniqlovchi metodlardan foydalanildi. Maqolada NER obyektlarini aniqlashda *rule based*(qoidalarga asoslangan yondashuv) va *dictionary based*(lug‘atlarga asoslangan yondashuv) asosida tadqiq etildi. Ushbu tadqiqot ishida, toponimlar ustida amalga oshirgan ilmiy tadqiqot ishlari haqida ma’lumotlar keltirildi. Shuningdek, toponimlarga xos asosiy xususiyatlardan biri – atoqli otlilik xususiyati va uning turdosh otdan farqi, toponim doirasidagi muhim atamalar taqdim etildi.

**Kalit so‘zlar:** NER, nom, nomlangan obyekt, turdosh ot, appelyativ, atoqli ot, joy nomi, toponomika, toponimiya, toponim, onomastika, lingvistika.

## TOPONIMLARNING O‘RGANILISH TARIXI

*Toponimlarning IX–XVI asrlarda o‘rganilishi*

Toponimlar bo‘yicha ilmiy va amaliy tadqiqotlar juda qadim zamonlardan boshlangan. Joy nomlarining paydo bo‘lishi, ularning ma‘nosi, grammatik tuzilishi to‘g‘risidagi ma’lumotlar buyuk olimlarimiz asarlarida aks etgan. Xususan, Yusuf Xos Hojibning *“Qutadg‘u bilig”* [Yusuf Xos Hojib, 1991, Abu Rayhon Beruniyning *“Qonuni Ma’sudiy”* [Абу Райхон Беруний, 1973], *“Hindiston”* [Абу Райхон Беруний, 1968] *“Saydana”* [Ubaydulla Karimov, 1974] Mahmud Koshg‘ariyning *“Devonu lug‘atit turk”*[Mahmud Qoshg‘ariy, 1963], Ibn Sinoning *“Dengiz qirg‘oqlari”* [Ibn Sino, 1996] asarlarida toponomiyaga oid ma’lumotlarni uchratishimiz mumkin.

Shuningdek, Mirzo Ulug‘bek, Abulg‘ozi Bahodirxon va Bobur ijodida ham joy nomlarining o‘rganilganligini kuzatish mumkin. Jumladan, professor H.Hasanovning ta’kidlashicha, birgina *“Boburnoma”-da* 1000 ga yaqin geografik nom tilga olingan [Zahiriddin Muhmmad Bobur, 2008]. Bobur o‘z asari *“Boburnoma”*da toponimlarning nafaqat etimologik tahliliga, balki grammatik tuzulishi ham ta’rif berib o‘tadi. Ushbu kitobda ikki mintaqa *Turkiston o‘lkasi* va *Hindiston diyori* toponimiyasi solishtirilib tadqiq etilgan. *“Boburnoma”*da ikki mam-

lakat tabiati, joylashish o'rnini va areali hisobga olingan bo'lib, asarda "Qarshi" toponimi etimologik tahlil qilinib, ushbu so'z mo'g'ul tilida "o'ttur(go'rxona)", turk tilida esa "saroy" deb nomlanishi keltirib o'tilgan. XV–XVI asrlarda esa bu so'z «mo'tabar shaxslar qabriga qurilgan dahma maqbara» ma'nosini bildirgan. Bobur toponimlarni etimologik tahlil qilish bilan bir vaqtda, ularni **affikslar** asosida ham tadqiq qilgan. Masalan, birgina **-an (yon), -ob, -tu, -ot**(masalan, *Ohangaron, Childuhtaron, Bog'izag'on, So'zangaron, Bishxoron*) qo'shimchalari juda xarakterli va ko'p qo'llaniladigan qo'shimchalar ekanligini ta'kidlangan.

#### *Toponimlarning XX asr boshlarida o'rganilishi*

Keyingi davrlarda toponomiyani o'rganish muhim sohalardan biriga aylanib bordi. *Xiva, Samarqand, Buxoro, Qo'qon, Toshkent* hujjatlarining toponimik va terminologik xazinalarini o'rganishda G'.G'ulomov, M. Yo'ldoshev, V.V. Bartold, P.P. Ivanov, V.L. Vyatkin, A.A. Semyonov, O.D. Chexovich, O.A.Suxareva, A.B. Ahmedov, A.R. Muxamadjonov, va R.G. Muqminova kabi olimlar ilmiy izlanishlarni amalga oshirgan.

Jumladan, E.M.Murzayev *O'rta Osiyo, Sinjon, Mongoliyadagi geografik nomlar va xalq terminlarini o'rganish jarayonida o'zining "Словарь народных географических терминов"* [Murzaev Э.М., 1984] asarida juda ko'p geografik terminlar izohini ham keltirib o'tgan.

#### *O'zbekiston toponomiyasining o'rganilishi*

Keyingi yillarda O'zbekistonda toponimiyasini o'rganishda bir qancha nazariy va amaliy ishlar amalga oshirildi. Jumladan, H. Hasanov, E. Begmatov, T. Nafasov, S. Qorayev, Z.Do'simov, B. Bafoyev, N. Husanov, T. Enazarov, N.Uluqov, M.N. Ramazonova, H.Uzoqov, L. Karimova, J. Latipov, X. Xolmo'minov, N. Oxunov, S. Naimov, Yo. Xo'jamberdiyev, Sh.M. Qodirova, Sh. Yoqubov, P. G'ulomov, M. Mirakmalov, A. Otajonova, A. Aslonov, M. Tillayeva, M. Almamatov singari tilshunoslik, geografiya, tarix, geologiya kabi fanlarining olimlari bu sohada ilmiy tadqiqotlar olib borishmoqda.

*Toponimlarni o'rgangan geograf olimlardan biri H.H.Hasanov "O'rta Osiyo joy nomlari tarixidan" (1965)*[Hasanov H, 1965], "Geografik nomlar imlosi" (1962) [Hasanov H, 1962], "Yer tili" (1977) [Hasanov H, 1977], "Geografik nomlar siri" (1985) [Hasanov H, 1985], "Geografiya terminlari lug'ati" (1964) [Hasanov H, 1964] asarlari, "O'rta Osiyolik geograf sayyohlar"(1964) [Hasanov H, 1964] maqolalari geografik toponomiya haqida qimmatli manbalardan hisoblanadi.

Geograf toponomistlardan biri hisoblangan Suyun Qorayev *toponimlarning etimologiyasini* va ular tarkibidan *etnotoponimlarni* chuqur o'rgangan. S. Qorayev "*Geografik nomlar ma'nosi*" (1978) [Qorayev S, 1978], "*Geografik nomlar ma'nosini bilasizmi?*" (1970) [Qorayev S, 1970], "*Toshkent toponimlari*" (1991) [Qorayev S, 1991] kitoblarida o'zining ilmiy xulosalarini bayon etgan. Birgina "*Toponomika*" [Qorayev S, 2006] o'quv-qo'llanmasida toponomiyaga oid ko'pgina ilmiy qarashlar, toponomikaning rivojlanishi, tarixiy aspektlar, topoformant va modellar, toponimlarning turlari bo'yicha tasnif, hududiy toponomiya haqida, amaliy toponomiya tog'risida ma'lumotlar hamda qo'llanmaning eng qimmatli qismi – oltita tarixiy asar: "*Hudud ul-olam*" [Hudud ul-olam, 2008] (982—983-yillarda fors-tojik tilida yozilgan, muallifi noma'lum), "*Buxoro tarixi*" [Abu Bakr Muhammad ibn Ja'far Narshaxiy, 1966] (Abu Bakr Muhammad ibn Ja'far-an-Narshaxiy), "*Devonu lug'atit turk*" [Mahmud Qoshg'ariy, 1963] (Mahmud Koshg'ariy), "*Boburnoma*" [Zahiriddin Muhammad Bobur, 2008] (Bobur), "*Toponimika ocherklari*" [Murzayev E.M., 1974] (E.M. Murzayev) asarlari tahlil qilingan. Shuningdek, S. Qorayev nafaqat geografiya, balki tilshunoslik nuqtayi nazaridan toponimlarni o'rgangan olimdir. Ushbu olim nomzodlik dissertatsiyasini filologiya fanlari yo'nalishida, doktorlik ishini esa geografiya fanlari sohasida amalga oshirgan.

XX asrning 90-yillarining ikkinchi yarmidan hozirgi vaqtgacha toponimlarni geografiya nuqtai nazaridan tahlil qilgan yana bir olimlardan biri Mirali Mirakmalovdir. Olimning (P.G'ulomov bilan hamkorlikda) yaratilgan "*Toponimika va geografik terminshunoslik*" universitetlar va pedagogika institutlarining geografiya, biologiya-geografiya, tarix-geografiya yo'nalishlari talabalari uchun o'quv qo'llanma sifatida yaratilgan asari *toponomika, geografik nomlarning paydo bo'lishi va geografik terminshunoslik tog'risida* qiziqarli ma'lumotlarni taqdim etadi. Shuningdek ushbu o'quv-qo'llanmada *O'zbek geografik terminshunosligining rivojlanishida tarjima(ruscha – o'zbekcha) adabiyotining ahamiyati* to'g'risida ham fikr yuritiladi.

Toponimlarni o'rgangan tarixchi olimlardan biri Zohid Madrahimov tarixiy toponomikani o'rganib, ilmiy asarlari hamda maqolalarida toponimlarning tarixiy aspektda o'rganilishi, nomning paydo bo'lishi va tarixiy nomlarning yuzaga kelishi haqida o'z fikrlarini bildirib o'tgan. Birgina "*Tarixiy toponomika*" [Madrahimov Z, 2017] asarida *O'rta Osiyo tarixiy toponomiyasining manbalari, yozma manbalardagi qadimiy toponimlar, Tarixiy toponimiyaning substrati – o'zagi hamda*



*nom yasovchi leksema – komponentlari, O‘zbekiston hududidagi etnonim, oronim, gidronimlar, Namangan viloyati toponomiyasi* haqida ma’lumotlar keltirib o‘tgan.

*Toponimlarni o‘rganan tilshunos(filolog) tadqiqotchilar* H. Egamov, Z.Do‘simov va T. Nafasovlar ishlari muhim va qimmatli manbalardandir. Xususan, Zarip Do‘simov va Hundaybergen Egamov hammuallifligida 1977-yilda “*Joy nomlarining qisqacha izohli lug‘ati*” [Do‘simov Z., Egamov X., 1977] nashr etilgan. Ushbu lug‘at alifbo tartibida bo‘lib, toponimlarni harf bo‘yicha qidirish qiyinchilik tug‘dirmaydi.

Ernest Begmatovning tadqiqotlari asosan o‘zbek antroponimlariga bag‘ishlangan holda, qisman antroponimika hamda toponimikaning nazariy va amaliy masalalariga ham munosabat bildirilgan.

A.Muhammadjonov esa XX asrning 50-yillari oxiridan mamlakatimiz toponimlarining amaliy masalalariga e’tibor qaratgan bo‘lsa, XX asrning 90 yillaridan esa toponimlarning etimologik tahliliga ham jiddiy e’tibor berib, bir qancha bahsli va qiziqarli maqolalarni e’lon qilib kelmoqda.

Mamlakatimizda toponimlarni ham nazariy, ham amaliy jihatlardan jiddiy o‘rganish XX asrning 60-yillaridan boshlangan. T.Nafasov XX asrning 60- yillaridan to bugungi vaqtgacha toponimlarning izohli tahlili bo‘yicha ilmiy ishlar olib borib, Qashqadaryo va Surxondaryo viloyatlaridagi bir qancha joy nomlarining izohining tahliliga kirishib, “*O‘zbekiston toponimlarining izohli lug‘ati*” (1988) [Nafasov T, 1988], “*O‘zbek nomnomasi*” [Haфасов Т, 1993], “*O‘zbek qishloqnomasi*” [Haфасов Т, 1994] singari bir qancha salmoqli ishlarni amalga oshirgan. T.Nafasovning “*Toponimika*” [Haфасов Т, 1988] nomli maxsus kurs uchun tuzilgan dasturi ham mavjuddir [Sodiqova T, 2018].

Shuningdek, 2007-yili (V.Nafasova bilan hammualliflikda) nashr etilgan “*O‘zbek tili toponimlarining o‘quv izohli lug‘ati*” [Nafasov T., Nafasova V, 2018] alifbo tartibida, lug‘at maqolalari izohli va izohsiz (havola) tarzda berilgan va turli darajadagi tovush va shakliy o‘zgarishga uchragan toponimlarga havola maqola taqdim etilgan.

Yana bir olim N.Oxunovning “*O‘zbek tili o‘qitishda joy nomlarini o‘rganish*” [Охунов Н, 1992], “*Joy nomlari va ta’lim*” [Охунов Н, 1993], “*Joy nomlari – tilimiz lug‘at boyligi*” [Охунов Н, 1995] nomli maqolalari ham toponimlar borasidagi salmoqli tadqiqot ishlaridan hisoblanadi.

#### *Toponimlarning jahonda o‘rganilishi*

Jahon tilshunosligida H.A. Smit [Smith, 1956], A.L. Dauzat [Алберт Даузат, 1960], G.J. Kopley [Copley, 1963], G.V. Lemon [Ле-

мон Г.Б, 1783] kabilarning ilmiy izlanishlari toponimikaning fan sifatida shakllanishida alohida e'tirofga loyiq[35].

*H.A. Smitning* tadqiqotida toponimlarning nazariy asoslari, *A.L. Duzat* ishida fransuz toponimlarining ma'no guruhlarini yoritilgan. *G.J. Kopleyning* ishida joylarning umumiy va mashhur nomlari bilan bog'liq masalalar tahlil etilgan bo'lsa, *G.V. Lemon* ingliz tilidagi toponimlarning etimologiyasiga oid qarashlarini bayon etgan. Toponim tushunchasiga ta'rif bergan mashhur rus tilshunoslari *N.V. Podolskaya* va *A.V. Superanskayalar* "toponimlar Yer planetasidan tashqari har qanday geografik nomlarni ifoda etuvchi barcha so'zlar uchun umumlashtiruvchi atamadir" [Подолская Н.В., Суперанская А.В., 1969], – deya qayd etishgan.

*V.A. Nikonov* tomonidan "toponimika geografik nomlarni o'rganish bilan shug'ullanuvchi; til tarixi, dialektologiya, etimologiya, leksikologiya sohalarini kesishadigan; tarix, geografiya, etnografiya bilan uzviy aloqadorlikda bo'lgan tilshunoslikning alohida bo'limi"[Никонов В.А., 1965] ekanligi bayon qilingan.

*D.E. Rozental, M.A. Telenkovalar* toponimikani "leksikologiyaning geografik nomlarni o'rganuvchi bo'limi; biror hududning geografik nomlari jamlanmasi" [Розентал Д.Э., Теленкова М.А., 1972], deb ta'riflagan [Qilichov B, 2023].

### METODLAR

NER obyektlarini aniqlash metodlari to'rtta yondashuv asosida amalga oshiriladi:

- **Rule based** – qoidalarga asoslangan yondashuv;
- **Dictionary based** – lug'atlarga asoslangan yondashuv;
- **Machine learning(ML)** – mashinali o'qitishga asoslangan yondashuv;
- **Deep learning(DL)** – chuqur o'rganishga asoslangan yondashuv.

NER obyektlari tarkibida – toponimlarni avtomatik aniqlash ham ushbu to'rtta yondashuv asosida amalga oshiriladi. **Rule based** (qoidalarga asoslangan yondashuv) va **Dictionary based** (lug'atlarga asoslangan yondashuv) dastlabki ma'lumotlarni olish uchun mukammal tizimlardan hisoblanadi.

**Dictionar based (lug'atlarga asoslangan yondashuv)**. O'zbek tili lug'atlarga boy til. Xususan, toponimlarni avtomatik aniqlash jarayonida ham toponimlar lug'ati qimmatli manba hisoblanadi.

Toponimlar asosida tuzilgan lug'atlar:

- "*Joy nomlarining qisqacha izohli lug'ati*" (Z.Do'simov, H.Egamov) [Do'simov Z., Egamov X., 1977]

– “O‘zbekiston joy nomlarining izohli lug‘ati” birinchi nashr (Tuzuvchilar: R.Y. Xudoyberganov (f.f.n., dots.), N. Uluqov (f.f.d., prof.), M.T. Mirakmalov (g.f.d., dots.), T.J. Enazarov (f.f.d., prof.), V.T. Nafasova (f.f.n., dots.), M.M. Avezov, Sh. Temirov (f.f.f.d.), O.Bo‘riyev (t.f.d.), X.O. Bo‘riyeva (t.f.f.d.), O. Boltabayev, Y. Ahmadaliyev (g.f.d., prof.), Q. Hakimov (g.f.n., prof.), Sh.Bekturdiyev, S.G‘aybullayev, K.M. Seytniyazov (g.f.n.), Y.Ne‘matova (f.f.f.d.)) 2022[Xudoyberganov R.Y....,2022]

– “O‘zbek tili toponimlarining o‘quv izohli lug‘ati” (T.Nafasov, V.Nafasova) [Nafasov T., Nafasova V., 2018]

**Rule based(qoidalarga asoslangan yondashuv)** metodidan foydalanish o‘zbek tili grammatikasi va til qonuniyatlari asosida shakllantiriladi. O‘zbek tili grammatikasida toponimlar “onomastika” – nomshunoslik bo‘limi doirasida tadqiq qilinadi. Ushbu bo‘lim qoidalardan – asosan, atoqli otlar xususiyatlaridan kelib chiqqan holatda tadqiq qilinadi. Grammatik qoidalar o‘rganilib ularga mos modellar yaratiladi. Toponimlarni tadqiq etishda, avvalo, lug‘atlarga tayanildi. Toponimlarning grammatik shakli lug‘atlar asosida shakllantirildi. So‘ngra toponimlarning paydo bo‘lish xususiyatlari, yasalish shakli va tarkibiga ko‘ra guruhlariga ajratildi va ularga mos modellar tuzildi.

Toponimlarning keyingi turlari **Machine learning**(ML - mashinali o‘qitishga asoslangan yondashuv) va **Deep learning**(DL - chuqur o‘rganishga asoslangan yondashuv) metodlari va ularga mos matematik modellar asosida kelgusi ishlarimizda aniqlanadi.

## ASOSIY QISM

### Turdosh va atoqli ot tushunchasi

Bizni o‘rab turgan har qanday geografik obyektning nomi bor va bu “nom” obyektlarni bir-biridan farq qilish uchun xizmat qiladi. Shu sababdan, tildagi ko‘pgina so‘zlar nomlarga(atoqli otga) aylangan[Hakimov Q, 2016]. Kundalik hayotimizdagi “kutubxona”, “maktab”, “bog‘cha” so‘zlari ham obyektни bildiruvchi so‘zlar, ammo ular turdosh ot(appellyativ – lotincha “atoqli ot aksi”))lardir. Turdosh va atoqli ot tushunchalari o‘zi anglatib kelgan obyektning “nomlanish” xususiyatiga ko‘ra bir-biridan farq qiladi. Ya‘ni, nomlangan hamda nomlanmagan obyekt mazmunning kengayishi va qayta nomlanish si-fati bilan ajralib turadi. **NE (Named Entity)** – ot so‘z turkumi doirasida nomlangan obyekt – “atoqli ot”ni ifodalaydi [43].

Nomlangan obyektlarni avtomatik aniqlovchi jarayon – **NER(Named Entity Recognition)** – nomlangan obyekttni tanib olish bo‘lib, matndagi barcha nomli obyektlar aniqlanadi. **NER tabiiy tilni qayta ishlash(NLP – Natural Language Processing)** ning ko‘plab sohalarida qo‘llaniladi[Rachna Jain..., 2020]. **NER** obyektlaridan biri – *joy nomlari(toponim)*dir.

### Toponimlar

Yuqoridagi turdosh ot sifatida keltirilgan misollar (“kutubxona”, “maktab”, “bog‘cha”) “joy”ni bildiruvchi obyektlarni ifodalaydi. Ushbu so‘zlarni eshitganimizda, ko‘z oldimizda umumiy “bino” ko‘rinishiga ega obyektlar tasvirlanadi. Agar ushbu joyni bildiruvchi obyektlar qayta nom bilan nomlansa(masalan, “Alisher Navoiy kutubxonasi”, “Avloniy maktabi”, “Lolazor bog‘chasi”), ular atoqli otga aylanadi va “toponimlar” deb yuritiladi.

Joy nomlari, **geografik nomlar** yoki **toponimlar** deb ataladi. Toponimlarni **toponimika** fani o‘rganadi. **Toponimika** yunoncha **topos** – joy va **onoma (yoki onima)** – nom so‘zlaridan tarkib topgan.

Toponimlar til lug‘at tarkibining bir qismi bo‘lib, til qonuniyatlariga bo‘ysunadi[Susan Li, 2018]. Toponimlar tilshunoslik(lingvistika) ning onomastika(yunoncha **“onomastike”** – *nomlash, nom qo‘yish san‘ati*)[Qorayev S, 2006] bo‘limida o‘rganiladi.

Toponimlar joy xususiyatini, uning tarixini, ushbu obyekt bilan bog‘liq voqea-hodisani anglatgan oddiy so‘z va nomning biruvidan hosil bo‘lgan atoqli otdir. Toponimlar doirasida uchta tushunchani farqlab olish lozim: **“Toponimika”, “Toponimiya”, “Toponim”**.

– **Toponimika** – joy nomlari, ularning tuzilishi, geografik nomlarning hosil bo‘lishi va rivojlanishi bilan shug‘illanuvchi *fan*;

– **Toponimiya** – *joy nomlari yig‘indisi*(masalan, Buxoro toponimi-yasi – Buxoro joy nomlari yig‘indisi);

– **Toponim** – bu bitta joyga tegishli *xususiy nom*.

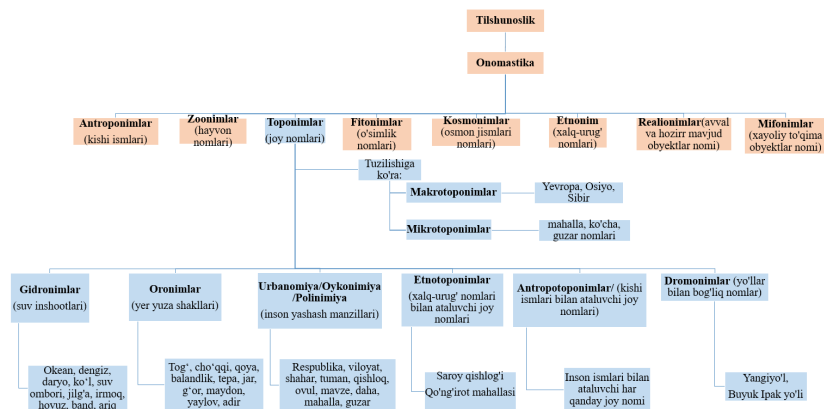
Toponimlar tilshunoslik nuqtayi nazari bilan o‘rganilganda nomning paydo bo‘lishi, uning tarixi, yaratilishi, leksik-semantik tasnifi, lingvistik strukturasi, so‘z turkumi, uning atoqli otligi, qaysi tilga mansub leksika ekanligi, toponim bilan bog‘liq til qonuniyatlari o‘rganiladi. Ammo, rivojlanib borayotgan asrimizda, lingvistik qarashlar kompyuter tiliga moslashtirilib, zamoniy elektron dasturlar yaratilayotgan bir davrda, qo‘lda joy nomlarini aniqlash va ular bilan bog‘liq statistik ilmiy tadqiqot ishlarini olib borish mushkul.

Ma‘lum bir hududdagi geografik nomlarning yig‘indisi shu hududda yashovchi xalqlarning asrlar davomida, nomlar yaratishdagi ijod

mahsuli hisoblanadi. Dunyoda qancha geografik nom borligini hech kim aniq bilmaydi. Taxminiy hisoblarga ko‘ra, butun yer sharida yarim milliarddan ortiq geografik nom bor. Holbuki, ular qatoriga soy, jilg‘a, buloq, quduq, jar, qir, mahalla, guzar, ko‘cha kabi mikrotoponimlar kirmaydi. Agar, yer yuzidagi barcha katta-kichik joy nomlarini hisobga olish imkoni bo‘lganda, son-sanoqsiz raqam hosil bo‘lar edi[https://qomus].

Ushbu zaruratlar matnni qayta ishlash jarayonida joy nomlarini avtomatik aniqlash ishini amaliyotga joriy etish kerakligini taqozo etmoqda.

### Toponimlar turi



### 1-rasm. Toponimlar va ularning turlari

#### Toponim(lar)ga xos xususiyatlar

Toponimlarga xos xususiyatlar quyidagilar:

– Doimo atoqli ot bo‘lishi;

Masalan, *Tolimarjon*, *Buxoro*, *Samarqand*...

– Doimo bosh harfda yozilishi;

Masalan, *Qo‘ng‘irot*, *Andijon*, *Toshkent*...

– Turdosh otning atoqli otga aylanishidan hosil bo‘lishi mumkin;

Masalan, *Saroy (mahallasi)*, *Tut (qishlog‘i)*...

– Toponimlardan ham yangi turdosh otlar paydo bo‘lishi yoki atoqli ot turdosh otga aylanishi (*detoponimizatsiya* – “*toponimlikdan voz kechish*”) (ko‘chishi) mumkin – Hosil bo‘lgan so‘z – *toponom* deyiladi. (masalan, *doka*, *tyul*, *boston*, *bolonya*, *saplin*, *krepdeshin*)[Hakimov Q, 2016];

## FOYDALANILGAN ADABIYOTLAR

1. Yusuf Xos Hojib. Qutadg' u bilig. Toshkent., 1991.
2. Абу Райхон Беруний. Қонуни Масъудий. I қисм. А.Расулов таржимаси. – Т.: Фан. 1973.
3. Абу Райхон Беруний. Ҳиндистон. Танланган асарлар. II том. Т., 1968.
4. Ubaydulla Karimov, Saydana. Fan., 1974.
5. Mahmud Qoshg'ariy. Devonu lug'atit turk". Fan., 1963
6. Ibn Sino. "Dengiz qirg'oqlari" 1996
7. Zahiriddin Muhmmad Bobur. Boburnoma. – Т., О'қитувчи. 2008
8. Мурзаев Э.М. Словарь народных географических терминов. – М., Мысль. 1984
9. Hasanov H. O'rta Osiyo joy nomlari tarixidan. – Т., Fan. 1965
10. Hasanov H. Geografik nomlar imlosi. – Т., Fan. 1962
11. Hasanov H. Yer tili. –Т., О'қитувчи. 1977
12. Hasanov H. Geografik nomlar siri. – Т., О'zbekiston. 1985
13. Hasanov H. Geografiya terminlari lug'ati. – Т., Fan. 1964
14. Hasanov H. O'rta Osiyolik geograf sayyohlar. – Т., 1964
15. Qorayev S. Geografik nomlar ma'nosi. – Т., 1978
16. Qorayev S. Geografik nomlar ma'nosini bilasizmi?.. – Т., 1970
17. Qorayev S. Toshkent toponimlari. – Т., Fan. 1991
18. Qorayev S. Toponomika. . – Т., О'zbekiston faylasuflari milliy jamiyati nashriyoti. 2006
19. Hudud ul-olam. – Т., О'zbekiston. 2008
20. Abu Bakr Muhammad ibn Ja'far Narshaxiy. Buxoro tarixi. – Т., 1966
21. Mahmud Qoshg'ariy. Devonu lug'atit turk". Fan., 1963
22. Zahiriddin Muhmmad Bobur. Boburnoma. – Т., О'қитувчи. 2008
23. Murzayev E.M. Toponomika ocherklari. Очерки топонимики, М., 1974
24. Madrahimov Z. Tarixiy toponomika. –., Navro'z. 2017
25. Do'simov Z., Egamov X. Joy nomlarining qisqacha izohli lug'ati. – Т., О'қитувчи, 1977
26. Nafasov T. О'zbekiston toponimlarining izohli lug'ati. – Т., О'қитувчи. 1988
27. Нафасов Т. Ўзбек номномаси. – Қарши: Насаф, 1993
28. Нафасов Т. Ўзбек қишлоқномаси. – Қарши: Насаф, 1994
29. Нафасов Т. Ўзбекистон топонимикаси. – Т., 1988. 22 б.
30. Sodiqova T. Sirdaryo viloyati toponimlari. – Т., 2018. 11b
31. Nafasov T., Nafasova V. О'zbek tili toponimlarining o'quv izohli lug'ati. – Т., Yangi asr avlodi. 2018.
32. Охунов Н. Ўзбек тили ўқитишда жой номларини ўрганиш. Олий ва ўрта махсус таълим юртларида ўзбек тилининг ўқитилишига бағи-

шланган 3-Республика илмий-амалий конференциясининг тезислари. – Урганч, 1992.

33. Охунов Н. Жой номлари ва таълим. “Таълим бўғинларида она тили ўқитиш мазмунини янгилаш асослари” мавзуидаги ўзбек тили доимий анжумани иккинчи йиғинининг тезислари. – Қарши, 1993.

34. Охунов Н. Жой номлари – тилимиз луғат бойлиги. “Таълим жараёнида сўз бойлигини оширишнинг асосий омиллари” мавзуидаги ўзбек тили доимий анжумани учинчи йиғинининг тезислари. – Тошкент: Ўқитувчи, 1995.

35. Smith, A. H. English Place-Names Elements. – Cambridge: 1956. – 163 p.;

36. Алберт Даузат. La Toponymie française. – Paris: Bibliothèque scientifique, Payot, 1960, Réimpression 1971 – 168 p.;

37. Copley, G. J. Names and Places with a short dictionary of common or wellknown place-names. – London: Phoenix House Ltd., 1963. – 226 p.;

38. Лемон Г.Б. English Etymology. – G.: Robinson, 1783. – 693 с.

39. Подолская Н.В., Суперанская А.В. Терминология ономастики// Вопросы языкознания. – 1969. – № 4. – С. 141.

40. Никонов В.А. Введение в топонимику. – М.: Наука, 1965. – С. 164.

41. Розентал Д.Э., Теленкова М.А. Справочник лингвистических терминов. – М., 1972. – С. 447.

42. Qilichov B. Onamastika. O‘quv qo‘llanma. –Buxoro., 2023. 28 b

43. Do‘simov Z., Egamov X. *Joy nomlarining qisqacha izohli lug‘ati.* – T., O‘qituvchi, 1977.

44. Xudoyberganov R.Y., Uluqov N., Mirakmalov M.T., Enazarov J., Nafasova V.T., Avezov M.M., Temirov Sh., Bo‘riyev O., Bo‘riyeva X.O., Boltabayev O., Ahmadaliyev Y., Hakimov Q., Bekturdiyev Sh., G‘aybulayev S., Seytniyazov K.M., Ne‘matova Y. O‘zbekiston joy nomlarining izohli lug‘ati. – Toshkent: Donishmand ziyosi, 2022.

45. Nafasov T., Nafasova V. O‘zbek tili toponimlarining o‘quv izohli lug‘ati. – T., Yangi asr avlodi. 2018.

46. Hakimov Q. Toponimika. – T., Mumtoz so‘z nashriyoti. 2016. 5 b

47. Rachna Jain, Abhishek Sharma, Gouri Sankar Mishra, Parma Nand, and Sudeshna Chakraborty. Named Entity Recognition in English Text. Journal of Physics: Conference Series. 2020.

48. Susan Li. Named Entity Recognition with NLTK and SpaCy. 2018.1p <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>

49. Qorayev S. Toponimika. – T., O‘zbekiston faylasuflari milliy jamiyati. 2006. 5–6 b.

50. <https://qomus.info/oz/encyclopedia/o/onomastika/>

51. Hakimov Q. Toponimika. – T., Mumtoz so‘z. 2016. 7 b.

---

## ФИЛОЛОГИЯ

УДК 81-11

### ИНДИВИДУАЛЬНОСТЬ И КОММУНИКАЦИЯ

*М. К. Абузалова, Л. Б. Авлиекулова*  
*Бухарский государственный университет,*  
*Бухара, Узбекистан*  
mehrinisoabuzalova@gmail.com,  
lolabehbutova@gmail.com

Если принять во внимание, что главным фактором в современном языкознании является человек, то при изучении человеческого языка и его общения с внешним миром изучаются также речевая ситуация, создающая это общение, и внутренние особенности человека. Индивидуальные особенности человека служат основой для их изучения. В статье показано значение этих отдельных признаков в организации процесса общения наряду с такими понятиями, как личность, личность, общение. Вопросы, которые мы хотим осветить в статье, рассматриваются на примерах различных речевых ситуаций, в том числе вербального и невербального общения.

**Ключевые слова:** индивид, человек, индивидуальность, вербальное общение, невербальное общение, кодирование, декодирование, речевая ситуация, паралингвистические средства, зрительный, акустический, тактильный, коммуникативный, коммуникативный, интерактивный, перцептивный.

### INDIVIDUAL AND COMMUNICATION

*Mekhriniso Abuzalova, Lola Avliyokulova*  
*Bukhara State University,*  
*Bukhara, Uzbekistan*  
mehrinisoabuzalova@gmail.com,  
lolabehbutova@gmail.com

If we take into account that the main factor in today's linguistics is a person, while studying the human language and its communication with the outside world, the speech situation that creates this communication and the internal characteristics of the person are also studied. Individual characteristics of a person serve as a basis for their study. The article shows the importance of these individual signs in organizing the communication process along with concepts such as individual, person, communication. The points we want to cover in the article are considered as examples of various speech situations, including verbal and non-verbal communication.

**Key words:** individual, person, individuality, verbal communication, non-verbal communication, coding, decoding, speech situation, paralinguistic tools, visual, acoustic, tactile, communication, communicative, interactive, perceptive.



## INDIVID VA KOMMUNIKATSIYA

*Abuzalova Mexriniso Kadirovna,  
Avliyoqulova Lola Bebutovna*

*Buxoro davlat universiteti*

*Buxoro, O'zbekiston*

mehrinisoabuzalova@gmail.com,

lolabehbutova@gmail.com

Bugungi kun tilshunosligida bosh omil inson ekanligini inobatga oladigan bo'lsak, inson tilini, uning tashqi olam bilan muloqotini o'rganish barobarida, shu muloqotni yuzaga keltiruvchi nutqiy vaziyat hamda shaxsning ichki xususiyatlari ham tadqiq etiladi. Ularni o'rganishda esa shaxsdagi individual belgilar asos bo'lib xizmat qiladi. Maqolada individ, shaxs, kommunikatsiya kabi tushunchalar bilan bir qatorda ushbu individual belgilarning kommunikatsiya jarayonini tashkil etishdagi ahamiyati ko'rsarib beriladi. Maqolada yoritmoqchi bo'lgan fikrlarimiz verbal hamda noverbal kommunikatsiyani o'z ichiga olgan turli nutqiy vaziyatlar misolida ko'rib chiqiladi.

**Kalit so'zlar:** individ, shaxs, individuallik, verbal kommunikatsiya, noverbal kommunikatsiya, kodlash, dekodlash, nutqiy vaziyat, paralingvistik vositalar, visual, akustik, taktil, muloqot, kommunikativ, interaktiv, perceptiv.

Manbalarda inson haqida “u borliqning oliy ko'rinishi”, “ijtimoiy – iqtisodiy taraqqiyot mahsuli”, “o'zini o'zi anglovchi eng aqlli mavjudot”, “antroposotsiogenez jarayonining bir ko'rinishi”, “tana, ruhiyat va ma'naviyatning yig'indisi” degan ko'plab ta'riflar keltirilgan. Inson so'zini izohlash, uning mohiyati, kelibchiqishi, tarixiy taraqqiyoti haqida ham yuzlab qarashlar mavjuddir. Bu tushuncha ijtimoiy fikr tarixida uch xil (mifologik, ilmiy, diniy-nazariy) nuqtayi nazardan o'rganilgan. Hozirgi kunda ham inson omili bir qancha fanlarning tadqiqot obyekti hisoblanadi. Jumladan, tibbiyot, psixologiya, antropologiya, sotsiologiya, pedagogika, psixolingvistika, tilshunoslik, adabiyotshunoslik va h.k. Biroq har bir fan unga o'zining qarashidan kelib chiqqan holda yondashadi. Jumladan, antroposentrik tilshunoslik rivojida tadqiqotlar markazida inson omili qo'yilmoqda. Masalan, psixologiya shaxsni sezgilar va idrok obrazlari, ong va tuyg'u, muomala, temperament kabi psixologik jarayonlar vositasida o'rgansa, falsafa fanining antropologiya bo'limida inson-individ- shaxs tushunchalari izohlangan. Unga ko'ra, inson tug'ulgan vaqtida individ bo'lib, u jamiyatning bir a'zosi bo'ladi, ulg'aygan sayin o'zining jismoniy holati, mehnat faoliyati, ruhiy dunyosi bilan asta-sekin individdan shaxs darajasiga ko'tarilib boradi.

Individ lotincha “individum” soʻzidan kelib chiqqan boʻlib, “boʻlinmaydigan” degan maʼnoni anglatadi. Bu atama mustaqil yashaydigan tirik vujud, har bir shaxsni bildiradi. Individ bu atrofimizdagi barcha odamlardir. Yaʼni, jinsi, yoshi, millati, ijtimoiy kelib chiqishi, dini, irqi, kasbi, jismoniy va ruhiy holatidan qatʼi nazar, individ hisoblanadi. Xalqimiz orasida yangi tugʻilgan chaqaloq “yoshiga yetgunicha qirq xil turlanadi” (yaʼni, qirq kishiga oʻxshaydi) degan fikr bor. Va bu qarash fan doirasida ham oʻz ifodasini topgan, desak yanglishmagan boʻlamiz. Yaʼni, har bir individ tugʻilganidan boshlab har kuni oʻsib ulgʻaygani sayin oʻzgarishda boʻladi. Bu oʻzgarish faqat tanadagi jismoniy holatlar boʻlib qolmay, balki aqliy rivojlanish hamda ruhiy holatida ham aks etadi. Statistik maʼlumotlarga qaraganda, dunyo aholisi soni sakkiz milliarddan oshiqdir. Eng hayratlanarlisi shundaki, bu sakkiz milliard individning hech qaysi bir-birini aynan takrorlamaydi. Hattoki, ona organizmida bir vaqtda rivojlanib, bir vaqtda dunyoga keluvchi egizaklar ham bir-biriga toʻlaligicha oʻxshash emas. Yaʼni, ular bir-birlariga qanchalik oʻxshamasinlar, ularning har biriga xos boʻlgan xarakter, temperament, iroda, qobiliyat, iqtidor va shu kabilar takrorlanmasdir. Chunki har bir insonda boshqalarda uchramaydigan, faqat uning oʻzigagina xos boʻlgan, takrorlanmas, xususiy belgilari boʻladi. Bir odamga xos boʻlgan iqtidor, qobiliyat, emotsiya, mimika, ovoz va shu kabilar boshqa bir odamda aynan uchramaydi. Xuddi mana shu belgi sifatlarining oʻziga xosligi bilan individuallik umumiylikka qarshi qoʻyiladi.

Bugungi kunda ommaviy talab darajasiga koʻtarilayotgan psixologiya sohasida har bir insondagi individual belgilarga alohida eʼtibor qaratish nechogʻli muhim ekaniga alohida urgʻu beriladi. Psixologlarning ish uslublariga eʼtibor qaratadigan boʻlsak, ular oʻz ishlarini, avvalo, mijozlarini tinglashdan, ularni kuzatishdan boshlaydilar. Bu kuzatuv jarayoni qancha davom etishini esa oldindan aytish qiyin. Bu uch kun, bir hafta, ikki oy va hattoki yillab davom etishi mumkin. *“Bu jarayon nihoyatda murakkab boʻlishi bilan birga qiziqarli ham. Chunki, shu vaqt mobaynida siz bemor bilan birga yashashni boshlaysiz. Uning bolaligini u bilan birga quvonch bilan esga olasiz. Xotira daftarlarini varaqlaysiz, maktabdagi doʻstlari, ilk muhabbati bilan tanishasiz. Ish faoliyatidagi yutuqlari-yu, “omadsiz” kunlarigacha u bilan birga boʻlasiz. Bemorni, uning muammosini oʻrganish faqat uni tinglash bilan oʻrganilmaydi, baʼzan siz bemorning qoʻn- u qoʻshnilari, oilasi, hamkasblari, doʻstlari bilan ham muloqot qilishingizga toʻgʻri kelib qolishi mumkin”*, deydi oliy toifali psixolog Donyor

Boboyorov. Bemor bilan muloqot jarayoni psixologga mijoz shaxsini o'rganishga, undagi o'ziga xos individuallikni ko'rishga xizmat qiladi. "Dastlab individuallik antik davrda atoqli yunon faylasuflari Levkipp va Demokrit tomonidan muayyan shakl va mazmunga, ya'ni o'ziga xos belgilarga ega bo'lgan narsa va hodisalarga, jumladan, atom yoki individ (ya'ni bo'linmas) kabilarga xos jarayonlarni aniqlash jarayonida ma'lum bo'lgan". O'rta asrlar falsafasida esa bu so'z aynan individ tushunchasini ifodalagan. Zero, inson tug'ilishidan tortib, toki uning so'nggi nafasigacha bo'lgan davrda bosib o'tgan umr yo'lida pog'ona-ma-pog'ona o'zligini topadi. Ya'ni, hayotda bo'ladigan har bir hodisa, uchraydigan har bir odam bir tajriba bo'lib, individdan individuallik-kacha o'zgaradi.

Yuqorida qayd etib o'tganimizdek, inson bir necha fan va sohalarning tadqiq obyekti sanaladi. Psixolingvistika esa shular jumlasidandir. Bu sohaning obyekti nutqiy faoliyatning subyekti bo'lgan shaxs, til egasidir. Ushbu soha psixologiya va tilshunoslik kesishgan nuqtada vujudga kelgan bo'lib, fanning predmetini belgilashda esa ayrim yevropalik psixolingvistlarning qarashlarini keltirish o'rinli bo'ladi. Jumladan, P.Fress psixolingvistikani "til orqali taqdim etiladigan bizning ekspressiv va kommunikativ ehtiyojlarimiz va vositalarimiz o'rtasidagi munosabatlar haqidagi fan", – deb izohlagan. T.Slamakazu psixolingvistikaning predmeti xususida "muloqot vaziyatining xabarga ta'siri" degan qisqacha xulosaga keldi. Bundan tashqari, Charlz Osgudning ta'kidlashicha, "Psixolingvistika so'zlovchi niyatining muayyan bir kodi (tili)da signalga aylanishi va bu signallar tinglovchining interpretiyasida qaytadan namoyon bo'lishi jarayonlarini o'rganadi. Boshqacha qilib aytganda, psixolingvistika KOMMUNIKATSIYANing kodlash va dekodlash jarayonlari bilan shug'ullanadi. Alqissa, ushbu olimlar qarashlarini bir joyga jamlaydigan bo'lsak, psixolingvistika sohasi muloqot va kommunikatsiya jarayoni asosida yuzaga kelgan de-yish mumkin. Zero, bularning negizi bevosita inson (individ)ga borib taqaladi.

Xo'sh, kommunikatsiya o'zi nima? Uning hosil bo'lishida individ tushunchasining o'rni qanday? Ushbu so'z lotincha "communicatio" so'zidan olingan bo'lib, "umumlashtiraman", "bog'layman" degan ma'noni bildiradi. Ya'ni, tirik (individlar o'rtasidagi o'zaro munosabat va hayvonlar o'rtasidagi signallar almashinuvi) va o'lik tabiatdagi (texnika olami) tizimlar o'rtasidagi o'zaro ALOQA. Individlar esa mana shu kommunikatsiyaning ishtirokchilaridan biridir. Individ tug'ilganidan so'ng tashqi olam bilan bevosita bog'liq bo'ladi. Oziqlanish, nafas

olish, ko‘payish, mehnat qilish, dam olish uning yashashi uchun zarur bo‘lgan tabiiy ehtiyojlar sanaladi. Xuddi shular singari bir-birlari bilan o‘zaro kommunikatsiya o‘rnatishlari ham zaruriy ehtiyoj bo‘lib, bu “ehtiyoj” asosida kishilik jamiyati shakllandi deyish mumkin. Neyropsixolog va neyrolingvist A.Luriyaning ta’kidlashicha, “Insoniyat paydo bo‘lgan davrlarda muloqot jarayoni tovushlarsiz amalga oshirilgan. O‘ylash mumkinki, birinchi muloqot faol mehnat jarayonida paydo bo‘lgan. Bunda ayni bir faoliyatga oid jestlar natijasida muloqot yuzaga kelgan va bu bevosita o‘sha faoliyat jarayonining ishtirokchilarigagina tushunarli bo‘lgan. Darhaqiqat, til hali shakllanmagan zamonlarda ham insonlar o‘zaro kommunikatsiya o‘rnatganlar. Masalan, xalqlar o‘rtasidagi urushlar misolida ko‘rsak, bugungi kuchli zamonaviy qurollardan tortib, hattoki uchli nayza va o‘tkir toshlar bilan jang olib borilgan davrlarda ham odamlar o‘zaro aloqa o‘rnatganlar. Faqat bunda kommunikatsiya vositasi so‘z emas, narsa-buyum (sovuq qurol) bo‘lgan. Yoki tarixiy san‘at asarlari hamda me‘moriy obidalarning bugungi kungacha saqlanib qolingani asosida bundan bir necha yuz, hattoki, ming yillar muqaddam yashab o‘tgan ajdodlar bilan o‘zaro aloqa yuzaga keladi.

Kishilar o‘rtasidagi kommunikatsiya jarayonining eng optimallashtirilgan ko‘rinishi bu ularning nutqiy faoliyat asosida muloqot qilishlaridir. “Nutqiy faoliyat – faol, aniq bir maqsadga qaratilgan, til vositasida amalga oshiriladigan va holatlardan kelib chiqqan holda insonlar o‘rtasida o‘zaro ta’sir qilish, muloqot qilish jarayonidir. Nutqiy faoliyat so‘zlash va tushunish jarayonlarining umumlashmasi, nutq yaratilishi va uning qabul qilinishiga oid barcha hodisalardir”. Aytaylik, so‘zlovchi, ya’ni 1-shaxsda ma’lum bir axborot paydo bo‘ladi va ushbu axborotni nutq organlari yordamida havo tebranishi bilan 2-shaxsga (tinglovchiga) yuboradi. O‘zaro axborot almashinish hodisasini vujudga keltirgan ushbu jarayon nutqiy muloqot deyiladi. Kommunikatsiya uchun faqatgina axborot uzatish va uning mazmunidagi idroki emas, balki nutqiy vaziyat, kommunikativ holat, individning tafakkuri, psixologik holati va boshqalar ham muhim hisoblanadi. Bu kabi nutq mazmunini oldindan belgilab beruvchi omillar, odatda, “muloqot manbayi” sifatida namoyon bo‘ladi. Muloqot manbayi esa gapdagi so‘zlar ifodalagan mazmunga emas, balki vizual vositalar (ko‘z qarashi bilan bog‘liq), mimika, imo-ishora, tana xatti-harakati, ovoz tempi, muloqot tempi, urg‘u, intonatsiya, ritm, tovushlar artikulyatsiyasi, tildagi dialekt, hat-toki, so‘zlovchining sog‘ligi bilan bog‘liq o‘rinlardir. Bularning bar-

chasi muloqotning noverbal ko‘rinishlari bo‘lib, “paralingvistik vositalar” deb ham yuritiladi. Noverbal vositalarining sohada uch ko‘rinishi o‘zaro farqlanadi. G. V. Kolshanskiy paralingvistik vositalarni quyidagicha ajratgan:

1. Vizual – ko‘z bilan ko‘rish orqali idrok etish mumkin bo‘lgan omillar (mimika, imo-ishora, xatti-harakat).

2. Akustik – eshitish orqali idrok etiladigan omillar (ovoz tembri, intonatsiya, tovushlarning baland yoki pastligi va h.k)

3. Taktill – ushlab ko‘rish, tegish orqali his qilish mumkin bo‘lgan omillar (turli belgi, ishora, ko‘rinish).

Muloqot – murakkab polifunksional tushuncha. Bu jarayonda individlar o‘zaro axborot almashadilar, bir-birlariga hissiy ta‘sir o‘tkazadilar, bir-birlarini idrok etadilar. Bu funksiyalarning barini o‘zida to‘liq mujassam etishda esa noverbal kommunikatsiya o‘zini namoyish etadi. Ushbu jarayonlarning ayrimlarini esa quyidagi misollar orqali ko‘rib o‘tamiz:

*“Agar kimningdir ko‘zlarida tomchi yosh aylanib, labida nim tabassum bo‘lsa, bilingki, bunday insonning ruhiy holatida ikki jarayondan biri hukmron bo‘ladi. Bularidan biri uning ruhi nihoyatda sokin bo‘lib, bu mamnuniyatni ifoda etsa, ikkinchisi kuchli ruhiy og‘riqning aksi bo‘lib, u kuchli alamzadalikni ko‘rsatadi”,* – deydi oliy toifali psixolog Nodira Fayziyeva. Ko‘rinib turibdiki, kishidagi bu kabi ruhiy holatlarni ifoda etishda vizual omillar yetakchilik qiladi. Yoxud birgina qoshlarning harakatida ham kishining jarayonga bo‘lgan hayratlanish, jahl, kuchli diqqat, xafalik singari bir qancha munosabatlari aks etadi. Ma‘lumotlarga qaraganda, imo-ishoralar(qo‘l motorikasi), mimika(yuz motorikasi), pantomimika (butun gavda motorikasi) butun axborotning 80 fozini ifodalab berar ekan.

Yig‘i, kulgu, qo‘rquv, jahl va boshqa hissiy emotsiyalar ham noverbal kommunikatsiyaning eng faol ko‘rinishlaridan biridir. Raqs tushish, jismoniy mehnat faoliyati bilan shug‘ullanish, yakka va jamoaviy sport mashg‘ulotlari esa butun gavda motorikasining yorqin namunasidir. Ayniqsa, jamoaviy sport o‘rinlari (futbol, basketbol, volleybol va boshqalar) orqali namoyon bo‘luvchi tana harakatlari asosida muloqotning kommunikativ (axborot uzatish), interaktiv (o‘zaro birgalikda harakat qilish), perseptiv(birgalikda tushunish) idrok qilinishi amalga oshiriladi. Bundan tashqari, qo‘l motorikasi bilan bog‘liq o‘rinlardan biri – salomlashish ham kishilarning milliy-madaniy odatlari jihatdan bir-biridan farqlanadi. Masalan, o‘ng qo‘lni ko‘ksning ustiga qo‘yish (o‘zbek xalqi), qo‘llarni o‘zaro juftlash (hind xalqi), qo‘llarni o‘zaro

juftlab bosh va gavdani biroz egish (koreys xalqi), kaftlarning orqasini o‘zaro bir-biriga tegizish (afrika xalqlaridan biri) milliy-madaniyatini aks ettiradi. Psixologizm hodisasini ilmiy-nazariy muammo sifatida tadqiq qilish borasida shuni ta’kidlash lozimki, psixologik matn markazida inson obrazi turar ekan, uning ruhiy olamini kashf etish, ichki olamida kechuvchi tizginsiz evrilishlarni tasvirlash, yozuvchining badiiy fantaziya imkoniyatlarini o‘rganish bo‘yicha izlanishlar olib borilmoqda. Jumladan, badiiy matnlarda psixologizm ko‘rinishlari, ularning obraz fe‘l-atvorini ko‘rsatib berishdagi badiiy vazifalari, badiiy lirik qahramonning ichki olami, ma’naviy dunyoqarashi, olam va odam haqidagi falsafiy qarashlari ham o‘z ifodasini topishi bilan bog‘liq muammolar yuzasidan ilmiy asoslangan masalalarni hal etishga jiddiy e’tibor qaratilmoqda.

Odatda, muloqotning verbal ko‘rinishi bilan noverbal ko‘rinishiga kommunikatsiyaning mukammal bir butunligi sifatida qaraladi. Chunki, ular o‘zaro bir vaqtda yoxud ketma-ket sodir bo‘lish orqali bir-birini to‘ldirib turadi. Ammo bu ikki muloqot shakli doim ham bir xil axborotni uzatish uchun xizmat qilavermaydi. Ba’zida ular bir-birini mutlaqo rad etuvchi axborotlarni ham ifodalashi mumkin. Bu hodisa odatda, nutq vaziyati va insonning (ba’zan axborot uzatuvchining, ba’zan axborotni qabul qiluvchining) ruhiy holati bilan bog‘liq holda amalga oshadi. Masalan, ko‘pchilik insonlar yolg‘on gapirgan vaqtlarida, turli imo-ishora va xatti-harakatlarni namoyish etadilar (ko‘zni tinglovchidan olib qochish, turli qo‘l va barmoq harakatlari). Bunday vaziyatlarda so‘zlovchining so‘z orqali uzatayotgan azboroti bilan xatti-harakatlari bir-birini rad etadi. Fikrimizni yanada tiniqroq ifoda etish maqsadida dunyoga mashhur detektiv filmlardan biri bo‘lgan “Erkyul Puaro” filmidagi quyidagi o‘ringa diqqat qaratamiz: Tasodifiy avtohalokat tufayli vafot etgan erining o‘chini olish maqsadida mashina haydovchisiga bir necha bor qasddan suyiqasd uyushtirgan Emma ismli ayolning navbatdagi tergov jarayonidan lavha:

**Tergovchi:** – *Aytingchi, xonim, qilgan ishingizdan chindan ham pushaymonmisiz?*

**Ayblanuchi:** – *Ha janob, pushaymonman. O’sha vaqtda kuchli depressiya ta’sirida bo‘lganman.*

*Ammo hozir qilgan ishim qanchalik noto‘g‘ri ekanini tushunib yerdim.*

**Tergovchi:** – *Demak, sizni ozod qilsak, bunday noma’qul harakatlarni boshqa amalga oshirmaysiz, shundaymi?*

**Ayblanuvchi:** – *Yo‘q, aslo!*

**Tergovchi:** – *Xo‘p yaxshi, ammo biz sizni qo‘yib yubora olmaymiz. Aksincha, oldingidan ham qattiq nazoratda ushlab turamiz. Chunki, siz qilgan ishingizdan mutlaqo pushaymon emassiz, aksincha, bu yerdan tezroq qutulib, yangi rejangizni amalga oshirishga shoshilmoqdasiz!*

Xo‘sh, ushbu lavhada tergovchi nima sababdan bunday xulosaga keldi?, degan savol tug‘iladi. Javob esa juda oddiy. Ayblanuvchi ayol “Yo‘q, aslo qilmayman”, degan vaqtda nutqdagi “**inkor**” ma’nosini ifoda etuvchi (boshni o‘ng va chapga qisqa-qisqa harakatlantirish) harakatni emas, balki “**tasdiq**” ma’nosini beruvchi (boshni yuqoridan pastga harakatlantirish) harakatni amalga oshirdi. So‘z va harakatning bir-biriga mutanosib emasligi esa uning miyasidagi chalkashliklar, ya’ni ikkilanishni ifodalaydi. Bundan tashqari, ayblanuvchi o‘zining pushaymonligini tushuntirayotgan vaqtda tergovchiga nihoyatda kuchli nigoh bilan qarab turishi ham tergovchida shubha uyg‘otadi. Chunki, odatda, insonlarga bunday nigoh qarash ularni nimagadir ishontirish uchun qo‘llaniladi. Pushaymonlik holatida esa ko‘pincha, nigohlar yerga qaratilgan, bosh esa xiyol egilgan holatda bo‘ladi. Ko‘pchilik tomoshabinlar tomonidan iliq kutib olingan “Chuqur” nomli turk serialidan olingan ushbu muloqot jarayoniga diqqat qiling:

*“Sen meni la‘natimsan. Sen tufayli boshimga kelmagan kulfat qolmadi, yoki qoldimi, aytchi?”*

*Qoldi, faqat jahannamning olovi qoldi. Ammo qara, men ketmadim, men hamon yoningdaman”.*

Qizning o‘z suhbatdoshiga aytgan so‘zlarini tahlil qiladigan bo‘lsak, undagi so‘zlar ma’no jihatdan **salbiy munosabatni** aks ettiradi, biroq qizning mimikasi, ya’ni yuz ifodasi matnni aks ma’noda tushunishga imkon yaratadi. Ya’ni, qizning samimiy tabassumi, xotirjam nigohlari uning ushbu nutqi qarshisidagi insonga nisbatan nafratni ifodalash uchun emas, balki **mehrni** ko‘rsatish uchun qo‘llanilganini tushunish qiyin emas. Kommunikatsiya jarayonidagi shu va shunga o‘xshash boshqa muloqot vaziyatlari psixolingvika hamda paralingvistika hamkorligida o‘rganiladi. Zero, inson “tana, ruhiyat va ma’naviyatning yig‘indisi”. Bundan ko‘zlangan yagona maqsad esa insonning muloqot jarayonini va u orqali inson shaxsini o‘rganishdir.

#### ADABIYOTLAR:

1. Fraisse P. La psucho-linguistique//Problems de psucho-linguistique. – Paris, 1965. – P.5
2. Jumayeva N. Individ, shaxs va individuallik tushunchalari. Maqola. – Toshkent, 2021-B.2

3. Лурия А. Основы нейропсихологии. – Москва: Академия, 2003 – С. 184
4. Osgood Che. E. Psychology: a Study of Science. – New York, 1963. V.6
5. Slama-Cazacu T. Indroton to Psucholinguistics.–The Hauge – Parij, 1973. – P. 57
6. Temirova F. Noverbal kommunikatsiya – paralingvistikaning oʻrganilish tarmogʻi sifatida. Maqola. – Toshkent, 2021
7. Xolmonova Z., Usmonova Sh. Psixolingvistika. – Toshkent, 2018 – B. 14
8. Абузалова М.К. Лингвистика ва тиббиёт ҳамкорлиги ижтимоий зарурат сифатида/”Soha lingvistikasi: tilshunoslik va tibbiyot integratsiyasining muammolari, yechimlari hamda istiqbollari” mavzusidagi xalqaro ilmiy-amaliy konferensiya materiallari.–Buxoro: Tadqiqot.uz, 2021.– B 19.
9. Юсупова Д.Ю. Ҳ.Худойбердиеванинг сўзларни кўчма маънода қўллаш маҳорати//НамДУ илмий ахборотномаси – Научный вестник НамГУ 2021 йил 6-сон. – 395-402- б.



УДК 81-11

**ЛЕКСИКА ОРХОНСКИХ ПАМЯТНИКОВ И ЕЁ ОТРАЖЕНИЕ  
В КАРАКАЛПАКСКОМ ЭПОСЕ «ҚЫРҚ ҚЫЗ»****Ш. Н. Абдиназимов**

*Каракалпакский государственный университет  
им. Бердаха, Каракалпакстан, Узбекистан  
shamshetdinabdinazimov@gmail.com*

В статье представлена информация об исследовании каракалпакского народного эпоса «Қырқ қыз», его связи с историей и лексический состав: тюркские слова, встречающиеся в орхонских памятниках. Эти слова были разделены на несколько тематических групп и проанализированы. То есть, родственные термины и слова, связанные с частями тела человека, общественно-политические слова, астрономические объекты, меры веса, слова, обозначающие время, слова, выражающие объем, действие и качество, слова, входящие в другие тематические группы.

**Ключевые слова:** каракалпакский язык, эпос, семантика, словарный состав, древние тюрки, тюркские языки.

***Abdinazimov Shamshetdin Nazhimovich***

*Karakalpak State University named after. Berdaha  
Karakalpakstan, Uzbekistan  
shamshetdinabdinazimov@gmail.com*

The article provides information about the study of the Karakalpak folk epic “Қырқ қыз”, its connection with history and lexical composition: Turkic words found in the Orkhon monuments. These words were divided into several thematic groups and analyzed. That is, related terms and words associated with parts of the human body, socio-political words, astronomical objects, measures of weight, words denoting time, words expressing volume, action and quality, words included in other thematic groups.

**Keywords:** Karakalpak language, epic, semantics, vocabulary, ancient Turks, Turkic languages.

Эпос «Қырқ қыз» – это произведение, принадлежащее каракалпакскому народу, которое не имеет аналогов у других тюркских народов. Оно было записано у известного каракалпакского жырау Қурбанбая Тажибаяева в период с 1938 по 1944 годы, после чего стало доступно широкой общественности. Запись эпоса была выполнена собирателями каракалпакского фольклора А. Бегимовым, Ш. Хожаниязовым и С. Мауленовым. Впервые эпос был опубликован на каракалпакском языке в 1949 году [Қырқ қыз 1949]. Поз-

же он переиздавался в 1956 и 1980 годы. В период с 2007–2015 годы был опубликован 100-томник

«Қарақалпақ фольклоры», в 9 томе которого вновь переиздан вновь. Эпос «Қырқ қыз» неоднократно переводился на русский, узбекский, казахский, туркменский, кыргызский и другие языки [Сорок девушек 1952; 1951; 1956; 1948; 1959; 1960; 1959; 1956]. С.П.Толстов, Т.А.Жданко, А.С.Морозова, Л.С.Толстова, П.П.Иванов, С.Камалов и др. исследователи, в своих научных работах оставили ценные сведения, касающиеся периода его создания, связи событий, отраженных в дастане с историческими вехами и ряд других проблем [Толстов 1948: 2; Жданко 1958: 113–114; Морозова 1960; Толстова 1984: 186–206; Иванов 2021].

На территории Элликалинского района Республики Каракалпакстан расположена крепость «Қырқ қыз». Памятник имеет почти прямоугольную форму, его длина составляет 300 м, ширина 240 м, высота 7–8 метров, на южной стороне расположен вход. По четырем сторонам городища имеются сторожевые башни. Основание памятника выложено кирпичами, размер которых составляет 50х50, 50х30 см. С.П.Толстов предполагал, что эпоса «Қырқ қыз» связан с фракийско-массагетской культурой. В дастане обнаруживаются отголоски массагетского эпоса о Тумарис. Византийские дипломаты писали, что «в VI веке на севере Хорезма, на побережье Арала находились тюркские племена, возглавляемые женщиной». Развалины крепости «Қырқ қыз» (пер. сорок девушек), относящиеся к I–VI векам н.э., в тюркской или иранской форме (Чильдухтаран), известны в Средней Азии по руинам (Мерв, Термез и др.) и ассоциируются со столь же распространенной легендой о царевне и её сорока подругах, выступающих в виде воительниц-амазонок, как в одноименном каракалпакском эпосе [Толстов 1948: 22]. Т.А. Жданко называя каракалпакский эпос «Қырқ қыз» памятником эпохи матриархата, отмечает, что он является бесценным произведением, позволяющим пролить свет на события V века до н.э. Она выдвигает предположение, что каракалпаки являются потомками амазонок прошлых времен [Жданко 1950: 101–102].

Дастан «Қырқ қыз» отличается богатством языка, в котором сохранились элементы, отражающих развитие тюркских языков. Особый интерес вызывает одна деталь дастана

«Қырқ қыз». В древние времена, правители тюркских племен имели обычай оставлять повествования о своей храбрости, под-

вигах во имя народа, военных походах на века, высекая их на камне. Древние тюрки имели развитую систему письма. Култегин, Билке Каган, Тонькок и др. памятники являются очевидными примерами тому. В дастане «Қырқ қыз» мы являемся свидетелями существования такой традиции. Так в одном из эпизодов, главный герой эпоса Гулайым, предводительница сорока девушек, ведет борьбу за освобождение народа, попавшего в руки врага. Одержав победу над врагом в бою она повелевает высечь это событие на камне:

Бул булақтың бойына,  
Ақ шатырды қурғызды.  
Жойылмастай мәңгиге,  
Тастан белги ойғызды.  
Кешегі болған урысты,  
Бул тасларға жаздырды  
[Қырқ қыз 1980: 331].

На берегу родника,  
Поставила белый шатер.  
Дабы не забылось во веки,  
На камне высечены слова.  
О победе во вчерашней сече

Словарный состав дастана «Қырқ қыз» богат и многослоен. Лексика дастана имеет ряд соответствий с семантическими группами памятника Култегин:

а) Родственные взаимоотношения и слова, связанные с частями тела человека: ака (49, КТБ) – аға (старший брат): Дарға асар алты бирдей ағанды (Повесит дарга всех шестерых родных ага) (48); ачим (30, КТМ) – ата (дедушка): Тап сол күни атаң минген торыға (48) (Гнедой, на которого сел в тот день ата); ини (31, КТМ) – ини (младший брат): Аға менен иниңди (62) (Старшего и младшего братьев); апа (предки, 28, КТМ) – апа (мать): Әне, апа деп мениң билгеним (Ну вот апа, это все, что я знаю) (49); қыз (29, КТМ) – қыз (девушка): Баяғыда қыз таппай, иним қайтып келген соң (Не найдя тогда девушку ту, брат вернулся ни с чем) (176); кәлин (33, КТБ) – келин (сноха): Хә, келинжан, келинжан (Эй, сношенька, сношенька) (143); йас (33, КТБ) – жас (слёзы): Көзден аққан жасыма (Слезы, льются из глаз моих) (166); көз (33, КТБ) – көз (глаза): Екі көзі аларып (Вытаращив глаза) (166); әр (32, КТМ) – ер (храбрый, сильный): Ер арысланның туқымы (Храбрых воинов потомок) (231); ат (31, КТМ) – ат (имя): Атам атап қойған атым Арыслан (Отцом наречен именем Арыслан) (98); қаш (33, КТБ) – қас (брови): Қара қаслы, қолаң шашлы (Чернобровая, с густыми черными волосами) (42); қан (30, КТМ) – қан (кровь): Қаны қашып жүзинен (Кровь отлила от лица) (55) х.т.б.

б) общественно-политические слова: ил (28, КТм) – ел (народ): Елден ерек болыўға (Быть в стороне от народа) (44); бай (28, КТм) – бай (богатый): Аллаяр деген бай болды (Богач по имени Аллаяр) (42); кул (29, КТм) – кул (раб): Кулы болып есикте от жаққаннан (Чем рабом быть и прислуживать у порога) (43); күң (29, КТм) – гүң (рабляя): Гүң қылып буны айдаңлар (Закройте ему рот и гоните прочь) (236); йағ (49, КТб) – жаў (враг): Жаўдың сести шыққанда (Когда послышится шум набега) (280); тәңри (27, КТм) – тәңир (бог): Мени неге бөндем дедиң тәңирим (Почему назвал рабом меня, мой Бог) (123).

в) астрономические объекты, меры веса, слова, обозначающие время: күн (27, КТм) – күн (солнце): Күн шықпастай қырқ жыл камал қылса да (Сорок лет не давал видеть солнца) (43); көк (28, КТм) – көк (небо): Шабытланып бәрха көкке қараса (С воодушевлением смотря на небо) (48); шаш (28, КТм) – тас (камень): Қәхәрленсем тасты кескен полатым (Стать, разрезающая в гневе камень) (186); тағ (29, КТм) – таў (горы): Таўларға шығып кетемен (На гору взберусь) (183); йир (30, КТм) – жер (земля): Атаўдың бәлент жерлерин (Высокие места острова) (44); күнтүз (31, КТм) – күндиз (днем): Кеше-күндиз түрли ойын ойнадың (Дни и ночи напролет играли в различные игры) (45); көл (32, КТм) – көл (озера): Шағыр көл менен Қорық көл (Озеро Шағыр и озеро Қорық) (236); йол (33, КТм) – жол (дорога): Жол үстінде зарлады (Горько плача на дороге) (253).

д) слова, выражающие объем, действие и качество: арық (28, КТм) – арық (худой): Арықларды бағыўға, семизлерди байлаўға (Худых выкармливать, жирных на привязь) (83); өлти (28, КТм) – өлди (умер, смерть): Өл десең бүгин өлемен (Если скажешь умру сейчас же) (223); йарығлығ (30, КТм) – жарақлы (с оружием): Жарақсыз батыр болмайды (Без оружия нет батыра) (277); йадағ (31, КТм) – жаяў (пешком): Атлы-жаяў келгенлер (Прибывшие на конях и пешими) (67); улуғ (31, КТм) – уллы (великий, большой): Саркоптың уллы батыры (Великийбогатырь Саркопа) (249); калти (33, КТм) – келди (пришел): Келди байдың қасына (Пришел к баю) (89); тыңла (27, КТм) – тыңла (слушай): Қулақ салып тыңлаңлар (Слушайте меня внимательно) (458).

ж) слова, входящие в другие тематические группы: көнул (28, КТм) – кеўил (душа): Кеўилде әрман қалмасын (Не оставляй в душе не исполненных желаний) (51); он (31, КТм) – он (десять): Он алты өгиз, он бир мал (Шестнадцать быков, одиннадцать голов

скота) (447); ат (31, КТм) – ат (лошадь): Киснейди ат, сести көкти жарады (Громогласно ржет конь) (46); оқ (31, КТм) – оқ (стрела): Сары жайдың оқлары қалмаққа тәсир етпеди (Стрелы желтого лука не повредили калмакам) (243); қылич (32, КТм) – қылыш (меч): Қылыш берип қолына (В руки меч вручил) (45); арик (32, КТм) – ерк (воля): Барсамдағы ерким бар, бармасам да ерким бар (Волен пойти и волен не пойти) (222).

Встречаются слова, имеющие изменения в семантическом строении: Если в памятнике Құлтегина слово «аруқ» передает значение «таза, пәк» – чистый, в языке дастана оно используется в значении «азғын, арық», что значит худой, истощенный: Арықларды, бағыұға, семизлерди байлаұға (Худых на откорм, жирных на привязь) (47). В памятнике Құлтегина слово «тоқыш-дым» обозначает высекать на камне, в каракалпакском языке слово «тоқыұ» означает вязать из шерсти, либо из хлопкового волокна. В дастане «Қырқ қыз» это слово использовано в переносном значении «кеуїлге тоқыұ» – вязать душу. В памятнике Құлтегина слово «уртым» (28, КТм) обозначает «тасқа жаздырдым» т.е. высек на камне, в дастане «Қырқ қыз» слово «урды» использовано в значении «түйреу, шаншыұ» проткнуть, уколоть: Найза менен урасаң (Коли пикой) (456).

Сопоставление лексики Орхонских памятников со словарным составом дастана «Қырқ қыз» показывает, что по прошествии веков, отдельные слова устарели и вышли из употребления, некоторые – изменили свое значение, в других произошли звуковые изменения. Однако при всем этом эти слова передают основное значение, заложенное в них, наиболее часто употребляемые из них, по прошествии веков, сохранились полностью в лексике дастана и современного каракалпакского языка.

#### ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА:

1. Қырқ қыз. Ташкент: Госиздат УзССР. 1949.
2. Сорок девушек. М-Л.: Детгиз.1952; Сорок девушек. М.: Гослитиздат. 1951; Сорок девушек. Второе исправленное издание. М.: Гослитиздат, 1956; «Қырқ қыз». Ташкент.: Уздавнашр, 1948; «Қырқ қыз» (на казахском языке). – Алма-ата.: ГИХЛ. 1959; «Қырқ қыз» (на киргизском языке). – Фрунзе.: Киргосиздат. 1960; «Қырқ қыз» (на туркменском языке). – Ашхабад. 1959; «Қырқ қыз» (на узбекском языке). – Ташкент. 1956.

3. Толстов С.П. По следом древнoхорезмской цивилизаций. – М-Л.: АН СССР, 1948, – С.2; Города гузов // Советская этнография. – 1947. – С. 99; К вопросу о происхождении каракалпакского народа// КСИЭ. – 1947. – №2. – С.72–74; Жданко Т.А. Каракалпакская эпическая поэма «Қырқ қыз» как истoрико-этнографический источник// КСИЭ. – М. – 1958. – Вып. XXX. С.113-114; Морозова А.С. К вопросу о происхождении сюжета каракалпакской поэмы «Қырқ қыз»// Труды АН Тадж. ССР. – Душанбе. – 1960. XX том; Толстова Л.С. Исторические предания южного при аралы. – М.: Наука, 1984. – С. 186-206; Иванов П.П. Қарақалпақлар тарийхының очерки (на каракалпакском языке) – Нукус.: Билим, 2021; Иванов П.П., Морозова А.С. Каракалпаки – (рукопись. 907). – Рукописный фонд КК отделения АН РУз. – Нукус.
4. Толстов С.П. По следом древне-хорезмской цивилизаций. – М-Л.: АН СССР, – 1948. – С.22.
5. Жданко Т.А. Очерки истoрической этнографии каракалпаков. -М-Л.: АН СССР. – 1950. – С. 101–102.
6. Қырқ қыз: – Нөкис.: Қаракалпақстан, 1980. – Б. 331.

УДК 81-11

**ЛИНГВИСТИЧЕСКАЯ ХАРАКТЕРИСТИКА УЧЕБНОГО  
ДИСКУРСА И ПРОБЛЕМЫ ПЕРЕВОДА  
(на английском и узбекском языках)**

***Н. Н. Абдурахманова***

*Андижанский государственный университет*

*Андижан, Узбекистан*

*n.n.abdurahmonova76@gmail.com*

В статье говорится о лингвистических особенностях учебного дискурса в общественных местах (театре и торговом центре) и проблемах перевода (на английском и узбекском языках). Дословный перевод инструктивного текста приводит к методическому маразму, говорят, что письменная инструкция не переводится буквально, а выражается так, как она дана на том языке, на который запрашивается перевод.

**Ключевые слова:** В общественных местах, оборудование, безопасность оборудования, запрещено, текст инструкции, помещения общего пользования, стилистические ошибки.

**LINGUISTIC CHARACTERISTICS OF EDUCATIONAL  
DISCOURSE AND PROBLEMS OF TRANSLATION  
(in English and Uzbek)**

***Abdurakhmanova Nargiza Nusratullaevna***

*Andijan State University*

*Andijan, Uzbekistan*

*n.n.abdurahmonova76@gmail.com*

The state talks about the linguistic features of educational discourse in the public space (theater and drama center) and the problem of translation (in English and Uzbek languages). A literal translation of an instructional text leads to methodological illness; it is said that written instructions are not translated literally, but are expressed as if it were the same language in which it is translated.

**Key words:** In public places, equipment, equipment safety, prohibited, text instructions, common areas, stylistic errors.

INSTRUKTIV DISKURSNING LINGVISTIK XUSUSIYATLARI  
VA TARJIMA MUAMMOLARI  
(ingliz va o'zbek tillarida)

*Abdurahmanova Nargiza Nusratullayevna*

*Andijon davlat universiteti*

*Andijon, O'zbekiston*

n.n.abdurahmonova76@gmail.com

Maqolada jamoat joylarida (teatr va savdo markazi) instruktiv diskursning lingvistik xususiyatlari va tarjima muammolari (ingliz va o'zbek tillarida) haqida gap boradi. Instruktiv matni so'zma-so'z tarjima qilish uslubiy g'alizlikka olib keladi, yozilgan yo'riqnoma so'zma-so'z tarjima qilinmasligi, balki tarjima talab qilinayotgan tilda berilish usulida ifodalanishi haqida so'z boradi.

**Kalit so'zlar:** texnika vositalari, texnika xavfsizligi, taqiqlanadi, yo'riqnoma matni, umumiy xizmat ko'rsatadigan xonalar, uslubiy g'alizliklar

Jamoat joylarida turli umumiy foydalanish jihozlari, umumiy xizmat ko'rsatadigan xonalar va umumiy texnika vositalaridan foydalanish bo'yicha yo'riqnoma matnlari ko'zga tashlanadi. Teatr, kutubxona va savdo markazlaridagi yo'riqnomalar shular jumlasidan.

Bu kabi yo'riqnomalarning lisoniy xususiyatlari rus va ingliz tillari misolida o'rganilgan. Jumladan, L.V.Rextinning texnika xavfsizligi qoidalari bo'yicha instruksiyalar matni xususiyatlari haqida "Заметки о речевом жанре инструкции", "Не влезай, убьет!" или "Особенности отечественного речевого жанра инструкции" maqolalari, xususan, instruksiyaning nutqiy janrlari haqidagi monografik tadqiqoti diqqatga sazovor. A.Ye.Belskaya esa rus va ingliz tili materiallari asosida instruktiv matnlarda aniqlik va qisqalikning kommunikativ me'yorga mos kelmasligini to'g'rilashga bag'ishlangan monografik tadqiqoti ham qimmatli fikrlar manbai sanaladi.

Mashina to'xtash (parkovka) joylarida quyidagi yozuv kuzatildi: *To'xtash taqiqlanadi // No parking*. O'zbek va ingliz tilidagi gap kontekstini qiyoslashdan ma'lum bo'ladiki, gaplar bir-birining aynan (so'zma-so'z) tarjimasi emas: o'zbek tilidagi gap ega va fe'l-kesimdan iborat, ingliz tilida ega+ kesim shaklidagi sodda gap emas, balki bo'lishsizlik shaklidagi gerundiy (+ing) bilan shakllangan. Bunday instruktiv matnlar odatda o'zbek va ingliz tillarida yonma-yon qo'llanadi, ammo ular o'zbek tilidan ingliz tiliga tarjima qilingan matn emas, balki ingliz lisoniy olamida qanday bo'lsa, shu shaklda ifodalangan gaplar keltirilgan. Demoqchi bo'lganimiz, o'zbek tili foydalanuvchilari yashaydigan joylarda o'zbek tilida yozib, ingliz tiliga tarjima qilinma-



gan, balki ingliz tilida shu ma'noni beruvchi ifoda usuli keltirilgan. Chunki instruksiya o'zbek tili tashuvchisiga emas, ingliz tili egalari o'qishi uchun yozilgan. Agar "To 'xtash taqiqlanadi" gapi so'zma-so'z ingliz tiliga tarjima qilinib, instruksiya sifatida berilsa, "Parking is prohibited" bo'lishi lozim. Ammo ingliz lisoniy olamida "Bu yerda parkovka qilish mumkin emas", "bu joyda mashina qo'yib bo'lmaydi" mazmuni No parking sifatida ifodalanadi. Bu ifoda vositasining qanday shaklda ishlatilishini ingliz tili korpusida kuzatamiz va tahlil qilamiz:

1. *That is why you would often see people parking in no parking areas, buying pirated movies without regrets, downloading music and software illegally and so on.*

2. *Parking: parking vehicles is only allowed in the parking spots at the camping compound only, no parking allowed or raising tents near the resort hoses or the village resort lawns.*

3. *No parking spaces are available at the venue; and 4.*

4. *There was no No Parking signage or yellow line(s) .*

Yuqorida ayni bir korpusdan olingan misollar va tasvir shuni ko'rsatadiki, "bu yerda to'xtash taqiqlangan" ma'nosini ifodalash uchun "No parking" ifodasi ishlatiladi. Ingliz tili qo'llanishda bo'lgan boshqa hududlarda shu ma'noni ifodalovchi yozuvlarni qidirganimizda No parking yozilgan instruksiyaga duch keldik. Demak, ingliz lisoniy olamida shu mazmundagi instruktiv matn "Parking is prohibited" emas, No parking shaklida uchraydi .

Toshkent shahri savdo markazidagi boshqa tasvirda rus tilida yozilib, keyin o'zbek tiliga tarjimasi berilgan yo'riqnoma matni ham uchraydi .

Rus tilida: *Уважаемые гости! В нашем ТРЦ много посетителей с колясками, а также людей пожилого возраста. Просим Вас оставлять лифты для них и пользоваться эскалаторами.*

O'zbek tilida: *Hurmatli mehmonlar! Bizning markazimizda yoshi katta va aravachalik yo'lovchilar ko'p. Iltimos, ularga liftni ishlatishga imkoniyat bering va eskalatoridan foydalaning!*

Ko'rinadiki, instruktiv matnni so'zma-so'z tarjima qilish uslubiy g'alizlikka olib kelgan: shu o'rinda yoshi katta va aravachalik yo'lovchilar birikmasi uslubiy xatolikka olib kelgan, bunda tarjimadagi xatolik bilan birga aravachalik so'zining qo'llanishi so'z qo'llash bilan bog'liq xatoni keltirib chiqargan. Bunday holatda birinchi tilda yozilgan yo'riqnoma so'zma-so'z tarjima qilinmasligi, balki tarjima talab qilinayotgan tilda berilish usulida ifodalanishi maqsadga muvofiq.

Instruktiv matnlarning uch til: o‘zbek, ingliz, rus tilida berilishi holati ham kuzatiladi.

*Chekish mumkin emas! No smoking! Курить запрещается!*

*Tozalikka rioya qiling! Keep clean! Соблюдайте чистоту!*

*Не бросайте мусор в унитаз! Axlatni unitazga tashlamang! Don't throw trash into the toilet .*

“*Chekish mumkin emas!*”, “*No smoking!*”, “*Курить запрещается!*” kabi uch tildagi gap asliyatdagi tilda qanday usulda berilsa, shu shaklda ifoda etilgan: so‘zma-so‘z tarjima qilinmagan.

Chekish mumkin emas gap so‘zma-so‘z va o‘sha strukturada tarjima qilinganida “Smoking is not allowed” bo‘ladi, ammo ko‘rinib turganidek, *No smoking* sifatida berilgan. Rus tiliga tarjimasi ham shunday: Курение запрещено bo‘ladi, ammo Курить запрещается o‘zbek tilida “*Chekish ta‘qiqlangan*” bo‘lishi kerak.

*Chekish mumkin emas!* (asliyatda berilishi)      *Chekish mumkin emas!* (So‘zma-so‘z tarjimada)

*No smoking! Smoking is not allowed Курить запрещается! Курение запрещено*

Bu uch misolni berilgani (asliyatdagisi) va so‘zma-so‘z tarjimasini qiyoslash shuni ko‘rsatadiki, instruktiv matnlar uch tilda shu mazmuni qanday ifodalash kerak bo‘lsa, shunday ifoda etilgan. Ingliz va rus tili korpuslarida bu gaplarning ishlatilish statistikasi ham buni isbotlaydi:

Mazkur gaplarning strukturasi va grammatikasini qiyoslash shuni ko‘rsatadiki, turli tillarda ayni ma‘noni ifodalashda turli grammatik qurilishli gaplardan foydalanilgan.

*Внимание! Опасность поражения электрическим током. Диққат! Электр токи уриш хавфидан эҳтиёт бо‘линг .*

Odatda, jamoat joylarida xavfsizlik bilan bog‘liq instruktiv matnlarning mazmunida e‘tiborni tortish, ehtiyotkorlikka chorlash mazmuni ustunlik qiladi. Ko‘pchilik doimiy foydalanadigan, ehtiyot bo‘lish talab qilinadigan jamoat joylariga eskolator va liftni misol keltirish mumkin. Bunday vositalardan foydalanish haqidagi instruktiv matnlarning o‘ziga xosligini quyidagi keltirilgan misollar vositasida o‘rganamiz.

### **Yo‘lovchi liftidan foydalanish qoidalari**

1. Lift kabinasini chaqirishda sizga kerakli bo‘lgan yo‘nalish tugmasini bosing.

2. Kabina qavatga yaqinlashganida ovozli signal chalinadi chiroqli ko‘rsatkich kabinaning keyingi yo‘nalishini ko‘rsatadi.

3. *Kabina va shaxta eshiklari avtomatik ravishda ochiladi.*

4. *Kabinaga kirgandan so'ng kerakli qavat tugmasini bosib.*

5. *Kabinaga berilgan harakat yo'nalishini tugmani bosib o'zgartirib bo'lmaydi.*

6. *Lift ishlamay qolsa, "telefon" tugmasini bosib va operatorga bu haqida xabar bering. Agar dispatcher javob bermasa, "trevoga" (tashvish) tugmasini bosib.*

*Chekish, tez yonuvchan, portlash xavfi bo'lgan moddalar, shuningdek, katta hajmli buyumlarni tashish.*

*Kabina to'liq to'xtashidan oldin kirish va chiqish. 13 kishidan ko'p yo'lovchi tashish.*

*Shaxta va kabina eshiklari yopilishini to'xtatib turish. Kabina va shaxta eshiklarini qo'llar yordamida ochish.*

Mazkur instruktiv matnga xos terminologiya sirasiga lift, kabina, shaxta, yo'nalish tugmasi, qavat, tugma kabi birliklarni ko'rsatish mumkin. Ushbu yo'riqnomaning sintaktik xususiyatlari sifatida sodda gapdan iborat bo'lish, harakat nomi bilan keluvchi sanaladigan qismlarning mavjudligi ko'rinadi. Masalan: "*TAQIQLANADI*" sanaladigan qismlar berilgan:

- Chekish, tez yonuvchan, portlash xavfi bo'lgan moddalar, shuningdek, katta hajmli buyumlarni tashish.

- Kabina to'liq to'xtashidan oldin kirish va chiqish.

- 13 kishidan ko'p yo'lovchi tashish.

- Shaxta va kabina eshiklari yopilishini to'xtatib turish.

- Kabina va shaxta eshiklarini qo'llar yordamida ochish.

Bunday qismlar temiryo'l xizmatidan foydalanish bo'yicha yo'riqnomalarda ham ko'p kuzatildi.

- Yo'lovchi tashish temiryo'l transportida taqiqlangan xavfli moddalar va jismlar ro'yxati:

- O'qotar qurollarning barcha turlari

- Portlovchi moddalar, portlatish vositalari va ular bilan to'ldirilgan buyumlar

- Tez alanga oluvchi suyuqliklar

- Zaharli va zaharlovchi moddalar

- O'tkir hidli va yemiruvchi moddalar

- Radioaktiv moddalar

- Giyohvand moddalar

- Boshqa noqonuniy xatti-harakatlar sodir etish uchun ishlatilishi mumkin bo'lgan modda va jismlar.

Mazkur ro'yxat ayni tabloda o'zbek va rus tillarida keltirilgan bo'lib, o'zbek tilidagi ro'yxatning rus tilidan tarjima ekanligi ko'rinib turibdi. Tarjimada esa bir necha uslubiy g'alizliklar uchraydi. Masalan, rus tilidagi ядовитые и отравляющие вещества birikmasi o'zbek tiliga zaharli va zaharlovchi moddalar deb tarjima qilingan. Zaharli va zaharlovchi so'zlari sinonim bo'lib, teng bog'lanishli bir so'z qo'shilmasida ularning yonma-yon ishlatilishi uslubiy xatolikni keltirib chiqaradi. Bunday vaziyatda (boshqa tildan tarjima qilinganda) faqat so'zma-so'z tarjima qilish yaxshi natija bermaydi, mazkur so'z qo'shilmasi ifoda usuli aniq bo'lish uchun transformatsiya vositasida adekvat tarjima qilish maqsadga muvofiq. Masalan: zaharli moddalar tarzida.

Shuningdek, другие опасные вещества, предметы и грузы, которые могут быть использованы для совершения акта незаконного вмешательства murakkab birikmasi boshqa noqonuniy xatti-harakatlar sodir etish uchun ishlatilishi mumkin bo'lgan modda va jismlar tarzida tarjima qilingan.

O'zbek tilida (tarjimada) boshqa noqonuniy xatti-harakatlar sodir etish uchun ishlatilishi mumkin bo'lgan modda va jismlar

Rus tilida (asliyatda) другие опасные вещества, предметы и грузы, которые могут быть использованы для совершения акта незаконного вмешательства

Rus tilidagi *вещества, предметы и грузы* teng bog'lanishli birikmasi o'zbek tiliga modda va jismlar sifatida tarjima qilingan bo'lib, ekvivalentlik buzilgan. Turli modda, predmet va yuklar deb so'zma-so'z tarjima qilinganda matnning mazmunidagi aniqlik ta'minlanardi. Ko'rinadiki, ayrim o'rinda so'zma-so'z tarjimaga, ayrim vaqtda transformativ tarjimaga, ya'ni adekvatlikka asoslangan tarjima maqbul sanaladi.

Xulosa sifatida aytish mumkinki, jamoat joylaridagi instruktiv diskursning mazmuniy va struktur tuzilishi o'ziga xos: sintaktik jihatdan egasiz gaplarning qo'llanishi, fikr ifodasida qisqalik va lo'ndalik shartiga amal qilish, o'ziga xos terminologiya va morfologik belgilarga ega bo'ladi.

#### ADABIYOTLAR:

1. Kanashuk S.A. Instruktivniy diskurs IT korporatsiy: sotsiologivsticheskiy aspekt: dissertatsiya na soiskanie uchenoy stepeni kandidata filologicheskix nauk : 10.02.01 / Kanashuk Sergey Aleksandrovich; nauch. ruk. N.A.Mishankina; Tom. gos. un-t. – Tomsk, 2012: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000432820>

2. Galperin I.R. *Tekst kak ob'ekt lingvisticheskogo issledovaniya*. M., 1981.
3. Sirotinina O.B. *Nekotore razmishleniya po povodu terminov «rechevoy janr» i « ritoricheskiy janr» // Janri rechi – 2. Saratov, 1999.*
4. Stepanov Yu.S. *Alternativniy mir, diskurs, fakt i prinsipi prichinnosti // Yazik i nauka konsa XX veka. – M., 1995. – Str. 35-73.*
5. Robert, P. *Le nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française. – P.: Dictionnaires le Robert, 2002. – P. 735.*
6. Xoroxordina O.V. *Instruksiya kak tip teksta / Mir russkogo slovo. – № 4 / 2013. – Str. 7–14.*
7. Kondrashkina Ye.Yu. *K voprosu o tekstosvyazuyushix kategoriyax v instruktivnom tekste / Mir russkogo slova. № 3 / 2018. – Str. 27-32.*
8. Nosova L.N. *Kommunikativno-pragmaticheskiy potentsial instruksii po primeneniyu lekarstvennykh sredstv v farmatsevticheskom diskurse: avtoref. diss. kand. filol. nauk. – M., 2013.*
9. Kasavin I.T., Щавелев S.P. *Analiz povsednevnosti. – M.: Kanon, 2004.*
10. Presnuxina I.A. *Delovoe obshenie v svete diatopicheskogo varirovaniya sovremennogo angliyskogo yazyka: diss. kand. filol. nauk. – M., 2005.*

УДК 81-11

**МЕСТО ЛИНГВИСКОКУЛЬТУРНЫХ ЕДИНИЦ,  
РЕПРЕЗЕНТИРУЮЩИХ УЗБЕКСКИЙ МЕНТАЛИТЕТ,  
В ХУДОЖЕСТВЕННОМ ТЕКСТЕ  
(На примере былин поэта Эркина Вахидова)**

*Н. И. Адизова*

*Бухарский государственный университет*

*Бухара, Узбекистан*

*nilufar.adizova71 @ gmail. com*

Каждый язык имеет свое значение, поскольку воплощает в себе древние ценности определенного народа, менталитет, важные аспекты истории традиций и обычаев. Единицы выражения, считающиеся культурным символом в одном народе, просто проявляются как средство общения для другого. В данной статье анализируются лингвокультуры, используемые для конкретных целей, как важный инструмент выражения образа мышления художника, образа нации, в том числе важная роль безальтернативной лексики как языковой единицы, отражающей менталитет Узбекский народ анализируется на примере эпосов Эркина Вахидова.

**Ключевые слова:** язык, культура, лингвокультурология, культурный феномен, национальный менталитет, антропоцентризм, объединение, духовный образ, национальный дух.

**THE PLACE OF LINGUISTIC-CULTURAL UNITS  
REPRESENTING THE UZBEK MENTALITY IN  
THE ARTISTIC TEXT  
(On the example of the epics of the poet Erkin Vahidov)**

*Adizova Nilufar Istamovna*

*Bukhara State University*

*Bukhara, Uzbekistan*

*nilufar.adizova71 @ gmail. com*

Each language has its own significance as it embodies the ancient values of ascertain nation, the mentality of important aspects of the history of traditions and customs. Units of expression that are considered a cultural symbol in one nation are simply manifested as a means of communication for another. In this article, lingvocultures used for specific purposes are analyzed as an important tool for expressing the artist's way of thinking, the image of the nation, including the important role of non-alternative lexicon as language units reflecting the mentality of the Uzbek nation is analyzed on the example Erkin Vahidov's epics.

**Keywords:** language, culture, lingvoculturology, cultural phenomenon, national mentality, anthropocentric, association, spiritual image, national spirit.

**O'ZBEK MENTALITETINING ADABIY MATNDAGI TIL-  
MADANIY BIRLIKLARNING O'RNI  
(Shoir Erkin Vohidov dostonlari misolida)**

*Adizova Nilufar Istamovna*  
*Buxoro davlat universiteti*  
*Buxoro, O'zbekiston*  
nilufar.adizova71@gmail.com

Har bir til o'ziga xos ma'noga ega, chunki u ma'lum bir xalqning qadimiy qadriyatlarini, mentalitetini, an'ana va urf-odatlar tarixining muhim jihatlarini o'zida mujassam etgan. Bir xalqda madaniy timsol hisoblangan ifoda birliklari boshqa xalq uchun oddiygina muloqot vositasi sifatida namoyon bo'ladi. Ushbu maqolada san'atkorning fikrlash tarzini, millat qiyofasini ifodalashda muhim vosita sifatida muayyan maqsadlarda qo'llaniladigan til madaniyatlari tahlil qilinadi. jumladan, muqobil lug'atning o'zbek xalqi mentalitetini aks ettiruvchi til birligi sifatidagi muhim ahamiyati Erkin Vohidov dostonlari misolida tahlil qilingan.

**Tayanch so'zlar:** til, madaniyat, lingvokulturologiya, madaniy hodisa, milliy mentalitet, antropotsentrizm, unifikatsiya, ma'naviy qiyofa, milliy ruh.

Linguistics is one of the important fields that are developing today. Its object appears in the process of interaction between culture and language, at the junction of linguistic and speech units as a sign of culture. He studies language as a cultural phenomenon, carrier. Culture is created by people who use language. In some nations, standards that express the usual norms and cultural characteristics may not correspond to the mentality of another nation. That is why the expression units that are considered as a cultural symbol in one nation are simply a means of communication for another. After all, each language acquires its own importance by embodying important aspects related to the ancient values, traditions, customs, and history of a certain nation. Yes, "...the way of thinking, lifestyle, and spiritual views of any people or nation are not formed by themselves, in a vacuum... If we look at the way of life and thinking of our people, it is unlike any other, formed over thousands of years, not only mutual interaction, but also we observe a number of specific features that appear as an integral part of our life.

For example, let's take the pairs of units in our language that enrich and complement each other with deep meaning, such as kindness, affection, kindness, and dignity. No matter how strange it may seem, the exact translation of these units into other languages is a difficult problem..." [ 1; 8]

In fact, there are such expressions and units reflected in our language, which are not just a beautiful means of speech, but a linguistic or verbal expression of the values and traditions of our nation, which have arisen as the basis of the age-old views and spiritual life of our people, are imprinted in our imagination, and are characteristic of our nation. In particular, linguoculturalisms that manifest the image of the nation, the identity and the artistic way of thinking of the creator, which have arisen at the heart of the methodological possibilities of the language in artistic texts – some lexical units, mythologems and archetypes, paremiological funds, phraseological units, symbols, similes, metaphors that express the national mentality (metaphor, metonymy, synecdoche...), precedent units, speech labels, periphrasis, etc. Below, we observe some lexical units reflecting the views and national image of the Uzbek nation based on examples:

Guzar. Passage; street, neighborhood. A prosperous, busy place with shops such as teahouses, butchers, grocers, located at the crossroads of villages or neighborhoods. In the Khanate of Bukhara: neighborhood. [2; 516].

This unity expresses the meaning of a crowded place, a neighborhood with unique values of the Uzbek people, which has the importance of uniting everyone, and shows the national-cultural spirit and solidarity. Through these verses, we can see the hint that there will be many people in the guzar and that the word will spread quickly:

Guzar. Long ears Man is a fool.

Caution!

Here

Trust. (From “Epic written on a board”)

Mulla. An educated, pious man. A person who received madrasa education in the countries of the Middle East and Central Asia. Religious figure, scholar. [2; 634].

In our nation, mullahs are valued as knowledgeable, considerate people who lead people on the right path, and their advice is listened to. Mulla urges people to be alert and patient based on the fact that tragedies do not happen in vain. Through this unity, the author also shows the mullahs as patient, intelligent people who bring people together:

A mullah swallowed Says the yuvvosh:

“It’s our fault, People are barren.

He does not know faith

Neither old nor young. (From “Epic written on a board”)



Home is a unit that unites the whole family for the Uzbek nation, representing a place where the memories of ancestors, traditions, and values are concentrated. In the author's epics, "home" is represented as a symbol of abundance, calmness, unity, and kindness. As a result of the terrible earthquake that occurred in Tashkent in 1966, many people became homeless. However, the Uzbek people are tolerant and do not allow anyone to stay on the street. In these verses, the author refers to:

It's not about housing, Long live.

But at such a time A known person...

In the Uzbek people, the tradition of doing something with help still continues as a value: The whole nation was stirred up,

Both a duty and a duty.

... He calls his friends

Hashar city. (From "Epic written on a board")

The terrible earthquake event that drives the plot of "Epic written on a board" did not bend the will of the nation, which was found in a thousand and one times of history. The poet Zucco expresses the patience, fortitude, strong will, cheerfulness, hope for the future embodied in the consciousness of our people through units such as tea house, green tea, askiya, anecdote, tanbur, tor, chirmanda, Navruz. These units represent the national identity, cultural life and openness of the Uzbek people. Including

It's been a terrible day Black page.

Students today An anecdote... Askiya Payrov Newest topic,

Again in the cafeteria Green tea and laughter...

He uses the concepts of "shepherd" (a shepherd) and "scarf" (light headwear for women) for the national image and unique expression of the characteristics of the Uzbek nation, such as humility, hard work, sincerity, loyalty, modesty, fortitude, and simplicity:

Still killing at one end, He didn't join in the fun Simple guy -

The poor shepherd... then,

A white scarf is on the head

A charming girl went into the water. (From the epic "Fountain of Dreams")

It is known that the cap has been valued as the national headwear of our people since ancient times. By using the word "doppi", the creator focuses not only on the external appearance of the hero, but also on showing his inner image (a person who embodies faith, education, and values):

A young clerk in a modern suit, tie, and a cap enters through the front door. (From the epic “Throne and Happiness”)

The author skillfully connects the important aspects of the Uzbek people with mythological views. A mythologeme can be a myth’s “protagonist, its important character, or its situation. It is basically based on an archetype. (“An archetype is a stable image that appears in people’s minds and spreads in culture.”).[4; 26.]

For example, bread is kissed and applied to the eyes or forehead as a blessing that has been appreciated by the Uzbek people since ancient times. It became a custom to leave a bite of bread before going to a distant land, sprinkle it to the birds, and pick it up immediately without pressing it when it falls on the ground. In the work of Erkin Vahidov, the archetype of bread, which our people have long cherished, admired, and source of sustenance and symbol of honesty, which cannot be compared with anything, is often found:

The alphabet and a piece of bread in my blue book,

I took my first step towards the school. (from the epic “Shout”)  
You write, bread comes from a pure soum,

The dirty million has blood fever. (From the epic “Throne and Happiness”)

Laughter also reflects the national-cultural image of the Uzbek nation. In Uzbek families, older people usually consider it rude to laugh out loud, or consider it “a state that does not bring good.” There are two approaches to laughter on the basis of mythologies: in fairy tales and myths, laughter is used to give additional color to the image of giants, beggars, cannibals, negative characters or the helpless state of the hero, while silent laughter, smiling is considered a sign of greatness characteristic of prophets and saints. In the epics of Erkin Vahidov, we can witness the use of “kah-kah kulgi” to reflect the helpless and unhappy state of the lyrical hero:

Hugging the unfortunate body, She washed her face with tears.  
Then suddenly He beat

He threw himself into the waves. (From the epic “Fountain of Dreams”) Unaware of the sorrows of the world,

Laughing and laughing

Lightning strikes. (from the epic “Shout”)

Units such as “Soil”, “Mother Earth”, “Earth”, “Zamin” also reflect the unique views and past of our people as archetypal images. Indeed, today’s peaceful and prosperous days were not easily achieved. In our people, taking a piece of soil from the land before riding to a dis-

tant place has a hidden symbolic meaning and is formed on the basis of mythological views. The poet uses units such as “Soil”, “Mother Earth”, “Yer”, “Zamin” instead of the concept of “Motherland”, which is repeated many times in the text of the epic, and achieves an increase in effectiveness:

Don't be surprised again From the soil

If you find a rusty helmet, There is a humming sound

If you listen to the heart of the earth. (From “Epic written on a board”) Mother Earth again

One day to live... (from the epic “Nido”) The Holy Land remains for you.

You keep it carefully in the palm of your hand... Hear, Mother Earth is moaning

From the pain of the crutches,

This is the heart of the earth. (from the epic “Shout”)

Erkin Vahidov's epic “The Fountain of Dreams” begins with the narrator's “Tale of the Black Sea” and the events in it mainly take place by the sea, water, and springs: The events begin with a simple shepherd telling a tragic love story in a remote corner of Crimea. The author connects the content of events to concepts such as “spring”, “spring”, “sea” based on mythological views. Through them, the poet sometimes shows signs of clarity, innocence, calmness, sympathy, sometimes reflects the unfortunate, poor, helpless situation of the hero, and the events that set the plot in motion related to the experiences of the lyrical hero are reflected more vividly:

Think sweet thoughts

He was killing under the eyes of Chashma... Softly gurgling waters

Soylar edi A story for a girl... (from the epic “Dream Fountain”)

...He came, Dry spring -

Tears flowed from his eyes...

... The blue sea comes into the language. Tell me a sad story

Hits the rocks... (from the epic “Fountain of Dreams”)

The poet reflects the notions of elegance, grace, beauty, spiritual and cultural views, which are valuable to our people, through the concept of “mermaid”. He uses it as a symbol of beauty, to embody supernatural and miraculous situations, and as a factor to reflect aspects typical of legends:

Months passed Like a running deer

Spring came after winter.

One day at night

Baby in hand

A mermaid came out of the sea. (From the epic “Fountain of Dreams”)

In general, it can be seen from the above examples that each speech unit used in the author’s epics acquires its own importance in the poet’s creative maturity as an important factor in awakening the national-cultural spirit. Knowledge of the science of language and speech allows the artist to create high-level artistic works, to create word sparks, to create colorful and unique expressions of movements, and most importantly, to become the owner of linguistic and artistic thinking.

#### REFERENCES:

1. Karimov I.A. High spirituality is an invincible force. – Tashkent: Spirituality, 2008. – В. 8.
2. An explanatory dictionary
3. of the Uzbek language. 5 volumes. – Tashkent: National Encyclopedia of Uzbekistan, 2006.
4. Vahidov E. Epics. – Tashkent: G’. Publishing house of literature and art named after Ghulam, 1973.
5. Usmanova Sh. Linguistic culture. Course of lectures. –Tashkent: University-2014.
6. DN Yuldasheva. BASIC NON-VERBAL COMPONENTS OF SPEECH. CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (2767–3758) 3 (01), 17–24. <https://masterjournals.com/index.php/crjps/article/view/661>
7. Юлдашева Д.Н. РОЛЬ МОЛЧАНИЯ В РЕЧИ// ОБРАЗОВАНИЕ И ИННОВАЦИОННЫЕ ИССЛЕДОВАНИЯ. № 1 (2020): <https://inter-science.uz/index.php/home/article/view/31>
8. Istamovna, Adizova Nilufar. “METHODODOLOGICAL POSSIBILITY OF INTERTEXTUALITY IN ARTISTIC TEXT.” INNOVATION IN THE MODERN EDUCATION SYSTEM 3.31 (2023): 182–191.
9. Istamovna, Adizova Nilufar. “The role of lexical units in expressing the spirit of national identity (based on the epics of Erkin Vahidov).” Conferences. 2021.
10. Adizova, Nilufar. “THE USE OF PHRASEOLOGICAL UNITS IN ERKIN VAHIDOV’S EPIC.” JOURNAL OF PHILOLOGY HORIZONS 6.6 (2021).

УДК 81-11

**КОММЕНТАРИИ К ЖАНРАМ, СВЯЗАННЫМ  
СО СКАЗКАМИ*****О. И. Адизова****Бухарский государственный университет**Бухаро, Узбекистон**obodon63@gmail.com*

В статье содержатся сведения о роли жанра сказки в фольклоре и его развитии. Изучались характерные для фольклора понятия традиции, обычаи, ценности и национальности, а также учитывалось отношение к ним.

**Ключевые слова:** фольклорист, фольклор, персонаж, традиция, исторический, этнография.

**COMMENTS ON GENRES RELATED TO FAIRY TALES*****Adizova Obodon Istamovna****Bukhara State University**Bukhara, Uzbekistan**obodon63@gmail.com*

The article contains information about the role of the fairy tale genre in folklore and its development. The concepts of tradition, customs, values and nationality characteristic of folklore were studied, and attitudes towards them were also taken into account.

**Key words:** folklorist, folklore, character, tradition, historical, ethnography.

**ERTAKLAR BILAN BOG‘LIQ JANRLAR HAQIDA  
MULOHAZALAR*****Adizova Obodon Istamovna****Buxoro davlat universiteti**Buxoro, O‘zbekiston**obodon63@gmail.com*

Maqolada xalq og‘zaki ijodida ertak janrining o‘rni va uning taraqqiyoti haqida ma’lumotlar keltirilgan. Folklorga xos an’ana, urf-odat, qadriyat va milliylik tushunchalari o‘rganilgan hamda ularga nisbatan munosabat ham e’tiborga olingan.

**Tayanch so‘zlar:** folklorshunos, xalq og‘zaki ijodi, xarakter, urf-odat, tarixiy, etnografiya

Ertaklar xalq ogʻzaki ijodining eng qadimiy janrlaridan biri boʻlib, ular yaratilishi jihatdan qadimiyligi, rang-barangligi, gʻoyaviy-badiiy yuksakligi, dunyo xalqlari adabiy merosiga munosib hissa boʻlib qoʻshiluvchi ekanligi bilan alohida ajralib turadi. Ertak umumfolklor hodisasi sifatida epik janrlardan boʻlib, M.Koshgʻariy oʻz «Devonu lugʻotit turk» asarida turkiy xalqlarda uning «etuk» atamasi bilan yuritilganini qayd etadi. Unga koʻra, etuk «biror voqeani ogʻzaki hikoya qilish»ni anglatadi.[1;12] «Ertak» soʻzi «er», aslida «ir» («yir», jir) soʻziga oʻxshatishni bildiruvchi «-tak» qoʻshimchasining qoshilishidan tarkib topgan boʻlib, qoʻshiqqa oʻxshash degan maʼnoni anglatadi. Negaki, ertaklarning boshlanmasi qoʻshiqqa oʻxshab ketadi. Shuningdek, baʼzi ertaklar tarkibida sheʼriy qismlar ham uchrab turadi. Lekin ertakda sheʼriy parchalar uchrashi odatiy hol emas. Shu sababli ertak xalq nasri namunasi hisoblanadi.

Qadimiy ajdodlarimiz tomonidan yaratilgan bu ogʻzaki badiiy meros mumtoz adabiyot, meʼmorchilik sanʼati va musiqa asarlari kabi bebahodir. Ularning mazmuni negizida soddalik va ayyorlik, ezgulik va yovuzlik, toʻgʻrilik va egrilik, muhabbat va nafrat, boylik va kambagʻallik kabi jahon xalqlari hayotiga xos boʻlgan umuminsoniy gʻoyalar yotadi.

Ertaklar janr sifatida uzoq muddatli shakllanish jarayonidan oʻtib kelgan. Ular ibtidoiy odamlarning turmushdagi biror voqeani oddiygina hikoya qilishlari asosida yuzaga kelgan. Davrlar oʻtishi bilan hikoya qilish ham takomillashib borgan. Soʻzga sigʻinish, ilohiy kuchlarga sigʻinish, animistik, totemistik, fetishistik eʼtiqodlar, gallisyunasiya va tush taʼsirida fantastik vositalarga toʻlisha borsa, hayvonlarni ovlash, xonakilashtirish, ular inonchlariga ishonish, hayvon mahsulotlaridan-gina emas, balki kuchidan ham foydalanish jarayonlarida hayvonlar haqidagi ertaklar paydo boʻla boshladi. Asta-sekin shu xildagi ertaklarda turmush tajribasini omuxtalashtira borish, u yoxud bu xildagi qusur va kamchiliklardan kulish ertakdagi obrazlarga majoziylik (allegorik) xususiyatni baxsh etdi.

Ertaklar oʻz janriy tabiatiga xos badiiy-kompozitsion qurilishga egaligi bilan alohida ajralib turadi. Chunki ular bir xil badiiy shakliy qoliplar doirasida yaratiladi va ijro etiladi. Kirish, boshlama, tugun, epik sarguzasht va tugallama ertak kompozitsion qurilmasining asosini tashkil etadi.[2;87] Ertaklar xalq ogʻzaki badiiy ijodiyotining epik turiga mansub boʻlsa-da, kirish qismi alohida boshlamaga asoslanishi bilan oʻziga xoslikka ega. Ayniqsa, unga xos qofiyadoshligi, anʼanaviyligi bilan alohida eʼtiborni tortadi. Masalan, “Bir bor ekan,

bir yo‘q ekan, bo‘ri bakovul ekan, tulki yasovul ekan, g‘oz karnaychi ekan, o‘rdak surnaychi ekan, toshbaqa tarozidor ekan, qurbaqa undan qarzdor ekan”. Bu parcha qofiyali ifodaga misol bo‘la oladi. [3;19]

Ertak og‘zaki milliy nasr namunasi sifatida o‘zining ifoda va bayon uslubiga egaligini asrlar davomida saqlab kelmoqda. Yozma adabiyotdagi nasr namunalarida uchraydigan ertaklar avvalo, an‘anaviyligi, bir ertakdan boshqasiga ko‘chib yurishi, “sayyor ekani” bilan ajralib turadi.

Yozma adabiyotda adiblar tomonidan yaratilgan asarlarda ma‘lum bir muallifning individual mahorati aks etsa, ertaklarda esa xalq topqirligi va zakovati namoyon bo‘ladi. Ular orqali butun boshli bir millat, xalq adabiyotining badiiyati baholanadi. Ertaklar og‘izdan-og‘izga o‘tib, sayqal topgan holda avloddan –avlodga etkazilar ekan, bu jarayonda ular syujetidagi ayrim motivlar, obrazlar tushib qolishi yoki akincha ko‘payib borishi tabiiy. Hatto goho bir xalq ertagiga xos syujet motivlari boshqa bir xalq ertagiga ko‘chib o‘tishi va buning oqibatida bir necha o‘xshash syujetli ertak paydo bo‘lishi mumkin. Folklorshunoslikda bunday ertaklar- sayyor syujetli ertaklar deb yuritiladi. Sayyor syujetli ertaklar ko‘pincha mazmunan o‘zaro o‘xshasalar-da, biroq milliy ruh, milliy urf-odat, milliy dunyoqarash, milliy tilga xos ifodaviy vositalar va geografik muhit tasviriga ko‘ra farqlanuvchi mustaqil adabiy –og‘zaki hodisa sifatida yashash huquqiga ega. Ertaklar xalq og‘zaki badiiy ijodiyotining epik turiga mansub bo‘lib, o‘ziga xos g‘oyaviy – mavzuviy yo‘nalishga, axloqiy – ta‘limiy va ijtimoiy – estetik vazifalarga ega. Ertaklar ham og‘zaki tarzda jamoa ijodi mahsuli sifatida anonim ko‘rinishda yaratiladi. Garchi uning to‘qilish ibtidosi individual ijodkorga borib taqalsa-da, og‘izdan – og‘izga, urug‘dan-urug‘ga, avloddan-avlodga o‘tish jarayonida dastlabki ijrochisiga xos belgilarini, obrazlarini, motivlari va badiiy vositalarini deyarli saqlab qoladi.[4;39]

Har bir xalqning ertagi o‘sha xalq tarixini, ma‘naviy-madaniy turmush tarzini, ichki dunyosini, imon-e‘tiqodini boshqa qardosh elu-ellatlar bilan ijtimoiy munosabatlarini, urf-odatlarini, yashash joyining iqlimi va tabiiy shart-sharoitlarini o‘rganishda muhim manba vazifasini o‘taydi.

Deyarli barcha ertaklarning g‘oyaviy yo‘nalishi yagona maqsadga-mehnat ahlining buyuk va yorqin kelajak uchun olib borgan kurashlarini, intilish va orzularini aks ettirishga qaratilgan. Shuning uchun ertaklar hamisha yaxshilik va murodga etishdan iborat umidbaxsh g‘oya bilan yakun topadi.

Ertaklar o'z janriy tabiatiga xos badiiy-kompozision qurilishga ega. Ular bir xil badiiy shakliy qoliplar doirasida yaratiladi va ijro etiladi. Kirish, boshlama, tugun, epik sarguzasht va tugallama ertak kompozision qurilmasining asosini tashkil etadi.

Ertaklarning an'anaviy kirish bilan boshlanishi jahondagi barcha xalqlar ertakchiligi uchun mushtarak xususiyat hisoblanadi. An'anaviy kirishning vazifasi tinglovchilar e'tiborini bir nuqtaga jalb etish, ertak tinglashga hozirlashdir. An'anaviy kirishlar syujet tabiatiga mos tushadigan xayoliy fon yaratishni, auditoriya, tinglovchilar qalbida ko'tarinki ruh, xushchaqchaq kayfiyat paydo qilishni ko'zlaydi.

Odatda, an'anaviy kirishlar qofiyalangan nasriy parchalar shaklida bo'lib, zamon va makon haqidagi ma'lumotlarni ifoda etadi. Ayrim hollarda kirish qism gina ertak boshlamasiga ulansa, ayrim hollarda o'zicha mustaqil holda qolaveradi. Ertakda boshlama syujet chizig'ida doimiy element hisoblansa-da, syujet rivojiga aloqador emas, balki harakatning tashkil topishiga turtkki bermaydi. U qahramonlar haqida ma'lumot beradi. Syujet chizig'ida sodir bo'ladigan voqea-hodisalarning qaysi zamon va makonda kechajagidan xabar beradi. Boshlama syujet yo'nalishida noaniqlik va umumiylik kasb etadi. «Bor ekanda, yo'q ekan. Och ekanda, to'q ekan. qadim zamonda bir podsho o'tgan ekan» kabi. Ertak boshlamalari ko'rinishiga ko'ra uzun yo qisqa bo'lib, voqea-hodisalarning kelajak o'rni va vaqti haqida noaniq, umumiy va mavhum ma'lumot beradi. Ba'zan boshlama ertak voqeligi bilan qo'shib yuborilishi ham kuzatiladi: «O'tgan zamonda uch aka-uka botirlar bo'lib, ular o'z omadlarini sinab ko'rish niyatida uzoq safarga yo'lga chiqibdilar». Bunday holatdagi boshlamalar ertakdagi voqealar rivojining tezlashishiga va keskinlashuviga, ziddiyatlarning kuchayishiga zamin hozirlaydi. [5;142] Shuni alohida ta'kidlash joizki, boshlama ekspozisiya emas, balki unga –ekspozisiyaga yo'l ochadi, o'rin beradi. Boshlama syujetga daxldor bo'lmay, ertakda syujet ekspozisiyadan boshlanadi. Ekspozisiya esa bevosita asarning tuguni bilan aloqador; syujet chizig'ini tashkil etuvchi hodisalar harakati, konflikt va voqealar rivoji ekspozisiyadan boshlanadi: «U podshoning yolg'iz bir o'g'li bor ekan, -yillar o'tib, podshoning o'g'li ulg'ayibdi. Bir kuni u tushida bir qizni ko'rib, oshiq bo'lib qolibdi». Ayonlashayotirki, podshoning yolg'iz o'g'li borligi aytilishi syujetga harakat baxsh etsa, tushida qizni ko'rishi-tugunni hosil qilgan. Ayni choqda syujet rivojiga – epik sarguzashtga asos bo'luvchi konfliktga ishoraga aylangan.

Ekspozisiya, asosan, ertak bosh qahramonlarini aniqlaydi, yo'lakay ularning axloq- odobi, qiyofasi, xarakter-xususiyati haqida ma'lumot beradi, o'zaro munosabatlarini oydinlashtiradi. Ekspozisi-



ya ana shu zaylda ertak kompozitsiyasining ma'lum elementi sifatida syujet voqealariga aniqlik va ravonlik baxsh etadi va tugunga ishora tarzida yakun topadi.

Xullas, ertaklar o'ziga xos badiiy-kompozitsion tizimga-xayoliy uydirmaga, syujet motivlariga, maxsus personajlarga, badiiy til vositalari va uslubiy jilolarga egaligi bilan ajralib turadi va xalq ertaklarining eng qadimiy namunalari sifatida nisbatan ko'pchilikni tashkil etadi. Shuningdek, ertaklar xalqning necha-necha ming yilliklar davomidagi hayotiy tajribalarini umumlashtirgan holda uning ijtimoiy ongida, estetik didida, axloqiy qarashlarida, e'tiqodida kechgan o'sish-o'zgarishlarning badiiy tarixi sifatida ayricha ahamiyat kasb etgan. Shu bois hozir ham miriqib tinglanadi, sevilib o'qiladi, eng muhimi, navqiron avlodning ma'naviy-axloqiy kamol topishida beqiyos ta'sir ko'rsatib kelmoqda.

#### ADABIYOTLAR:

1. Adizova O. I. The exceptional in the works of Van Gennep. Special issue ijssir april 2023 issn 2277–3630 (online), published by international journal of social sciences & interdisciplinary research., comparative study of literature, comparative linguistics, translation studies, 2023y, 121–126 b

2. Adizova O.I.Folklorshunoslikka bir nazar. Zamonaviy psixologiya, pedagogikada fan, ta'lim va amaliyot integratsiyasi: muammo va yechimlar. Xalqaro ilmiy-amaliy konferensiyasi tezislar to'plamini 30-aprel 2023-yil127-131b

3. Adizova O. I. Les contes biographiques et esthétiques. Innovative Development in Educational Activities ISSN: 2181-3523 VOLUME 2 | ISSUE 6 | 2023 Scientific Journal Impact Factor (SJIF): 5.938

4. Adizova O. I. Medieval Tales and Fairy Tales in the 17th Century. Journal of Survey in Fisheries Sciences 10(2S) 1689–1697, 2023 Skopus

5. Arnold Van Gennep Le folklore. Croyances et coutumes populaires françaises. Paris : Librairie Stock, 1924, 16 p. Édition complétée le 3 août, 2009 à Chicoutimi, Ville de Saguenay, Québec

6. В.Я.Пропп. Фольклор и действительность. Избранные статьи. Москва: издательство « Наука», 1976,16стр

7. Yuldasheva D.N., Chullieva G. T. THE ROLE OF NOVERBAL MEANS IN SPEECH MEANING. International Engineering Journal For Research & Development, Vol.6, Issue 2 published May 16 2021. – Pag.1956-1961. www.iejrd.com <http://www.iejrd.com/index.php/%20/article/view/1956>

8. Narzullaeva, D. B. (2023, January). HISTORY OF THE TRANSLATION OF THE QURAN INTO FRENCH. In INTERNATIONAL CONFERENCES (Vol. 1, No. 2, pp. 116-118). <http://erus.uz/index.php/cf/article/view/1495>

УДК 81-11

**ЗНАЧЕНИЕ ПАМЯТИ В ХУДОЖЕСТВЕННОМ ПОРТРЕТЕ***Ш. Н. Ахмедова<sup>1</sup>, Д. Назарова<sup>2</sup>**<sup>1</sup>БухГУ, Бухара, Узбекистан**<sup>2</sup>Научный сотрудник БухГУ  
Бухара, Узбекистан*

shoiraaxmedova@gmail, d.nazarova@gmail.com

В данной статье рассматривается значение воспоминаний в создании литературного портрета, их роль в создании целостного образа человека. В портрете Откира Хошимова, созданном литературоведом А. Расуловым, он обратился к воспоминаниям писателя в создании полноценного образа писателя.

**Ключевые слова:** литературный портрет, память, критика, жанры критики, литературный процесс, литературная критика, образ писателя.

**THE SIGNIFICANCE OF MEMORY IN LITERARY PORTRAIT***Akhmedova Shoira Ne'matovna<sup>1</sup>, Nazarova D.<sup>2</sup>**<sup>1</sup>BukhsU, Bukhara, Uzbekistan**<sup>2</sup>Researcher BukhsU  
Bukhara, Uzbekistan*

shoiraaxmedova@gmail, d.nazarova@gmail.com

This article discusses the importance of memories in creating a literary portrait, their role in creating a complete image of a person. In the portrait of O'tkir Hoshimov created by the literary scholar A. Rasulov, he turned to the memories of the writer in creating a full-fledged image of the writer.

**Key words:** literary portrait, memory, criticism, genres of criticism, literary process, literary criticism, image of a writer.

**BADIIY PORTRETDAGI XOTIRA***Axmedova Shoira Nematovna<sup>1</sup>, Nazarova D.<sup>2</sup>**<sup>1</sup>BuxDU, Buxoro, O'zbekiston**<sup>2</sup>Ilmiy xodim BuxDU  
Buxoro, O'zbekiston*

shoiraaxmedova@gmail, d.nazarova@gmail.com

Ushbu maqolada adabiy portret yaratishda xotiralarning ahamiyati, shaxsning yaxlit qiyofasini yaratishdagi o'rni ko'rib chiqiladi. Adabiyotshunos A.Rasulov

tomonidan yaratilgan O‘tkir Hoshimov portretida yozuvchining to‘laqonli obrazini yaratishda yozuvchi xotiralariga murojaat qilgan.

**Tayanch iboralar:** adabiy portret, xotira, tanqid, tanqid janrlari, adabiy jarayon, adabiy tanqid, yozuvchi obrazi.

**Introduction.** Literary portrait is one of the important genres of literary criticism. When creating it, the portraitist pays attention to the person chosen as an object, and devotes a lot of space to his observations, opinions, and memories if he knows him closely. The literary portrait genre differs from other genres in its purpose, size, scope of analysis, and tasks. Also, a literary portrait requires attention to almost all works of the creator and expresses his attitude to them. In addition, literary portraits are written about artists who have served in the field of literature for a long time and who could please readers with their works. In the literary portrait, author’s goodwill towards the creator is clearly felt. From this point of view, the potential of literary portraits as a unique way of realizing a specific personality is wide enough. The main goal of the author of a literary portrait is to create an artistic image based on the artist’s unique, inner world, the life and biography of a particular person. A portrait is always “the story of a person about a person” stated in the language of a portrait painter [ 4, 29].

**The main part.** The main purpose of portrait navigation is to try to “perpetuate the image of this or that person”. Depending on the main intention of the author, there are colorful, completely different types of literary portraits, from the whole set of reflections and memories about one or another contemporaries (K. Fedin’s book “Gorky is among us”, works in the “Memory of Contemporaries” series) in articles and memoirs “ silhouettes”, up to lines in a portrait can be found [2, 32]. But in this case, as well as in other cases, we have the right to consider “literary portrait” as a phenomenon that creates the image of a unique “person”, expands the imagination about his life and work, and complements the biography of the writer being described. [1,46]. The author of a literary portrait finds the personal aspects of this person, the special features of a specific image, not only from his appearance, but also from his way of thinking, biography, and creative activity. This is manifested in the life mood of a writer or poet, episodes of his biography, life, relationship with others, at the same time, the non-repetition of his work, the language, style, construction, etc. of his works. For example, I. Karamzin, the founder of this genre in Russian literature, presents a critical genre – a literary portrait, that is, an essay about the

writer's life and work, he gives a detailed description of the writer, primarily as a person. In this, he relies not only on creativity, but also on the memories and thoughts of the writer's relatives: "They say that the life and character of a writer can be seen in his work; and yet, while we love subsequent notices, always inquire about the author from the first persons who knows or knew him personally." [ 1,76].

Therefore, it is necessary to take into account that memories have a great role in creating a literary portrait. Because memoirs allow one to learn deeply about the personality of the writer. The presence of such features in Uzbek portrait studies is only beneficial. For example, when writing a literary portrait of G. Gulyam, his warm attitude towards his wife, Muharram Aya, in addition to appearing as a great poet and philosopher, his inner image was revealed brighter and is more respected by the reader. The poet's contemporaries wrote about details such as the poet's kindness to his wife, bringing her a basket of fruits: peaches, pomegranates, with pleasure, no matter which city he went to. [ 6,80]. In the literary portrait, the reflection of such memories, the detailed description of the lines of the human figure of the poet gives it vitality and integrity.

Academician B. Nazarov also says about this relatively new trend in Uzbek literary studies: "Examination of the writer's creativity in close connection with his personality is one of the positive trends in literary criticism and literary studies in recent years. Studying the artist's life path, nature, youth, interests, and dreams serves as a key to reveal the secrets of skill and artistic aesthetic principles in his literary works." [3, 86].

Literary critic A. Rasulov's work dedicated to the popular writer O. Hoshimov stands out among the literary portraits created in the 90s of the 20th century due to the variety of means of expression and its unique construction. It is characterized by a lively conversation with the writer himself, describing the critic from the eyes of others, and striving to draw a holistic image of him. So, first of all, the author's sympathy for the personality and work of the writer O'Hoshimov is clearly felt, which is one of the factors that ensure the full output of the literary portrait. Literary portrait is full of vivid memories. Attention is paid to the writer's human qualities: hospitality, thoughtfulness, and humility. Assessment of his attitude to literature from the point of view of his friends, his confessions, sincerity in lively conversations, simplicity of the language of the work – all these combined caused an attractive literary portrait.

In short, a portrait is a description of a person and at the same time an assessment of him. But in the portrait expressed through words, the artist directly refers to the evidence of creativity and biography, giving him a direct description of the author. His personal “I” is felt especially vividly and clearly in literary portraits, where the author appears as an interlocutor to the contemporary depicted, and as an active person. The use of vivid memories, firstly, provides an interesting and readable output of the literary portrait, and secondly, helps to embody the complete image of the writer in the eyes of the reader.

## REFERENCES:

1. Ахмедова Ш. Ўзбек адабий танқидчилиги жанрлари. Тошкент. “Фан”. 2008. 46 -бет.
2. Барахов В. С. Литературный портрет. – М.: Наука, 1985. – 32 с.
3. Назаров Б. Ўзбек танқидчилиги. Тошкент. 1978.
4. Смирнова В. Современный портрет. М.: 1964. с.337.
5. Современная литературная критика. К вопросы теории и методологии. М.: 1977.
6. Саид Аҳмад. Назм чорраҳасида. – Тошкент, Ёш гвардия, 1982. – 80 б.

УДК 81-11

**ЭТНО-ТЕРРИТОРИАЛЬНАЯ И КУЛЬТУРНАЯ  
ХАРАКТЕРИСТИКА ПЕРИФРАСОВ**

*Ахмадова Умидахан Шавкат кизи*  
*Бухарский государственный университет,*  
*Бухара, Узбекистан*  
u.s.ahmadova@buxdu.uz  
uaxmadova477@gmail.com

В статье представлены и анализируются перифрастические единицы, воплощающие этнотерриториальные и культурные особенности. Доказано, что понятия язык, речь, культура, нация, традиция и ценности сходятся в одной точке, и одно требует другого. Обосновано, что перифразы являются одним из методических средств, оказывающих воздействие и повышающих уровень слышимости речи.

**Ключевые слова:** языкознание, перифраз, ремесло, национальное, культурное, ценность, мышление, традиция, ремесло, дискурс, содержание, сущность, народ, социальное.

**ETHNO-TERRITORIAL AND CULTURAL CHARACTERISTICS  
OF PERIPHRASES**

*Ahmadova Umidakhan Shavkat qizi,*  
*Bukhara State University,*  
*Bukhara, Uzbekistan*  
u.s.ahmadova@buxdu.uz  
uaxmadova477@gmail.com

In the article, periphrastic units embodying ethno-territorial and cultural features are presented and analyzed. It has been proven that the concepts of language, speech, culture, nation, tradition and values converge at the same point, and one requires the other. It is justified that periphrases are one of the methodological tools that have an impact and increase the level of audibility of speech.

**Key words:** linguistics, periphrasis, craft, national, cultural, value, thinking, tradition, craft, discourse, content, essence, people, social.

PERIFRAZALARNING ETNO-HUDUDIIY, MADANIY  
XUSUSIYATLARI

*Ahmadova Umidaxon Shavkat qizi*

*Buxoro davlat universiteti,*

*Buxoro, O'zbekiston*

*u.s.ahmadova@buxdu.uz*

*uaxmadova477@gmail.com*

*<https://orcid.org/0000-0002-6395-9812>*

Maqolada etno-hududiy, madaniy xususiyatlarni o'zida mujassamlashtirgan perifrastik birliklar keltirilib tahlil qilingan. Til, nutq, madaniyat, millat, an'ana va qadriyat tushunchalari ayni bir nuqtada birlashishi, biri ikkinchisini taqozo etishi dalillangan. Perifrazalar nutqning ta'sir ko'rsatuvchi, eshitimlilik darajasini oshiruvchi uslubiy vositalardan biri ekanligi asoslanilgan.

**Kalit so'zlar:** lingvistika, perifraza, hunarmandchilik, miliy, madaniy, qadriyat, tafakkur, an'ana, hunar, nutq, mazmun, mohiyat, xalq, ijtimoiy.

Til – ijtimoiy hodisa. U ijtimoiy rivojlanish jarayonida paydo bo'ldi. Kishilarning, har bir jamiyat a'zosining faoliyatini muvofiqlashtirish kabi muhim vazifani bajaradi. Til ayni vaqtda ongning shakllanishida muhim rol o'ynaydi. Til qobig'idan tashqarida ong mavjud bo'lmaydi. O'z fizik tabiatiga ko'ra o'zi bildiradigan narsaga nisbatan shartli bo'lgan til belgisi, shunday bo'lishiga qaramay, real voqelikni bilish jarayoni bilan bog'liq. Demak, til jamg'arilgan bilimlarni qayd etish va saqlash hamda ularni avloddan avlodga yetkazish vositasidir. Til insonlar faoliyatida fikrlash va o'zaro muomala vositasi sanaladi. Til muomala vositasiga aylanishi bilan imkoniyat doirasi ham kengayib boradi. Endi u nutq, nutqiy jarayon quroliga aylanadi. Bu orqali tilning ijtimoiylik imkoniyati ham kengayib boradi. Nutqiy faoliyatda til ko'magida suhbatdosh yoki tinglovchiga ma'lum axborot ham yetkaziladi. Ana shu suhbatning eshitimli va yoqimli bo'lishini ta'minlovchi unsurlardan biri perifrazalardir. Til mulki sanalgan perifrazalarni til va nutq bosqichida tahlilga tortilishi ham til va jamiyat munosabatlarining yana bir karra aloqadorligini ko'rsatadi.

Perifraza narsa-hodisani, obyektini tavsiflash uchun qo'llaniladigan stilistik figura sanaladi. Mavjud tavsiflash har qanday yorqin va ahamiyatli belgi yoki sifatni ajratish asosida yuzaga keladi. Perifrazalar hodisani shunday yorqin ifodalaydiki, jamiyat ahli kundalik hayotda undan keng foydalanishi oqibatida barqarorlik kasb etadi. Jumladan, Imom Buxoriy deganda *hadis ilmining sultoni*, mutafakkir Alisher

Navoiy deganda *eski o'zbek adabiy tilining asoschisi*, Margarit Tetcher haqida gap ketganda, shubhasiz, *temir ayol* perifrastik baholashlar qo'llaniladi.

Perifrazalar ma'lum tushunchani ifodalashdan tashqari unga qo'shimcha (ijobiy yoki neytral, ba'zan salbiy) ma'no ham yuklaydi, bu ularda pragmasema mavjudligini isbotlaydi. Masalan, yomg'ir so'zining o'rniga *osmonning ko'z yoshlari* tasviriy ifodasi qo'llangan-da fikr ancha ta'sirchan, emotsional bo'ladi. Perifrazalar badiiy asarlarda, ommaviy axborot vositalari sarlavhalarida yoki maqolalarda juda keng tarqalgan lingvistik hodisadir. Darhaqiqat, yorqin tavsiflashlar kitobxonni o'ziga jalb qilmay qo'ymaydi. Ayrim perifrazalar o'quvchilar orasida keng tarqalganligidan kim yoxud nima haqida so'z bo'rayotganligi aniq idrok qilinadi va bu, ko'pincha, muayyan perifrazani qo'llayotgan millat, elat yoki hudud bilan ham bog'liq bo'ladi.

Masalan, disfemistik xarakterdagi ba'zi ruscha perifrazalar o'zbek nutqida mutlaqo uchramaydi: *домашний убийца* (кухонный нож), *играть в гробу* (умирать), *книжный характер* (равнодушный), *колыбель законности* (законы) kabi.

Yoki o'zbek tilidagi davlat ustunlari (rahbarlar), raqs malikasi (Mukarrama Turg'unboyeva), osmon tirnog'i (oy), baraka urug'i (chigit), kumush tola (ipak), oq oltin ijodkorlari (paxtakorlar), farovon hayot bunyodkori (fermer) kabi perifrazalar rus yoki ingliz tillarida uchramaydi va hokazo.

Aytish joizki, oltin boshqoq (bug'doy), zangori ekran (televizor), zangori olov (gaz), po'lat qush (samolyot), mo'yqalam sohiblari (rasomlar), bahor elchisi (qaldirg'och), o'rmon malikasi (archa), qanotli do'stlarimiz (qushlar), salomatlik posbonlari (shifokorlar), charm qo'lqop ustasi (bokschi), hayvonlar podshosi (sher) kabi yuzlab perifrazalar o'zbek va rus tillarida bir xilda ishlatiladigan universal tasviriy ifodalardir.

O'zbek tili perifrazalarining o'zi ham hududiy jihatdan farqlanishini kuzatamiz. Masalan:

*El Oqpodsho qarorgohilab kelgunichayin olaqaroq tun bo'ladi* [1,6].

– *Taqsir... Siz to'pga tutajak dargoh masjid...*

– *Masjid bo'lsa nima bo'пти?*

– *Masjid Xudoi taolo dargohidur, taqsir. Xudoi taolo dargohidur* [1,10]. *Direktorimiz qop-qora chigitni ko'z-ko'z etadi.*

– *Paxta aslida mana bunday ichi qora bo'ladi! – deydi. – Paxta – ichi qora mato! Paxta odamzotni indamay boplaydi! U dunyo-yu,*



*bu dunyo o'ngarilmas qiladi! Byuroda o'rtoq Mavlonov ana shunday dedi!* [1, 33].

*Direktorimiz u yerda-bu yerda yotmish bolalarga baqiradi:*

– *Bu nima yotish?! – deydi. – Paxta fronti enangni uyimi?! Men hozir byurodan keldim!*

*Senlarni deb o'rtoq Mavlonovdan eshitmaganlarimni eshitib keldim!* [1, 35].

*Ayolimiz boshi uzra gumbaz tugun qo'ndirdi. Ayolimiz bag'ri aro qo'shqo'llab bola quchdi. Ayolimiz qishloq yo'l oldi* [1, 70].

*Olam tandir taft-tandir taft bo'ldi. Sonsiz ajinlarim-da dosh berolmadi. Saraton telpagim-da dosh berolmadi. Men lablarim yorilib-lablarim yorilib dalaladim* [1, 94].

Surxondaryo shevasi barq urib turgan mazkur asarda Tog'ay Murod o'ziga xos hududiy perifrazalardan foydalanganligini kuzatamiz. Jumladan, *Xudoi taolo dargohi – masjid; ichi qora mato – paxta; gumbaz tugun – bo'xcha; tandir taft – jazirama (issiq)*. Keltirgan misollarimizdan bittasi – paxta fronti (dala) barcha hududlarda bir xilda ishlatiladi.

– *Ebi, ebi, – dedi Rahim. – burodaring ahvolini gapur-da axi. Bar padari uyquki, peshinga chikin bo'lsa; ko'chaga bo'lib yotkan to'palangni go'shingga ilasanmi?* [2, 137].

– *Pul topiladi? – deb kuldi Rahim. – Bu bepadar boyvachchangizni qara, Sharifboy.*

– *Voy-voy, – dedi Rahim, – nazari baland bo'yoqchining o'g'lidan xafa bo'llem* [2, 137].

– *Qani, ining bilan ko'rishdigmi, Qobel?*

– *Ko'rishdim.*

– *Odami nag'z bo'lg'an?*

– *Odami nag'z* [2, 137].

– *Endi tutmakchilar...*

– *Endi tutmakchilar? – deb zaharxanda qildi Rahim. – Dardi xarini tutadi* [2, 139].

Mazkur misollarda ikki tilda so'zlashuvchi hudud vakillari – buxoroliklarning nutqida uchrovchi *bepadar boyvachcha, nazari baland bo'yoqchi, dardi xarina* kabi perifrazalar ham aynan hududiy xarakterdadir.

Til o'zining serjilo, nozik semasi bilan inson qalbini zabt etadi yoki butkul vayron qiladi. Bilamizki, o'zbek tili lug'at boyligi, qo'llanish doirasining kengligi bilan alohida ahamiyat kasb etadi. Bu imkoniyatlarni voqelantiruvchi til vositalaridan biri – perifrazalar bugungi

kunda jamiyatning barcha sotsial qatlamlari doirasida soʻzlashuvchi shaxslar nutqini uslubiy jihatdan farqlashiga guvoh boʻlamiz. Asosan, badiiy uslubning komponenti hisoblangan perifrastiklar bugungi davr talabi yuzasidan ham publitsistik, ham ilmiy, ham soʻzlashuv uslublarining badiiy estetik effekt beruvchi, ruhiyatga taʼsir qiluvchi asosiy vositasiga aylandi. Perifrastiklar barcha til egalari nutqida keng qoʻllaniladi. Har bir qatlam oʻz nutqida perifrastiklar maʼlum darajada foydalanadi. Chunki u nutqimizni bezovchi, oʻrinsiz qaytariqlarning oldini oluvchi, kishi ruhiyati va kayfiyatini koʻtaruvchi, nutq obyektining dolzarbligini koʻrsatuvchi tasviriy ifodadir. Perifrastik yoki tasviriy ifodalarning hosil boʻlishi predmet, voqea va hodisalarni oʻz nomi bilan emas, balki uning xarakterli belgi-xususiyatlarini tasviriy usul orqali ifodalash jarayoni hisoblanar ekan, yuqoridagi dalillar fikrimizga oydinlik kiritadi. Badiiy asar qiymatini oshiruvchi perifrastiklar, odatda, ham tasviriy vosita, ham biror soʻzni almashtiruvchi tasviriy element hamda ekspressivlik va emotsionallikni kuchaytiruvchi koʻchim turi sifatida qoʻllaniladi. Yozuvchilar oʻz asarlarida maʼlum bir predmet yoki inson faoliyati bilan bogʻliq hodisani yangicha nomlash orqali fikrni aniq tasvirlashga erishadi. Bunda perifrastik maʼno nomlangan soʻzning maʼnosiga qaraganda ancha taʼsirchan va ifodaviyligi bilan ajralib turadi. Masalan, «*Tun qushi ohista uchib oʻtdi, uning qizgʻish koʻzlari koʻrindi*» (Bodom qishda gulladi, 130-b.) yoki «*Yana toʻrt oyoqli bola... Bola! Cholgʻon, haqiqatan ham enam bola boʻlib qolibdi-ku?*» (Ogʻir tosh koʻchsa, 121-b.). Keltirilgan matnlarda ***tun qushi, toʻrt oyoqli bola*** perifrastik qoʻshilmalari boyqush, yoshi anchaga borib qolgan kampir maʼnosida ishlatilgan boʻlib, yozuvchining perifrastikdan foydalanish mahoratini anglab olishimizga imkoniyat yaratadi. Badiiy adabiyot tilida har bir badiiy-tasviriy vosita maʼlum bir voqea va hodisaga ekspressiv tus berish uchun ham ishlatiladi.

Toshkent, Fargʻona, Jizzax hududlarida yashovchilar nutqida *buloq suvidek tiniq* (zilol), *oftobni toʻsgan togʻ* (bahodir), *kun koʻzgu-si* (tong), *kumush poyandoz* (qor), *saxovat momosi* (yomgʻir), *bagʻri otash moʻjiza* (tandir); Xorazmda yashovchi aholi nutqida *oq buloq* (sigir), *qora tayloq* (shoʻx bola), *uzunning suvi* (chiroyli), *yigit doʻsti* (ot) kabi hududiy perifrastiklar qoʻllaniladi.

Til jamiyat koʻzgu-si, millat gʻururidir. Tilsiz jamiyat mavjud emas. Tildagi barcha boyliklar nutqda voqelanadi, nutq orqali yetkaziladi. Jamiyatning bir muhim boʻgʻini sanalgan hunarmandlar tili ham oʻzbek tili leksikasining asosiy bir qismi sanalib, uni boyituvchi manba sirasiga kiradi. Hunarmandchilik sohasida oʻzbek xalqining

milliy-madaniy xususiyatlari, o'ziga xos nomoyon bo'ladi. Tilning bunyodkorlik qudratidan jamiyatdagi har bir shaxs o'z shaxsi, salohiyati va qobiliyatidan kelib chiqqan holda foydalanadi. Shu jumladan, hunarmandchilik sohasida ham perifrazalarning paydo bo'lishi hamda qo'llanishi ham inson tafakkuriga borib taqaladi. Bilamizki, har bir mamlakatning milliyligi, an'ana va urf-odatlarining uning ya'ni xalqning hunarmandchilik sohasida ham o'z aksini topadi. Masalan:

### **O'zbekiston kashtachiligini noyob turi – zardo'zlik**

«Qadimda ham, hozir ham zardo'zlik bilan shug'ullanadigan hunarmandlarni **«zardo'z»** deb atashadi. *O'zbekiston kashtachiligini noyob turi zardo'zlik* san'atimiz to'la tasavvur qilib bo'lmaydi. Bu hunarmandchilik turi o'tgan asrning o'rtalarida Buxoroda gurkirab rivojlangan» [3].;

### **«Otashin kulolchilik» – kulolchilik**

Shahrisabz kulolchilik maktabi ustalari quyosh va olov rangdan samarali foydalanadi. Shu bois Yevropada bu uslubga «Otashin kulolchilik» deya ta'rif berilgan, deb yozadi O'za. [4].;

### **Hunarmandlar yurti – Shahrixon**

*Shahrixon azaldan hunarmandlar yurti sifatida dovrug' taratgan. Ayniqsa, tumanda pichoqchilik, do'ppido'zlik, beshiksozlik, temirchilik, tikuvchilik, milliy choponlar tayyorlash, kosibchilik, yog'ochdan turli buyumlar yasash kabi hunar turlari juda rivojlangan. Shahrixonlik hunarmandlar mahsulotlariga xorijda ham qiziqish katta* [5].;

### **Tadbirkor va hunarmandlar yurti – Buxoro**

«Buxoro azal-azaldan tadbirkor va hunarmandlar yurti bo'lib shuhrat qozongan. Bu salohiyatni ishga solish uchun aholiga dastak bo'ladigan sharoitlar yaratib, kasbga o'rgatsak va ozgina ko'mak bersak, bemalol 90 ming aholining bandligini ta'minlasa bo'ladi», deydi Shavkat Mirziyoyev [6].;

### **Kulollar yurti – Rishton**

«Ayni kunlarda Farg'ona vodiysining eng qadimgi maskanlaridan biri, kulollar yurti deya e'tirof etiladigan Rishtonda O'zbekistonda yagona bo'lgan kulolchilik majmuasida so'nggi pardoqlash ishlari ketmoqda. Majmua usta kulollar uchun barcha shart-sharoitlarga ega ustaxona va turarjoylar bo'lib qolmay, chet ellik sayyohlarning ham sevimli joyiga aylanishi kutilmoqda» [7].

Hududlar doirasida ham hunarmandchilik sohasi farqlanishi, kasb-korija, turmush tarziga qarab maskan nomlari perifrazalanganligini kuzatish mumkin.

Inson o‘zi mansub bo‘lgan millati so‘zlaydigan til hamda milliy-madaniy mentallik va voqelikni individual idrok etish usuli bo‘lgan o‘z tilining egasi ekanligi, til va nutqda insonning ichki dunyosi, aqliy saviyasi aks etishi, lisoniy shaxs sifatida namoyon bo‘lishini ko‘rish mumkin.

#### ADABIYOTLAR:

1. Мурод Т. Отамдан қолган далалар. – Тошкент: Шарқ, 1994. – 168 б.
2. Қодирий А. Ўткан кунлар. Мехробдан чаён. Рўмонлар. – Тошкент: Фафур Гулом номидаги Адабиёт ва санъат нашриёти, 1994. – 656 б.
3. <https://bozurgoniy.ulkansayt.uz/zardozlik-sanati>
4. <https://bozurgoniy.ulkansayt.uz/news/ganjkorli-va-kulolchilik-sanati.html>];
5. ([https://uza.uz/ru/posts/hunarmandlar-markazi-foydalanishga-topshirildi\\_277862?q=%2Fposts%2Fhunarmandlar-markazi-foydalanishga-topshirildi\\_277862](https://uza.uz/ru/posts/hunarmandlar-markazi-foydalanishga-topshirildi_277862?q=%2Fposts%2Fhunarmandlar-markazi-foydalanishga-topshirildi_277862)).
6. DY Yusupova. Poetess Khalim Khudoyberdieva and his skill in use of the word. International Scientific Journal ISJ Theoretical & Applied Science Philadelphia, USA issue 01, volume 93 published January 30, 2021. – Pag. 238–241. <https://www.elibrary.ru/item.asp?id=44644284>
7. (<https://qalampir.uz/uz/news/280-mingta-yangi-ish-urni-makhsus-turizm-dasturi-mirziyeev-bukhoroni-ng-keyingi-5-yili-%D2%B3ak-ida-47610>).
8. (<https://uzreport.news/society/ozbekistondagi-yagona-kulolchilik-majmuasi-faoliyatini-boshlashga-tayyorgarlik-korilmoqda>).

УДК 81-11

**ДЕТСКИЕ ПЕСНИ, СОЗДАННЫЕ НА МОТИВЫ  
НАРОДНЫХ ПЕСЕН***З. А. Амиркулова**БухГУ**Бухара Узбекистан**zarinaamirkulova@gmail.com*

В статье высказываются мнения, касающиеся детских песен, которые чрезвычайно просты по форме и имеют неповторимую мелодию.

**CHILDREN'S SONGS BASED ON FOLK SONGS***Zarina Azamat kizi Amirkulova**BukhSU**Bukhara, Uzbekistan**zarinaamirkulova@gmail.com*

The article expresses opinions regarding children's songs, which are extremely simple in form and have a unique melody.

**Key words:** folk songs, children's folklore, games, songs, ritual and seasonal songs, ethnographer.

**ХАЛҚ ҚЎШИҚЛАРИ ОҲАНГИДА ЯРАТИЛГАН БОЛАЛАР  
ҚЎШИҚЛАРИ***Амиркулова Зарина Азамат қизи**БухДУ**Бухоро Ўзбекистон**e-mail:zarinaamirkulova@gmail.com*

Мақолада болалар қўшиқларининг шакл жиҳатдан ниҳоятда оддий, ўзига хос оҳангдорликка эгаллиги билан боғлиқ фикрлар билдирилган.

**Таянч сўзлар:** халқ қўшиқлари, болалар фольклори, ўйинлар, қўшиқлар, маросим ва мавсумий қўшиқлар, этнограф.

Халқ оғзаки бадиий ижодиёти хилма-хил адабий тур ва жанрларда намоён бўлган ҳодиса сифатида кишилиқ жамияти тарихида кўп вазифали ижтимоий-эстетик моҳиятга эга сўз санъати ҳисобланади. У ҳамма замонларда ҳам ўз ижодкори бўлган халқнинг орзу-армонларини ифодалаб келади, қолаверса, халқнинг

Ўз-ўзинигина эмас, балки болаларнинг ҳам маънавий-ахлоқий жиҳатдан шаклланишларида муҳим тарбиявий восита вазифасини бажармоқда. Болалар халқ ижодиётидан ота-боболарининг ҳаёт тажрибасини, меҳнат ва кураш илмини ўрганиб келаётирлар. Бинобарин, халқ ижодиёти қадим замонлардан ҳозиргача болалар учун ҳам эзгуликни қадрлаш ва ёвузликдан нафратланиш сабоғи бўлиб, авлоддан авлодга ўтгани сайин муттасил тўлишиб, мукаммаллашиб, оғиздан-оғизга кўчгани сайин тобора сайқалланиб, халқ миллий руҳини теранроқ акс еттирган ҳолда асрлар давомида авлодлар маънавий-ахлоқий ворислигини таъминлаб келди, халқ учун ўзлигини англаш, ўз тарихини севиш воситасига айланиб қолди, энг муҳими, ёзма адабиётнинг бунёдга келишида генетик асос вазифасини бажарди. Бу силсилада болаларнинг ўз фольклори ҳам муайян рол ўйнади [1,8].

Катталар фольклоридаги кўшиқ, эртақлар болалар фольклорида фарзандларга мослаштирилган ҳолда яратилган. Уларни болалар жуда севиб тинглайдилар, ўқийдилар ва ижро этадилар. Болалар фольклори халқимиз оғзаки ижодининг таркибий қисмини ташкил этади. Унинг ўзи топишмоқ, тез айтиш, ялинчоқ, хукмлагич, санамачоқ каби кўп жанрлардан ташкил топган. Халқ оғзаки поэтик ижодидаги образларнинг ёрқинлиги, сўз маъноларининг аниқлиги, матннинг оҳангдорлиги, мусиқийлиги, жозибadorлиги каби хусусиятлар бола руҳига ором, хузур бағишлайди, уни ўзига маҳлиё этади.

Болалар кўшиқлари шакл жиҳатдан ниҳоятда оддий, ўзига хос оҳангдорликка эгадир. Болалар тилидан айтилган кўшиқлар ўз маъно-моҳияти билан айтувчиларнинг руҳий оламида юз бераётган турли хилдаги товланишлари, кўшиқ мазмунини ифодалаган содда, самимий, беғубор туйғуларни тараннум этишиши билан ажралиб туради.

Бинобарин, болалар фольклори – кичкинтойлар олами билан катталар дунёсининг ўзаро уйғунлашуви оқибатида юзага келган ўйинлар, кўшиқлар ва мусиқий поетик жанрларнинг яхлит бир тизимига айланган халқ оғзаки ижодиётининг ўзига хос тармоғидир.

Шундай кўшиқларнинг асл намунаси “Читтигул” туркумига кирадиган бир туркум болалар кўшиқларидир. Читтигул туркумига кирадиган кўшиқлар мавзусининг ниҳоятда ранг-баранглиги ва жуда жозибалилиги билан ажралиб туради. Чунончи:

*Оқ шолига кўк шоли,  
Оқ шолини оқлайлик,  
Кўк шолини кўклайлик,  
Яхши кунга сақлайлик,  
Ҳа-ю, читтигул,  
Ҳа-ю, читтигул*

сингари мисраларга назар соладиган бўлсак, тури, рангини эслатиш орқали шоли ҳосили жуда мўл бўлганлигини, уни эҳтиётлаб, тежамкорлик билан ишлатиш кераклиги болалар тилидан чиройли тасвирланган.

Энди бу мисраларга разм солинг:

*Жўхорилар бўлди оқ,  
Қўнма, шум қораялоқ!  
Агар қўнсанг, урайин,  
Бўлар оёгинг чўлоқ.  
Ҳа-ю, ҳувв!*

Кўриниб турибдики, кийин бир вазиятда катталар меҳнати билан жўхори экилган ва ҳосилга кирган жўхори агар қўриқланмайдиган бўлса, қарға, кузгунлар еб, экинни барбод қилади. Бу ишни эса кичкина авлод – болага топширилган. Боланинг ғайрат-шижотати, экинни сақлаб қолиш учун уриниши, ўйлаймизки, бугунги кун болаларига ҳам намуна бўлади. Юқорида кўриб ўтганимиздек, қўшиқлар болаларнинг дунёқарашини нафосатини ўстиришга, уларни ҳаёт ва меҳнатга муҳаббат, ватанпарварлик, дўстлик, инсонпарварлик каби олийжаноб фазилатлар руҳида тарбиялашга ёрдам беради. «Читтигул» – гулни, баҳорни эслатгани учун мавсум қўшиқлари қаторига киритиш одат бўлиб қолган, лекин фольклоршунос олим Ғ.Жаҳонгиров «Аслида бу қўшиқ баҳордан қишгача болалар, хусусан қизлар ўйнаб туриб айтадиган севимли қўшиғидир» [2, 54] – деб ёзади. «Читтигул» – мукамал ишланган, гўзал қўшиқ. Қўшиқнинг шакли ҳам, сўзлари ҳам энгил. Шу сабаб бўлса керак мазкур қўшиқ республикамызнинг деярли барча вилоятларида «Читтигул»га тақлидан айтилган қўшиқлар

«Банги-банги», «Лайзангул» шаклида ҳам учрайди. Бу қўшиқни куйлаш учун: “Қизлар жуфт- жуфт бўлишиб, кафтларини бир вақтда бир-бирига уриб айланишиб, қўйлақларининг этагини тугтиб, қўшиқ айтадилар. “Читтигул-о читтигул” дейилганда қизлар юзма-юз турадилар ва қўлларини маълум тартиб билан бир-бир-

ларига қарсиллатиб теккизиб, чиройли оҳанг касб этадилар. “Ҳа-ю читтигул, ҳа-ю читтигул” дейилганда қизлар бир хил мақом билан турган жойларида гир айлананиб куйлайдилар.

“Читтигул” кўшиғининг қизиқарли бўлиши болаларнинг ижодкорлигига кўпроқ боғлиқ бўлади. Бунда ўзига хос ўйин элементлари ва рақс ҳаракатлари бир-бирини тўлдириб боришади.

Болалар фольклори билан боғлиқ кўшиқларини ўргатиш орқали болаларда халқимизнинг ўзига хос анъаналарга бўлган ҳурмати, кадриятларни эъзозлашни, тарихимизга эҳтиром билан муносабатда бўлиш каби ҳисларни уйғотади.

#### АДАБИЁТЛАР:

1. Жумабобоев М. Ўзбек болалар адабиёти. – Т.: Ўзбекистон, 2002
2. Султонов Й, Раҳмонов Н, Турдимов Ш. Читтигул. – Т.: Ўқитувчи, 1992.
3. Жўраев М. Ўзбек мавсумий маросим фольклори. – Т.: Фан, 2008.
4. Юсупова Д.Ю. Ҳ.Худойбердиеванинг сўзларни кўчма маънода қўллаш маҳорати//НамДУ илмий ахборотномаси – Научный вестник НамГУ 2021 йил 6-сон. – 395–402-б.
5. Жаҳонгиров Ғ. Болалар эртақлари. – Т.: Чўлпон, 1972. – В.20.
6. Жаҳонгиров Ғ. Ўзбек болалар фольклори. – Т.: Чўлпон, 1975. – В. 88.



УДК 81-11

**ВЫРАЖЕНИЕ СЛОВ В СЛОВАРЯХ С НАРЕЧНЫМИ  
-ОНА АФФИКСАМИ**

*Асадов Тулкин Хамроевич*  
*Бухарский государственный университет,*  
*Бухара Узбекистан*  
*t.asadov@gmail.com*

В данной статье дан сложный по структуре и по содержанию наречный аффикс -она, подробно изучается возможность этих аффиксов в образовании слов, структурно-семантическое содержание слов с этим аффиксом. Выражения этих слов в словарях наблюдались. Каждое теоретическое положение подробно доказано и проиллюстрировано примерами.

**Ключевые слова:** словообразование, словообразовательный аффикс, продуктивный и непродуктивный аффикс, лексическая единица, способы образования слов, значение способов образования слов, структура слова.

**EXPRESSION OF ADVERBS WITH AFFIX -ONA IN THE  
DICTIONARIES**

*Asadov Tulkin Hamroyevich*  
*Bukhara state university*  
*Bukhara, Uzbekistan*  
*t.asadov@gmail.com*

The article is dedicated to adverb affix *-ona* which is complex in structure and meaning, and the possibility of these affixes in word formation, structural-semantic composition of words with this affix have been thoroughly researched. The expression of these words in dictionaries has been observed. Each theoretical view is adequately proved and responded with examples.

**Key words:** word formation, word-forming affix, productive and unproductive affixes, lexical unit, derivative, derivative meaning, word structure

**RAVISHGA XOS -ONA AFFIKSLI SO'ZLARNING  
LUG'ATLARDAGI IFODASI**

*Asadov To'liqin Hamroyevich*  
*Buxoro davlat universiteti,*  
*Buxoro, O'zbekiston*  
*t.asadov@gmail.com*

Mazkur maqolada ravishga nisbat beriladigan, tarkiban va mazmunan murakkab bo'lgan -ona affiksi, bu affikslarning so'z yasashdagi imkoniyati,

mazkur affiksli soʻzlarning struktur-semantik tarkibi atroflicha tadqiq etilgan. Mazkur soʻzlarning lugʻatlardagi ifodasi kuzatilgan. Har bir nazariy fikr misollar bilan keng dalillangan va munosabat bildirilgan.

**Kalit soʻzlar:** soʻz yasalishi, soʻz yasovchi affiks, unumli va unumsiz affiks, lugʻaviy birlik, yasama soʻz, yasama soʻz maʼnosi, soʻz tarkibi.

Oʻzbek tilshunosligida ravishga xos -ona soʻz yasovchi affiksi, uning xususiyati haqida bir qator fikrlar eʼtirof etilgan[3;5;6]. Mazkur affiksning soʻz yasashga munosabati – imkoniyati borasida fanimizda munozarali fikr-mulohazalar ham uchraydi [2,41-44].

-ona qoʻshimchali soʻzlar ravishlar guruhida oʻrganiladi. Buning bosh sababi mazkur affiksli soʻzlarning morfologik oʻzgarasligidir. Fors-tojik tilidan oʻzlashgan -ona affiksi bilan hosil boʻlgan 100ga yaqin soʻz nutqimizda u yoki bu darajada ishlatiladi. -ona affiksi bilan hosil boʻlgan yasama soʻzlar muayyan bir soʻz yasash qolipi, yaʼni [shaxs oti/sifat+ona] qolipi hosilalari sanaladi. Toʻgʻri, bu kabilar yasalishning dastlabki holatida (masalan, *qodirona pafos*, *qahhorona soʻz qoʻllash usuli* va h.k. misollarda) eshituvchiga gʻayritabiiy tuyulishi mumkin, biroq bu kabi soʻzlar davr bilan bogʻli holda umumisʼtemoldagi soʻzlar qatoridan joy olib, ular safi ortib bormoqda.

-ona affiksli soʻzlar [shaxs oti/sifat+ona] qolipi hosilalari hisoblanadi. Mazkur qolip oʻta unumli emas, albatta. Shu bois -ona affiksini oʻrta unumli soʻz yasovchi deb baholash mumkin. Bu affiksning oʻta faol emasligini quyidagicha izohlash mumkin. Maʼlumki, -ona soʻz yasovchisi affiks sifatida grammatik jihatdan -larcha, -chasiga kabi ravish yasovchi affikslarga teng keladi, yaʼni oʻzaro sinonim boʻladi. Masalan: *mardona* – *mardchasiga* – *mardlarcha* kabi. Demakki, -larcha, -chasiga affikslarida soʻz yasovchilik rolining balandligi [shaxs oti/sifat+ona] qolipini unumsizlashtirib turadi[1,67].

Tarixiy mansublik jihatdan forsiy tiliga tegishli -ona affiksi va shu shakl bilan yuzaga kelgan soʻzlar nutqimizda miqdoran yuzga yaqin boʻlib, ters lugʻatda mazkur affiksli 87 ta soʻz keltirilgan. Bular: *beadabona*, *boadabona*, *adabona*, *muhibbona*, *gʻoyibona*, *gʻolibona*, *gʻaroyibona*, *raqibona*, *aybona*, *donishmandona*, *rindona*, *mardona*, *ojizona*, *fojiona*, *goʻdakona*, *dadilona*, *odilona*, *razilona*, *oqilona*, *johilona*, *muloyimona*, *olimona*, *zolimona*, *maʼsumona*, *xoinona*, *pahlavona*, *qahramonona*, *majnunona*, *dilbarona*, *paygʻambarona*, *dilovarona*, *firibgarona*, *igʻvogarona*, *birodarona*, *kamtarona*, *muzaffarona*, *mutakkabirona*, *magʻrurona*, *ayyorona*, *mugʻombirona*, *bahodirona*, *botirona*, *faqirona*, *tahqirona*, *mohirona*, *hiylakorona*, *devkorona*, *riyo-*

*korona, tajavuzkorona, makkorona, sitamkorona, isyonkorona, fidokorona, xaloskopkorona, tashabbuskorona, xiyonatkorona, san'atkorona, ehtiyotkorona, lutfkorona, namoyishkorona, oshkorona, jinoyatkorona, ma'yusona, mastona, do'stona, badbaxtona, betarafona, zaifona, xarifona* [4,22-24]. Bu qo'shimchanning tilimizga o'zlashishi nisbatan keyingi davrlarga to'g'ri keladi.

Lug'atdagi bu so'zlarning ayrimlari nutqiy yasamalar sanaladi. Vaholanki, lug'atlarda lug'aviy birliklarga qamrab olinmog'i lozim, aslida.

Ravish yasovchi -an elementli so'zlar (*lutfan, tasodifan, qalban*) dan farqli o'laroq -ona qo'shimchali yasama so'zlarning yasalish asosi ma'lum bir qat'iy ma'no guruhiga ham ega. -ona affiksli so'zlarning lug'at qatlamimizda paydo bo'lishi o'zbek lug'atchiligining turli davrlariga xos bo'lib, ularning katta qismi shu holicha o'zlashgan, ayrimlari o'zbek tili so'z yasalishi usuliga ko'ra paydo bo'lgan. Biroq -ona affiksli so'zlarning qaysi biri o'zlashgan qaysi biri o'z qatlamda, yani so'z yasalishning ichki imkoniyati (affiksatsiya usuli)ga ko'ra hosil bo'lganligini aniqlash mushkul.

“O'zbek tilining izohli lug'ati”da -ona affiksli so'zlarning aksariyati mavjud. Faqat ters lug'atda berilgan to'qsonga yaqin so'zdan o'n yettitasi izohli lug'atda o'z belgisiga ega emas[7]. Bular: *adabona, muhibbona, g'aroyibona, donishmandona, go'dakona, johilona, muloyimona, majnunona, dilbarona, dilovarona, fribgarona, ig'vogarona, botirona, oshkorona, badbaxtona, mushtoqona, maxfiyona*. Mazkur so'zlarning amaldagi O'TILdan o'rin egallamaganligiga bosh sabab, bizningcha, eski, ya'ni 1981-yilda nashr etilgan O'TILda ushbu so'zlarning o'z ifodasiga ega bo'lmaganligi bo'lsa kerak. Chunki bu kabilar eski O'TILda ham kiritilmagan. To'g'ri, izohli lug'atlar keng iste'moldagi so'zlarni qamrab olishini, ters lug'atda keltirilib, biroq amaldagi O'TILda o'z belgisiga ega bo'lmagan bu kabi so'zlarning ko'pchiligi qo'llanish doirasi torligi bilan (masalan: *adabona, muhibona, muloyimona* kabi) alohidalik kasb etishini inkor etmoqchi emasmiz. Biroq nutqimizda tez- tez qo'llanuvchi *donishmandona, johilona, muloyimona, majnunona, fribgarona, botirona, oshkorona, g'aroyibona, maxfiyona* kabi so'zlarning amaldagi O'TILda keltirilmaganligi kishini o'ziga soladi.

Mazkur qolip hosilalari yuqoridagi so'zlar bilangina cheklanmaydi. O'zbek xalqi nutqida *sharqona, o'zbekona, turkona, alpona, qodirona* va h.k. kabi so'zlar ham mavjud. Bu kabi so'zlarni ham o'zbek tiliga oid izohli lug'atlardan topib bo'lmaydi. Tilshunos olim A.Hojiyev

bu kabi (jumladan, *sayrona*, *shekspirona* kabi) soʻzlarni okkazonal (nutqiy) yasalish deya taʼkidlaydi. Olimning taʼkidicha, bu kabi yasalish sunʼiy boʻlib, yasalishning meʼyoriy holatiga boʻysunmaydi [2,43]. Bilamizki, har qanday nutqiy yasalmalar (okkazonal yasalmalar), individual nutq koʻrinishlari, muallif neologizmlari lugʻatlarga kiritilavermaydi. Bizningcha, oʻzbekona, turkona, sharqona, gʻarbona, donishmandona, johilona, muloyimona, majnunona, fribgarona, botirona, oshkorona, gʻaroyibona, maxfiyona kabi OʻTILda keltirilishi lozim soʻzlar allaqachon lugʻaviy birlikka aylanib ulurgan. Buni mazkur soʻzlarning kitobiy uslubda keng qoʻllanilayotganligi ham dalillaydi. Misollarni kuzatamiz: 1. *Mungli tovushiga osurman quloq/ Dovdir majnunona xayolga asir/ Varaqlar ustida mixlandi nigoh/ Yongʻin changalida turibman axir.* (R.Parfi) 2. *Davlat rahbarining qatʼiyati, jasorati, donishmandona strategiyasi va eng muhimi xalqimizning birdamligi, ...butkul yangi Oʻzbekistonda shodiyona kunlarni nishonlashga oʻzimizni haqli sanasak boʻlar.* (“Xalq soʻzi” gazetasi, 2021) 3. ... *johilona ish tutishlar millatimiz tarixi va taqdirida xunuk va zararli iz qoldirishi tayin.* (“Xalq soʻzi” gazetasi, 2021)

*Qodirona* (*pafos*), *qahhorona*, *shekspirona* yoki *yusufona* (*usul*) kabi nutqiy yasalmalar ilk davrda har qanday kishi uchun gʻayritabiiy tuyulishi tabiiy. Biroq har bir yasama soʻz ehtiyoj mahsuli tufayli yuzaga kelishini nazardan qochirmaslik kerak, qolaversa, yasama soʻz yasalishning dastlabki davrida gʻayritabiiy (okkazonal) tuyulishi mumkin. Misollarni yana kuzatamiz: 1) *Gʻafur Gʻulom bejizga “qahhorona qisqalik” iborasini ishlatmagan. Abdulla Qahhor bir varaqlik hikoyalari ham butun kitoblarning zalvorli yukunini jo qila olgan.* (K.Azzamov) 2) *Sharq va Gʻarbda yashab oʻtgan koʻplab mutafakkirlar asarlarida ham navoiyona donishmandlikni koʻramiz.* (R.Jabborov)

Yuqorida keltirilgan misollar, umuman, -ona affiksli yangi-yangi soʻzlarda kuzatiladiki, bu affiks bugungi kunda shaxs nomlaridan yangi yasalmalar vujudga keltira oladi. Demak, [shaxs oti+ona] qolipi faoliyati davom etyapti, deya xulosa chiqarishga asos bor. [sifat +ona] qolipi hosilalari (*botirona*, *gʻolibona*, *mardona* va h.k.) oʻta unumsiz boʻlib qolgan. Aniqrogʻi, bu kabi yasalmalarning yuzaga kelishi uchun yangi-yangi sifatlar kerak boʻladi, avvalo.

-ona affiksli soʻzlarning -larcha, -chasiga affikslariga nisbatan nutq tasirchanligini kuchaytirishi, soʻz jozibadorligini taʼminlashdagi xizmati, uslubiy imkoniyati yuqori. Yana misollarga murojaat qilamiz: 1. *Nutqning eng haroratli, eng kuchli joyiga kelganda, artistona harakat qilib, raisga maʼnodor qarab qoʻyardi.* (S.Nazar) 2. *Qodiri-*

*yona* – *Chexovona hikoyalar yo'q darajada*. (Mirmuhsin) 3. *Uchinchi Rennsans deyilmish muhtasham qo'rg'on sari alpona qadam tashlash vazifasi turibdi*. (“Yangi O'zbekiston” gazetasi. 2021-yil oktabr soni) 4. *Xalqning og'ir ahvoli Fitratning “Abulfayzxon” asarida zo'r badiiy mahorat va shekspirona ko'lamda tasvirlangan*. (“Tarixdan hikoyalari”, 81-bet) 5. “*Oltin zanglamas*”ni o'qiganimda sezganman, undagi *godirona ifodalarga tan berganman*. (O.Mo'minxo'ja “Ustozlar va zamondoshlarim haqida”) 6. *Ma'naviy tahdid sifatida baholanayotgan g'arbona yashash bizning muqaddas qo'rg'onimizga qanchalik salbiy ta'sir ko'rsatmoqda*. (“Oila tinchligi – jamiyat farovonligi asosi” maqola) 7. *O'zbekona kiyinish madaniyatini targ'ib qilish, zamonaviy kiyinishimizga rioya qilishni yoshlar o'rtasida keng targ'ib qilish lozim* (“Milliy liboslar – ma'naviyat ko'zgusi”)

Xullas, -ona affiksi ravish yasashda faol, o'ta unumli bo'lmasa-da, mazkur shaklni o'rta unumli so'z yasovchi affikslar qatorida o'rganish mumkin. -ona affiksli yasama so'zlarning yasalish asosi esa shaxs nomi bo'lib keladi. Kuzatilgan misollar ham buni to'liq dalillaydi.

#### ADABIYOTLAR:

1. Асадов Т. Сўз туркумлари тизимида равиш. Монография. – Т.: Мухаррир, 2010.
2. Ҳожиёв А. Ўзбек тилида равиш ясашиши тизими масаласига доир// Ўзбек тили ва адабиёти. 2001, 2-сон, 41–44-б.
3. Кононов А.Н. Грамматика современного узбекского литературного языка. – М.-Л., 1960.
4. Кўнғуров У., Тихонов А. Ўзбек тилининг терс луғати. – С., 1968.
5. Юлдашева Д.Н. РОЛЬ МОЛЧАНИЯ В РЕЧИ// ОБРАЗОВАНИЕ И ИННОВАЦИОННЫЕ ИССЛЕДОВАНИЯ. № 1 (2020): <https://interscience.uz/index.php/home/article/view/31>
6. Rahmatullayev Sh. Hozirgi adabiy o'zbek tili (darslik). – Т.: Universitet, 2006.
7. Ўзбек тили грамматикаси. I том. Морфология. – Т.: Фан, 1975.
8. Ўзбек тилининг изоҳли луғати. 5 жилдди. – Т.: Ўз. мил.энцикл., 2006–2008.

## УДК 81-11

## ИЗУЧЕНИЕ АРАБСКИХ СОГЛАСНЫХ ГРУППАМИ

*Г. А. Астанова**Бухарский государственный университет**Бухара, Узбекистан**gulnoraastanova@gmail.com*

В условиях, когда общественно-политические, культурные и экономические связи нашей страны с зарубежными странами Востока крепнут, становится актуальной подготовка потенциальных филологов, переводчиков, синхронных переводчиков, свободно владеющих иностранными языками, в том числе восточными. один из актуальных вопросов современности. В связи с этим важно внедрить в систему непрерывного образования Узбекистана международные образовательные стандарты преподавания иностранных языков и базировать ее на уровнях «общеευропейских компетенций владения иностранными языками: обучение, преподавание, оценка». Соответственно, данная статья посвящена изучению классификации согласных при изучении фонетики арабского языка.

**Ключевые слова:** арабский язык, согласные, звуки, классификация, гласные, узбекский язык.

## LEARNING ARABIC CONSONANTS BY GROUPS

*Astanova Gulnora Aminovna**Bukhara state university**Bukhara, Uzbekistan**gulnoraastanova@gmail.com*

At a time when the social, political, cultural and economic relations of our country with the foreign countries of the East are getting stronger, training of potential philologists, translators, simultaneous translators who are fluent in foreign languages, including Eastern languages, is one of the urgent issues of the day. In this regard, it is important to introduce international educational standards for teaching foreign languages into the continuous education system of Uzbekistan and to base it on the levels of “pan-European competences for mastering foreign languages: learning, teaching, assessment”. Accordingly, this article focuses on studying the classification of consonants in the study of phonetics of the Arabic language.

**Key words:** Arabic language, consonants, sounds, classification, vowels, Uzbek language.

## ARAB TILIDA UNLILAR VA UNING GURUHLARINI O'RGANISH

*Astanova Gulnora Aminovna*  
*Buxoro davlat universiteti*  
*Buxoro, O'zbekiston*  
gulnoraastanova@gmail.com

Mamlakatimizning Sharqning xorijiy davlatlari bilan ijtimoiy-siyosiy, madaniy va iqtisodiy aloqalari mustahkamlanib borayotgan bir sharoitda chet tillarini, jumladan, sharq tillarini mukammal biladigan salohiyatli filologlar, tarjimonlar, sinxron tarjimonlarni tayyorlash dolzarb ahamiyat kasb etmoqda. . zamonamizning dolzarb masalalaridan biri. Shu munosabat bilan O'zbekistonning umrbod ta'lim tizimiga chet tillarini o'qitishning xalqaro ta'lim standartlarini joriy etish va uni "Chet tillari bo'yicha umumiy Yevropa kompetensiyalari: o'rganish, o'qitish, baholash" darajalariga asoslash muhim ahamiyatga ega. Shunga ko'ra, ushbu maqola arab tili fonetikasini o'rganishda undoshlar tasnifini o'rganishga bag'ishlangan.

**Tayanch so'zlar:** arab tili, undoshlar, tovushlar, tasnif, unlilar, o'zbek tili.

From the point of view of the Uzbek language, three of the 25 consonants in our language have been borrowed from other languages. They are: "f", sliding "j" and "ts" sounds.

The sound "f" first entered the Uzbek language through Persian and Arabic words. M.: fan, opinion, fatwa, secret, mufti (Arabic), decree, raftor, week, zarboft (Persian).

The sliding sound "j" first entered through Persian words. M.: mujda (good news), ajdar, gijda.

These two sounds, i.e. "f" and the sliding "j" began to enter Russian-international words from the end of the 19th century. M.: factory, buffet, candy, fufaika, garage, giraffe, magazine, dubbing, reportage, landscape. The sound "ts" entered our language only through Russian-international words. M.: constitution, conference, intuition, lyceum.

The remaining 22 consonants in our language belong to their own layer, that is, they have existed in the Uzbek language for a long time. But there are only 2 variants of the "v" sound. "v", which is considered a lip-lab sound, belongs to its own layer, and the lip-tooth variant of the "v" sound is found only in Arabic and Russian-international words. M.: homeland, and, wagon, currency.

Consonants are classified according to the following aspects:

I. According to the source of origin.

II. According to the participation of the vocal cords.





|                            |        |           |  |   |          |          |           |   |  |   |           |          |  |
|----------------------------|--------|-----------|--|---|----------|----------|-----------|---|--|---|-----------|----------|--|
| Slider                     | noisy  | difficult |  |   | <b>z</b> | <b>z</b> | <b>z.</b> |   |  | , | <b>'a</b> |          |  |
|                            |        | soundless |  | f | <b>ʃ</b> | <b>s</b> | <b>c.</b> | h |  |   | <b>x.</b> | <b>h</b> |  |
|                            | sounds | ʃ         |  |   |          |          | l         |   |  |   |           |          |  |
| Vibrating sounds           |        |           |  |   |          |          | r         |   |  |   |           |          |  |
| Mixed sounds or affricates |        |           |  |   |          |          | j         |   |  |   |           |          |  |

Classification of consonant sounds according to the participation of the voice. Consonants are divided into 2 groups according to the participation of the voice:

1. Sonors: m, n, l, r.
2. Voiceless: the remaining 21 consonants belong to voiceless.

In the composition of sonors, flour occupies a greater position in relation to noise. According to their sonority, they are close to vowel sounds.

In noisy consonants, the noise takes a greater position compared to it. Classification of consonants according to the place of formation.

The existing 28 consonant sounds are formed in the throat zone, tongue zone or lip zone.

Therefore, in this respect, consonants can be divided into 3:

1. A guttural consonant.
2. Language consonants.
3. Lip consonants.

This division table is shown above.

Classification of consonants according to the method of formation.

Consonants are divided into 2 according to the way consonant sounds are formed or according to the way they are pronounced:

1. Explosive: b, d., d, k, m, n, t, t., q, 'hamza. (10).
2. Sliding: z, z., z, 'ayn, y, l, s., s, c, f, x, x., sh, g', h, o', r. (16) .

According to the method of formation, some sounds have certain characteristics. For example, the sound "r" is a vibrating sound. Therefore, it can be considered as sliding only conditionally. And the sound "L" belongs to the sliders. But the air flows out from both sides of the tongue. That is why it is called a side sound.

Types of consonant sounds according to their composition.

According to the composition of consonants, it is divided into 2:

1. Single consonants.
2. Mixed consonants or affricates.

There is only 1 affricate in the Arabic literary language, which is the explosive sound “j”. The remaining 27 consonants are single consonants. They are simple in composition.

Description of Arabic consonants.

Consonants s and z are sliding and noisy sounds formed between the teeth. Consonant s is voiceless, and consonant z is voiced. When pronouncing them, the tip of the tongue is placed between the teeth and glued to the upper teeth. Then these sounds slip out of the small slit formed between the tip of the tongue and the lower teeth. There are no such sounds in the Uzbek language. It is similar to the pronunciation of th in English, for example (to) thank, there is there.

Explanation: Arabic words with such letters become simple s and z when used in Persian and Turkish languages.

The consonant “r” is formed in the process of regularly hitting the hard palate with the tip of the tongue. The consonant “R” is sometimes pronounced hard, sometimes soft, depending on the vowel that comes after it. Arabic “r” resembles the hard “r” in Russian, but in some cases it is pronounced soft. “r” before the vowels “a” and “u” is pronounced very hard. For example, my god, there are many rubbas. However, when accompanied by “i”, it resembles the Russian r. For example, “ribā” is usury.

The consonant “vav” is a lip-lip, sonorous sliding sound. In its pronunciation, the lips come closer to each other and move forward, creating a round airway. There is no such sound in Russian and Uzbek. This sound is similar to w in English. For example: war war. M: minister, minister.

Consonant “y” is a middle palatal sliding sound. Its slip occurs between the middle wall of the tongue and the hard palate. In this case, compared to the pronunciation of “y” in the Russian language, the language tends more forward. Therefore, the Arabic “y” is softer than the Russian “y”. For example: yadū my hand, yasibu it seems to be jumping.

Also, the consonants “o” and “y” are similar to vowel sounds. That is why they are also called semi-vowels.

Consonant “j” – before the tongue, sonorous and noisy sound. It is similar to the sound “j” in Uzbek. For example: Jadūd is a new, prosperous merchant.

The sound “k” is a middle, explosive, noisy, voiceless consonant. It is formed as a result of an explosion between the wall of the palate and the hard palate. It is pronounced a little softer, just like in Uzbek, for example: write katab, kabūr is big.

Also, the consonants “o” and “y” are similar to vowel sounds. That is why they are also called semi-vowels.

Consonant “j” – before the tongue, sonorous and noisy sound. It is similar to the sound “j” in Uzbek. For example: Jadūd is a new, prosperous merchant.

The sound “k” is a middle, explosive, noisy, voiceless consonant. It is formed as a result of an explosion between the wall of the palate and the hard palate. It is pronounced a little softer, just like in Uzbek, for example: write katab, kabūr is big.

The sounds s and sh are letters of the same shape, with s without a dot and sh with three dots.

### **Summary.**

The sounds described above have their own characteristics. Because learning the pronunciation of such sounds requires a lot of work and effort from an Uzbek student. In the same way, every subject teacher who teaches the Arabic language needs great responsibility and attention in teaching these sounds.

### REFERENCES

1. Abdullaev M. Arab tili. T., 2006.
2. Ibrohimov B. O'zbekcha-arabcha so'zlashgich. – T., 1994.
3. Ibrohimov N., Yusupov T. Arab tili grammatikasi. 1,2-jildlar. – T., 1998, 2003.
4. Talabov E. Arab tili. – T.: “O'qituvchi”, 1993.
5. Халидов Б.З. Араб тили дарслиги / Тарж. Остонова Г. А. – Т.: «Фан», 2007.
6. Astanova G.A. Arab tili saboqlari. Buxoro, 2023.
7. Astanova G.A. Arab tilidan boshlang'ich kurs. Buxoro, 2023.

## УДК 81-11

## ОКАНЬЕ В СИСТЕМЕ ГЛАСНЫХ УЗБЕКСКОГО ЯЗЫКА

*A. P. Axmedov**БухГУ**Бухара, Узбекистан**atouллоaxmedov@gmail.com*

В данной статье представлена информация об оканье в тюркском языке, в том числе в узбекском языке. Обобщены взгляды учёных тюркологов. Утверждается, что оканье образовалось на основе внутреннего закона тюркских языков (узбекского языка). Комментированы спорные моменты.

**Ключевые слова:** тюркский, тюркский язык, оканье, нелабиализованный, литературный язык, диалект, персидско-таджикский язык

## PHONETICS “O” SYSTEM OF THE UZBEKI LANGUAGE

*Akhmedov Atoulo Rakhmatovich**BukhSU,**Bukhara, Uzbekistan**atouллоaxmedov@gmail.com*

This article provides information about phonetics “o” in the Turkish language, including the Uzbek language. The views of learned Turkologists are generalized. It is argued that phonetics “o” was formed on the basis of the internal law of of the Turkic languages (Uzbek language). Controversial issues are commented on.

**Key words:** Turkish language, phonetics “o”, non-labialized, literary language, dialect, Persian-Tajik language.

## O‘ZBEK TILI UNLILAR TIZIMIDA “O”LANISH

*Axmedov Atoulo Raxmatovich**BuxDU**Buxoro, O‘zbekiston**atouллоaxmedov@gmail.com*

Ushbu maqolada turkiy tillarda, jumladan, o‘zbek tilida “o”lanish hodisasi borasida ma’lumot berilgan. Bunda tirkiyshunos olimlarning fikrlari umumlashtirilgan. “O”lanish hodisasi turkiy tillarning (o‘zbek tilining) ichki qonuniyati asosida hosil bo‘lganligi bayon qilingan. Munozarali fikrlarga sharhlangan.

**Kalit so‘zlar:** turkiy, turkiy til, “o”lanish, lablanmagan, adabiy til, sheva, fors-tojik tili

Turkiyshunoslikdagi dolzarb muammolardan biri o‘zbek tilida «o»lanish hodisasidir. Bu boradagi fikrlarni bayon qilishdan oldin, «o»lanish borasida to‘xtalish lozim.

«O»lanish deganda, umumturkiy keng orqa qator lablanmagan unlining yarim lablangan keng orqa qator unliga o‘tishi, ya’ni lablanmagan unlining ma’lum darajada lablanish hodisasiga aytiladi. Masalan: *at-ot, ata-ota, baba-bobo*. Turkiy tillar orasida umumturkiy «a»ning «o»ga o‘tishi bizning tilimizda (o‘zbek tilidan) eng keng tarqalgan. Lekin bu «a»ning «o»ga o‘tishida qat’iy va buzilmas qoida uchtaadir:

1. Lab-lab «v» va chuqur til orqa «q» dan oldin kelgan «a» muntazam ravishda «o»ga o‘tadi. Masalan: *oq-aq, sana-sanoq, qayra-qayraq, yara-yaroq*;

2. -v qo‘shimchasi «a» unlisi bilan tugagan so‘zlarga qo‘shilganda «o» «o»ga o‘tadi: *so‘ra- so‘rov, sayla-saylov*;

3. «a» ning «o»ga o‘tishi «g» dan oldin ham uchraydi; *yarog‘, bog‘, tog‘, sog‘, chog‘*...

Shu uch holatdan tashqari umumturkiy «a»ning o‘zbek tili va shevalarida «o»ga o‘tishida qat’iy qonuniyat yo‘q. O‘zbek adabiy tilida «o» bilan talaffuz etiladigan so‘zlar qarluq-uyg‘ur shevalarining o‘zida «a»li variantga, shevadagi «o»li so‘zlar adabiy tilda «a»li shaklda kelishi mumkin:

| Adabiy til  | Sheva       |
|-------------|-------------|
| <i>Bola</i> | <i>Bala</i> |
| <i>Qora</i> | <i>Qaro</i> |
| <i>Olma</i> | <i>Alma</i> |
| <i>Anor</i> | <i>Onor</i> |

Shevalarda *oka, katta* adabiy tilda *aka, katta* kabilarda «a»ning «o»ga o‘tish sabablari turkiyshunoslikda haligacha muammodir. Bir qism tadqiqotchilar «o»lanishni eroniy tillar ta’siri deb baholashsa (chunonchi, E.Polivanov, A.K.Borovkov), ko‘pchilik turkiyshunoslar, jumladan, A.K.Borovkovning o‘zi ham, «o»lanish yuqorida ko‘rib o‘tilganidek, akkomadatsiya – lablangan va chuqur til orqa undoshlarning ta’siri ostida shakllangan, degan taxmini oldinga suradilar. Turkiy arealogiyada, jumladan, N.Z.Gadjeva, N.A.Baskakov tadqiqotlarida «o»lanish markazi hozirgi o‘zbek tili lisoniy areali sanalsada, birinchidan «o»lanish – «a»ning «o»ga o‘tishi – o‘zbek shevalarining o‘zida

yuqorida ko‘rib o‘tganimizdek bir xil emas. Ikkinchidan, bu hodisa o‘zbek tili bilan yondosh bo‘lgan qirg‘iz, turkman, qozoq, qoraqalpoq tillarining shevalarida ham, o‘zbek tili bilan hech qachon bevosita aloqada bo‘lmagan ozarbayjon tili, xususan, Boku shevasida, G‘arbiy

Sibir totor shevalarida, Shimoliy Kavkazdagi qarachay, bolqar va qo‘miq tillarida, totor tilida, boshqird tilida bundan tashqari paradik ravishda chuvash shevalarida ham uchraydi. Shuning uchun akademik A.N.Kononov «*«o»lash faqat o‘zbeklargagina xos emas va hamma o‘zbeklar ham «o»lamaydi. Bundan tashqari barcha turkiy «o»lovchilarda «o»lash bir xil emas», -degan fikrni oldinga surgan. V.V.Reshetov o‘zining «Узбекский язык» kitobida o‘zbek tili shevalarida 3-4 xil «o» tovushi borligini tavsiflagan. Demak, o‘zbeklar «o»si ham bir xil emas. Shuningdek, totor, qo‘miq, qarachay va boshqa tillardagi «o» ham bir-biriga o‘xshash emas. Shuni alohida ta’kidlash lozimki, turkiy tillarda «a» qaysi holatlarda «o»lanadi va qaysi holatlarda «o»lanmasligini belgilovchi qat’iy qonuniyat mavjud emas. «O» tovushi mavjud bo‘lgan tillarda «o» bilan talaffuz etiladigan so‘z ikkinchi bir turkiy tilda «a» shaklida uchraydi. Turkiy arealogiyaning hozirgi holatida «o»lanish markazi o‘zbekzabon areal sanalishiga qaramay «o»lanish hodisasi o‘zbekzabon markazdan Sibir, Paloje, Shimoliy Kavkaz va Zakavkazyaga tarqalgan, deb arealogiya nuqtayi nazaridan xulosa chiqarish mutlaqo noto‘g‘ridir. Chunki o‘zbekzabon arealdan bu mintaqalarga migratsiya ham, kuchli ta’sir ham hech qachon bo‘lmagan. Akademik A.N.Kononov ma’lumotiga qaraganda, qadimgi turkiy (ehtimol qarluq toifasiga mansub) xalqlarning katta bir bo‘linmasida «a»ning «o»ga o‘tish hodisasi qat’iy qonuniyat bo‘lgan. Tarix, yozma yodgorliklarda qayd etilgandan ancha oldin bu qabila (bo‘linma) parchalanib turli toifalarga bo‘linib ketgan. Nisbatan kattaroq qismi hozirgi o‘zbek xalqi tarixiy asosiga, ayrim-ayrim nomayondalari ozmiko‘pmi boshqa turkiy xalqlarning tarixiy asosidagi tarkibiga kirgan. Shuning uchun «o»lanish siporadik (uzuq yuluq) shaklda xilma-xil sheva va tillarda uchraydi. Hozircha turkiyshunosliklar «o»lanish hodisasini qaysi fokus zonadan («o»lanishni tarqalish markazidan) qaysi yo‘nalishlarga yo‘nalganligi muammo bo‘lib qolmoqda. Chunki «o»lanishning sixronik markazi keng tarqalgan mintaqasi o‘zbekzabon areal tarixiy markaz vazifasini hech qachon bajarmagan. Tabiiyki, okkomadativ «o» (ya’ni lab va chuqur til orqa undoshlari ta’siri ostida «a»ning lablanishi) turkiy tillarda mustaqil ravishda har bir tilda (lahjada) alohida-alohida ko‘rinishda shakllangan va rivojlangan bo‘lishi mumkin. Turkiy sheva va tillarda «o»ning bir xil emasligi, ehtimol shu*

bilan – «o» tovushining tillarda har bir til uchun oʻziga xos qonuniyat asosida rivojlanganligi bilan izohlanishi mumkin boʻlsa kerak.

«O»ning kelib chiqish manbayi sifatida qadimgi turkiy birlamchi choʻziq «a» manba boʻlgan degan faraz ham mavjuddir. Haqiqatan ham «o»lanish keng tarqalgan oʻzbek tilida «o», eksperimental fonetistlar Ahmadjon Mahmud, Sora Otamirzaeva tadqiqotlariga koʻra, barcha boʻgʻinlarda choʻziq talaffuz qiladi. Choʻziqlikning kamayishi jihatidan oʻzbek tili unililari quyidagi darajalanish qatorini tashkil etadi: o→oʻ→e→a→u→i.

Bundan tashqari arab va eron tillaridagi choʻziq «a» oʻzbek tilida muntazam ravishda «o»ga oʻtadi: *Allah-Olloh, kitab-kitob, katib-kotib...* Lekin arabcha qisqa «a»ni ham oʻzbek tilida «o»ga oʻtish holatlari oz emas: *avqat-avqot, alov-olov-olav*. Lekin bu faraz ham eʼtirozga oʻrin qoldiradi. Zeroki, «o»lanishning manbayini qadimgi turkiy birlamchi «a» boʻlganda, bu «a» «o»lanish xususiyatga xos boʻlgan til va shevalarning barchasida «o»ga oʻtishi lozim edi. Vaholanki, ahvol bunday emas. Jumladan: *a: d, a: y* kabi qadimgi birlamchi choʻziq «a»li soʻzlar muntazam ravishda «o»lovchi til va shevalarda «o»ga oʻtmaydi. Bu hodisani XI asrdayoq Mahmud Koshgʻariy qayd etgan edi. U ikki alif bilan berilgan *az, al, at* kabi soʻzlarni qattiq va qisqa talaffuz qilish maʼqul deb alohida taʼkidlagan. Shuning uchun «o»lanish turkiy tilshunosning ham tarixiy fonetikasi, ham arealogiyasining chigal muammolaridan bir boʻlib qolaveradi.

Endi bir-ikki ogʻiz soʻz turkiy tillarda, jumladan, oʻzbek tilida «o»lanish hodisasi qoʻshni eroniy til taʼsirida yuzaga kelgan, degan talqin yuzasidan. Gap shundaki, «o»lanishni boshqa bir tilning taʼsiri deb bohalash (oʻzbeklarga tojiklarning taʼsiri, totor va boshqirdlarga rus «o»lovchilarining taʼsiri va x.k.) oʻrinli emas. Chunki, eroniy tillar ichida tojik tilining oʻzi «o»lovchidir. Bu hodisa eronshunoslikda xuddi turkiyshunoslikda boʻlgani kabi oʻzbeklarga toʻntariladi. Lekin eron tillaridagi (ularning tiliga oʻzlashgan arabiy soʻzlardagi) choʻziq «a»ni (alif mamdudani – maddali alifni) oʻzbek tilida «o»ning boshqa unililardan koʻra bir oz choʻziqligi asosida, arab va fors tillaridagi choʻziq «a»ning muntazam ravishda «o» bilan berish oʻzbek tilida «o»lanishning keng tarqalishiga mos boʻlgan boʻlishi mumkin. Lekin bu oʻzbeklarda «o»lanish eroniy va arab tilining taʼsiri deb aytishga asos boʻla olmaydi.

#### ADABIYOTLAR:

1. AR Akhmedov, Sevara Poʻlatova. Theoretical and practical issues of creating uzbek areal linguistics. Middle European Scientific Bulletin, 2021

2. A.R. Akhmedov, Sevara Po‘latova. O ‘zbek shevalarini lingvo-geografik xaritalashtirish. Linguistics of different fields: problems, solutions and prospects of linguistics and medical integration. 2021.

3. A.R. Akhmedov. Vobkent tumani shevasida so‘zning morfemik tarkibi. scientific reports of bukhara state university. 4/2022.

4. Sevara Pulatova. Intelligentsia of uzbekistan as a social strata. Himalayan and Central Asian Studies. 2015/7/1

5. Dildora Yusupova, Dilorom Yuldasheva. Halima Xudoyberdiyeva she’riyati lingvopoetikasi. – Germaniya: “Globe Edit international Publishing Group” ISBN: 978-620-0-60879- 6. – 2020. – 125 p.



УДК 81-11

## ПЕРЕЁНИМЫ – ОДИН ВИД ОНОМАСТИЧЕСКИХ ЕДИНИЦ

**С. К. Ахмедова**

*Бухарского государственного университета  
Бухара, Узбекистан  
sevaraxmedova@gmail.com*

В этой статье рассматривается переёнимы, как один из видов ономастических единиц, место и значение их в языке и речи, некоторые лексико-семантические свойства и вопрос классификации их.

**Ключевые слова:** Ономастика, ономастическая единица, переёним, транспорт, названия транспортного средства, классификация.

## PEREYONYMS IS A TYPE OF ONOMASTICS

**Axmedova Sevar Qudratilloevna**

*Bukhara State Universite  
Bukhara, Uzbekistan  
sevaraxmedova@gmail.com*

This article discusses the names of transports as one of the types of onomastic units, their place and meaning in language and speech, some lexico-semantic properties and the question of their classification.

**Key words:** Onomastics, onomastic unit, pereyonym, transport, transport names, classification.

## PEREYONIMLAR – ONOMASTIK BIRLIKLARNING BIR TURI

**Axmedova Sevar Qudratulloevna**

*Buxoro davlat universiteti  
Buxoro, O'zbekiston  
sevaraxmedova@gmail.com*

Ushbu maqolada onomastik birliklarning bir turi bo'lgan pereyonimlarning til va nutqdagi o'rni, ayrim leksik-semantik xususiyatlari, tasnifi masalasiga e'tibor qaratilgan.

**Tayanch so'zlar:** onomastika, onomastik birlik, pereyonim, transport, transport nomlari, tasnif.

Pereyonimlar – onomastik birliklar qatorida turuvchi leksik birliklardan. Ular o‘ziga xos lisoniy- semantik, milliy-madaniy, stilistik xususiyatlarda namoyon bo‘ladi. Badiiy matnlarda poetonim sifatida ham qo‘llanib keladi.

Dunyodagi barcha tillar tizimida pereyonimlarning o‘z o‘rni bor. Ular ko‘pincha qisqartma so‘zlar ko‘rinishida uchraydi. Bir tildan ikkinchi bir tilga o‘zlashma (olinma) so‘z sifatida kirib borib, uning lug‘at tarkibini boyitishga xizmat qiladi. Bunda ular bir tildan boshqa bir tizimdagi tilga kirib borar ekan, tarjima qilinmaydi va o‘z asliyati saqlab qolishi kuzatiladi.

Pereyonim deganda, transport vositalarining atoqli nomlari tushuniladi. Pereyonim yunoncha – “periov” so‘zidan olingan bo‘lib, “harakatlanish vositasi” degan ma‘noni anglatadi. Ular tilda alohida leksik tizimni hosil qiladi. Ya.I.Avlakulov pereyonimlarning turlarini o‘z izlanishlarida ko‘rsatib o‘tgan. [1,98–99]

Pereyonimlarni, avvalo, qayerda qo‘llanishiga qarab to‘rtga bo‘lib o‘rganish mumkin:

1. Quruqlikda yurishga mo‘ljallangan transport vositalarining nomlari: “Ikarus” avtobusi, “Kaptiva” mashinasi, “Vosxod” mototsikli, “Tigr” tanki.

2. Suvda suzishga mo‘ljallangan transport vositalarining nomlari: “Titanik” kemasi, “Avrora” kreyseri. Bu turdagi atoqli otlarga nisbatan yunoncha “nautuxos” – flot, “onoma” – nom so‘zlaridan olingan “nautonim” atamasi ham qo‘llanadi. “Brig”, “Karavella”, “Shlyupka”, “Frechat” kabi transport kemalari, harbiy kemalar nomi bunga misol bo‘la oladi.

3. Havoda uchishga moslangan transport vositalarining nomlari: “Kukuruznik” samolyoti, “TU-24” samolyoti.

4. Rels (temir izlar) ustida yuruvchi transport vositalari: “Afrosiyob” tezyurari, “Sharq” poyezdi.

Pereyonim onomastik birlik sifatida bu kabi transport vositalarining alohida bir tipiga nisbatan qo‘llanadi. Ko‘rinadiki, ular atoqli otlar turkumiga kiradi.

O‘zbek tilida pereyonimlarning o‘zlashtirilishi o‘tgan asrning ikkinchi yarmidan kuchaygan. To mustaqillik yillarigacha tilimizga “Pobeda”, “Moskvich”, “Zaporjets”, “Zil”, “Kamaz”, “Jiguli”, “Volga” (va hokazo) kabi pereyonimlar o‘zlashtirilgan bo‘lsa, istiqloq yillarida “Mercedes-Bens”, “Damas”, “Tiko”, “Nexia”, “Matiz”, “Lacetti”, “Spark”, “Gentra” “Maserati”, “Toyota-Prado”, “Maybach”, “Ferrari”, “Ford”, “Chevrolet Captiva” kabi yengil avtomobillarning

nomlari kirib keldi. Shuni ham aytish kerakki, ushbu avtomobillarning aksariyati yurtimizda ishlab chiqarildi.

Ularning til va nutqda qoʻllanilishiga davr ham taʼsir koʻrsatadi. Xususan, oʻtgan asrning 80-yillarida Vengriyada ishlab chiqarilgan “Ikarus” avtobuslari juda ommabop boʻlgan. Oʻsha paytlarda Yevropada ishlab chiqarilgan avtobuslarning yarmidan koʻprogʻini “Ikarus”lar tashkil etgan. Har yili oʻrtacha oʻn ikki ming dona avtobus ishlab chiqarilar va uning asosiy qismi SSSRga eksport qilinardi edi. Shu sababli sobiq ittifoqda, jumladan, Oʻzbekistonda shahar va shaharlararo yoʻnalishlarda, asosan, “Ikarus” avtobuslari xizmat qilardi.[6] Shunday ekan, bu pereyonimning oʻtgan asrning oxirida oʻta faol, hozirgi kunda nutq faoliyatida kam qoʻllanishi davr bilan bogʻliqligini dalillaydi.

Hozirgi paytda “Tiko” rusumli avtomobillar ishlab chiqarilmayapti. Ulardan aholi kam foydalanyapti, shuning uchun bu pereyonim nafaol soʻzga aylanib bormoqda. “Tracker”, “Epica” kabi avtoulovlar esa ayni kunlarda xaridorgir boʻlganligi sababli ularning nomi ham tilda faol qoʻllanishda. Bu esa pereyonimlarning tarixiy (diaxron) va zamonaviy (sinxron) qatlama egaligidan dalolat beradi.

Pereyonimlarni shakllantirishda uni kashf qilgan olimlar, konstruktorlar nomi yoki oʻsha transport vositasi ishlab chiqilgan zavod, shahar nomi asos qilib olingani koʻp kuzatiladi. Jumladan, “AN-24”, “An-124” passajir samolyotlari [7], “Il-76” harbiy-transport samolyoti [8], “Il-86” “Il-96” transport samolyotlari [9], “TU-104”, “TU-154” avialaynerlari, “Moskvich”, “Volga-24” avtomobillari, “Belarus”, “Altay” traktorlari, “Kamaz” yuk mashinasi kabilar buning yorqin misollaridir.

“Nexia” soʻzining maʼnosi “Qaldirgʻoch” degani. Hozirda uning Ravon Nexia, Chevrolet Nexia kabi turlari mavjud.

“Chevrolet Malibu” pereyonimi tarkibidagi “Malibu” soʻzi inglizcha boʻlib, AQSh shaharlaridan birining nomidir. U Kaliforniya shtatiga qarashli Los-Anjeles okrugi gʻarbida joylashgan Malibu shahri sharafiga nomlangan. Qayd etish kerakki, birinchi marta “Malibu” nomi 1962-yilda paydo boʻlgan.

Abbreviatsiya – pereyonimlar yasalişining keng tarqalgan usuli sifatida namoyon boʻladi. Baʼzi pereyonimlar metaforik koʻchma nomlar bilan parallel ishlatiladi. Masalan, “Chevrolet Captiva”ga nisbatan “Tartan oti” nomi qoʻllanadi.

Transportlar har xil sohalarda va turfa maqsadlarda qoʻllanishiga qarab ham turlanadi. Jumladan, ular madaniy turmushda shaxsiy avtoulov, jamoa transporti, qishloq xoʻjaligi va qurilishda, harbiy soha, tibbiyot texnikasi sifatida qoʻllanib keladi.

Xulosa shuki, dunyo tilshunosligida turli tizimdagi tillarning leksikoni, onomastikasini qiyosiy aspektda o'rganishga bag'ishlanayotgan tadqiqotlar salmog'i kundan kunga ortib bormoqda. Turli onomastik birliklar qatorida pereyonimlarning tilda leksik-semantik, milliy-madaniy, badiiy matnda esa lingvostilistik, poetonimlik xususiyatlarini, bir tildan boshqa bir tizimdagi tilga o'zlashish omillarini, tarjimada o'z asliyatini saqlab qolish sabablarini aniqlash hozirgi tilshunoslik ilmi oldida turgan dolzarb vazifalardan biridir.

#### ADABIYOTLAR:

1. Avlakulov Ya.I. O'zbek tilining onomastik ko'lami. – Toshkent: Fan va texnologiya, 2017. – B. 98–99. Avlakulov Ya.I. (2017). Onomastic degree of Uzbek language. – Tashkent: Science and Technology, – 138 p.
2. Begmatov E., Do'simov Z., Nafasov Z., Qorayev S. (1991). O'zbek nomshunosligi: tadqiq yo'shnalishi va usullari. – Xiva, – № 1, 12 b.
3. Begmatov E., Uluqov N.M. (2006). O'zbek onomastikasi terminlari izohli lug'ati. – Namangan, – 88 b.
4. Superasnkaya A.B. (1973). Obshaya teoriya imeni sobstvennogo. – Moskva: Nauka. – 159 p.
5. Sizranova G.Yu. (2013). Onomastika. – Tolyatti: TGU, – 248 p. 6.
6. <https://kun.uz/14836825>
7. <https://ru.wikipedia.org/wiki/%D0%90%D0%BD-124>
8. <https://ru.wikipedia.org/wiki/%D0%98%D0%BB-76>

УДК 81-11

**СЕМАНТИКА ОБРАЗОВ В ПРОИЗВЕДЕНИИ  
«МЫШЛЕННЫЕ ДНИ» АБДУЛЛЫ КАДИРИ**

***Н. А. Бафоева***

*Бухарский государственный педагогический институт*

*Бухара, Узбекистан*

*bafoevanigora@buxdpi.uz*

В данной статье статистически анализируются ойконимы являющиеся неотъемлемой частью топонимии романа Абдуллы Кадири «Мнувшие дни», и изучено семантика некоторых ойконимов.

**Ключевые слова:** топоним, ойконим, махалля, кишлак, лексика, статистический анализ.

**SEMANTICS OF IMAGES IN THE WORK “THOUGHT DAYS”  
BY ABDULLAH QADIRI**

***Bafoeva Nigora Azimovna,***

*Bukhara State Pedagogical Institute*

*Bukhara, Uzbekistan*

*bafoevanigora@buxdpi.uz*

This article statistically analyzes oikonyms, which are an integral part of the toponymy of Abdullah Qadiri's novel “Bygone days” and studies the semantics of some oikonyms.

**Keywords:** toponim, oykonym, makhalla, village, vocabulary, statistical analysis.

**ABDULLA QODIRIYNING “O‘TKAN KUNLAR” ASARIDAGI  
OYKONIMLAR SEMANTIKASI**

***Bafoeyeva Nigora Azimovna***

*Buxoro davlat pedagogika institute*

*Buxoro, O‘zbekiston*

*bafoevanigora@buxdpi.uz*

Mazkur maqolada Abdulla Qodiriyning “O‘tgan kunlar” romani toponimiyasining ajralmas qismi hisoblangan oykonim birliklar statistik tahlil etilgan va semantikasi misollar asosida yoritilgan.

**Kalit so‘zlar:** toponim, oykonim, mahalla, qishloq, mavze, leksika, statistik tahlil.

Onomastik leksika til lugʻat tizimidagi mavjud birliklarning bir qismini tashkil etadi. Toponimlar onomastik leksikaning alohida bir qatlamini tashkil qilsa, oykonimlar toponimlarning alohida bir leksik qatlami sanaladi.

Hududlarning aholi istiqomat qiladigan, joylarning atoqli otlari sanaladigan oykonimlar kishilik jamiyati, uning tarixi va rivojlanishi bilan uzviy bogʻliqdir. Aholi yashash joylariga nom berish kundalik hayot tarzimizning zaruriy ehtiyojlaridan biridir. Oykonimlar oʻzi taalluqli boʻlgan hududning turli jihatlarini aks ettiradi. Bunda asosiy eʼtibor aholining etnik tarkibi, kasb-kori, hududning geografik oʻziga xosligi, relyefi, unga yondosh boʻlgan boshqa obyektlar va hokozolar-ga qaratilgan boʻladi [1; 98].

Murakkab bir tizim sifatida oykonimlar guruhi oʻz tarkibiga oʻlka, viloyat, mamlakat, shahar, posyolka, qishloq, mahalla, guzar, qalʼa, ovullarning nomlarini qamrab oladi. Ishimiz obyekti hisoblangan “Oʻtgan kunlar” asarida koʻchalar, mavzelar, dahalar, mahallalar, qoʻrgʻonlarning nomlarini uchratish mumkin [2].

“Oʻtgan kunlar” asarida aniqlangan oykonimlarning aksariyati qadimiy boʻlib, ularning ayrimlari bir necha asrlik tarixga ega.

Mahalla nomi: *Toshkandda...mahallasi, B...mahallasi, Ch... mahallasi.*

Mavze nomi: *Ming Oʻruk, Yormozor, Oqtepa.*

Daha nomi: *Beshyogʻoch, Sebzor, Koʻkcha, Laylakxona.*

Qoʻrgʻon nomi: *Koʻkcha.*

Ularning semantikasiga quyidagicha izoh beramiz:

### **1. Tarkibida miqdor bildiruvchi sonlar mavjud boʻlgan nomlar.**

Asarda miqdor bildiruvchi sonlar mavjud boʻlgan oykonimlar-ga Ming oʻrik va Beshyogʻoch nomlari misol boʻla oladi. Toponim qatlamidagi bu nomlarga yuqorida izoh berib oʻtdik.

### **2. Tarkibida narsa-predmet nomlari mavjud boʻlgan oykonimlar.**

**Eski Joʻva** – bozor va mahalla. Bu nomda asosan ikkinchi asos ustunlikka ega boʻlib oʻz etimologiyasiga ega. H.Hasanovning fikriga koʻra, bu yerda qurol-yarogʻlar ombori joylashgan – juva (xususan, **jiba** – “qurol”, “ovchilar qurollari”) va shu yerdan mahallaning nomi ham kelib chiqqan [3; 40].

Buxoro xonligida *jibachi* – “ovchi” shu tarmoqdagi toʻrtinchi amaldor deb hisoblangan. Tarixiy manbalarda *juba* – bozor maydoni, bozor rastalari sanalgan. Buxoroda oʻrta asrlardan *Jubayi baqqolon* – “Baqqollar bozori” boʻlgan.

**Laylaxona** – qishloq. Asarda daha. (*Rahimbiy Qozoqboybiy o'g'li Beshyog'och daha, Laylaxona dahalik.*) Surxondaryo viloyatining Sho'rchi tumanida shu nomdagi joy bor. *Laylak* + *xona* affiksi (aynan “uy”, “xona”), toponimlar tarkibida tez-tez nimanidir ko'pligini ko'rsatadi. **Bo'rixona** – “*bo'rilar ko'p joy*”, **Qulonxona** – “*qulonlar ko'p joy*”, **Kaptarxona** – “*kaptarlar ko'p joy*”, **Laylaxona** – “*laylaklar ko'p joy*”.

**Ko'kcha** – Toshkent shahridagi mahalla nomi. Mahalla nomi Ko'kcha darvozasi va shu nomdagi qabila bilan bog'liq. Muhammad Solihning yozishicha, Ko'kcha darvozasining nomi qabila nomidan olingan, ya'ni ko'kcha qabilasidan. Tarixchi olim A.R. Muhammadjonovning fikriga ko'ra, “**ko'kcha**” so'zi aslida “*kohicha*” so'zining o'zgartirilgan shakli, “*kohicha*”, “*kohak*” so'zi “*qal'acha*” degan ma'noni anglatadi va bu yerda o'rta asrlarda harbiy qal'a bo'lganidan dalolat beradi. Mahalla aholisi asosan ko'nychilik, kosibchilik va to'qimachilik, mavzelerde dehqonchilik bilan shug'ullangan. Shahar tashqarisida mavzeleri bo'lgan.

Ushbu nom respublikamizning boshqa joylarida ham uchraydi [4; 225].

**Yormozor** – asar tarkibidagi mavze.

Toponim ikki so'zdan: *yor* va *mozor* so'zlaridan kelib chiqqan. Toponim tarkibidagi *yor* so'zi “*jar*” ma'nosini anglatishini bundan qariyb X asr oldin M.Koshg'ariy aytib o'tgan edi. “*Jar*” so'zining “*yor*”, “*yar*” shaklida talaffuz etilishidan Farg'ona vodiysida Yorboshi, Yortepa, Yorqatoy, Yorqin, Yorqigiloq, Yorqo'rg'on, Oqjar, Oqyor, Kuygan-yor, Ko'kyor, Yortagi, Yorko'cha kabi joy nomlari vujudga kelgan.

Mozor – O'zbekistonda toponimlar tarkibida ko'p uchraydigan termin. Bu so'z arabcha bo'lib, lug'aviy ma'nosi “*sig'inish joyi*” demakdir. Joy nomlarini tanlashda yaxshi mo'ljal vazifasini bajarib, saqlanish vaqti uzoq bo'lganligi uchun ko'p nomlar tarkibida mazkur so'z uchraydi. Masalan: *Arabmozor, Buvimozor, Ko'kmozor, Mozorboshi, Mozortagi, Oqmozor, Sarmozor, Terakmozor, Tolmozor, Xurmozor* va boshqalar [4; 534].

Asardagi ushbu parchaga e'tibor qaratadigan bo'lsak, muallif Yormozor mavzesidagi o'rda haqida keng izoh berib o'tganiga guvoh bo'lamiz.

### 3. Belgi-xususiyat asos bo'lgan nomlar.

*Oqtepa* – balandlik joyda bunyod etilgan qadimgi qo'rg'on o'rni. **Oq** so'zi bu o'rinda ham *katta, baland, yuksak* ma'nosiga ega. *Oqtepa* – yon-atrofida shu tipdagi boshqa balandliklardan hajman katta bo'lgan tepa.

Madaniy qatlamlarning qoldiqlari va feodal dehqonlar o'sha paytlari ilk o'rta asrlar qal'a qurilishlari buzilgan joylar ham shunday

nomlangan. Demak, yaqin-yaqinlargacha Toshkent shahrining o‘zida sakkizta Oqtepa bo‘lgan.

#### 4. Sir saqlash bilan bog‘liq mahalla nomlari.

“Sir saqlash” badiiy usulini Abdulla Qodiriy ham o‘z tarixiy romanlarida juda o‘rinli, mahorat bilan qo‘llagan. Asarning “Isitma orasida” bobida maktublarni almashtirish jarayoni aks etgan tasvirlarda muallif hatto joy nomlarini ham sir saqlaydi. Bu orqali Homid boshchiligidagi manfur guruhning noqobil ishlari hech kim yo‘q, kimsasiz bir joyda ro‘y berayotganini kitobxonga ruhiy tasvirlar orqali ta’sir etadi. Bu orqali zukko yozuvchi kitobxonni keying voqealar rivoji qaysi joyda davom etishiga qiziqtiradi. Ayni paytda hozir bo‘layotgan voqealar joyi Toshkentning qayerida sodir bo‘layotganiga kishida qiziqish uyg‘otadi. Bilishga harakat qilganing sari asardagi kichik – kichik detallarga ham qiziquvchanlik bilan e’tibor qaratishdan boshqa chora qolmaydi.

*Saratoning ikkinchi kuni... Marg‘ilonning shimolida bo‘lg‘an B... mahallasining kunchiqar tomonida janubga qarab burulgan tor ko‘chaning... [5; 190].*

“Jodugar Hindi” bobidagi mahallada nomning o‘zi ham yo‘qligi Otabekni xatning haqiqatdan ham sirli ekanligiga shubha uyg‘otadi. Bu xat bugungi kundagi “yumaloq xat”larga ishora qilingandek go‘yo.

*Otabek maktubning unvoniga ko‘z tashladi: «Toshkanda... mahallasida turg‘uchi Otabek Yusufbek hoji o‘g‘lig‘a tegsin». Kim tomondan yozilganligi unga ma’lum bo‘lib, qo‘lidagi kitobni toqchaga qo‘ydi va shosha-pisha xatni ochdi [5; 178].*

“Baxt va baxtsizlik” bobidagi Usta Alim hikoyasida Saodatning mahallasi nomini sir saqlashida ayollarga xos pishiqlik va ayni damda odob doirasi saqlash ma’nosi ifodalanadi.

*Qiz menga ishonmay boshlag‘an edi:*

*– Menga ishonasizmi?*

*Men ishonch bildirib o‘lturmadim:*

*– Uyingiz qayerda?*

*– Ch... mahallasida.*

*– Kimning qizisiz?*

*– Sharif muliqchining qiziman [5; 212].*

Shu tariqa Abdulla Qodiriy asaridagi “sir saqlash” usuli kitobxonni voqealar rivojiga qiziqtirish, uni voqea tizimiga tezroq jalb etishga xizmat qilgan.

Umuman olganda asar tarkibida mavjud oykonimlar asar semantik-uslubiy xususiyati, badiiyatini oshirishga xizmat qilgan.



Toponimlar til lugʻat tarkibining ajralmas qismlaridan biri sifatida maʼlum oʻzgarishlarga ham uchrab turadi. Toponimiya jamiyatning ijtimoiy hayoti, mafkurasi va maʼnaviyatiga yaqin aloqada boʻlgan lugʻaviy qatlam boʻlgani uchun jamiyatning siyosati, xalqning dunyoqarashida aks etuvchi yangilanish va oʻzgarishlar mavjud joy nomlarini doimo baholab turishga olib keladi. Bunga yaqqol misol sifatida Oʻzbekiston davlatining mustaqillik davrida olib borayotgan toponimik siyosatini keltirish mumkin [6; 33-34]. Bu oqilona siyosat tufayli Oʻzbekistondagi joy nomlari tarkibida tubdan oʻzgarishlar sodir boʻldi. Shoʻro davri gʻoyalari bilan uzviy bogʻliq boʻlgan koʻpgina nomlar isteʼmoldan chiqdi, aksincha, kommunistik mafkura tazyiqi ostida siqib chiqarilgan, unutilib boʻlingan koʻpgina oʻzida tarixiy, milliy ruh saqlagan oʻzbekona, xalqona nomlar qayta tiklandi.

Abdulla Qodiriyning “Oʻtgan kunlar” romani toponimiyasi yuzasidan olib borilgan kuzatishlarimiz toponimlar, asosan, affiksatsiya, sintaktik hamda onomastik konversiya usulida yasalgan.

Tadqiqodimiz jarayonida “Oʻtgan kunlar” romani toponimik qatlamida -ligʻ (-lik), -iston yasovchi koʻmakchi morfemalarini uchratdik va asarda **affiksatsiya usuli** orqali nom yasash xarakterli emasligining guvohi boʻldik.

Birdan ortiq asosni qoʻshish orqali yangi soʻz hosil qilish soʻz yasashning **sintaktik usuli** deb yuritiladi. Ilmiy manbalarda ushbu usul sintaktik usul, soʻz qoʻshish, negizlarni qoʻshish, analitik usul, morfologik-sintaktik usul, kompozitsiya usuli deb ham nomlanib kelinmoqda. Qoʻshma toponimlarning toponegizi birdan ortiq soʻzdan tashkil topgan tufayli, bu negizni tashkil etgan komponentlar toponimlikkacha boʻlgan maʼlum sintaktik munosabat asosida uyushgan boʻladi. Buning asosiy koʻrinishlari quyidagilar:

1) sifatlovchi va sifatlanmish (bitishuv) munosabatidagi toponegizlar: *Oqtepa, Eski Joʻva, Eski Nomoz*;

2) son va ot munosabatidagi toponegizlar: *Ming Oʻrik va Beshyogʻoch*;

3) ot va ot munosabatli topogenezislar: *Laylakxona, Yormozor, Xoʻja Maʼoz, Avliyo Ota*;

Baʼzan toponim tarkibida uch qatnashganligiga guvoh boʻlish mumkin: *Shayxontohur-Shayx*”+ “*Xovand*”+ “*Tohur*”;

Qoʻshma toponimlarning toponegizlari, yaʼni komponentlari moshlashuv, bitishuv, boshqaruv yoʻli bilan birikadi.

Bir obʼekt atoqli otini ikkinchi yoki bir necha obʼektni nomlash uchun koʻchirish onomastik tizimda keng tarqalgan usuldir. Atoqli otlarning

vazifaviy jihatdan mana shu kabi ko‘chirilishi (ko‘chishi)ni A.V.Superanskaya o‘z ishlaridan birida onomastik konversiya deb nomlaydi.

So‘zlarning nominatsion, ya’ni funktsional ko‘chishi tildagi keng tarqalgan hodisadir. Bunda bir so‘z turkumiga mansub leksemaning boshqa so‘z turkumi vazifasiga ko‘chishi ko‘zda tutiladi. Lekin bu jarayon shunchaki almashinish emas, balki yangi so‘z yasash usullaridan biri deb qaraladi: “Konversiya – biror lug‘aviy asosni so‘z hosil qilishning boshqa paradigmasiga o‘tkazib, yangi so‘z yasashdir” [7; 92-93].

Konversiya so‘z yasashning leksik-semantik usuli yoki leksik-grammatik usuli deb yuritib kelinadi. Oddiy so‘zlar, ya’ni turdosh leksikaning konversiya usulida yasalishini atqli otlarning shu usulda yasalishidan farqlash maqsadida atqli otlar tizimidagi konversiya – onomastik konversiya deb yuritiladi.

Turdosh so‘zning otga ko‘chishi birinchi bosqich konversiya bo‘lsa, atqli otning boshqa atqli ot vazifasiga o‘tishi ikkinchi bosqich konversiya hisoblanadi. Ikkinchi bosqich konversiya, ya’ni atqli otning boshqa atqli ot vazifasiga o‘tishi usulini juda ko‘p kuzatdik.

Toponim va oykonimlar qatlami asar tarkibiy qismining antroponimlardan keyimgi ikkinchi o‘rinni egallovchi onomastik birlik bo‘lib, ular o‘ziga xos yasaliş va uslubiy- semantik xususiyatga ega.

#### ADABIYOTLAR:

1. Bayramali Ergashevich Kilichev A way of making words on proper nouns. International Journal on Integrated Education. Volume 3, Issue II, Feb 2020, pp. 96–98.
2. Kilichev B.E., Bafoyeva N. Some characteristics of onomastic units in Abdullah Qodiriy’s novel “Bygone days”. Asian Journal of Research in Social Sciences and Humanities. ISSN: 2249-7315 Vol. 12, Issue 01, January 2022.
3. Ҳасанов Ҳ. Ўрта Осиё жой номлари тарихидан. – Т., 1965. – Б.40.
4. О‘zbekiston joy nomlarining izohli lug‘ati. I nashr. – Toshkent, 2022. – B. 225.
5. Абдулла Қодирий «Ўтган кунлар» романи. «Адабий барқарор асарлар туркуми». – Т., 2019. – Б.190.
6. Dildora Yusupova, Dilorom Yuldasheva. Halima Xudoyberdiyeva she’riyatı lingvopoetikasi.– Germaniya: “Globe Edit international Publishing Group” ISBN: 978-620-0-60879- 6. – 2020. – 125 p.
7. Бегматов Э. Жой номлари – маънавий кўзгуси. – Тошкент, 1998. – Б. 33–34.
8. Суперанская А.В. Структура имени собственного. Фонология и морфология. – М., 1969. – С. 92–93.

УДК 81-11

**GENDER INEQUALITY AND VIOLENCE: A COMPARATIVE ANALYSIS OF KOREAN AND UZBEK CULTURES*****Bakhronova Dilrabo<sup>1</sup>, Virna Velázquez<sup>2</sup>, Dilnavoz Khamidova<sup>3</sup>, Umida Djalilova<sup>4</sup>****<sup>1</sup>DSc, professor of Uzbekistan State World Languages University  
<https://orcid.org/my-orcid?orcid=0000-0002-2012-7426>**<sup>2</sup>Professor of Faculty of Languages,  
Autonomous University of the State of Mexico, Toluca,  
México <https://orcid.org/0000-0002-5630-9360>**<sup>3</sup>Tashkent State University of Oriental Studies**<sup>4</sup>Tashkent State University of Uzbek language and literature  
[dilrabobakhronova@gmail.com](mailto:dilrabobakhronova@gmail.com)*

Gender-related concerns have been significant throughout history, persisting across various societies. The mental and physical harassment of women, unequal pay, and the preferential appointment of men to high-ranking positions have persisted over time. In contemporary democratic societies where women's dignity and rights are safeguarded, their status is now on par with that of men. However, issues such as sexual violence, the degradation of women, and mental and physical abuse remain alarmingly prevalent. Even in the 21st century, sexual discrimination and the humiliation of women continue to be pressing issues in highly developed countries. For instance, according to reports, a woman is killed by her husband every four days in the UK.

**Key words:** women in the Timurid dynasty, polygamy, Cape reforms, Japanese annexation, Confucianism, cyberbullying, gender violence in literature.

**ГЕНДЕРНОЕ НЕРАВЕНСТВО И НАСИЛИЕ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ КОРЕЙСКОЙ И УЗБЕКСКОЙ КУЛЬТУРЫ*****Бахронова Дилрабо<sup>1</sup>, Вирна Веласкес<sup>2</sup>, Дильнавоз Хамидова<sup>3</sup>, Умида Джалилова<sup>4</sup>****<sup>1</sup>Узбекского государственного университета мировых языков  
<https://orcid.org/my-orcid?orcid=0000-0002-2012-7426>**<sup>2</sup>Профессор факультета языков Автономного университета штата Мехико, Толука, Мексика  
<https://orcid.org/0000-0002-5630-9360>**<sup>3</sup>Ташкентского государственного университета востоковедения**<sup>4</sup>Ташкентского государственного университета узбекского языка и литературы**[dilrabobakhronova@gmail.com](mailto:dilrabobakhronova@gmail.com)*

Гендерные проблемы были значительными на протяжении всей истории и сохранялись в различных обществах. Психическое и физическое притеснение женщин, неравная оплата труда и предпочтительное назначение мужчин на руководящие должности сохраняются с течением времени. В современных демократических обществах, где достоинство и права женщин защищены, их статус теперь приравнивается к статусу мужчин. Однако такие проблемы, как сексуальное насилие, деградация женщин, а также психическое и физическое насилие, остаются тревожно распространенными. Даже в XXI веке сексуальная дискриминация и унижение женщин продолжают оставаться актуальными проблемами в высокоразвитых странах. Например, по имеющимся данным, в Великобритании каждые четыре дня муж убивает женщину.

**Ключевые слова:** женщины династии Тимуридов, многоженство, капские реформы, японская аннексия, конфуцианство, киберзапугивание, гендерное насилие в литературе.

## JINSIY TENGIZLIK VA ZO'RAVLIK: KOREYA VA O'ZBEK MADANIYATINING QISOSLI TAHLILI

*Baxronova Dilrabo<sup>1</sup>, Virna Velaskes<sup>2</sup>, Dilnavoz Xamidova<sup>3</sup>,  
Umida Jalilova<sup>4</sup>*

*<sup>1</sup>O'zbekiston davlat jahon tillari universiteti  
Toshkent, O'zbekiston*

*<https://orcid.org/my-orcid?orcid=0000-0002-2012-7426>*

*<sup>2</sup>Meksika shtati Toluka avtonom universiteti tillar fakulteti professori,  
Meksika <https://orcid.org/0000-0002-5630-9360>*

*<sup>3</sup>Toshkent davlat sharqshunoslik universiteti ilmiy xodimi*

*<sup>4</sup>Toshkent davlat o'zbek tili va adabiyoti universiteti ilmiy xodimi  
[dilrabobakhronova@gmail.com](mailto:dilrabobakhronova@gmail.com)*

Gender muammolari tarix davomida muhim bo'lgan va turli jamiyatlarda saqlanib qolgan. Ayollarga nisbatan ruhiy va jismoniy tazyiqlar, tengsiz maosh va erkaklarning rahbarlik lavozimlariga imtiyozli tayinlanishi vaqt o'tishi bilan davom etib kelmoqda. Xotin-qizlar qadr-qimmatini va huquqlari himoyalangan zamonaviy demokratik jamiyatlarda ularning maqomi endi erkaklarnikiga teng. Biroq, jinsiy zo'ravonlik, ayollarning tanazzulga uchrashi, ruhiy va jismoniy zo'ravonlik kabi muammolar keng tarqalgan bo'lib qolmoqda. 21-asrda ham ayollarni jinsiy kamsitish va tahqirlash yuqori rivojlangan mamlakatlarda dolzarb muammolar bo'lib qolmoqda.

**Kalit so'zlar:** Temuriylar ayollari, ko'pxotinlilik, Keyp islohotlari, Yaponiya anneksiyasi, konfutsiylik, kiberbulling, adabiyotda jinsga asoslangan zo'ravonlik.

Despite considerable efforts, gender inequality remains a persisting challenge for both Korean and Uzbek women, awaiting resolu-

tion. Historically, in both nations, preserving the dignity and chastity of women within the family has been paramount, often resulting in restrictions on their freedom to venture out alone. In ancient Korea, for example, girls were only permitted to leave their homes on three occasions in their lifetime: when they got married, when their husband passed away, or when they themselves passed away. Korean women have traditionally been excluded from public life, thereby denied the opportunity to hold public office or earn additional income. Often, the perspectives of Korean and Uzbek women were disregarded, and their consent was not sought in marriage arrangements. It was not uncommon in the histories of both nations for girls from low-income families to be sold to wealthy nobles (boylar/양반 – yangban) in exchange for substantial sums of money or land.

During Korea's Choson Dynasty, the number of women working as 기생 – *kiseng* increased. *Kisengs* were involved in entertaining affluent patrons at pubs and brothels, performing at grand events, and even engaging in espionage. Intriguingly, among the lower classes, *kisengs* were the only literate women, highlighting their specialized training. Centuries later, with the introduction of the 'Kabo Reform' in Korea, the *kiseng* profession was entirely abolished. However, the arrival of Japanese troops ushered in a dark era for the Korean populace, particularly women and girls, who were subjected to enslavement, forced labor, and sexual exploitation by Japanese soldiers. These grim circumstances persisted until the end of Japanese colonization, resulting in years of suffering and humiliation for Korean women. As U.S. forces liberated Korea, the nation underwent industrialization, with both men and women working in factories on an equal footing.

In Uzbekistan, collective farming commenced in the 1930s and 1940s, with women laboring in collective farms and factories under challenging conditions. During the 1930s in Uzbekistan and the 1950s in Korea, both men and women were sent to Germany for education aimed at rebuilding their respective countries. Tragically, enlightened women returning from Germany to Uzbekistan faced repression under the Soviet government. In Uzbekistan and Korea, women began assuming roles in the public sector, occupying various high-ranking positions. The country witnessed a surge in female poets and writers who championed feminism. Gender equality moved to the forefront of societal concerns, leading to women's participation in government and administration, including presidential roles in various associations.

However, contemporary challenges persist, as the active involvement of women in economic and political processes has given rise to issues such as sexual and cyberbullying. A 2015 in Korea survey revealed that eight out of ten female workers experienced some form of sexual harassment, often occurring during corporate dinners (회식). Alarmingly, the study found that most victims chose not to report harassment for fear of job loss or workplace discrimination. In today's culture, praising or critiquing women's appearance or attire is viewed as a form of violence. Furthermore, cyberbullying among young individuals, including the distribution of pictures and videos captured through hidden cameras on social platforms, has become increasingly commonplace.

This article delves into the artistic interpretations of the treatment of women in Korean and Uzbek cultures, shedding light on the historical evolution of gender dynamics and their portrayal in literature.

## INTRODUCTION

History has shown that respect women and listening to their opinions in the family and in society are inherent in the culture and traditions of Eastern countries. In particular, during the reign of the Temur dynasty, the great commander Amir Temur had 18 wives and Saraymulkhanum was considered as the greatest of all princesses and was awarded the title of "Great Lady" or "Bibikhanum". As Bibikhanum was highly intelligent and enterprising woman, she accompanied Amir Temur on many military campaigns and achieved the rank of Great lady. Amir Temur relied on Bibikhanum's advice in conducting the affairs of the kingdom, in ruling, in solving the problems that arose in the management of the kingdom. Bibikhanum was allowed to attend banquets, drink wine, talk to palace guests, and even receive ambassadors in tents, which set her apart from other women in terms of status (Fayziyev, 1994). She was also allowed to meet Amir Temur from the battles. Bibikhanum could sentence to death or save from death. Since Bibikhanum was able to influence Amir Temur's attitude and decisions, the palace officials tried to establish a good relationship with her. In addition, women of the Timurid period made a great contribution to the development of science and construction. Although the Sufi khanaqah was a favorite type of building patronized by Timurid women, it was far from the only type of religious building they funded. We know from architectural evidence and historical records of the time

that Timurid women's engagement in shaping the religious landscape of the medieval period was more expansive in scope. So, for example, Timurid women were engaged in developing Islamic education through their patronage of madrasas. Similarly, Gawhar Shad, the most famous Timurid woman, supported normative Sunni Islam by funding the building of two major Friday mosques. As already mentioned, other Timurid women of the ruling class helped the solidification of Sufi Islam through their patronage of khanaqahs and mausoleum shrines. It is equally fair to say that as a side effect of the even-handed distribution of their patronage, Timurid dynastic women fulfilled a diplomatic function, creating stability for the court by balancing out the power of rival religious establishments. We should not, however, conclude from our Timurid example that the patronage of alternative religious buildings such as khanaqahs and madrasas was an activity exclusive to the women of this particular dynasty, because across the centuries, the ruling-class women of various Islamic dynasties engaged in the same activity, patronizing alternative religious buildings (Arbabzadah, 2016). Relying on historical facts mentioned above, in the East women were honored and respected, they had a certain role in government. However, the rights mentioned above were given only for some women, and in general, Uzbek and Korean women were mainly engaged in household chores and educating children. From an early age, they were trained only to do housework well, and from a very young age, there were many cases of marriage without their consent. Only daughters of wealthy family were trained by hiring special teachers.

It can be said that the study of anthropocentric interpretation in various linguistic phenomena, as well as the interpretation of male and female images in the nominations of gender stereotypes, is aimed at the manifestation of ethnocultural qualities inherent in a particular nation (personality) (Bakhronova 2015, Bakhronova and all 2022). Below, we substantiate the gender issue through examples from Uzbek and Korean literature and historical works.

In the ancient national culture of Korea, the status of women was low in society during the reign of Three kingdoms, and this was commonplace. But for the first time, during the period of Silla kingdom attitude toward women changed. During the Silla period, women had the right to ascend the throne which was called songol<sup>11</sup>. According to songol, Queen Sondok was the only woman to ascend the throne.

---

<sup>11</sup> Songol: (Korean: □□) the upper class had the right to ascend to the throne during the Silla period.

During the Koryo period, equality of women in the family was recognized and their status in society was adequately protected. According to the *hojok* (호적 – family tree) belonging to this period, the woman gained the right to be head of the family and inherit after the death of her husband. This shows that there was a big difference in the situation of Indian women whose husbands were died (i.e. in India women were burned along with their dead husbands). In Korea, if a marriage was annulled, the woman had the right to receive all the property that she had owned in her husband's house. The fact that a husband's violence against a woman sometimes led to the dissolution of a marriage indicates that the status of women in Korea was relatively good. The children of women from their second marriages were not excluded from society. During the Koryo period, sons and daughters were included in the family tree without separation, and the inheritance from their parents was equally distributed. Because of women had a freedom during that time, they could even ride horses outside.

However, during the Choson dynasty, the role of women in society was severely curtailed and the status of women in the family began to be considered lower than that of men. Such views were based on some ethical principles of Confucianism. Women could not even go out at any time, let alone their career. In Korea, from 1477, women's divorce their husbands was also prohibited. It was forbidden for a woman to talk with her children if she asks for a divorce. As the country encouraged people who followed the rules of Confucianism, it was commonplace that women committed suicide or did not remarry after the death of their husbands. The inability of women to remarry after the death of their husbands, and the strong psychological or physical pressure exerted by society on remarriage, continued until the Cabo Reforms of 1894 were put into practice. Particularly, one of the Korean writers, O Yong Su, in his book called "Seaside Village" arose issues such as women had to work in difficult conditions, widows should have spent their whole lives alone, and that they should have not remarry. Chu Yo Sup's "The Houseguest and My Mother" is ideologically similar to this work.

We can see that women were forced to do the same heavy works as men with housework and child rearing, which negatively affected their morale in the works of Uzbek writers as well. Togay Murad's novel "Otamdan qolgan dalalar (The Fields Left by My Father)" is a case in point. The novel reveals the hardships of the Uzbek people during the former Soviet regime, and the fact that women were involved in the



same hard work as men. The protagonist, Dehkonkul, was engaged in growing cotton in the fields, and his wife had to work in the fields from early morning till night in difficult conditions, enabling to devote enough time to household chores and child-rearing.

The fact that the farmer's wife, despite her young age, has wrinkles on her face, no time to look at herself, and become very nervous is a testament to how much women suffered at that time. She can not endure such hardships and humiliation and eventually burns herself. The situation that suppressed by authorities which even blamed women as crazy, was considered as a manifestation of a rotten and savage society. Mikhail Dudin's poem, *Women are burning themselves in Uzbekistan, it is impossible to write in our country*, was called as "a work against to the method of socialist realism" by the governor. During the Soviet era, the deaths of men at the front and in the army led many women to become widows in their early age, to spend their entire lives without remarriage, and to raise their children alone.

Another Uzbek writer, Abdullah Qadiri's novel "Last Days", discusses the issue of gender in a *qo'shxotinlik* (polygamy). The protagonists of the novel, Otabek and Kumush, marry with love, but mother-in-law do not like Kumush as she come from village. She looks for a bride from city and marries a girl named Zaynab to his son as a second wife without the consent of Otabek and Kumush. Eventually, Zaynab, understanding that she is between two lovers, poisons Kumush and goes crazy. In this novel the tragedy of women in the early 19th and early 20th centuries, issues such as marrying without the consent of the girls, marrying as a second wife and discriminating as urban and rural were revealed.

Abdullah Qahhor's story called "Dahshat (Horror)" is another example of the problem of polygamy. The main hero of this story, Olimbek Dodho, who is in his 70s, has eight wives, the youngest of them, Unsin, is described as a young bride who has married only five months. Unsin wants to go home of her parents' and asks her husband's permission, saying that she is ready to any task. The task is to go to the cemetery, light a fire, make tea and return home. Unsin, who really wants to go home, agrees and goes to the cemetery, despite the terrible wind. She tries to find the strength to fulfill the task and overcome her fear, but she really fears when the monkey climbs on her shoulder, which is sent by Olimbek Dodho, as she thinks that it is devil. Unsin, who fulfills the task in fear and panic, returns home, but dies before dawn. In this short but meaningful story, the author openly describes the age-old

problems of such Uzbek women, and the fact that a young woman's life is ruined because of a cruel husband. Unfortunately, such kind of cases were common among rich families in ancient times.

From 1939, Koreans began to be sent to Sakhalin Island, Japan, and the Japanese colonies in Southeast Asia, first on a recruitment basis and then as forced labor. In 1942, a decree was issued requiring Korean youth to receive military training in special places. Beginning in January 1944, college students were called up for military service. The girls and young women were sent to the front, where they were enslaved by Japanese soldiers and officers for sexual purposes. During this period, about 180,000 Korean women became victims of such cases. In addition, during the Japanese colonial period, many young women were forcibly removed from their homes, enslaved, and tortured by Japanese soldiers. Women who lost their fathers, husbands, or siblings in the war had to make money for themselves to be alive and even earn money through prostitution. In particular, in the book named "Potatoes" written by Kim Donin in 1921, exposed a 15-year-old girl from a poor yangban family is sold to an elderly man for only eighteen won, as her husband is lazy and work nowhere, they suffer from poverty, starvation for several years, and eventually find the easy and ultimately way to make money that became common among women at the time – prostitution. The protagonist, Ponnyo, do not want to serve because he come from a wealthy family, and she can not earn much by begging. One summer day, city officials invited Ponnyo and the women in her neighborhood to work to manually clean the silkworms. But Ponnyo does not satisfy from this work, hence she earned only forty-two chons. One day the company's supervisor invites Ponnyo to have intercourse with him and she agrees. As a result, she begins to earn money, just like other women in a society where adultery and crime are rampant. Although her husband is aware of his wife's actions, which she is earning money through having sexual relations with men, he does not pay attention for these things because he is full and has no worries. One day, when Ponnyo picked up a basket of sweet potatoes from the yard of the Chinese Wang without permission, he catch her and offered to have sex with him in exchange for sweet potatoes. Ponnyo returns home with a basket of sweet potatoes and three won. In this way, the Chinese Wang would often visit Ponnyo's house. When the Chinese Van comes, Ponnyo's husband will leave the house, and when he left, he will return home and enjoy the money left over for his "wife's service". When Ponnyo hears that

Van is getting married, he goes to Van's house and threatens him with a sickle. She herself dies from this crooked Chinese sickle. To cover up his actions, Van closed the case by giving Ponnyo's husband thirty won and the doctor twenty won. Ponnyo is buried in a cemetery with a diagnosis of cerebral hemorrhage.

During the Japanese annexation, many girls were abused, both mentally and physically, due to poverty. The tragic situation of women's status and inequality are also described in Ke Yongmuk's story "Adada" written in 1935. Adada, whose real name is Hwaksiri but can not say anything apart from "A-da-da" because of her dumbness, constantly beaten, insulted and humiliated by her parents. When she is a girl only under the age of 20, her parents marry her for poor man. The young man, who is married to Adada, gets fertile land for marrying to Adada and respects her at the beginning. However, once he gets good harvest and becomes rich, he starts to mock Adada for her dumbness and beat her every day. Tired of the beatings, Adada runs away one day and goes home, but poor Adada suffers from her mother's torture at home. Exiled from her parents' house, Adada comes to Suron's house and begins to live there. As soon as they move to Sinmido Island and start a new life there. Suron shares his plans getting new land for money, which he saved for his marriage. Upon hearing this, Adada terrifies of what happened in her previous life and fears that if her husband buys the land and becomes rich with it, he will beat her like her ex-husband and drive her out of the house. So one day while Suron is asleep, she throw all the money into the sea. As Suron knows it, he kicks Adada and then throws her into the sea from a cliff. From the works that realistically express the real way of life, pain and suffering of the people, we see that the women of both nations have been living in inequality for many years, enduring spiritual and physical oppression and injustice. However, over the years, as the two countries have gained independence and reforms, women's dignity, rights, and security have been protected and many opportunities have been created for them.

By the 1980s, female writers began to appear in Korean literature to promote feminism. Lee Xi Jo, In Xi Gyon, Yang Kwi Ja, Park Wan Suh, and Shin Kyon Suk promoted the ideas of early feminism through their works. In Xi Gyeong's "My Wife's Box", Lee Xi Jo's "Liberty Bell" and Kong Jiyong's "Mackerel" are some of the most famous works written in the spirit of feminism.

## CONCLUSION

Even in today's modern world, it is difficult to say that all gender issues have been resolved. Particularly, intimate partner violence (IPV) is a worldwide public health problem as well as a serious social problem in South Korea. IPV is characterized as any behaviour within an intimate relationship that causes physical, psychological, or sexual harm of one partner to another. According to the 2010 national survey in South Korea, the prevalence of reported IPV was 53.8%, and 81.9% of this violence was perpetrated by husbands against their wives. Women were significantly more likely than men were to report being a victim of IPV (verbal: 28.2% vs. 24.4% [ $p < .0001$ ], physical: 6.9% vs. 3.4% [ $p < .0001$ ]). In addition, women tended to perpetuate verbal violence against their spouse more often than men were (26.7% vs. 25.3%). However, 5.1% of men and 3.4% of women reported perpetrating physical violence against their spouses. A similar proportion of men and women reported male-to-female violence, yet more women than men reported female-to-male violence. What is more, the distribution of videos and photos shot by cameras in lavatories has been increasing among young people. This condition can lead girls to feel under depression or suicidal. A similar situation occurs in Uzbekistan, where women are being subjected increasingly to psychological harassment in order to get money through the distribution of sexually explicit images or text messages. Moreover, the beating of brides by husbands, even fathers-in-law, are the result of the fact that some feudal men in society continuing to offend women. An analysis of individual convictions for such crimes by region revealed that most of the perpetrators live in Tashkent, Fergana, Samarkand, and Tashkent. The majority of victims of gender violence live in these areas. Of the 2,300 victims, 353 live in Tashkent region, 393 in Samarkand region and 228 in Tashkent. Despite Uzbekistan taking efforts such as making a system of prevention of oppression and violence against women in order to put an end to such situations, the fact that these issues are being continued shows the need for a lot of psychological, sociological and psycholinguistic research in this area.

Gender violence in the works we have mentioned also actually occurred as a result of the verbalization of life events.

## REFERENCES

1. Anthology of modern Korean prose. Compiled and prepared by: Saidazimova U.T., Choi So Yong. – Tashkent: Istiqlol Publishing House, 2014. – p.520.
2. Arbabzadah, Nushin. “2 Women and Religious Patronage in the Timurid Empire”. *Afghanistan’s Islam: From Conversion to the Taliban*, edited by Nile Green, Berkeley: University of California Press, 2016, pp. 56–70. <https://doi.org/10.1525/9780520967373-007>
3. Fayziyev, Turgun. *Temuriy malikalar*. – Toshkent: A. Qodiriy nomidagi xalq meros nashriyoti, 1994. <https://ziyouz.uz/ilm-va-fan/tarix/temuriy-malikalar/>
4. Khan, V.S. “History of Korea”. – Tashkent: Baktريا Press, 2013. – p. 128.
5. So Yong, Choi. “History of Korean culture”: monography. – Tashkent: Istiqlol, 2019. – p.120.
6. Torkunova, A.V. “History of Korea”. – Moscow: Moscow state institute of international relations; “Russian political encyclopedia”, 2003. – p. 430.
7. Baxronova, D.K. Gender i tselevye kontseptual’nyegendernye stereotypy [Gender and conceptual target gender stereotypes]. // *Gendernye issledovaniya v gumanitarnykh naukax: sb. st. po mater. III mejdunar. nauch.-prakt. konf. № 3*. – Novosibirsk: Sibak, 2015. <https://sibac.info/conf/gender/iii/43241>
8. Bakhronova, D., Abduraimov, S. (2022) The Influence of the National-Cultural Image of the World on the Construction of Metaphors in the Poetic Discourse. <https://iopscience.iop.org/article/10.1149/10701.13113ecst>

УДК 81-11

## СОСТАВ УЗБЕКСКОЙ МЕДИЦИНСКОЙ ТЕРМИНОЛОГИИ

*Г. З. Бозорова*

*БухГУ*

*Бухара, Узбекистан*

*gulmirabozorova@gmail*

В данной статье рассматривается медицинская терминология и ее интеграционные отношения с лингвистикой. Анализируются плеонастические единицы, образующиеся при появлении в речи медицинских терминов. Также классифицируются особенности образования медицинских терминов.

**Ключевые слова:** лингвистика, терминология, медицина, медицинские термины, медицинская аббревиатура, плеонастическое сочетание, плеоназм.

## COMPOSITION OF UZBEK MEDICAL TERMINOLOGY

*Bozorova Gulmira Zayniddinovna*

*BukhSU*

*Bukhara, Uzbekistan*

*gulmirabozorova@gmail*

This article covers medical terminology and its integration relations with linguistics. Pleonastic units formed when medical terms occur in speech are analyzed. Also, the features of the formation of medical terms are classified.

**Key words:** linguistics, terminology, medicine, medical terms, medical abbreviation, pleonastic combination, pleonasm.

## O‘ZBEK TILI TIBBIYOT TERMINOLOGIYASI TARKIBI

*Bozorova Gulmira Zayniddinovna*

*BuxDU*

*Buxoro, O‘zbekiston*

*gulmirabozorova@gmail*

Mazkur maqolada tibbiy terminologiya, uning tilshunoslik bilan integratsion aloqalari yoritilgan. Tibbiyotdagi atamalar nutqda voqelanganda hosil bo‘ladigan pleonastik birliklar tahlilga tortilgan. Shuningdek, tibbiyotga oid terminlarning shakllanish xususiyatlari tasniflangan.

**Kalit so‘zlar:** tilshunoslik, terminologiya, tibbiyot, tibbiy atamalar, tibbiy abreviatura, pleonastik birikma, pleonazm.

Har qanday fanning, jumladan, tilshunoslik fanining ham rivojlanishi o'zgarishlar jarayoni shaklida sodir bo'ladi – paradigmalari, ular asosida g'oyalar, usullar, ko'pincha, tadqiqot sohalari o'zgaradi. Shu bilan birga, tilshunoslikda an'anaviy paradigmalari ham mavjud bo'lib, ilm-fan periferiya (chegara)siga chekinishi mumkin va yangi paradigmalari uchun maydon bo'shatib beradi. O'ylab qarasa, kognitiv, gender kabi yangi yo'nalishdagi tilshunoslik, pragmalingvistika, lingvomadaniyatshunoslik, etnolingvistika, huquq tilshunosligi va boshqa yo'nalishlari yangi deb ham bo'lmaydi. Chunki istiqbol davrida gurkiragan mazkur tilshunoslik sohalari o'zbek lingvistikasida ham mustahkam o'rin egalladi va diqqatga sazovor tadqiqotlar yaratildi. Yangi yo'nalishni belgilashda esa vaqt omili muhim ahamiyatga ega. Mutlaqo yangi yo'nalish sifatida ekolingvistika, internet lingvistikasi, biolingvistika, tilingvistika, lingvoiktisodiyot, neyrolingvistika kabi sohalarni ajratish mumkinki, ularda hali nazariy va uslubiy munosabatlar to'la shakllanmagan, ammo taraqqiy qilish bosqichida turibdi. darhaqiqat, ayni kunlarda tilni uni egasi bilan dialektik birlikda o'rganadigan antropotsentrik yo'nalishda bajarilishi lozim bo'lgan tadqiqotlar tugal o'z yechimini topgani yo'q.

Bu esa o'z navbatida, tibbiy nutq alohida tadqiq obyekti ga aylanishiga sabab bo'lmoqda. Bugungi kunda terminologik tizimlarning shakllanishi va sohaga xos terminologik lug'atlarni yaratishning lingvistik muammolari bilan nafaqat tilshunoslar, balki professional mutaxassislar shug'ullanadi, shuning uchun fanlarning amaliy jihatidan muqarrar integratsiyasi mavjud. O'rganilayotgan bilim sohasining konseptual apparatiga ega bo'lish kasbiy muvaffaqiyat va terminologik savodxonlikning yuqori darajasiga erishishning muhim shartidir. Shifokor jamiyatning elita qismiga mansub bo'lib, uning nutqi me'yoriy, kommunikativ, axloqiy va estetik jihatlarga javob bera olishi kerak[1].

Barchamizga ma'lumki, tibbiyot sohasida ham abbreviaturalar keng qo'llanadi. Qisqartmalar, asosan, turli kasalliklar nomi, tibbiy jihozlar yoxud atamalar shaklida namoyon bo'ladi. Tibbiyotga oid abbreviaturalarni o'rganish jarayonida ularda ham ortiqchalik elementlari ko'zga tashlandi. Xususan, *OITS sindromi*, *O'RVI infeksiyasi*, *O'RK kasalliklari*, *UTT tekshiruv*i kabi birliklarda qisqartmaning ma'nosini to'la bilmaslik yoki bosh harflardan tashkil topgan abbreviaturaning mazmuniga e'tiborsizlik natijasida shifokorlar nutqida, tashxis varaqalarida, kasallik tarixlarini yozishda mana shunday pleonastik birikmalar shakllanganiga guvoh bo'lamiz. Ma'lumki,

uslubiy vositalar sanalmish evfemizm va perifraza oʻzaro oʻxshash hodisalar boʻlib, ularning har ikkalasida ham voqelikni inson koʻz oldiga keltiradigan darajada tasvirlashga asoslanish xususiyati mavjud, ammo biri nutqni muloyimlashtirishga xizmat qilsa, biri qayta badiiy, taʼsirchan nomlashga asoslanadi. Tibbiy nutqda termin bilan uning badiiy, xalqona, milliy ifodasi ketma-ketlikda qoʻllanish holatlari tez-tez kuzatiladi va bu ommaga tushunarliklik kasb etadi. Chunonchi, “*giyohvandlik – asr vabosi; moxov – rangsiz maydon; amneziya – teshik xotira; asab kasalligi – asr xastaligi; ogʻriq – organizm hayqirigʻi; astma – hushtakli tovush; bavosil – shohlarning kasali, gijja – ichki dushmanlar; immunitet – beminnat tansoqchi*” [2] kabi evfemik perifrazalarni shifokorlar nutqida tez-tez uchratishimiz mumkin. Kasalliklarni (*xususan ichki kasalliklarni*) toʻgʻri aniqlashga, ularning qanday kechishlarini toʻla va aniq tasvirlab berganligi va bu sohadagi qator kashfiyotlari Ibn Sinoni **hozirgi zamon terapiyasi asoschisi**, deb atashga imkon beradi (“Tib qonunlari” kitobi. I jild. 7-bet.) “Barkamol avlod – 2014” sport musobaqalarida namanganlik **salomatlik posbonlari** – shifokorlar zimmasiga musobaqa ishtirokchilarining sogʻligini muhofaza etish vazifasi yuklatilgan.

Shuningdek, tibbiyotda ayrim dori vositalarining izohi, uning xususiyatlari keltirilgan maʼlumotlarda ham pleonastik birikmalar koʻzga tashlanadi. Masalan, televideniya orqali “Espumizan” dori vositasining reklamasida **“qorindagi meteorizm”** birligi qoʻllangan boʻlib, mazkur birikmada qorindagi qismi ortiqcha hisoblanib, tibbiy jihatdan meteorizm faqatgina qorin boʻshligʻi uchun tegishli hodisadir. Aniq xulosa meteorizm soʻzining izohida oydinlashadi:

**Meteorizm** – (*yun. meteorismos* – shishish, qappayish) ich dam boʻlishi – ichakda koʻp yel toʻplanishi natijasida qorinning taranglashishi.

Izohli lugʻatda keltirilgan maʼlumotlar hamda tibbiyot mutaxassislari tomonidan berilgan xulosalarga tayanib, meteorizm faqat qorin boʻshligʻi hamda ichaklarda kuzatilgan hodisaligini inobatga olgan holda, mazmuniy takror sifatida pleonazmga misol qilib olish mumkin.

Yuqoridagi tahlil natijalariga koʻra, tibbiy pleonastik birliklarni uch guruhga ajratdik:

**1. Qisqartma (abbreviatura) shaklidagi.**

**2. Kasallik nomini ifodalovchi tibbiy atamaning semasini takrorlovchi.**

**3. Evfemik perifraza shaklidagi.**

Mazkur uch guruhdan birinchisi – qisqartma shaklidagi pleonastik



birliklar tamoman ortiqcha nutqiy kamchilik sifatida namoyon bo'lsa, ikkinchi va uchinchi guruh pleonazmlari diskurs talabidan kelib chiqib muayyan uslubiy vazifa bajaradi. Bu esa quyidagi jadvalda aniqroq ko'rinadi:



2.1.-rasm. Tibbiy pleonastik birliklar tasnifi

Tahlillar natijasi tibbiy nutqda pleonastik birliklarning qo'llanish shakllari turlichaligini ko'rsatdi. Xususan, boshqa sohalarga qaraganda, tibbiyotda, xususan, tibbiy atamalarda semantik ortiqchalik qisqartma (abbreviatura)lar bilan birgalikda, uning o'zida izohlangan birlikning mazmunini takror tarzda qo'llash holati ko'p uchrashi qayd etilgan.

#### ADABIYOTLAR:

1. Абузалова М.К. Лингвистика ва тиббиёт ҳамкорлиги ижтимоий зарурат сифатида// Soha lingvistikasi: tilshunoslik va tibbiyot integratsiyasining muammolari, yechimlari hamda istiqbollari mavzusidagi xalqaro ilmiy-amaliy konferensiya materiallari . – Buxoro, 2021. – B. 17–22.

2. Кобилова А.Б. Инглиз ва ўзбек лингвомаданиятида тиббий перифразаларнинг хусусиятлари: Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. Бухоро- 2022. — Б. 41.

3. Shavketovna, Y. A., & Zayniddinovna, B. G. (2023). Lexical-Semantic Characteristics of Medical Abbreviations. Best Journal of Innovation in Science, Research and Development, 2(6), 98–101. Retrieved from <http://www.bjisrd.com/index.php/bjisrd/article/view/289>

4. Vozorova, G. (2023). Лингвопрагматическая характеристика плеонастических единиц используемых в политической речи. ЦЕНТР

НАУЧНЫХ ПУБЛИКАЦИЙ (*buxdu.Uz*), 34(34). извлечено от [https://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/9752](https://journal.buxdu.uz/index.php/journals_buxdu/article/view/9752)

5. Zayniddinova B. G. Ortiqchalik tamoyilining jahon tilshunosligida organilishi. – 2021.

6. Bozorova, G. (2021, February). Tautology and pleonasm: synchronic and diachronic views. In *Конференции*.

7. Yarashovna, S. G., Istamovna, A. N., & Zayniddinova, B. G. (2023). Methods of Using Metaphorical, Pleonastic and Paremic Units in Uzbek Linguistics. *resmilitaris*, 13(2), 5560–5572.

8. Samandarova, G. Y. (2021). Fundamentals of folk proverbs formed on the basis of the lexical-spiritual group of insects. *current research journal of philological sciences*, 2(05), 39–42.

9. Bozorov Z.A International scientific and practical conference “Modern psychology and pedagogy: problems and solutions” “Imperial College London” 2021–2022. 165–169 б.

10. Бозоров З.А. The system of formation of civic culture among students in higher educational institutions. *ISJ Theoretical & Applied Science*, 01 (81), 455-458. Soi: <http://s-o-i.org/1.1/TAS-01-81-81> Doi: [https://dx.doi.org/10.15863/TAS\\_Scopus\\_ASCC:3304](https://dx.doi.org/10.15863/TAS_Scopus_ASCC:3304). (2020).

11. Kadyrovna A. M., Zayniddinova B. G. Description of Events Which is Near to Pleonasm //Spanish Journal of Innovation and Integrity. – 2022. – T. 5. – С. 281–285.

УДК 81-11

**TALMEHNING ERKIN VOHIDOV SHE'RLARIDAGI BADIY  
VAZIFALARI**

***Bozorova Gulmira***  
*Buxoro davlat universiteti,*  
*Buxoro, O'zbekiston*  
g.bozorova@gmail.com

В статье рассматривается роль поэтического искусства в поэтике и его поэтические функции в стихах. В этом плане анализируются внутренние аспекты искусства тальмех в стихах Эркина Вахидова.

**Ключевые слова:** поэзия, искусство тальме, имена исторических деятелей, названия известных произведений, герои произведений, названия небесных тел.

**TALMEHNING ERKIN VOHIDOV SHE'RLARIDAGI BADIY  
VAZIFALARI**

***Bozorova Gulmira***  
*Buxoro davlat universiteti,*  
*Buxoro, O'zbekiston*  
g.bozorova@gmail.com

The article examines the role of poetic art in poetics and its poetic functions in poetry. In this regard, the internal aspects of the art of talmeh in the poems of Erkin Vakhidov are analyzed.

**Key words:** poetry, talmeh art, names of historical figures, names of famous works, heroes of works, names of celestial bodies.

**TALMEHNING ERKIN VOHIDOV SHE'RLARIDAGI BADIY  
VAZIFALARI**

***Bozorova Gulmira***  
*Buxoro davlat universiteti,*  
*Buxoro, O'zbekiston*  
g.bozorova@gmail.com

Maqolada she'riyatda she'riy san'atlarning tutgan o'rni va uni she'rdagi poetik vazifalari haqida fikr yuritilgan. Shu jihatdan Erkin Vohidov she'rlaridagi talmeh san'ati ichiki ko'rinishlari tahlil etilgan.

**Kalit so'zlar:** she'riyat, talmeh san'ati, tarixiy shaxslar nomi, mashhur asar nomlari, asar qahramonlari, osmon jismlari nomi.

She’rda qo’llangan har bir she’riy san’at ijodkorning mahorati ni belgilab beradi. Jumladan, talmeh san’ati shoirdan kuchli tarixiy bilimlarni talab etadi. O’z navbatida, she’riyatda talmeh san’atidan foydalanish orqali ijodkor bir necha maqsadlarni ko’zlaydi:

1. She’r g’oyasini yanada oydinlashtirish maqsadida tarixiy voqealar (shaxslar) bilan qiyoslash, farqlash va o’sha voqeaga o’xshatish;

2. She’rda ma’lum bir asar, qahramon, hodisaga ishora qilish orqali she’rxonni mulohaza qilishga undash (albatta, bunda she’r g’oyasini chuqurroq anglash uchun o’sha asar mazmuni yoxud qahramon taqdiridan xabardor bo’lmoq lozim.);

3. Bir vaqtning o’zida she’rxonni ham tarix haqiqatidan, ham she’rning badiiy jozibasidan bahramand bo’lishiga imkon yaratish.

Albatta, bu maqsadlardan foydalanish uchun shoir mumtoz adabiyotni chuqur o’rganmog’i, o’tmish voqeligidan xabardor bo’lmog’i, shu bilan birga, xalqining qadim qadriyatlarini, folklori bilan tanish bo’lmog’i lozim.

Talmeh san’atining yorqin namunalari Erkin Vohidov ijodida kuzatiladi. O’zbek xalqining sevimli ijodkori, she’riyatimizning yorqin namoyondalaridan biri Erkin Vohidov ijodi nihoyatda go’zal va o’ziga xosdir. Shoir Erkin Vohidov ijodida bir qator yangiliklar qatorida she’riy san’atlarning ham o’rni nihoyatda katta. Xususan, talmeh san’atidan ham nihoyatda keng foydalanilgan. Shoir she’rlaridagi talmeh san’ati namunalari ma’no jihatdan quyidagi turlarga ajratish mumkin:

**1. Tarixiy shaxslar** nomini keltirish orqali talmeh san’atini hosil qilish.

Chulg’anib sokin xayolga

Jim yotar sog’onalar.

Ne ko’rar **Mirzo Ulug’bek**

Besh asrlik xobida?

Bosh egib ustoz **Ulug’bek**

Qoshida turdim bu tong,

Charx urib ko’nglim xayolning

Otashin girdobida. [1, 67]

Shoir o’zining “Samarqand” she’rida Samarqand tarixi, qadimiy obidalarini tilga olib, u yerda yotgan buyuk bobokalonlarimiz qabri, xususan, Mirzo Ulug’bek sog’onasi haqida so’z yuritadi. “Ilk muallimdir Samarqand ilmi aflok bobida” deb, astronomiya sohasida Ulug’bekning qo’shgan hissasi nechog’lik katta ekanligini ta’kidlaydi.

**Navoiy** sham kabi yonib,  
 Xazondek sarg'ayib so'lmish,  
 Varaqlar ham olibdur  
 Sohibi devoni rangidan.

**Fuzuliy** Karbalo dashtin  
 Quyundek kezdi **Qays** birlan,  
 Dili hech topmadi taskin  
 Davrning boda, bangidan [1, 91]

Ushbu parcha Erkin Vohidovning “Xazina” she’ridan keltirilgan. Shoir xazina deyish orqali turkiy adabiyot rivojiga hissa qo’shgan va uning dovrug’ini dunyoga sochgan buyuk bobokalonlarimizni nazarda tutadi.

2. **Mashhur asarlar** nomini keltirish orqali talmeh san’atini hosil qilish.

“**Yosh Verter**”ni o’qib bir kun meni ham-  
 Sevarmikin kimdir, deya o’ylading  
 G’oyibona yo’llaringga intizor,  
 Allakimning kutishini bilmaysan [1, 12]

Ushbu she’rda ulug’ nemis shoiri Iogann Wolfgang Gyotening “Yosh Verterning iztiroblari” nomli asari tilga olinadi. Shoir hali sevgisini uchratmagan qiz qo’liga “Yosh Verterning iztiroblari” kitobini tutqazib o’yga toldiradi. She’rda asar qahramonlari singari chin sevgisini uchratishiga umid qilgan qiz o’y xayollari va lirik qahramonning unga murojaati kuylangan.

3. **Asar qahramonlari** nomini keltirish orqali talmeh san’atini hosil qilish.

Sevgi sahrosida qolmish  
 Necha **Majnundan** g’ubor,  
 Necha **Farhod** gardi yotgan  
 Bistunning tog’ida. [62]

Majnun – Sharq adabiyotida keng tarqalgan oshiq obrazi. Farhod – ishqning eng yuqori cho’qqisiga chiqqan komil inson obrazi. Tabiiyki, har bir shoir, xususan, Erkin Vohidov ham sevgi haqida jo’shqin she’rlar bitgan. Bu she’rlarda Layli, Majnun, Farhod, Shirin obrazlari ko’p uchraydi. Eng qizig’i esa shoir bu nomlardan har gal yangidan yangi o’xshatishlar, faqatgina shoir ko’zi ko’ra oladigan xususiyatlarni topa olgan.

4. **Osmon jismlari** nomini keltirish orqali talmeh san’atini hosil qilish.

Ana **Hulkar**-yetti qiz imlar meni yiroqdan,

Oydin ko'chalar bo'ylab shu'lalarda oqurman.  
 Ulug'bek qadam qo'ygan bu muqaddas tuproqdan  
 Ulug'bek nigoh tikkan yulduzlarga boqurman [1, 23]

O'zingsan-ku mening yulduzim,

**Surayyodan, Zuhrodan** ortiq.

Ko'k toqidan olmadek uzib,

Birin senga qilaymi tortiq? [1, 26]

Ushbu misrada lirik qahramon yorini Surayyoga, Zuhroga o'xshatadi. Mumtoz adabiyotda ham yorni oyga, yulduzlarga o'xshatish an'anasi mavjud bo'gani barchamizga ma'lum.

Rost g'azal avjida barcha

**Oy** ila **Zuhro** emas,

Ko'p erur somonchilar ham

She'riyat osmonida... [61]

Erkin Vohidov ushbu mashhur "Yoshlik devoni"dan olingan parchada Oy va Zuhro deyish orqali adabiyotimizdagi g'azalnavislikda yetuk bo'lgan siymolarni nazarda tutmoqda. Har yerda bo'lgani kabi sohasida iste'dodsiz, o'xshamagan va g'aliz g'azallar bitgan shoirlar ham yo'q emas. Ularni esa somonchilarga qiyoslaydi. Sababi somonchilar fanda yulduz turi emas, osmonni o'rab olgan xira nur hisoblanadi.

Erkin Vohidov she'riyatida badiiy san'atlarning rang-barang ko'rinishlari mavjudki, ular shoir she'rlarining hali o'rganilmagan ko'p qirralari borligidan dalolat beradi.

#### ADABIYOTLAR:

1. Эркин Воҳидов. Танланган асарлар. Тошкент, 2019.
2. Sulton I. Adabiyot nazariyasi. – T.: O'qituvchi, 2005.
3. Куронов Д., Мамажонов З. ва бошқалар. Адабиётшунослик луғати. – T.: Akademnashr, 2010.
4. Dildora Yusupova, Dilorom Yuldasheva. Halima Xudoyberdiyeva she'riyati lingvopoetikasi.– Germaniya: "Globe Edit international Publishing Group" ISBN: 978-620-0-60879-6.– 2020.–125 p.

УДК 81-11

**SOME THOUGHTS ON TEACHING VISUAL EXPRESSIVENESS  
OF THE UZBEK LANGUAGE IN RUSSIAN-SPEAKING GROUP***Fattakhova D. f.f.f.d. (PhD)*

fatakhova@gmail.com

This article examines the use of means of verbal expression in works of art. Means of verbal expressiveness serve not only to clarify, wider and deeper disclosure of many of the depicted pictures and phenomena, but are also an indicator of the level of creative thinking; their use is important in achieving imagery and accuracy of speech.

The article also gives a brief description of the means of verbal expressiveness of comparison, personification and provides examples. Students' work on studying these means will help them consolidate their knowledge and skills in studying Uzbek as the state language.

**Key words:** state language means of speech expressiveness, personification, comparison, skill, speech culture, expressiveness, imagery, accuracy of speech

В данной статье рассматривается использование средств речевой выразительности, в художественных произведениях. Средства речевой выразительности служат не только для уточнения, более широкого и глубокого раскрытия многих изображаемых картин и явлений, но и являются показателем уровня творческого мышления; их использование важно при достижении образности, точности речи.

В статье также дается краткая характеристика средств речевой выразительности сравнения, олицетворения и приводятся примеры. Работа студентов над изучением этих средств поможет им закрепить знания и навыки изучения узбекского языка как государственного.

**Ключевые слова:** государственный язык средства речевой выразительности, олицетворение, сравнение, навык, культура речи, выразительность, образность, точности речи

**Introduction**

It is customary to distinguish the following main functions of language, which are reflected in speech: communicative (communication function); thought-forming (function of embodiment and expression of thoughts); expressive (function of expressing the internal state of the speaker); aesthetic (the function of creating beauty through language).

From the point of view of the culture of speech, the communicative function of language presupposes the orientation of participants in speech communication towards fruitfulness, mutual usefulness of communication, and a general focus on the adequacy of understanding

speech. However, achieving functional communication efficiency is impossible without knowledge and compliance with the norms of the literary language. This, in turn, requires the presence of the most necessary and elementary ideas about the lexical, grammatical and stylistic norms of the Uzbek language. They are indispensable when working on constructing correct Uzbek speech, which must be distinguished by clarity and accuracy of presentation of thoughts, consistency and information content. It must also be characterized by richness, expressiveness, appropriateness, purity, accessibility, correctness, logic, accuracy, etc.

### **Methodology**

Language has the units necessary to construct messages, the rules for their organization, and ensures the emergence of similar images in the minds of participants in communication. Language also has special means of establishing and maintaining contact between participants in communication. Achieving functional communication efficiency is impossible without knowledge and compliance with the norms of the literary language.

Speech culture includes a number of signs and properties, the totality and system of which speaks of its communicative perfection: accuracy of speech; consistency, mastery of logical reasoning; cleanliness, i.e. the absence of elements alien to the literary language and rejected by moral norms; richness – variety of speech, absence of the same signs and chains of signs; the appropriateness of speech is such a selection, such an organization of language means that make speech consistent with the goals and conditions of communication. Appropriate speech corresponds to the topic of communication, its logical and emotional content, the composition of listeners or readers, informational, educational, aesthetic and other objectives of the speech. Expressiveness is a feature of the structure of speech that maintains the attention and interest of the listener or reader.

The more different and consciously distinguished linguistic signs and their combinations fall on the same “speech area”, on the same “speech space”, the richer and more diverse the speech. There are also structural and linguistic conditions that ensure the richness of students’ speech: a large volume of active vocabulary, a variety of morphological forms and syntactic structures used in speech. The richer the vocabulary, the more developed the grammatical structure of a student’s speech, the more favorable the conditions for working on its accuracy



and expressiveness. Let us clarify that the ratio of a student's active and passive vocabulary is the most important measure of the quality of speech and its level of culture. Speech culture presupposes the regular reproduction of vocabulary, taking into account polysemy, phraseology, figurative meanings, etc. Consequently, the process of teaching students the culture of speech and its expressiveness is of particular importance. The culture of speech is an important part of the national culture as a whole. Therefore, an urgent need is to work to improve the speech culture of Russian-speaking students. Formation of future philologists' skills in mastering the means of speech expressiveness of the Uzbek language is a pressing problem today. Expressive speech serves not only to express the emotions of the speaker, but also to influence the addressee.

### Results and analyses

As E. Dzhanzhakova notes, "expressiveness necessarily presupposes the possibility of influence, in other words, the emotional-expressive aspect of the utterance is focused on both the speaker and the listener, therefore attempts to contrast expressiveness as the ability of units to convey additional meaningful information and expressiveness as the ability of such units to "strengthen" author's influence on the addressee" do not seem appropriate [1,112]. She also notes the existence of a number of figures and tropes specifically designed to influence the emotions of the addressee [1,113].

Means of verbal expressiveness serve not only to clarify, wider and deeper disclosure of many of the depicted pictures and phenomena, but are also an indicator of the level of creative thinking; their use is important in achieving imagery and accuracy of speech. When teaching Russian-speaking students the means of speech expressiveness of the Uzbek language, the knowledge and use of the following groups of tropes become essential: metaphors, metonymies, epithets, paraphrases and comparisons; and figures: rhetorical questions, rhetorical appeal, repetitions, personification, silence, exclamation, because their frequency of use in texts, as statistical analysis shows, is very high.

Comparison is a type of trope that consists of comparing one object with another for the purpose of artistic description; it is "a figurative expression built on the comparison of two objects, concepts or states that have a common feature, due to which the artistic meaning of the first object is enhanced. This is one of the most commonly used means of expression in Uzbek speech. Comparisons are varied in form, type

of evaluation, degree of emotionality, imagery. Comparisons are distinguished between traditional and original, commonly used and new, as well as updated, complete or incomplete. The comparisons also differ in the scope of their use. [2, 555].

Comparison is a comparison of two phenomena in order to explain one of them with the help of the other. In the Uzbek language, comparison as a speech case is formed by various lexical means. Comparison has two components: the object of comparison and what it is compared with: o'hshash, teng, kabi, singari, go'yo, huddi, misoli, bamisoli, aynan. Misoli gul, misoli bulbul; go'yo nurli oftob (conjunctions: as, as if, as if; that, all the same; as if, exactly; as, so (and); words: similar, similar; similarly, similar, like; resembles; etc.) [3.93].

For example: "Ariqlarning bo'ylarida, marzalarda ko'klam nishonasi- yalpizlar bodroq-bodroq bo'lib una boshlagan, baqalarning "vaq-vaq"i avjida. Chimzorlardagi kuzda o't qo'yib kuydirilgan ajriqlarning tomirlaridan chiqqan yangi, yashil barra giyohlar yer bag'ridagi hayotning qaytadan jonlanganidan dalolat berardi. So'ppaygan qovjir-oq qamishlar izg'rinli bahor shamolida bir-biriga urilib noxush ovoz taratardi. Ikki labi ko'karib qolgan Dimariq hali hech kim va hech nima loyqalab ulgurmagan uchun tip-tiniq, bilinar-bilinmas mayin chayqalib yotardi" [4, 227].

"Suv tiniq bo'lgani uchun uning harakatlari aniq ko'rinar edi. To'g'ri borib po'sti qolmagan, suv yalab o'tayotgan tol to'nkasi tagiga qo'l suqdi. Amfibiya odamday suv ostida har zamonda oyoqlarini siltab, muallaq holda ancha turib qoldi, keyin orqasiga tisarilib, suv betiga otilib chiqdi... O'n beshtaga borganda akam suvdan chiqdi... Baliqni opam ikkalamiz ayvonda tozalab berib turdik, onam qovurishga tushdi. Otam tayyor bo'lganini peshma-pesh ichkariga olib kirib ketardi. Mehmonlar juda xursand: xoxolashar, ayniqsa, "baliqxo'r"ning kulgisi momoguldirakday uyning derazalarini zirillatib yuborardi."

Personification, as a special type of metaphor, is the endowment of inanimate objects with the properties and characteristics of a person. This is the most common trope that goes back to folk poetry. For example, in the story "Fear" A. Kakhhar gives the following description of the wind:

"Yaqin ikki haftadan beri ko'z ochirmayotgan kuzak shamoli yaydoq daraxtlar shoxida chiyllaydi, g'uvillaydi; tomlarda vishillaydi, yopiq eshik va darchalarga bosh urib uf tortadi. Bunday kechalarda odamzod qo'yimijoz g'uj bo'lib va nimanidir kutib jimgina o'tirishni xohlab qoladi."

For a moment, it seems to us that we seem to hear the sound of the wind, it's dark around, not a soul, and this picture comes to life before our eyes. Although Unsina fulfilled what she promised, her heart could not bear the fear she experienced. After all, she was still a very young girl.

“Shamol bir xuruj qilganida nimanidir keltirib darchaga urdi. U narsa darchani tirmalaganicha sidirilib pastga tushib ketdi. Hamma o'tirgan yerida go'yo bir qarich cho'kkanday bo'ldi va tin olmay bir-biriga qaradi.”

### Conclusion

The author's skillful use of such expressive means of speech as comparison, personification, emotionally affecting the reader deserves his attention, makes the work even more interesting.

Mastery of the means of verbal expressiveness presupposes taking into account the communicative-activity approach to learning the language of the specialty. At the same time, one should also take into account the specific national characteristics of the audience and the degree of proficiency in the Uzbek language of philology students. This is due to the linguistic environment itself, which has formed to date in our republic. It is the linguistic environment that inevitably contributes to the mastery of the Uzbek language as a necessary means of communication. The majority of the Russian-speaking audience, as is known, are students of Uzbek nationality who speak Uzbek as their native language. Among them there are also many Russian speakers who have a good command of the Uzbek language, which is not their native language. However, the level of proficiency in both languages usually turns out to be insufficient when mastering the means of verbal expression that are most often used in communication.

So, in the presentation of educational material it is necessary to take into account the reproductive, adaptive, local-modeling, system-modeling levels of its presentation. And only then can the teacher achieve his goal. Namely, teaching students the correct, competent use of means of verbal expressiveness in speech.

### REFERENCES:

1. Джанжакова Е.В. О разграничении тропов и фигур // Риторика и синтаксические структуры. – Красноярск, 1988. – С. 112–113.
2. Выразительные средства современной русской речи Тропы и фигуры. Терминологический словарь. – М.: Феникс, 2007. – С.869.

3. Бегматов Э., Жиянова Н. Нутқ маданияти асослари. – Тошкент, 2006. 93 б.
4. Бегматов Э. Нутқ маданияти тушунчаси ҳақида // Ўзбек тили ва адабиёти. – Тошкент, 1975. № 5
4. XX аср Ўзбек ҳикояси антологияси “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, Тошкент, 2009. – 233
5. Маҳмудов Н. Ўхшатишлар ва миллий образ ЎТА, Т., 2013 – №1.– Б. 3–8;
6. Маҳмудов Н., Худойберганова Д. Ўзбек тили ўхшатишларининг изоҳли луғати. – Т.:Маънавият, 2013;
7. Худойберганова Д Семантические и стилистические особенности конструкций уподобления в узбекском языке: Дисс. канд. филол. наук. Т., 1989. – 120с.
8. Ожегов С.И. О нормах словоупотребления // Лексикология. Лексикография. Культура речи. – М.: Наука, 1984. – С. 285–287.
9. Хайитов Т.А. Методика обучения выразительному чтению с применением аудиовизуальных средств обучения в национальном языковом педвузе. Автореф. дис. ... канд. пед. наук. – Ташкент, 1991. – 19 с.

УДК 81-11

**REALIYALAR MILLIY MADANIYATNI ANGLATUVCHI  
BIRLIK SIFATIDA***Ishanjanova Munosibxon Saidahmadovna**Andijon davlat chet tillari institute**Andijon, O'zbekiston*

munosibxon@gmail.com

Mazkur maqolada realiyalarning milliy xos soʻzlar sifatidagi jihatlari va uni olimlar tomonidan yaratilgan tasniflari oʻrganilgan. Shuningdek, realiyalarning tarjimada oʻziga xos xususiyat kasb etishi yoritilib oʻtilgan.

**Kalit soʻzlar:** tarjimashunoslik, atama, til, madaniyat, ekvivalentsiz soʻzlar

Realiyalar tarjimashunoslikda oʻziga xos oʻrin egallaydi. “Realiya” soʻzining oʻzi lotincha – “moddiy”, “haqiqiy” maʼnolarini kasb etadi. Realiyalar lugʻat taʼriflariga koʻra, “moddiy madaniyat obʼektlari” sanaladi. O.S. Axmanova oʻz lugʻatida realiyalarni oʻzgarimas, maʼlum bir millat madaniyatiga talluqli soʻzlar ekanligi aytadi [1;381]. Tilshunoslik va tarjimashunoslikda realiyalar maʼlum millatlarga va madaniyat obʼektlariga, tarixiy faktlarga, milliy qahramonlarning ismlariga, mifologik mavjudotlarga xos boʻlgan nomlar sanaladi.

S. Vlaxov va S. Florin realiyalar – bu bir xalqning hayoti, turmush tarzi, madaniyati, ijtimoiy va tarixiy rivojlanishiga xos boʻlgan obʼektlarni nomlaydigan soʻzlar va iboralar, ular boshqa xalq tilida mavjud boʻlmasligi, shuning uchun ular umumiy asosda tarjima qilinmay, oʻziga xos yondashuvni talab qilishini eʼtirof etadi [2;55].

Ushbu taʼrif olimlar tomonidan realiyalarni tarjima qilish nuqtai nazaridan berilgan boʻlib, tarjima jarayonida ular nafaqat tillar, balki matnlar, madaniyatlar va vaziyatlar ham bir-biriga qarama- qarshiligiga urgʻu beradilar.

G.D. Tomaxin turli tillar va madaniyatlarni taqqoslashda realiyadan foydalanish quyidagi xususiyatlarni ajratib koʻrsatishga imkon berishini aytadi [3;13]:

1. Faqat bitta til jamoasiga xos realiyalar,
2. Ikkala til jamoasida ham mavjud, ammo ulardan birida bu qoʻshimcha maʼnoga ega realiyalar,
3. Turli jamiyatlarda shunga oʻxshash realiyalar maʼnosi shakldoshlaridan farq qiluvchi realiyalar.

Voqelikda til va madaniyat oʻrtasidagi yaqinlik aniq namoyon boʻlib, jamiyatning moddiy va maʼnaviy hayotida yangi realiyalarning

paydo bo'lishi tilda tegishli so'zlarning paydo bo'lishiga olib keladi. Realiniyning o'ziga xos xususiyati uning ob'ektiv tarkibining tabiatidir. Realiniyalar individual yoki ma'lum jamoa doirasi bilan cheklanishi mumkin. Shuningdek, realiniyalar madaniyat bilan chambarchas bog'liq bo'lgan lingvistik hodisa sifatida ushbu leksik birliklar jamiyat rivojlanishidagi o'zgarishlarga tez moslahadi, ular – neologizmlar, tarixiylik, arxazimlardan farqlanishi lozim. Realiniyalarning o'ziga xos xususiyati madaniyat bilan bog'liqligidadir. Ayrim hollarda realiniyalarni nomlardan, atamalardan farqlash qiyin kechadi, lekin ularning nafaqat stilistik maqsadlar, balki milliy, mahalliy va tarixiy ranglarni qayta tiklash uchun ham xizmat qilishi, atamalar esa, asosan, sun'iy ravishda ma'lum tushunchalar va ob'ektlarni nomlash uchun yaratilishi bilan ajratiladi. Realiniyalar tabiiy ravishda, xalq so'z ijodkorligi natijasida paydo bo'ladi. Atamalar har qanday narsalarning nomi bo'lib, ushbu ob'ektlarning tarqalishi bilan tarqaladi. Realiniya esa ular paydo bo'lgan xalqning madaniy mulki hisoblanadi. S. Vlahov va S. Florin realiniyalar ekvivalentsiz so'zlar qatoriga kirishini ta'kidlaydilar. Realiniya badiiy madaniyat va ommaviy axborot vositalarining sublingualizmiga xos bo'lib, ma'lum bir xalq madaniyati bilan uzviy bog'liqdir, bu xalq tili uchun keng tarqalgan va boshqa tillar uchun begona sanaladi. Atamalar milliy rangga ega bo'lmay, asosan, fan sohasiga tegishli bo'lib, sun'iy ravishda, faqat ob'ekt yoki hodisaning nomi uchun yaratilgan bo'lib, tilda keng qo'llaniladi [2;17] Ba'zi hollarda realiniya adabiy me'yordan chetga chiqadi, masalan, dialektizmlar, jargonlarda ushbu holat kuzatiladi. Realiniyada til va madaniyat o'rtasidagi yaqinlik munosabati aniq namoyon bo'ladi: jamiyatning moddiy va ma'naviy hayotida yangi urf-odatlarining paydo bo'lishi tilda realiniyalarning paydo bo'lishiga olib keladi. Realiniyalarning o'ziga xos xususiyati, bir tomondan milliy, ikkinchi tomondan tarixiy zamonga aloqador. Realiniyalar milliy va tarixiylik xususiyatlari bilan ajralib turadi. Realiniyalar mazmuni etnik-madaniy lug'atlarda aks etadi. Realiniyalarni tadqiq etgan olimlar ularning turlari haqida to'xtalib, ularni mavzuviy tasniflarini yaratganlar. Bu borada A.E. Suprun, V.I. Repin tasniflari haqida gapirish mumkin, shuningdek, A.A. Reformatskiy tasnifi mavzuviy printsiptiga asoslanadi, unda chet tilidagi so'zlar qaysi tillardan lug'atlarga kirib kelganligi qayd etilgan. Realiniyalarning mavzuviy bo'linishi quyidagilarni tashkil etadi: 1) tegishli ismlar, 2) shaxslarning mavqei va ijtimoiy statuslari, 4) kostyum va zargarlik buyumlari tafsilotlari, 5) oziq-ovqat va ichimliklar, 6) murajaatlar kabilar [4].

Hozirgi kungacha madaniy jihatdan belgilangan birliklarning yagona tasnifi mavjud emas va tadqiqotchilar ma'lum printsiplarga asoslanib realiyalarning turli tasniflarini taklif qilishadi.

E.M. Vereshchagin va V.G. Kostomarov milliy-madaniy semantikaga ega bo'lgan so'zlarni yetti guruhga ajaratadi [5;60]. Olim nazariyasiga asoslanib realiyalar quyidagicha tasniflandi:

1. Mustaqillik davrida paydo bo'lgan tushunchalarni ifodalovchi so'zlar (masalan: Vazirlar Mahkamasi, Senat, deputat).

2. Yangi hayot so'zlari (masalan: madaniyat bog'i, ekopark, Yangi Andijon).

3. An'anaviy hayot ob'ektlari va hodisalarining nomlari (masalan: palov, doira, do'ppi, sunnat to'y).

4. Tarixiy davrlarning shaxslari, ob'ektlari va hodisalarini bildiruvchi so'zlar (masalan: yig'och, bekota, lashkarboshim sarkarda).

5. Frazeologik birliklarning so'z boyligi (masalan: sog'doqqa solg'ucha, ikki o'q otquncha fursat, so'zlari bir biriga qovushmoq).

V.S. Vinogradovning so'zlariga ko'ra, milliy-madaniy ma'lumotlarning mazmuni, birinchi navbatda, milliy tarix va davlat tuzilishining o'ziga xos jihatlari, geografik muhit xususiyatlari, o'tmish va hozirgi moddiy madaniyatning o'ziga xos ob'ektlari, etnografik folklor tushunchalari kabilarni qamrab oladi [6;87]. Realialarni aks ettiruvchi tushunchalar milliy xarakterga ega va ekvivalenti bo'lmagan so'zlar deya ta'riflanadi. V.S. Vinogradov realialarni o'rganar ekan, ularni bir qator tematik guruhlariga bo'ladi:

1. Kundalik hayotni aks ettiruvchi realiyalar: a) uy-joy, mulk; b) kiyim-kechak; c) oziq-ovqat, ichimliklar; d) mehnat va kasb turlari; e) banknotalar, o'lchov birliklari; f) musiqa asboblari, xalq raqslari va qo'shiqlari, ijrochilar; g) Xalq bayramlari, o'yinlar; h) murojaatlar;

2. Etnografik va mifologik realiyalar: a) etnik va ijtimoiy jamoalar va ularning vakillari; b) ajoyib mavjudotlar, afsonaviy joylar;

3. Tabiat dunyosini aks ettiruvchi realiyalar: a) hayvonlar; b) o'simliklar; c) landshaft;

4. Davlat – ma'muriy tuzilish va ijtimoiy hayot realialari: (aktual va tarixiy)

a) ma'muriy birliklar va institutlar; b) jamoat tashkilotlari, partiyalar va boshqalar, ularning ishtirokchilari; c) sanoat va agrar korxonalar, savdo muassasalari; d) harbiy bo'linmalar va darajalari;

e) lavozimlar, kasblar, unvonlar;

5. Onomastik realiyalar; a) antroponimlar; b) toponimlar; v) adabiy qahramonlarning ismlari; d) kompaniyalar, muzeylar, teatrlar, restoranlar, do'konlar, plyajlar, aeroportlar va boshqalar nomlari;

6. Assotsiativ voqelikni aks ettiruvchi realiyalar: a) vegetativ belgilar; b) hayvonlarning ramzlari; c) rang ramzi; d) folklor, tarixiy va adabiy alyuziyalar. Ularda turmush tarzi, xulq-atvori xususiyatlari, tarixiy, folklor va adabiy qahramonlarning harakatlari, tarixiy voqealar, afsonalar, urf-odatlar aks etadi. e) til ishoralari. Ular, odatda, frazeologik birlik, maqol, maqol, iborani o'z ichiga oladi.

Assotsiativ realiyalar turli xil milliy tarixiy va madaniy hodisalar bilan bog'liq va tilda o'ziga xos tarzda mujassamlangan. G.D. Tomaxin shu turdagi realiyalarni o'rganadi va ularni konotativ realiyalar deb ataydi. Konotativ realiyalarning denotativ realiyadan farqi semantik tuzilishi o'zgacha bo'lib, ular hissiy – ekspressiv ohanglarda o'zlarining moddiy ifodasini topadilar, shuningdek, taqqoslanadigan tillarda kontseptual o'xshash so'zlarning axborot nomuvofiqliklari aniqlangan. Bunda kundalik qo'llaniluvchi “quyosh”, “oy”, “dengiz”, “qizil” kabi so'zlar mamlakatshunoslik bilimlar bilan assotsiativ realiyalar tarkibida birga keladi.

Tadqiqotchilar madaniy markerlangan leksikani realiyalar ekvivalenti bo'lmagan so'zlar deya ifodalaydilar. Jumladan, L.S. Barxudarov ekvivalentsiz so'zlarni quyidagicha tasniflaydi [7; 94-95]:

1. Boshqa til lug'atida doimiy muvofiqlikka ega bo'lmagan tegishli ismlar, geografik nomlar, muassasalar, tashkilotlar, gazetalar va boshqalar nomlari.

2. Realiyalar-boshqa tilda gaplashadigan odamlarning amaliy tajribasida mavjud bo'lmagan narsalar, tushunchalar va vaziyatlarni bildiruvchi so'zlar.

3. Lakunalar-ba'zi sabablarga ko'ra boshqa tilning leksik tarkibida mos kelmaydigan til birliklari.

Xullas, realiyalar ekvivalentsiz so'zlar sanalib, tarjimada mos, qisman mos va mos kelmagan holda izoh orqali birliklar bilan o'g'iriladi.

#### ADABIYOTLAR:

1. Ахманова, О. С. Словарь лингвистических терминов / О. С. Ахманова. – М., 1966. – 401 с.

2. Влахов, С. И. Непереводимое в переводе / С. И. Влахов, С. П. Флорин. – Москва : Валент, 2009. – 360 с.



3. Томахин, Г. Д. Реалии в культуре и языке. Реалия – предмет и реалия – слово / Г. Д. Томахин // Иностранные языки в школе. – 2007. – № 8. – С. 19–28.

4. Реформатский, А. А. Введение в языковедение : Учебник для филологических специальностей высших педагогических учебных заведений / А. А. Реформатский ; под ред. В. А. Виноградова. – Москва : Аспект-Пресс, 2001. – 536 с.

5. DN Yuldasheva. BASIC NON-VERBAL COMPONENTS OF SPEECH. CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (2767-3758) 3 (01), 17–24. <https://masterjournals.com/index.php/crjps/article/view/661>

6. Верещагин, Е. М. Лингвострановедческая теория слова / Е. М. Верещагин, В. Г. Костомаров. – Москва : Русский язык, 1980. – 320 с.

7. Виноградов, В. С. Перевод : Общие и лексические вопросы / В. С. Виноградов. – 2-е издание, перераб. – Москва : Книжный дом Университет, 2004. – 235 с.

## УДК 81-11

**INTONEMES ARE REPRESENTATIVES  
OF THE INTERNAL STATE OF THE COMMUNICANT**

*Gulchekhra Toshpulatovna Chullieva*  
*Bukhara State University, Uzbekistan*  
*ORCID ID (ORSID 0000-0002-7642-8893)*  
*gchulliyeva@bk.ru*

The following article deals with intonation and its units. Types of intonema, the functions performed by intonema in speech, the meanings it expresses, and its semantic types are also given in the article.

In modern linguistics, it is important to analyze the language phenomena based on the state of speech, because speech is complex and multidimensional. Language units are specific to perform different tasks in speech. In speech that results from the communicators' common goal of "informing" or "receiving information", tone, or intonation, is one of the tools that help to understand this "purpose of speech".

**Introduction**

Any speech is dynamic, and this movement is accomplished by a certain intonation: the tone, speed, pauses, pitch and power of the speech change and glow depending on the requirements of the content and emotion of the sentence. We cannot make up a sentence without intonation: that's why every sentence has its own intonation. [6] For example, if we tell the sentence "*Grandmother has come*" in question, exclamatory or affirmative form, we will be able to change the aim of the sentence. The word "Sky..." (incomplete tone, i.e. the continuation of an idea),

"*Sky*" (informative tone); "*Sky!*" (endless excitement, enjoyment); "*Sky?*" (interrogative tone); pronouncing in this way serves to reveal the essence of the speech. Thus, melody is the main nonverbal tool for the correct understanding of the "purpose of speech".

It is clear that, language is a "system of systems". Prosody is an integral part of the phonological system of a language. Each component has its own units, rules and methods.

**Literature review**

The word "prosodica" is being used in recent researches. Prosody is a unit of phonology that reflects the physiological properties of speech as a measure of the substantial level. The linguistic-functional features and functions of phonetic-phonological units in language, the compo-

nents of phonology that study their role in the mechanism of language are called phonemes, syllobema, accentema and intonema. Phonology itself is divided into segmental and supersegmental phonology. Prosodica is a generalization of syllables, axintems, and intonemes that are the subject of supersegmental phonology [1].

Intonation is derived from the Latin word “intonation” (“intono, intonare” – “to pronounce aloud”). Intonation is a set of interrelated rhythmic and melodic components of speech. As the American scholar Kington indicated, “Intonation is the lifeblood of the speech process”. Intonology is the study of intonation in linguistics. Scholars focus on intonation tools such as accent, melody, tempo, and timbre. The tempo, timbre, and emphasis of these tools are used to express the main mood, emotions, and mood of the speaker. The pace of speech can vary depending on the purpose: slow, medium, fast. Changing the pace of speech depends on the emotional state of the speaker. It’s played by actors who play more roles, because that’s how they describe the situation of the actor they’re playing. [7]

## Analysis

Alexander Alexandrovich Reformatsky, a representative of the Moscow School of Phonology, was one of the scholars who made a significant contribution to the role of intonation in the phonological system. He revealed the peculiarities of intonation in his book “Prolegomena to the study of intonation”. A.M. Peshkovsky, in his article “Intonation and Grammar”, emphasized the need to study the intonation structure of sentences in order to build a grammatical theory [3].

Looking at intonation from the point of view of the speech process seems that its function as a way of organizing the whole meaning of the communicative speech expression and connecting it logically with the next sentence. Intonation reveals the informational nature of speech, reflects the interconnectedness of speech. The description of intonation allows us to look at the sentences as a meaningful one, a grammatical tool, and a speech phenomenon, connecting them in terms of logical sequence when analyzed based on the above function. H. Jamolkhanov has defined the role of intonation in language and speech as follows:

- it serves as an important tool for expressing the intonation of the sentence;
- it participates in distinguishing types of speech according to the purpose of expression and emotionality;

– it is an important phonetic device representing syntactic connections [6].

All the functions of intonation together help to express the rich nature of oral discourse, which is called “speech in action”.

In the “age of innovation” with the development of computer technology, the telephone and various communication programs (such as telegrams, IMO) are actively used as a means of communication between people, as well as long-distance means of communication. People use the “sms” service to send messages, the possibilities and conveniences of which are not new to today’s man. Interestingly, in sms messages, we use “smile” because the words do not fully reflect our thoughts, that is, the purpose of our speech. In addition to the main “information” of the speaker, they have a variety of additional meanings, such as his personal opinion, mood in the process of speech, age, nationality. This means that the letters of the alphabet, the words in our dictionary, “fail” to fully explain the “purpose of speech.” In oral speech, the tone of speech can express the function of these “smiles”. For example, consider the following passage from the work: “*It’s cold, – said my mother in **thoughtful tone**. There is not anything to put on children’s body and head, especially difficult for the little ones*” [5]. The author’s “**thoughtful tone**” reflects the woman’s helplessness and the family’s support for her husband. A change in the tone of the sentence could change the tone of the speech. (E.g. anger, indifference, threat, etc.).

## Discussion

Individual intonation research in world linguistics has begun not so long ago – in the 1940s, with great advances in linguistics and the collection of a number of experimental data from the study of many world languages. The use of modern precise methods of phonetic analysis has led to astonishing results, which have made it possible to declare intonology as an independent branch of linguistics with its own object of study and its own method of analysis. Intonation began to be studied from different angles: linguistics, paralinguistics, sociology, didactics, aesthetics, and others.

Intonation is a universal linguistic tool, all languages have their own intonation, but its form and function inevitably differ from the point of view of the situation.

The research that began today with the study of the difference between the English dialect and the British and American dialects has

led to the discovery of many interesting scientific discoveries, such as the Azerbaijani dialect, the Italian dialect, and the Turkish dialect. It should be noted that the Uzbek dialect has not yet been studied.

The unit of intonation is the intonema (tone phoneme, intonation structure), which is the intonation period, the intonation characters (parameters) sufficient to distinguish the statement or part of it and the communicative type of the statement, the semantic significance of the syntagms, and the division into topics and parts.

The term intonema first appears in textbooks and manuals on Russian linguistics published in the 1940s. Russian scholar V.A. Artyomov, who directly involved in intonation, linked the origin of the term to the name Reformatsky. [2]

According to the opinion of V.A. Artyomov, intonema is a functional unit with a typical pattern of various intonation elements, in this regard it is important to distinguish between the concepts of intonem and intonation invariant. Typically, the first term is a socially non-linguistic unit of speech in the context of a language system, while the second term refers to speaker speech as a relative model (structure) of the acoustic features required in a particular language.

An intonema is a unit of intonation, an intonation model, the result of the generalization of intonational qualities related to the meaning of a statement. Intonema provides intonational regularity, which helps to identify intonation-semantic segments in speech. The material of an intonema consists of intonation elements (stop, syntagma, rema, etc.), but this intonema does not mean a simple combination of these elements. The physical aspect of an intonema is usually characterized by the formation of generalizations based on the similarities of the pure intonation sounds present in the speech of certain language speakers. According to experts, a generalized set of intonemes is stored in the memory of speakers of a particular language (for example, Uzbek, Russian or German) and serves as a norm in the pronunciation or “comprehension” (listening) of specific parts of speech.

Linguists classify intonemes semantically and divide them into four groups. These are:

1) **mental intonema** (true intonema division, connection intonema level, significance intonema level, question intonema, affirmation intonema) serve to distinguish syntactic meanings and categories, to separate the flow of speech, and to compare the question with affirmation. Mental intonemes are always present in speech.

The values of intellectual intonemes are the moments of mental activity reflected in the language associated with the formation of speech,

that is, the use of intonation in one of two selected forms of expression: statement (transmission of information) or question (desire to receive information).

True intoneme division serves to divide a text into paragraphs, sentences, and syntagms, and is followed by a pause.

The communicative intoneme refers to the relationship between the actual units of division by component. It usually consists of melodious music, which slows down the pronunciation. True intoneme division and communication intoneme often interact in the flow of speech.

Significant intoneme is used to emphasize the most important parts of speech. It is characterized by slow speed and high density as a secondary elevation. Significance intonation usually refers to the logical effect of one or more words on a syntagm. For example,

*You are right. All conditions are well. We know that the conditions are right for any whirlwind,*

– said Omonullo, emphasizing the word “we”. It was a special sound that struck the boss in the brain like a hammer and shook him. But he laughed out loudly.

2) **arbitrary intonema** – directly, depending on the position of the speaker (communicator) in the speech environment. Naturally, the speech conditions are intertwined with the listener’s influence on the communicator during the discourse, the fact, the external influences (health, social status, level, worldview of the informant). The melody of voluntary intonemes is, of course, higher than the norm accepted by society, because even if the communicator’s mood is not reflected in the words, it is difficult not to express it in the tone of speech. For example,

*We left in the yard only two with my Kimsan aka. I shamed to be standing in front of him, and I was walking toward the yard for laundry, when Kimsan aka gestured with his hand: Sit down, Robiya... I have something to talk to you...*

*I have understood everything from the **word tone**. My heart pounded. He sat in front of me on the porch. There was a deep, thoughtful meaning in his eyes. Sweat dripped down his sunburned face, and his lips quivered. It’s as if he’s grown up one day ...*

It is clear from the passage that the speaker (Kimsan) knew something terrible about the listener’s father (Robiya) from his tone of voice.

Arbitrary intonemes are divided into two groups:

1) *arbitrary intonema* – neutral intonema (My heart pounded. He sat in front of me on the porch. There was a deep, thoughtful meaning

in his eyes. Sweat dripped down his sunburned face, and his lips quivered. It's as if he's grown up one day...

2) *stimulating intonema* (order, strong density, high range). The order is pronounced, the words are clearly separated and emphasized, and the tone is in a wide range of ups and downs. Options for implementing the order intonation: decision, desire of interest, persuasion, warning, request, etc. This intonation is characterized by wave-like music. The music is placed in a high range so that the interlocutor adapts to the interlocutor as usual. For example,

– “Comrade Captain, if you have anything to say, come and talk.”  
*There's some information on the phone that you can't tell, you know.*

*After a brief explanation of the address, the man forgave the connection. Omonullo understood from **their tone** that the case of the “**boy who speaks directly to Moscow**” was related to the State Secret Service. He also realized that it would not be easy to find and talk to someone with a “helicopter underneath”.*

In discourse, the speaker uses a stimulus intonation even when the goal is to turn an idea expressed in the speech into a basic speech reality (to provoke, to draw attention, sometimes to gain information through it). For example,

*Omonullo told him about his new views on Samandar Ochilov and the conclusions of the examination. Mels Khodjaev had promised to deal with the next case, which would be resolved through a military prosecutor. After Omonullo summed up the trip, he began to report, “**We didn't sit idle either, brother**”.*

– “I spoke to the district military prosecutor personally,” said Mels Khodjaev, with childish joy and childish pride, and looked at Omonullo as if expecting applause.

If we say something in order to prove or explain something (the content of which does not correspond to orthoepic norms), all the sentences can be an example of arbitrary intonation.

## Conclusion

A communicator is able to use all types of intonemes effectively in speech, based on his inner experiences, and emotional state of speech.

## REFERENCES:

1. Abduazizov A. Introduction to Linguistics. – T.: Sharq, 2010. 178 p., 22-p.

2. Artyomov V.A. Speech intonation // Collection “Experimental Phonetics”. Issue Number 3. – Minsk, 1971. – 201 p.
3. Yuldasheva Dilorom Nigmatovna, Gulchekhra Toshpulot qizi Chullieva. The role of nonverbal means in speech meaning//International Engineering Journal For Research & Development, Vol.6, Issue 2 published May 16, 2021.– Pag. 1956–1961. [www.iejrd.com](http://www.iejrd.com) <http://www.iejrd.com/index.php/%20/article/view/1956>
4. DN Yuldasheva. Basic non-verbal components of speech// Current research journal of philological sciences (2767-3758) 3 (01), 17–24. <https://masterjournals.com/index.php/crjps/article/view/661>
5. Chullieva G.T. Intonation. Intonema, its structure//Functions of intonema. Middle European Scientific Bulletin. Volume 8, January 2021. Pag. 141–148. <https://cejsr.academicjournal.io/index.php/journal/article/view/179>
6. G.T.Chullieva. Intonema and its Types//Middle European Scientific Bulletin. Volume 10, March 2021. ISSN 2694-9970. Pag. 91–95. <https://cejsr.academicjournal.io/index.php/journal/article/view/314>
7. Dilorom Nigmatovna Yuldasheva, Gulchekhra Tospulatovna Chullieva. Expression Of Euphemistic Meaning Through Silence And Intonation// Published/ publié in Res Militaris (resmilitaris.net), vol.13, n°2, January Issue 2023.–5596-5606-p. <https://resmilitaris.net/menu-script/index.php/resmilitaris/article/view/3187>
8. D.N.Yuldasheva, G.I. Toirova, D.Y.Yusupova. Silence is a Pragma linguistic event taking medical Ephemeric significance// Philosophical Readings XIII.4 (2021), pp. 2290-2297. 2290 [https://www.researchgate.net/publication/365726821\\_Silence\\_is\\_a\\_Pragma\\_linguistic\\_event\\_taking\\_medical\\_Ephemeric\\_significance](https://www.researchgate.net/publication/365726821_Silence_is_a_Pragma_linguistic_event_taking_medical_Ephemeric_significance)
9. Jamolkhonov H. Theoretical phonetics of Uzbek language. – T., 2009. 224 p.
10. uzsmart.uz. Tohir Malik. Corpses do not speak. Short story. – 162 p.



УДК 81-11

**THE IMPORTANCE OF PRAGMALINGUISTIC  
FEATURES OF LANGUAGE TASKS IN INTERNATIONAL  
COMMUNICATION**

*Khamroyeva Nafisa*  
*Bukhara State University*  
*Bukhara, Uzbekistan*  
nafisa@gmail.com

From the point of view of substantiality (dialectic), it can be said that any source, thing that is studied is more diverse, therefore its properties are also more diverse. amoral linguistics, the polyfunctionality of language – it serves to fulfill several tasks – is its general feature, which is unanimously recognized by all linguists. The article talks about speech communication and its importance, functions of language, dialogic speech.

### **1. Introduction**

Currently, in world linguistics, attention is being paid to the practical use of language capabilities, the realization of linguistic units, i.e., their use in speech, and the clarification of their meanings and functions. This process is explained by the fact that the human factor is becoming more important due to the development of technology around the world, the culture of speech, the effectiveness of communication, the view of language as one of the sources of raising our spirituality is growing.

The fact that new directions have already formed in world linguistics that investigate the organic relationship of language with speech, society, culture, national thinking, and even artificial intelligence puts new tasks before our science.

### **2. Literature analysis.**

In linguistics, language is interpreted as a diverse, multifaceted, multitasking (functional) social-spiritual-ethnic phenomenon. Therefore, different definitions of language have been given in European science. For example, Vladimir Zvegintsev in his book “Очерки по общему языкознанию” cited Wilhelm Humboldt’s three different definitions of language, along with Friedrich Hegel, August Schlicher, Heyman Steinthal, Alexander Potenia, Philip Fortunatov, Ivan Baudouin de Courtenay, Hugo Schuhardt, Otto Espersen, Benedetto

Croce, Cites various definitions by 21 other famous scientists, including Eduard Sepir, Anton Marty, Ferdinand de Saussure (2 definitions), Henry Ebbinghaus, Friedrich Keynes, Carl Fossler, Antoine Mee, Joseph Vanries, Vittore Pisani, Louis Elmslev. Naturally, the 25 different definitions of the language presented in this book by V. Zvegintsev are not all the definitions that exist in our science as a source of linguistic research. Because:

First of all, as you can see from the given list:

a) the author is limited to the definitions of Wilhelm Humboldt (1767–1835) and later linguists. However, from our grandfather Aristotle to Humboldt, dozens of linguists (including Franz Bopp, Jacob Grimm, Rasmus Rask, etc.) gave different definitions of language;

b) from this list, the definitions given by Eastern scholars were not paid attention to at all.

Therefore, this list is not complete. It should be noted that the author himself (Zvegintsev) does not claim that the list is complete or at least complete; he cites 25 different definitions of the language just to emphasize its versatility.

Secondly, after the 60s of the last century, empirical-pragmatic directions (socio- and pragmalinguistics, speech linguistics, text linguistics, text grammar, and finally culturology...) became popular in our science. With the transfer of its units and construction, system and structure, colorful relationships to actual language – discourse and its effectiveness, 8dozens of new definitions were given to the language – now from a pragmatic point of view, and these, *a priori*, we can say, are not the end – a new era, new conditions, including the information age and computer analysis technologies, open up new aspects of language, and this will continue indefinitely

### 3. Results

Speech is a powerful factor of a person's mental maturity, formation as a person. It is impossible to form, educate, and develop the intellect of a person without speech communication. For this reason, communication helps to organize the work of people together and to carry it out. All our social behavior in the objective world is reflected in our speech activity. Speech communication is the pronunciation of a certain sentence in a specific communication environment. The formation of the content of speech communication is the "enrichment" of the meaning of the expressed sentence in relation to the communication

text created by the communicator (speaker and listener), a product of perception. Thanks to speech communication, people perform various practical and theoretical activities. In addition, processes such as information exchange and mutual understanding also occur due to speech communication. It also helps to build interpersonal relationships and put them into practice. The main tool of speech communication is, without a doubt, language. After all, the main task of language is communicative task. We will return to the tasks of language in more detail.

Every creature and event/phenomenon that takes place in existence (nature and society), can affect our senses, can be studied and described (philosophical *uzb* .narsa, *ar*. zot, *russian*. вещь, *german* .die Sache.) in nature and in essence it is brilliant. Comprehensively is a common inherent (ontological, natural, divine...) characteristic of beings and events in existence. The reason for diversity is that the desired philosophical thing stands in various relationships with various things and events in existence, and the same thing shows its own edge with each type of its various relations.

One of the most important requirements of the theory of dialectical knowledge (gnoseology) is to always consider the diversity of the thing (the source of study/research). Therefore, in dialectics, it is never possible to make an abstract and firm judgment, “ This is such and such “ . The truth of each judgment should be reflected in the definition in relation to which thing, in what relations and in what cases. Only then will the verdict be clear. That’s why in dialectics, the specification of judgment is evaluated as a *rise from* abstraction to concreteness . However, we are used to evaluating going from specific (particular) to abstract (general) as *an ascent (induction)* , and going from abstract (general) to specific (particular) as *a descent (deduction)* . In the dialectical scientific thinking, the situation is evaluated as a particularization, the opening up of relationships, and the emphasis on the inherent diversity of things. Values disconnected from relationships are abstract, groundless, scientific error, disconnection from the nature of things, disconnection from the chain of continuous development, and absolutization is a denial of the richness of things.

When the source of the research is diverse, no judgment about it (except for the recognition of its diversity) can be absolute: each judgment must necessarily indicate the truth of the particular side of the thing and the type of communication process. Now let’s go to the description of **the tasks of the language , which is our immediate goal.**

#### 4. Discussion

From the point of view of substantiality (dialectic), not only the source itself, but also every part of it is rich. Therefore, if the language itself is versatile, the task of the language, which is one of its aspects, is also versatile, that is, the language is multifunctional. Therefore, we will briefly dwell on this aspect, which should be explained in our section, because the fact that is the goal of our research is related to one of the important tasks of the language. modern linguistics , the polyfunctionality of a language – it serves to fulfill several tasks – is its general feature, which is unanimously recognized by all linguists. The function of the language means what the language serves, what is its importance for the person who owns it, and its place in the society. In this, the nature and structure of the language, the nature of its units, the methods and results of their interconnection, and hundreds of other matters (language aspects) are disconnected, and attention is paid only to this direction – the functions of the language.

Alisher Navoi also expressed very valuable thoughts about the versatility of language and its performance in society in the introduction to the epics “Khamisa”, “Mahbub ul-qulub” and, in particular, in his special work dedicated to linguistics – “Muhokamat ul-lughatayn”:

***The day turned into a happy day,  
Kilgan is always a classic with his speech***

(Description: *He demonstrated the activity of creating truth and made man superior to all creatures with speech* ). After this verse, the Holy Qur’an says “***allamal adam asma’a kullaho***” (Sura 2, verse 27). Referring to the verse “He taught the name of all things to man”, he emphasizes that the language is a great divine gift to man and writes: “*The name “almutakallim” made me happy, so that this knowledge became a blessing to the entire creation and this visit was a privilege*” (Description): *Truth gave man the ability to speak and express his thoughts, and with this ability, man became superior to all creatures and became the head of them with this honor* ).

While describing various features and aspects of the language, the researchers define a specific function of the language in connection with the goals and research methods and state their conclusions about it. B.Norman in his book called “Tasks of language” and a special chapter that occupies about 30 pages lists 15 different tasks – features of language one by one: communicating (communicative task), forming an opinion (cognitive task), educational (accumulative) task , nam-

ing/naming (nominative) task, encouraging-directing (regulatory) task, communication initiation (phatic) task, etc. He concludes his opinion with the following words: “In addition to the tasks specifically described above, a number of other socially important tasks of language can be distinguished. In particular, in addition to certain functions such as *ethnicity*, which unites a certain people, *aesthetic*

, which transforms ordinary text/information into a work of art, and *emotional-expressive*, which serves to express the feelings of the speaker, the *magic function is also characteristic* of language. But these are not all manifestations of the social status of the language. “It is clear from this that the tasks of the language are many, colorful and diverse.

Later, many linguists, including N. Slyusaryova, in the article “Language functions” emphasize the need to determine the functions of the language on the basis of hierarchy (hierarchy). 1) the main tasks of the language; 2) main tasks; 3) three levels are distinguished as additional tasks. The main tasks *of* language are the following two tasks:

1) to be “worldly medium”;

2) we understand the formation of the product of consciousness, its expression in the form of words and sentences.

The first task is called **a means of communication** (*communicative task* /abbreviated SF/), and the second is called an **educational task** (*cognitive task* /abbreviated KF/).

Language *are* as follows:

a) to be a means of expressing the feelings and attitudes of the speaker – the task of expressiveness (*emotional task* /abbr. EF/);

b) a generality that unites the three features listed above, the function of gender as a translation of the essence/genus (*metalingual function* /abbr. MF/).

*additional functions of the language*, it is necessary to dwell on the relation of the functions of the main and main two stages, which are unanimously separated by scientists and listed above. If we compare the tasks that are divided in these two stages, it can be seen that the characteristics given as the main task are connected with the *generality ~ particularity dialectic relations shown as the main task*. In particular, the first main task assigned as /EF/ is essentially based on /KF/, which is specified as one of the main tasks. Only in this, not the events of existence, but the mental state, experiences and attitude of the person (speaker) lie on the ground of the content formed by the language. Language forms these situations and relationships based on

/KF/, which is the main task. The next main task is to communicate to others based on /CF/.

*Meta-linguistic function of the language /MF/* and its connection with the main tasks do not cause any problems from the point of view of the dialectic of generality ~ particularity (including mysticism), because dialectic (mysticism) affects the desired meaning (specificity and individuality, arial realization, sense organs) evaluates the realization in the form it takes) as a reflection, reflection of a certain whole (generality, essence, breed...). In this case, naturally, each feature (tajalli) reflects the general features of the breed (in modern linguistic terms, system relations and structure in part). Therefore, the general outline (units, systemic/system-structural/ relationships between them) of the common language (metalanguage), which is a form of socio-psychological phenomena, is reflected in the desired language. That's why, regardless of the nature of the language (the level of distribution, how many people speak this language, and the characteristics of development, etc.), the content expressed by the means of one language can be easily expressed by the means of another language. Because a desired language is the interpretation, perception, form of realization, appearance of a common language (metalanguage). Therefore, content in a second language can be provided in any desired language (of course, using borrowed words from other languages). The basis of this is the same – the two main tasks of the language listed above. Therefore, /MF/, which is considered as the second main feature, is not an independent separate task, but one of the special manifestations of those main tasks, essentially, it is the exaggeration of some aspect in a certain situation, the strengthening of the situation (required by the context, conditions...).

## 5. Conclusion

The dialectic unity, wholeness, and complex of language aspects (i.e. tasks, sides, features) has the ability to embody the signs and characteristics characteristic of all the listed aspects in a natural (natural, divine, substantial) manner, and it can manifest the desired aspect based on the requirements of the situation. But in this process, its other aspects do not disappear completely – it strengthens the manifestation of its characteristic for this relational system under the influence of a certain condition/situation and its various means while keeping its essence in this system, and weakens (neutralizes, excludes) the other aspects. but it does not lose completely – the brighter manifestation

(intensification, exaggeration...) of a certain side comes at the expense of weakening, darkening, relaxation of another side. The language is the same. The two main (genital, financial, substantial...) tasks that determine its essence – /CF/ and /KF/) – are never completely disconnected from each other, and occur in a hundred different conditions in as many forms (that is, additional tasks) based on the requirements of specific conditions. Therefore, it can be boldly said that the additional functions of the language, which appear to be essentially unlimited, complement, decorate, specify its two main functions – /CF/ and /KF/, and reveal new aspects of the breed. Therefore, completing the description of the diversity of language tasks, the two main tasks of the language – /CF/ and /KF/ are “decorated” and “decorated” in various contexts with a number of additional tasks, one of such additional tasks is the task of initiating communication (phatic task /q. A general conclusion can be drawn that FF/).

## REFERENCES:

1. Khamroeva N. N. THE IMPORTANCE OF COMPLIMENT IN THE NORMS OF SPEECH ETHICS //Scientific progress. – 2021. – Т. 2. – №. 7. – С. 1213-1217.
2. Низомиддиновна Ҳ. Н. ФАТИК МУЛОҚОТ ТУРЛАРИ ВА ИЛТИФОТ ҲАҚИДА //BOSHQARUV VA ETIKA QOIDALARI ONLAYN ILMIY JURNALI. – 2022. – Т. 2. – №. 2. – С. 27–31.
3. Слюсарева Н.А. Функции языка. Лингвистический энциклопедический словарь. – Москва: СЭ., 1990 С. 564.
4. Kuldoshev R. et al. Mathematical statistical analysis of attainment levels of primary left handed students based on pearson’s conformity criteria //E3S Web of Conferences. – EDP Sciences, 2023. – Т. 371. – С. 05069.
5. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019. – P. 352–355. (scopus)
6. Toirova G, Abdurahmonova N., Ismoilov A., Applying Web Srawler Teshnologies for Sompiling Parallel Sorpora as one Stage of Natural Language Possessing. 2022 7th International Conference on Computer Science and Engineering (UBMK) Sep. 14–16, 2022, Diyarbakir /Turkey pp. 73–75. (scopus)

УДК 81-11

**ЭПИСТОЛЯР ПУБЛИЦИСТИКА ВА МАКТУБОТНИНГ  
БОЛАЛАР АСАРЛАРИДАГИ БАДИИЙ-КОМПОЗИЦИОН  
ВАЗИФАЛАРИ**

**Жамилова Башорат Сатторовна**  
*Бухоро давлат университети Ўзбекистон*  
bashoratjamilova@gmail.com

Мақолада болалар матуотида эпистоляр публицистика намунаси бўлган хат жанридаги мақолалар, шунингдек, болалар адабиётида мактубот шаклида ёзилган қисса ва ҳикояларнинг долзарб аҳамияти ҳамда бадий-композицион вазифалари ҳақида фикр юритилган.

**Калит сўзлар:** хат, эпистоляр, публицистика, мактубот, қисса, ҳикоя, мақола, дидактика, тамсил, кундалик, суҳбат.

**ARTISTIC COMPOSITION TASKS OF EPISTOLARY  
PUBLISHING AND LETTRISH IN CHILDREN'S WORKS**

**Djamilova Bashorat Sattorovna**  
*Bukhara State University, Uzbekistan*  
bashoratjamilova@gmail.com

The article reflects on the actual meaning and artistic and compositional functions of articles in the genre of letters, which are an example of epistolary journalism in children's literature, as well as stories and stories written in the form of letters in children's literature.

**Key words:** letter, epistolary, journalism, letter, story, novella, article, didactics, example, diary, conversation.

**ХУДОЖЕСТВЕННО-КОМПОЗИЦИОННЫЕ ЗАДАЧИ  
ЭПИСТОЛАРНОЙ ПУБЛИЦИСТИКИ И ЛЕТРИСТА  
В ДЕТСКИХ ПРОИЗВЕДЕНИЯХ**

**Б. С. Джамилова**  
*профессор Бухарского государственного университета,*  
*Узбекистан*  
bashoratjamilova@gmail.com

В статье размышляется о актуальном значении и художественно- композиционных функциях статей в жанре писем, являющихся образцом эпистолярной публицистики в детской литературе, а также рассказов и рассказов, написанных в форме писем, в детской литературе.



**Ключевые слова:** письмо, эпистолярный, публицистика, письмо, рассказ, повесть, статья, дидактика, пример, дневник, беседа.

Жаҳон публицистикаси ва адабиётда хат шаклида ёзилган шеър, поэма, ҳикоя, қисса, романлар борлиги эътироф этилган. “Эпистоляр сўзи, – деб ёзади М. Худойкулов, – лотинча “epistola” – хат, юборилган маълумот деган маънони билдиради. Бу жанр адабий-бадиий ёки публицистик асарнинг хат, мактуб тарзида ифода этилишидир”[6,205].

Жумладан, болалар публицистикасида ҳам **хат жанри** услубида ёзилган мақолалар алоҳида ўрин ва мавқега эга. Публицистик хат газета ва журналларда материалларнинг таъсирчанлигини оширишга хизмат қилади. Умумпублицистикада хат шаклининг қўлланиш тарихи, турли кўринишлари ва аҳамиятли жиҳатлари атрофича ўрганилган [2, 60]. О.Тоғаев газета-журнал материалларининг ёзилиш услубида хат жанри хусусиятларининг мавжудлиги ижтимоий пафоснинг шахсий ҳис-туйғу, кечинмалар орқали ифодаланиши, газетхонга бевосита мурожаат кабилар ижтимоий-сиёсий воқеликка жозиба бахш этиши ва ўқимишли қилишини алоҳида таъкидлайди [4,10].

Худди шу усул болалар матбуотининг ҳам негизини ташкил этади ва аксарият мақолалар кўпроқ хат шаклида ёзилгани билан эътиборни тортади. Масалан, сарлавҳа ўрнида қўлланган “Бувимга мактублар” хати “Тонг юлдузи” газетасининг бир неча сонида берилган.

Мактублар Хитойнинг Сзинзянь шаҳрида ота-онаси билан чет эллик талабалар ётоқхонасида яшайдиган Комрон Аминов томонидан тошкентлик буви ва бобоси номига йўлланган бўлиб, ижтимоий характерга эга. Хусусан, хат муаллифи Конфуций ватани бўлмиш Сйуйфу (Syuyfu)нинг машҳурлиги сабабларини ойдинлаштиради.

Болалар ижодкорлари ёхуд умуман адибларнинг болаларга ёзган мақолаларида хат шаклини танлашлари бежиз эмас. Чунки хат шакли муаллифга ўқувчини яқин дўст, маслакдош тутиб, дилдан ёзиш имкониятини туғдиради. Бунда табиийлик устувор. Бинобарин, “Муборак кунларнинг азиз фарзанди”[7], “Билиб қўйки, сени Ватан кутади”[8], “Ойнинг ўн беши ёруғ”, “Она юртинг – олтин бешигинг”[9]; “Иншо ёки ўқитувчимга хат”[10] каби мактубларда айни шундай ижтимоий пафос етакчилик қилади.

Бундан ташқари, болалар адабиётда ҳам мактубот, хат шакли-

дан фойдаланиш кўп кузатилади. Масалан, Э.Маликнинг “Онаизор” кассасининг композицияси эпистоляр негизга эга. Асарнинг ҳар бир боби онанинг ўғлига йўллаган мактуби шаклида ёзилган. Шу орқали мактубларда келтирилган ранг-баранг воқеалар ягона махражда боғланади.

Асардаги Она образи барча ёлғиз аёлларнинг умумлашма образига айланган бўлса, отасизлик нафақат оила тўқислигига, балки, бола руҳиятига озор етказиши кичкинтой Шернинг болалик изтироблари мисолида очиб берилган. Бола оиласини ташлаб, ҳамма нарсани арокқа алмаштириб кетган бўлса-да, дадаси борлигига ўртоқлари билан гарров ўйнайди, бунга ишонмаган бир ўртоғини судраб отасининг вайрона кулбасига олиб келади, каттакон кулф осилганига қарамай, кечгача шу ерда қолиб кетади: “Кетмайсан, кетмайсан, дадамни кўрасан, дадам бор... ахир, ҳозир келадилар”, – дея ўртоғининг ёқасига ёпишади; йиғлаб туриб, ёлбора бошлайди. Боланинг ўртоғи олдида изза бўлиб, папқасининг устига ўтириб, бошини тиззалари орасига олиб ҳиқ-ҳиқ йиғлай бошлаши ҳам Она мактубида эсланиб: “Сен ҳеч қачон, ҳа, ҳеч қачон ўз фарзандингни бунақа аҳволга солмайсан. Бу азобни гўдак бошингдан ўтказдинг”, – дейилиши асар марказида кўпроқ бола руҳияти, ўксик қалб жароҳатлари туришини ойдинлаштиради.

Онанинг мактубларида шу тариқа оилада отадан болага ўтадиган фазилатлар, гап сўзларидаги дадиллигу мустақилликни аёл барибир бера олмаслиги, билъакс, уни тергайвериб, ўта кўнгилчан қилиб кўйганидаги пушаймонлик аёнлашади.

Муҳими, хатларнинг умумлашмасида ҳам композицион бутунлик кўзга ташланади: мактубдан-мактубга ўтилгани сайин воқеалар ривож кульминацион нуқтага кўтарилади. Шер бола бўлиб, отасиз ўсганлиги учун бахтиёрликни ҳис қилмаган эди, ўсмир ёшида отасини излаб бориб, ундан кўнгли совийди. Энди фақат онасига суянчиқ, елкадош бўламан, деганда эса ҳарбий хизматга жўнайди... Муштипар онасининг орзуларини рўёбга чиқариш учун ҳамма нарсага рози бўлади... Гарчи, киссада фақат онанинг ўғлига ёзган мактублари келтирилган, ўғилнинг онага ёзган бирорта жавоб хати бўлмаса ҳам, фарзандининг онасига меҳрибонлиги, сирдошлиги англашилиб туради. Дамно Афғон фожиаси она ва боланинг бу умидини ҳам саробга айлантиради. Даставвал, аёл кизи Сабоҳатни безорилардан қутқарган, кейин унга кўнгил бериб, совчи кўйган Улуғжон исмли ҳарбий йигит учун кўк кияди, сўнгра эса... бунинг ёнига Шернинг азаси қўшилади...

Э. Маликов ижодида қиссанинг яна бир кўриниши кундаликда битилган **мактублар** шаклидадир. Муаллифнинг “Етим қолган хотиралар” қиссасида ўсмир қиз Маъсуданинг мактуб-хотиралари орқали болалар уйларидаги кўнгли ўксик ўғил-қизларнинг қалб кечинмалари, ота-оналари, оила соғинчи, жамият ҳақидаги та-саввурлари акс эттирилган. Натижада бу қисса ҳам янгича ғоявий композицион асосга эга бўлган [5,76].

Кундалик ёзиш ўсмир ёшдаги қиз-йигитларга хослиги назарда тутилса, танланган услубий йўналиш реалликка эга эканлиги аёнлашади. Шунингдек, болалар уйи, майшатга берилган ота-оналарнинг ўз болаларини тирик етим қилиб қўйишгани ўзбек болалар адабиётида қаламга олинмаган мавзулардандир. Отаси чет элга ишлашга кетиб, унинг касал бўлганлигини эшитган она бориб олиб келиш баҳонасида болаларини шу ерга ташлаб кетади.

Ёзувчи кундаликда муҳрланган қалб изҳорлари – **хатлар** баҳонасида етимхонада тарбияланаётган Ҳалима, Мадина, Солиҳа, Наргиза исмли қиз ва Одил каби ўғил болаларнинг аччиқ қисматларини тилга олади. Бу тақдир эгаларининг ҳар бири алоҳида ҳикоя. Шу сабабли муаллиф асосий воқеа ривожидида бу ҳикояларни қолипловчи ҳикоя тарзида беради. Маъсуданинг ойижонисига ёзган кундалигидаги мактубларида Дармонжон, Наргиза, Одил қисматлари жуда таъсирчан ҳикоя қилинган. “Дармонжон дунёдаги ҳамма оналарни ёмон кўриб қолган экан. Бечора, бегона одамни ота дея олмагани учун бу ерларга келиб қолган экан- да. Хўрлан-саям, ўлимга рози бўлсаям, барибир бегонани “ота” демапти. Мен ҳам худди шундай қилган бўлардим...” Маъсуданинг мактубларида Дармонжоннинг онаси мисолида ўз фарзандини эртақлардаги ўғай оналардек калламушли ертўлага қамаб, бўғиб ўлдирмоқчи бўлгани, бундай воқеалар таъсирида етимхонадаги ҳамма болалар ҳам барча оналарни ёмон кўра бошлаши таъкидланади. Ёки Наргиза исмли қиз тимсолида “етимхона касали”га йўлиққан болалар аҳволи баён қилинади. “У соғиниш касали! – уқтиради қизалоқ. Болалар оналарини ўйлаб-ўйлаб, ичларидан куйиб-куйиб, охири ётиб қоладилар... Биз ўн тўққиз қиз- кулсақ-куламыз, йиғласак – йиғлаймиз. Дардимиз битта-да, дардимиз. Етиммиз!”

Болалар адиби Э.Маликнинг “Набираларимга мактублар”[3] асари ҳам шу анъананинг новаторона давоми сифатида эътиборга лойиқ. Адиб сўзини неварога мактуб ёзишдан бошлаган бўлса-да, асарнинг бутун композициясида уларнинг ҳар бири алоҳи-

да сарлавҳалардан иборат. Унда келтирилган мустақил сюжетли ҳикояларда бобонинг – муаллифнинг неварасига ғойибона айтган гаплари, хотирасига келган воқеалар орқали шу замон ва маконнинг долзарб мавзулари, ахлоқий сабоқлар, тарбиявий мақсадда изчил баён этиб борилади. Айниқса, болаларга мавҳум тушунчалар, масалан, “Ишонч”, “Диққат”, “Ҳадя”, “Ҳавас ва ҳасад”, “Суянчиқ”, “Меҳр” кабилар реалистик воқеалар, бадий деталлар орқали талқин этилиши, деярли ҳар бир ҳикоя-мактубнинг “набиграгинам”, “болажоним” деб бошланиб, асосан, барча болаларга қаратилган ўғиту насихатлар экани асарнинг дидактик аҳамиятини таъминлайди.

Маълумки, жаҳон болалар адабиётида кичкинтойлар тасаввурининг ва ҳаёлот оламининг серқирра бўёқларини, бадий ифодасини кўриш мумкин. Бинобарин, болалар шоири ва носири А.Обиджон ҳам ўзининг “Ғалати мактублар”и билан эпистоляр адабиётнинг болаларбоп завқовар янги туркумини яратди, дейиш мумкин. Шоир мазкур туркумдаги йигирма еттита шеърда архаик ва тарихий сўзлар, бугунги ёшлар деярли ишлатмайдиган, ҳатто уларнинг нималигини ҳам унутаёзган буюмлар, миллий қадриятлар, уй анжойлари ва жиҳозларини жонлантиради. “Калишнинг ботинкага”, “Чайнамайшимингнинг хўрозқандга”, “Кўранинг қумғонга”, “Сополкосанинг чинникосага” сингари кўплаб “ёзилган хат”ларидаги киноя, кесатик, истехзо ва мажоз, тамсил сингари бадий унсурлар кулгининг турли хилларини намоён қила олганидан ташқари, шеъринг мактубнинг бир намунаси сифатида эътирофга лойиқ. Яъни, айнан шу предмет ва буюмларнинг мактуби орқали асори-атиқага айланиб бораётган тарихий нарса-буюмлар, сўзлар замондош ўқувчига етиб келади. Албатта, уларнинг бадий ифодасида хилма-хиллик масалан, “Карнайнинг сурнайга”, “Қутичанинг сандиққа” “Пўстиннинг айиқчага ёзган хати” кабилар мактуб композициясини ўзида тўлиқ намоён қилса, “Шамнинг лампочкага ёзган хати”, “Тандирнинг чакмоққа ёзган хати” кабиларда киноявий тамсиллар устувор. Шунингдек, шеърлардаги “Чайнамайшиминг”, “Чориғ бобо”, “Шиппак опа”, “Хиром ака”, “чўтир Кирза тоға”, “Маҳсибиби”, “Амиркони хола” сингари антропонимлар ҳам аслида ана шу композицион шакл моҳиятини уйғунлаштириш билан бирга болалар тилини намоён этади. Бу эса ўқувчида ижтимоий ҳаётдан узоқлашаётган воқелик, тарихий деталларни жонли мулоқотга хос эпистоляр шаклда тафаккурида қайта тикланишига ёрдам беради.

Дарҳақиқат, хат жуда қадимий, ёши улуғ, “мўйсафид” жанрлардан бўлиши билан бирга, кўплаб адабий-бадий, публицистик асарларда кенг ва новаторона қўлланилаётгани билан бирга болалар адабиётида ҳам ўзига хос бадий-композицион вазифаларни бажариб келаётгани эътиборга лойиқдир.

#### АДАБИЁТЛАР:

1. Карасев П.С. Проблемы теории публицистики. – Л.: ЛГУ, 1973;
2. Муҳаммадиев Р. Публицистика жанрлари. – Т.: Ўзбекистон, 1965;
3. Малик Э. Набираларимга мактублар. – Тошкент: Шарқ, 2019. – Б.368.
4. Тоғаев О. Публицистика жанрлари. – Т.: Ўқитувчи, 1976.
5. Сафаров О., Жамилова Б. Болалик боқийлик тимсоли ёхуд “мухбирликнинг олис йўллари”. – Тошкент: Шарқ, 2011. – Б. 76.
6. Худойкулов М. Журналистика ва публицистика. – Тошкент: тафаккур, 2011. – 205-б
7. Муборак кунларнинг азиз фарзанди // Тонг юлдузи. – 1992. – 8 янв.
8. Билиб қўйки, сени Ватан кутади // Гунча. – 1992. – № 1. – Б. 1.
9. Ойнинг ўн беши ёруғ // Гулхан. – 1993. – № 1. – Б. 1.
10. Она юртинг – олтин бешигинг. Иншо ёки ўқитувчимга хат // Тонг юлдузи. – 1993. – 29 сент.

## УДК 81-11

## LINGUISTIC CULTURE UZBEK CULTURE

*Jumayeva Dilnoza Bakhshulloyevna*

*Navoi State Pedagogical Institute teacher*

*d.jumayeva@gmail.com*

This state provides information about the linguocultural nature of the name in the Uzbek language, and also explains why each name has its own characteristics. Names have social significance and are explained because they reflect the culture and social environment of a society.

**Introduction:** prangmalinguistics, linguistic culture, anthroponym, cognitive linguistics, names, linguistic stylistics, anthropocentric paradigm.

Much attention is paid to the primacy of the anthropocentric paradigm in studies of pragmalinguistics, cognitive linguistics, psycholinguistics, and sociolinguistics in world linguistics. The pragmatic, cognitive and linguo-cultural significance of anthroponyms requires a deep study of the place of anthroponyms in the literary text and the disclosure of conceptual information.

Names have social significance, and they reflect the culture and social environment of the community. Names also inform about the generation of a certain person, his place in society. Each name is unique for a reason. It can be popular or underused. Each name has its own history, “biography” and “geography”. Names can be linguistically real or fictional. Names can give certain information about their object.

This information takes three forms:

- 1) linguistic appearance;
- 2) speech and appearance;
- 3) overview.

As a language unit (linguistic expression), they name individuals and separate individuals from each other.

Public information is based on the appellative meaning based on the anthroponym: the names Gulnara, Gulandom, Gulrukhsar are based on the symbols of delicacy and fluidity characteristic of flowers.

Nationality is reflected in the names. Names have a sense of antiquity or freshness. For example: Alpomish, Amir Temur, Jalaluddin Manguberdi, Barchin, Tomaris.

Names derived from appellatives have the phenomenon of synonymy (Stop, Tursun), but there is no polysemy. The phenomenon of synonymy is quite common. For example, *Abdul Mutallib ibn Hashim – Shauba, Abdul Mutallib – Abu Lahab*.

It seems that anthroponyms are, first of all, words, they are a part of language. Names have the following pragmatic features: according to the perlocutionary feature of the names: *ijobiu* (beautiful name, beautiful name, beautiful name, decent name, meaningfully beautiful name, good name) or *salbiu* (meaningless / ugly) name, pompous name, humiliating names, inappropriate names, inappropriate names, ugly names, names with ugly meaning, uncharacteristic names: *Dunbau* (junk), *Turob* (soil) *Tabdila* (replaced), *Anzur* (untimely), *Sotiboldi* (unrelated to the family); according to gender: boys' names (*Nishonbek*, *Artiqbou*, *Koshokbou*, *Tursun*, *Turdi*, *Ushar*), girls' names (*Ziouda*, *Ogilou*, *Khumor*, *Khumora*, *Orzigul*, *Orzuniso*); has a national and ethnic background: *June*, *Euphemia*, *Frances*, *Nick* (English); *Alexander*, *Marina*, *Sereja*, *Valois* (Russian); *Khumouon*, *Oubek*, *Alpomish* (Uzbek), *Elmira* (Tatar); according to religion: *Ibrahim*, *Ismail*, *Ishaq*, *Ya'qub*, *Yusuf*, *Lut*, *Auuub*, *Uunus*, *Musa*, *Harun*, *Iluos*, *al-Yasa'*, *David*, *Solomon*, *Zakariuo*, *Yahuah*, *Jesus*, *Muhammad*.

The development of linguistics, such as cognitive linguistics, prefigured the development of linguistic research in the last quarter of the 20th century. Learning the language in connection with the culture and thinking of a certain ethnic group prevailed. The study of the triad "Language – culture of the nation" has become the center of a relatively new field of linguistics – linguocultural studies. The unit of study of *lingvokulturologiua* is *lingvomadaniuat*, which is a dialectal unit of linguistic and extralinguistic (conceptual and subject) content.

Therefore, the synthesis of linguistic meaning and cultural meaning is the plan of expressing *lingvomadaniuat*. Cultural linguistics is a science related to sociolinguistics, and its scope includes cultural facts and their influence on the activity of the language sign.

The composition and part of the anthroponymic formula is less common, such as patronymic (name according to the person), paternal name). Today, most peoples are characterized by two forms of formula – personal name (or several personal names) and family name (or several surnames).

The anthropological formula "name + patronymic + family name" was established in all the nations of the CIS, despite the initially established naming system. Only a few peoples have managed to preserve the uniqueness of their national names, for example, Azerbaijanis: *Aliuev*, the son of *Husaun Israfil*, for a man, *Mammadova*, the daughter of *Zeunab Qurbanali*, and others.

Although the official anthroponymic model is preserved, the na-

tional identity is realized. Anthroponymic models of modern peoples of Europe and America and many peoples of Asia consist of binomial, name, personal name and surname. Undoubtedly, this model together with personal data (data on gender, marital status, last name, etc.) provides an appropriate measure of individualization of a person. At the same time, the structure of personal name and surname may be unclear. For example, the number of names for Germans is not officially limited. Usually, a newborn baby is given one or two (rarely three or more) names, one of which is the asosiu and is usually called a person when addressed. Two surnames are also one surname.

In Russian, surnames with consonants can be taken by both men and women (Volk, Korotich, Gusak, etc.), but the difference is made by shortening the male surname (Volka, Korotich, Gusak) and the surname. Immutability of Auol family name. Exceptions are surnames in -ih, -ih, -o (Shchevchenko), which are not declined in any case. However, in full naming, there is a significant difference between personal names and patronymics.

Many Western nations have different forms of first names, but not surnames.

Many peoples of the world have their own semantics of nouns – male and female – according to different semantic models. The roles of men and women, father and mother, husband and wife in naming are not equal. Often, the role of a person in the naming process is important, as evidenced by the patronymic of peoples who have already adopted the patronymic system. Usually, a woman takes her last name from her father or husband. On the contrary, although it is allowed by law, it is less common.

The degree of individuation through anthroponymic models of men and women is asymmetric in the Uzbek people (as well as in other nations of the world), but currently there is a tendency to preserve the family name of women in marriage.

Although the anthroponymic model has a second name or more personal names, it is mainly two- component (first name + last name), because it did not include an intermediate component – the patronymic. Modern English anthroponymic formulas were formed primarily on the basis of Old German onomasticons, which included layers of Old German ideas about beautiful and meaningful, as well as zoophoric groups of names. The reason for giving a name was to preserve the name of the nearest relative or to change it. Modern changes of ancient onomastics lead to giving new meaning to old concepts.



Personal names of Uzbeks are distinguished by their uniqueness among other Turkic-speaking peoples and among other peoples of Islamic culture. An important role was played in the development of the list of modern Uzbek names by the dictionary of the ancient Turkish and Persian languages.

Language and culture are an integral part of the structure of knowledge that reflects the history, religion, mythological events, customs and traditions of the people of this nation like a mirror. It includes, first of all, famous names among the representatives of a language – well-known persons, mythological and religious heroes, characters of works of art. Over time, these names are recognized as a set of national characteristics (character, behavior, appearance, national status) that are fixed in their composition and have social significance for that community. These anthroponyms are precedents, elements of national-linguistic consciousness, general views of cognitive awareness and understanding specific to a linguistic community. On the one hand, allusive anthroponyms refer to the owner of the name and express an indexical meaning. On the other hand, the relation and connection of this meaning with the expressed object is special, and the allusive anthroponym acts as a symbolic sign. In addition, it is also considered an iconic symbol, modeling a particular state or situation.

In the Uzbek language, lexemes such as mother, father, mother, aunt, uncle, uncle, grandfather, grandmother, etc., have a social sign, meaning the concept of kinship; husband, wife, daughter-in-law, kuuov, kaunona, kaunota, kaunisingil, ovsin, meaning the social role of a person in the family; lexemes such as old man, old woman, old woman, which mean the characteristics of washing, also have a complex semiotic structure, and when used in speech, the central meaning of the word is expressed directly, and the border meaning is expressed through the presupposition. For example:

1. ***Onamiz ketimdan tosh otib qoldi.*** (T. Murod. “Otamdan qolgan dalalar”)

2. ***Momolar o‘gimagan-da.*** (T.Murod. “Otamdan qolgan dalalar”)

The lexeme mother used in the first sentence is “1 A woman who has children (in relation to her own children); means a woman with a child. Through this lexeme, the central presupposition in the form of “Our mother has children” is restored. At the same time, this lexeme refers to the presupposition “Our mother is auol” because it describes a person in terms of gender.

The lexeme *momo* used in the Keuingi sentence is a unit specific to the dialect and is used in relation to mother or grandmother (mother or father's mother). When this lexeme is applied to a mother, it means a person, as well as the presupposition "Momo bu – auol" that refers to her gender; and when applied to grandmother, it brings up the presuppositions "Momo – old woman" indicating that she is older in terms of gender and age.

In the Uzbek language, lexemes such as *kampir*, *chol*, *kariua* also have a multi-component semitic composition, and presuppositions are expressed through their semitic components in speech. For example: *Ortiqalining engashib yuradigan kampir ammasi bor edi...* (X.To'xtaboev. "Shirin qovunlar mamlakati")

The presupposition "Old aunt who leans over and hits" is expressed through the semiotic component "old" which forms the dictionary meaning of the lexeme "old woman" used in this sentence.

In conclusion, it can be said that the personal names of Uzbeks are unique among other Turkic-speaking peoples and among other peoples of Islamic culture. In all languages of the world, individualization of the person can be distinguished as a universal. For example, common nouns (appellatives) according to the following parameters: gender, age, family status, social status, features of appearance, physical condition, character, intellect, etc.

#### ADABIYOTLAR:

1. Мирзиёев Ш.М. Эркин ва фаровон, демократик Ўзбекистон давлатини биргаликда барпо этамиз // Ўзбекистон Республикаси Президенти лавозимида киришиш тантанали маросимида бағишланган Олий Мажлис палаталарининг қўшма мажлисидаги нутқ. – Тошкент: Ўзбекистон, 2016. – 56 б.

2. Абдуллаев А. Ўзбек тилида экспрессивлик ифодалашнинг синтактик усули. – Тошкент: Фан, 1987. 15-б.

3. Абдухамидова Д. Япон ва ўзбек тилларида қариндош-уруғчиликни ифодаловчи лексик бирликларни системалаштиришга оид // "Шарқшунослик" илмий журнали. – 2020. №2 – Б.34

4. Степанова Ф.В. Русский женский антропонимикон в культурно-генетическом аспекте: Автореф...дисс.канд.филол.наук. Лесосибирск, 2006.-23 с.

5. Сатторов Ф. Ўзбек исмлирининг туркий қатлами: Филол. фанлари номзоди ... дисс. автореф. – Тошкент, 1990. – 21 б.

6. DY Yusupova. Poetess Khalim Khudoyberdieva and his skill in use of the word. International Scientific Journal ISJ Theoretical & Applied Science

Philadelphia, USA issue 01, volume 93 published January 30, 2021.–Pag. 238–241. <https://www.elibrary.ru/item.asp?id=44644284>

7. Худойназаров И. Антропонимларнинг тил луғат тизимидаги ўрни ва уларнинг семантик-услубий хусусиятлари: Филол. фанлари номзоди ... дисс.автореф. – Тошкент, 1998.

8. Jumayeva Dilnoza, Toirova Guli. Pragmatic peculiarities of anthroponyms// International scientific and practical conference cutting edge-science March, April 2021 Shawnee, Usa Conference Proceedings.

9. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019. –P.352-355. (scopus)

10. Toirova G, Abdurahmonova N., Ismoilov A., Applying Web Srawler Teshnologies for Sompiling Parallel Sorpora as one Stage of Natural Language Prosessing. 2022 7th International Conference on Computer Science and Engineering (UBMK) Sep. 14 – 16, 2022, Diyarbakir /Turkey pp. 73–75. (scopus)

## УДК 81-11

HUQUQIY TA'LIM VA UNING O'ZBEKISTON HUDUDIDA  
SHAKLLANISHI.

*Jumayeva Feruza Faxriddinovna*  
*Buxoro davlat pedagogika institute*  
feruzajumayeva@gmail.com

Mazkur maqolada huquqiy ta'lim tushunchasi, dunyo hamjamiyatida huquqiy ta'limning shakllanishi va rivojlanishi, O'zbekiston hududida huquqiy huquqiy savodxonlikni shakllantirishning ilk ko'rinishlari hamda huquqiy ta'lim, huquqiy tarbiya, huquqiy savodxonlik kabi tushunchalarning mazmun-mohiyati to'grisida fikr yuritilgan.

**Kalit so'zlar:** siyosiy-huquqiy qarashlar, huquqiy ta'lim tizimi, huquqiy ong, huquqiy madaniyat, huquqiy savodxonlik, axloqiy tarbiya, ijtimoiy munosabatlar, fuqarolik jamiyati, huquqiy me'yorlar, odat huquqi.

LEGAL EDUCATION AND ITS FORMATION IN THE  
TERRITORY OF UZBEKISTAN

*Jumayeva Feruza Fakhridinovna*  
*Bukhara State Pedagogical Institute*  
feruzajumayeva@gmail.com

In this article, the concept of legal education, the formation and development of legal education in the world community, the first formation of legal legal literacy in the territory of Uzbekistan, and the content and essence of concepts such as legal education, legal upbringing, legal literacy are discussed.

**Key words:** political and legal views, legal education system legal awareness, legal culture, legal literacy, moral education, social relations, civil society, public organizations, legal norms, customary law.

ПРАВОВОЕ ОБРАЗОВАНИЕ И ЕГО ФОРМИРОВАНИЕ  
НА ТЕРРИТОРИИ УЗБЕКИСТАНА

*Ф. Ф. Джумаева*  
*Бухарский государственный педагогический института*  
feruzajumayeva@gmail.com

В данной статье рассмотрены понятие юридического образования, становление и развитие юридического образования в мировом сообществе, первые проявления формирования правовой правовой грамотности на территории Узбекистана, а также содержание и сущность таких понятий, как

юридическое образование, обсуждаются юридическое воспитание, правовая грамотность.

**Ключевые слова:** политические и правовые взгляды, образовательная система преподавания права, правосознание, правовая культура, правовая грамотность, нравственное воспитание, общественные отношения, гражданское общество, правовые нормы, обычное право.

Huquqiy ta'lim – mas'ul davlat organlari tomonidan tasdiqlangan ma'lum bir huquqiy tushuncha, bilim va ko'nikmalarni aniq maqsad asosida tizimli, rejali tarzda ta'lim oluvchilarda shakllantirishga qaratilgan faoliyat jarayonidir. Huquqiy ta'lim ta'lim oluvchilarda na faqat huquqiy savodxonlikni shakllantirish balki huquqiy bilimlarni hayotiy vaziyatlarda qo'llay olishni ham nazarda tutadi. Shu jihatdan yurtimiz olimasi H. To'ychiyeva o'z dissertatsiyasida huquqiy ta'limga quyidagicha ta'rif beradi "Huquqiy ta'lim – hayotiy-amaliy faoliyatda qo'llaniladigan va egallamishi zarur bo'lgan huquqiy ma'lumot va ko'nikmalarni shakllantirishga qaratilgan tizimli, rejali va muayyan maqsadga yo'naltirilgan nazariy va amaliy faoliyat jarayonidir"[1]. Xo'sh, dunyo hamjamiyatida huquqiy ta'lim qachondan boshlangan? Albatta bu savolning javobi ilk davlatchilik tuzumi boshlangan davrlarga borib taqaladi. Sababi shundaki davlat paydo bo'libdiki uni boshqarish uchun ma'lum bir ijtimoiy-huquqiy normalar paydo bo'lgan, garchi bu asosan "odat huquqi" tarzida namoyon bo'lgan bo'lsada, hamda bu normalar avloddan avlodga o'rgatilib, xalq, aholi munosabatlari shu normalar orqali tartibga solingan. Qadimgi Rimda maxsus ritorika maktablarda o'quvchilarga alohida huquqiy ta'lim berilgan, lekin hali bu jarayon tizimli bo'lmagan. Miloddan avvalgi 3-asrda birinchi plebey pontifex maximus-ruhoniy amaldorlar boshlig'i Tiberiy Korunkaniy ruhoniy bo'lmagan o'quvchilarga ommaviy huquqiy yo'riqnomalar berdi va ular jurisprudentes- yuridik maslahatchilar deb ataldi. Jurisprudentlar huquqiy adabiyotlarni o'qish bilan birga ustozlari bilan turli xil huquqiy masalalarni muhokama qilish orqali qonunni o'rganishi mumkin edi. Keyinchalik aynan shu o'quvchilar o'zlarini huquqshunoslar sifatida namoyon etdilar.[2] Shu o'rinda huquqshunos olim Z. Islomovning quyidagi fikrlarini eslatib o'tish joiz "Huquqiy bilim berishni ta'lim muassasalarida maxsus huquqiy fanni o'qitish, shuningdek, maxsus ta'lim uslublarini qo'llash orqali amalga oshirish yaxshi samaralar beradi"[3] Darhaqiqat, yuqoridagi huquqiy ta'lim vujudga kelishi tarixi haqidagi fikrlarimiz ham olimning fikrlarini tasdiqlayapti hamda hozirgi jamiyatimiz uchun ham dolzarbligi bilan ahamiyatlidir.

Qadim sharq ayniqsa Markaziy Osiyo hududi davlatchilik tarixi juda qadim zamonlarga borib taqaladi. Va aynan yurtimiz xalqlari chuqur davlatchilik g'oyaviy-ma'naviy negizlariga asoslangan o'ziga xos siyosiy-huquqiy madaniy boylik egasidir. O'rta Osiyo xalqlari o'z madaniyati bilan jahon sivilizatsiyasiga butundunyo huquqiy madaniyati rivojlanishiga munosib hissa qo'sha olgan. Milodda avvalgi bir necha yuz yillar oldin O'zbekiston hududida Katta Xorazm, Parkana, So'g'diyona kabi shahar davlatlar bo'lib, ulardagi huquqiy munosabatlar tartiboti haqida biz uchun asosiy manba "Avesto" kitobidir, eramizning boshlari va milodiy VII asrgacha bo'lgan davr huquqiy munosabatlar tartiboti to'g'risida unchalik ko'p manba mavjud emas, asosan "So'g'd yozma manbalari" hamda "Turkiy xalqlar yodnomalari" asosida bilishimiz mumkin.[4]

O'zbekiston hududida shakllangan siyosiy-huquqiy qarashlar, davlatchilik va huquq tarixi deyarli to'liq o'rganilmagan va bu borada ilmiy tadqiqotlar olib borish, tarixiy-huquqiy fanlar oldida eng dolzarb muammodir. Chunki mustaqillikka erishgan kunimizdan boshlab shaxdam qadamlar bilan bardavom rivojlanib kelayotgan respublikamiz uchun hozirgi zamon siyosiy-huquqiy institutlari hamda davlatimiz huquqiy tartibot munosabatlari rivojlanishining nazariy asoslarini tushunib yetish mumkin emas. Zero, tarixni bilmay turib, kelajakni qurib bo'lmaydi.

Huquqshunos olim A. Saidov ta'kidlaganidek "Huquqiy demokratik fuqarolik jamiyatida huquqiy savodxonlik axloq darajasidagi mavqega ega bo'ladi. Shunga ko'ra yuksak savodxonlik, ma'naviylik, vatanparvarlik, halollik kabi xislatlarni o'zida mujassamlashtirgan kadrlarni tayyorlashga alohida e'tibor berish zarur.[5] Shuning uchun ham O'zbekistonda mustaqillimizning ilk kunlaridan huquqiy demokratik davlat asoslarini mustahkamlashga doir islohotlar, ayni paytda yurtimizda aholining huquqiy ongi va huquqiy madaniyatini oshirish bo'yicha tizimli chora- tadbirlarni qat'iy olib bordi. Jumladan, 1997- yil 29-avgustda O'zbekiston Respublikasi Oliy Majlisi tomonidan qabul qilingan "Jamiyatda huquqiy madaniyatni yuksaltirish milliy dasturi", 2001 yil 4- yanvarda e'lon qilingan "O'zbekiston Respublikasi Konstitutsiyasini o'rganishni tashkil qilish to'g'risidagi" Prezident farmoyishi kabi normativ hujjatlarda aholi huquqiy madaniyat darajasining oshirilishini ta'minlash uchun avvalo huquq-tartibot tizimida ishlaydigan xodimlarning huquqiy malakasini rivojlantirishga, fuqarolarga bevosita huquqiy yordam ko'rsatish tizimini yaxshilash, ta'lim tizimida huquqiy bilimlarni targ'ib qilish uchun maxsus

huquqiy fanlarni o'qitish, jumladan yuridik adabiyotlat, darsliklar va qo'llanmalar nashr etish, huquqiy ilmiy tadqiqotlar olib borish ishlarini qo'llab quvvatlash va boshqa masalalar o'z aksini topgan edi. Ammo yillar o'tishi, jamiyatimizdagi ijtimoiy, iqtisodiy, siyosiy munosabatlarning o'zgariganligi, davlatimiz fuqarolarning fikr yuritish darajasining o'zgarishi ushbu milliy dasturni yangilash vazifasini hukumat oldiga ko'ndalang qo'ydi. Bundan tashqari 2017 –yil 7- fevraldagi “O'zbekiston Respublikasini yanada rivojlantirish bo'yicha harakatlar strategiyasi to'g'risida”gi Prezident farmoni, 2017-yil 7- sentabrda qabul qilingan “Huquqiy axborotni tarqatish va undan foydalanishni ta'minlash to'g'risida”gi qonun, 2018- yil 13- apreldagi “Davlat huquqiy siyosatini amalga oshirishda adliya organlari va muassasalari faoliyatini tubdan takomillashtirish chora – tadbirlari to'g'risida”gi Prezident farmoni, 2019- yil 9- yanvardag “ Jamiyatda huquqiy ong va huquqiy madaniyatni yuksaltirish tizimini tubdan takomillashtirish to'g'risida”gi Prezident farmoni [6] va boshqa me'yoriy

– huquqiy hujjatlarning qabul qilinganligini ko'rish mumkin.

Xulosa o'rnida shuni aytish mumkinki, yurtimiz hududida huquqiy normalar shakllanishi uzoq tarixga ega bo'lib, ularni naqadar chuqur o'rganib borsak hozirgi davr siyosiy-huquqiy munosabatlarga doir muammolarga javob topgan bo'lar edik. Huquqiy ta'lim-tarbiya tizimini yanada rivojlantirish davr talabi hisoblanadi. Shu o'rnida huquqshunos olimlarning quyidagi fikrlarini ta'kidlash o'rinlidir “Huquqiy ta'lim va huquqiy tarbiya kompaniya tarzida emas, balki ma'lum maqsadni ko'zlab, rejali va samarali tarzda, turli ijtimoiy guruxlarning xususiyatlarini inobatga olgan holda olib borilishi lozim[7]. Darhaqiqat, huquqiy savodxonlikni axloq darajasiga ko'tarish uchun, avvalo maktabgacha va boshlang'ich ta'lim o'quvchilarida huquqiy tushunchalarni shakllantirish bilan birga vatanparvarlik, insonparvarlik, halollik kabi axloqiy xususiyatlarni ham shakllantirib borish hayotiy ehtiyojdir.

#### ADABIYOTLAR:

1. To'ychiyeva Hamida Nabiyevna. Shaxs huquqiy ongini shakllantirishda huquqiy ta'limning o'rni va uning ilmiy-amaliy muammolari. 12.00.01- Davlat va huquq nazariyasi va tarixi. Siyosiy va huquqiy talimotlar tarixi. Yuridik fanlar nomzodi ilmiy darajasini olish uchun yozilgan dissertatsiya-2009-24-b

2. <https://www.britannica.com/topic/legal-education> (Ko'rilgan sana: 19.09.23y)

3. Z.M. Islomov. Davlat va huquqning umumnazariy muammolari: huquqni tushunish, huquqiy ong va huquq ijodkorligi / Mas'ul muharrir akademik Sh.Z. Urazayev – T.: TDYI, 2005. 158 b
4. Davlat va huquq tarixi: IIV oliy ta'lim muassasalari uchun darslik // H. Odilqoriyev, N. Azizov, X. Madirimov. – T.: O'zbekiston Respublikasi IIV Akademiyasi, 2012. – 12–30 b
5. Tadjixanov U., Saidov A. Huquqiy madaniyat nazariyasi. Darslik. 2 tomli. 1-tom/Mas'ul muharrir akademik Sh.Z. Urazayev – T.: O'zbekiston Respublikasi IIV Akademiyasi, 1998, – 76 – b.
6. O'zbekiston Respublikasi Vazirlar Mahkamasining “Jamiyatda huquqiy madaniyatni yuksaltirish bo'yicha tadbirlarni monitoring qilish va baholash to'g'risidagi nizomni tasdiqlash haqida”gi 342-son qarori. 2019 yil 20 aprel, O'zbekiston Respublikasi Qonun hujjatlari ma'lumotlari milliy bazasi, 22.04.2019 y., 09/19/342/2995-son]
7. Tadjixanov U., Saidov A. Huquqiy madaniyat nazariyasi. 2 tomli. 1-tom / Mas'ul muharrir akademik Sh.Z. O'razayev – T.: O'zbekiston Respublikasi IIV Akademiyasi, 1998, – 56-b



УДК 81-11

**“TUYA” ZOONIMI ISHTIROK ETGAN MAQOLLARDA  
FREYMLAR TADQIQI***Jumayeva Marjona Erkinovna**BuxDU*

marjonajumayeva@gmail.com

Mazkur maqolada “tuya” zoonimi ishtirok etgan maqollar kognitiv tilshunoslikning o‘rganish obyekti hisoblangan freym hamda slotlar nuqtayi nazaridan tahlilga tortilgan. Dastlabki freym – tuyaning fiziologiyasi va uning tarkibidagi kichik slotlar turli misollar vositasida ochib berilgan.

**Kalit so‘zlar:** zoonim, obyekt, kognitiv tilshunoslik, freym, slot, fiziologiya, sintaktik gap, subfreym.

**STUDY OF FRAMES IN PROVERBS WITH THE  
PARTICIPATION OF “CAMEL”***Jumayeva Marjona Erkinovna**BuxSU*

marjonajumayeva@gmail.com

In this article, the proverbs with the zoonym “camel” are analyzed from the point of view of frame and slots, which are the object of study of cognitive linguistics. The first frame – the physiology of the camel and its sub-slots – is revealed through various examples.

**Key words:** zoonym, object, cognitive linguistics, frame, slot, physiology, syntactic sentence, subframe.

**ИЗУЧЕНИЕ КАДРОВ В ПОСЛОВИЦАХ С УЧАСТИЕМ  
«ВЕРБЛЮДА»***М. Э. Джумаева**БухГУ*

marjonajumayeva@gmail.com

В данной статье пословицы с зоонимом «верблюд» анализируются с точки зрения рамки и слотов, которые считаются объектом исследования когнитивной лингвистики. Первый кадр – физиология верблюда и его подслоев – раскрывается на различных примерах.

**Ключевые слова:** зооним, объект, когнитивная лингвистика, фрейм, слот, физиология, синтаксическое предложение, подфрейм.

Kognitiv tilshunoslikning o'rganish obyekti inson egallagan bilim va tajriba hisoblanishi bilan qanchalik ahamiyatli bo'lsa, shu egallangan bilimlarni ma'lum bir qolip hamda sxemaga solishinishi- yu, o'ziga xos assotsativ maydonlarni hosil qilishi bilan freym hamda slotlar ham shunchalik ahamiyatlidir.

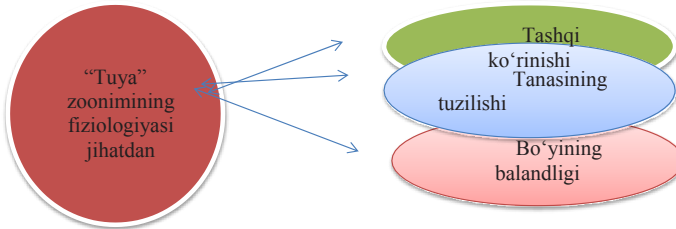
Inson hayoti davomida o'zlashtirgan bilimlarni ma'lum bir qolipga solsagina u voqelanish jarayonida izchillik hamda mukammallik kasb etadi. Xususan, birgina non konseptini tasavvur qilishimiz bilan bir qancha freymlar to'pi, assotsiativ maydon darhol shakllanadi. Nonga bog'liq bo'lgan uning qator turlari, non bilan aloqador jarayonlar, so'zlovchining mazkur mahsulot bilan bog'liq bo'lgan o'tmish hayoti, xalqimiz tarixida urushga ketayotgan askargan non tishlatish-u, beshikdagi chaqaloq boshiga non qo'yishimiz, kelinchakning yangi uyga non bilan kirib kelishi kabi lingvokulturologik xarakterga ega bo'lgan jarayonlar ham yaqqol xotirada jamlanadi.

"Freym" tushunchasini tilshunoslikka birinchilardan bo'lib Ch. Filmor tadbiq etdi (Fillmore 1985; Fillmor 1988). Ammo oldingi tajribaning yangi axborotni 42 qabul qilish va idrok etishdagi rolini maxsus o'rgangan psixolog F.Bartlett ushbu faoliyat «bilishning stereotip namunalari hamda ma'lum ijtimoiy guruh uchun xos bo'lgan normalar (me'yorlar)» zahirida kechishini ta'kidlagan (Qarang: Velichkovskiy 1982: 42). F.Bartlett fikricha, inson har qanday axborotni qabul qilishda oldindan tayyor strukturalardan foydalanadi va ushbu andozaviy tuzilmalar axborotni eslab qolish, xotirada saqlash vazifasini osonlashtiradi. Bundan tashqari, freymga monand bo'lgan sxema nazariyasiga asos solinishi xotirada iz qoldirgan oldingi tajribaning yangi axborotni qanday yo'l bilan faol idrok etilishiga ta'sir ko'rsatishi masalasini oydinlashtirishga imkon beradi [3, 41].

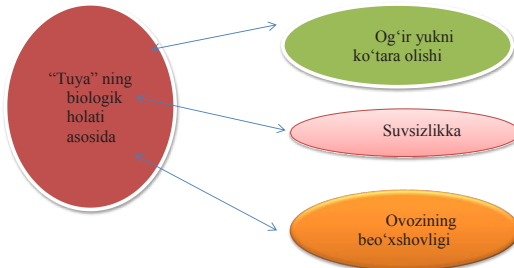
Freym haqida tilshunoslikda turlicha qarashlar mavjud. Bir qator tilshunoslar freymni faqatgina sintaktik gap tarkibidagi yagona predikat sifatida qabul qilsalar, yana bir guruh olimlar freymlar subfreym(kichik freymlar) dan tashkil topgan, voqelik haqidagi umumiy bilimlarni o'zida jamlagan hodisa sifatida qabul qiladilar.

Freymalar tilda bir qancha slotlardan tashkil topadi. Shundan kelib chiqib mazkur jonivor qatnashgan maqollarni freymlar guruhidan tashkil topgan deyish mumkin.

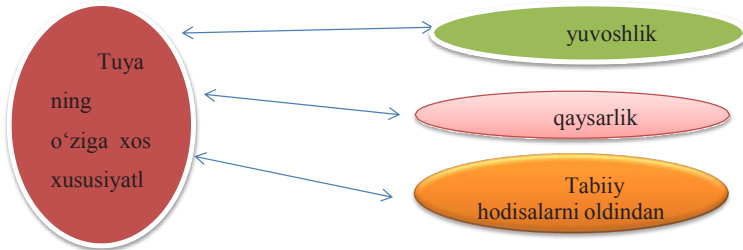
I. “Tuya” zoonimining fiziologiyasi jihatdan slotlardan tashkil topgan freym yukni



II. Tuya” ning bajaradigan biologik vazifasi jihatdan mavjud freym-lar toʻpi



III. Tuyaning oʻziga xos xususiyatlari nuqtai nazaridan tarkib topgan slotlar



**Dastlabki freym-lar toʻpining shakllanishida tuyaning fiziologiyasi asos boʻlib xizmat qilgan. Xususan, koʻpgina maqollarda tuyaning fiziologik holati bilan bogʻliq maqollar uchraydi.**

**Dastlabki slot: Tuyaning tashqi koʻrinishi:**

*Ilon oʻz egriligini bilmay tuyaning boʻynini egri der.*

Tabiatan ilon yerda turli egri harakatlar qiladi, bu esa unga oson va tez harakatlanish imkonini beradi. Tuya esa mavjudotlar ichida

bo‘ynining hamda tanasining ustki qismi egriligi bilan ajralib turadi. Bu esa unga og‘ir yukni bemalol ko‘tarish imkonini beradi. Ustidagi tepachalar esa unga issiqda soya qilib turish imkonini beradi.

***Tuya ustida qatiq uyimas.***

Ijtimoiy hayotda qatiq uyishi, uning sutdan qatiqqa aylanishi uchun tekis hamda issiq joy kerak bo‘ladi. Bu esa unga tezroq hamda sifatli uyish imkonini beradi. Tuya tanasining turli egriliklari sababidan ham xalqda mazkur maqolning yaratilgan.

**Ikkinchi slot: Bo‘yining balandligi.  
*Baxtsiz kishini tuyaning ustida ham it qopadi.***

Azaldan qishloq xo‘jaligiga va chorvachilikka asoslangan o‘zbek xalqida tuya jonivori qadrlangan. Boshqa jonivorlardan farqli ravishda tuya zoonimi ulkan vazni va biologik holati jihatidan ajralib turadi. Uzoq masofa uchun ulov vositasida sifatida qadimda va O‘rta asrlarda faol ishlatilgan. Shu sababdan ham mazkur jonivor yo‘qsil oilalardan kamdan-kam topilgan. O‘ziga to‘q xonadonlarda esa nafaqat ulov vositasi sifatida, balki foydali suti sababidan ham boqilgan. Mazkur maqol “kambag‘alni tuyaning ustida ham it qopar” variantiga ham ega. O‘zbek madaniyatida qadimda ochlik va iqtisodiy tanglik davrida *baxtsizlik* atamasi ostida *kambag‘allik* ham nazarda tutilgan. Ijtimoiy turmushdagi qiyinchiliklar va O‘rta asr tuzumida yuz bergan kamchiliklar jabrini eng ko‘p yo‘qsil xonadonlar tortishgan. Shu sababdan ham baxtsiz kishini baxtsizligi ustiga bir daydi it ham qopishi hech gap emas. Mazkur maqolda it obrazining berilishi bu jonivorning o‘zbek turmush tarziga juda yaqinligi bilan belgilanadi. Qolaversa, it xonaki bo‘lishiga qaramay, hurish va tishlash xususiyatiga ham ega. Yuqoridagi maqolda iqtisodiy jihatdan qiynalgan, ishi yurishmagan, peshonasi sho‘r insonni kichkina it ham o‘zidan bir necha baravar katta tuya ustida qopa olishi haqida aytilgan. Mazkur tushunchaning lingvokulturema sifatida shakllanishida tuya va it zoonimlarining o‘zbek xalqi turmush tarziga yaqinligi va xonaki hayvonlar sifatida istiqomat qilganligi, xalqimiz tafakkurida bir kelsa qo‘sha keladi, degan tushuncha, omadi yurushmagan insonni har qanday kulfatlar ketma-ket qarshi olishi haqidagi qarashning mavjudligi, tuya zoonimi o‘zbek madaniyatiga tanish bo‘lgan hayvonlar ichida kattaligi jihatdan farq qilishi hamda ulkan hajmga ega ekanligi, it zoonimi bilan kattaligi jihatdan ziddlikda turishi asos bo‘lgan.

Variantlari: ”Yomon kishi bolasin tuyada ham it qopar”, “Baxtsiz kishi bolasin tuyada ham it qopar”. Xalqimizning yaxshilik va yomon-

lik, yovuzlik va ezgulik haqidagi qarashlari mazkur maqol misolida yoritilgan. Maqolda bировга yomonlikni ravo ko‘radigan kishi o‘zi tugul, oila a‘zolari ham aziyat chekishi mumkinligi ta‘kidlangan [7, 152].

Bari kalga qaytadan jingala soch bitdi. Mening kalam yaltirab qoldi. Bitta-da soch bitmadi! Do‘xtirlar ajablendi, kamdan kam uchraydigan voqea, dedi. Xo‘rligimdan yig‘lab-yig‘lab uyga keldim. Onamiz tap-taqir kallamni ushlab-ushlab ko‘rdi. Ich-ichidan kuydi. Aytib-aytib yig‘ladi.

– **Kambag‘alni tuyaning ustida it qopadi**, degani shu-da, – dedi.

Katta telpak sotib oldim. Qishin-yozin quloqlarimgacha bostirib kiyib yurdim. Maktabda-da telpagimni olmadim. Matematika muallimimiz oyog‘ini tirab talab etdi:

– Bosh kiyimingni olmasang, dars o‘tmayman!– dedi. Old qatorda o‘tirgan sinf oqsoqolimiz muallimga bir nimalar dedi. Ammo muallim o‘z so‘zida

turdi:

– O‘quvchi darsda yalangbosh o‘tirishi lozim! Qoida shunday! [5, 56].

Bundan tashqari xalqimizning milliy-madaniy fondida tuyaning fiziologiyasi bilan bog‘liq quyidagi maqollar uchraydi.

*Yiqilsang, nordan yiqil.*

*Tuya mingan qo‘y orasiga yashirinolmas.*

*Tuya mingan bo‘yin yashirinolmas.*

*Sog‘ kishi kuyani ham ko‘rar, ko‘r kishi tuyani ham ko‘rmas.*

*Ko‘rmas tuyani ham ko‘rmas.*

**Uchinchi slot: Tanasining kattaligi**

*Ignadek teshikdan tuyadek sovuq kirar.*

*Dushmaning kuya bo‘lsa ham, tuya bil.*

*Tuyadek bo‘ying bo‘lguncha, tugmadek o‘ying bo‘lsin.*

*Tuyani shamol uchirsa, echkini osmonda ko‘r.*

*Tuyaday qiz uzatib, tugmaday jaz yemadim*

Yuqoridagi maqolda birgina tuyaning fiziologiyasiga ko‘ra o‘ziga xosliklar turli maqollar misolida ochib berilgan. Freymalar to‘pi sifatida tuyaning fiziologiyasi, uning slotlari sifatida esa tanasining tashqi ko‘rinishi, bo‘yining balandligi, tanasining kattaligi kabilarni kiritishimiz mumkin.

## ADABIYOTLAR :

1. Imom AL Buxoriy. Al jomi as – sahih. 2-qism. T.: Qomuslar bosh tahririyati, 1997. – 432 b.
2. Маслова В. Лингвокультурология: Учебное пособие. М.: Academia, 2001. 208 с.
3. Сафаров Ш. *Когнитив тилишунослик*. Рисола. – Жиззах: “Санзор” нашриёти, 2006. – Б. 41.
4. Schmidt-Nielsen B., Schmidt-Nielsen K. The camel, facts and fables // UNESCO Courier. – 1955. – № 8–9. – С. 29.
5. То‘g‘ay Murod “Ot kishnagan oqshom” qissa, T.: “Sharq” 2008-y, 1-b
6. O‘zbek xalq maqollari. Toshkent: Gafur Gulom nomidagi nashriyot-matbaa ijodiy uyi, 2009. – 127 b.
7. Xolmonova Z, Tosheva D. Zoonim komponentli paremalarning izohli lug‘ati. T.: “Turon zamin ziyo” – 2016, – 359b.
8. Шомаксудов Ш., Шорахмедов Ш. Хикматнома. Т.: Ўзбек совет энциклопедияси, 1990. 524 б.

УДК 81-11

ZAMONAVIY LUG‘ATCHILIK TARAQQIYIT  
BOSQICHLARI*Jo‘raqulova Mushtariy Abdumalik qizi**O‘zMU*

mushtariy00712@gmail.com

*Anorboyeva Dilshoda Ibrohim qizi**O‘zMU*

mushtariy00712@gmail.com

**Kalit so‘zlar:** Elektron lug‘at, tarjima lug‘at, lingvistik ta‘limotlar bazasi, ochiq manbali lingvistik resurslar.

**Key words:** Electronic dictionary, translation dictionary, database of linguistic doctrines, open source linguistic resources.

**Ключевые слова:** Электронный словарь, переводной словарь, база данных лингвистических учений, лингвистические ресурсы с открытым исходным кодом.

**Annottatsiya.** Bugungi kunda lug‘atlarga ehtiyoj ortgan sari lug‘atlar ustida ishlash va oson topish maqsadida turli metodlardan foydalanilmoqda. Bugungi davr axborot- texnologiya davri bo‘lganligi sababli hamma kompyuter va smartfonlardan foydalanadi. Shu sababli lug‘atlarni kitoblardan varaqlab izlagandan ko‘ra elektron izlaganni afzal ko‘rishdi. Shunga qarab tilshunos olimlar ham lug‘atlarni elektron variantini ishlab chiqishmoqdalar. Izlab ko‘rilayotgan so‘zlar esa tez va oson, vaqtni ham tejashga asqotmoqda.

**Abstract.** Today, as the need for dictionaries increases, various methods are used to work on dictionaries and find them easily. Today’s era is the era of information technology, so everyone uses computers and smartphones. For this reason, they preferred to look up dictionaries electronically rather than flipping through books. Accordingly, linguists are developing electronic versions of dictionaries. Searching for words is quick and easy, saving time.

**Абстрактный.** Сегодня, когда потребность в словарях возрастает, для работы со словарями и облегчения их поиска используются различные методы. Сегодняшняя эпоха – это эпоха информационных технологий, поэтому все используют компьютеры и смартфоны. По этой причине они предпочитали искать словари в электронном виде, а не листать книги. Соответственно, лингвисты разрабатывают электронные версии словарей. Поиск слов осуществляется быстро и легко, экономя время.

Zamonaviy ilm-fanning barcha sohalariga zamonaviy axborot-kommunikatsiya texnologiyalarining tatbiq etilishi, xususan, tilshunoslik sohasiga kirib kelishi tilning turli aspektlarini o‘rganishda kompyuter

lingvistikasi nazariyasi asosida shakllangan korpuslardan foydalanish boshlandi. Jumladan, elektron lugʻatlar fikrimizning isboti. H.Nesi "Dictionaries in electronic form" maqolasida korpusga asoslangan elektron lugʻatlarda soʻzlarning yozilishi, maʼnosi yoki qoʻllanilishi har qanday lingvistik taʼminotga nisbatan qoʻllanilishi mumkin [1], deb taʼkidlaydi. Kompyuter leksikografiyasi – koʻp asrlar davomida mavjud boʻlgan anʼanaviy qoʻlda yozilgan leksikografik amaliyotdan yangi qogʻozsiz axborot texnologiyalariga oʻtish, bu lugʻatlarni yaratish uchun matnli maʼlumotlarni qayta ishlash usullari va dasturlari toʻplami bilan ifodalanadi.

Kompyuter leksikografiyasining markaziy obyekti bu kompyuter lugʻati boʻlib, u avtomatik qayta ishlash va toʻldirish dasturlari bilan jihozlangan mashina tashuvchilaridagi har qanday leksikografik ish sifatida tushuniladi.

S.Zagainov korpus asosidagi elektron lugʻatga kirish orqali leksik birliklarning parallel korpuslardan tarjima qilingan ekvivalentini izlash imkonini beruvchi maxsus dastur yordamida tashkil etilgan leksikografik maʼlumotlar tizimi sifatida taʼrif berib, elektron lugʻatni yaratishda korpuslarga oʻziga xos dasturiy taʼminot sifatida qaraydi .

N.Abduraxmonovanning "Kompyuter lingvistikasi" darsligida lingvistik taʼminot tushunchasiga quyidagicha taʼrif beriladi: "Lingvistik taʼlimotlar bazasi tilga oid barcha maʼlumotlar tushuniladi. Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga asosan lingvistik taʼminotida quyidagilarning mavjud boʻlishi talab qilinadi: matn- leksik asos, leksema shakl-soʻz birikmasi, sintaktik gap strukturalari. Tarjimadan avvalgi tahlil jarayonida uchta asosiy ierarxaik bosqichda amalga oshiriladi: matn-jumla-leksema shakl". Elektron lugʻat qogʻoz lugʻatning qulay qidiruv tizimi bilan jihozlangan elektron shakli, degan fikr keng tarqalgan. Biroq koʻplab tadqiqotchilar tomonidan turlicha qarashlar ham mavjud boʻlib, Y.Pervanov fikricha, elektron va qogʻoz lugʻat oʻrtasidagi tub farqni elektron belgilash va gipermatnda emas, balki foydalanuvchini virtual leksikografiyaga butunlay boshqacha tarzda jalb qiladigan til semantikasining leksikografik talqinining yangi usuli sifatida eʼtirof etadi .

"Korpusga asoslangan elektron lugʻatlarda soʻzlarning yozilishi, maʼnosi yoki qoʻllanilishi har qanday lingvistik taʼminotga nisbatan qoʻllanilishi mumkin [2], deydi H.Nesi "Dictionaries in electronic form" maqolasida. Ushbu fikrlardan kelib chiqib shuni ayta olamizki, elektron lugʻatlarni yaratishda korpuslar dasturiy taʼminot vazifasini



o'taydi. Elektron lug'at qog'oz lug'atning qulay qidiruv tizimi bilan jihozlangan elektron shakli, degan fikr ham yo'q emas.

Lekin N.Forget "Korpus texnologiyasi asosidagi elektron lug'atlar qog'oz lug'atlardan qo'llanilishi, taqdimoti, parametrlari, texnik jihatlari bilan farqlanishini aytadi. Muallif elektron lug'atlarda gipermatnlardan foydalanishga va qog'oz lug'atlarga nisbatan so'zlarning ma'nosini yoki imlosini topish uchun qidiruv tizimiga asoslanganligiga e'tibor qaratadi [3]. Ya'ni elektron lug'atlarning afzal tarafi ham so'zlarni tez va oson yo'l bilan topa olishimizdir. Bunday elektron lug'atlarni yaratishda mashina tarjimasining alohida o'rni bor. Mashina tarjimasi-maxsus kompyuter dasturi yordamida matnlarni (yozma va ideal og'zaki) bir tabiiy tildan ikkinchi tabiiy tilga tarjima qilish jarayonidir.

"Mashina" o'rniga ba'zan avtomatik so'z ishlatiladi, bu ma'noga ta'sir qilmaydi. Biroq, mashina tarjimasini Avtomatlashtirilgan tarjima bilan aralashtirib yubormaslik kerak, u butunlay boshqacha ma'noga ega – bu bilan dastur oddiygina odamga matnlarni tarjima qilishga yordam beradi.

Tarjima uchun elektron kompyuterlardan (kompyuterlardan) foydalanish g'oyasi 1947- yilda AQShda birinchi kompyuterlar paydo bo'lgandan so'ng darhol ifodalangan. Mashina tarjimasining birinchi ommaviy namoyishi 1954- yilda bo'lib o'tdi. Ushbu tizimning primitivligiga qaramay, bu tajriba keng rezonansga ega bo'ldi. Bugungu kunga kelib esa zamonaviy tilshunoslikning leksikografiyada lug'atlarni yaratish va ulardan foydalanish uchun kompyuter tizimlarini sezilarli darajada kengaytirdi va mustahkamladi. Lug'atga kirishni tashkil qilish usullari, lug'atlarning tuzilishi va ularni yaratish texnologiyasi uni qiziqtiradigan markazdir. Leksikografiyada kompyuterdan foydalanish to'g'risida gap ketganda, uning uchta bosqichni ajratib ko'rsatdi:

1. Kompyuter yordamida (qog'ozli) leksikografiya,
2. Mavjud qog'oz lug'atlarni elektron vositaga o'tkazish,
3. Elektron muhit uchun yangi yaratilgan elektron lug'atlar.

Ushbu bosqichlarning har biri davomida lug'at tuzuvchilar va metalexikograflar o'zlarining "kelajak lug'atining" versiyasini tuzdilar va kengaytirib bordilar.

CD-ROM va qo'lda ishlaydigan qurilmalardan tashqari, Internet, smartfonlar va iPad-larning tez sur'atlarda rivojlanishi va ma'lumot olishning boshqa usullari tufayli tadqiqotchilar ham onlayn lug'atlarga qiziqishni bildira boshladilar.

Elektron lug'atlarning paydo bo'lishiga ob'ektiv holatlar sabab

bo'ladi, ularning yordami bilan lug'at hajmi muammosini hal qilish, lug'atda kerakli ma'lumotlarni tezda topish va tarkibni hozirgi vaqtga moslashtirish mumkin. V.P.Berkovning fikriga ko'ra, XXI asr lug'ati quyidagi xususiyatlarga ega bo'lishi kerak:

1. O'zining hajmida ulkan.
2. Foydalanuvchining ehtiyojlariga qarab ma'lumotlarning bosqich-ma-bosqich chiqarilishini ta'minlash.
3. Lug'atning sog'lom bo'lishi qobiliyati.
4. Kirish va chiqish tillari so'zlari haqida to'liq grammatik ma'lumotlarni, xususiyl nomlar bo'yicha keng materiallarni transkripsiya dasturlari bilan ta'minlash qobiliyati.
5. Grafik tasviriy materialni kiritish imkoniyati.

N. Abduraxmonova muallifligidagi "Mobil o'quv lug'atchiligini rivojlantirishda lingvistik resurslar va raqamli texnologiyalarning o'rni" mavzusidagi maqolada quyidagi fikrlar keltirilgan: "Ta'limiy korpus uchun lingvistik bazani yaratishning umumiy qoidalarini quyidagicha shakllantirishni tavsiya etamiz:

- *birinchidan*, ma'lumotlar bazasining betarafligini ta'minlash;
- *ikkinchidan*, ma'lumot natijalari oynasini standartlashtirish;
- *uchinchidan*, lug'at ma'lumot tavsifining mukammalligiga erishish;
- *to'rtinchidan*, ma'lumot hajmini cheklamaslik;
- *beshinchidan*, ma'lumotlarni imkon qadar giperhavola usuli bilan bog'lash.[4]

#### FOYDALANILGAN ADABIYOTLAR:

1. Nesi H. Dictionaries in electronic form // The Oxford history of English lexicography. Vol. II: Specialized dictionaries. Oxford: Oxford University Press, 2009. P. 458–478.
2. Загайнов С.С. Проект универсального электронного военного словаря [Электронный ресурс] // Электронная библиотека (репозиторий) Томского государственного университета. 2020. URL: <https://bit.ly/3d3fNd0> (дата обращения: 12.03.2020)
3. Forget N. Les dictionnaires électroniques dans l'optique de la traduction. Ottawa: 2000
4. Abduraxmonova N., Latipova G., Abduvahobov G'.. Mobil o'quv lug'atchiligini rivojlantirishda lingvistik resurslar va raqamli texnologiyalarning o'rni. Konferensiya materiallari 2023.

УДК 81-11

**QORAQALPOQ QAHRAMONLIK DOSTONLARI TILINING  
O'RGANILISHI**

*Kdirbaeva Ulzada Kayrbaevna*  
*Qoraqalpoq davlat universiteti*  
*ulzada@gmail.com*

Maqolada qoraqalpoq qahramonlik dostonlari tilining o'rganilishi masalalari so'z etilgan. «Qırıq qız», «Alpamis» va «Edige» dostonlari tilini o'rgangan filolog olimlar va ularning ilmiy ishlari haqida ma'lumot berilgan.

**Kalit so'zlar:** qahramonlik dostonlari, xalq o'gzaki ijodi, «Qırıq qız» dostoni, «Alpamis» dostoni, «Edige» dostoni, variant, doston tili

**STUDY OF THE LANGUAGE OF KARALPAK HEROIC EPICS**

*Kdirbaeva Ulzada Kayrbaevna*  
*Karakalpak*  
*State University*  
*ulzada@gmail.com*

The article discusses the issues of learning the language of Karakalpak heroic epics. Information is provided about the philologists who studied the language of the epics "Forty girls", "Alpamis" and "Edige" and their scientific works.

**Keywords:** heroic epics, folklore, the epic "Forty Girls", "Alpamis" epic, "Edige" epic, option, the language of the epic

**ИЗУЧЕНИЕ ЯЗЫКА КАРАКАЛПАКСКОГО  
ГЕРОИЧЕСКОГО ЭПОСА**

*У. К. Кдырбаева*  
*Каракалпакского государственного университета*  
*ulzada@gmail.com*

В статье рассматриваются вопросы изучения языка каракалпакского героического эпоса. Приводятся сведения о филологах, изучавших язык эпосов «Сорок девушек», «Алпамыс» и «Едиге», и их научных труда.

**Ключевые слова:** героический эпос, фольклор, вариант, эпопея «Сорок девушек», эпос «Алпамыс», эпос «Едиге», эпический язык

Xalq o'zaki ijodi xalqning tarixi, madaniyati, urf-odat dasturlari va tili haqida ma'lumot beruvchi asosiy manba hisoblanadi. Jahon tilshunosligida epos tili va uning tarkibidagi leksik birliklarni o'rganish, leksikografik asoslarni yaratish, leksik qatlamlarning tarkibini aniqlash va badiiylikni isbotlash lingvistika sohasining rivojlanishida

katta ahamiyat kasb etadi. Songi yillarda ota- bobalarimiz tomonidan birnecha asrlar mobaynida yaratilgan bebaho ruhiy va madaniy me'roslarni tiklash mamlakat siyosati darajasiga ko'tarilgani ham bejiz emas. Xususan, xalq tarixini so'zlovchi dostonlarning jamiyat va millat tarixidagi o'rnini belgilash, ko'p jildli me'roslarni tayyorlash va nashr etish, ularning badiiy-estetik qiymatini tanitish dolzarb masalalarning biri sifatida qaraldi.

O'zbekiston respublikasi Prezidenti Sh.M.Mirziyoyev ushbu masala bo'yicha quyidagicha fikr bildirgan edi: «Tengsiz ruhiy boyligimiz bo'lgan xalq og'zaki ijodining bebaho namunalarini asrab- avaylash va rivojlantirish, uni kelajak avlodga toliq yetkazish butun dunyo mutafakkir olimlari va san'atkorlarinig, mamlakat va jamiyat arboblarning, barcha madaniyat vakillarining yuksak vazifasi hisoblanadi. Folklor san'ati bu – insoniyatning bolalik qo'shig'i. Agar biz bu bebaho san'atni saqlab qolish uchun barchamiz birgalikda harakat qilmasak, ertaga kech bo'ladi, kelajak avlodlar, tarix bizni kechirmaydi» [1.1].

Qoraqalpoq til bilimida xalq og'zaki ijodining tili bo'yicha dastlabki ilmiy ish O.Bekbawlovning 1979 yili nashr etilgan «Qahramonlik eposning arab-forsi leksikasi va uning tarixiy-lingvistik xususiyati» nomli asari bo'lib, unda epos tilida qo'llanilgan arab-forsi leksikasi klassik shoyirlar Ajiniyaz, Ku'nxoja va Berdaq asarlarida qo'llanilgan misollar bilan taqqoslab o'rganilgan. Olim bu asarida epos va uning tili haqida quyidagicha fikr bildirgan: «Epostagi boshqa da bir qator xodisalar kabi bu ham o'zidan-o'zi paydo bo'lgan emas, holbuki ular bizni o'tkan davrning ancha tubiga yo'naltiradigan, o'z tarixiga ega ko'pgina sirli faktlar hisoblanadi» [2. 19].

«Qırıq qız» dostoni tilining o'rganilishi. Qoraqalpoq til bilimida folklor asarlari tilining o'rganilishiga bag'ishlangan birqancha ilmiy ishlar mavjud. Masalan, 1992 yili Sh.Abdinazimov ««Qırıq qız» dostonining leksikasi» mavzusida kandidatlik dissertatsiyasini himoya qildi. Ushbu ilmiy ish qoraqalpoq til bilimidagi folklor asarlari tilining alohida o'rganilgan dastlabki ilmiy ishi hisoblanadi. Asarda xalq og'zaki ijodi tilining o'rganilish holati, «Qırıq qız» dostonining leksik tarkibini o'rganishga bag'ishlangan bir qator maqolalari ham e'lon qilindi: ««Qırıq qız» dostonidagi eski so'zlar qatlami», «Qahramonlik eposdagi xitoy tili elementlari», ««Qırıq qız» dostonidagi toponimlar», ««Qırıq qız» dostonining leksik qatlami haqida», ««Qırıq qız» dostonidagi arabizmlar», ««Qırıq qız» dostonidagi eskirgan so'zlar» va boshqalar.

1997 yili 10–12 sentabr kunlari Nukusda Qoraqalpoq xalq dostoni «Qırıq qız» va turkiy zabon xalqlarning folklorini tadqiq qilish masalalari»ga bag'ishlangan xalqaro konferensiya o'tkazildi. Bul

konferensiya materiallarida «Qırıq qız» dostonining tilini oʻrganishga bagʻishlangan M.Qalenderovning ««Qırıq qız» qoraqalpoq ogʻzaki xalq poeziyasining durdonasi», Sh.Abdinazimovning ««Qırıq qız» dostoni leksikasining ayrim masalalari», O.Dospanovning ««Qırıq qız» dostonining lingvistik tavsifi», A.Karibaevaning ««Qırıq qız» eposidagi milliy oyin terminlarining etimologiyasi», A.Saburovaning ««Qırıq qız» dostonida qoraqalpoq milliy musiqa terminlarining qoʻllanilish masalalari» nomli maqola va tezislar nashr etildi [3. 56-58]. Yuqoridagi qayd etilgan ilmiy ishlarda «Qırıq qız» dostonining faqat leksik va stilistik taraflari oʻrganilgan boʻlib, doston tilini lingvopoetik va lingvokulturogik aspektlarda oʻrganish lingvist olimlar oldida turgan dolzarb muammolardan hisoblanadi.

«Alpamis» dostoni tilining oʻrganilishi. Qoraqalpoq xalqida «Alpamis» dostonining 8 varianti bor boʻlib, boshqa turkiy xalqlarda bu dostonning variantlari bunchalik koʻp uchramaydi. 1956 yili sentyabr oyida Toshkent shahrida «Alpamis» dostoni boʻyicha butun ittifoq boʻylab ilmiy-nazariy konferensiya oʻtkazildi. Ushbu konferensiyada «Alpamis» dostonining paydo boʻlish manbasi, genezisi, dostonning gʻoyasi va lingvistik xususiyati haqida ilmiy fikrlar bildirilgan. Konferensiya materiallari 1959 yili eʼlon qilindi. 1999 yil noyabr oyida Surxondaryoda «Alpamis» dostoni paydo boʻlishining 1000 yilligi belgilandi. Bular mamlakatimizda folklor asarlarini oʻrganishga katta eʼtibor qaratilganligidan darak beradi.

«Alpamis» dostoni qoraqalpoq tilida A.Abdievning ««Alpamis» dostoni tilining leksik-semantik va stilistik xususiyati» nomli dissertatsiya ishida dastlabki marotaba har taraflama oʻrganildi. Folklor asarlarining tili dastlabki marta funksional-stilistik yoʻnalishda oʻrganildi. Dissertatsiya ishining asosiy obyekti etib «Alpamis» dostonining Oʻgiz jirov varianti foydalanilgan. Mazkur ishda dostonning leksik tarkibi, doston tilining leksik-semantik va funksional-stilistik xususiyati oʻrganilgan.

2008 yil 27–28 noyabr kunlari «Markaziy Osiyo xalqlarining milliy folklori va jamiyatning madaniy rivojlanish muammolari» mavzusidagi xalqaro ilmiy-amaliy konferensiyalarning oʻtkazilishi mazkur sohadagi qoʻlga kiritilgan yutuqlar hisoblanadi. Ushbu konferensiya materiallarida lingvist olimlardan Sh.Abdinazimov, A.Abdiev, M.Ayimbetov, Z.Alniyazov, D.Oʻrinboeva va I.Xalmuratovlar oʻzlarining ilmiy maqola va tezislari bilan qatnashdi [4.153–155].

Bundan tashqari yosh olim O.Embergenovning qoraqalpoq qahramonlik dostonlaridagi sinonimlarni oʻrganishga bagʻishlangan «Qoraqalpoq qahramonlik dostonlaridagi leksik sinonimlarning stil-

istik xizmati», «Qahramonlik dostonlarda frazeologik soʻz birikmalari orasidagi sinonimlik bogʻliqlik» kabi va boshqa maqolalari eʼlon qilindi.

«Edige» dostoni tilining oʻrganilishi. Qoraqalpoq qahramonlik dostonlarining til tarafidan eng badiiyi «Edige» dostoni hisoblanadi. Ayrim sabablarga koʻra «Edige» dostonini oʻrganish masalasiga yetarli darajada eʼtibor qaratilmadi. Lekin, qoraqalpoq folkloristlari dostonni yigʻish va oʻrganish masalasi bilan shugʻullanishni davom etdi. Masalan, I.Sagʻitovning 1963 yili nashr etilgan

«Qoraqalpoq xalqining qahramonlik eposi» kitobida «Edige» dostonining ahamiyati, uning tarixiy munosibati va folklor asari sifatida tarixdan farqi haqida noyob fikrlar bildirgan. Q.Ayimbetov 1968 yili nashr etilgan «Xalq danaligi» kitobida Erpolat jirov va u ijro etgan «Edige» dostoni haqida toʻqtalib, unga yuqori baho bergan.

Mazkur ilmiy ishlarning paydo boʻlishiga taniqli olim V.M.Jir-munskiyning 1958 yili Moskva nashr etilgan «Некоторые итоги изучения героического эпоса народов Средний Азии», 1962 yili Moskva-Leningradta nashr etilgan «Народный героический эпос» kitoblari asosiy manba hisoblanadi.

«Edige» dostoni qoraqalpoqlarda, qozoqlarda, bashqurtlarda, nogʻoylarda, tatarlarda va boshqa da xalqlarda bor. Harbir xalqda «Edige» dostoni oʻziga xos milliy xususiyatga ega boʻlib, u qoraqalpoq, qozoq va nogʻoy xalqlarining ortiq mulki hisoblanadi.

Qoraqalpoqning «Edige»sini 1903 yili Moskva universitetining talabasi Belyayev Chimboyli Bekimbet jirovdan eshitib yozib olgan. Lekin, mazkur nusxa yoʻqolib ketgan. 1929 yili Qalli Ayimbetov Chimboy shahridagi Erpolat jirovdan yozib olgan. Lekin, mazkur nusxasi ham yoʻqolib ketgan. 1934 yili Qalli Ayimbetov ikkinchi marotaba oʻsha Erpolat jirovdan yozib olib, Moskvadagi badiiy adabiyot bosmasiga bergan va 1937 yili nashr etilgan.

Oʻtgan asrning 60–70 yillari «Edige» dostonining Oʻgiz jirov, Qiyas jirov, Jannazar jirov, Esemurat jirov, Jumabay jirov variantlari yozib olindi. Mazkur qoʻl yozmalar Oʻzbekiston Ilmlar akademiyasining Qoraqalpogʻiston qoʻl yozmalar fondida saqlanadi. Har bir jirov «Edige» dostonin oʻzicha tafavut bilan ijro etadi. Mamlakatimizda songi yillarda folklor asarlarining lingvistik xususiyatlarini oʻrganishga alohida eʼtibor qaratilmoqda. Qahramonlik dostonlarni xalq orasida targʻibot qilishga bagʻishlangan konferensiyalarning oʻtkazilishi bunga yaqqol misol boʻla oladi. 2001 yili 5-6 sentyabr kunlari Nukus shahrida Qoraqalpoq xalq dostoni «Edige»ni oʻrganish masalalari»-

ga bag'ishlangan xalqaro ilmiy-nazariy konferensiya o'tkazildi. Mazkur ilmiy-nazariy konferensiyada «Edige» dostoni tilining u yoki bu masalalarini o'rganishga bag'ishlangan M.Qalenderovning ««Edige» dostonining lingvistik xususiyati», X.O'teniyazovaning ««Edige» dostonida hayotiy leksika», Z.Tlewovaning ««Edige» dostonida ko'makchilarning qo'llanilishi» nomli maqolalari e'lon qilindi [5. 46–48].

Qoraqalpoq til bilimida mazkur dostonning tillik xususiyatlari xaligacha alohida tadqiqot obyektiga aylanmadi. Faqat ayrim maqola va konferensiya materiallarida nashr etilgan birqancha tezislar uchraydi. Xususan, U.Qdirbaevaning bir qator maqola va tezislari: ««Едиге» дастанындағы метафоралардың қолдану ерекшеліктері», ««Едиге» дәстаны тилиндегі фразеологизмлердің структура-семантикалық анализі», ««Едиге» дәстанында нақыл-мақалалардың қолланылуы өзгешелігі» kabilar «Edige» dostonining lingvistik xususiyatlarini o'rganishga qaratilgan. Dostonning leksik-semantik va stilistik xususiyatlarini o'rganish, dostonni lingvopoetik va lingvokulturologik tahlil qilish qoraqalpoq til bilimidagi ahamiyatli vazifalardan hisoblanadi. Sababi, xalq og'zaki muloqat tilini, xalq dostonlarining leksik tarkibini har tomonlama va teran o'rganish qoraqalpoq adabiy tilining manbalarini yana ham o'rganishga, uning shakllanish va rivojlanish yo'llarini kuzatishga, xar xil lingvistik hodisalarning qonuniyatlarini bilishga imkoniyat yaratadi.

#### ADABIYOTLAR:

1. Мирзиёев Ш.М. Миллий ўзлик, тил ва ҳаёт тарзи, анъана ва урф-одатларимиз ифодаси // Халқ сўзи. 2019 йил 8 апрель, № 69 (7299).
2. Бекбаўлов О. Қаҳарманлық эпостың араб-парсы лексикасы ҳәм оның тарийхий-лингвистикалық характеристикасы. Нөкис. «Қарақалпақстан», 1979.
3. Халық аралық илимий теориялық конференция тезислери, – Нөкис, 1997. – С. 56–58.
4. Yuldasheva D.N., Chullieva G. T. THE ROLE OF NOVERBAL MEANS IN SPEECH MEANING. International Engineering Journal For Research & Development, Vol.6, Issue 2 published May 16, 2021. – Pag.1956–1961. [www.iejrd.com](http://www.iejrd.com) <http://www.iejrd.com/index.php/%20/article/view/1956>
5. Материалы международной научно-практической конференции. – Нөкис, 2008. – С. 153–155.
6. Қарақалпақ халық дәстаны «Едиге»ни изертлеу мәселелери. Халықаралық илимий-теориялық конференцияның тезислери. – Нөкис, 2001. – С. 46.

УДК 81-11

## ETHNOOYKONYMS OF BUKHARA

*Kilichev Bayramali Ergashovich**BukhSU*

b.e.kilichev@buxdu.uz

Mazkur maqolada Buxoro viloyati hududidagi etnooykonimlar tahlil etilgan. Ularning semantik-etimologik, hamda areal xususiyatlari o'rganilgan. Etnooykonimlarga asos bo'lgan etnonimlar, ularning asoslari va yuzaga kelish sabablariga doir qarashlar bayon etilgan. Asosiy e'tibor Buxoro viloyat etnooykonimlarining asosini tashkil etgan turkiy etnonimlarga qaratilgan.

**Kalit so'zlar:** nom, atoqli ot, joy nomi, geografik obyekt, etnonim, etnotoponim, etnooykonim, etimologiya, urug', qavm, qabila, qishloq, aholi punkti

This article analyzes ethnooykonyms on the territory of the Bukhara region. Their semantic-etymological and areal features are studied. The ethnonyms underlying ethnooykonyms, views on their foundations, and the reasons for their occurrence are described. The main attention is paid to the Turkic ethnonyms, which form the basis of the ethnooykonyms of the Bukhara region.

**Keywords:** name, proper noun, toponym, geographical object, ethnonym, etnotoponym, ethnooykonym, etymology, tribe, village, locality.

В данной статье анализируются этноойконимы на территории Бухарской области. Изучены их семантико-этимологические и ареальные особенности. Описаны этнонимы лежащие в основе этноойконимов, взгляды на их основы и причины их возникновения. Основное внимание уделяется тюркским этнонимам, составляющим основу этноойконимов Бухарской области.

**Ключевые слова:** имя, имя собственное, топоним, географический объект, этноним, этнотопоним, этноойконим, этимология, племя, село, деревня, населенный пункт.

Geographical objects are called oikonyms – inhabited places, and settlements [3,58]. On the other hand, oikonyms are divided into several groups, and a large part of them are ethnotoponyms. Ethnotoponyms are based on place names formed by ethnonyms (names of people, nations, tribes, clans).

Ethnonym is derived from the Greek words ethnos – “people” and “onim” – “name”, their combination is called ethnonym, and the science that studies ethnonyms is called ethnonomics [16,75]. Vocabulary units representing ethnos belong to the oldest layers of any language. Most of them are names of places, in particular, as toponyms, hydronyms, necronyms, godonyms, and have been preserved to this day as a structural element of toponyms.



Any ethnonym has its meaning. It is difficult to understand the meaning of some ethnonyms because they are very ancient words. Usually, the names of peoples, nations, as well as clans' names are ancient. For example, several opinions have been expressed about national names such as Uzbek, Kyrgyz, and Kazakh, and clan names such as Kungyrot, Mitan, Kangli, Kipchak, Uyshun, Karluq, Khalaj, but the etymologies of these ethnonyms are based on science and not conclusively resolved.

Names of places – ethnotoponyms – based on the names of the race, tribe, and people have their characteristics and their research, semantic-etymological analysis is of great importance in the deep study of the history of our nation. Researcher N. Adizova noted that there are two aspects of the transformation of ethnonyms into place names: firstly, the ethnonym directly expresses the meaning of the place name, and secondly, some names are other words of the ethnic element (ethnonym) formed by joining with. [1, 80].

In the Lower Zarafshan oasis, in particular, in the territory of the Bukhara region, there are ancient settlements of people – cities, and fortresses, where different peoples, tribes, and nations lived. Many place names in the territory of Bukhara are named after Turkic tribes. Below we will analyze some Turkic ethnonyms found in the region semantically and etymologically.

Olot. This ethnonym is found in the forms of *Olot*, *Orlot*, *Arlot*, and *Olotkhana* in the Bukhara, Kogon, and Olot districts of the region. Olot is a phonetically modified form of the ethnonym Arlot. Olot was one of the clans of the Uzbek people, and in fact, it was a common tribal name for the Turkic and Mongolian peoples. Arlot was one of the four tribes that Genghis Khan presented to his son Chigatay. This tribe settled in the northern part of Afghanistan. In Movarounnahr, the Arlots, although few, were considered influential tribes.

Arlots were one of the most widespread Uzbek people. This tribe was widespread in Bukhara, Samarkand, Surkhandarya, Khorezm, and Kashkadarya regions, as well as in the north of Afghanistan. Abulghozi Bahadir Khan mentioned the Orlot clan in the section called “Mongol descendants” in his book “Shajarai Turk” and said that the original meaning of the word Arlot means “beloved son of his mother”. [7, 176–177].

The Arlot tribe was mentioned among the 92 Uzbek “peoples” in “Jome at-Tawarikh”, “Ravzat us-Safo”, “Shayboniynoma”, “Abdullanoma”, “Shajarayi Turk” memoirs, also Alisher Navoi’s works. [16,

164]. The ethnonym of Arlot was recorded in the form of *olot* in the work of Kh. Doniyorov [5, 100].

The name of Alot district of Bukhara region, the name of Alot settlement in Peshku district is related to arlot. There are also toponyms of Old Olot and Upper Olot in Olot district.

Bahrin. The name of a place found in the Vobkent and Romitan districts of the region. This clan had branches such as Karakalpak, Bahrin, Uyali Bahrin, Naiman Bahrin, and Baganagli Bahrin. Bahrin was one of the Turkic Mongolian tribes. During the reign of the Mongols, the Baharins came to Dashti Kipchak and settled. From the 18th to 19th centuries, a large part of the Bahrins settled in the Tashkent oasis, Fergana Valley. The Bahrin tribe is a descendant of the Mongols [4, 49].

Bahrin is the name of a bird, a type of falcon, which was trained and used as a hunting bird in the past. [12, 669]. So, Bahrin was a totem and was considered a sacred creature of the tribe and was worshipped. Totem became the name of this tribe. Based on the ethnicity of the population, the village was called Bahrin.

Kalon. The settlement in the Bukhara district is called Kalon. Among the people, there are opinions that kalon means big, huge (Persian), and it was named because the majority of the population was made up of large-bodied people. This toponym was based on an ethnonym. Because the names of clans belonging to the Khoshtamgali and oyinni sections of kungirats from the Turkic peoples were called *kal*. In the speech of speakers of the Persian-Tajik language in Bukhara, *kal* people (people belonging to the clan *kal*) were called kalon (-on – plural form). From this, the village of Kallar (Kallar) was created.

Mangit. A settlement in the Bukhara, Jondor, and Gijduvan districts of the province is called *Mangit*. The Mangit ethnooikonym arose based on the toponymization of the ethnonym. Mangit is the oldest and largest tribe of the Uzbek people, and its names include divisions like Oqmangit, Mangit, Cholmangit, Boygundamangit, Tokmangit, Chalamangit; and branches of tribes such as Temirkhoja, Esaboy, Esat, Guvalak, Kosa, Toz, Karabayir, Baqirchi, Chovkay, Kulatamgali, Mangitkozaq, Galobotir, Beshkal, Chobakchak, Uz, Uvolai, etc. [11, 160–162]. There is information about Mangit a word related to the Turkic, Mongolian, and Tungus-Manchurian languages, that is, their Altaic Bobotil of the past. The Mongolian form is “myangad”, which means thousands. [2, 63–65].

In the toponymy of Uzbekistan, there are ethnonyms such as Mangit (Buvaida, Dangara, Khojaabad), Mangitguzar (Guzor district),

Mangittop (Qamashi district), Mangitabad (Urgut district), Mangittepa (Ishtikhan District), Okmangit (Pastdargom, Urgut Districts), Mang'it-solma, and Mangityop (Gurlan, Hazoras). In addition, the Republic of Karakalpakstan, as well as Moldavia, Bashkortostan has the toponym Mangit. [6, 92].

Urganji. This name is found in the toponymy of the region in the form Urganjiy, Urganjiylar, Urganjlik, Urganjiyon, and Urganjikhana. Toponym Urganji is an ethnonym. Because this name served to name a group of people (tribe) from Urganch, people who moved from this city, and later their place of residence. On this basis, the ethnonym was formed from the urganji ethnonym. S. Qorayev said, "Urganjis are the descendants of craftsmen who came and settled in Bukhara after Genghis Khan razed Urganch, the capital of Khorezm. [16, 225], according to other sources, Uzbeks left Urganch and settled in Bukhara and other places at the end of the 17th century. Based on this, the meaning of Urganjikhana – "place, village where Urganji people" emerged.

Yobukhana. Detailed information about this ethnonym is recorded in the dictionary of N. Adizova [1, 77-78]. Toponym consists of two parts: yobu+khana. Yobu (jobu) is an Uzbek clan. Yobu were also part of the Kazakhs, and lived scattered in the middle and lower reaches of Zarafshan (Payariq, Pastdargom, Kattakorgon, Narpay, Gijduvon), in the Karshi desert. They were engaged in animal husbandry and agriculture. The descendants of Yobu were absorbed into the Uzbek people. This village was named so because people belonging to the Yabu clan lived in it. Yobu is a Turkic ethnonym, the name of one of the Uzbek tribes. In the sources, it is found in the forms yobu, yovu, yovi, jobili, and jobu. The following poem, written by the poet Ubaydi, is about the position of the Yabu in Sughd:

*Murderers will be found in Sughd,  
The horse on which the Yobus are riding closes.  
At the beginning of the works, the hand did not stop,  
Let the people stay here, or these.*

One of the titles of the supreme ruler of the Turks was called Yabgu (Yabgu). The term "yabgu" is also found in Orhun-Enasoy monuments. In the ancient Turks, this title was given to the head of the tribe or the head of the union of tribes. The Qarluq Khans, one of the ancient peoples of Central Asia, were also called Yabgu. The Yobus formed a small group among the Uzbeks (according to the data of 1926, about 4 thousand people).

Qarga. There is also a Kargali form of this toponym. This name is found in the Kogon and Karakol districts of the region. Qarga is a Turkic ethnonym. In Western Siberia, Turkic peoples have a clan. In Kyrgyz, qarga is one of the ancient clans. The clan of the same name is also recorded in the Bashkirs. Qarga is one of the most common clans among Uzbeks. It is found in various tribes and clans. Khorezm kungirats are a branch of Karagorsok clan of Qajjirali section of southern Uzbekistan kungirats – karaga, Surkhan oasis Turkmens are a karaga clan of Jhilon tamgali section. There is information about the presence of this clan in the Fergana Valley.

Qarga is a totem, a sacred animal (bird). Its Mongolian form is xeree(n) kerey, which means crow [8, 593]. The qarga totem became a clan name. The word representing a totem (qarga) was added the suffix -li during the transition to an ethnonym. The name of the village was probably based on the name of the clan of the Turkomanclus [11, 241].

We also encounter different interpretations of some place names among the people. Among such names, it is possible to include the toponym of Qataghon in the Vobkent district. Suppression of this toponym is a ban; cancel, destroy; to persecute; cases of connection with the word in the sense of repression are observed. This toponym was formed based on an ethnonym. Qatagon is one of the ancient and large tribes of the Uzbek people. Kazakhs and Kyrgyz also have the Katagan tribe. The repressors spoke the Qarluq dialect of the Uzbek language. The place where he lived (village) and other places there (ditch, tepa dasht, kyr...) were named after the tribe. The village of Qatagon was named in this way [11, 360–361].

The phonetic composition of some toponyms has led to the misinterpretation of their etymology due to their pronunciation being close to another commonly used word. In particular, it is possible to express such an opinion about the toponym of the *Khahramon* in the Romitan district. This name has the Kakraman form.

Researcher I. Kholmuratov thought about the Karaman toponym in the Republic of Karakalpakstan and Khorezm province, and its formation 1) The word *khahramon* expresses the meanings of a hero and a person's name. But in the pronunciation of representatives of the dialect, the consonant h in this word is dropped and it is pronounced and written in the form of karaman; 2) Karaman and Aqman – names of Turkmen tribes. Karamanlar is the name of one of the tribes and clans that are part of the Uzbek people; 3) noted that a type of alder tree is called Karaman (Black Oak) in Khorezm [17, 111–112].

Regarding the detailed etymology of this toponym, the following can be stated: “*Karaman is one of the Uzbek clans. I will build in Fergana qipchaks. Other Turkic peoples also have this ethnonym. The Qtai tribe of the Karakalpaks included the Karamanjuli clan. Turkmen have Karaman and Akman clans. It is known that Uzbeks also have an Akman clan. In the ethnonym, the word “black” forms an integral pair*”. [10, 243–244]. Therefore, the toponym of Khahramon is an ethnonym, and it is appropriate to write and pronounce it as Karaman.

In conclusion, it should be said that the toponyms of the Bukhara region have a special place in the toponymy of the Lower Zarafshan oasis due to their division into many internal meaning groups, the reasons for their appearance, and their motivation. Among regional toponyms, there are many place names formed based on ethnonyms. Ethnonyms related to Turkic peoples occupy the main place. Studying them, and analyzing their lexical-semantic, etymological, and motivational features is of great importance for the study of the toponymy of Uzbekistan and serves to inculcate the values related to the centuries-old history of our nation into the minds of the young generation.

#### REFERENCES:

1. Adizova N.B. Lexical-semantic study of microtoponyms of Bukhara district: (Ph.D.) Dissert. – Bukhara, 2021. – p. 80.
2. Begaliev N. From the history of Uzbek ethnonyms. – Samarkand: SamSFLI publication, 2005. – P. 63–65.
3. Begmatov E., Ulukov N. Explanatory dictionary of Uzbek onomastics terms. – Namangan, 2006. – 103 p.
4. Vladimirtsov B.Ya. Social system of the Mongols. Mongol nomadic feudalism. – L. 1934. – P.49.
5. Daniyurov Kh. Genealogies and dialects of the Uzbek people. Monograph. Second edition. – T.: “Navroz”, 2017. – P. 100.
6. Dosimov Z., Egamov Kh. Brief explanatory dictionary of place names. – T.: O’qituvchi, 1977. – 174 p.
7. Karmysheva B.H. Essays on the ethnic history of their regions in Tajikistan and Uzbekistan. – M. 1976. – P. 176–177.
8. Mongolian-Russian dictionary. – M. 1957. – 716 p.
9. Naimov S. Brief explanatory dictionary of place names of Bukhara region. – Bukhara, 1997. – P. 44.
10. Nafasov T. Explanatory dictionary of toponyms of Uzbekistan. – T.: O’qituvchi, 1988. – 288 p.
11. Nafasov T. Kashkadarya rural district. Research of names of villages of Kashkadarya region. Tashkent, “Editor”, 2009, P. 46.

12. National Encyclopedia of Uzbekistan. 12 vols., 1 vol. State Scientific Publishing House “National Encyclopedia of Uzbekistan”. Tashkent, 2000, 388-p.

13. National Encyclopedia of Uzbekistan. 12 vols., 10 vol. State Scientific Publishing House “National Encyclopedia of Uzbekistan”. Tashkent, 2000, 388-p.

14. Kilichev B.E. Bukhara ethnotoponyms and their etymological features//”Scientific information of Bukhara State University”, 2018, issue 1, pp. 64–67.

15. Kilichev Bayramali, Safarova Maftuna The Concept of Totemism and Ethnonym//CENTRAL ASIAN JOURNAL OF LITERATURE, PHILOSOPHY AND CULTURE Volume: 03 Issue: 05 | May 2022, ISSN: 2660-6828 © 2022, CAJLPC, Central Asian Studies, All Rights Reserved

16. Karayev S. Toponymy. – Tashkent: Publishing House of the National Society of Philosophers of Uzbekistan, 2006. – P.75.

17. Kholmuratov I. Linguistic study of the idioms of Southern Karakalpakstan and their lexical bases: filol.fanl.fals.doct...(PhD) dissertation. – Bukhara, 2020. – P. 111–112.

УДК 81-11

## “TEMUR MALIK” ASARIDAGI GIDRONIMLAR

*Kilichev Bayramali Ergashovich,  
Sharopova Ferangiz Nurmatovna  
Buxoro davlat universiteti*

b.e.kilichev@buxdu.uz, f.n.sharopova@buxdu.uz

Ushbu maqolada Mirmuhsinning “Temur Malik” asaridagi gidronimlar, xususan, Xazar (Kaspiy) dengizi, Amudaryo, Sirdaryo, Zarafshon kabilar leksik-semantik va tarixiy- etimologik jihatdan tahlil etilgan.

**Kalit soʻzlar:** “Temur Malik” romani, toponim, toponimika, gidronimlar, gidronimiya, Xazar (Kaspiy) dengizi, Amudaryo, Sirdaryo, Zarafshon, Koʻhak, Narshaxiy, Bobur, leksik-semantik, tarixiy-etimologik.

This article analyzes hydronyms in Mirmuhsin’s work “Temur Malik”, in particular, the Khazar (Caspian) Sea, Amu Darya, Syr Darya, Zerafshan, from a lexical-semantic and historical-etymological point of view.

**Keyword:** “Temur Malik”, novel, toponym, toponymy, hydronyms, hydronymy, Khazar (Caspian) Sea, Amu Darya, Syr Darya, Zerafshan, Kohak, Narshakhi, Babur, lexical-semantic, historical-etymological.

В данной статье с лексико-семантической и историко-этимологической точки зрения анализируются гидронимы в произведении Мирмухсина «Темур Малик», в частности, Хазар (Каспий)ское море, Амударья, Сырдарья, Зерафшан.

**Ключевые слова:** “Темур Малик”, роман, топоним, топонимия, гидронимы, гидронимия, Хазарское (Каспийское) море, Амударья, Сырдарья, Зерафшан, Кохак, Наршахи, Бабур, лексико-семантический, историко-этимологический.

**Mavzuning dolzarbligi.** Badiiy asar tilini oʻrganish orqali muallifning bosh gʻoyasini anglash imkoni paydo boʻladi. Bunda asarda qoʻllangan atoqli otlar, xususan, gidronimlar muhim oʻrin egallaydi. Badiiy, ayniqsa, tarixiy mavzuda yaratilgan asarlarda xalqimizning boy oʻtmishi, millatimizning shakllanish jarayonlari, shuningdek, bugungi kunlargacha yetib kelgan nomlar haqida qimmatli maʼlumotlar yigʻilgan boʻladi. Ularni har tomonlama oʻrganish tariximiz va madaniyatimiz, qolaversa, maʼnaviyatimiz uchun ham muhimdir.

**Muammoning oʻrganilish darajasi.** Tilimizda suv nomlari bilan bogʻliq koʻplab lugʻaviy birliklar mavjud. Ular uzoq davrlardan beri qoʻllanilib kelmoqda. Fanda suv bilan bogʻliq nomlar gidronimlar deb

ataladi. Gidronimlar joy nomlari sanalgan toponimlardan ham qadimiyroq bo'ladi. Chunki tarixdan ma'lumki, kishilar suv havzalari, daryo, ko'l bo'ylarida o'troqlasha boshlagan. O'tqor hayot kechirish esa toponimlarning paydo bo'lishi uchun asosiy omil hisoblanadi. Shu sababli diunyoning deyarli barcha hududida aholi yashash maskanlari daryolar yoki boshqa suv havzalari yaqinida joylashgan.

Aytish o'rinliki, badiiy asar tilidagi gidronimlarni o'rganish borasida sanoqli ishlargina amalga oshirilgan. Bular sirasiga N.Begaliyevning dissertatsiyasini, T.Nafasov, Sh.Nafasov, N.Uluqov, A.Muhammadjonov va boshqalarning maqolalarini kiritish mumkin. Ko'chilik tadqiqotchilar gidronimiyani hududiy toponimiya tarkibidagi bir guruhlar, nomlar qatlami sifatida o'rganish bilan cheharalangan.

Tarixiy gidronimlar, xususan, o'zbek romanchiligi namunalardagi suv nomlari hozirga qadar lingvistik jihatdan tahlil etilmagan. Ularni o'rganish tilimizning atoqli otlar fondi uchun ham, tariximizni o'rganish uchun ham muhim sanaladi [1;314–317].

Gidronim toponimlarning bir turi bo'lib (hidro – suv+onoma – atoqli ot), suvliklar nomini ifodalaydi. Gidronim toponim turi sanaladi. Aytish lozimki, O'zbekiston toponimiyasi gidronimlarga juda boydir. Unda daryo, ko'l, ariq, kanal, soy, buloq, chashma, jilg'a va boshqa komponentli atoqli otlar ko'p uchraydi. Gidronimlarning hosil bo'lishida ushbu komponentlar faol qo'llaniladi va turkiy tillarning deyarli barchasida uchraydi. "Devon-u lug'otit turk", "Tafsir", "Boburnoma" asarlarida *su*, *suu*, *suv*, *sub*, *suy* shakllarida ishlatilganligiga guvoh bo'lish mumkin. Toponimlar tarkibidagi suv so'zi "ariq", "daryo", "buloq" semalariga ham ega: *Qorasuv*, *Oqsuv*, *Kattasuv* kabi.

Xuddi shuningdek, ko'l komponentli gidronimlar ham deyarli barcha turkiy tillarda uchraydi: *Issiqko'l*, *To'dako'l* (*Buxoro*), *Ayg'irko'l* (*Jizzax*), *Aydarko'l*.

Qozoq tilshunos olimi T.Jonuzoqov Qozog'istonning ulkan hududida 85 mingdan ortiq daryo bor, ulardan 8 mingi o'n kilometrdan uzun; ko'llar soni 50 ming atrofida [2; 27], deb qayd etgan.

Biz ushbu axborotda Mirmuhsinning "Temur Malik" romanida uchraydigan Xazar (Kaspiy), Amudaryo, Sirdaryo, Zarafshon kabi ayrim tarixiy gidronimlarni semantik-etimologik jihatdan tahlilga tortamiz.

Xazar. Dengiz nomi. Ushbu gidronim haqida fikr yuritgan P.V.Jilo o'z maqolasida *yer sharida birona dengiz shuncha nomning yarmiga ham ega bo'la olmagan*, deb qayd qilib, dengizning yetmishta nomini sanagan [3].



H.Hasanov dengiz haqida gapirib, uning o‘z tevaragidagi viloyatlarning nomi bilan, sohillardagi mashhur shaharlarning ismi bilan, o‘ziga quyiladigan daryolarning nomi bilan, atrofda yashagan xalqlarning oti bilan atalgan. Kaspiy va Orol dengizlarining tarixiy toponimik kartalari bunga yaqqol misol bo‘la oladi [4; 5], deya qayd etib, P.V..Jilo aniqlaganlariga yana o‘nta nom qo‘shgan.

Ushbu gidronimning etimologiyasi borasida, bizningcha, T.Jonuzoqovning fikri asoslidir. Olim o‘zining maqolasida Xazar – Kaspiy dengizining nomi haqida fikr yuritib, nomning ikki qismdan iboratligi, kas – xalq nomi, pi (fonetik variantlari: bi, bu, ba) – ko‘l, daryo, buloq. Demak, Kaspiy gidronimi ko‘l ma‘nosiga ega bo‘lib, *kas xalqining ko‘li* tarzida sharhlanishi mumkin [2; 28].

Demak, Kasbiy gidronimi *Kas xalqiga tegishli ko‘l* degan ma‘noni bildiradi. Xuddi shunday fikr S.Ataniyazov tomonidan ham bildirilgan: *kasbiy so‘zi kas va (pi) elementlaridan tashkil topgan. Kas (xalq nomi) pi (bi) ko‘l, buloq ma‘nolarini bildiradi* [5; 26].

Hazar so‘zi urug‘ bilan bog‘liq bo‘lishi mumkin. Shu dengiz atrofida yoki yaqinida yashagan qabila nomidan olingan bo‘lishi ehtimoli bor [6; 43].

Amudaryo. O‘zbekiston chegarasidan o‘tuvchi Amudaryo bilan bog‘liq qaydlar qadimgi grek va Xitoy, turkiy yozma yodgorliklarda uchraydi. Ilk yozma ma‘lumot Oks (Oksos) nomi bilan grek manbalarida uchraydi. Oks mahalliy o‘kuz – daryo so‘zining grek tilidagi talaffuzi asosida vujudga kelgan.

Ya.G‘ulomov o‘zining asarida O‘kuz – Amudaryoning mahalliy sak – massaget nomi bo‘lib, o‘rta asrning eng so‘nggi davrigacha saqlanib qolgan [7; 37], deya qayd qilgan.

V.V. Bartoldning “К истории орашения Туркестана” (СПб, 1914. – С. 71) kitobida Amudaryoning yana bir nomi Jayxun haqida fikr yuritilib, uning Tavrotdagi Gixon nomini arabcha shakli asosida paydo bo‘lganligi ko‘rsatilgan.

Daryo nomining hozirgi Chorjo‘y shahrining avvalgi nomi Amuya bilan bog‘liqligi bor [8;13].

Bu fikrni H.Hasanov ham quvvatlab, *daryoning ayrim bo‘laklari-Jar‘yob, Balx daryosi, Termiz daryosi, Kolif daryosi, Urganch daryosi, Xorazm daryosi nomlari bilan yuritilgan. Jumladan, u o‘rta oqimidagi Omul, Omulya (XV asrdan Chorjo‘y) shahri yonida Omuy daryosi deb atalgan, qisqarib Omul bo‘lib qolgan (bu so‘zning asl ma‘nosi deb emas, amurd ismli xalqqa dahli bo‘lsa, ehtimol)* [4; 46], degan.

Demak, Amudaryo nomi Amuya toponimi asosida shakllangan.

Sirdaryo. Daryo nomi. Antik davrda daryo mahalliy tilda Yaksarta deb atalgan, yunon manbalarida qisqacha Yaksart deyilgan. Pliniy *Yaksart daryosini skiflar Silis deb ataydi. Sir – daryoning qadimiy sak tilidagi nomi. Sir nomi dastlab daryoning quyi oqimiga, Yaksart nomi esa oʻrta va yuqori oqimiga tegishli boʻlgan* [9].

Boshqa manbada gidronim hind-yevropa tillariga xos *sal – tuz* soʻzidan olingan. daryo oʻzanida suv yigʻilib, etaklari oqarib turadi, yaʼni *tuzli daryo* demakdir [10; 5–8].

Sirdaryo haqida H.Hasanov fikr bildirib, bu nomni eshritish bilanoq “Sirli daryo” deb fahmlaysiz. Biroq bu daryoning hech siri yoʻq, ammo nomining tarixi fan uchun rostdan ham sirdir, deya qayd etib, uning *sir, yaʼni rang, boʻyoq; togʻ oralaridan, sirtlaridan oqib keluvchi daryo; sariq – loyqa suvli daryo; koʻp suvli daryo* kabi maʼnolarga egaligi haqidagi taxminlar borligini koʻrsatgan. Qadimiy manbalar, shuningdek, yunon tarixchilari fikrlaridan kelib chiqib, Sirdaryo nomi *sir* qabilasi nomidan olinganligi uqtirgan [4; 51].

Bobur oʻzining memuarida *Sayhun daryosikim Xoʻjand suyigʻa mashhurdir, yaʼni Sayhun daryosi Xoʻjand nomi bilan yuritilishini* qayd etgan.

Sirdaryoning Fargʻona daryosi, Banokat daryo, Sayhun, Oʻror daryo nomlari bilan atalgani ham maʼlum. Bu hol gidronimlarda polunimiya hodisasining mavjudligini koʻrsatadi.

Zarafshon. “Zarafshon daryosikim, uni Koʻhak daryosi deb yuritadilar, boshlab kunchiqar togʻlaridan chiqib, unga Iskandar bulogʻi – Iskandar koʻlining suvi va Maschoh togʻlaridan chiqadigan buloqlar qoʻshiladi. Boshqa suvlar va buloqlar bilan birikib, daryo boʻlib Buxoro va Qorakoʻlgacha yetib boradi [11; 122].

Xulosa. Oʻrta Osiyo hududidagi Amudaryo, Zarafshon, Sirdaryo kabilar uzoq davrlar davomida kishilar uchun ichimlik manbayi, chorvalik, dehqonchilik va bogʻdorchilik uchun sugʻorish taʼminoti vazifasini bajargan. Xazar dengizi va unda oqib keladigan suv tarmoqlari xalqlar orasidagi turli-tuman munosabatlarni rivojlantirishga xizmat qilgan.

Ushbu suv manbalarining koʻplab nomlar bilan atalishi esa ularning kishilar hayotida muhim oʻrin tutganligi bilan izohlanadi.

Xullas, asar tilida qoʻllangan gidronimlar oʻzining bir necha asrlik tarixi bilan ham tariximizni, ham madaniy merosimiz izchil oʻrganish uchun poydevor vazifasini oʻtaydi. Ularni semantik-etimologik jihatdan tadqiq etish, paydo boʻlish sabablarini oydinlashtirish kelajak avlodimiz tafakkurini rivojlantirish uchun zarur omildir.

## ADABIYOTLAR:

1. Kilichev B.E., Bafoyeva N. Some characteristics of onomastic units in Abdullah Qodiriy's novel "Bygone days". Asian Journal of Research in Social Sciences and Humanities. ISSN: 2249-7315 Vol. 12, Issue 01, January 2022.

2. Джанузаков Т. Общeturкские основы казахских гидронимов древнейшего периода // Ономастика Узбекистана. – Т., 1989. – С. 27–29.

3. Жило В.П. О названиях Каспийского моря. Известия АН Азербайджана. № 4. 1960.

4. Hasanov H. O'rta Osiyo joy nomlari tarixidan. – Т.: Fan, 1965. – В. 46.

5. Атаниязов.С. Туркменистаннинг топонимияси – Ашхабад, 1981.– Б. 26.

6. Kilichev B.E., Safarova M.Z. Bukhara region's typical toponyms transformed by means the names of nations/ International Scientific Journal Theoretical & Applied Science. Philadelphia, USA. 2021 02 94.

7. Фуломов Я. Хоразмнинг суғорилиш тарихи. – Т., 1959. – Б.37.

8. Дўсимов З., Эгамов Х.Жой номларининг қисқача изоҳли луғати. – Т.: Ўқитувчи, 1977. – Б.13.

9. Кляшторный С.Г. Древнетуркские рунические памятники. – М.: Наука, 1964.

10. Албаров Б.А. Этноним “царцат” осетинских нартских сказаний// IV Всесоюзная научная конференция по иранской филологии (тезисы докладов) –Т., 1964. – С. 5–8.

11. Абу Тоҳирхўжа Самарқандий Самария. Форсчадан Абдулмўмин Сатторий таржимаси. – Т. : Янги аср авлоди, 2009. – Б. 122.

12. Bayramali Ergashevich Kilichev A way of making words on proper nouns. International Journal on Integrated Education. Volume 3, Issue II, Feb 2020, pp. 96–98.

13. Kilichev B.E., Safarova M.Z. Toponyms Formed On The Basis Of Nation Names. International Journals of Sciences and High Technologies. Vol. 25 No. 1 February 2021, pp. 104–107.

14. Kilichev B.E., Safarova O. Study of toponyms in the language of the work. Galaxy International Interdisciplinary Research Journal (GIIRJ) ISSN (E): 2347-6915. Vol.10/ Issue 1, Jan. (2022) 92–95.

15. Bayramali Kilichev, Timur Khojaoglu, Nodira Adizova Classification of Oykons (On the Example of Bukhara Region). International Journal of Culture and Modernity ISSN 2697-2131, 2022, Volume 16.

## УДК 81-11

MAIN GENRES AND FEATURES OF VIRTUAL  
COMMUNICATION*Khalilova Ruxsora Raupovna**Bukhara State University**ruxsoraxalilova@gmail.com*

The article focuses on the ability of verbal communication to convey information and the linguistic essence of non-verbal linguistic means-their place in the grammatical system. Digital speech communication has brought people's communication to a new level. Virtuality, equality of communicators; observation of the peculiarities of machine voice, such as the hypertext form. The article focuses on address (mass or personalized direction), contact direction, time direction (synchronous/asynchronous communication), level of interactivity (speed of response of recipients to messages), regulation (presence of strict specifications for message form/content), taking into account the parameters of the communication organization of well-founded opinions.

**Key words:** speech communication, interactive speech communication, communication mechanism, networking field, forum, social network, blog, audiovisual communication, internet.

Communication is impossible without the contact and interaction of members of the group with each other (sharing information and receiving information). For this reason, the idea of 'visual contact' seems to be obsolete. Communication is an important part of speech, of the exchange of ideas and of knowledge. But verbal communication (Russian: «речевое общение», English: «discourse») is one of the most popular functional words. In certain instances, of course, the method of reasoning relates to the material realization of linguistic possibilities in writing or orally. However, there are also several forms of contact not related to the occurrence of linguistic units. [14]. Non-materialized modes of communication of linguistic means are also called non-verbal communication. In our culture, and also in the rules of Sharia, there is an ethical and esthetic concept of "Silence – a sign of consent" which should also be used as a means of unconditional communication. Take a look at the story "Sharoit" by Tilavoldi Juraev:

...There was a peculiar silence in the car. It was by no means a silence of bitterness, rage, resentment, remorse. This silence has been a secret silence (Juraev T. Villages).

Another example of this would be Yu. Solntseva's "Seventeen Moments of Spring," in which Stirlitz, a spy, meets his wife, who came to see her as a casual companion at the Elephant Café in Berlin. This contact was a very necessary meeting for both Stirlitz and his wife.

This meeting informs that both parties are safe and that they miss each other. In fact, this information was given, accepted without words-language means.

It is clear from this record that speech activity is inextricably linked to a set of factors such as linguistic, ethical, aesthetic, national, spiritual-cultural, social, causal (cause, effect, purpose) and so on. However, this is not the only basis that needs a particular pragmatic term of speech communication – speech communication is often distinguished as a specific form of use of semiotic systems. What's the feature? While it is a semiotic system based on the structure of the language[11], it varies significantly from other semiotic systems[11]. As the main aspect of these variations, scientists unanimously point out three characteristics. These are followings:

1. Language can be the basis of all other social semiotic systems in the sector. Thus, language can be called the general, primary social semiotic system, the main basic system that easily replaces it.

2. In language, unlike all other semiotic systems, the relationship between expressive and non- expressive sign is asymmetrically dualistic in nature. While the form/representation may be linked to multiple content/expressions, the same content/representation may be related to different forms/expressions[4].

3. Language signs are self-evolving.

There is another form of communication that could be called “virtual speech communication.” Various global communication opportunities have emerged with the development of information technology. With the advent of the Internet, brokerage connections have increased considerably and various methods of data transmission have emerged. Emails, forums, social networks, blogs, chats, etc. [7], for example. Virtual speech communication is taking people's communication to a new level. This allows you to store and transmit large amounts of data, communicate online, and use audio-visual communication channels. Digital speech communication is wider than other forms of communication.

Virtual dialog can be defined as follows: 1) the mass nature of communication; 2) the direction of communication: from many to many; 3) the synchronous communication; 4) the average speed of notification to the recipient; 5) the lack of strict criteria for the form and content of the message. Today, a modern electronic communication environment has been created, such a communication service being introduced in two different ways: direct (PC, chat) and indirect (email, forum, teleconference)[1,2,5].

Chat is one of the methods of direct contact. It's a networking program that is a way of rapidly exchanging messages over the Internet in a real-time system. Communication and messaging in the forum-related chat system takes place in the real-time system. The word chat is derived from English and means "friendly chat, hangoma"-electronic chat, real-time machine contact. Chat is a conversation between many Internet users at the same time[5]. In this case users are talking via the Internet, that is, exchanging news via email, discussing a subject or talking to each other. In such a virtual communicative space, all users communicate by exchanging messages in the form of peer-to-peer writing. In the field of virtual communication, speech discourse means the simultaneous communication between two or more users through the Internet. It does not matter where the users are located, i.e. where the users are in the virtual communication space. You only need one of the following special programs: Skype, Mail Agent, Google Talk, ICQ. Users can communicate in a virtual communicative process by connecting a microphone and a headset to a device and using apps. Chat, like regular speech communication, has three components: beginning a conversation, saving a conversation, and finishing a conversation. The user communicates on the basis of the language and pragmatic rules of the Uzbek language chat, which convey their presence in the virtual communication space.

If we say video communication over the Internet, users see each other on a computer screen (online), which means that the moving images of users are transmitted to each other. The computer that connects to this video is called a webcam. In the course of video contact over the Internet, not only do all the users involved in the interaction hear each other, but they also see each other.

In short, the issues of oral and written communication between people are repeatedly discussed by philosophers, sociologists and linguists, contributing to the advent and further growth of the theory of speech communication. As the global computer network opens the door to new possibilities for contact between members of different ethnic and cultural groups, there is a need to further deepen scientific studies on the topic of language and culture in Internet linguistics.

#### REFERENCES:

1. Galichkina, E. N. Kompyuternaya konferentsiya v sisteme janrov kompyuternogo obshcheniya. Mezhvuzovskiy sbornik nauchnykh trudov – Saratov: Izd-vo «Nauchnaya kniga», 2003. – Выпуск 9. – С. 276–280.

2. Demyankov, V. 3. Interpretation, understanding and linguistic aspects of modeling – M.: Izd-vo Mosk. un-ta, 1989. – 172 p.

3. Jichkina, A.E. Sotsialno-psychologicheskije aspekty obshcheniya v In-ternete [http:// fioqiston.ru/projects/articles/strategy.shtm/](http://fioqiston.ru/projects/articles/strategy.shtm/).

4. Kartsevskiy S. Ob asymmetrichnom dualism lingvisticheskogo znaka // Zvegintsev V.A. History of writing in essays and excerpts. Ch.II. – Moscow: Prosveshchenie, 1965. – 85 p.

5. Kremleva, S. «ChAT» as the innovation of virtual communication // <http://sociology.extrim.ru/dip/start.htm>

6. Loshmanova L.T. Jargonizirovannaya lexika v bytovoy rechi molodyoji 50–60-x godov. Avtoref.diss... kand.filol.nauk. – Leningrad, 1975. – P. 57–59.

7. Ovcharova, K. V. Electronic discourse: general characteristics // Linguistic organization discourse: functional and content aspects: Mejevuz. sb. tr. molodyx uchenyx. – Krasnodar, 2004. – Vyp.1. – S. 108–113.

8. Mirtojiev M., Mahmudov N. Language and culture.– Tashkent: Uzbekistan, 1992. – 110 p.

9. Muminov S.M. Socio-linguistic features of Uzbek communication behavior: Filol.fan.d-ri. ... diss. –Tashkent, 2000. – B. 198–204.

10. Pocheptsov G.G. O meste pragmaticheskogo elementa v lingvisticheskoy opisaniy // Pragmaticheskije i semanticheskije aspekty sintax. – Kalilin: KGU, 1985. – S. 12–18.

11. Solntsev V.M. Yazyk kak sistemno-strukturnoe obrazovanie. – Moscow: Nauka, 1971. – S. 95.

12. Saussure F. Trudy on yazykoznaniiyu. –Moscow: Progress, 1977. – S. 52.

13. Straten V.V. Argon and argotism. – V sb.: «Trudy komissiiy on the Russian language». – Leningrad, 1931. – P. 91–98

14. Toirova G. About ethnosociopragmatism or pragmatic sociolinguistics. Scientific information of Bukhara State University. 2008, № 1. – P. 53–55.

15. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277–3878, Volume-8 Issue-2S10, September 2019. – P.352–355. (scopus)

16. Toirova G, Abdurahmonova N., Ismoilov A., Applying Web Srawler Teshnologies for Sompiling Parallel Sorpora as one Stage of Natural Language Prossessing. 2022 7th International Conference on Computer Science and Engineering (UBMK) Sep. 14–16, 2022, Diyarbakir /Turkey pp. 73–75. (scopus)

17. Toirova G. The importance of linguistic models in the development of language bases. // Buxoro Davlat universiteti ilmiy axboroti. – Buxoro, 2020. – №6. – B. 98–106.

## УДК 81-11

FUQAROLIK JURNALISTIKASIDA TIL VA USLUB  
MASALALARI*Malikova Go'zal<sup>1</sup>, Sharofiddinova Mahtob<sup>2</sup>*<sup>1</sup>*Buxoro davlat universiteti, Buxoro, O'zbekiston*<sup>2</sup>*BuxDU jurnalistika yo'nalishi 2-bosqich talabasi*<sup>1</sup>*gmalikova@gmail.com*

Ushbu maqolada ayni paytda fuqarolik jurnalistikasining erkin minbariga aylangan ijtimoiy tarmoqlar tili va uslubi haqida so'z yuritiladi, ba'zi kanallar tavsifi misollar orqali tahlil qilinadi. Onlayn kontentlar o'rganilib, ularda adabiy til me'yorlarining buzilish holatlari, uslubiy jihatdan yo'l qo'yilgan xatolik va g'alizliklar, feyk xabarlar va ularning ijtimoiy tarmoqlardagi ulushi borasida mulohaza qilinadi.

**Kalit so'zlar:** Internet, ijtimoiy tarmoq, feyk, faktcheking, onlayn kontent, adabiy til me'yori, auditoriya, nutq.

ISSUES OF LANGUAGE AND STYLE IN CITIZEN  
JOURNALISM*Malikova Guzal<sup>1</sup>, Sharofiddinova Makhtob<sup>2</sup>*<sup>1</sup>*Bukhara State Universitet, Bukhara, Uzbekistan*<sup>2</sup>*2nd-level student of Journalism department of BukhSU*<sup>1</sup>*gmalikova@gmail.com*

At the same time, this article talks about the language and style of social networks, which have become a free platform of citizen journalism, the description of some channels is analyzed through examples. Online content is studied, cases of violations of literary language standards, methodological errors and idiosyncrasies, fake news and their share in social networks are discussed.

**Key words:** Internet, social network, fake, fact-checking, online content, standard of literary language, audience, speech.

ПРОБЛЕМЫ ЯЗЫКА И СТИЛЯ В ГРАЖДАНСКОЙ  
ЖУРНАЛИСТИКЕ*Маликова Гузал<sup>1</sup>, Шарофиддинова Махтоб<sup>2</sup>*<sup>1</sup>*Узбекское языкознание и журналистика**преподаватель кафедры*<sup>2</sup>*Студентка 2 курса факультета журналистики БуксГУ*<sup>1</sup>*gmalikova@gmail.com*



В то же время в данной статье говорится о языке и стиле социальных сетей, ставших свободной площадкой гражданской журналистики, на примерах анализируется описание некоторых каналов. Изучается онлайн-контент, обсуждаются случаи нарушения норм литературного языка, методические ошибки и особенности, фейковые новости и их доля в социальных сетях.

**Ключевые слова:** Интернет, социальная сеть, фейк, фактчекинг, онлайн-контент, уровень литературного языка, аудитория, речь.

Bugun dunyo bo'ylab internet foydalanavchilarining aksariyati Facebook, You tube, WhatsApp, Instagram, Twitter, Snapchat, Pinterest, line, Telegram kabi ijimoiy tarmoq va messenjerlardan foydalanishmoqda. Ular jahon bo'yicha qamrovi va auditoriyasi keng ijtimoiy muloqor platformalari hamdir. Bir asr ilgari ommaviy kommunikatsiyaning bu tur yo'nalishi faqatgina xayollarda yoinki fantastik yo'nalishdagi adabiyotlarda uchragan. Ammo bugun biz bir necha o'n yillar ilgari inson xayolotining mahsuli bo'lmish muloqot turlaridan real hayotda faol foydalanmoqdamiz. Hatto yuz ko'rishib diydorlashish, ovozli xabarlashish, har qanday hajmdagi ma'lumotlarnida yashin tezligida o'rtog'imiz, qarindoshimiz yoki birodarimizga yetkazishda bugun ijtimoiy tarmoqlarga yetari yo'q, albatta.

“Sovuq urush” ning mahsuli bo'lmish Internet (lotincha: “inter” – “aro” va “net” – “tarmoq” – standart minimal internet protokoli orqali ma'lumot almashuvchi kompyuter tarmoqlarining butun jahonda omma uchun ochiq to'plami) ayni paytda insoniyatning uzog'ini yaqin, og'irini yengil qilayotgani rost. Ammo shunday jihatlari ham borki, ba'zan foydasidan ko'ra ko'proq zararini his etamiz. Masalan, jurnalistikada sohaning asosiy prinsiplaridan biri bo'lmish xolislik tamoyili bor. Har qanday axborot haqqoniy, to'g'ri va obyektiv bo'lishi lozim. Unda jurnalistning tarafkashligi yoki “men”i yoqlanmaydi. Xolisligini yo'qotgan axborot esa o'z-o'zidan jamiyat orasida turli tushunmovchilik va nizolarni, salbiy oqibatlarni keltirib chiqaradi.

Ijtimoiy tarmoq – bu qiziqishlari o'xshash yoki onlayn aloqaga ega bo'lgan odamlar o'rtasida muloqot qilish, tanishish, ijtimoiy munosabatlar yaratish uchun, shuningdek, ko'ngilochar ( musiqa, film yoki videoroliklar) va ish maqsadlarida ishlatiladigan onlayn platforma. Ijtimoiy tarmoqlarning eng avvalgi shakli elektron e'lonlar taxtasi bo'lib, ularning birinchisi CBBS nomi ostida IBM xodimi

U. Kristensen tomonidan 1978-yilda yaratilgan. 1983-yilda allaqachon dunyoda 800 ta, 1988-yilda esa 5000 ta elektron doskalar mavjud edi. Shu sayin katta tezlik bilan ijtimoiy tarmoqlar va ularning foydalanuvchilari soni keskin ko'payib boraverdi.

Yaqinda “We Are Social” va “Hootsuite” kompaniyalari “Digital 2021” hisobotini taqdim etishdi. Unda jahon bo'yicha Internet va ijtimoiy tarmoqlar foydalanuvchilari soni qanchaga yetgani haqida ma'lumot berilgan. Worldometers.com ma'lumotlariga ko'ra, bugungi kunda dunyo aholisi soni 8 mlrd.dan ortiqni tshkil etmoqda [1]. Statista.comning ko'rsatishicha, Internetdan foydalanuvchilar soni ayni paytda 5,16 mlrd.ni (bu sayyormizdagi insonlarning 64,4% degani), ijtimoiy tarmoq foydalanuvchilari esa 4,76 mlrd.ni tashkil etmoqda[2]. Demak, har kuni bo'lmasa ham kunora yuqorida raqamlarda ko'rsatilgan insonlar axborotlarni aynan onlayn platform orqali qabul qilishadi. Siz-u bizga ma'lumki, Internet jurnalistika masalasida barcha ma'lumot ham, afsuski, to'g'ri bo'lib chiqavermaydi, ularning orasida yolg'oni qancha. Ammo ko'pchilik tarqatilayotgan xabarlarining rost-yolg'onligini saralamaydi, faktcheking (faktlarni tekshirish) olib bormaydi: qanday axborot uzatilayotgan bo'lsa, uni boricha qabul qilib, atrof-dagilarga aynan o'sha mazmundagi yangilikni tarqatayotganlar ham ko'plab topiladi. Sababi jamiki foydalanuvchilarda axborotni saralash, uni tanlab o'qish ko'nikmasi shakllanmagan. Ayni mana shu muammo bugun ko'pgina jamiyatlarining “oyog'idan chalmoqda”. Birgina axborot xuruji, axborot urushi sabab o'zaro va millatlararo nizolar, kelishmovchiliklar kelib chiqmoqda.

O'zbek jurnalistikasi ham yuqoridagi muammolardan bebahra qolayotgani yo'q. ayniqsa, internet jurnalistika va ijtimoiy tarmoqlar masalasida. Bir soat oldin onlayn platformalarorqali tarqatilgan “o'ldi” yoki “ishdan olindi” qabilidagi xabarlar zum o'tmay, “feyk” (yolg'on) maqomini oladi. Sababi bu kabi xabarlarini tarqatayotgan mualliflar, bloggerlar, kanal egalari ularni tekshirib ko'rish haqida bosh ham qotirib ko'rishmaydi. Shov-shuv bo'lsa, kanali reytingi ko'tarilsa bas. Albatta, barcha kanallarni bunday deyoilmaymiz, lekin aksariyatida bugun shu hol.

Masalan, kuni kecha **“maktablar 12 yillik ta'lim tizimiga o'tishi”** xususida xabarlar tarqab ketdi. Vaholanki, Maktab va maktabgacha ta'lim vazirligi bunga rasmiy munosabat bildirib, bu kabi xabarlar asossiz ekanini ta'kidladi. Yoki **“Jizzaxda yangi oltin koni topildi”** sarlavhasi ostida bir necha kanallar postlar tarqadi, alaloqibat bu ham yolg'on bo'lib chiqdi. Shu bois ham bugun ommaga axborot tarqatayotgan har bir inson haqqoniylik va xolislik prinsiplariga tayanishi, foydalanuvchini chalg'itmosligi loizm. Bu uchun albatta jurnalistika sohasidan xabardorlik va professionallik ham talab etiladi.

Dilni xira qiladigan yana bir holat borki, bugun ko'pgina yengil-yelpi, shov-shuvli xabarlar (matnli, ovoqli, lavhali) tarqatishga ixtisoslashgan ayrim content va kanallar qaysidir ma'noda xalqimizning savodxonligiga ham putur yetkazyapti. Postlarda shevadan foydalanish, adabiy til me'yorlarini buzish, jumalalarda g'alizlik, fikr tarqoqligi kabi bir necha salbiy holatlar kuzatilmoqda. Gapimizning isboti sifatida quyidagilarni ilova qilsak:

*“Har bir ayol eri uni yaxshi ko'rishligini xohlaydi. Er-xotin o'rtasida muhabbat bo'lmasligi oilaviy hayotni zerikarli, sovuq qilib qo'yadi...” ko'rishligini* – ko'rishini. Yoki *“1-maydan 11-maygacha Yerda kuchli magnit bo'ronlari bo'ladi, davleniyasi borlar ogoh bo'ling” Davleniya* – qon bosimi (Uzbek Gold kanalidan).

*“Qaynonasini qaynonasini qaynonasi bor ekan...”, “Sizdan ugina – bizdan bugina. Zato qorin to'yvotti”* (Uzbek video channel).

*“Voyaga yetmagan o'smirlar Saudiya Arabistoniga mustaqil yoki ota-onalari yoxud o'rnini bosuvchi shaxslar bilan borganlari haqida ma'lumot yo'q. suratda ular bilan birgan o'zbekistonlik jurnalist Karimberdi To'ramurodovni ko'rish mumkin.”* (Qoraxabar)

*“Ukraina prezidentini so'zlariga ko'ra, Putin har qanday keliguvni buzadi...”* (Kun.uz)

*“Bakuga borganimda mug'om konsertiga bilet olgandim shu saytdan, onam Ozarbayjon mug'omlarini yaxshi ko'radi. Biletlarni Uzb-dan turib olganman. Shunchun pochtaga xat kepturadi. Tak vot, mono xoramilar kelyaptikan Bakuga. Bunaqa yaxshi yaqin shans bo'maydi xali beri qiziqqanlarga. Viza kerakmas do'vlat topshiringlar”.*

*“Mashetiga kino tavsiya qilib ketay. Pryam estetik zavq olib, o'ylanib, nostalgiyalar ko'zlarimga kirib ketib ko'rdim. Tverdo tavsuyu”*(Alimoff live).

Yuqoridagi misollarda arzimadek tuyulgan imloviy, jumla bilan bog'liq xatoliklar, hatto haqoratli so'zlarning ishlatilishi bora-bora o'sha kanal foydalanuvchilari uchun normal holatga aylanadi. Vaholanki, ulardagi obunachilar soni 10 mingdan 155 mingtagacha. Bular birgina Telegram messenjerida ko'zimiz tushganlari. Facebook, twitter, Instagram kabilardagi post va izohlarni gapirmasa ham bo'ladi. Bevosita ommaga axborot yoki ma'lumot tarqatayotgandan keyin uning o'ziga yarasha mas'uliyati va me'yorini bilish lozim, chamamizda.

Oliy Majlis Senatining axborot siyosati va davlat organlarida ochiqlikni ta'minlash masalalari qo'mitasi bergan ma'lumotga ko'ra, mamlakatimizda eng ko'p foydalaniladigan ijtimoiy tarmoq

„Telegram” messenjeri boʻlib, hozirda undan 18 mln.dan koʻp inson foydalanib kelmoqda. Shuningdek, “Facebook”dan 4,7 mln., “Odnaklassniki”dan 16,7 mln., “Instagram”dan 3,7 mln., “Linkedln”dan 288 ming, “Twitter”dan esa 51,6 mingdan ortiq foydalanuvchi foydalanib kelmoqda va bu koʻrsatkich kundan-kunga ortib bormoqda.

Kun sayin uslubi oʻzgarib borayotgan global ijtimoiy tarmoq tilga taʼsiri oqibatida ikki xarakterli tamoyil yuzaga kelmoqda:

Birinchidan, adabiy til meʼyorlari bilan solishtirganda, baʼzi til vositalari soddalashib, baʼzilari esa murakkablashib bormoqda. Masalan, hislarni ifodalash ijtimoiy tarmoqlarda koʻpincha smaylik yoki emoji-lar tarzida beriladi.

Ikkinchidan, yozma va ogʻzaki nutqning qorishuvi va aksincha almashinuvi oqibatida til meʼyorlariga putur yetmoqda. Koʻpgina ijtimoiy tarmoqlardagi postlar soʻzlashuv yoʻnida koʻcha tilida yozilganiga guvoh boʻlamiz. Nutqni toʻgʻri tuzish uchun tadqiqotchilar quyidagicha fikr bildiradilar:

“Nutqni grammatik jihatdan toʻgʻri qurish, yaʼni soʻz yasash, soʻz oʻzgartirish, soʻz birikmalari va gaplardagi toʻgʻri grammatik bogʻlanishlarning buzilishi grammatik xatolar sifatida tan olinadi. Nutqiy xatolar, grammatik xatolardan farqli oʻlaroq, til birliklarining tarkibi yoki tuzilishida emas, balki ularni qoʻllashda, aniqroq aytganda, koʻproq toʻgʻri tuzilgan soʻz va soʻz birikmalarini notoʻgʻri, nooʻrin, nojoʻya qoʻllashda kuzatiladi”[3.1].

Yuqoridagi mulohazalardan shunday xulosaga kelish mumkinki, yoʻl qoʻyilayotgan har bir xatolik – u xoh imloviy boʻlsin, xoh uslubiy, bora-bora oʻquvchini oʻsha xatolikka koʻzini oʻrgatib qoʻyadi, auditoriyani savodsizlik sari yetaklaydi. Shuningdek, xabar yo matnlarda jurnalistik pozitsiya bildirilmaydi, ammo ijtimoiy tarmoqlardagi postlarda bu meʼyorlar buzilishiga koʻp bora guvoh boʻlamiz.

“Odatda, matnda muallifning fikri va bahosi mavjud boʻlmaydi, chunki jurnalistning asosiy maqsadi tinglovchilarga tezkor, aniq va obyektiv maʼlumotlarni berish va xabardor qilishdir.”[3.2]

Xulosa oʻrnida shuni aytish mumkinki, ijtimoiy tarmoqlarda oʻz kanalini yuritayotgan jurnalist yoki blogerlar reyting koʻtarish yo shov-shuvli xabarlar, xalqqa yoqadigan soda tilda axborot yetkazish xayolida adabiy til meʼyorlarini buzmaslik kerak. Zero, soʻz yuki, soʻz zalvorini his eta bilish jurnalistikaning chinakam yuklaridan biridir.

## ADABIYOTLAR :

1. <https://www.worldometers.info/world-population/>
2. <https://www.statista.com/statistics/617136/digital-population-worldwide/>
3. Тешабаева Д.М., Бакиева Г.Х., Израил М.И., Тошмухамедова Л.И., Нуритдинова М.Ч. Журналистика. III жилд. Медиалингвистика ва тахрир. – Т.: O'zbekiston, 2019.
4. Yuldasheva D.N., Chullieva G. T. THE ROLE OF NOVERBAL MEANS IN SPEECH MEANING. International Engineering Journal For Research & Development, Vol.6, Issue 2 published May 16, 2021. – Pag. 1956–1961. [www.iejrd.com](http://www.iejrd.com) <http://www.iejrd.com/index.php/%20/article/view/1956>
5. Malikova Go'zal, Sharofiddinova Mahtob. (2023). IJTIMOIIY TARMOQLAR TILI VA USLUBI. Innovations in technology and science education 2/10. <https://doi.org/10.5281/zenodo.7979519>

## УДК 81-11

MUHAMMAD YUSUF SHE'RIYATIDAGI AYRIM MUALLIF  
NUTQI NEOLOGIZMLARINING IFODASI XUSUSIDA*Mardonova Sarvinoz Oripovna**Buxoro davlat universiteti**sarvinoz@gmail.com*

Mazkur maqolada Muhammad Yusuf she'riyatidagi ayrim muallif nutqi neologizmlari, mazkur hodisa tadqiqi, ya'ni affiksatsiya, kompozitsiya usulida hosil bo'lgan muallif nutqi neologizmlari, ularning semantik-morfologik xususiyatlari, matndagi ahamiyati o'rganilgan va shoir ijodidan olingan she'riy parchalar orqali dalillangan.

**Kalit so'zlar:** lingvopoetika, individual neologizmlar, matn lingvistikasi, affiks, yasama so'z, affiksatsiya usuli, kompozitsiya usuli, ot, sifat, lug'aviy birlik, murakkab so'z.

ON THE EXPRESSION OF SOME NEOLOGISMS OF AUTHOR'S  
SPEECH IN MUHAMMAD YUSUF'S POETRY*Mardonova Sarvinoz Oripovna**Bukhara State University**sarvinoz@gmail.com*

The article describes some neologisms of the author's speech in Muhammad Yusuf's poetry, a study of this phenomenon, that's to say, neologisms of the author's speech formed by the method of affixation, composition, their semantic-morphological characteristics, their importance in the text have been studied and proved through poetic samples of the poet's works.

**Key words:** linguopoetics, individual neologisms, linguistics of text, affix, derivative, affixation method, compositional method, noun, adjective, linguistic unit, compound word.

О ВЫРАЖЕНИИ НЕКОТОРЫХ АВТОРСКИХ РЕЧЕВЫХ  
НЕОЛОГИЗМОВ В ПОЭЗИИ МУХАММАДА ЮСУФА*С. О. Мардонова**Бухарский государственный университет**sarvinoz@gmail.com*

В данной статье рассматриваются некоторые неологизмы авторской речи в поэзии Мухаммада Юсуфа. Исследованы неологизмы авторской речи, образованные по способу аффиксации, композиции, их семантико-морфологи-

ческие характеристики, их значение в поэзии и доказаны фрагментами из творчества поэта.

**Ключевые слова:** лингвопоэтика, индивидуальные неологизмы, лингвистика текста, аффикс, производное слово, способ аффиксации, способ композиции, существительное, прилагательное, лексическая единица, сложное слово.

Badiiy matnni lingvopoetik tahlili jarayonida taqiqotchi oldida ijodkor qalamiga mansub individual jihatlar, ijodkor individual uslubiga xos boʻlgan struktural oʻzgarishlar, ijodkorning individual nutq uslubini ochib berish vazifasi ham koʻndalang turishi tabiiy. Oʻzbek lingvopoetikasida bu borada, yaʼni badiiy uslubda individuallik, uning koʻrinishlari borasida ayrim ijodkorlar misolida u yoki bu darajada tadqiqotlar amalga oshirilgan [2;3;5;6;8;11]. Lekin mazkur tadqiqotlar ijodkor individualizmi (individual neologizmlar)ning oʻrganilishini maʼlum bir foizinigina tashkil etadi xolos. Qolaversa, bugungi kun matn tilshunosligida badiiy matnning antroposentrik talqini borasidagi tadqiqotlarning ommalashib borayotgani[7] ham individual nutq neologizmlarining chuqurroq tahliliga ehtiyoj tugʻdirmoqda.

Sir emaski, haqiqiy ijodkor soʻz ustida tinimsiz mehnat qiladi, uning yangi-yangi maʼnolarini kashf etadi, turli-tuman koʻchimlar yaratadi. Bu haqda, yaʼni badiiy adabiyotda soʻzning oʻrni borasida H.Gʻulom quyidagicha fikr bildiradi: "Til ustidagi ish yozuvchi ijodining tub asosidir. Mavzu tanlangach, maqsad aniqlangach, qahramonlar, ularning portretlari va xarakterlari loʻnda tasavvur etilgach, adib qoʻliga qalam olib, asarni boshlar ekan, soʻz ustida davomli, mashaqqatli, benihoya ish boshlanadi. Bu ishda yozuvchining bisoti – uzoq yillar davomida yigʻib toʻplagan soʻz xazinasidan u qunt bilan jumla shodalari teradi, oʻy-fikrdagi boylik ham, yozuv stollaridagi yashirin yon daftarlar ham ishga tushadi. Soʻz kerak! Yangi, ohori toʻkilmagan, voqeaga, qahramon ichki dunyosiga, uning nozik feʼl-atvorini ochishga xizmat qiladigan, oʻquvchini dafʼatan hayratga solib qoʻyishga qodir soʻz kerak!.. Oʻz asariga tilning kalitini topa bilgan yozuvchi oʻz uslubida ravon ijod qila oladi. Dunyoda til yaratishdan murakkab sanʼat yoʻq"[10, 171–172].

Shu nuqtayi nazardan qaraydigan boʻlsak, sevimli shoir Muhammad Yusuf ham "tilning kalitini topa bilgan" ijodkorlardan sanaladi. Uning sheʼriyati jozibadorligi, koʻngilga xush yoqishi, kuy va qoʻshiqlarda silliq uygʻunlasha olishi bilan barchani oʻziga rom etadi. Bunda, albatta, shoirning ona tilimizning kuchli bilimdoni ekanligi,

shoirning til sezgirligi o'ta yuqori bo'lganligi qo'l kelgan. Shoirning o'quvchilar qalbidan joy egallashi ham uning so'z qo'llash mahorati bilan bog'liq.

Muhammad Yusuf yangi-yangi so'zlar ijodkori ham. U ona tilimiz lug'at qatlamini boyitishda ham samarali ijod qildi. Shoirning muallif nutqi yasalmalari turli shakl va mazmunda bo'lib, ularning yasalishida hech qanday sun'iylik, zo'rma-zo'rakilik sezilmaydi. Muhammad Yusuf o'zbek tilining ichki imkoniyatlari asosida, ya'ni ham affiksatsiya, ham kompozitsiya usulida betakror so'zlar hosil qildi. Shularning ayrimlarini kuzatamiz.

Ma'lumki, o'zbek tili lug'at qatlamida *-boz* affiksi bilan yasalgan 50ga yaqin so'z bor [4, 51–52]. Mazkur affiks, odatda, fors-tojik lug'at qatlamidan yasama so'zlar tarkibi bilan o'tgan bo'lib, o'zbekcha so'zlardan oz miqdorda yangi so'z yasashi aytiladi. A.Hojiyev *-boz*, *-navis*, *-xon*, *-paz*, *-do'z*, *-boz* shaxs oti yasovchi affikslari qatori *-boz* shaklini ham yangi so'z yasay olmasligi, deyarli mahsulsiz ekanligini ta'kidlab o'tadi [1, 75]. Muhammad Yusuf ijodini kuzatish davomida guvohi bo'lamizki, *-boz* affiksi bilan so'z yasash davom etyapti. Shuning o'ziyoq *-boz* affiksining so'z yasovchi bo'lolmasligi haqidagi aytilgan fikrlarni rad etadi. Jumladan:

Men shoir emas, men *nazmboz* kosib,  
She'rim shuhratimga emas munosib.

\* \* \*

Eslamasa nima, men bir to'pori,  
Qishloqi *so'zbozni*, osmon yiqilmas.

\* \* \*

Shoir o'lsa, kim ham kuyunar deysiz  
*Madhiyaboz* hozir suvarakdan ko'p.  
Ular misli qishda tarnovdagi muz,  
Barmoq tegsa yerga to'kilar to'p-to'p.

Muhammad Yusuf "Maktub" she'rida *-vachcha* elementi bilan *shoirvachcha* so'zini hosil qilganini kuzatamiz:

Ko'ngil ayvonida kuz hushtak chalar,  
Ko'zimdan yosh bo'lib yog'adi yaproq.  
Do'stlarim – takabbur *shoirvachchalar*  
O'zlarini sevishar yoridan ko'proq.



”-*vachcha*” aslida tojik tilida yosh bola ma’nosidagi *bachcha* so’zidan olingan bo’lib, bu element otlarda qavm-qarindoshlik (amaki-vachcha, xolavachcha) hamda jonivorlarga nisbatan qo’llanib (itvachcha, mushukvachcha) uslubiy bo’yoqdorlikni ta’minlashda ishlatiladi. Tahlilga tortilgan misolda *shoivachcha* so’zi salbiy ma’no ifodasi uchun qo’llangan.

Muhammad Yusuf she’riyatida -roq affiksining, ya’ni lug’aviy shakl yasovchi affiksning so’zga o’zgacha holda qo’shilganligiga duch kelamiz, ya’ni mazkur affiks o’z turkumi doirasidagi so’zga emas, balki o’zga turkum so’zga birikib kelgan: So’rsang, sevinchimdan o’kinchim ko’proq,

Sendan-da *kumushroq* sochimdagi oq.

Yilim kelmay turib, yilim ketmoqda,

Ko’zda tabassum-u kiprikda titroq. (“Yangi yil kechasi”)

-roq affiksi qiyosiy darajadagi sifatlarni yuzaga keltiradi. Bu affiksning ot lug’aviy birliklariga birikishi tabiiy hol emas, aslida. *Kumushroq* shaklidagi so’zning yuzaga kelishida *kumush* so’zining belgi ma’nosida ham qo’llanishi sabab bo’ladi. Ma’lumki, ayrim otlar sifatiy belgi ham ifodalay oladi. Masalan: *cho’yan pech*, *taxta ko’prik*, *to’sh yo’l*, *oltin yaproq*, *xonatlas ko’ylak* va h.k. Ayrim manbalarda bu kabi holatga nisbatan sifatlashish termini ishlatiladi. Aslida esa mazkur holat, ya’ni otlarning belgi ma’nosi uchun qo’llanishi “ko’chish” hodisasiga hech qanday aloqasi yo’q. Aytmoqchimizki, otlarning belgilik ma’nosini, sifatlarning predmetlik ma’nosini ifoda etishi mazkur turkumdagi ayrim so’zlarning tabiiy holati bilan belgilanadi. Bu holat so’z turkumlarining davrlar osho yondosh “yashab kelishi” bilan bog’liq holat. Shu bilan bog’li holda *kumush* so’zining ham rang-tus, ham bu so’zning qiymatlilik darajasi nutqimizda turli o’xshatishli birikmalarni hosil bo’lishiga sabab bo’lgan, deyish mumkin. Shoir Muhammad Yusuf kumushning shu tabiiy holati (rang-tusi)ga tayanib *kumushroq soch* epitetini yuzaga keltirgan, deyish mumkin. Ko’rinadiki, shoir muallif nutqi neologizmlari faqat so’z yasovchi affiklar bilan emas, balki lug’aviy shakl yasovchilar bilan ham yuz bergan.

Badiiy asar ijodiy jarayon mahsuli bo’lib, ham shaklan, ham mazmunan yangicha hosila ma’noni yaratish esa ikki karra ijodkorlik, izlanish, topqirlikni talab etadi. Bu jarayonda ijodkorlar til fondida tayyor, mavjud bo’lgan birliklardan foydalanishdan ko’ra o’zlarining badiiy-estetik idealiga mos keladigan so’zlarni yaratadilar. Maqsadli,

o'rinli, jo'yali "yaratmalar" o'quvchini o'ziga chorlaydi, unga estetik zavq bag'ishlaydi.

Muhammad Yusufning muallif nutqi neologizmlari qatorida murakkab komponentli birliklar salmoqli o'rin egallaydi. Masalan, *bulbulzabon*, *bulbulnavo* so'zlari nutqimizda, xalq tilida tez-tez qo'llanib turadigan qo'shma so'zlar sanaladi. "O'zbek tilining izohli lug'ati"da *bulbulnavo* so'zi o'z ifodasiga ega, biroq *bulbulzabon* so'zi lug'atda keltirilmaydi[9]. Shoir "Andijon" she'rida "*Bulbulzabonlaring* xazon yopdimi/ Dunyo ularning hech tengin topdimi" tarzidagi misralarni yaratadi. Bu bilan ifodaviylik, ta'sirchanlik, obrazlilik ta'minlanadi. Quyida ko'rib o'tiladigan qo'shma so'zlar haqida ham shu fikrni aytish mumkin.

*Ot+ot shaklidagi qo'shma otlar.*

Rayhon, aylanay rayhoningdan sening

*Oyqaboq* go'zalim, joningdan sening. ("Rayhon")

Misoldagi *oyqaboq* so'zi butunlay yangi sifatlash sanaladi. Bu so'z boshqa ijodkorlarda uchramaydi. Keyingi misol:

Qadim *O'zbekmomo* kuyi yodimda,

Yodimda chechaklar kulgan chambili.

She'rdagi *O'zbekmomo* shaxs oti o'zbek ayollarining o'tmishdagi umumlashma obrazi bo'lib, qo'shma so'z tarkibidagi *momo* so'zi "hurmatli, ko'pni ko'rgan, e'zoz-e'tiborli ayol" ma'no ifodasiga ega bo'lib kelgan.

She'riyat bu qadim buyuk so'qmoq ekan,

*Kiyikso'qmoq* ekan, *kiyikso'qmoq* ekan.

Shoir she'riyat yo'li mushkulligini, hamma ham bu yo'ldan yura olmasligini *kiyikso'qmoqqa* qiyoslashi bilan ifoda etadi. O'z o'rnida topib aytilgan ushbu obrazli tasvir to'laqonli estetik vazifa bajargan.

Muhabbatim, qaydasan, qalb ardog'im,

*Jayronko'zim*, qirlarga qolgan ohim.

Mazkur misrada *jayronko'z* individual nutq neologizmi muhabbat so'zining sifatlashi bo'lib kelgan.

Bedapoya to'ridagi juftin qidirib,

Sayray-sayray jimib qoldi bedana dilgir.

*Suvsumbulni* va'dasiga ko'mib ko'ndirib,

O'ynatgani olib ketdi irmoq qaygadir.

Boshqa bir misol:

Uyimizga bir kuni  
Qarindoshlar to'lishdi.  
*Chaqmoqtuyoq* toyimdan  
Ayirmoqchi bo'lishdi.

*Sifat+ot shaklidagi qo'shma otlar.*

Changimagin, tuproq yo'l,  
Bo'lg'in menga o'rtoq yo'l.  
*Yolg'izboshman, yolg'izqo'l,*  
Chopolmasam netarman. ("Netatman")

Tahlilga tortilgan misollardagi *oyqabog*, *o'zbekmomo*, *kiyikso'qmoq*, *jayronko'z*, *suvsumbul*, *chaqmoqtuyoq yolg'izbosh*, *yolg'izqo'l* kabilarning lug'atlarda, xususan, "O'zbek tilining izohli lug'ati"da[9] o'z ifodasiga ega emasligi ham ularning muallif neologizmlari bo'la olishi dalili bo'la oladi. Bu kabi *ot+ot va sifat+ot shaklidagi* yangi-yangi yasama so'zlar ot va sifat turkumiga oid.

Umuman, lingvopoetikaning o'rganish obyekti matn, yani badiiy matn sanalib, matnni tahlil qilish orqali shoir va yozuvchilarning badiiy mahorati, matn yaratishdagi iste'dodi, bunda til birliklari, xususan, nutqiy yasalmalar va so'z tabiatidagi turli-tuman yangi-yangi ma'no ifodalar, ijodkorlarning umumxalq tiliga munosabati o'rganiladi. Bu jarayonda tadqiqotimiz asosi bo'lgan Muhammad Yusuf she'riyati va uning tadqiqi ham xarakterli sanaladi.

#### ADABIYOTLAR:

1. Ҳожиёв А. Ўзбек тили сўз ясаши тизими. –Т.:“Ўқитувчи”, 2007.
2. Муҳиддинов М., Каримов С. Фафур Фуломнинг поэтик маҳорати. Монография. – Самарқанд, 2003.
3. Қосимова М. Бадий нутқ индивидуаллигининг лингвистик хусусиятлари (Тоғай Мурод асарлари асосида). НДА. – Т., 2007.
4. Қўнғуров У., Тихонов А. Ўзбек тилининг терс луғати. – С., 1968.
5. Тўхтасинова О. Ўзбек тилида лексик окказионализмлар ва уларнинг бадий-эстетик хусусиятлари. НДА. – Т., 2007.
6. Умирова С. Ўзбек шеъриятида лингвистик воситалар ва поэтик индивидуаллик. (У.Азим шеърияти мисолида). НД. – Самарқанд, 2019.
7. Худойберганова Д. Ўзбек тилидаги бадий матнларнинг антропоцентрик талқини. ДДА. – Т., 2009.
8. Юлдашев М. Бадий матннинг лингвопоэтик тадқиқи. ДД. – Т., 2009.

9. Ўзбек тилининг изоҳли луғати. 5-жилдли. – Т.: “Ўзбекистон миллий энц. си”. 2006–2008.

10. Yusupov D. Y. Shoir so‘z aytgani kelar dunyoga...(Shoira Halima Xudoyberdiyevaning hayot va ijod yo‘liga chizgilar). «Сўз санъати» халқаро журналі (ISSN: 2181-9297 DOI: 10.26739/2181-9297 www.tadqiqot.uz) – 2021, 1-сон, 4-жилд. – В.: 21–28. <https://tadqiqot.uz/index.php/article/view/377>

11. Шодиева Д. Муҳаммад Юсуф шеъряті лингвопозітасі. НДА. – Т., 2007.

УДК 81-11

## XALQ OG'ZAKI IJODIDA MAKTUBLAR

*Murotova Ismigul Bobomirzayevna**BuxDU*

ismigul@gmail.com

Ushbu maqolada badiiy adabiyotning muhim janrlaridan bo'lmish maktublarning xalq og'zaki ijodidagi o'rni va ahamiyati haqida atroflicha to'xtalilgan bo'lib, maktublarning xalq og'zaki ijodidagi jumladan, epik turga mansub bo'lgan dostonlar tarkibida qay tartibda, qanday badiiy kompozitsiyalardan foydalanilgani haqida fikr yuritilgan.

**Kalit so'zlar:** maktub, xat, xalq og'zaki ijodi, mayna, doston, genesis.

Maktub va xatlar kishilar o'rtasidagi axborot almashinuvi vazifasini o'tagan, bir so'z bilan aytganda, ommobop janrdir. Avvallari, maktub deb qog'ozga yoki boshqa vositalar (loy bo'lagi, pergament varog'i yoki qayin po'stlog'i) qo'lda yozilgan xabarni atashgan. XVII-XIX asrlar asosan, yozma maktublar asri bo'lgan o'qimishli kishilar muntazam ravishda xatlar qabul qilib, jo'natib turishgan. Tarixan harflarda qo'llaniladigan so'zlashuv so'z birikmasi so'roq va undov formulalari tizimi rivojlangan.(1)

Tarixda muhim ahamiyatga ega bo'lgan maktublar orqali jumladan, podshohlarning farmonlari, xalqaro munosabatlar, murojaatnomalar, kishilar o'rtasidagi dil izhorlar, axborotlar, muhim voqealar o'z egalariga yetkazilgan. Ushbu maktublarda badiiy til unsurlaridan mohirona foydalanilgan. Masalan miloddan avvalgi III asrda Markaziy Osiyoda hukmronlik qilgan Botir Tangriqutning Xitoy malikasiga yozgan maktublari va malikaning Tangriqutga yozgan javoblari ma'lum.

Maktubning ilk ko'rinishlari O'rta Osiyo tarixida yozuv paydo bo'lmasdan ilgari buyumlar vositasida ham ifodalangan. Bunga Misol sifatida skiflar shohi Iskandarning Eron shohi Doroga yuborgan buyumlar- kamon o'qi, baqa, sichqon va qushdan iborat o'ziga xos "maktub"ni eslash kifoya.(3)

Asosan, maktubning ilk ildizlarini, xalq og'zaki ijodida ham ko'rishimiz mumkin. Ayniqsa, epik turga mansub bo'lgan dostonlarda ko'proq uchratamiz. Masalan "Ravshan" dostonida maktubning yorqin namunasi ko'ramiz. Ravshan bandi bo'lganda uni qutqarish uchun Go'zal yori Zulxumor uning yurtiga maktub yozadi. Maktub quyidagi ko'rinishda: "Hasan mardning o'g'li, Xon dallining joni-di-

li, oti bek Ravshan, Shirvon mamlakatiga, Qoraxon viloyatiga kelib bandi bo'ldi. Qirq kun muhlat qilib, zindonga soldi. Shu qirq kunning o'n to'qqiz kuni o'tkanda shu xat bo'ldi, namozshomda bitildi. Tong otmay, Chambilbelga,

to'ramning eliga yetadi. Nomozi bomdod shu xabar bo'lsa, Ravshanning murabbiylari bo'lsa, kelsa, ayirib olsa. Shu xat tekkan-dan so'ng, yigirmanchi kunida Shirvon shahrining bozori, Bozorining o'rtasida dori, shu yerda Ravshanbekning osilib o'ladigan yeri. Agar Ravshanbekning Xesh- tabarlari, g'am-guzorlari, ichkuyarlari bor bo'lsa, kelib Ravshanbekning gavda-suyagini olib ketsin, deb xat bitildi". Asarda maktubning o'rni beqiyos. Ravshanning omon qolishi aynan shu maktub tufaylidir. Shu o'rinda "O'tkan kunlar" romanida-gi Kumushning jasoratini eslash o'rinli. Xullas, Zulxumor Maktub-ni bog'lab, issiqqa chidamli, sovuqqa bardoshli qilib tayyorlaydi va mayna qushi orqali manzilga jo'natadi. Maktub orqali biz Ravshanning islom diniga e'tiqod qilishini ham bilib olamiz. Dalil sifatida Zulxu-mor "Nomozi bomdod" degan jumlaning maktub tarkibida keltiradi. Dos-tondagi voqealarning ketma-ketligini ta'minlash aynan ushbu maktub asosida amalga oshirilgan. Sababi Shirvon eli Ravshanbekning elidan ancha uzoqda bir oy deganda yetib olish mumkin. Ammo Ravshanning osilishga yigirma kun qolgan edi. Demak, Zulxumorning xati masofani inobatga olingan holda doston tarkibiga kiritilgan. Maktub juda sodda va ravon tilda bitilgan. Xabarlarining ketma-ketligi uni o'qishda biron muammoga duch kelishga yo'l qo'ymaydi. Maktubning haqiqiyli-gi, aynan, qahramonimiz Ravshan haqida gap ketayotganini bildirish uchun (chunki Ravshanlar ko'p bo'lishi mumkin), otasining ,bobos-ining ismlarni ham keltiradi. Dostonda maktubni manzilga yetkazish mayna qushning zimmasida edi. Odatda, maktub va xatlarni tashish vazifasini kaptar qushlar amalga oshirgan. Dostondagi masofalarning uzoqligi hamda vaqtning kam qolganligi sababli mayna qushga ushbu vazifa yuklatilgan. Mayna qushning biologik tarifi bilan tanishadigan bo'lsak, ular issiq va sovuqqa o'ta chidamli va qolgan qushlarga qara-ganda juda tez uchganligi uchun dostonda o'rin egallagan. D ostonda ham yozma ham og'zaki maktub bor desak, mubolag'a bo'lmaydi. Sababi, Zulxumor maktubni mayna qushga bo'g'laganidan keyin ham, og'zaki Ravishda unga murojaat qilib xabar aytadi. Masalan, Zulxumor beka mayna qushini osmonga uchirib: "Jonivor! To'ramning xon Dalli enasiga ikki pari enasiga salom degin"- deb zor-zor yig'lab maynasiga aytgan so'zi:

Oqar ko'zdan qonli yoshim  
Cho'q savdoga qolgan boshim  
Arzim, eshit mayna qushim  
Bor, Chambilning xoni kelsin,  
Yovmutning beklari kelsin  
Xasim bo'lib tengi to'shim  
So'zim eshit ,mayna qushim Bor,  
Chambilning xoni kelsin  
Yovmitning erlari kelsin..... (4)

Zulxumorning maynaga og'zaki qilgan fig'onlari dostonida o'n to'rt banddan iborat. Ushbu og'zaki maktub orqali biz Zulxumorning Ravshanni qanchalar sevishini, Ravshanning jindondaligi unga juda qattiq ta'sir qilganligini bilib olamiz. Qolaversa, pok muhabbat har doim jasorat talab qilishiga amin bo'lamiz. Zulxumor shohning qizi bo'lsa ham hech narsadan tap tortmay ushbu maktublarni yozadi. Agarda dostonga maktub kiritilmagan bo'lsa edi, Ravshan qanday qilib qutqarilar edi.? Albatta, biron chorasi topilardi, ammo maktub orqali olgan taassurotni o'quvchilar ololmas edi. E'tibor bersak, maktub aynan Ravshanning ota-onasiga yozilgan. Bu orqali inson Har qanday vaziyatda birinchi bo'lib ota-onadan yordam oladi. Insonning eng yaqin do'sti uning ota-onasi ekanligiga ushbu naktub yaqqol misoldir. Yangi o'zbek adabiyotida ham xuddi shunga o'xshash misol bor. Masalan "O'tkan kunlar" romanidagi Otabekni ham otasining maktubi qutqarib qoladi. Kumush ham xuddi Zulxumor kabi Otabekning joniga oro kiradi. Ko'ryapmizki, asarlarda keltirilgan maktub va xatlar shunchaki kiritilmaydi. Ularning ma'lum bir o'ziga xos vazifasi bo'ladi.

Xalq og'zaki ijodining yana bir mahsuli bo'lgan "Kuntug'mish" dostonida ham maktub shaklidagi berilgan unsurlar mavjud. Ammo ushbu makubda yozma matndan emas, balki buyumlardan foydalanilgan. Masalan, Kuntug'mish va Xolbeka bir-birlarini tushida ko'rganidan keyin Xolbeka bitta sandiqqa o'zining rasmini chizdirib, sochiniing tolasidan qo'yib, Chiltonlar almashtirgan uzukni ham joylab sandiqni yaxshilab tanballab daryoga oqizib yuboradi. Bu yerda maktubni yetkazish vazifasini suv bajarmoqda. Yuqorida aytganimizday maktublilar asosan, xabar vazifasini bajaradi. Bu dostonida aynan yozma shaklda bo'lmasa ham, ammo buyumlar orqali xabar yetkazilgan. Shu maktubdan keyin Kuntug'mish Xolbekani izlab yo'lga tushadi. Nimaga Xolbeka yozma matndan foydalanmadi. Mening fikrimcha, ular bir-birini tushida ko'rdi va ular bir-birining yuzini taniy oladilar. Shuning uchun Xolbeka matnli maktubni xohlamadi. Xuddi shunday buyumlar orqa-

li o'zining fikr mulohazalarini bildirish Alisher Navoiyning "Saddi Iskandariy" dostonida ham bor. Doro bilan Iskandarning elchilar orqali muzokara olib borganida, Doro Iskandarni mot qilish uchun "Chavgon bilan go'ya va kunjut" berib yuboradi. O'zining fikrlarining bayonini shu buyumlarga yuklaydi. Asarlarda mana shunday o'xshashliklar uchrab turadi.

Dostonlar ichida o'zining syujetga boyligi bilan ajralib turadigan "Alpomish" dostonida ham qahramonlar o'rtasida maktub almashinish jarayonlari berilgan. Masalan, Barchin o'zining nochor axvolini bayon etib, Hakimbekdan yordam so'rab, unga maktub yozadi va bu nomasini o'nta navkar orqali yuboradi. Xalq og'zaki ijodida maktubning turli xil ko'rinishlari mavjud. Yuqoridagi misollar bunga yaqqol dalildir. Ayniqsa, dostonlardagi maktublar ayollar tomonidan yozilganligi yanada diqqatga sazovor. Xalq og'zaki ijodi o'zi adabiyotning boshlanishi desak ham bo'ladi. Shuning uchun maktublarning genezisini xalq og'zaki ijodida deb ko'rsatamiz.

Xulosa o'rnida shuni aytish mumkinki, ma'lum bir asarga kiritilgan xoh she'r, xoh hikoya, xoh maktub, xoh qo'shiq, vaholanki, oddiygina detal ham o'zining bajaradigan vazifasiga ega. Jumladan, maktublar ham asarni ochib beruvchi, yoki yakunlovchi, qahramonning ichki kechinmalarini, oshiq va mashuqalarning bir-birlariga qilgan dil izhorlarini kitobxonga yetkazishga xizmat qiladi. Maktublar asarlarda kimningdir hayotini qutqarib qolishi ham mumkin. Buni yuqorida bir necha misollar yordamida ko'rdik. Xalq og'zaki ijodi esa man shunday misollarga boydir. Xullas, badiiy adabiyotda maktublar o'zining munosib o'rniga ega.

#### ADABIYOTLAR:

1. Горбачева Е. Ю., Мигулина А. В., Епостолярный в историческом контексте. Молодой ученый. 2016 N28.1 (132.1) В. 6–8.
2. Abdurayimova Irodaxon Iqboljon qizi. Maktublar tarixi haqida 2023-yil
3. Alpomish- Toshkent. Ma'naviyat. 1987.
4. Yuldasheva D.N., Chullieva G. T. THE ROLE OF NOVERBAL MEANS IN SPEECH MEANING. International Engineering Journal For Research & Development, Vol.6, Issue 2 published May 16, 2021. – Pag.1956-1961. [www.iejrd.com](http://www.iejrd.com) <http://www.iejrd.com/index.php/%20/article/view/1956>.
- 5 Dostonlar aytuvchi Ergasg Jumanbulbul og'li; yozib oluvchilar Muammadisa Ernazar o'g'li, Hodi Zarif; mas'ul muharrir T.Mirzayev-T. "Sharq", 2011-yil.
6. Rahim Vohidov, Husniddin Eshonqulov. O'zbek mumtoz adabiyoti tarixi. Toshkent- 2006-yil.



УДК 81-11

**KONSEPT VA NOVERBAL FIKRLASH MUOMMOSI**

*Nazarova Nargiz*  
*Andijon davlat universiteti*  
*Andijon, O'zbekiston*  
nargiz@gmail.com

Ushbu maqolada fikrlashni ikkita asosiy turga: noverbal va verbal fikrlash turlarga bo'lib tadbiiq etilishi keng yoritiladi. Fikrlashni rivojlanish darajasiga qarab har xil turlarga bo'lish mumkinligi hamda noverbal fikrlash o'ziga xos xususiyatlarga ega ekanligi talqin etiladi. Konsept va noverbal fikrlash muommosini hal qilish haqida ham keng ma'lumot beradi.

**Kalit so'zlar:** konsept, kognitiv, lingvistik, visual, verbal, noverbal, ob'ekt, mexanizm, psixologiya.

**Ключевые слова:** концепт, когнитивный, лингвистический, визуальный, вербальный, невербальный, объект, механизм, психология.

**THE PROBLEM OF CONCEPTUAL AND NOVERBAL THINKING**

*Nargiz Nazarova*  
*Andijan State University*  
nargiz@gmail.com

**Abstract.** This article broadly describes the application of thinking into two main types: non-verbal and verbal thinking. It is interpreted that thinking can be divided into different types depending on the level of development, and that non-verbal thinking has its own characteristics. It also provides extensive information on solving the problem of concept and non-verbal thinking.

**Key words:** concept, cognitive, linguistic, visual, verbal, non-verbal, object, mechanism, psychology.

**ПРОБЛЕМА КОНЦЕПТУАЛЬНОГО И НОВЕРБАЛЬНОГО МЫШЛЕНИЯ**

*Наргиз Назарова*  
*Андижанский государственный университет*  
nargiz@gmail.com

В этой статье рассматривается применение мышления в двух основных типах: невербальное и вербальное мышление. Толкуется, что мышление можно разделить на разные типы в зависимости от уровня развития, невер-

бальное мышление имеет свои особенности. Он также предоставляет обширную информацию по решению проблемы концептуального и невербального мышления.

Fikrlashni ikkita asosiy turga bo'lish mumkin. Bular noverbal va verbal fikrlashdir. Noverbal fikrlash, xususan, vizual fikrlash, aqliy tasavvur va ob'ektlarning tasvirlariga tayanadi. Boshqa tomondan, verbal fikrlash mavhum belgi tuzilmalari bilan ishlaydi. Odamlarning aqliy mehnati haqida gap ketganda, ular turli xil imtiyozlarga ega ekanligi aniqlandi. Ba'zi odamlar vizual yoki xayoliy stimullarni talab qiladi. Masalan, ko'rilgan yoki tasavvur qilingan narsalar, boshqalari esa mavhum belgi va tushunchalar bilan ishlashni afzal ko'rishadi.

Olimlarning ta'kidlashicha ham, fikrlashni aqlning turli shakllari bilan bog'liq bo'lgan verbal va noverbal turlarga bo'lish mumkin. Verbal intellekt (aql-idrok) so'z yordamida bilimlarni tizimlashtirish va tahlil qilish ko'nikmasini hamda o'qish, yozish va og'zaki muloqot qobiliyatlarini o'z ichiga oladi. Bu verbal va mantiqiy fikrlash bilan bog'liq. Boshqa tomondan, noverbal intellekt ob'ekt va vizual tasvirlarni manipulyatsiya qilishni o'z ichiga oladi.

Fikrlashning bu ikki turi o'zaro bog'liq bo'lib, birining yetishmasligi boshqasiga ta'sir qilishi mumkin. Zaif verbal fikrlash ma'lumotni idrok etish va tushunishga to'sqinlik qilishi mumkin, noverbal fikrlashning past darajasi esa nutqning rivojlanishiga to'sqinlik qilishi mumkin. Masalan, agar inson ob'ektlarning shakli, o'lchami va nisbatlarini tanib olishda qiynalsa, ular vizual o'xshash harflarni farqlashda ham qiynalishi mumkin.

Noverbal fikrlash vizual stimullarga tayanadi. Ushbu jarayon tufayli odamlar tasvirlangan ob'ektlar orasidagi o'xshashlik va farqlarni to'g'ri aniqlashlari mumkin. Ob'ektlarning aqliy tasavvurlari va vizual-hissiy tasvirlarning shakllanishi idrok jarayonida sodir bo'ladi.

Fikrlash til bilan ham, tilsiz ham sodir bo'lishi mumkin, degan g'oya tarafdorlari fikrlashning ushbu har xil turlari yoki shakllariga murojaat qilishadi. Biroq, bu atamalar o'rtasida aniq o'xshatishlarni kuzatish ba'zan qiyin vazifa bo'lishi mumkin.

Umuman olganda, fikrlash muloqot kabi quyidagi shakllarni o'z ichiga olishi mumkinligi ta'kidlanadi: 1) verbal: fikrlashning bu turi og'zaki yoki mavhum shaklda amalga oshiriladi; 2) noverbal: fikrlashning bu shakli tasvir, diagramma va boshqa noverbal vositalardan foydalanishni o'z ichiga oladi.

Noverbal fikrlashga turli xil quyidagi hissiy shakllar kiradi: his qilish (1); idrok etish (2); tasavvur qilish (3).

His qilish deganda, miya ob'ekt va hodisalarning inson sezgilariga ta'siri asosida voqelikdagi xususiyatlarini aks ettiradigan aqliy jarayon tushuniladi.

Idrok o'ziga xos xususiyatlar to'plamini qamrab olgan ob'ekt yoki hodisani bir butun sifatida bevosita aks ettirishni o'z ichiga oladi. Bu ob'ekt yoki hodisaning faqat individual xususiyatlarini aks ettiradigan hissiyotdan farq qiladi.

Tasavvur qilish ilgari idrok etilgan ob'ekt yoki hodisalarni aks ettirishni, shuningdek, samarali tasavvur yordamida tasvirlarni yaratishni o'z ichiga oladi.

Fikrlashni rivojlanish darajasiga qarab har xil turlarga bo'lish mumkin: 1) vizual-samarali; 2) vizual-majoziy; 3) verbal-mantiqiy; 4) mavhum-mantiqiy.

Vizual-samarali fikrlash – bu noverbal fikrlash shakli bo'lib, muloqot jarayonida ob'ektlarni bevosita idrok etishga tayanadi. Bu fikrlashning eng asosiy shakli bo'lib, u yanada murakkab fikrlash turlari uchun asos bo'lib xizmat qiladi. Ushbu turdagi tafakkur bilan fikrlash jarayoni insonning jismoniy dunyoni idrok etishi bilan chambarchas bog'liq va uning ishtirokisiz unga erishib bo'lmaydi. U haqiqiy ob'ekt bilan amaliy mashqlarni o'z ichiga oladi.

Vizual-xayoliy fikrlash ham mental tasvir va obrazlar bilan ishlaydigan noverbal fikrlash turidir. Vizual-xayoliy fikrlash bilan insonning fikrlari uning atrofdagi voqelikni idrok etishi bilan bevosita bog'liqdir. Ushbu turdagi fikrlash insonning bevosita va ishchi xotirasi bilan chambarchas bog'liq bo'lib, u yerda kerakli tasvirlar saqlanadi. Aksincha, nazariy xayoliy fikrlash tasvirlarning uzoq muddatli xotirasiga tayanadi.

Verbal-mantiqiy fikrlash – bu tildan foydalanishga va tushunchalar bilan mantiqiy jarayonlarga tayanadigan fikrlash turi. Mantiqiy tushunchalardan foydalanib, odamlar o'zlari o'rganayotgan voqelikdagi muhim qonuniyat va kuzatilmaydigan munosabatlarni ajrata oladilar.

Mavhum-mantiqiy fikrlash, shuningdek, mavhum fikrlash deb ham ataladi. U verbal fikrlash shakli bo'lib, muhim bo'lmagan jihatlarni e'tiborsiz qoldirib, ob'ektning muhim xususiyatlari va aloqalarini aniqlashni o'z ichiga oladi.

Vizual-samarali, vizual-majoziy, og'zaki-mantiqiy va mavhum-mantiqiy fikrlash shaxsning hayot tsiklidagi fikrlash rivojlanishining ketma-ket bosqichlarini ifodalaydi.

Yu.S.Maslovning so'zlariga ko'ra, "noverbal fikrlashni real dunyo taassurotlarini idrok etish natijasida hosil bo'lgan, xotirada saqlanadi-

gan va tasavvur yordamida qayta tiklanadigan vizual- hissiy tasvirlardan foydalanish jarayoni deb ta'riflash mumkin. Noverbal fikrlash ba'zi hayvonlarda ma'lum darajada mavjud bo'lib, ularga atrof-muhitda harakat qilish va tegishli qarorlar qabul qilish imkonini beradi. Odamlar noverbal fikrlashni verbal fikrlash bilan birgalikda yuqori darajada rivojlantirdilar. U, asosan, texnik xarakterdagi ijodiy vazifalarni, masalan, mexanik qismlarning harakatini hal qilishda qo'llaniladi. Bunday muammolarni hal qilish, odatda ichki yoki tashqi nutq bilan bog'liq emas, aksincha "texnik" yoki "muhandislik" fikrlash deb nomlanuvchi maxsus tafakkur shakliga tayanadi".

Bir qator tilshunos va faylasuflarning fikriga ko'ra, til tushunchalarini aniq va izchil ajratib olishimizga imkon berishda hal qiluvchi rol o'ynaydi. Til yordamisiz bir tushunchani boshqasidan samarali ajratish qiyin bo'lar edi.

Noverbal fikrlash esa real dunyo taassurotlarini idrok etish natijasida hosil bo'lgan vizual-hissiy tasvirlardan foydalanishga tayanadi. Ushbu tasvirlar xotirada saqlanadi va tasavvur yordamida qayta yaratilishi mumkin. Noverbal fikrlash ba'zi hayvonlarda ma'lum darajada mavjud bo'lib, ularga atrof-muhitda harakat qilish va tegishli qarorlar qabul qilish qobiliyatini ta'minlashga xizmat qiladi.

Noverbal xatti-harakatlar bo'yicha tadqiqotlar fanning turli sohalarini qamrab olgan uzoq tarixga ega. Qadimgi yunon va rim yozuvlari biz noverbal fikrlash deb ataydigan narsalarni allaqachon muhokama qilgan. Jumladan, Kvintilianning milodiy I asrdagi "Nastavlenie oratoru" ("Ma'ruzachiga ko'rsatma") esesida odamlar tomonidan ishlatiladigan imo-ishoralar tilga olingan. O'shandan beri noverbal xatti-harakat va fikrlash hayvonlarning xulq-atvori, antropologiya, raqs, tilshunoslik, falsafa, psixiatriya, psixologiya, shuningdek, nutqni shakllantirish va idrok etish kabi turli sohalarda o'rganilmoqda. Tarix davomida ko'plab taniqli olimlar zamonaviy noverbal xulq-atvor fanining rivojlanishiga hissa qo'shdilar.

Noverbal fikrlash muammosi bu jarayonlar qanday sodir bo'lishi, qanday mexanizm va strategiyalar qo'llanilishi va ularning verbal fikrlash bilan o'zaro ta'siri haqida ma'lumot beradi.

Noverbal fikrlash o'ziga xos xususiyatlarga ega. Jumladan:

1. Vizual idrok. Noverbal fikrlash vizual idrok va tasvirlar bilan ishlash qobiliyatiga asoslangan. Noverbal fikrlashda kuchli odamlar odatda vizual ob'ekt va tasvirlarni osongina idrok etadilar va tahlil qiladilar.

2. Orientatsiya (yoʻnalish). Noverbal fikrlash obʼektlar orasidagi munosabatlarni aniqlashga imkon beradi.

3. Obrazlarni aniqlash. Noverbal fikrlash tasvir va obrazlarni tanib olishga yordam beradi. Noverbal fikrlash qobiliyatiga ega odamlar shakllar, obrazlarni tezda taniy oladilar, bu ularga tanib olish va tasni-flash bilan bogʻliq muammolarni osongina hal qilish imkonini beradi.

4. Hissiy idrok. Noverbal fikrlash ham hissiy idrok bilan bogʻliq. Noverbal fikrlash boshqa odamlarning hissiy holatini yaxshiroq tushunishga yordam beradi.

5. Ijodkorlik. Noverbal fikrlash ijodkorlik va intuitiv fikrlashni rivojlantiradi.

Noverbal fikrlash muammolarini hal qilish uchun hissiy intellekt va kommunikativ kompetensiya koʻnikmalarini rivojlantirish muhimdir. Bunga noverbal signallarni tanib olish va talqin qilish, shuningdek, turli kontekst va madaniy xususiyatlarga moslashish qobiliyati kiradi.

Noverbal fikrlash xatolarga duch kelishi mumkinligini ham hisobga olish kerak. Masalan, baʼzi odamlar beixtiyor notoʻgʻri noverbal signallarni uzatishi yoki ularni notoʻgʻri talqin qilishi mumkin. Shuning uchun tushunmovchilik ehtimoliga diqqatli boʻlish muhimdir.

Olimlar noverbal fikrlashni va uning boshqa odamlarning muloqotiga, shuningdek, tushunishiga taʼsirini oʻrganadilar. Ular imo-ishora, yuz ifodalari, ovozli intonatsiya kabi noverbal fikrlashning turli jihatlarini, shuningdek, noverbal fikrlash qanday qilib buzilish va xatolarga duchor boʻlishi mumkinligini ham tadqiq etmoqdalar.

Noverbal fikrlash boʻyicha tadqiqotlar boshqa odamlar bilan qanday munosabatda boʻlishni va muloqot qobiliyatini qanday yaxshilashni tushunishga yordam beradi. Shuningdek, ular psixologiya, sotsiologiya, shaxslararo munosabatlar va boshqa koʻplab sohalarda amaliy ahamiyatga ega boʻlishi mumkin.

Turli sohalardagi olim va tadqiqotchilar fikrlashning yangicha turiga qiziqib, oʻrganmoqdalar. Uning mexanizmlari, jarayonlarini va kognitiv funktsiyalarga hamda muammolarni hal qilishga taʼsiriga eʼtibor qaratmoqdalar. Jumladan, Xovard Gardner aqlning turli shakllari, shu jumladan, noverbal fikrlash gʻoyasini oʻz ichiga olgan koʻp intellekt nazariyasini ishlab chiqqan amerikalik psixologdir. Rodjer Sperry miyaning chap va oʻng yarmiga boʻlinishini kashf etgani uchun Nobel mukofotini olgan amerikalik nevrolog. Uning tadqiqotlari shuni koʻrsatdiki, miyaning har bir yarmi har xil fikrlash turlariga, shu jumladan, noverbal fikrlashga ixtisoslashgan. Jeyms Karsen esa vizual fikrlash va uning qaror qabul qilish hamda ijodiy fikrlashga taʼsirini

o'rgangan amerikalik neyrofiziolog. Rodjer N. Shepard bo'lsa, vizual idrok va fazoviy fikrlash sohasidagi tadqiqotlari bilan tanilgan amerikalik psixolog.

Shuningdek, nemis olimlari, noverbal fikrlash va uning kognitiv jarayonlarga ta'sirini faol o'rganmoqdalar. Ulardan ba'zilari ushbu tadqiqot sohasiga katta hissa qo'shgan. Masalan, noverbal fikrlash bilan shug'ullangan mashhur nemis olimlaridan biri Volfgang Kyolderdir. U noverbal kommunikatsiya va vizual fikrlash sohasida tadqiqotlar olib bordi.

Vizual idrok va vizual fikrlashni tadqiq qilishga katta hissa qo'shgan nemis psixologi Kurt Koffkani ham ta'kidlash joiz.

Ularning tadqiqotlari fikrlash jarayonlarini va til hamda tafakkur o'rtasidagi o'zaro ta'sirni yaxshiroq tushunishga yordam beradi.

Yuqoridagi fasllarda batafsil ta'kidlanganidek, konsept – bu ma'lum ob'ekt, hodisa yoki g'oyani tavsiflash va tasniflash uchun ishlatiladigan mavhum tushuncha. Tushuncha bizga ma'lumotni tartibga solish va tuzishda, shuningdek, turli elementlar o'rtasidagi munosabat va aloqani tushunishda yordam beradi.

Konsept oddiy yoki murakkab, mavhum yoki aniq bo'lishi mumkin. Ular jismoniy ob'ekt, jarayon, xususiyat, munosabatlarni tasvirlash uchun ishlatilishi mumkin. Konsept so'z, belgi, grafika, model shaklida taqdim etiladi.

Konsept insonning fikrlashi va muloqotida muhim rol o'ynaydi, chunki u fikr almashish, bilim olish va atrofda dunyoni tushunish imkonini beradi. Shuningdek, konsept fan, san'at, falsafa va boshqa bilim sohaslarini rivojlantirish uchun asosdir.

Umuman olganda, konsept ma'lumotni tartibga solish va tuzishga yordam beradi, bu esa atrofda dunyoni yaxshiroq tushunish va ular bilan muloqot qilish imkonini beradi.

“Noverbal fikrlash” tushunchasi va muammosi psixologiya hamda kognitiv fanlarda ikkita muhim jihatni ifodalaydi. Konsept ongimizdagi ma'lumotlarni tartibga solish va tasniflash uchun foydalanadigan bilimlar toifasidir. Konsept atrofda dunyoni tushunishga yordam beradi, ulardan ob'ekt, hodisa, g'oya va ular o'rtasidagi munosabatlarni aniqlash hamda tavsiflash uchun foydalaniladi. Tushuncha mavhum (masalan, adolat) yoki aniq (masalan, mushuk) bo'lishi mumkin. Ular til va tafakkur asosidir.

Konsept va noverbal fikrlash sohasidagi tadqiqotlar atrofda dunyoni qanday qabul qilish va tartibga solishni, shuningdek, boshqa odamlar bilan qanday munosabatda bo'lishni yaxshiroq tushunishga

yordam beradi. Ushbu mavzular kognitiv jarayonlarni anglash va inson aqlini rivojlantirish uchun muhimdir.

Noverbal fikrlash og'zaki yoki og'zaki ifoda ishlatmasdan ma'lumotni o'ylash va tushunish qobiliyati ekan, bu imo-ishora, yuz ifodalari, ovozli intonatsiya va boshqa noverbal iboralar kabi noverbal signallarni idrok etish, qayta ishlash va talqin qilishni o'z ichiga oladi.

Noverbal fikrlash tushunchasi psixologiya va kognitiv fanlar sohasida chuqur ildizlarga ega. U odamning noverbal signallar orqali ma'lumotni tushunish va uzatish qobiliyatini o'rganadi. Noverbal fikrlash muloqotda muhim rol o'ynaydi, chunki u har doim ham so'z bilan ifodalab bo'lmaydigan his-tuyg'ularni, niyatlarni va boshqa murakkab tushunchalarni yetkazishga imkon beradi.

Biroq, noverbal fikrlashning bunday ahamiyatiga qaramay, u muammo va tushunmovchiliklar manbai bo'lishi ham mumkin. Masalan, noverbal signallarni madaniy, ijtimoiy va individual xususiyatlariga qarab turli odamlar turlicha talqin qilishlari mumkin. Bu muloqotda tushunmovchilik va nizolarga olib keladi.

Noverbal fikrlash muammolarini hal qilish uchun turli xil texnika usullaridan ham foydalanish mumkin. Masalan, faol tinglash va aniqlovchi savollar berish noverbal signallarni aniqlashtirishga va tushunishni yaxshilashga yordam beradi. Noverbal signallarga va boshqa odamlarning hissiy holatlariga nisbatan sezgir bo'lish uchun hissiy va ijtimoiy intellektni rivojlantirish ham foydalidir.

Umuman olganda, noverbal fikrlash boshqa odamlar bilan muloqot qilish va tushunishning muhim jihati hisoblanadi. Bu so'z bilan ifodalab bo'lmaydigan ma'lumotlarni uzatish va idrok etish imkonini beradi. Biroq, noverbal fikrlashdan muvaffaqiyatli foydalanish uchun bor mahoratni rivojlantirish va yuzaga kelishi mumkin bo'lgan muammolar hamda tushunmovchiliklarga tayyor bo'lish kerak.

#### ADABIYOTLAR:

1. [https://studbooks.net/903136/psihologiya/ponyatie\\_osobennosti\\_neverbalnogo\\_myshleniya\\_d\\_etey\\_patologii\\_rechi\\_detey\\_obschim\\_nedorazvitiem\\_rechi](https://studbooks.net/903136/psihologiya/ponyatie_osobennosti_neverbalnogo_myshleniya_d_etey_patologii_rechi_detey_obschim_nedorazvitiem_rechi)
2. [deepcloud.ru/articles/chto-takoe-neverbalnoe-myshlenie](https://deepcloud.ru/articles/chto-takoe-neverbalnoe-myshlenie)
3. <https://vplate.ru/myshlenie/verbalnoe-i-neverbalnoe/>
4. <https://deepcloud.ru/articles/chto-takoe-neverbalnoe-myshlenie/>
5. Grechko P.K. Sotsialnaya realnost: kontseptsii i metodologiya issledovaniy. – M.: Institut filosofii RAN, 2008. – 500 s.

6. [https://studbooks.net/903136/psihologiya/ponyatie\\_osobnosti\\_neverbalnogo\\_myshleniya\\_d\\_etey\\_patologii\\_rechi\\_detey\\_obschim\\_nedorazvitiem\\_rechi](https://studbooks.net/903136/psihologiya/ponyatie_osobnosti_neverbalnogo_myshleniya_d_etey_patologii_rechi_detey_obschim_nedorazvitiem_rechi)

7. Maslov Ÿu.S. Vvedenie v yazikoznanie: Uchebnik dlya filologicheskix spetsialnostey vuzov. – M.: Visshaya shkola, 1987. – 300 s. – S. 97–101.

8. Girutskiy A.A. Vvedenie v yazikoznanie: Ucheb. posobie. Mn.: TetraSistems, 2001; 2-e izd., 2003; 3-e izd., 2005; 4-e izd., dop. 2008. – 288 s.

9. Feldman R.S. (Ed.) Applications of nonverbal behavioral theories and research. Hills-dale, NJ: Erlbaum, 1992. – 322 p.

10. Nepp M., Xoll D. Neverbalnoe obshenie: poln. ruk. – SPb.: Praym-EVROZNAK, 2006. – 512 s.



УДК 81-11

**XALQ SHE'RIYATIDA GUL SO'ZI ASOSIDA OBRAZLILIKNING  
VUJUDGA KELITIRILISHI*****O'rayeva D. S.<sup>1</sup>, Teshayeva M. X.<sup>2</sup>****<sup>1</sup>BuxDU professori, f.f.d.**Buxoro, O'zbekiston**<sup>2</sup>BuxDU magistranti**darmonoy@gmail.com*

Dunyodagi barcha xalqlar adabiyotida gul turlari inson tasviri va xususiyatlari, ichki kechinmalarini ifodalovchi an'anaviy poetik obrazlardan biri bo'lib, muayyan badiiy-stetik vazifalarda namoyon bo'ladi. Bunda gulning xalq millay qarashlari, madaniyati, yashash sharoiti bilan bog'liqlikda talqin etilishi kuzatiladi. Shaxs tashqi ko'rinishi, his-tuyg'ulari, xarakter- xususiyatlari tasvirida gul obrazining qo'llanilishi qo'shiqlar ta'sirchanligini yuzaga keltiruvchi muhim omillardan biridir.

Xalqning o'zi yashab turgan makondagi flora va fauna olami manzarasiga oid yiqqan bilimlari asosida gullar obrazi yaratilar ekan, bunda insonlarning tabiat – jamiyat – inson omili o'zaro bog'liqligi haqidagi tasavvur-tushunchalari yuzaga chiqadi. Gullar olamining badiiy adabiyotda obrazlantirilishi o'ziga xos mavqega egaligi ravshanlashadi.

Gullar xalq og'zaki ijodi, jumladan, xalq lirik hamda mavsumiy va oilaviy-maishiy marosim qo'shiqlaridagi an'anaviy poetik obrazlardan biridir. [2] Qo'shiqlarda ular yo turg'un, yo vaziyat ramzi sifatida keltirilishi ko'rinadi. Gullar lirik qahramonning ichki kechinmalari, turli ruhiy holatlarini badiiy fonda ochib berish bilan birga peyzaj, ideal shaxs tasvirini yaratish, lirik qahramon qiyofasini gavdalantirish, inson xususiyatlarini tavsiflash uchun ham keng qo'llanadi.

Gullarning poetik obraz darajasiga ko'tarilgani, ularning inson obrazini berishda, shaxs sifatlarini poetik ifodalash va baholashda keng qo'llanganiga guvoh bo'lish mumkin. Ularning xalq og'zaki badiiy ijodida obrazlantirilishi, turli motivatsion ko'rinishlarda ifoda etilishi e'tiborni tortadi.

Gullar obrazining ramziy, timsoliy, metaforik, sifatlovchilik ma'no va vazifalarida qo'llanishi, genezisining qadimgi totemistik miflarga borib taqalishi, murakkab tarixiy-folkloriy jarayonni bosib o'tganligidan dalolat beradi. Keyinchalik ularga ijtimoiy talqin ham berilgan.

Gullar xalq she'riyatida ko'pincha o'xshatish bo'lib kelib, insonning ko'rinishini, ruhiy holatini yoki ma'lum bir davrdagi vaziyatini, tana a'zolarini, umrini qaysidir bir gulga mengash asosida yuz beradi.

Xalq orasida “guldek umr”, “gul yuzli yor”, “qo‘li gul”, “gul yorim”, “gulday bolam”, “nargis ko‘z” birikmalari bot-bot quloqqa chalinadi. [4]

Nikoh to‘yiga aloqador “Xush kelibsiz” turkumidagi qo‘shiqlarda kelinchak ta‘rif-tavsifi uchun gullar obrazidan juda keng qo‘llanilishi kuzatiladi:

O‘zi bir gul, xush kelibsiz,  
 Sochi sumbul, xush kelibsiz,  
 Guli ra‘no, xush kelibsiz.  
 Gulshan aro, xush kelibsiz.  
 Payvasta gul, xush kelibsiz.  
 Shikasta gul, xush kelibsiz.  
 Raftori gul, xush kelibsiz,  
 Guftori gul, xush kelibsiz.  
 Gulchambarak, xush kelibsiz  
 Gul shaydosi, xush kelibsiz.  
 Gul chamanlig‘, xush kelibsiz,  
 Gulxonalig‘, xush kelibsiz.  
 Gul bargi tar, xush kelibsiz.  
 Guli bahor, xush kelibsiz,  
 Olam gulzor, xush kelibsiz,  
 G‘uncha dahan, xush kelibsiz.  
 Guli rayhon, xush kelibsiz.  
 Gul bahori, xush kelibsiz.  
 Guli na‘no, xush kelibsiz. [1]

Ko‘rinadiki, gullar obrazi ko‘pincha badiiy adabiyotda inson tasviri va xususiyatlarini, shaxs go‘zalligini yoki biror-bir fazilatini poetik ifodalash uchun ishlatilishi kuzatiladi. Bunda gullarning turli nomlaridan foydalaniladi.

Gullar insonning turli ruhiy holatlari, sifatleri, ichki kechinmalarini o‘tmish va bugun (diaxron- sinxron), jinslararo farq (gender) nuqtai nazaridan ham talqin qilishda o‘ziga xoslikka ega.

Mening qizim gulchehra,  
 Qizil gulingni asra. [1]

Ushbu qo‘shiqdagi “gulchehra” so‘zi “gul yuzli” birikmasiga sinonim bo‘lib, asosan, xotin- qizlarga nisbatan ishlatiladi.

Gullar bilan bog‘liq atoqli ismlar (antproponimlar) ham xalq qo‘shiqlarida uchraydi:

Kulib, kulib gapirganning  
 Oti Anor, yor-yor.  
 Anorgulning har soʻzidan  
 Bollar tomar, yor-yor. [1]

Oʻzbek xalqi orasida gulga eʼtiqod bilan bogʻliq holda shakllangan “Qizil gul sayli”. “Lola sayli”, “Gul bazmi” singari marosimu urf-odatlar, gul bayramlari ham borki [3], ularning kelib chiqishida totemistik miflarning alohida taʼsiri seziladi. Gullarning marosimiy yoki nomarosimiy qoʻshiqlarda ramziy-timsoliy obraz darajasida namoyon boʻlishi muayyan badiiy evolyusiyasiga erishgan. “Gul bazmi” qoʻshiqlari ham yozib olingan. Ularning birida shunday kuylanadi:

Gul berdi, koʻngil berdi,  
 “Gul berganni sev”, – derdi.  
 Yurib bogʻda u goʻzal,  
 Menga guldasta terdi.  
 Oldiniga uyaldim,  
 Qiz boqardi qoʻlimga.  
 Oʻzi kabi guli ham  
 Yoqib qoldi koʻnglimga.

Oʻzbek xalq sheʼriyatida shaxs tasviri va sifatlari, ruhiy kechinmalarining gullar vositasida poetik ifodalanishiga xos xususiyatlar ajdodlarimizning gullarni obrazlantirish badiiy-estetik tajribasi anʼanaviylikka egaligidan dalolat beradi. Bunda shaxs tasviri va xususiyatlarini gullar orqali ochib berish, sifatlash muayyan badiiy-estetik tajriba darajasiga koʻtarilgani kuzatiladi.

Gullarni poetik obraz sifatida obrazlantirishda ularning har biri inson tana aʼzolarini (gulbadan, gulyuz, gʻuncha lab, sunbul soch, nargis koʻz, qoʻli gul va hokazo kabi) tavsiflash, taʼriflash uchun oʻziga xos ramziy-timsoliy maʼnolarda, badiiy-estetik vazifalarda keltirilishi namoyon boʻladi.

Kulol yigit qoʻli gul,  
 Gʻazalxon goʻyo bulbul.  
 Hunarmandlar toʻplaydi,  
 Bogʻbon qoʻlida jambul.

Gullarning obrazlantirilishi misolida kishilarning aqlu zakovati, estetik didi kuzatiladi. Chunki ularni obrazlantirish dastlab epik asar-

larda (asotir, afsona, rivoyat, ertak, dostonlarda), undan soʻng esa xalq qoʻshiqdarida, anʼanaviy va zamonaviy yozma sheʼriyatda ham yuz bergan.

#### ADABIYOTLAR:

1. Бухор элда гул сайли. Бухоро халқ кўшиқлари ва термалари. Тўплаб нашрга тайёрловчи ва кириш сўз муаллифлари Д.Ўраева, Д.Ражабов. – Т.: Мухаррир, 2010. – 188 б.
2. Мусакулов А. Ўзбек халқ лирикаси. – Т., Фан, 2010. – 307 б.
3. Сафаров О., Маҳмудов М. Оила маънавияти.–Т.: Маънавият, 1998. – 110 б.
4. Пазлитдинова Н.З. Ўзбек тили фитонимларининг лексик-семантик хусусиятлари: Филол. фанлари бўйича фалсафа доктори (PhD) дисс. автореф. – Фарғона, 2018. – 53 б.

УДК 81-11

**BOLALAR SHE'RIYATIDA JONIVORLAR OBRAZINING  
BADIY TALQINI (ANVAR OBIDJON SHE'RLARI MISOLIDA)**

*Ochilova Nozima*

*BuxDU*

nozima@gmail.com

Maqolada Anvar Obidjon ijodining o'ziga xosligi to'g'risida so'z yuritilgan. Sho'rning badiiy mahorat qirralari, bolalarga bag'ishlangan she'rlarida jonivorlar obrazining badiiy talqin etilishi masalasi o'rganilgan.

**Kalit so'zlar:** bolalar adabiyoti, she'riyat, Anvar Obidjon ijodi, ijodiy niyat, badiiy mahorat, hayvonlar obrazi, obrazlilik.

**ARTISTIC INTERPRETATION OF THE IMAGE OF ANIMALS IN  
CHILDREN'S POETRY  
(EXAMPLE OF ANVAR OBIDJAN'S POEMS)**

*Ochilova Nozima*

*BuxSU*

nozima@gmail.com

The article talks about the uniqueness of Anvar Abidjon's work. The artistic skills of the poet, the issue of artistic interpretation of the image of animals is studied in his poems dedicated to children.

**Key words:** children's literature, poetry, work of Anvar Obidjon, creative intention, artistic skill, image of animals, imagery.

**ХУДОЖЕСТВЕННАЯ ИНТЕРПРЕТАЦИЯ ОБРАЗА  
ЖИВОТНЫХ В ДЕТСКОЙ ПОЭЗИИ  
(НА ПРИМЕРЕ СТИХОВ АНВАРА ОБИДЖАНА)**

*Очилова Нозима*

*БухДУ*

nozima@gmail.com

В статье говорится об уникальности творчества Анвара Абиджона. Художественное мастерство поэта, вопрос художественной интерпретации образа животных изучаются в его стихах, посвященных детям.

**Ключевые слова:** детская литература, поэзия, творчество Анвара Обиджона, творческий замысел, художественное мастерство, образ животных, образность.

Hayvonlar va jonivorlar haqida yozilgan barcha she'rlar va asarlar katta-yu kichikka birdek yoqadi. Xalqimiz azal-azaldan bunday ijod namunalari sevib mutolaa qilib kelmoqda.

Badiiy ijod namunasi o'z davrining voqea va holatlarini ko'rsatadi. Har bir hayvon majoziy suratda qahramonni aks ettiradi. Jumladan, Ezop masallari, Alisher Navoiy dostonlarining bir necha epizodlari, "Sher bilan Durroj", mashhur masalnavis Gulhaniyning "Maymun bilan Najjor" masali, Nasafiyning "Bahoriyat" ("Hayvonotnoma") asarlari. Ushbu asarlar bolalarbop, ular insonlarning yaxshi xislatlarini tarbiyalashga, ularda do'stlik, rostgo'ylik kabi fazilatlarni shakllantirishga xizmat qiladi. Bundan ko'rinib turibdiki, hayvonlar obrazi tasvirlangan asarlar majoziyliigi, tarbiyaviy ahamiyatga egaligi bilan bugunga qadar xalq og'zaki ijodi, shuningdek, buyuk shoir va yozuvchilarimiz tajribalari orqali shakllanib, sayqallanib kelingan.

Bugungi kunda ham bu an'ana o'z ahamiyatini yo'qotgani yo'q. Zamonaviy o'zbek she'riyatining taniqli namoyondalari Po'lat Mo'min, Shukur Sa'dulla, Orif To'xtash, Ma'mur Qahhor kabi ijodkorlar tomonidan olib kirilgan ilg'or an'analar shor, nosir, bolalar adabiyotining taniqli vakili Anvar Obidjon ijodida izchillik bilan davom ettirildi. Chindan ham "Anvar Obidjon o'zbek bolalar adabiyotida o'ziga xos maktab yaratgan shoirlardan biri hisoblanadi. Bu ko'proq uning turli buyumlar, jonzotlar, o'simliklar haqida yaratgan asarlarida ko'zga tashlanadi" [2;209].

Anvar Obidjon she'riyatida turli-tuman hayvonlar tasvirini uchratish mumkin. Bularidan, tulki, bo'ri, xo'roz, baqa, sigir, qarg'a va boshqa hayvon hamda parrandalarning salbiy, ijobiy xislatlarini shoir majoziy ma'noda o'xshatishlar orqali ko'rsatib bergan.

Anvar Obidjon "Xo'roz qo'shig'i" she'ri orqali ayrim insonlarning mansabparastligi, ularning o'z amllari ortidan maqtanib yurishlari xo'rozning yurishi va uning boshidagi toji tasviri orqali gavdalantirib bergan.

Men ketyapman qaddim g'oz,  
 Ququq-qu.  
 Chetroq turing, iltimos,  
 Ququq-qu.  
 Axir, shohman-hukmdor,  
 Ququq-qu.  
 Boshda oltin tojim bor,  
 Ququq-qu.

Oramizda shunday insonlar borki, ular hech narsani o‘z kuchlari va mehnatlari bilan uddalay olmasalar ham boshlaridagi “Oltin toj” orqali eplay olishlariga ishonadilar.

Tepakalman, sochim yo‘q,  
Ququq-qu.  
Toj kiymay ilojim yo‘q,  
Ququq-qu.

Inson hamisha jamiki orzulariga o‘z kuchi va mehnati orqali erisha, uning kelajagi, albatta, porloq bo‘ladi. Shunday ekan maqtanchoqlik yomon odat ekanligi Xo‘rozvoy orqali mohirona tarzda ko‘rsatib berilgan. Qolaversa, shoir xo‘rozning nuqsonlari va kaltaahmligini ham ifodalab bergan.

“Dastlabki muvaffaqiyat” she’rida ham maqtanchoqlik aks ettirilgan. Xulosani esa shoir o‘quvchining o‘ziga qo‘yib bergan.

Qarang Qo‘ng‘iz  
Uchganmish Qaytib  
Yerga tushganmish.  
Gerdayvolib so‘ng debdi:  
– Samalyoting  
Kim bo‘pti!

Anvar Obidjon har bir hasharot, har bir hayvonning kichik belgilarigacha e’tibor beradi. Haqiqatan ham qo‘ng‘izning uchishi samalyotga o‘xshaydi. Uchgan paytida esa samalyotga o‘xshab ovoz chiqaradi. Adib bolalarga maqtanchoqlik qilmaslikni, har bir muvaffaqiyatni kamtarona qarshi olish kerakligini, atrofdagilarini esa hurmat qilish kerakligini Qo‘ng‘izvoy orqali yetkazmoqchi bo‘ladi. Ya’ni, hali endi uchishni o‘rgangan Qo‘ng‘izvoyning dastlabki parvozi muvaffaqiyatli o‘tgani uchun atrofdagilarni mensimasdan maqtanchoqlik qilishni boshlaydi. Inson har doim ham yutuqlarga birdek erishavermaydi. Balandlarga ko‘tarilgan har qanday inson, yerga tushishi kerakligini ham unutmastligi kerak. Maqtanchoqlik qilib hech qachon do‘st orttirib bo‘lmaydi. O‘quvchi she’rni o‘qir ekan, kerakli xulosani Qo‘ng‘izvoyning xatti-harakati orqali o‘zi chiqarishga harakat qiladi. Maqtanchoqlik xususiyati nafqat bolalarda, balki, kattalarda ham uchraydi. Inson esa bunday yomon illatlarga qarshi kurashmog‘i lozim.

Adib “Sichqonlar qo‘shig‘i” she’ri orqali sichqonlar zararkunanda kemiruvchi jonivor bo‘lsa-da, ular mushukning taomi ekanliklarini yaxshi bilishlari mohirona, quvnoq, hazilga boy tarzda tasvirlab bergan.

Devorlarni yemiramiz,  
 Chi-chi-chiy.  
 Sandiqlarni kemiramiz,  
 Chi-chi-chiy.  
 Bug‘doylarni o‘maramiz,  
 Chi-chi-chiy.  
 Mushuk uchun, semiramiz  
 Viy-viy-viy!

Sichqonlar barcha narsani kemirib, olib qochishini barchamiz yaxshi bilamiz. Shoir sichqonlar orqali o‘g‘rilikning oqibati qanday bo‘lishini va uning yomon odat ekanligini, bunday yomon illatlarning jazosi borligini mushuk orqali bolalarga sho‘x, quvnoq tarzda yetkazib bergan. Demak, azaldan o‘g‘rilik jazosiz qolmagan va bunga yomon illat sifatida qarab kelingan.

Anvar Obidjonning “Tekinxo‘r” she‘rida insonga xos yana bir yomon illat Qumursqavoy va Chumchuq orqali mohirlik bilan she‘riy tasvirlab berilgan.

Qumursqavoy  
 Topib uvoq,  
 Terga botib  
 Ketdi uzoq.  
 Chumchuq kelib  
 Sekingina,  
 Yutdi-qo‘ydi  
 Tekingina.

Adib bolalarga mehnatkash va tekinox‘r insonlarning xarakter-xususiyatini Qumursqavoy va Chumchuq orqali qiziqarli tarzda she‘riy tarzda gavalantirgan. Qumursqa tabiatan juda sekin harakat qiladi va shu bilan birga juda nimjon. Chumchuq esa qumursqaga nisbatan ancha chaqqon hamda parvoz qila oladi. Yuqoridan turib pastdagi ozuqasini oson topa oladi. Demak, chumchuq mehnat qilsa, qumursqa topgan ozuqadan yaxshiroq o‘ljaga ega bo‘lishi mumkin. Ammo u mehnatni emas, yalqovlikni, ishlamay ovqat topishni afzal ko‘rdi. Shunday ekan inson hech qachon o‘zidan zaif kishilarga ozor bermasligi, aksincha ulardan o‘z yordamini ayamasligi lozim. Har bir bola she‘r bilan tanishgan paytda, albatta, o‘z-o‘zidan Chumchuqning ishi to‘g‘ri emasligini anglaydi va yaxshi- yomonni o‘zi ajrata oladi. Xalqimizda shunday naql bor: “Mehnat qilib topganing, qand-u asal totganing”.



Anvar Obidjon – serqirra ijodkor. U serqirra mavzularga, insonlarning turli xususiyatlariga alohida to‘xtalib o‘tadi. Bu esa uning hayvonlar haqidagi she‘rlari orqali yaqqol ko‘rinib turibdi. Uning she‘rlarida kichkina ertak kabi voqeabandlik mujassam. Ijodkor har bir hayvonning ko‘rinishi, tabiati orqali insonning xarakter-xususiyatini to‘g‘ri va aniq ko‘rsata olgan. Bolalarga insondagi yaxshi va yomon xislatlarni ko‘rsatishda chiroyli tasvirlardan mohirlik bilan foydalangan. Mahoratli ijodkor, iliq, yorug‘, go‘zal narsalarni yaxshi ko‘radi va ijobiy sifatlarga ega etib tasvirlaydi. Bolalarga ham adib o‘zining bolalarcha qalbidagi quvonchni, beg‘uborlikni takrorlanmas ijodi orqali ulashadi. Anvar Obidjon o‘z she‘rlari orqali o‘quvchilarga tarbiyaning yuksak cho‘qqilarini ko‘rsata oladi.

#### ADABIYOTLAR:

1. Adashboyev T., Abdurahmon A. O‘zbek bolalar adabiyoti antologiyasi. I jild. She‘riyat. – Toshkent: O‘qituvchi NMIU, 2009. – 288 b.
2. Jumaboyev. M. Bolalar adabiyoti va folklor. Darslik. O‘zbekiston Yozuvchilar uyushmasi Adabiyot jamg‘armasi nashriyoti, 2006. – 216 b.
3. Юсупова Д.Ю. Ҳ.Худойбердиеванинг сўзларни кўчма маънода қўллаш маҳорати//НамДУ илмий ахборотномаси – Научный вестник НамГУ 2021 йил 6-сон. – 395–402- б.
4. Obidjon A. O‘g‘irlangan pahlavon. – Toshkent: Cho‘lpon nomidagi nashriyoti, 2006. – 136 b.

## УДК 81-11

## RAUF PARFI IJODIDA HAYVONOT RAMZLARI

*O'roqova Nafosat Yoriyevna**Buxoro davlat universiteti*

n.y.uroqova@buxdu.uz

Ushbu maqolada XX asr o'zbek adabiyotining yorqin namoyandalaridan, sohir she'riyat olamining taniqli vakillaridan biri Rauf Parfi – O'zTurk qalamiga mansub hayvon obrazlari aks etgan qator she'rlarning muhim xususiyatlari masalasida so'z boradi. Rauf Parfi she'riyatining asosiy mavzusi bo'lmish erk tushunchasini ifodalashda hayvon obrazlaridan foydalanishi shoir uslubini ifodalovchi vosita sifatida ko'zga tashlanadi.

**Kalit so'zlar:** Rauf Parfi, XX asr o'zbek adabiyoti, shaxs erki, ko'chim, ramz, hayvon ramzlari, badiiy tasvir vositalari

## ANIMAL SYMBOLS IN THE CREATIONS OF RAUF PARFI

*Uroqova Nafosat Yorievna**Bukhara State University*

n.y.uroqova@buxdu.uz

This article deals with the important features of a series of poems with animal images written by Rauf Parfi, one of the prominent representatives of Uzbek literature of the 20th century, one of the famous representatives of the world of poetic poetry. The main theme of Rauf Parfi's poetry is the use of animal images to express the concept of manhood as a means of expressing the poet's style.

**Key words:** Rauf Parfi, Uzbek literature of the 20th century, personal freedom, migration, symbol, animal symbols, artistic image tools.

## ЖИВОТНЫЕ СИМВОЛЫ В ТВОРЕНИЯХ РАУФ ПАРФИ

*Н. Ё. Урокова**Бухарский государственный университет*

n.y.uroqova@buxdu.uz

В данной статье рассматриваются важные особенности цикла стихотворений с образами животных, написанных Рауфом Парфи, одним из ярких представителей узбекской литературы XX века, одним из известных представителей мировой поэтической поэзии. В качестве средства выражения стиля поэта выделяется использование образов животных для выражения концепции мужественности, которая является основной темой поэзии Рауфа Парфи.

**Ключевые слова:** Рауф Парфи, узбекская литература XX века, свобода личности, миграция, символ, животная символика, средства художественного изображения.

## KIRISH

Butun Turk dunyosining, qolaversa, zamonaviy o‘zbek adabiyotining yirik vakili, shoir va tarjimon Rauf Parfi adabiy merosini o‘rganish va targ‘ib qilish, uning asarlari misolida adabiyotshunoslikning dolzarb muammolarini tahlil qilish bugungi kun adabiyotshunosligi oldida turgan dolzarb vazifalardan biridir. Shoir ijodi XX asrning 70-yillaridan boshlab hozirgi kungacha ham kitobxonlar e‘tiboridan tushmay kelmoqda. Yangi o‘zbek adabiyotini Rauf Parfi she‘riyatisiz tasavvur etib bo‘lmaydi.

## ASOSIY QISM

Har bir davr adabiyoti, xususan, she‘riyatining nozik qirralarini ifodalovchi xususiyati, uslubiy yo‘nalishining she‘riy vaznlarga, janrlarga xos munosabati namoyon bo‘ladi. Bu jarayonda ramz ham she‘riyatga ta‘sirini o‘tkazadi. Barcha davrlarda individual xususiyatdan kelib chiqib ijod qilinishini inkor qilib bo‘lmaydi. Zamonaviy o‘zbek she‘riyati ham mana shunday jarayonni boshidan kechirdi va bu jarayon davom etmoqda. Rauf Parfi ijodi ko‘chim va metaforalarga, ramziy obrazlarga boy she‘riyat hisoblanadi. Ushbu kichik tadqiqotimiz doirasida uning hayvon obrazlari qo‘llangan she‘rlarini o‘rganishni maqsad qildik.

Ma‘lumki, yillar davomida qayta o‘zlashtirilgan va yaxlitlangan simvol va ramzlarga yangicha yondashuvlar paydo bo‘ldi. Ijodkorlar ushbu shakllar, simvollar orqali mavzu va g‘oyani ochib berish yo‘lidan yurdilar. Asarning badiiy-estetik qimmatini, ta‘sir kuchini oshirishga xizmat qiladigan vositalardan biri bu – ramzdir. Sharq folklori va adabiyoti namunalarida ramzning turli shakllari ko‘plab uchraydi. Rauf Parfining “Ostonamda yotar do‘stimning iti” deb nomlanuvchi she‘ri bu jihatdan e‘tiborga molik. Bunda shoir bir band, ya‘ni to‘rt qatordan tashkil topgan kichkina hajmli she‘rga bir olam ma‘no singirdira olgan:

Ostonamda yotar do‘stimning iti  
Yulduzlarga qarab,  
“Eh, o‘limgur do‘stim-ey.  
Marhum  
Do‘stimning iti [3, 68].

“Ostonamda yotar do‘stimning iti” deb boshlanuvchi ushbu she‘rda it uy hayvoni obraz darajasiga ko‘tarilgan. U sodiqlik, vafodorlik ramz-

ini ifodalab kelgan. Qadimdan it xuddi shu sheʼrdagi kabi vafodorlik ramzi sifatida qoʻllanib keladi. Bunda R.Parfi anʼanalarga tayangan holda itni doʻst tushunchasi bilan teng koʻradi.

Koʻchalarda yurgayman sarson,  
Adashgan it singari uvol.  
Goʻyo kimsa hol soʻrmas mendan  
Joʻralarim soʻramaslar hol [3, 138].

“Mana shundoq kechar kunlarim” sheʼrida kelgan ushbu bandda ham it obraz darajasiga chiqarilgan. Bunda it roʻshnolik koʻrmagan, xoʻrlangan, qadr topmagan kimsalar ramzini ifodalagan. “Adashgan it” iborasini shoir Furqat ijodida anʼanaviy hol tarzida uchratishimiz mumkin.

Zamonaviy sheʼriyatdagi ramzlar misrada yuzaga kelishi, fikrni toʻla ochib berishga xizmat qilishi va jarayonga aloqasi nuqtayi nazari-dan alohidalik hosil qiladi. Ramz barcha sohalarida uchrashi mumkin, uning sheʼriyatga singishida ham shoirning mahorati jiddiy oʻrin egal-laydi. Poetik ramzlar tashqi koʻrinish jihatidan intim tuygʻularni bel-gilar orqali ifodalashga xizmat qiladi. Shu bilan bir qatorda, ramz oʻzi ifodalayotgan tushunchani aniq va badiiylik meʼyorlariga amal qilgan holda koʻrsatishi ham shart hisoblanadi.

Rauf Parfi “Arava tortayotgan otga” sheʼrida insonlarga xos xusu-siyatlarni hayvonlarga koʻchiradi. Bunda ot obrazi yordamga keladi:

– Uf, buncha uf tortasan bunday,  
Yemishing soz, hansirama, bas!  
– Oʻzim shunday, odatim shunday,  
Men shunaqa olaman nafas [3, 68].

Shoirning “Chingiz Aytmatov” nomli sheʼrida ham ot obraziga duch kelamiz:

U manzil bagʻrida soʻngi yoʻq sirlar,  
Tashna dunyo yotar hali oʻksinib,  
Oʻqchir oʻtim va tishlari gʻijirlar,  
Ohista tuproqqa bosar koʻksini,  
Uning iztirobi meni gʻijimlar,  
Uning koʻzlariga koʻmdim oʻzimni [3, 132].

Bu yerda shoirning insoniy his-tuygʻularni, unga xos hassos ke-chinmalarni ot obrazi orqali ochib berganligiga guvoh boʻlamiz. Dun-

yoning butun tashvish-u muammosini yelkasiga ortgan va boshqa ilgirilashga majoli yetmayotgan ot obrazi shu kabi fazilatlarga ega insonlar qiyofasini yoritgan. She'ning oxirida lirik qahramon haqiqiy insonog'liklarga o'zini baxshida qilish darajasida shakllantirilgani, hatto, u otning ko'zlariga o'zini ko'mish bosqichigacha yetib boradi.

Zamon shoirlari ramzlarga asoslangan ko'chimlarni qo'llashda (ular yaratilmaydi, matnga tayyor holda olib kiriladi) an'anaviylikdan chetga chiqmaydi, poetikada shakllangan qoliplar asosida ish ko'radi. Ramziy ko'chimlar hosil qilishda shakliy jihatdan ularga ta'sir etmay yangi ma'naviy qirralarni kashf etishlari tabiiy. Yuqorida buni bir qancha misollar orqali ko'rib o'tdik.

She'riyatda ramzlarning tuzilishi va uning turli o'rinlarda turli ma'nolarda kelishini majoziylik nuqtayi nazaridan tahlil qilib to'laroq anglash mumkin. Har bir ramz ma'lum bir poetik ifoda va shu ifodaga asos bo'luvchi ma'lum bir unurni ifodalaydi. Shoirning ruhiy holati obrazlarni qanday ramziy ma'noda qo'llash uchun asos bo'lishini ta'minlab beruvchi unsur sanaladi. Shu orqali bir-birini takrorlamaydigan, bir-birini inkor qilmaydigan, o'zaro yaqinlikka ega, ammo jiddiy farqlarga uchragan bir qator ramziy obrazlar she'riyatga kirib keladi.

## XULOSA

Butun turkiy adabiyot maydonida yorqin o'zligini namoyon etgan Rauf Parfi yangi, ohori to'kilmagan, noan'anaviy majoziy obrazlari bilan kitobxonlar muhabbatiga erishdi, she'rlari ramziy obrazlilik asosiga qurilgan novatorona badiiy tafakkur tarzi bilan individuallik kasb etdi.

Uning hayvon obrazlari aks etgan talaygina she'rlari mavjud bo'lib, shoir ushbu ramziy obrazlardan badiiy tafakkur tarzini ifodalashdagi asosiy vosita sifatida unumli va o'rinli foydalana olgan.

## ADABIYOTLAR:

1. Атоуллох Хусайний. Бадойиъу-с-санойийъ. Адабиёт ва санъат – Т.: Адабиёт ва санъат, 1981. – Б. 67.
2. Жабборов Н. Замон. Мезон. Шеърят. – Т.: – Фафур Фулом номидаги нашриёт – матбаа ижодий уйи. 2015.
3. Rauf Parfi. Sakina (she'rlar). – Т.: Muharrir, 2013. – 376 b.
4. Саримсоқов Б.И. Бадийлик асослари ва мезонлари. – Т.: Ўз ФАТАИ, 2004. – Б. 9.
5. Ҳамдам У. Янги ўзбек шеъряти. – Т.: Адиб, 2012. – Б. 304.

## УДК 81-11

## ЧЎЛПОН ШЕЪРИЯТИДА ТАБИАТ РАМЗЛАРИ

*Қувватова Дилрабо<sup>1</sup>, Раҳима Шарипова<sup>2</sup>*

<sup>1</sup>*БухДУ профессори, ф.ф.д.*

<sup>2</sup>*НавДПИ ўқитувчиси, ф.ф.ф.д. (PhD)*

Мақола Абдулхамид Чўлпон шеърларидаги шамол ва булут образларининг рамзий-тимсолий маънолари таҳлилига бағишланган. Бунда шамолнинг лирик қахрамон кайфиятини ифодалаб келиши, булутнинг табиат манзараси фониди мавжуд тузумга ва воқеликка муносабатни ифодалаши асосланган.

**Калит сўзлар:** табиат рамзлари, шамол, булут, лирик қахрамон, ижтимоийлик, ташхис.

Статья посвящена анализу символических значений образов ветра и облаков в стихотворениях Абдулхамид Чулпана. Он основан на том, что ветер выражает настроение лирического героя, облако отношение к существующему строю и действительности на фоне пейзажа.

**Ключевые слова:** символы природы, ветер, облако, лирический герой, социальность, олицетворение.

The article is devoted to the analysis of the symbolic meanings of the images of wind and clouds in the poems of Abdulkhamid Chulpan. It is based on the fact that the wind expresses the mood of the lyrical hero, the cloud expresses the attitude to the existing system and reality against the backdrop of the natural landscape.

**Key words:** symbols of nature, wind, cloud, lyrical hero, social, personification.

Янги ўзбек шеърлятида ҳам шамол образи турли рамзий маъноларни билдириб келганлиги кузатилади. Бу жиҳатдан Абдулхамид Чўлпоннинг «Шарқ қизи», «Кўклам қайғуси», «Гўзал», «Баҳор келди», «Кўклам келадир!», «Яна кўклам», «Кўклам ёмғири», «Кўкламдан хабар», «Баҳорни соғиндим...», «Қор», «Уйқу», «Яна қор» «Рақамлик севги» каби шеърлари эътиборни тортади:

Шунда, шаббода – қиз билан бирга,  
 Қувлашиб ўйнаганда япроқлар,  
 Шу жувон толдаги қуруқ бир шох,  
 Қуп-қуруқ гавдасин тутиб елга –  
 Қайси кўкламни! Билмадим, йўқлар!  
 Ким ҳасратини англатар?! [1, 191]

Муҳаббат мавзусидаги бу шеърда шамол қизга қиёсланган. Яъни бунда шаббода – қиз билан бирга “қувлашиб ўйнаган япроқлар” ошиқни англатади. Шу жувоннинг “қуруқ бир шох”, “қуп-қуруқ гавда” сини шамолга тутиши эса баҳорга интилган лирик қаҳрамон кайфиятини ифодалаб келади.

Эрта тонг шамоли сочларин ёзиб,  
 Ёнимдан ўтганда сўраб кўрамен.  
 Айтадир: «Бир кўриб, йўлимдан озиб,  
 Тоғу тошлар ичра истаб юрамен!  
 Бир кўрдим мен уни шунчалар гўзал,  
 Ойдан-да гўзалдир, кундан-да гўзал!...» [1, 31]

Шоирнинг “Гўзал” шеъридан олинган юқоридаги сатрлар икки хил ҳолат тасвирини кузатиш мумкин: 1. Табиат манзараси. Яъни лирик қаҳрамон – ошиқ ўз суюқлисини эрталаб эсган майин шамолдан сўрайди. 2. Шамол ташхислантирилган. Инсондаги сочининг ёйилиши ҳолати ва шамолнинг эсишида уйғунлик бор.

Ариқда сувларнинг ўйноқи кўшиғи  
 Шохларда ухлаган баргларни уйғотди.  
 Айниқса, шамолнинг у юмшоқ шўхлиғи  
 Шохларда баргларни титратди, ўйнатди [1, 155].

“Барг” шеъридан олинган юқоридаги сатрларда ҳам шамол ташхислантирилган, яъни инсонга хос хусусият – шўхлик табиатга кўчирилган.

Чўлпон ижодида булут образи ҳам турли рамзий маъноларни ифодалаб келганлиги кузатилади. «Кўклам келадир» шеъри ҳам бу жиҳатдан аҳамиятли:

Тиниқ ҳаво... Кўк юзинда оппоқ ҳарир пардадек,  
 Оқ булутлар унда-бунда ёйилганлар паришон.  
 Шуълаларнинг акси билан йилтиллаган игнадек  
 Кун тигълари қарашларга ўткир-ўткир санчилган [1, 491].

Кўклам осмони мусаффо бўлади. Табиатнинг бу манзарасини шоир янада жозибали тасвирлайди. Баҳор осмонида сузиб юрган оқ булутлар оппоқ ҳарир пардага ўхшатирилган. Айни чоқда бунда ташхис ҳам бор: булутларнинг ҳар ёнга ёйилиши инсондаги паришонлик ҳолати билан уйғунлик касб этган.

Сенинг қаттиқ сир – бағрингни кўп йиллардир эзганлар, Сен безсанг-да, қарғасанг-да, кўкрагингда кезганлар.

Сенинг эркин тупроғингда хеч хәкки йўқ хўжалар,  
 Нега сени бир кул каби қизғанмасдан янчалар?  
 Нима учун ўч булути селларини ёғдирмас?  
 Нима учун куч тангриси бор кучи-ла солдирмас? [1, 88]

Чўлпоннинг “Бузилган ўлкага” шеърдан олинган юкоридаги сатрлар рамзий ифодаларга бой. Жумладан, шоир “ўч булути” истиорасини қўллайди. Лирик қаҳрамон эрксизликка қарши кураша олмайди. Ўч булути селлари эса кураш воситаси. Лирик қаҳрамон – ватанпарвар инсон озодликни қўлга киритиш йўлларини излаб ёзғиради. Шеърда булут шу жихатдан ижтимоий мазмун ташиган.

– Ҳаво бузук!..  
 – Кўк юзида коп-қора  
 Кўмир каби қатор-қатор булутлар;  
 Иблисларнинг қўзларидан оп-ола ёнган нима?  
 – Олтин ўчоқда ўтлар...  
 Ҳаво бузук ... Тўғри, Шарқнинг хавоси  
 Неча йилдир жуда ёмон бузилган;  
 Уни бузган оппоқ қўллар жафоси  
 Ёзмишига қора чизик чизилган  
 Кўк юзида қатор-қатор чизилган [1, 45].

Шоирнинг “Қизил байналмилал” шеърдан олинган юкоридаги сатрларда булут рамзий маъно ташиган. Яъни “кўк юзидаги коп-қора кўмир” каби “қатор-қатор булутлар” кўчма маънода озодлик ва эрк душманлари образини ифода этади. Рамз бунда ижтимоий мазмун касб этган. Шоир бу билан кифояланмайди: уларни иблисларга тенглаштиради. Қолаверса, эрксевар ижодкор юрт келажигидан ҳам хавотирда: Шарқнинг хавоси кўпдан бузилганлиги ҳам унинг зулм исканжасида бўлганлигини билдириб туради. “Ёзмишига қора чизик чизилган, кўк юзида қатор-қатор чизилган” булутларда юрт қисматини олдиндан кўриш хусусиятини англаш мумкин. Жуда мураккаб даврда яшаб, ижод қилган Чўлпон бундай хавотирларга тушиши табиий эди. Шу билан бирга, шоир мустамлака зулми остида эзилаётган халқининг дардларини бошқа йўсинда айтолмас ҳам эди.

Қора булут тўдаси ким кўкларни –  
 Шарқни ёпган парда янглиғ ёпмушдир.  
 Куз кўшини – оғу тўлуғ ўкларни  
 Ёз бағрига хеч саноксиз отмишдир!



Куз манзараси акс этган юкоридаги мисралар ҳам ижтимоий мазмун билан суғорилган. “Қора булут тўдаси” кўчма маънода Туркистон озодлигига соя солган кучларни англатади. Ёхуд ёз бағрига отилган “куз кўшини – оғу тўлуғ ўқлар” тасвирида ҳам табиат манзараси фонида мавжуд тузумга ва воқеликка муносабат яширинган.

Кўринадики, Чўлпон шеъриятида шамол ва булут образлари табиат манзарасини ифодалашдан ташқари, турли рамзларни, ижтимоий мазмунни ҳам билдириб келган.

#### АДАБИЁТЛАР:

1. Чўлпон. Асарлар. 4 жилдлик. 1 – жилд. Шеърлар, мансуралар, шеърый хатлар. – Тошкент: Академнашр, 2013.

2. Quvvatova D., Karimova Sh. Artistic Functions of Anaphora in Modern Uzbek Poetry// Journal of Positive School Psychology <http://journalppw.com> 2022, Vol. 6, No. 5, 6270 – 6274.

3. DN Yuldasheva. BASIC NON-VERBAL COMPONENTS OF SPEECH. CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (2767-3758) 3 (01), 17–24. <https://masterjournals.com/index.php/crjps/article/view/661>

4. Quvvatova D., Sharipova R. The graduality of nature symbols in folk lyrics and classical poetry // A Journal for New Zealand Herpetology. BioGecko. Vol 12 Issue 03 2023 ISSN NO: 2230-5807.

## УДК 81-11

ETYMOLOGICAL EXPLANATION OF THE PHENOMENON  
OF SONG AND THE PROFESSION THAT DENOTES IT

*Rajabov Dilshod Zaripovich*  
*Bukhara State University*  
rajabovdz@gmail.com

Despite the fact that the song was created as a folk poem in the scales of finger and aruz, it differs in that it is intended for solo and collective, refined (sung) and unorthodox (sung in performance) performance. For example: “jugs”, some types of songs on Navruz, “Let’s go play”, “song about sickles”, songs performed at funeral ceremonies were performed solo and by a team. Previously, most seasonal ritual songs, the baker’s song, ufo-jakala, family ritual songs were performed with a melody (sung), while labor songs were performed without rhythm (sung), boar, embroidery, hoe, etc. Folk songs, however, require a certain musical tone, regardless of which genre they are performed – solo, collective, oral, sung without melody. This is another of the main features of the song.

**Basic concepts:** motive, art, genre, song, belief, image, prayer, finger, ritual

As you know, art consists of three components: CREATOR – PERFORMER – LISTENER. A characteristic feature of this triad, perhaps, was the “presentation at a high level of human capabilities and skill” in art. Therefore, in folk art there is an absolute distance from artificial forms, in which there are no artistic structures that do not have a natural basis. In such a folk song we do not see anything too expressive, too polished, and shaved. However, in the samples of folk oral art, especially in works related to epic genres, there are often motifs built on the basis of magic, fiction and fiction. Nevertheless, a person is always able to enrich his simple, correct sincere opinion by returning to himself. No matter how much she is draped in a dress, nevertheless, aggressiveness, anger are also striking, as are such noble qualities as tolerance, affection, loyalty, compassion in her. M.Koshgari’s reflections on the word *kög terani* confirm that he has existed in this for a long time. In addition, the “Explanatory Dictionary of the Uzbek language” states that “a song is a composed song performed by a singer, a group or a choir.” If we rely on this idea, it turns out that “song” and “song” intrusions are terms that stand in the same synonymous row. However, the words of the song and the chant, although they stand in the same synonymous row, represent two separate phenomena in their meaning and content.

Professor O. Safarov in his textbook “Uzbek folk oral art” writes: “Many centuries have passed since the term “song” was introduced into scientific circulation. After all, it represents folk poetry in a broad sense, while in a narrow sense it represents the form – genre of folk lyrics. This can be observed in Tajik folklore, as well as in Karakalpak, Azerbaijani and Uighur folklore. The invasions of Chunonchi, Tajik “surud”, Karakalpak “koshik”, Azerbaijani “koshma” and Uighur “Koshak” have such an aesthetic essence. In M. Brown’s “DLT” there is such a quatrain:

Turkan qatin qutyog’a,  
Tergur mendin qo’shig’  
Ajpgul sizni tabug’chi,  
Otnur jag’i tabug’.

Mazmuni:  
Turkon xotin qoshig’a,  
Mendan yetkur she’r.  
Xizmatda xodimingiz,  
Rost, deb payom ber.

In English:  
Turkan katin kutyoga,  
Tergur Mendin ‘s Hut  
Aigul you tabugchi,  
From nur jaw tabug.

Content:  
Turkon’s wife’s spoon,  
a poem from me.  
Your employee in the service,  
Give Payom what is true.

This is the poetic literature of Academician A.Kayumov. Indeed, the word “song” was originally used to mean “praise, deception.” It is noteworthy that this activity took the form of a “song” in Yusuf Hos Khojiib’s book “Kutadgu bilig” (“Knowledge of the messenger of happiness”), which underwent phonetic change, while in Khorezm it took the form of a “song”. On the works of Yusuf Khos Khojiib:

Bu turkche qo’shiqlar tuzettim seg’e,  
Oqo’rda uno’tma duo qil meg’e.

Content: I composed these Turkish songs for you (I wrote),  
bless me, don’t forget when you read.

In Kharezmiy's poems:

Kel, ey oy yuzli dilbar, tut bir qo'sh,  
Biror qo'sh birla, qilgil bizni madxush.

In this byte, the word “ko'sh” has the meaning of ending, creating a poem.

The 12th-century linguist Mahmoud Umar Zamakhshari comments in his Arabic-Persian-Turkish-Mughal quadrilingual dictionary “Muqaddimat ul-adab” that the Arabic “poem” is considered a “song” during the Turkish occupation, noting that the word “qo'shiqchi (singer)” formed with the suffix “-chi” attached to it, is essentially equivalent to the term “poet”: “the poem by kotho guft is a poet-short singer”. (Alisher Navoi also uses the “song” occupation in his work “mezon ul- avzon” to refer to “poetry”, again noting that it is also referred to as feminine, lan, Turkic, where they occur as internal forms of folk poetry.

Alisher Navoi writes, paying special attention to the fact that “song” has a genre concept in a narrow sense: “Yana is a song of durkim, an orgushtak method, and some supporters of kutubida ul usul olubtur and ul surud a'rob “teva surur gudi Vaz bila madidi musammani solimda voke olur, aning asli bu na”. - vdurkim, beit:

Vahki ul oy hasrati dardi dog'i firqati,  
Ham erur jonimga o't ham hayotim ofati”

A.Navoi imagines a folk song composed in such a way, finished in the aruz valley, that now such a phenomenon differs from the song by the invasion of “song” and is perceived as the internal monotony of the song.

Babur also writes in his “Risolai Aruz”, confirming this point of view: “During the time of Sultan Hussein Mirza, there was another shooting season of thing appeared “Turkish”.

Since the days of Alisher Navoi and The Beavers, it has become a principle to distinguish between the internal genres of the song.

In general, the term “song” is derived from the Turkish verb “add”, which means to add and sing line by line. After all, a folk song is born together with a melody and this differs from poetry in written literature. In addition, folk songs have historically also been accompanied by dancing and, of course, differ in their consistency with singing. A.Musakulov takes this feature into account and considers folk songs as “a kind of poetic art that arose on a historical basis, inextricably linked with ancient ideas, customs, especially with the cult of productivity,

accompanied by singing and dancing, and steadily maintaining its connection with singing.” In general, traces of beliefs are also important in folk songs that do not need the magic of it, and the secret. Therefore, scientists who have specifically studied the types of folk art related to art, as you know, came to the conclusion that beliefs, rituals played an important role in the creation and development of works of folk art. M. According to Khadyrov, the art of dance has proved its origin on the basis of symbolic actions related to the labor process, ritual and ancient beliefs, by the example of the development of traditional folk dances, such as the “big game”, “status game”, “dance game”. But even on the basis of these dance patterns, the unity of man and nature served as an important factor. The dance elements in the game “five pops” grew out of ancient rituals based on imitation of hungry wolves and village dogs. Similar views on the history of folk dance have also been put forward by major folklorists. Chunonchi, Khodi Zarif, in his article “On the comparative study of folklore and archaeological materials” claims that the “clapping game” appears in connection with the hunting rituals of our ancestors and is the remnants of raccoon dances, which later merged with totemic beliefs. He points out that the dancers performing this dance are represented by a wolf in the middle, and people belonging to a tribe that worships a wolf in a circle.”...Because the image of the Wolf is considered positive, even sacred, by some Turkic peoples, including several Uzbek Semenov, whose concept is due to the fact that they once believed in the wolf, knew it sacred as a totem,” writes. Based on these considerations, we can say that a person expresses his feelings, thoughts, joys and sorrows, so to speak, colorful experiences in a ready state, often in four and in certain situations in binary, ternary, quintuple, hexadecimal forms, singing, intonation and dancing in various versions based on “from mouth to mouth” is a popular folk poem among the el. “Due to the fact that it is usually performed by men, often teenage boys, the style of performance observed in dance games also has characteristics associated with the expression of certain emotions through male dance movements.” As it becomes clear, the song performed in the direction of “applause” was performed mainly by young people, which generated a harmony of specific content and meaning. Despite the fact that the song was created as a folk poem in the scales of finger and aruz, it differs in that it is intended for solo and collective, refined (sung) and unorthodox (sung in performance) performance. For example: “jugs”, some types of songs on Navruz, “Let’s go play”, “song about sickles”, songs performed at

funeral ceremonies were performed solo and by a team. Previously, most seasonal ritual songs, the baker's song, ufo-jakala, family ritual songs were performed with a melody (sung), while labor songs were performed without rhythm (sung), boar, embroidery, hoe, etc. Folk songs, however, require a certain musical tone, regardless of which genre they are performed – solo, collective, oral, sung without melody. This is another of the main features of the song.

## REFERENCES:

1. Ёқуббекова М. Ўзбек халқ кўшиқларида ўхшатиш. – Т.: Фан, 2003 – Б. 6.
2. Мамажонов С. Давр оҳанглари, асримиз образи, замондошимиз портрети. ЎТА 1982 й №2, – Б. 9.
3. Умумий психология. Т.: ўқитувчи., 1992, – Б. 338.
4. Винокур Г.О. Понятие поэтического языка . Избранные работы по рускому языку. Москва.: Просвещение, 1959. 390 бет.
5. Панов Е.Н. Знаки. Символы. Языки. М.: Знание, 1980.
6. Орипов А. Танланган асарлар. – Т.: Ғофур Ғулом, 2001 – Б. 227.
7. Юсуф М. Сайланма. – Т.: Шарқ. 2007. – Б. 58.
8. Йўлдошев М. Бадий матн лингвопоэтикаси – Т.: Фан, 2008 – Б. 8.

УДК 81-11

**O'QUVCHILARDA O'ZBEK MUSIQA FOLKLORI HAQIDAGI  
TASAVVURLARNI SHAKLLANTIRISHNING  
IJTIMOIY-PSIXOLOGIK OMILLARI**

*Rajabova Go'zal Zarifovna*  
*Buxoro davlat pedagogika institute*  
rajabovag@gmail.com

Maqolada umumiy o'rta ta'lim maktablarining musiqa darslarida o'zbek musiqa folklori haqidagi tasavvurlarni shakllantirishning ijtimoiy-psixologik muammolari haqida fikrlar bildirilgan.

**Kalit so'zlar:** Folklor qo'shiqlari, o'zbek musiqa folklori, ijtimoiy-psixologik muammolari, bolalar folklor, marosimlar.

**SOCIO-PSYCHOLOGICAL FACTORS FORMING STUDENTS'  
CONCEPTS ABOUT UZBEK MUSICAL FOLKLER**

The article expresses opinions on the socio-psychological problems of forming ideas about Uzbek musical folklore during music lessons in secondary schools.

**Key words:** Folk songs, folklore of Uzbek music, socio-psychological problems, children's folklore, rituals

**СОЦИАЛЬНО-ПСИХОЛОГИЧЕСКИЕ ФАКТОРЫ  
ФОРМИРОВАНИЯ ПРЕДСТАВЛЕНИЙ СТУДЕНТОВ  
ОБ УЗБЕКСКОМ МУЗЫКАЛЬНОМ ФОЛЬКЛОРЕ**

В статье высказаны мнения о социально-психологических проблемах формирования представлений об узбекском музыкальном фольклоре на уроках музыки в общеобразовательных школах.

**Ключевые слова:** Народные песни, фольклор узбекской музыки, социально- психологические проблемы, детский фольклор, обряды.

O'zbek xalqining boy musiqa merosi xazinasida katta yoshdagi kishilar dunyoqarashi bilan kichik yoshdagi bolalar ma'naviy olamini o'zida mujassamlashtirgan bolalar folklor qo'shiqlari (aksariyat o'yin bilan uzviy bog'liq holda ijro etiluvchi) salmoqli o'rin tutadi.

Folklor qo'shiqlari mazmunida bolalarning o'ziga xos quvnoq, o'yinqaroq, sodda va beg'ubor hayoti, dunyoqarashlari, bolalarbop o'yinlar, aytishuvlar va muayyan voqelik bilan bog'liq jarayonlarni ifoda etilishi bilan o'ziga xos xususiyatga ega.

Bolalar folklor qo‘shiqclarini o‘rganish va tahlil qilish shuni ko‘rsatadiki, ular ham xalq musiqasining ommaviy turlari singari badiiy va g‘oyaviy mazmunga ega bo‘lib, sodda shakli, ijro uslubi, nihoyatda ixcham, o‘yin-raqs xarakteriga egadir. Shularga ko‘ra ulardan ta‘limiy va tarbiyaviy maqsadlarda keng foydalanishda qulaylik tug‘diradi.

Xalq o‘g‘zaki ijodida yaratilgan qo‘shiq namunalari bolalarning ruhiy, aqliy, badiiy tafakkuri va dunyoqarashiga mos soddaligi, kuy-lashga oson va qulayligi bilan bolalar diqqatini o‘ziga tez jalb qila oladi. Shuning uchun ham maktabgacha ta‘lim muassasalari hamda umumta‘lim maktablarining boshlang‘ich sinflarida o‘quvchi-yosh-larda milliy musiqa san‘atimizga qiziqish va ishtiyoq uyg‘otish, bolalar folklor qo‘shiqclarini o‘rganish, kuylash jarayonida ularning g‘oyaviy-badiiy mazmunini anglash, idrok etish, tasavvur etish bilan ularga o‘z munosabatini bildira olish, ya‘ni tavsif berishga o‘rgatib borish va shu asosda bolalarda musiqa san‘atining asta-sekinlik bilan jiddiy hamda murakkab turlariga ijodiy, faol munosabatni, ularda ijtimoiy-psixologik omillarni shakllantirish mumkin. Bu esa bugungi kunda musiqa ta‘limi va tarbiyasi uchun mas‘ul tarbiyachilar, pedagoglar va to‘garaklar rahbarlari oldidagi eng muhim va ahamiyatga molik vazifalardan hisoblanadi [1,53].

Umumta‘lim maktablarining boshlang‘ich sinflarida o‘qitiladigan fanlar dasturi va darsliklarida ko‘plab xalq og‘zaki ijodiyoti namunalari bilan bir qatorda xalq qo‘shiq va aytimlari hamda bolalar folkloriga oid namunalari ham kiritilgan. Ta‘limning ilk davrlaridan boshlab bolalarga bunday namunalarni o‘rgatish, ijro ettirish jarayonida ularga og‘zaki an‘anadagi milliy folklor musiqasi, uning yetakchi janrlari, tuzilishi va an‘anaviy ijro uslublari haqida eng zarur tushunchalarni berish asosiy maqsad qilib qo‘yilgan.

Har bir xalqning milliy o‘ziga xosligini ko‘rsatadigan an‘ana va marosimlari, urf-odat va bayramlari bor. Asrlar bo‘yi yaratilib, zukko xalqning yuksak ijodiy salohiyati tufayli sayqallanib kelgan ana shunday an‘analari aslida o‘sha millatning o‘zligini, mentalitetini, milliy qiyofasini belgilaydi.

O‘zbek xalqining paydo bo‘lish tarixi qanchalik qadimiy bo‘lsa, uning milliy an‘analari, urf-odatlari, marosimlari ham shunchalik uzoq tarixiy ildizga egadir. Ular xalqimizning ma‘naviy ehtiyojlari zamirida juda qadim zamonlardayoq paydo bo‘lib, shakllangan va zukko ajdodlarimizning benazir ijodiy salohiyati tufayli asrlar mobaynida sayqallanib kelgan.

Xalqimiz an‘analarining go‘zalligi va betakrorligini his etish uchun birgina dehqonchilik madaniyati bilan aloqador urf-odatlar, mavsumiy



marosimlar va serhosillik g'oyasini o'zida mujassamlashtirgan qadriyatlarini eslashning o'zi kifoya. Misol uchun, hut yaqinlashishi bilan, "qoziq boshida qor turmas" erib ketadigan davr boshlanganda qish uyqusidan uyg'ongan dala-qirlarda, tog' ungarlarida ko'klam darakchilari – boychechaklar bo'y ko'rsatishi bilan oq xalqimiz an'anasiga ko'ra, "boychechak xabari" marosimi o'tkazilgan [2,18].

Birgina bahor faslida o'tkazilgan udum va marosimlarning o'zida bir olam ma'no aks etgan: "loy tutish" – bahor faslining ilk kunlarida loyqa o'tirib qolgan ariq-zovurlarni mahalla-qishloq ahli birgalashib, ko'pchilik bo'lib tozalashi, ariq qazish va ularni yangi mavsumga tayyorlash maqsadida o'tkazilgan an'anaviy hashar; "qozon to'ldi" – ro'zg'orga qut-baraka yog'ilishi, mahallalar obod, ko'ngil xotirjam bo'lishini tilash maqsadida navro'z bayramidan bir kun avval o'tkazilgan oilaviy marosim; "shox moylar" – bobodehqonlarimizning mo'l hosil olish niyatida erta bahorda dalaga qo'sh chiqarib, baraka urug'i qadashi udumi; "qo'sh oshi" – erta bahorda dalaga birinchi marta qo'sh solinadigan kuni momolar tomonidan bug'doy yormasidan maxsus taom pishirish odati; "yilboshi sayli" – quyoshning hamal burjiga o'tishi, yangi mehnat mavsumining boshlanishi, inson ruhiyatida ham, tabiatda ham yangilanish, tozarish jarayoni ro'y berishi va bahoriy tengkunlikni nishonlash maqsadida o'tkazilgan qadimiy bayram; "kadi sindirdi" – navro'z sayli kuni mahalla ahli bir yerga jam bo'lib, yangi yilda barcha ezgu niyatlar amalga oshsin, ro'zg'orimizdan qut-baraka arimasin degan niyatda oilada foydalanib kelingan suv va moyqovovqlardan bir nechtasini yerga urib sindirish udumi; "navro'z go'ja" – g'alla ekinlarining hosili barakali bo'lsin uchun yilboshi bayrami kuni yetti xil donli ekin masallig'idan go'ja tayyorlash an'anasi; "olma otish" – navro'z bayrami saylida qiz-juvonlar halinchak uchishayotganda yigitlar o'zlariga unashtirib qo'yilgan qayliqlariga olma o'tishi odati; "sumalak sayli" – har yili ko'klamda yil barakali, serhosil, to'kin-sochin kelsin degan niyatda bug'doy maysalatib, bahoriy taom tayyorlash marosimi; "sumalak bayti" – sumalak saylida chechan momolarning ko'klamni vasf etuvchi xalq qo'shiqlarini aytishi an'anasi; "halinchak uchish" – ko'klamda qiz-juvonlarning bog'lardagi mevali daraxtlarga arqon tashlab, arg'imchoq uchishi udumi; "buqcha tashladi" – ko'klam kunlari dehqonlar dala hovlisiga ko'chib chiqmoqchi bo'lganda o'zlari tanlagan joyga pichoq suqib belgi qilishi va band qilish uchun biror buqchani tashlab kelishi odati; "abruk chiqarish" – boshqa chorvadorlar bilan birgalikda belgilangan kuni bahorgi yaylovga ko'chib ketolmagan cho'pon oilasi o'zining birorta ro'zg'or anjomini ko'chuvchilardan berib yuborishi qa'dasi; "gulbarra" – ko'klamning

birinchi chechagi – “gulbarra” ochilganda o‘tkazilgan xalq sayli; “lola sayli” – qir-adirlar, dala-tuzlar qip-qizil lolalarga burkangan paytda tabiat qo‘yniga saylga chiqish udumi; “qizil gul sayli” – har yili ko‘klamda qizil gullar ochilgan paytda an’anaviy tarzda o‘tkazilgan sayl; “darveshona” – yil barakali kelsin degan umidda erta bahorda aholidan ixtiyoriy asosda to‘plangan pul yoki oziq-ovqat hisobiga o‘tkazilgan xayriya ehsoni; “Sust xotin” – yil qurg‘oq kelganda o‘tkazilgan qadimiy yomg‘ir chaqirish marosimi va hokazo.

Xullas, ijodkor xalqimizning bahor fasli bilan bog‘liq bunday marosim va udumlari g‘oyat ko‘p bo‘lib, bundan bir-ikki yil burun men ularni ro‘yxat qilganimda birgina navro‘z bayrami tantanalari doirasida an’anaviy tarzda o‘tkazib kelingan yetmishdan ortiq turli xil udum, rasm-rusum va marosimlar mavjud bo‘lganligini ko‘rib, xalqimiz ma’naviy qadriyatlari tizimining qanchalar serqirra va mukammaligidan qalbmida o‘zgacha bir g‘ururni tuygan edim[3,41].

O‘zbekistonning har bir hududi o‘ziga xos folkloriga ega. Har bir viloyatlarning o‘ziga hosligi, viloyatlarning o‘zida tumanlari folklorining betakrorligi o‘zbek xalqining qadimdan dunyoqarashi boy va san’atsevar ekanligidan darak beradi. Ajablanarlisi hech bir hududning folklori bir biriga o‘xshamaydi, o‘xshash jihatlari bo‘lgan taqdirda ham, albatta o‘sha hududninggina o‘ziga hos bo‘lgan jihati albatta ajralib turadi.

Xulosa qilib aytganda, Milliy istiqlolimiz sharofati tufayli o‘zbek musiqiy folklorshunosligida ham yangi bir ilmiy yo‘nalish vujudga keldi. Bu yo‘nalishning asosiy maqsadi har bir hudud folklorini alohida-alohida tadqiq etish, har tamonlama chuqur o‘rganish, uning o‘ziga xos badiiy olamini va psixologik asoslarini keng tahlil qilishga imkon berdi. O‘zbek musiqiy xalq folklori san’ati darhaqiqat boy va betakror sanaladi. Uni yanada chuqurroq o‘rganish va qatida yashiringan boy ma’naviy merosni kelajak avlodga o‘rgatish va saqlab qolish bizning bugungi kundagi eng dolzarb vazifalarimizdan biri sanaladi.

#### ADABIYOTLAR:

1. Mahmudov M.H, Safarov O.T. Oila ma’naviyati. – T.: Ma’naviyat, 1999.– 210 b.
2. Qo‘chqorov G‘. Folklor qo‘shiqchiligi. – T., 2005. – 48–50 b.
3. G‘oziyev E. Psixologiya. – T.: O‘qituvchi, 1994. – 224 b.
4. Yuldasheva D.N., Chullieva G. T. THE ROLE OF NOVERBAL MEANS IN SPEECH MEANING. International Engineering Journal For Research & Development, Vol.6, Issue 2 published May 16, 2021. – Pag. 1956–1961. [www.iejrd.com](http://www.iejrd.com) <http://www.iejrd.com/index.php/%20/article/view/1956>.

УДК 81-11

GENESIS AND DEVELOPMENT OF THE CONCEPTION AND  
THE CONCEPT “ROAD”*Rakhimov Muzaffar Muradivich,  
Ahrorova Nigina Anvar's daughter**Bukhara State University*

muzaffar@gmail.com, nigina@gmail.com

**Basic concepts:** conception, myth, motif, mythological, multifunctional, genesis, episode, path, characteristic.

**Abstract:** This conceptosphere has frozen in the structure of the epic plot, moving from myth to fairy tale as a synthesis of primitive ideas. In fact, although the heroes are preceded by a three-sided conceptosphere, artistic interpretations of this epic motif retain its archaic layer – the ancient core: the epic land to which the older brothers went is opposed to “the one that does not come” from which the younger one went. The general aspect inherent in the people, including culture, is verbalized through language, it is language that drives the basic concepts of culture, literally, culture is fully realized in language and comes to the surface in speech. The realization of the desired advantage inherent in culture is carried out through the trichotomy “thinking-language-speech”.

This section highlights that language expresses understanding, the emergence of true and figurative meanings in speech reality, its relation to a specific text and dialect; the communicative nature of language, as well as the fact that culture is a semiotic sign system capable of absorbing information in the same way as language. Yu.M.Lotman believes that although culture is a multifunctional, lush unit, it cannot self-organize, unlike language, since it, first of all memory, is considered a source of activity of a complex semiotic system acting as a collection, accumulation.

The genesis of the conceptosphere of the “road” begins with folk oral art. Mainly in fairy tales, it acquires a special meaning and meaningfulness, including a leading place in the adventure journeys of the heroes of the work. In most Uzbek and Russian fairy tales, the three-eyed exudes three-eyes in front of the bathers, the three-way exit opens the way for the Wanderer and makes it interesting. The three paths, in turn, open the separation of good and evil, meanings that cause the mouths of brothers to move away from each other, determining the path in the next life. Fairy tales also act as a tool to ensure the development of the plot. Thus, the traditional three-sided sema in many folk-

lore tales has been thoroughly analyzed by researchers. In our opinion, the traditional three-sided motif goes back to the mythological ideas of the ancient Turks. In the pre-existing examples of mythological legends that trace the origin of the material world, primitive ideas about the structure of the universe are completely shaken. They say that at some point the whole being was clouded by the boundless water, in which a good ship was floating, in which the son put the old one on his shoulders. Ulgan created this land with the support of his nanny. To strengthen the ground above the water, three bales were enough for the giant: two were placed on the edge, and one in the middle. The fish's head in the middle was pointed to the North and was considered the main pillar of the Earth. Ancient Turkic legends also initially reflected two worlds – their own and someone else's. This mythological imagination is also embodied in the image of an epic space in Uzbek folk tales. The analysis of the motif of the traditional tripartite conceptuosphere shows that when describing one mythological space, the three worlds (blue, terrestrial, underground) are never compared with each other. Because in fairy tales there are always two ways – two worlds are compared. In the fairy tale “Bulbuliguyo” three paths go ahead of the princes, who go on a journey in search of the bird that stole the emerald leaf of the tree. The older brother says “come,” the middle one says “take a chance,” and the younger one says “don't go.” Then the middle one said: “It's ‘dangerous if it happens,’ chat, won't it happen if I come with my brother and talk,” he followed his brother and got to him. Sometimes it is not shown that these two paths are comparable from birth to birth. In particular, in the fairy tales “The Swan”, “Abulkosim”, “The King of Egypt” the three brothers were divided into three ways. The episode is about how the average one returns from the road and follows in the footsteps of his older brother cargo. This conceptuosphere has frozen in the structure of the epic plot, moving from myth to fairy tale as a synthesis of primitive ideas. In fact, although the heroes are preceded by a three-sided conceptuosphere, artistic interpretations of this epic motif retain its archic layer – the ancient core: the epic land to which the older brothers went is opposed to “the one that does not come” from which the younger one went. The general aspect inherent in the people, including culture, is verbalized through language, it is language that drives the basic concepts of culture, literally, culture is fully realized in language and comes to the surface in speech. The realization of the desired advantage inherent in culture is carried out through the trichotomy “thinking-language-speech”.

It is obvious that researchers have different approaches to the concept of linguacultural words. For example, Emil Benveniste gives the following definition to the concept of language and culture: “Humanity imagines linguacultural words in three units. This is culture, language and human personality. The researcher perceives linguacultural word as a “lens” and studies material and cultural values through it. This “lens” is based on a linguistic concept.”

It can be seen that one of the irreplaceable conceptual units characteristic of the Uzbek nation, the “path”, has been improving over the years, becoming a multi-valued unit of national values. This conceptual sphere is multifaceted, it is a path that instills in the soul of a representative of this nationality feelings of pride, dignity, affection – his whole life. Currently, there are two different approaches to the term “conceptual sphere” in linguistics: it is approached by the categories of cognitive linguistics and linguaculturology.

In cognitive linguistics, the conceptual sphere is studied as a tool representing human knowledge and experience, the mental lexicon, the landscape of the entire universe, the conceptual and linguistic system of the brain, and the human psyche. The conceptual sphere represents cultural universals and exists in thinking, containing cognitive-rational and emotional-mental content.

Each expression within the framework of the conceptual sphere “path” is a specific organizer of the lexico-semantic field inherent in it. The main source for a thorough analysis of their meanings related to external and internal culture is fiction.

In particular, Said Ahmad Baho in his book about kishsalar tuplami “From kishnagan oksham” by Togai Murad says: “After reading these lines, you become disembodied for Hael. You will only be smeared, as if Mangli was listening to a song.

The song was a long journey and came from Chennai, the Southern Surkhana of our Uzbekistan” (p.5). This is quoted in the lines: “a long way”, the song gives an idea of the homeland of those units that have gone far. And also proudly praises the motherland – the sennok of our Uzbekistan, who arrived from Southern Surkhana. That is, without walking, the bulk of the boat took off. Ogir Hayat miammolar mungli song bulib flew from Surkhan on a remote way. The “path” concentrate in these sentences is a breathtaking simple distance, a place of decline, but also scraps of life.

Note that only one of these quoted mujaz Misra was able to bring to the surface a number of internal meanings of the lexical unit “road”.

At the same time, the conceptosphere of the “path” sings like a song about great love – love desires in our national values. For example, “the great epics about sharq – Leyli and Majnun, Farhad and Shirin, Tokir and Zuhras – have been living for several hundred years. They depict the unfortunate Oshik-Mashuk, who, unable to reach each other, gave their lives on the path of love” (the same work. 5 p.) In honor of the heroes of the epic about sharks, brought by the oppressor of the Uzbek people, Said Ahmad Yukori, he connects them with the story of Togai Murad “People who walked on the moon”. Said Ahmad called this verse as follows: the verse “people walking on the moon” became known as “A love song”. Togai Murod kissa Har sings the mutual love of the two heroes – the old woman chol, as a song. Generally speaking, they are unfortunate oshik -Mashuk, who gave their lives on the path of life-love. The relationship between these two heroes is dominated by the national culture of bebaho sharks. Look.”..According to the plot, Koplun and Oimomo live, connecting night with dawn, day with the moon, year after year in the hope of a child. Coplon convinces his wife Momoshi that Oimomo is her husband’s grandfather. They are the grandfather of the loader child, the mother of the loader child is blurring, connecting years after years.”(It works.5B.). (when spouses are not used to each other’s name).

In the Uzbek language and culture, the concept sphere “road” is important and occupies a special place in the system of linguistic units. It, like all linguistic conceptual spheres, has manifested itself as a sub-model of culture.

#### REFERENCES:

1. Yuldasheva D.N., Chullieva G. T. THE ROLE OF NOVERBAL MEANS IN SPEECH MEANING. International Engineering Journal For Research & Development, Vol.6, Issue 2 published May 16, 2021. – Pag. 1956–1961. [www.iejrd.com](http://www.iejrd.com) <http://www.iejrd.com/index.php/%20/article/view/1956>.
2. Yuldasheva D. N. Problems of national language education at school. Volume 4 Issue 3 BSU 2020 (3) Published by 2030 Uzbekistan Research Online, 2020 <https://uzjournals.edu.uz/buxdu/vol4/iss3/15/>.
3. Юсупова Д. Ю. Х. Худойбердиеванинг сўзларни кўчма маънода қўллаш маҳорати//НамДУ илмий ахборотномаси – Научный вестник НамГУ 2021 йил 6-сон. – 395–402-б.

УДК 81-11

**ОБРАЗНЫХ ПРИЕМОВ И ПОЭТИЧЕСКИХ ФИГУР,  
ХАРАКТЕРНЫХ ДЛЯ ТЕКСТОВ ЭССЕ*****Рахимова Мохинабону Абдухамид кизи****доктора философии (Phd) по филологическим наукам  
старший преподаватель национального центра обучения  
педагогов новым методикам Самаркандской области  
тел: +998933339980 mohinabonu1992@mail.ru*

В данной статье подробно анализируются фоностилистический анализ текстов сочинений на узбекском языке и их специфика. В частности, с помощью примеров объясняются такие фонетические явления, как аналогия и диатез в текстах эссе.

**Ключевые слова:** эссе, поэтические фигуры, аналогия, диатез, фонетическая стилистика, своеобразие текстов эссе.

This article analyzes in detail the phonostylistic analysis of the texts of essays in the Uzbek language and their specificity. In particular, with the help of examples such phonetic phenomena as analogy and diathesis in essay texts are explained.

**Key words:** essay, poetic figures, analogy, diathesis, phonetic stylistics, originality of essay texts.

Фонетическая стилистика при использовании звуков в стилистических целях изучает их особенности речевой организации: ритм, тон, аллитерацию и т. д. [1, 11]. «В публицистике есть такие средства, как антитеза и риторический вопрос, риторическое обращение, которые служат для усиления интонации в речи» [2, 28].

Основные описательные средства и поэтические фигуры, характерные для языка текстов эссе, – сравнение, метафора, эпитет, преувеличение, литота, антикульминация, диатез, умолчание, перифраз – также служат повышению экспрессивно-эмоционального воздействия текста.

**Анафора** (анафора – повторение слова или сочетания в начале стихов в поэтическом произведении [3, 55]) была лингво-поэтическим средством для яркого выражения мыслей автора о сердце, истине, справедливости. Например: Борлиқда нимаики мавжуд – барида, барчасида қалб бор, қалбсиз нарса ҳам, жондор ҳам, воқеаю ходисот ҳам йўқ. Қалб – марказ, қалб – калит, қалб – ўлчов, мезон. (Х.Дўстмухаммад. Қалб хамиша ўртада). То, что существует в существовании – все, все имеет душу, нет бездуш-

ной вещи, нет живого, нет события или события. *Сердце* – центр, *сердце* – ключ, *сердце* – мера, критерий (К.Дустмухаммад. Сердце всегда посередине).

В текстах эссе, в **акромнограмме** четверостишие или предложение заканчивается вторым, второе начинается с этой единицы. В других его формах слова и фразы повторяются в этом порядке. Например: Уйдан *ишга, ишдан бозорга, бозордан богчага, богчадан дўконга, дўкондан шифохонага, шифохонадан дорихонага, дорихонадан оихонага, оихонадан бошқа...* хоналарга шошасан (А. Жўраев. Шошаётган одам). Ты носишься *из дома на работу, с работы на рынок, с рынка в садик, из садика в магазин, из магазина в больницу, из больницы в аптеку, из аптеки на кухню, из кухни в другие комнаты* (А. Джураев. Спешащий человек).

**Аналогия.** Тексты эссе также чрезвычайно богаты сравнениями, которые создают образы. Среди показателей сравнения **дай/-дек/-дайн, гўё, худди, бамисоли, каби** (-как/-как/-как, как бы, впрочем, как будто, подобно) преобладает частота употребления **-дек / словно**:

| индика тор<br>общего<br>объема<br>с повт<br>рением | У.Азим             | Ш.Ортиков       | М.Мирзо            | А.Абдиев          | Н.Жўраев        | Н.Жонузок       | Н.Боқий         |
|--|--------------------|-----------------|--------------------|-------------------|-----------------|-----------------|-----------------|
|  | 2932<br>//<br>1656 | 2719 //<br>1795 | 5612<br>//<br>3102 | 1552<br>//<br>732 | 4595 //<br>2159 | 2194 //<br>1447 | 7899 //<br>4013 |
| -дай   | 9                  | –               | 1                  | –                 | 5               | 18              | 1               |
| -дек   | –                  | 4               | 5                  | –                 | –               | –               | 4               |
| -дайн  | –                  | –               | 1                  | –                 | –               | –               | –               |
| гўё  | –                  | 2               | 1                  | 2                 | 2               | 6               | 1               |
| каби   | –                  | 14              | 1                  | –                 | –               | –               | –               |
| бамисол<br>и                                       | –                  | –               | 1                  | –                 | –               | –               | –               |
| мисоли   | –                  | –               | –                  | –                 | 1               | –               | –               |
| худди  | 1                  | 4               | 3                  | –                 | 1               | –               | –               |
| ўхшаб  | –                  | 2               | 1                  | –                 | –               | –               | –               |



Среди них показатель *-дай* (как) употребляется больше всего 18 раз: *булбулдай, боларидай, беш кўлдай, нурдай, шабнамдай, ёқутдай, яшилдай, ёшдай, боладай, хурсанндай, ходимдай, одамлардай, қайиқдай, подшоҳдай*; как соловей, как ребенок, как пять рук, как свет, как роса, как рубин, как зелень, как ребенок, как ребенок, как рабочий, как человек, как лодка, как король. *-дек* (как) употребляется 4 раза: *гўрўглидек, қиёматдек, рамзидек, паҳлавондек, жилгаларидек, шеъридек, учидек, гавжумдек, адабиётдек, шоҳдек*; как бык, как судный день, как символ, как борец, как джилга, как поэма, как совет, как толпа, как литература, как царь. Каби (как) 14 раз: *ҳикоялар бўлгани каби, яхши томонга ўзгариши каби, Станиславский каби, “Артамоновлар иши” каби, барча нариги дунёчилар каби; как истории, как перемены к лучшему, как Станиславский, как Дело Артамонова, как все остальные миры.*

В эссе сравнением является в основном человек, а сравнение предстает как животные и птицы: *булбулдай, боларидай, тутиқушдай, турналардай, тойчоқдай*; как соловьи, как дети, как попугаи, как журавли, как кобылы. Творчество уподобляется предмету, природному явлению: *гугурт чўпининг учидек, шоир мисраларга шодадай тизаркан, ёқутдай бебаҳо; тоғнинг мусаффо жилгаларидек, гўё сув, боғи ҳам худди адабиётдек*; как кончик спички, поэт как цветок в стихах, бесценный как рубин; как чистые холмы горы, как вода, и сад, как литература. Например: Тонглар совуб, чўпон-чўликнинг кўл ва лаблари баайни *совуқ очган кўсақдай ёрилади* (Ш.Салом. Қорхат); Утро прохладное, руки и губы пастуха *трескаются, как волдыри от холода* (Ш. Салам. Кархат).

Такие стилистические фигуры называются антиклимаксами, если слова с похожим значением или слова с противоположным значением используются рядом неоднократно в соответствии с их градуальными характеристиками. Примеры: *Бевафо, бедаво, бераҳм* Севги (А. Умирова. Ватан ҳақида кўшиқ). Элас-элас, олис-олис соғинчлар, *ҳижрону фироқ, айрилиқ ва дийдор* манзиллари (С. Сайид. Олис ишқ соғинчи ёхуд ватан сурати). *Неверная, неизлечимая, безжалостная* любовь (А. Умирова. Песня о Родине). Элас-элас, далекие тоски, *эмиграция и разлука, разлука и дружеские* обращения (С. Сайид. Тоска по далекой любви или картина родины).

**Диатеза.** В этой стилистической фигуре смысл подтверждения исходит из двух негативов: Эҳтимол турли даврлардаги аёл-

ларнинг *на ташқи гўзаллик, на ички гўзаллик*да жинсдошларидан ажралиб турган эркаклар учун бунчалик қурбонликка боришганини тушунманганингиз ҳам бор гапдир (Ш.Ортиқов. Истеъдод балоси). *На бировнинг юзини хотир қиламан, на бировнинг кўзини* (У. Азим. Рауф Парпи). Возможно, вы не понимаете, что женщины разных эпох шли на такие жертвы ради мужчин, не отличавшихся от сверстников *ни внешней красотой, ни внутренней красотой* (Ш. Артиков. Бич таланта). Я не помню *ни чьего-то лица, ни чьих-то глаз* (У. Азим. Рауф Парпи).

**Ономатопея** – это прием художественного подражания звуковым явлениям в поэтической речи, стихах или прозе [4, 65]. Например: Умрзоқ ака шундай мисолларни айтар экан, менинг юрагим *гурс-гурс* уриб, жуда тўлқинланиб кетдим: ҳе, ёшлигимизни эслаيمان... Шу Барбад ҳам ўша ясаган асбобини ўнглаб чалса тингловчилар йиғлашган-а, тескарисини чалганда тингловчилар *қаҳ-қаҳ* отиб кулаверишган-а... Жўрабек ака саволга жавоб бериш учун “*дик- дик*” этиб ўрнидан тураётгани (Ш.Бўтаев. Шукур, айтчи, бухаёт нима?). Когда брат Умрзак рассказывал такие примеры, мое сердце колотилось, и я был очень взволнован: эй, я вспоминаю нашу молодость... Когда этот Барбад играл на инструменте, который он сделал правильно, публика плакала, а когда он играл наоборот кругом публика залилась смехом...

...Жорабек вставал, чтобы ответить на вопрос (Ш. Ботаев. Шукур, скажи, что это за жизнь?).

В приведенных примерах гармония звуков в словах типа *гурс-гурс, ках-ках, дик-дик* возводится на уровень не только гармонии звуков, но и лингвопоэтического фактора, определяющего стиль сочинения. В диссертации проанализированы такие фонетические приемы, как **аллитерация**, ассонанс, анафора, эпифора, акромнограмма, звукоподражание, парономазия, которые использовались для повышения художественной ценности текстов сочинений, усиления их эстетического воздействия, показа гармоничности соотношения формы и содержания, создавать тексты со своей внутренней музыкой, выражать мысли писателя о человеке, наблюдениях, глубоких суждениях, его личная точка зрения служила выражению мыслей философа в своеобразной художественной форме.

Итак, анализ жанра эссе, являющегося одним из художественно-публицистических жанров, с лингвокультурологической точки зрения также помогает выявить специфические стороны этого

стиля. Каждый язык возникает и развивается как продукт определенной культуры. Каждый текст отражает традиции, обычаи, нормы национального общения, символы и стереотипы народа, к которому принадлежат говорящие. На основе соотношения между языком + культурологией, культурологией + языковыми моделями признается объект изучения лингвокультурологии. С этой точки зрения создатели текстов эссе являются, с одной стороны, представителями и пропагандистами той культуры, к которой они принадлежат, а с другой стороны, они создают через слово национально-культурный образ своего народа.

#### ЛИТЕРАТУРА:

1. Каримов С. Ўзбек тилининг фонетик стилистикаси. – Самарқанд: СамДУ нашри, 2016. – Б. 11.
2. Ўзбек тили стилистикаси. – Тошкент: Ўқитувчи, 1983. – Б. 28.
3. Каримов С. Бадиий услуб ва тилининг ифода тасвир воситалари. Ўқув қўлланма. – Самарқанд: СамДУ нашриёти, 1994. – Б. 55–63.
4. Ҳотамов Н., Саримсоқов. Б. Адабиётшунослик терминларининг русча-ўзбекча изоҳли луғати. – Тошкент: Ўқитувчи, 1983. – 376 б.
5. Сабитова З.К. Лингвокультурология. Учебник. – М: Флинта: Наука, 2013. – С. 10.
6. Караулов Ю.Н. Общая и русская идеография. – Москва: Наука, 1976.
7. Ҳамроева Ё. Ўзбек тилининг идеографик луғатини тузиш тамойиллари: Филол. фан. номзд. дис... – Тошкент, 2009. – Б. 133–134.
8. Маклакова Е.А. Национальная языковая специфика и лингвокультурологические особенности семантики слова: <http://journals.uspu.ru>

## УДК 81-11

## ‘ZBEK XALQ LATIFALARI

*Safarova Nigora Oxunjonovna<sup>1</sup>, Tuyeva Zulfiya Nozimjonovna<sup>1</sup>,  
Zayniddinova Munisa Ibrohim qizi<sup>2</sup>*

<sup>1</sup>*Buxoro davlat pedagogika instituti.*

<sup>2</sup>*Buxoro innovatsiyalar universiteti 2- bosqich magistranti  
nigor.safarova@mail.ru; тел.: (99891)445-20-78*

## ПРО УЗБЕКСКИХ НАРОДНЫХ АНЕКДОТОВ

*Н. О. Сафарова<sup>1</sup>, З. Н. Туева<sup>2</sup>,  
Зайниддинова Муниса дочь Иброхим<sup>3</sup>*

<sup>1</sup>*Бухарский государственный педагогический институт*

<sup>2</sup>*к.п.н., доцент Бухарского государственного педагогического  
института, преподаватель*

<sup>3</sup>*учитель начальных классов, магистр 1-курса*

*Бухарского инновационного университета  
nigor.safarova@mail.ru; тел.: (99891)445-20-78*

## ABOUT UZBEK FOLK JOKES

*Safarova Nigora Okhunzhonovna<sup>1</sup>, Tuyeva Zulfia Nozimjonovna<sup>2</sup>,  
Zayniddinova Munisa daughter of Ibrohim<sup>3</sup>*

<sup>1</sup>*candidate philology of sciences, professor  
Bukhara state pedagogical institute*

<sup>2</sup>*Ph.D., Associate Professor,*

*Bukhara State Pedagogical Institute, teacher*

<sup>3</sup>*primary school teacher, 1st year master  
of Bukhara Innovation University*

*nigor.safarova@mail.ru; тел.: (99891)445-20-78*

O'zbek folklori epik janrlari orasida xalq latifalari g'oyat omma-viyiligi va hozirjavoblighi bilan ajralib turuvchi mustakil an'anaviy janr hisoblanadi. «Latifa» so'zi arabcha «latif va lutf» so'zlaridan olingan bulib, nozik qochirimli, mayin kinoyali kichik kulgili hikoyani anglatgani sababli istilohiy mohiyat kasb etgan. Shuni alohida ta'kidlash joizki, fors-tojik tilida bunday kulgili epizodik hikoyalarni maxsus to'plab, 1223 yilda «Javome'ul-hikoyot va lavome'ur-rivoyot» (Hikoyatlar to'plami va rivoyatlar ziyosi) nomi ostida alohida kitob sifati tartib bergan Muhammad Avfiy Buxoriy (1172/1176–1233/1242) ham ularni latifa demagan, balki hikoyat degan.

Sirasini aytganda, aslida hikoyat deyilsa-da, davrlar o'tishi bilan ma'lum g'oyalarni ifodalashga moslashtirilib qayta ishlanishi nati-

jasida latifa namunalariga aylangan bunday asarlar Majduddin Muhammad binni Adnom Surhakatiyning «Tarixi Turkiston», Abu Ja'far Muhammad binni Jarir Tabariy (838–923)ning «Tarixi Tabariy», Abu Rayhon Beruniy (973–1048)ning «Osor-ul-boqiyah», «Kitob-ul Hind», saljuqiylar vaziri Nizomulmulk (1018–1092)ning «Siyosatnoma», Unsurul Ma'aliy Kaykovus (1021-vafot yili noma'lum)ning «Qobusnoma», Nizomi Aruzi Samarqandiyning (XII asr) «Chahor maqola», Zahiriy Samarqandiy (XII asr)ning «Sindbodnoma», Hakim Sanoiy (XII asr)ning «Hadiqat-ul haqiq» (Haqiqatlar bog'i), Sa'diy Sheroziyning «Guliston», Zayniddin Vosifiyning «Badoe' ul-vaqoe'» (Badiiy voqealar), Davlatshoh Samarqandiy (1435/36–1495)ning «Tazkirat ush shuaro» (Shoiralar bo'stoni), Xondamirning «Makorim ul-axloq» (Yaxshi hulqlar) va «Habib us-siyar» (Yoqimli xislatlar), Mirxondning «Ravzat us-safo» (Safo bog'i) va boshqalarda istagancha keltirilgan esa-da, barchasida hikoyat istilohi istifoda etilgan.

Faqat Nosiruddin Burhoniddin Rabg'uziyning 1310 yilda intihosiga yetkazilgan «Qisasi Rabg'uziy» asaridagina latifa istilohi ostida ikkita voqeot bayoni berilgan. Voqealar bayoni garchi jiddiy bo'lsa-da, vale kinoyaga moyilroq ohangga ega. Ana shu dalillarning o'ziyoq latifa istilohining nisbatan keyinroq, aniqrog'i, 1532–33 yillarda Faxriddin Ali Safiy tomonidan tartib berilgan «Latoyif ut-tavoyif» (Turli toifadagilarning latifalari) kitobining el orasida keng shuhrat qozona borishi tufayli nainki adabiy, balki xalq ijodida ham taomilga kira boshladi, istilohiy mohiyat kasb etgan holda ilmiy muomalada barqarorlasha borganini kuzatish mumkin.

To'g'ri, xalq orasida XVI asrdayoq keng shuhrat qozongan bu kitob goh «Latoyif ut-tavoyif», goh «Latoyifuz-zaroyif», gohida esa shunchaki «Latofatnoma» tarzida atalib kelingan esa-da, uni o'zbek tiliga o'girgan tarjimon Hoji Sotimxon Xoja In'om Munavvar qayd etganidek: «Latoyif so'zi garchi latifa so'zining ko'pligi bo'lsa ham, «latoyif» so'zi garchi latifa mazmunini inkor etmasa ham, «latoyif» – bu latifa emas. Latoyif turli tuman dilkash, maroqli, xarakterli, xullas, kimmatli voqea va hodisalarning latifanamo ixchamlikdagi bayonidir. Shuning uchun ham bu hikoyalarni latifa deb emas, balki latofatlar, lutflar deb atash mantiqqa yaqindir. Chunki bu latoyiflarni o'qigan kishi xuddi latifalarni o'qigandek ham qah-qah solib kuladi, ham chuqur fikrga toladi, sidqidildan yum-yum yig'laydi ham. Safiyning latoyiflari orasida, garchi xalq latifalariga aylanib ketgan hikoyatlar uchrab tursa ham, bu asar, eng avval, chindan ham dog' qiluvchi latifalar majmuidir». Bu fikrdan uch narsa oydinlashayotir: birinchisi va muhimi,

latoyif soʻzining latifa mazmunini anglatishi va koʻplikda qoʻllanilganligi; ikkinchisi, boshda u hikoyat maʼnosida, aniqrogʻi, hazilkashlikka yoʻgʻrilgan lutflardan iborat kulgili epizodik hikoyalarni anglatganligi; nihoyat, uchinchisi esa, Ali Safiy latoyiflarining ancha-munchasi bora-bora xalq latifalariga aylanishi va shu jarayonda latifa xalqning shu xildagi zarofatli ogʻzaki epizodik hikoyalarni ifodalovchi istilohiy mohiyat kasb etganligi. Shu zaylda, aytish mumkinki, xalq ijodiyotida latifa janri garchi taxminan IX–XI asrlarda paydo boʻlib, taraqqiy eta borganligi va Mahmud Qoshgʻariy maʼlumoticha, turkiy xalqlar orasida saw (sav) deb atalganligi maʼlum esa-da, folkloriy hodisani ifodalovchi arabcha «latifa» istilohi asta-sekin uni siqib chiqara boshladi va XIII asrdan muomalaga, kira borib, XIU–XUI asrlar davomida toʻla barqarorlashuvga erishdi. Buni ozar folklorshunosi V.Valievning «latifa» soʻzi bir termin kabi ozarbayjon foʻlklorida keyinroq paydo boʻlgandir. Umuman, xalq orasida bu termin ishlatilmagan. Xalq orasida bu xildagi namunalar «bazama» deyiladi. Holbuki, folklorshunoslar arab tilidagi kalima «latifa» soʻzini bir termin kabi qabul qilganlar. Bizga qolsa, latifa oʻrnida «bazama» ishlatilsa, yanada munosib boʻlardir, – degan fikri ham maʼlum maʼnoda tasdiqlaydi.

Alisher Navoiy ham latifa oʻrnida «bazla» va uni aytuvchini «bazlagoʻy» tarzda qoʻllaydi:

Foydasiz qilmasa zohir karam,  
 Bermasa bir bazlaga yuz ming diram.  
 («Hayratul-abror», 100-b) Va:  
 Zarofat holatida bazlagoʻlar,  
 Atoridqa qilib har dam gʻulular.  
 («Farhod va Shirin», 30-b) va boshqalar.

Dastlabki paytda foydasiz narsalar uchun qalam «zohir etish» sarflashga hojat boʻlmaganidek, buhuda, mantiqsiz, kulgu qoʻzgʻamaydigan «bazla» (latifa)ga «yuz ming diram» sarflamoqqa ne hojat deyilsa, ikkinchi baytda nozikfahmlilik, zariflik kayfiyatida «bazlagoʻ» (latifagoʻy)lar aytayotgan quvnoq latifalardan kah-qahalar avji Atorud (Yupiter)gacha yetardi degan maʼnoni uqish mumkin. Koʻrinayotirki, «bazama» va «bazla» soʻzlari yagona «baz» oʻzagi asosida yaratilganligi bilan oʻzaro hamohangdir.

Shunga qaramay, turkiy tildagi «bazla» va «bazama» istilohlari oʻrnida arabcha «latifa» istilohi ilmiy taomilga oʻzlashib ketdi. Biroq xalq orasida «hikoyat», «mutoyiba» istilohlarini «latifa» maʼnosida qoʻllash hamon davom etayotir.

Latifa umumfolklor hodisasi bo'lib, ruslarda anekdot, polyaklarda perekklad, italiyanlarda fassetsiya, frantsuzlarda fablio va nemislarda shvank deb yuritiladi. Biroq unutmazlik kerakki, ruslardagi «anekdot» mazmunan «latifa» istilohiga teng ekvivalent bo'la olmaydi. Chunki «anekdot» so'zi «chop etilmagan asar» yoki «chop etilishi mumkin bo'lmagan asar» ma'nosida bo'lib, uni ilk bor vizantiyalik tarixchi Prokopiy Kesariyskiy qo'llagan edi. U o'zining Vizantiya imperatoriga bag'ishlab yozgan kitobini «Anekdot» deb nomlagan edi. Hozirgacha fanda «Taynaya istoriya» («Yashirin tarix») nomi bilan mashhur bo'lib kelayotgan bu kitobni yozishda maxfiy hujjatlardan foydalanilganligi tufayli uni shunday atash zaruriyatga aylangan edi. Anekdotlarda ham tagdor qochirimlar, piching va kesatidlarni ifodalovchi ma'nolar borligi sababli XVIII asrdan e'tiboran xalq og'zaki nasrida latifaga o'xshagan hodisani ifodalovchi istiloh sifatida qo'llana boshlandi.

O'zbek xalq latifalari favqulodda boshlanishi va favqulodda naql etilishi bilan e'tiborni tortadi. Ularning asosida yengil mutoyiba va o'tkir hajv yotadi. Ular hajman muxtasar, mazmunan ixcham bo'ladi. Asosan, bir yoki ikki-uch epizoddan tashkil topadi. Yakka ijroda aytiladi. Xalq hajviyoti va mayin yumorga asoslangan latifalarda hayot haqiqati uydirma, so'z o'yini, kinoya va qochimlar vositasida aks etiriladi. Latifalar yechimida satira va yumor hal qiluvchi rol o'ynaydi. Yechim tasodifan ko'tarilgan kulgu, kah-qaha yoki so'z o'yiniga asoslanadi. So'z o'yini latifa kulminatsiyasini tashkil etadi.

Latifalarda nutqning dialog shakli keng qo'llanadi. Ular ijtimoiy-siyosiy va maishiy hayotiy ma'no kasb etadi. Kishilar fe'l-atvorida uchraydigan salbiy axloqiy xususiyatlarni inkor etib, soddalik, go'llik, kaltafahmlik, savodsizlik, omiylik, merovlik, g'afloyinlik, befarosatlik ustidan kulgi uyg'otadi. Odamlarni hushyorlikka chorlaydi.

Latifalar ortiqcha tasviriylikdan xoliligi bilan ham ajralib turadi, ularda birorta ortiqcha so'z yoki detal uchramaydi. Binobarin, kompozitsiyaning soddaligi, syujetning ixcham va lo'ndaligi, yechimning favqulodda yuz berishi va tasodifan ko'tarilgan kulgiga asoslanganligi, yagona markaziy qahramon obraziga bog'liqligi latifaning doimiy va barqaror janriy belgilari sanaladi.

Latifalarning paydo bo'lishi masalasi folklorshunoslikda anchayin bahslarga sabab bo'lib kelayotir. Chunonchi, A.P.Peltser bu masalada A.Potebnya va A.Gluxovlar fikrini quvvatlab, latifalar ertaklar, massallar yoki naqlarning qayta ishlanishi, ixchamlashtirilishi, qochirimli ifodaviylik asosida kulgi uyg'otish darajasiga keltirilishi natijasida

paydo bo'radi deb hisoblaydi . O'zbek folklorshunosi H.Razzoqov ham «Satira, yumor xarakteridagi hajviy ertaklarning kichik hajmdagi-lari tobora ishlanib, ixchamlanishi natijasida bora-bora latifa janriga aylanib ketganidek, ulardagi epik komik qahramon Kallar, Aldarko'ssa, Mashrab, Nasriddin afandi obrazlarining takomillashib, badiiy tip darajasiga ko'tarilishi uchun yordam beradi», deya yuqoridagi fikrga tarafdorlik qiladi . Bu fikrni «Latifaning janr sifatida shakllanishida ertak, rivoyat, lof, askiya va maqollarning ta'siri kuchli bo'lgan»ligini ta'kidlab, prof. B.Sarimsoqov ham tasdiqlaydi .

Tojik folklorshunosi S.Maxdiev bu qarashlarga «har bir janr o'zicha mustaqil bo'lib, boshqa vositalarsiz, o'ziga xos qonuniyatlar asosida ijod etiladi. Binobarin, latifaning paydo bo'lishini ham folklorning u yoki bu janri bilan bog'lab qo'yish durust emas», deya e'tiroz bildiradi. Biroq bu fikri bilan u o'z tadqiqotining 31-38-sahifalarida ifoda etilgan latifa va ertak (afsona), latifa va aksiya, latifa va maqol munosabatlariga to'xtalganida o'z-o'zini inkor etganday yo'l tutadi. Aniqrog'i, bu janrlarning bir-biriga ta'siri va uzoq muddat davomida og'izdan og'izga o'tib sayqallanishi natijasida bir-biriga o'tib turishiga oid o'z karashlarini rad etayotganiga ahamiyat bermay, afandining podshoga nashvati sovg'a qilishi, xotini qovoq sovg'a qiling deganiga qaramay, nashvatini olib borishi, podshoning esa uni o'sha nashvati bilan urib jazolashi, shunda afandi «Xayriyat, qovoq sovg'a qilmaganim», deya o'z-o'ziga tasalli berishi haqidagi latifaning genetik ildiziga nazar sola turib, G.Kuzmichevskiyning uni milodiy II asrda paydo bo'lganligini qayd etganini, biroq u Gerodotning «Tarix» kitobidan rivoyat shaklida eramizgacha bo'lgan V asrda mavjud bo'lganligini aytadi va o'sha namunani aynan keltiradi.

«Afandining sovg'asi» va «Birinchi anjir» latifalarida baliq detali bodring va anjir bilan almashinib, afandining jazolanishiga sabab bo'lgan. Shuningdek, o'zbek folklorida afandining xotini yuvgan kirni otib, so'ngra uning ichida o'zining bo'lmaganidan quvonishini yoki bozori kasod bo'lgan sigirini dallol maqtovidan so'ng xaridorlar o'rtasida talashga aylanganidan iyib, sotmay qaytarib olib kelishi haqidagi latifalar aslida xuddi shu motiv ta'sirida yuz aga kelgan. Shunday misollar munosabati bilan latifalar paydo bo'lishining ikki xususiyatini ta'kidlashga to'g'ri keladiki, S.Maxdiev buni ham inkor etadi:

Birinchisi – «Polikart uzugi rivoyati» davrlar o'tishi bilan latifa janriga o'tgach, ortiqcha tafsilotlardan tozalangan. Demak, latifaning paydo bo'lishida janrdan janrga (ayni misolda: rivoyatdan latifaga) ko'chishdan iborat folklorga xos migratsion jarayon sodir bo'lganki,



buni inkor etish mumkin emas. Bunday hodisa har bir milliy folklorida o'z ichki imkoniyatlari doirasida ham kechishi an'anaviydir.

Ikkinchisi – bu latifa boshqa xalqlarga o'tib, shunchaki baynalminal folkloriy hodisagagina aylanib qolmagan. Aksincha, milliy folklorlarga xos xususiyatlar bilan sug'orilib, jahon xalqlari og'zaki ijodiyotining o'zaro ta'siri va samarali hamkorligi hosilasiga aylangan. Aniqroq qilib aytilsa, yunon va rim latifasidagi baliq detali tojik latifasida nashvatiga, o'zbek latifasida bodringga almashinib, milliylik ruhini kuchaytirgan, natijada qadimiy yunon rivoiyati zamirida betakror o'zbek va tojik latifalari yuzaga kelgan. Yuqoridagi mulohazalardan ayonlashadiki, latifalar tubandagi uch omil asosida paydo bo'lib, rivojlanib kelmoqda: Bu omillarning barchasiga, albatta, ijtimoiy hayotda sinfiy munosabatlarning kuchaya borishi va feodal zo'ravonlikning chuqurlashuvi ma'lum darajada ta'sir ko'rsatgan. Zero, latifalarda nisbatan ta'sir ko'rsatish, norozilik bildirish kayfiyatlarining mushtarak yetakchi tamoyilga aylanganligi sababini shunday izohlash mumkin.

Birinchi omil – u yoki bu ayrim mashhur tarixiy shaxslar xarakteri, xatti-harakati va ko'p qirrali ijtimoiy-siyosiy va amaliy faoliyatining yo ijobiy, yo salbiy badihaviy baholanishi oqibatida yuzaga kelgan latifalar. Aflotun, Iskandar, Arastu, Luqmon, Buqrot, Bahromgo'r, Buzurjmehr, Juho, Bahlul, Tufayliy, Abunuvvos, Husrav Parvez, Anushervon, Birbal, Mirali, Mushfiqiy, Mashrab, Balakirev, Skovoroda, Kuyruchuk, Behnazar, Kamina va boshqalar nomlari bilan bog'liq latifalar ana shunday mohiyatga ega.

Ikkinchi omil – xalq asrlar davomida o'z quvnoq ruhiga va idealiga mos kelgani uchun satirik va yumoristik ertaklarning to'qima qahramonlari Kal, Aldarko'sa, Ko'sa va Laku Paklarning nomlari bilan aloqador talay latifalar ijod qilgan. Bunday latifalarning ayrimlari shu qahramonlarga oid ertaklarning qayta ishlanishi, ixchamlashtirilishi negizida bunyod bo'lgan, yana bir qismi esa tarixiy shaxslar obrazlariga daxldor latifalarning Kal, Aldarko'sa, Ko'sa va Laku Paklarning nomlariga o'tishi yoki o'tkazilishi tufayli yuzaga kelgan.

Uchinchi omil – ayrim urug'lar, yo ayrim shahar yoki qishloqlarda yashovchi qavmlarning fe'l-atvori, tabiati va hodisalariga, odamlarga, hayvonlarga, qushlarga, hasharotlarga, mehnat va hayotga soddavor munosabatlarining badihaviy bahosi sifatida yuzaga kelgan latifalar. Bu toifaga gabrovoliklar (Bulg'oriya), poshexoнbeliklar (Rossiya), mil'daliklar (Olmoniya), gotemliklar (Angliya), qazvinliklar (Eron) qahramoni bo'lgan latifalar mansub. Prof. V.Valievning guvohligicha, bunday holatni Ozarbayjon kent va dahalari ahliga daxldor latifalar

misolida ham kuzatish mumkin. (Qarang, eslatilgan asar, 290-6.) O'zbek va tojik xalqlarida ham bu silsilada qadim zamonlardan beri latifalar to'qish va aytish an'anasi mavjud. Bu an'ahanuzgacha so'nmagan, aksincha, faol rivoj topib kelayotir. Chunonchi, hozirgi O'zbekiston hududida Buxoro muzofotida shirinilar, emarilar, Samarqand shahrida mahallotilar, Xorazm viloyatida xonqalıklar,

Farg'ona vodiysida g'irvonliklar, oltiariqliklar, Qashqadaryo voahasida pochvonliklar, Surxon muzofotida g'urutliklar, Tojikistonda esa darvozliliklar, xufliklar, rumonliklar qaxramoni bo'lgan yumoristik latifalar bisyor. Bu guruhga mansub latifalarda joy yo urug' nomi umumlashtiruvchi mohiyatga ega bo'lgani bois latifa kahramoni nomi ham umumlashtiruvchi mazmun kasb etgani holda, asosan, lokal ijrochilik doirasida aytiladi.

Xullas, xalq latifalari qahramoniga aylangan Xoja Nasriddin afandi obrazi sarchashmasida ana shu ikki buyuk siymo – Xoja Nasriddin Oqshahariy va Xoja Hasan Nasriddin Tusiylar turadi. Biroq aynan shu tarixiy shaxslarning o'zi bo'lmay, balki asrlar davomida xalq tomondan badiiy qayta ishlangan va uning orzu- armonlarini ifodalovchi epik qahramon timsoliga aylangan. Uning nomi xalq latifalarida Mullo Nasriddin, Xoja yoki Xo'ja Nasriddin, Nasriddin afandi va Afandi shakllari uchraydi.

Xullas, latifalarda Nasiriddin afandi va boshqa personajlar nutqi juda katta badiiy estetik vazifani o'taydi; qahramon va personajlar xarakteriga xos xususiyatlarni ochishga xizmat qiladi. Ayniqsa, afandi nutqidagi qochirimlar, tagdor kinoyalar, kulgi yaratuvchi ko'p ma'noli so'zlar uning o'tkir didli, so'zga chechan, qiziqchilikka moyil tabiatini belgilaydi. Shu fazilati bilan hajman siqiq va ixcham bu latifalar xotiraga tez o'rmashib o'zlashadi va og'izdan-og'izga o'tib, sayqal topgan holda elga kulgi ulashib yashayveradi.

#### ADABIYOTLAR:

1. Dildora Yusupova, Dilorom Yuldasheva. Halima Xudoyberdiyeva she'riyat lingvopoetikasi.– Germaniya: “Globe Edit international Publishing Group” ISBN: 978-620-0-60879- 6.– 2020. – 125 p.

2. Юлдашева Д. Н. Роль молчания в речи// Образование и инновационные исследования. № 1 (2020): <https://interscience.uz/index.php/home/article/view/31>

3. Сабитова З. К. Лингвокультурология. Учебник. – М: Флинта: Наука, 2013. – С. 10.

УДК 81-11

**M.KOSHG‘ARIYNING “DEVON”IDAGI BA’ZI ETNONIMLAR  
XUSUSIDA*****M. Safarova, Z. Sattorova****BuxDU*

maftunasafarova@gmail.com, zsattorova@gmail.com

Ushbu maqolada M.Koshg‘ariyning “Devoni lug‘otut turk” asarida keltirilgan turkiy qabila va urug‘larning ba‘zilari xususida fikr yuritilgan. Ularning leksik-semantik, etimologik xususiyatlari bayon etilgan.

**Kalit so‘zlar.** Etnonim, xalq, millat, qabila, urug‘, nom, o‘g‘uz, qiniq, qayig‘, bayunlar kabi.

**ON SOME ETHNONAMES IN “DEVON” BY M.KOSHG‘ARIY*****M. Safarova<sup>1</sup>, Z. Sattorova<sup>2</sup>****<sup>1</sup>(Teacher of BuxSU)**<sup>2</sup>(4 th stage student of BuxSU)*

maftunasafarova@gmail.com

This article discusses some of the Turkic tribes and clans mentioned in M.Koshg‘ariy’s “Devoni lug‘otut turk” work. Their lexical-semantic, etymological features are described.

**Keywords.** Such as ethnonym, people, nation, tribe, clan, name, oguz, kinik, kayik, bayuns.

**О НЕКОТОРЫХ ЭТНОНИМАХ В «ДЕВОН» М. КОШГАРИЙ*****М. Сафарова<sup>1</sup>, З. Сатторовва<sup>2</sup>****<sup>1</sup>(преподаватель БухГУ)**<sup>2</sup>(студентка 4 курса БухГУ)*

maftunasafarova@gmail.com

В данной статье рассматриваются некоторые тюркские племена и роды, упомянутые в труде М. Кошгарий «Девони луготут тюрк». Описаны их лексико-семантические, этимологические особенности.

**Ключевые слова.** Такие как этноним, народ, нация, племя, род, имя, огуз, киник, кайик, баюны.

Ma‘lumki, Mahmud Koshg‘ariyning “Devoni lug‘otut turk” asari qardosh turkiy tillar haqida yozilgan qomusiy asar hisoblanadi. Asarda XI asrning 2-yarmida Markaziy Osiyoda va G‘arbiy Xitoy hududida

yashagan turkiy urugʻ va qabilalar, ularning turmush tarsi, ijtimoiy kelib chiqishi, tili, dini, tarixiga oid maʼlumotlar berilgan.

Muallif bu asarni xalq orasida yurib, xalqlar va qabilalar bilan bevosita suhbatlashib yozganini taʼkidlaydi.

U qabila va urugʻlarning turmush tarzi, anʼanalari, suhbatlarini kuzatishi natijasida “Devonu lugʻotut turk” asari dunyoga kelganligini eʼtirof etadi. Olim asarida koʻpgina qabila va urugʻ nomlari keltirib oʻtgan. Masalan, oʻgʻuzlar, qiniqlar, qayigʻlar, bayunlar, ivalar, salgʻurlar, avsharlar, begtililar, bukduzlar, bayatlar, yazgʻirlar, eymurlar, qoraboʻluklar, igdarlar, uragirlar, tutirqalar, ulayundlugʻlar, toʻgarlar, pachanaklar, chuvaldarlar, chepnilar, charuqlugʻlar, yagʻmo, tuxsi, yamak, argʻu, chigil, udunlar (Xoʻtanda yashovchilar), ogʻrak, (Qorayigʻach deb atalgan chegarada yashovchi bir turk aymogʻi), elkabuloq, arabut (Tamar hindiy), aramut (Uygʻurlar yaqinida yashovchi bir turk qabilasi), qarluq, kanjak, sumlum (Turk tilini bilmaydiganlar), sumlim tat (turk tilini bilmaydigan fors), argʻu, bulaq va boshqalar.

Biz yuqorida sanab oʻtilgan qabila nomlarining izohini topishga harakat qildik. Aytish joizki, 1960-63-yillarda S.Mutalibov tomonidan nashrga tayyorlangan “Devonu lugʻotut turk” asarining izoh qismida asarda keltirilgan barcha qabila nomlariga alohida toʻxtalib oʻtilmagan. Faqatgina oʻgʻuz, chigil, yagʻmo, tuxsi, qarluq, kanjak, argʻu kabi qabilalar haqida maʼlumotlar keltirilgan.

*Oʻgʻuz qabilasi.* Oʻgʻuz qabilasi eng qadimgi va nufuzli qabilalardan biri sifatida Oʻrta Osiyo xalqlari va ularning tillari tarixuda alohida oʻrin tutadi. Shu sababdan unga tarixchi olimlar va tilshunolar juda ilgari zamonlarda jiddiy qiziqib keldilar. Bu masala yuzasida muhim ilmiy tekshirish amalga oshirildi. Ayniqsa, bu urugʻ etimologiyasiga juda koʻp olimlar qiziqish bildirganlar. Chunonchi, V.Bang, V.Radlov, A.N.Bernshtam, D.Sinor, G.Ramsted ishlari shu jumladandir. Biroq bu tekshiruvlarda masala toʻla hal etilmagan. Hatto Oʻgʻuzxon toʻgʻrisida ayrim uydirma, afsonalar ham bugungi kunga qadar saqlanib qolgan. Oʻgʻuz qabilasi haqida “Risolutus safo” va Bichurin asarida ham maʼlumot keltirilgan. Lekin bu asarlarda ham masala toʻla toʻkis yoritilgan emas. Shuning uchun bu haqida yetuk tadqiqotchi M.Koshgʻariyning maʼlumoti benihoya muhimdir. M.Koshgʻariy oʻz asarida bu soʻzning etimologiyasini aniqlashga emas, balki uning juda qadimgi ota-bobolarga borib taqalishini aytish bilan cheklangan. Bunday cheklanish tasodif emas, albatta. Aksincha, bu hol oʻgʻuz soʻzining antroponim sifatida yuzaga kelgani, soʻng maʼlum sharoitda qabila nomiga aylangani bilan bogʻliq. M.Koshgʻariy oʻz asarida oʻgʻuzlarn-

ing u davrda tutgan mavqeyi, tili, kasb-u kori, urug'-aymoqlari, turar joylari va boshqalarni aniq faktlar asosida yoritib beradi.

*Chigil qabilasi.* Ushbu qabila "Devon"da eng asosiy o'rinni ishg'ol qiladi. Mahmud Koshg'ariyning fikricha, o'g'uzlar Jayhundan quyi Chingacha bo'lgan xalqlarni chigillar deb hisoblar edilar. Devonda chigil ismi Iskandar Zulqarnayn tomonidan berilgani aytilgan. Chigillar eng eski xalqlardan biridir. Navoiy ma'lumoti ham M.Koshg'ariy fikrlarini tasdiqlaydi. U Iskandar Zulqarnayn yurishida, Movaroun-nahrda chigil, yag'mo nomlarinigina tilga oladi:

Chigil birla Yag'modan aylab ubur,  
Nechunkim chamandin sabovu dabur. ("Xamsa"dan)

Ibnul Asir tarixi qoraxoniylar davrida Samarqandga qilingan hujumda chigil askarlari boshchilik qilganini yozadi. Haqiqatda ham chigillar O'rta Osiyodagi eng eski xalqlardan ekanini devondagi ko'p ma'lumotlar isbotlaydi. Mahmud Koshg'ariy chigillarning o'g'uzlar bilan juda qadimdan to'qnashib kelganini, ularning bir-biriga qarshiligi devon yozilgan davrda ham davom etishini yozar ekan, o'g'uzlar tili-dagi kamchiliklarni ham ko'rsatib beradi. Shu bilan birga ularga qarshi til jihatdan ishonchli qabilalar boshida chigillar turadilar.

*Yag'mo qabilasi.* Yag'mo Tiroz yaqinidagi shahar. U yerda yashovchi qabila ham yag'mo deb atalgan. Uni qora yag'mo deb ham yuritilgan edi. Yag'molarning bir qismi Ila vodiysining o'ng qirg'og'ida, ba'zilari jikillar bilan birga yashar edi.

"Devon"dagi ma'lumotlar yag'molarni jangovar xalq sifatida tanitadi. "Devon"da ularning urf-odatlariga va til xususiyatlariga doir ma'lumotlar ancha bor. Yag'molarda savdo-sotiq ishlari ancha taraqqiy qilgan edi. Ulardagi saroybonlarning ish usuliga doir ma'lumot ham xarakterlidir. Ularning tillari jikil va tuxsilar tili bilan yaqin bo'lsa ham, o'ziga xos ayrim xususiyatga ega bo'lgan.

*Tuxsi qabilasi.* Tuxsi qabilasi haqida asarda ma'lumot yetarli. Ular e'tiborli qabilalardan sanaladi. Tuxsilarning tillari yolg'iz turkcha bo'lib, asosan, Qayos shahrida yashaydilar (Qayos 3ta katta rayondan iborat bo'lgan shahar: 1) Soblig' Qayos; 2) O'rung Qayos; 3) Qora Qayos).

Ular til jihatdan chigil, yag'moga yaqindir. "Devon"da ularning yashash joylari ham bir-biriga yaqin, aralash bo'lgani to'g'risida ma'lumotlar keltirilgan.

Mahmud Koshg'ariy "Devon"i 3-tomining 52-betida Bista so'zini izohlar ekan, bunday deb yozadi: Bista chetdan kelgan savdogarni

uyga qoʻndiradi, mollarini sotib beradi, sotguncha oʻzini ham, molini ham boqadi; ketar vaqtida mol boshiga maʼlum miqdorda haq oladi. Men buni, – deydi, Mahmud Koshgʻariy, chigil, yagʻmo, tuxsilarda oʻz koʻzim bilan koʻrdim.

Tuxsilar haqida asarda yana koʻp yerda maʼlumotlar bor boʻlib, ulardan ham Tuxsilar tilda, shaharda, feʼl-u odatda chigil, yagʻmolarga yaqin, lekin ularga nisbatan kichikroq qabila ekani koʻrsatiladi.

*Qarluq qabilasi.* Qarluqlar turklarning bir turi boʻlib, vabarlilar (koʻchmanchilar) va oʻgʻuzlardan boshqadir. Ular aslida turkmanlardandir.

Mahmud Koshgʻariy qabilalarni bayon qilgan maxsus faslda qarluqlarni koʻrsatmaydi. Lekin qarluq asarning toʻqqiz yerida uchraydi.

Bu oʻrinda keltirilgan maʼlumotlar ularni yetakchi qabila tariqasida emas, aksincha, boshqa qabilalarga tobe, ikkinchi darajali qabila, turkmanlarning bir toifasi sifatida koʻrsatadi.

“Devon”da qarluqlarning Choʻgʻilan degan boshliqlari ismidan boshqa bir muhim maʼlumot berilmagan. Ularning maxsus shaharlari ham koʻrsatilmagan. Til jihatdan qarluqlarning bir qismi oʻgʻuzlarga, bir qismi yagʻmo, chigillarga qoʻshilishi mumkin, Tarixiy asarlarda qarluqlar nomi boshqa qabilalarning nomlarig nisbatan mashhurroq ekanligini nazarga olsak, ularning ilgari mashhur boʻlgani va devon yozilgan davrda ikkinchi darajali qabilalarga birikib ketgani anglashiladi. *Kanjak qabilasi.* Kanjaklar ikki tilda soʻzlashuvchi qabilalar hisoblanadi. Shuning uchun ular tillarida ayrim kamchiliklar uchraydi. Kanjaklarning biror maxsus shaharlari xususida maʼlumot keltirilmagan. Qoshgʻarda kanjakcha soʻzlashuvchi sulola borligi haqida aytilgan “Devon”da kanjaklarga nisbatan berilgan maxsus soʻzlar asosan tokchilik, bogʻchilikka oidligi xarakterlidir. Masalan: bulduni uzum – mayiz saqlanadigan joy; boʻshunjak – uzum urugʻi; nabzi – uzum uzilgandan keyin tokda qolgan shingil-shapoq; lujnat – xirmon qilishda qishloq ahlari bir-biriga odam yuborib, hashar qilishi; tabchang – balanddagi toklarni kesishda ustiga chiqishga qoʻllaniladigan uch oyoqli asbob.

*Argʻu qabilasi.* Asardagi maʼlumotiga koʻra, argʻular Oʻzgand va Azgish shaharlarida yashar edilar. Argʻu til xususiyatlarining bir qismi hozirgicha yetib kelishi, ular juda keng hududda yashovchi, hozirgi Chimkentdan Toʻqmoqqacha boʻlgan joylarda va butun Fargʻona vodiysida yashagan bir qancha qabilalar orasida koʻp nufuzga ega boʻlgani va mamlakatni idora etish u davrda koʻproq ularning qoʻlida boʻlgani xususida fikrlar keltirilgan.

Asarda arg‘u qabilasining keyinchalik XI asrda mavqeyi ancha o‘zgarishi bilan bog‘liq ma’lumotlar keltirilgan. So‘nggi davr tarixiy asarlarida arg‘u qabilasi nomining uchramasligi ular mavqeyining o‘zgarishi bilan bog‘liqdir. Lekin arg‘ularning yuqorida ko‘rsatilgan hududlarda yashab kelgani ma’lum. Mahmud Koshg‘ariy arg‘ular bilir so‘zini bilir tarzida talaffuz qilishini ta’kidlaydi, ya’ni o‘tgan zamon fe’l yasovchi affiksni “u” unlisi bilan aytadilar: bardum, keldum kabi.

Xulosa qilib shuni aytish joizki, Mahmud Koshg‘ariyning “Devonu lug‘otut turk” asari turkiy xalqlar tili, dini, madaniyati, turmush tarzi xususida qimmatli manba bo‘lishi bilan juda ahamiyatlidir. Asarda ko‘plab qabila va urug‘ nomlarining atalish sabablari, semantik xususiyatlari, ularning turmush tarzi ifoda etilgan. Bu esa tilshunoslik, tarix, geografiya va boshqa fanlarning rivojlanishiga xizmat qiladi.

#### ADABIYOTLAR:

1. Кошғарий М. Девону луғотит турк. I том. – Т.: 1960. 499 б.
2. Dosimov Z., Egamov Kh. A brief annotated dictionary of place names. – Т.: Teacher, 1977. – В. 48.
3. Kiličev B.E. Onomasti cs // Study guide. – Bukhara, Durdona, 2023. – 180 p.
4. Kiličev B.E., Safarova M.Z. Toponyms Formed On The Basis Of Nation Names. International Journals of Sciences and High Technologies. Vol. 25 No. 1 February 2021, pp. 104–107.
5. Kiličev B.E., Safarova M.Z. Bukhara region’s typical toponyms transformed by means the names of nations. International Scientific Journal Theoretical & Applied Science, <http://t-science.org/> p-ISSN: 2308-4944 (print) e-ISSN: 2409-0085 (online) 2021.02.05.
6. Kiličev B.E., Safarova M.Z. The Concept of Totemism and Ethnonym. Central Asian Journal of Literature, Philosophy and..., 2022.
7. Safarova M.Z. Alisher Navoiy ijodida etnonimlarning qo‘llanilishi. Filologik tadqiqotlar: muammo va yechim” mavzusida xalqaro ilmiy konferensiya. 2021.

## УДК 81-11

## YOZUVLAR TARAQQIYOTIDA BOBURNING O'RNINI

*Saidov Yoqub Siddiqovich**Buxoro davlat universiteti*

saidovyqub@gmail.com

Maqolada Zahiriddin Muhammad Boburning arab alifbosiga asoslangan eski o'zbek yozuvini isloh etish borasidagi ijtimoiy faoliyati o'rganilgan, ilmiy-nazariy qarashlari talqin etilgan, "Xatti Boburiy" risolasi yozilishining ijtimoiy va ma'rifiy sabablari aniqlangan, uning ilmiy- tarixiy ahamiyati ochib berilgan, adibning turkiy tillarning talaffuz tabiatiga moslashtirish maqsadida bu yozuvni yaratganligi asoslangan hamda yozuvlar taraqqiyotidagi o'rni qisman belgilangan.

**Kalit so'zlar:** til, tafakkur, nutq, yozuv, arab alifbosi, turkiy yozuv, forsiy yozuv, tovush, harf, isloh, imlo, xat turi, eski o'zbek yozuvi.

The article examines the social activities of Zahiriddin Muhammad Babur in the reform of the old Uzbek script based on the Arabic alphabet, interprets his scientific and theoretical views, identifies the social and educational reasons for writing the treatise "Khatti Baburiy", and reveals its scientific and historical importance. It is based on the fact that the writer created this writing in order to adapt it to the pronunciation of Turkic languages, and its place in the development of writing is partially defined.

**Key words:** language, thinking, speech, writing, Arabic alphabet, Turkish writing, Persian writing, sound, letter, correction, spelling, letter type, old Uzbek writing.

В статье рассматривается общественная деятельность Захириддина Мухаммада Бабура по реформированию старой узбекской письменности на основе арабского алфавита, интерпретируются его научно-теоретические взгляды, выявляются социально- просветительские причины написания трактата «Хатти Бабурий», раскрываются его научные и историческое значение. Оно основано на том, что писатель создал это письмо с целью приспособления его к произношению тюркских языков, и частично определено его место в развитии письменности.

**Ключевые слова:** язык, мышление, речь, арабский алфавит, турецкая письменность, персидская письменность, звук, буква, исправление, орфография, тип письма, староузбекское письмо.

Yuksak darajadagi iste'dodli shoir, qonunshunos, adabiyotshunos, san'atshunos, tarixchi, etnograf, tilshunos Zahiriddin Muhammad Bobur jahon ijtimoiy ongi, badiiy tafakkuri, fani, madaniyati va adabiyoti taraqqiyotida alohida o'ringa ega, dunyo ahli orasida o'zining nihoyatda nodir ma'lumotlarni o'zida qamrab olgan «Boburnoma» asari, nafis g'azal va ruboiylari, aruz, harb ishi, islom qonunshunosligi,



she'riyat va til nazariyasi sohalariga doir risolalari bilan ma'lum va mashhur adiblardan biridir. Shuningdek, u yangi yozuv asoschisi ham bo'lib, bu borada katta muvaffaqiyatga erishgan alloma hamdir.

Tarixdan ayonki, VII asrdan boshlab islom dini bilan birga turkiy xalqlar hayotiga arab yozuvi ham kirib keldi va bu o'lkada shu paytgacha mavjud bo'lgan barcha xat turlarining o'rnini deyarli egalladi. Mazkur xat turi o'zining taraqqiyot bosqichlarida ayrim o'zgarishlarga uchradi. Mahalliy aholining madaniy va ma'rifiy ehtiyoji sabab arab yozuvi soddalashtirishga harakat qilindi. Shunday xat turlaridan biri *nasta'liq* (yoki xatt-ut-ta'liq) bo'lib, u fors mamlakatlari hududlarida keng tarqaldi. Shu sababli uni muomalada *forsiy xati* deb ham atashadi. Mazkur yozuvning asoschilaridan biri Xoja Mir Ali Tabriziy (1330-1405) bo'lib, u "Qiblat ul-kutub" ("Kotiblarning peshvosi") unvoni bilan sharaflangan. Arab yozuvi asosida bundan tashqari "nasxi buxoriy" (asoschisi Mir Ubayd Buxoriy), "Mavlaviy Soqiy" (asoschisi Mavlaviy Soqiy). "noxun xati" (asoschisi Nizomiddin Buxoriy) kabi xat turlari ham qo'llanishda bo'lgan [1, 4].

Alohida qayd etish o'rinliki, o'zbek tili tarixida o'n ikki asrdan ortiqroq davrda amalda bo'lgan arab yozuvini o'zgartirish, isloh etish harakati ko'lamli keng emas. Ayrim ziyoli yoki guruhlarning bu boradagi sa'y-harakatinigina alohida ta'kidlab o'tish mumkin, xolos. Bunday harakatning eng dastlabki yorqin namunasi sifatida Mahmud Koshg'ariyning ilmiy va ma'rifiy faoliyatini misol tarzida aytsa bo'ladi. U "Devonu lug'otit turk" (XI) asarida arab alifbosiga asoslangan turkiy yozuv xususiyatlari bo'yicha mantiqli fikr bildirib, birinchilardan bo'lib tovush va harf munosabatini aniqladi, turkiy tillarda joriy etilgan arab alifbosidagi harfiy belgilar soni ushbu tillarning barcha tabiiy talaffuz xususiyatlarini to'liq ifodalay olmasligiga doir o'rinli qarashlarini ilgari surdi, arab alifbosi tizimiga yana yettita shakliy belgini qo'shish kerakligi bo'yicha mantiqli tavsiyani berdi [4, 8].

Turkiylar, chunonchi, o'zbeklar qo'llagan arab yozuvini islohi bo'yicha eng jiddiy va ahamiyatli harakatni Zahiriddin Muhammad Bobur amalga oshirdi, desak xato qilmagan bo'lamiz. Bobur tabiatan hayotda har tomonlama ijodkor bo'lganligi ma'lum. Buni uning "Boburnoma" asari aniq dalillaydi. Unda keltirilishicha, alloma umrining oxirigacha samarali ijodiy ish bilan shug'ullangan, har qanday og'ir sharoit va qiyin vaziyatlarda ham ijod etishdan to'xtamagan. Shu uchun undan har jihatdan muhim, boy va ahamiyatli ilmiy-adabiy meros qoldi. Adibning merosida "Xatti Boburiy" risolasi alohida qiymat va ahamiyatga ega. Tarixiy-adabiy manbalarda qayd etilishicha, ushbu

asar taxminan 1504-yillarda, Boburning yigirma yoshlarida Qobulda bo‘lib turgan vaqtida yozilgan. “Xatti Boburiy” yozuvi “sig‘noqiy” yozuvi deb ham nomlangan. Shoir buni mashhur g‘azallaridan birida quyidagicha ifoda etadi:

Turklar xatti nasibing bo‘lmasa ne tong Bobur,  
*Boburiy xatti emasdur, xatti Sig‘noqiydur.*

Baytda keltirilgan Sig‘noq qadimiy turkiy shaharlardan biri bo‘lib, u Sirdaryo bo‘yidagi Talas vodiysi hududi tarkibida bo‘lgan. Bobur qadimgi turkiy alifbo va uyg‘ur yozuvidan xabardor bo‘lgan, ularning ma’qul joylarini o‘z alifbosiga kiritgan [8, 218]. Boburning “Xatti Boburiy” risolasi xorijiy va o‘zbek olimlari tomonidan qisman o‘rganilgan [6]. Bu borada o‘zbek olimlaridan S.Azimjonova va I.Nizomiddinovlar kabilarning xizmatlarini alohida e’tirof etish o‘rinlidir. Ular o‘tgan asrning 50- 60-yillarida mazkur risola bo‘yicha jiddiy izlanishlarni amalga oshirishdi. Ma’lumki, YUNESKO qarori bilan 1983-yilda Bobur tavalludining 500 yilligi jahon bo‘yicha keng nishonlandi, asarlari turli tillarga tarjima qilindi, uning hayoti va asarlari bo‘yicha tadqiqotlar amalga oshirildi. Parij shahrida “Boburnoma” asarining yangi fransuzcha tarjimasini “Bobur kitobi” nomi bilan chop etildi. Tarjimani mashhur va iste’dodli turkshunos Jan Lui Bakke Gramon amalga oshirdi. 1985-yil mazkur tarjima ba’zi tuzatishlar, yangi ilmiy manbalar va ko‘p sonli miniatyura bezaklari bilan nazarda tutilgan tilda yana qayta nashr etildi. Shuningdek, yangi nashrga bir qator ilovalar ham kiritildi. Ilovalarda Bobur ajdodlarining shajarasi, “Xatti Boburiy” haqidagi maqola, shu asarning qo‘lyozmasi fotofaksimilesi aks ettirildi [7].

Bobur imlo masalalariga, ayniqsa, arabcha so‘zlarning yozilishi me’yorlariga alohida e’tibor qaratgan. Uning bu boradagi say’-harakatlari o‘g‘li Xumoyunga yozgan maktubida yaqqol namoyon bo‘lganligini ko‘rish mumkin. Alloma bu maktubida dastlab til va tafakkur masalasiga, ularning o‘zaro munosabati va aloqadorligiga diqqat qaratadi, fikrlashi yuqori darajada bo‘lgan shaxsning xati ham chiroyli va savodli bo‘lishini aytib o‘tadi. So‘ngra esa so‘zlarni yozuvda to‘g‘ri ifoda etishni, ismlar imlosi va ma’nosiga e’tiborli bo‘lish kerakligini, imlo me’yorlariga to‘liq amal qilishlikni, yozuvni, albatta, husnixatda bitishlikni, bunday yozuvlar jamiyat hayotida katta ahamiyatga egaligini alohida ta’kidlaydi.

Adib o‘g‘lining xatidagi kamchiliklarni ko‘rsatib, uni xatosiz yozishga undaydi. Bobur fikricha, muallifning xati savodli va maz-

munli yozilishi, uslubiga ko‘ra ravon, ochiq va sodda bo‘lishi kerak. U bu xususda quyidagilarni yozadi: “Xat biturda takalluf qilay deysen, ul jihatdin mug‘laq bo‘ladir. Bundan nari betakalluf va ravshan va pok alfoz bila biti, ham sanga tashvish ozroq bo‘lur ham o‘qug‘uchig‘a”. Bobur arab yozuvining ifoda imkoniyatlari va shakliy tuzilishiga doir xususiyatlarini, foydali, ahamiyatli, qulay tomonlarini va ayni paytda kamchiliklarini ham puxta bilgan [5, 161]. Shunga ko‘ra u arab yozuvini soddalashtirish va osonlashtirish, bu yozuvda bitilgan muayyan so‘zni har xil yozish va o‘qishdagi qiyinchilik hamda noaniqliklarni bartaraf etish maqsadida nazarda tutilgan yozuv turini ishlab chiqdi. “Boburnoma” asarida o‘zining yangi yozuvini yaratganligini eslatib o‘tadi: “Ushbu mahallarda (1504-yillarda – Y.S.) boburiy xatini ixtiro qildim” [3, 421].

Bobur ushbu risolasida arab alifbosini turkiy tillar, xususan, o‘zbek tili xususiyatlaridan kelib chiqib, uni birmuncha soddalashtirishga harakat qiladi. Adib kishilar o‘rtasidagi muloqat, bilim va og‘zaki ijod namunalarini kelajak avlodga yetkazish uchun yozuv muhim ahamiyatga egaligini yaxshi anglagan. Ma’lumki, tilning o‘zaro aloqa, estetik, dunyoni bilish, kishilik jamiyati tomonidan egallangan tajriba-bilimlarni saqlash va avlodlarga yetkazish kabi vazifalari yozuv vositasida amalga oshiriladi. Allomaning yozuv islohi bo‘yicha harakatlarini o‘rganib turib, u alifbodagi bir harf muayyan bir tovushni ifodalashi, bir tovush muayyan bir harf bilan ifodalanishi lozimligini chuqur anglagan, degan xulosaviy fikrga kelish mumkin.

Adib mazkur xat turida arab yozuvini turkiy tillarning talaffuz xususiyatlariga yanada moslashtirish uchun arab alifbosiga turli o‘zgartirishlar kiritib, uni isloh etdi, bir necha asrlar davomida foydalanib kelingan yozuvda turkiy tili tovush qurilishini to‘laroq ifodalash, unga moslashtirish maqsadida bir qator harfiy o‘zgartirishlar kiritdi, arab alifbosini birmuncha soddalashtirishga harakat qildi. Boburning hayoti va ijodini o‘rganish jarayonida uning iste’dodli kalligraf, husniyat qoidalariga to‘liq rioya etib, chiroyli yozuv sohibi, xattot ekanligiga amin bo‘lish mumkin. Adib, aytilganidek, arab alifbosining yozilishi va o‘qilishida murakkabliklar mavjudligi bois nisbatan yangi alifbo – «Xatti Boburiy»ni ixtiro etdi, uni xalq orasida keng joriylanishi uchun kurashdi, keng ommalashtirish uchun zaruriy harakatlarni amalga oshirdi.

Ilmiy-tarixiy manbalarda keltirilishicha, Bobur Hindistondagi hukmronlik davrida ushbu xat turidan foydalanib turli maktublari va mushaf yozadi. Alloma Qur’oni karimning ikki nusxasini “Xatti Boburiy”da ko‘chirtirib, ulardan birini Makkaga, ikkinchisini esa Teh-

ronga yuboradi. Bu nusxalar fanda «Mushafi Boburiy» deb ataladi. Asarning qo‘lyozmalaridan biri Eronning Mashhad shahridagi “Ostoni qudsi Rizoviy” kutubxonasida saqlanadi. “Bobur xalqaro fondi” tomonidan 2004-yilda “Mushafi Boburiy” foto nusxasi O‘zbekistonga olib kelindi. Ushbu fotonusxa hozirda Andijonning “Bobur va Jahon madaniyati” muzeyida saqlanadi. Ma’lumki, *mushaf* so‘zi arabcha bo‘lib, *sahifalangan* degan lug‘aviy ma’noni ifodalaydi. Xalifa Usmon ibn Affon davrida ko‘chirilgan Qur‘on nusxalari shu nom bilan atalgan. Keyinchalik kitob shaklidagi Qur‘onning har bir nusxasi *mushaf* deb nomlangan.

Ayonki, Bobur hirotlik mashhur ulamolardan Qozi Ixtiyor va Muhammad Miryusuflarga “Xatti Boburiy” alifbosini o‘rgatadi. O‘g‘illari, jumladan, Komron Mirzo va Hindol Mirzolarni mazkur yozuvda maktublar bitishga undaydi. Biroq adibning ijtimoiy-madaniy, ma’rifiy taraqqiyotda muayyan darajada vosita bo‘luvchi bu ahamiyatli yangiligi o‘sha davrda, ayrim kishilarni e’tiborga olmaganda, ko‘pchilik tomonidan e’tirof etilmadi va qo‘llab-quvvatlanmadi. Adib bu xat turini, yuqorida aytilganidek, arab alifbosi asosida tuzgan. Uning alifbosi yigirma sakkiz harfdan iborat tarzida shakllantirilgan. “Xatti Boburiy”ning arab alifbosi bilan o‘xshash va farqli jihatlari mavjud. Bobur o‘z alifbosini arab alifbosidagi zeru zabarlardan xoli qildi, ayni vaqtda, arab alifbosiga keyinchalik kiritilgan to‘rtta forscha harflarni undan chiqarib tashladi [2, 31].

Umumiy o‘xshash jihatlari shundaki, “Xatti Boburiy”da arab alifbosidagi ayrim harflar aynan qabul qilingan. Bunday harflar jumlasiga *alif, yo*, so‘z o‘rtasidagi *be* kabilarni aytib o‘tish mumkin. Farqli, ayirma jihatlari esa quyidagilardan iborat: 1) to‘liq o‘zgartirilgan harflar (*kof, nun, vov, fe, sin, shin* kabi harflar); 2) ayrim jihati o‘zgartirilgan harflar (*be, jim, xe, to, zo* kabi harflar); 3) qisman o‘zgartirilgan harflar (*mim, he, dol, ayn* kabi harflar).

Bobur nazarda tutilgan risolasining biror o‘rnida mazkur xat turini qaysi maqsadda kashf etganligini yozmagan. Nazarimizda, u ushbu xat turini shakllantirishda o‘z oldiga asosiy maqsad sifatida quyidagilarni belgilagan bo‘lishi mumkin:

1) xalqning savodlilik darajasini oshirish maqsadida arab alifbosini isloh etgan;

2) arab alifbosini turkiy tillarning talaffuz me‘yorlariga moslashtirish maqsadida uni o‘zgartirgan;

3) eng muhimi, turkiy xalqlarning madaniyati va ma‘rifati taraqqiyoti uchun mazkur xat turini yaratgan.

Bobur amalda boʻlgan arab alifbosida xat-savod chiqarish qiyinligi va juda koʻp vaqt hamda mehnatni talab etishini eʼtiborga olib, aholining savodini osonroq va tez fursatda chiqarish, oʻqish va yozishni qulaylashtirish, turkiy tillarga xos xususiyatlarni yozuvda muayyan darajada aks ettirish, ushbu tillarning talaffuz tabiatiga moslashtirish maqsadida bu yozuvni yaratgan, desak xato qilmagan boʻlamiz. Adib yashagan ijtimoiy-siyosiy davrda necha asrlardan beri muqaddas Qurʼon yozuvi sifatida amalda boʻlgan alifboni isloh etish, oʻzgartirish va yangilashga harakat qilish bilim, isteʼdod, topqirlikni talab etishi bilan birga, ayni paytda, eng muhimi, madaniy-maʼrifiy jasoratni ham talab qiladi.

Xulosa sifatida aytish mumkinki, Bobur oʻzbek xalqi uchun oʻzlashtirish oson va qulay boʻlgan, turkiy tillarning oʻziga xos tovush tizimi xususiyatlari eʼtiborga olingan arab alifbosi negizida maxsus xat turini yaratib, oʻzbek tilshunosligi tarixida oʻchmas iz qoldirdi. Adibning bu xat turi bilan bogʻliq ilmiy qarashlari va amaliy faoliyatini chuqur oʻrganish, uning ilmiy-tarixiy ahamiyatini ochib berish, tadqiqot natijalari bilan jahon ilm ahlini xabardor etish mazkur faning dolzarb masalalari sirasiga kiradi, desak xato qilmagan boʻlamiz. Bunday ishni amalga oshirish orqaligina Boburning yozuvlar tarixidagi oʻrnini toʻgʻri belgilash mumkin boʻladi.

#### ADABIYOTLAR :

1. Абдулғафур Раззоқ Бухорий. Хуснихат дурдоналари. – Тошкент: Мовароуннахр. 2008.
2. Азимжоновна С. Бобурнинг хинд девони. – Тошкент: Фан, 1966.
3. Бобир. Бобурнома. – Тошкент: Фан, 1960.
4. Махмуд Кошғарий. Девону луғотит турк (таржимон С.М. Муталлибов). I-III. –Т.: Фан, 1960–1963. – Т. I. 1960. – 499 б.
5. Саидов Ё. Ўзбек адабий тили тарихи. Дарслик. – Бухоро: Дурдона, 2021. -376 б.
6. Султонов Ў. Хатти Бобурий // Мозийдан садо. – Тошкент: Ўзбекистон, 2003.
7. Холбеков М. Бобур ва “Бобурнома” француз шарқшунослиги саҳифаларида // Жаҳон адабиёти. –Тошкент: 2017. – № 2. – Б. 35–46.
8. Қодиров П. Тил ва эл. –Тошкент: Маънавият, 2010. – 294 б.
9. Ўзбекистон тарихи. Дарслик. –Тошкент: Янги аср авлоди, 2003. 187 б.

УДК 81-11

**“HASHAROT” LMG ASOSIDA SHAKLLANGAN O‘ZBEK XALQ  
PAREMALARINING NAZARIY TADQIQI***Samandarova Gulnoz Yarashovna**Buxoro davlat universiteti*

gulnozsamandarovna1102@gmail.com

Ushbu maqolada “Hasharot” lug’aviy ma’noviy guruhi asosida shakllangan o‘zbek xalq paremalarining o‘rganilishi, ularning lingvistik asoslari hamda nazariy tadqiqi yoritilgan. Shu yo‘nalishda ilmiy izlanish olib borgan jahon va o‘zbek olimlarining tadqiqotlari tahlil qilingan.

**Kalit so‘zlar:** parema, paremiologiya, maqol va matallar, hasharot nomi, entomologiya, lingvistik asos.

**THEORETICAL STUDY OF THE UZBEK PAREMS FORMED ON  
THE BASIS OF LSG OF THE “INSECTS”***Samandarova Gulnoz Yarashovna**BukhSU*

gulnozsamandarovna1102@gmail.com

This article covers the study of Uzbek folk songs formed on the basis of the lexical and semantically group “Insects”, their linguistic foundations and theoretical research. Researches of world and Uzbek scientists who conducted scientific research in this direction were analyzed.

**Key words:** parema, paremiology, proverbs and sayings, insect name, entomology, linguistic basis.

**ТЕОРЕТИЧЕСКОЕ ИЗУЧЕНИЕ УЗБЕКСКИХ НАРОДНЫХ  
ПАРЕМ, СОЗДАНЫХ НА ОСНОВЕ ЛСГ «НАСЕКОМЫХ»***Г. Я. Самандарова**БухГУ*

gulnozsamandarovna1102@gmail.com

В данной статье рассматривается исследование узбекских народных парем, сформировавшихся на основе лексико-семантической группы «Насекомых», их лингвистические основы и теоретическое исследование. Проанализированы исследования мировых и узбекских учёных, проводивших научные исследования в этом направлении.

**Ключевые слова:** парема, паремиология, пословицы и поговорки, название насекомых, энтомология, языковая основа.

Dunyo tilshunosligida xalq donishmandligining hosilasi boʻlgan maqol, matal, maqol-matal tipidagi birliklar shaklida yuzaga kelgan paremlar obrazlilik va ekspressivlik, troplar va stilistik figuralar asosiga qurilganligi bilan ajralib turadi. Ularning lisoniy xususiyatlarini oʻrganish ona tilimiz xazinasidagi botiniy boyliklarni yorqin tasavvur etishimizga koʻmaklashadi. Shu nuqtayi nazardan “hasharot” lugʻaviy maʼnoviy guruhi asosida shakllangan oʻzbek xalq paremlarining lingvistik asoslari va leksik-semantik xususiyatlarini tadqiq qilish paremiologiyaning ilmiy tadqiqi doirasini kengaytirgani holda amaliy jihatdan ularning nutqiy qoʻllanish darajasini oshiradi.

Oʻzbek tilshunosligida soʻnggi yillarda lingvistik va turli sohalarining integrativ munosabatiga, lisoniy birliklarning til xususiyatlarini milliy asoslarda tadqiq etish masalalariga alohida eʼtibor berilmoqda. Paremlarning kelib chiqishi, tarixiy rivojlanishi va maʼno xususiyatlarini oʻrganuvchi paremiologiya boʻlimi ham mana shunday eʼtibor mahsuli sanaladi.

Fanda maqollarni xalq ogʻzaki ijodi, tilshunoslik obyekti sifatida oʻrganish, ularning lingvistik xususiyatlarini tadqiq etish va, umuman olganda, lingvistik obyekti sifatida qarash uzoq tarixga ega. Tilshunoslikda bu yoʻnalish doirasida koʻplab tadqiqotlar olib borildi va muayyan natijalarga erishildi. Jahonda bir qator olimlar bu borada izlanish olib borishgan. Jumladan, N.V.Zimovets, I.Sirot, S.Basharan, Lafta Adnan Habib, O.A.Kornilov, E.V.Kaledina, O.G.Xabarova, O.V.Cherian, Omer Bicher, V.Boktayeva, Sh.Mazhitayeva, O.A.Dmitriyeva, Y.A.Krivoshapova, A.Sichugova, Chje Inkuy, V.Budziszewska, V.Vazniy, S.Utesheny, N.Bunko, A.Gura, T.I.Vendinalarning tadqiqotlari eʼtiborga molik.

Oʻzbek tilshunosligida paremiologik birlik sifatida maqol va matallar ustida tadqiqotlar olib borilishi, koʻplab olimlarning paremlar boʻyicha shoir va yozuvchilar ijodidagi oʻziga xosliklar borasidagi izlanishlari, nomzodlik va doktorlik dissertatsiyalarida ularni qoʻllash bilan bogʻliq uslubiy qarashlarning oʻrganilishi diqqatga sazovor. Oʻzbek tilshunosligida maqolning sintaktik xususiyatlari X.Abdurahmonovning nomzodlik va doktorlik ishida oʻrganilgan boʻlsa; oʻzbek maqol va matallarining lugʻaviy xususiyatlari M.Sadriddinovaning tadqiqotida; maqollarning koʻp maʼnoliligi, sinonimligi, variantliligi va ularning uslubiy xususiyatlari B.Joʻrayevaning nomzodlik ishida, oʻzbek, rus va qozoq tilidagi maqollarning semantik va struktur xususiyatlari P.Bakirovning tadqiqotlarida batafsil oʻrganildi.

Soʻnggi yillarda oʻzbek tilshunosligida B.Joʻrayevaning “Oʻzbek xalq maqollari shakllanishining lingvistik asoslariva pragmatik xusu-

siyatlari” nomli doktorlik dissertatsiyasi, D.Turdaliyevaning “O‘zbek xalq maqollarining lingvopoetik xususiyatlari”, D.Toshevaning “Zoonim komponentli maqollarning lingvokulturologik xususiyatlari”, Sh.Qalandarovning “O‘zbek lingvomadaniy muhitida xalq maqollari evfemizatsiyasi”, Y.Nurovaning “O‘zbek xalq paremalarining etnolingvistik xususiyatlari” (kiyim va oziq-ovqat nomlari asosida), M.J.Sharipovning “Chorva nomlari asosida shakllangan o‘zbek xalq maqollarining lingvopoetik va uslubiy xususiyatlari” (“qo‘y” lug‘aviy ma‘noviy guruhi asosida) kabi qator tadqiqotlar amalga oshirildi.

Hozirgi kunda har jihatdan taraqqiy etib borayotgan yurtimizda aynan ilmiy sohada yuz berayotgan katta o‘zgarishlar o‘zbek tilining taraqqiyotiga, ayniqsa, uning leksikasi rivojiga sezilarli ta’sir ko‘rsatmoqda. Bu o‘zgarishlar tilimiz lug‘aviy tarkibining nafaqat miqdor jihatdan, balki uning sifat jihatdan ham rivojlanishiga sabab bo‘lmoqda.

Ma’lumki, xalqlar o‘z asriy hayotlari natijasida juda katta tajribaga ega bo‘ladi va mana shu tajribalar asosida ulkan ma’naviy boylik sanalgan paremalar kabi birliklarni shakllantirib, ularni kelajak avlodga meros qilib qoldiradi. Bu durdonalarning yuzaga kelishiga juda uzoq yillar, asrlarni o‘z ichiga olgan har bir millatning etnik, kundalik-maishiy, hayotiy tajribalari, urf-odatlarini asos bo‘lar ekan, ularni asrab-avaylash, boyitish va o‘rganish har bir xalq zimmasidagi eng dolzarb vazifa bo‘lib qolaveradi. Tabiatda hasharotlar turli-tuman bo‘lib, ular asosan, quruqlik bo‘ylab keng tarqalgan. Ayniqsa, sernam iqlim hududlarida ularning ko‘plab turlari uchraydi. Hasharotlarning katta qismi yerda, ba’zi turlari suvda yashasa, yana bir qismio‘z hayotini tuproq bilan bog‘liq holatda yashab o‘tadi. Bu mitti mavjudotlar xilma-xil mahsulotlar bilan oziqlanib, tabiatda moddalar almashinuvida ishtirok etadi.

Hasharotlarning juda ko‘p turlari o‘simlik zararkunandalari; hayvonlar va odamlarga zarar keltirishi bilan birga, o‘simliklarni changlatish, zararkunanda hasharot va begona o‘tlarni yo‘qotishda ham faol ishtirok etib, o‘zlarining foydali jihatlarini namoyon qiladi. Masalan, hasharotlardan asalari, ipak qurtlari qimmatbaho mahsulot bersa, ayrimlari ovlanadigan hayvonlar uchun ozuqa manbayi hisoblanadi. Hasharotlarning turlari va ularning foydali hamda zararkunandalik jihatlarini, ulardan amaliyotda foydalanish usul, texnologiyalarini o‘rganuvchi fan entomologiya deb yuritiladi.

**Entomologiya** (lot. entos – hasharotlar va logos – fan) – hasharotlar to‘g‘risidagi fan bo‘lib; ularning tuzilishi, hayot kechirishi, individual va tarixiy rivojlanishi, xilma-xilligi, yer yuzida tarqalishi, yashash muhiti bilan munosabatlari va boshqalarni o‘rganadi.



Vazifasiga binoan, nazariy, ya'ni umumiy entomologiya va amaliy entomologiya farqlanadi. Umumiy entomologiya hasharotlar morfologiyasi, embriologiyasi, fiziologiyasi, biokimyosi, etologiyasi, entomogeografiyasi, paleontologiyasi, sistematikasi va boshqa fanlarga ajratiladi. Bu fanlarni o'rganish obyektiga ko'ra yanada kichikroq bo'limlarga bo'lish mumkin.

Masalan, sistematika tarkibida koleopterologiya – qattiq qanotlilarni, lepidopterologiya – kapalaklarni, mirmikologiya – chumolilarni o'rganadi.

Amaliy entomologiyaning o'rganish obyekti qishloq xo'jaligi o'simliklari va mahsulotlari zararkunandalari, odam, hayvonva o'simliklarning parazitlari hamda xalq xo'jaligi hamda tabiatda foydali ahamiyatga ega bo'lgan hasharotlardir. Amaliy entomologiya ham o'z vazifasiga ko'ra, o'rmon zararkunandalari (o'rmon entomologiyasi), qishloq xo'jaligi ekinlari zararkunandalari (qishloq xo'jaligi entomologiyasi), xonaki va yovvoyi hayvonlarda parazitlik qiluvchi hasharotlar (veterinariya entomologiyasi), odamda parazitlik qiluvchi (tibbiyot entomologiyasi) hamda odamlar foydalanadigan mahsulotlarni beradigan tut va dubdagi ipak qurti (ipakchilik), asalarini (asalarichilik) o'rganadigan fanlarga ajratiladi.

Barchamizga ma'lumki, turkiyshunoslikda zoonimlarni lingvistik nuqtayi nazardan tadqiq etish M.Koshg'ariyning “Devonu lug'otit turk” asaridan boshlangan bo'lib, o'zbek tilshunosligida hozirga qadar ham izchillik bilan davom etib kelmoqda. D.M.Yuldashevaning nomzodlik dissertatsiyasida alla, ovutma, erkalama, ertak, topishmoq, maqtov, masxaralama va tez aytishlarda qo'llangan zoonimlar tahlili amalga oshirilgan. G.Y.Hakimova tadqiqotida zoonim komponentli frazeologik birliklar so'z turkumlari bo'yicha tahlil qilingan va foiz miqdori aniqlangan. Zoonim komponentli frazeologik birliklar tarkibidagi zoonimlarning ifoda semalari 17 ta hayvon nomi misolida 11 ta sema qirralari nuqtayi nazaridan tahlilga tortilgan. Olima D.Toshevaning tadqiqotida bo'ri komponentli maqollardagi 14 xil, arslon komponentli maqollardagi 3 xil, it komponentli maqollardagi 21 xil, ot komponentli maqollardagi 19 xil, tuya komponentli maqollardagi 8 xil, tarkibida qush nomlari ishtirok etgan maqollardagi 5 xil kontseptlar tahlili amalga oshirilgan. Ammo hozirga qadar hasharot nomlari ishtirok etgan maqollarning batafsil tahlili amalga oshirilmagan.

Jamiyat a'zolarining kundalik hayotida keng qo'llaniladigan narsa-buyum, voqea-hodisa, ularga xos belgi-xususiyatlarning ijtimoiy shartlangan qo'shimcha ma'nolari mavjud bo'ladi.

Shuningdek, “hasharot” LMG asosida shakllangan xalq paremalarining keng tarqalganligi, ayniqsa, kishilar hayoti, yashash tarzi (jamoadi, yolgʻizlikda), xatti-harakati (yeyish, yoʻrgʻalash, sakrash) hamda xususiyati (gʻamxoʻrlik qilish, mehnatkashlik, dangasalik) bilan bevosita bogʻliq boʻlganligi sababli ularning maʼno qurilishida ham qoʻshimcha semalar ajratiladi.

Umuman olganda, jahon tilshunosligida zoonimlar, xususan, hasharot nomlari turli aspektlarda tadqiqilingan. Etimologik, lingvokulturologik, leksik-semantik jihatdan tahlil etilgan ushbu ishlar mazkur tadqiqot uchun nazariy asos vazifasini oʻtaydi.

#### ADABIYOTLAR :

1. Аникин В.П. Русские пословицы и поговорки. – Москва: Художественная литература, 1988. – 432 с.
2. Бердиёров Х., Расулов Р. Ўзбек тилининг паремиологик луғати. – Т.: Ўқитувчи, 1984. – 286 б.
3. Кошғарий М. Туркий сўзлар девони (Девону луғотит турк). Уч томлик. I том. – Т.: Ўзбекистон ССР Фанлар академияси нашриёти, 1960. – 500 б.
4. Корнилов О.А. Языковые картины мира как производные национальных менталитетов. – Москва: КДУ, 2013. – 348 с.
5. Шомақсудов Ш., Шораҳмедов Ш. Маънолар махзани (ўзбек мақолларининг изоҳли луғати). – Т.: Ўзбекистон миллий энциклопедияси Давлат илмий нашриёти, 2001. – 448 б.
6. Ўзбек халқ мақоллари. 1-жилд. А–О. – Т.: Фан, 1987. – 368 б.
7. Shavketovna, Y. A., & Zayniddinovna, B. G. (2023). Lexical-Semantic Characteristics of Medical Abbreviations. *Best Journal of Innovation in Science, Research and Development*, 2(6), 98–101. Retrieved from <http://www.bjisrd.com/index.php/bjisrd/article/view/289>
8. Bozorova, G. (2023). Лингвопрагматическая характеристика плеонастических единиц используемых в политической речи. *Центр научных публикаций (buxdu.Uz)*, 34(34), извлечено от [https://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/9752](https://journal.buxdu.uz/index.php/journals_buxdu/article/view/9752)
9. Yarashovna, S. G., Istamovna, A. N., & Zayniddinovna, B. G. (2023). Methods of Using Metaphorical, Pleonastic and Paremic Units in Uzbek Linguistics. *resmilitaris*, 13(2), 5560–5572.
10. Samandarova, G. Y. (2021). Fundamentals of folk proverbs formed on the basis of the lexical-spiritual group of insects. *Current research journal of philological sciences*, 2(05), 39–42.
11. Bozorov Z.A International scientific and practical conference “Modern psychology and pedagogy: problems and solutions” “Imperial College London” 2021–2022. 165–169 б.

---

12. Бозоров З.А. The system of formation of civic culture among students in higher educational institutions. ISJ Theoretical & Applied Science, 01 (81), 455-458. Soi: <http://s-o-i.org/1.1/TAS-01-81-81> Doi: <https://dx.doi.org/10.15863/TAS> Scopus ASCC: 3304. (2020).

13. DN Yuldasheva. BASIC NON-VERBAL COMPONENTS OF SPEECH. CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (2767–3758) 3 (01), 17-24. <https://masterjournals.com/index.php/crjps/article/view/661>.

## УДК 81-11

**J. K. ROULINGNING FANTASTIK ASARLARIDAGI  
ANTROPONIMLARNING LINGVO-PERSPEKTIV  
MUAMMOLARI***Solijonov Juraali Kamoljnovich**Termez State University**Termez, Uzbekistan**stuard\_redman@mail.ru**ORCID: 0000-0001-6394-295X**UDK 81.2/4481.25*

Mazkur maqola J.K.Rouling tomonidan yozilgan “Garri Potter” turkumidagi romanlar seriyasining tarkibidagi antroponimlarning ingliz tilidan o‘zbek tiliga tarjimasini va ularning mohiyatiga lingvistik nuqtai nazardan o‘quvchi uni anglashi uchun ta’sir qiluvchi omillarni hisobga olgan holda tahlil qilingan. Obyekt sifatida tanlangan ismlar to‘qima bo‘lgani tufayli personajlarning ismlaridan oldin ularning asardagi vazifasi hamda xarakteri o‘ylab topilgani, so‘ngra uni lingvistik qoliplarga solingan. Shu tufayli maqolada tarjimon ularni o‘zbek tiliga o‘girishdan oldin etimologik, leksikologik, semantik va fonetik tahlil qilishi uchun qoliplar taklif qilingan.

**Kalit so‘zlar:** antroponimlar, transliteratsiya, lingvo-perpektiv, fantastic asarlar, Salazar Slizerin, ingliz-o‘zbek tarjima, leksikologiya.

**LINGUA-PERSPECTIVE PROBLEMS OF TRANSLATING  
ANTHROPONYMS IN THE FICTION BOOKS OF J.K.ROWLING**

This article analyzes the English-Uzbek translation problems of anthroponyms in the Harry Potter series written by J.K. Rowling from the point of the linguistic factors that influence the reader’s understanding of their essence. The names chosen as objects in the article are fictive, the role and character of the personages in the work were created before the names of the characters, and then it was put into linguistic patterns. That is why; this article proposes templates that include etymological, semantical and phonetical analysis before the act of translation.

**Keywords:** anthroponyms, transliteration, lingua-perspective, fiction, Salazar Slytherin, English- Uzbek translation, lexicology.

**INTRODUCTION**

Lexic problems have always been one of the most central points of translation; we have to limit the scope of the term as onomastical problems of translation. The translation process of a J. K. Rowling novel is so disputable that one could get deep into the lines and live in the

wold of eloquent powerful stylistics. Word building of the novel put the great tasks that require enormous research and analysis. In all culture, throughout the beginning of literature, about 2000 years of BC or earlier, people wanted to choose the proper name for the new member of their family relating to his prosperous future, which they picked up by considering the name's deep meaning carefully. This is still an important part of our life, simultaneously, as they called literature as "the mirror of life," writers have been trying hard to name their characters in books in that way. The criteria are hard to accomplish on that part of writing. When we look at the case inductively, names of characters in fairy tales are the best examples: "Bo'g'irsoq" or "Bilmasvoy" in Uzbek – "Незнайка" in Russian relates the personage's lacking knowledge which can be translated as "Dunno," "Know-Nothing" or "Ignoramus" into English. Main aim to get them translated is to reach the point where a child should be aware of at least one behavioural temper of the character that helps a child to imagine a vivid picture of the personage in the tale.

The process may not always be repeated with all names in literature, especially when the name is not as straightforward as given above as "Alice" in Wonderland and "Alpomish" (an Uzbek epic). They express a specific cultural background and carry special history in them, although "Sim-sim" and "Qaqnus" were meant to determine that they belong to another fictive world that is apart from us that hooks a reader to a story or a personage name more than a simple name.

#### LITERATURE REVIEW AND RESEACH METHODOLOGY

Comparative literature, contrastive and descriptive translation and interpretation analysis of linguistic terms and stylistically analysis of literary translation during the research for the article.

The inventive crafting of the author – J. K. Rowling creates inevitable questions theoretically to develop in the research. The outcomes can be seen from three angles: the first one is to examine if the names are supposed to be translated or some translators translated them, if yes, how, in the target language. As it comes, the second angle is to see the strategies and methods used to translate the objectives and to select the right ones to show the answer. Finally yet importantly, is to determine the names' carrying the semantic load or any descriptive meaning by comparing and contrasting the English-English variations, as well, during the study.

Anthroponyms are studied with their etymologies and are analyzed with controversial ideas of two translators of a work. By contrasting them in English, Russian and Uzbek, we intend to get a fragile line of semantic connection a bit stronger in the Uzbek version.

To investigate the Russian versions of Harry Potter books, we have browsed some authentic comments and official articles on the internet, such as personal opinions and announcements, and official articles on Russian magazines. Some are given below:

*Official translation of Harry Potter is considered to be poorly done among Russian fans. That is why we have at least two other translations. One was done by Maria Spivak (Мария Спивак) and the other is called simply 'public translation' (народный перевод). – says Mishka Koala [1]*

A Russian newsagent BFM.RU says “Гарри Поттер и философский камень” вышла в новом переводе” – Harry Potter and the Philosopher’s Stone have been republished with a new translation” by Maria Spivak in Azbooka-Atticus: Machaon publishing house. [2] Respectively, one can observe that The Blair Partnership on Wikipedia has accepted Maria Spivak’s version.

## RESULTS

The names examined in the present article are one of the most common and important personages, despite the fact that, we had much more to analyze in the first book of Harry Potter with more than 50 names. We have seen the etymological and semantical background of names: examined 19 character names, 2 nouns and a solution for the issue of interlanguage phonetical transformation: when to use “g” and “h” in Uzbek instead of “h” in English. Next research goes deeper in the current field and continues with the names of plants and meals, which are problematic, but ‘page-turner’ fields of onomasiology.

## DISCUSSION

Names of personages are supposed to be transliterated in “Harry Potter” book 1. However, what if a character loses its features for changing its name? Each name is chosen due to the role, aim, character and importance of a personage in the novel. For example, Professor Quirrell is one of the main characters in the book 1, but it is known with negative features as aggression, confusion and scare. The word seems to be derived from “Quirrel” the abbreviation of “squirrel,” but

it is only a one-sided hypothesis. However, Shokir Dolimov, a translator of Russian-Uzbek, gave “белка” for “Quirrel.”

– Профессор Белка! – таништирди уни Хагрид, – Гарри, профессор Белка “Хогварц”да сенга мураббийлик қилади. [3, P.57]

We consider that the translation of Quirrel to ‘Belka’ which means ‘a squirrel’ in Russian is a total mistake and connected to Mr Sh. Dolimov in Tashkent on the talk about the Uzbek translation of the book to correct his misunderstanding of Maria Spivak’s first unedited English–Russian translated version where “Белка” was used initially. He admitted that and accepted the transliterated version ‘Kvirel’ into Uzbek.

Surprisingly, the Russians accepted the version of Maria Spivak who saved the original name... later in the latest editions:

– Профессор Квиррелл! – представил его Хагрид. – Гарри, профессор Квиррелл – один из твоих будущих преподавателей. [4]

Another fact “Professor Quirrell” is told to come from “querulous” which means “fearful,” and “irritable”. The author claims to have created the name imaginary. Some say this appeared after removing “s” from “squirrel” – “quirrel.” [5]

J. K. Rowling has always used the art of onomastics for the series of Harry Potter. This means one can see the features of characters through their names. Most names are alliterative, whimsical, and have a kind of Latin-based roots.

To continue with, Ronald Bilius Weasley, the word ‘Ronald’ is proper name, from Old Norse Rögnvaldr “Having the Gods’ Power,” from rögn “gods,” literally “decreeing powers” (plural of reign “decree”) + valdr “ruler,” he is said to have ‘the advisor of the king’ meaning as one of the best friend of a hero. The word “Weasley” As for “Weasley,” which sounds a lot like “weasel,” which is not a popular animal. In addition, a man named John Wesley was the founder of Methodism in England in the 1700s. Rowling is supposed to like the animal, though, ‘the rumours came up after deleting the post on Twitter on weasel’ according to The Insider. The personage is positive and he is a favourite of children to be secretly courage and patient as most of them. We approve the “Uizley” version of the name into the Uzbek language, although it can change to “Wizley” in speaking.

“Garri birdan Ronga qaradi, uning angrayib qolganligini ko’rib u ham hech qanday kitob o’qimaganligini tushunib oldi.

– Men Ron Uizleyman, – dedi Ron.

– Garri Potter, – dedi Garri.” [6]

“Гарри Роннинг ханг-манг бўлиб қолган башарасига қараб, у ҳам дарсликларни тўла- тўқис ёд олмаганлигига ишонч ҳосил қилгач, ўзини энгил тортди.

– Мен Рон Уэслиман, – пўнғиллади Рон.

– Гарри Поттер, – ўзини таништирди Гарри.” [3, P.108]

In Uzbek, we still have a confusion on the change “s” to “z” as in “Weasley” to “Uizley” or “Uesley” which are stated in two fragments of two Uzbek versions. It is claimed that it would be phonetically pure to accept it as “Uzley” in Uzbek according to some philologists.

Ron Weasley’s best friend Harry Potter has always been “Garri Potter” in the Uzbek language, however, an official newsagent KUN.UZ is calling him now as “Harri Potter” which does not fit to the norms of phonetics in Uzbek. The Uzbek language was strongly affected by the Russian language in the period of Soviets for at least 40 years. Therefore, Uzbeks have transliterated any name starting with the letter “H” in any Germanic language as “G” as in “guide” in English. For instance, “Genri” for “Henry”, “Gitler” for “Hitler” etc.

The article “Harri Potter aslida ikkita bo’lganmi?” [7] or “Harri Potter haqidagi birinchi ertakka 20 yil to’ldi” [8] are really intriguing that one of the most outstanding agency has still had controversial principles on transferring the phenome “h” to “g” with its earlier news “Deniyel Redkliff Garri Potter rolini ijro etishni inkor etdi” [9]. Here is the proper solution for the case: when “H” comes directly before the letters ‘a,’ ‘u,’ ‘o’ we keep the /h/ sound, such as English “Doctor House” to Uzbek “Doktor Haus” in pronunciation. When the letter “h” comes before ‘e,’ ‘i’ or ‘y’ we have to use ‘g’ instead:

**TABLE 1. The phonemic change norms in transferring the letter “h” [h] from English to “g” [g] into Uzbek.**

|    | Latin or English based names | Transliteration into Uzbek |
|----|------------------------------|----------------------------|
| 1. | <b>H</b> ermione             | <b>G</b> ermiona           |
| 2. | <b>H</b> itler               | <b>G</b> itler             |
| 3. | <b>H</b> ydro                | <b>G</b> idro              |
| 4. | <b>H</b> ouse                | <b>H</b> aus               |
| 5. | <b>H</b> agrid               | <b>H</b> agrid             |

The table above proves our theory in our translation version which means KUN.UZ agency is wrong to use “h” or “g.” As the aim of the



master's article, we conclude that the word has to be as "Garri" in Uzbek as 90% of the people know it from Uzbek translated version of Harry Potter movies which have been broadcasted since 2005 on the Uzbek TV channels.

Another bright character is Albus Percival Wulfric Brian Dumbledore, which is the director of Hogwarts Witchcraft and Wizarding School in the Harry Potter universe. In the opening chapter of the first novel of the series, Dumbledore arrives at number four, Privet Drive in Little Whinging, Surrey. When the evil wizard Lord Voldemort kills Harry's parents before being reduced to a lesser form, Dumbledore decides to place the now-orphaned Harry in the home of Vernon and Petunia Dursley. Historically, "Albus" meant "white, bright" in Ancient Roman in the Latin language. The name used to be given to boy babies. Several European magazines mention it to be a hint to his long white beard. "Percival" was one of King Arthur's Knights of the Round Table, which strengthens his position in the novel by his being in search for Voldemort's Horcruxes (advanced dark magic to attain immortality in the novel).

"Wulfric" is an Anglo-Saxon male name that means "wolf ruler" or "wolf power", doubtlessly, he has a wolf power in the novel. "Dumbledore" is used for an old English word to mean "bumblebee" which is an insect – a large hairy social bee which flies with a loud hum, living in small colonies in holes underground. Rowling herself said the reason for the name "Dumbledore" in her interview in 1999:

*"Because Albus Dumbledore is very fond of music, I always imagined him as sort of humming to himself a lot," Rowling said in a 1999 interview.*" [10]

We accepted the word "Dambldor" – to get pronunciation form of "Dumbledore" in Uzbek by omitting "e" at the end:

– Biz buni faqatgina taxmin qila olamiz, balki, hech qachon bila olmasmiz ham, – dedi Dambldor." [6, P.12]

– Тахмин қилишдан бошқа чора йўқ, – жавоб қайтарди Дамблдор, – Буни биз ҳеч қачон билмаймиз. [3, P.13]

Lord Voldemort is a mortal and the scariest character of all. The personage always thinks of government (governing both: wizarding and Muggle world) and fame. He wants to achieve this using immortality and dark arts. His followers are made to obey his rules. The only punishment of him is death: the spell of "Avera Ke Davra." Nobody dares to tell his name because of fright. The name of him adds more

power and scarcity for readers, as well. In Latin the word “valde” means “powerful,” “de morte” is “death.” It keeps this meaning in all European languages. Especially, in French this word is “flight from death,” in Norwegian “violence” is “valde.” [11] Unfortunately, the word “Voldemort” means nothing in the Uzbek language, it is just an attractive word. To transform it as “Ajal elchisi” or “Ajal hukmdori” could be the biggest mistake.

To most people’s astonishments, Tom Marvolo Riddle, which can be rearranged as “Voldemort”, also has its background. The shortened version of “Thomas” is “Tom” which means “twins” because Harry and Voldemort were twins at the end of the plot. “Marvolo” is derived from “mar” – “desire to damage”, “volo” means “I want” due to the Behind The Name web resource.

“Riddle” is the real riddle of Tom’s turning out to be Voldemort, Harry had to find that out in “Harry Potter and the Chamber of Secrets.”

Another case with the name Rubeus Hagrid, as we found, Rowling, the author revealed in a 1999 interview that “Hagrid” is a dialect word to mean “having a bad night”, as he is a heavy drinker, he has a lot of bad nights which opens another secret behavior of the personage, however, few readers are aware of the fact. “Rubeus” is reported to mean “turn red,” from the Latin origin “Rubeo,” as his face is getting red when he starts drinking by Pottermore project of Rowling’s web page. [12] It would be nice to transfer the semantics of the name into Uzbek, which is, unfortunately, impossible.

In both Uzbek translation versions of the book, we can see the word as “Hagrid” without a change.

“Dambldor bu sayyoraldan nimanidir aniqladi, shekilli, soatni cho’ntagiga qaytarib solib qo’ydi va: – Hagrid kech qolyapti, axir, uning o’zi vaqtida etib boraman demaganmidi? – dedi.” [6, P.12]

“Шундай бўлса-да, Дамблдор ўз соатини яхши тушунади шекилли, уни чўнтагига қайтариб солиб қўйди. – Хагрид кечикяпти, – деди у, – Мени бу ерга келишимни ўша айтган бўлса керак Сизга?” [3, P.13] Draco Malfoy is meant to be and to sound a bad-tempered personage in the novel. “Draco” is “dragon” for its Latin roots from ancient times to express power besides him and his family. “Mal” is known to be a French word meaning “evil-hearted” plus “foy” is the result of changed “foi” in French for “belief.” Therefore, he has always had an evil belief in the idea “mud- blood.” They accepted “Drako Malfoy” as an Uzbek version.

Another hero of Harry Potter, Sirius Black, is the quirk (paronymy) of linguistics in literature. “Sirius” is known to be the brightest star, which can be seen from our planet in the sky. In addition, it is a part of Canis Major stars, the Great Dog constellation. So the character can turn into a dog in the full moon. The name has also accepted without any serious changes into Uzbek.

Finally yet importantly, the most famous surname (in the first story) – the Dursleys, the author says she chose this only for her dislike towards the name. Intriguingly, the writer picked the “Dudley”, the son of the Dursleys, “Dursley” looking at the map, areas where she has never been to. Going inside the writer admits she chose each name of the Dursleys only why she did not care about them or feel unpleasant when she heard. Anyway, we accepted “Dadley” for “Dudley” and “Durzleylar” for “Dursleys”, “Petunia”, and “Vernon” with no change. There is only little change in the pronunciation of the phone “s” to “z” and the affix “lar” are added to the family names to mean the whole family in Uzbek.

A funny character – Argus Filch – the caretaker of Hogwarts school has its own story going back to Greek mythology where “Argus” – a watcher giant with a hundred eyes. And “filch” seems to be “to steal something” [13, P. 334] in informal speaking in the UK which is alleged to have a connection to the behaviour of the character. The translators transfer the name without any phonemic and graphic alteration into Uzbek.

Gregory Goyle and Vincent Crabbe are two friends of Draco Malfoy whose name does not mean more than his bodyguards as they are described to stand on two sides of Draco. “Gregory” is the result of a Greek word to mean “observant” and “watcher”, as it comes “Goyle” sound much more like “Gargoyle” which is a Gothic-style building to refer his body and vibe to the image of Malfoy.

“– Aytgancha, bular – Kreb va Goyl, – bepisand gapirdi u Garrining nigohini payqab, – mening ismim esa Malfoy, Drako Malfoy.” [6, P.102]

A reader witnesses in the novel Vincent Crabbe and Goyle are always together, especially, with Draco as it given in the extract above. Vincent derived from the Latin “Vincere,” meaning “to defeat” and is compatible with what he wants to accomplish with Goyle and Malfoy.

Salazar Slytherin is a co-founder of four houses of Hogwarts School of Witchcraft and Wizardry with Godric Gryffindor, Helga Hufflepuff and Rowena Ravenclaw. The name of the House was unique

as it was named after a fictitious personage, hence we only used transliteration here to accept it as “Salazar Slizerin” into Uzbek. The semantic meaning and etymology of the word may not be transformed into Uzbek with for at least two reasons: at first, the word “Slytherin” is the changed form of “slithering” in English which means to move like a snake. If we could translate the meaning behind, the word would be “sudralish” as the result “Salazar Sudralerin” (“sudralish” is “slide” and “slither”) needs editing. Because “Sudralar” itself is often used alone, and when it is alone the complex alliteration does not happen. “Sudralar” is a verb in Uzbek so we have to get the adjective form of the word. As it seems, the process is so confusing that the result is not as expected.

Moreover, the Hogwarts founder’s names – Godric Gryffindor, Helga Hufflepuff, Rowena Ravenclaw also make use of alliteration, which gives them additional poetic and aesthetic functions: the reader can enjoy the sound and wordplay of the names. Thus, these names are an instance of both semantic and phonesthetic meaning, which the translations variously maintain. The Uzbek, Russian and German translations offer the poetic enjoyment of the rhyming names without transferring their semantic meaning, as the names are not translated [14]. The Finnish translators, in contrast, has access to the semantic levels without the poetic enjoyment of alliteration, as the names are rendered as Godric Rohkelikko (rohkea – ‘brave’), Helga Puuskupuh (puuskuttaa ja puhalttaa ‘to huff and puff’), Rowena Korpinkynsi (korppi ‘raven’, kynsi ‘claw’) and Salazar Luihuinen (luihu ‘guileful, sly’). And it would be acceptable and would draw a vivid picture of the characters (cofounders of Hogwarts) to translate their names with alliteration like, Helga Hufflepuff to “Helga Hansiruff”, derived from “hansiramoq” and “uff” tortmoq” in Uzbek, Rowena Ravenclaw to “Rovena Roqarg’achangal”, from “qarg’a” and “changal” in Uzbek, and Salazar Slytherin to “Salazar Sudralar.”

By the way, the word “House” is “uy” in Uzbek. Despite we do not use “uy” for various fields of the study; we could choose “fakultet” (faculty) for “house.” However, the word refers to the different field of study from others like “The faculty of mathematics” and “The faculty of languages.” In spite of the semantic limit of the word “faculty,” we have to accept it, as we do not have equivalent to the word [15]. Though we have “yo’nalish” in Uzbek, it is the same as the “faculty.” Originally, it means “branch of knowledge”, in an academic sense, from Old French.

Considering that the “house” is the building or accommodation, we can render it as “uy” (house) which could not be equivalent.

“Шахсан мен ўзим ҳақимда сўраб билдим. Умид қиламан-ки, “Триффиндор”га қабул қилишади мени. Айтишларига қараганда, энг зўр факультет “Триффиндор” экан.” [3, P.78]

“– Namunali talabalar! – norozi ohangda soʻz boshladi u. – Fakultetingiz talabalarini shu zahotiy oq oʻz yotoqxonalariga olib boring.” [6, P.102]

“– А я играю. Отец говорит, что будет преступлением, если меня не возьмут в сборную факультета, и я тебе скажу: я с ним согласен. Ты уже знаешь, на каком будешь факультете?” [4, P. 38].

### CONCLUSION

The names in the article are from two fields of literature: fantasy fiction and literature aimed at children. Translators could have more opportunity to translate the names in literature aimed for children. Fantasy fiction is more difficult, though. Nevertheless, what happens when the writer prepares a tight package with the mixture of names for fantasy fiction and literature aimed at children? The answer comes as the mixture, as well, it is positive and negative; some can be translated and some not. This depends on the position of a writer and a translator: a writer is the ruler of a translator, the names should stay the same, in some cases, and translators endeavour to translate them into their mother tongue to make it predictable competing with children’s literature. Moreover, sometimes they have to make inter-lingual transformations not only on personage names but also on names of plants, wizarding books, meals and spells.

### REFERENCES:

1. <http://www.gazette-du-sorcier.com/De-nouvelles-traductions-pour-Harry-Potter>.
2. <http://www.hachette.com/en/maison/azbooka-atticus>.
3. Dolimov Sh. Z. Garri Potter va falsafiy tosh. – Toshkent, 2008. – 108 b.
4. Спивак М. В. Дж. К. Ролинг. Гарри Поттер и философский камень. – Россия: 2000. С – 160.
5. <http://www.merriamwebster.com/dictionary/querulous>.
6. Solijonov J. K. Garri Potter va Afsungarlar toshi. – Termiz: Surxon-Nashr.
7. <https://kun.uz/uz/news/2017/06/23/arri-potter-aslida-ikkita>

8. ulganmi?q=%2Fnews%2F2017%2F06%2F23%2Garri-potter-asli-da-ikkita-bulganmi
9. <https://kun.uz/uz/news/2017/06/26/arri-potter-akidagi-birinci-ertakka-20-jil-tuldi>
9. <https://kun.uz/uz/114051?q=%2F114051>
10. <http://www.accio-quote.org/articles/1999/1099-connectiontransc2.htm>
11. [https://harrypotter.fandom.com/wiki/List\\_of\\_characters\\_in\\_translations\\_of\\_HarryPotter](https://harrypotter.fandom.com/wiki/List_of_characters_in_translations_of_HarryPotter)
12. <https://www.pottermore.com/features/the-definitive-rubeus-hagrid-quiz>
13. Oxford University. Oxford Dictionary of Current English. – USA: Oxford University press, 2006.
14. Худайкулов, Абдулла Эшкуватович. «Семантическая адаптация некоторых английских заимствованных слов в русских письменных памятниках xvii-xix веков.» Вестник Челябинского государственного университета 28 (2011).
15. Shirinova, Raima, et al. “Representation of the national picture of the world in literary translation.” *International Journal of Advanced Science and Technology* 29.5 (2020): 144 6.

УДК 81-11

## THE PRAGMALINGUISTIC FEATURES OF COORDINATING CONJUNCTIONS

*Abduqodirova Shohista Tohir qizi*  
*The student of filology faculty of BSU*  
 shohista@gmail.com

In this article the pragmatic meaning and function of the conjunctions that highly effect to the meaning of the sentence were learnt. The types of conjunctions were informed. The coordinating conjunctions were emphasized.

**Key words:** Conjunctions, coordinating conjunctions, correlative conjunctions, subordinating conjunctions, conjunctive conjunctions, negative conjunctions opposite conjunctions, pragmatics, presuppose, speech situation, figurative meaning.

N. Abdurahmanov divided and classified the words used as conjunctions and conjunctions into three groups – coordinating conjunctions, correlative conjunctions, subordinating conjunctions:

In the old Uzbek language conjunctions are divided into two types according to their function: 1) coordinating conjunctions 2) conjunctive conjunctions. B. Abdushukurov in his work (textbook) entitled “History of the Uzbek language” also noted that conjunctions in the old Uzbek language are divided into 2 types: 1) coordinating conjunctions 2) conjunctive conjunctions. In the work “Morphology of the Uzbek language of the 19th century” by A. Matgoziyev, coordinating conjunctions used in the old Uzbek language are divided into the following groups according to their function and meaning: a) coordinating conjunctions: va, -u(-yu, -vu), ham (ham..., ham), taqi // dag‘i, yema, ma ; b) correlative conjunctions: ya (yi..., ya), yaki, yaxud (yaxud..., yaxud...), xih..., xah (xih..., xah...), azu, gah (gah..., gah...) v) opposite conjunctions: amma, lekin, lek, vaekin, valek, vale; g) negative conjunctions: ne..., ne..., chi..., chi...

Currently the following can be cited as the grammatical features (meaning) of conjunctions in the Uzbek language:

- a) does not have an independent lexical meaning;
- b) cannot be grammatically formed;
- c) is an answer to a specific question and cannot act as a part of a sentence;
- d) is used in various auxiliary meanings and tasks;
- e) serves to connect the united parts of the sentence and the simple sentences in the compound sentence.

In our opinion, conjunctions differ from other word groups by the above-mentioned grammatical features, and these grammatical features serve as a basis to define conjunctions. A conjunction is a word that serves to connect the parts of united clauses, sentences in a compound sentence, and in this way to express various relationships between them. In other words, a conjunction is one of a group of auxiliary words that can be used to connect clauses and simple sentences together.

Conjunctions in the Uzbek language are covered in a monographic aspect and in dissertations and a number of scientific articles. Linguist A. Mukhtorov has carried out significant work in the field of studying the grammatical features of connectives for many years. Scientific articles of the scientist such as “Conjunctions in Uzbek”, “Subjunctive conjunctions in Uzbek”, “Usage of subjunctive conjunctions in simple sentence structure” were published in scientific collections of SamSU. In 1953, A. Mukhtorov defended his candidacy thesis on “Conjunctions in the modern Uzbek language”. Conjunctions in Uzbek are divided into two types according to their function in the sentence: a) coordinating conjunctions b) conjunctive conjunctions

The definition of coordinating conjunctions is as follows: conjunctions that connect clauses and clauses that have an equal relation to each other are called coordinating conjunctions. Coordinating conjunctions not only connect united clauses and sentences, but also create the meaningful relationship between them. For example: Boyvachcha kissed the ground, repented, and immediately rushed to his elderly parents to please them (N. Aminov). The conjunction “and” is used in this sentence. In the work “Current Uzbek Literary Language”, coordinating conjunctions are divided into four groups according to their meaning in the text. a) linking conjunctions: and, as well as, as well as with; b) contradictory connectives: but, but, but as; c) disjunctive conjunctions: or, or, or, of course, occasionally, occasionally, occasionally, one by one, sometimes; g) negative conjunction: na-na. The performance of pragmatic tasks of Uzbek language conjunctions in connection with factors such as specific text-speech situation has been partially reported in some studies. For example, A.B. Pardayev’s monograph “The role of auxiliary word groups in the Uzbek language in the linguistic system and linguo-pragmatic analysis” analyzed the presupposition of some conjunctions. The author writes the following about conjunctions creating presuppositions. “Our observations have shown that a certain number of conjunctions can be included among the means of expressing



active presupposition.” The scientist explains with examples that some conjunctions refer to presupposition. It can be seen that conjunctions can also perform a pragmatic function by pointing to a presupposition. In a connecting sentence, it is used to connect some simple sentences, connected parts, and syntactically-semantically relate to each other. Their appropriate and purposeful use in the process of speech prevents possible stylistic insanity, and their use in speech creates pragmatic (derivative) meanings.

In this chapter of our work, issues related to the emergence of pragmatic (derivative) meanings arising from the use of connectives in speech are covered. There must be at least two participants in the process of information exchange. The communicative action of one participant necessarily requires a response action: the question does not remain unanswered, it requires a request, agreement, denial, threat, reaction to the opinion. The interpersonal relationship that occurs in the process of speech communication cannot be without an object, because the subject and the object are interdependent and cannot be used separately. Consider the following text:

– ”Allah is a witness, may your services for the happiness of the country lead Afghans to the path of happiness.

– Say so... Then what do your young people want?

– Young people? Tarzi took his eyes off Amir and stroked his beard.

– Tell the truth, don't be afraid of my anger. I need to be told the truth.

– Young people want to see Afghanistan independent. Good time is passing, state asylum.

– The opportunity has not yet come.

– The opportunity came with a German delegation.

– Is that so? What is the reason why I missed the opportunity? Am I scared?

– No, my great emir, the reason is different: you get two million four hundred thousand rupees a year and fifty million rupees to be given after the war. Amir glanced at the chair across the room. The letter written by Tarzi demanding independence and fifty million is there”.

In the quoted passage, the interpersonal relationship that arose in the process of speech communication did not emerge only with lexical units. The units that are the objects of study of pragmatics also have a “share” in the emergence of interpersonal relations. That is, the elementary (speech act – linguistic act, any elementary meaning intended by the speaker) of language units that occurs only during speech com-

munication is also expressed. Relationships that arise in the framework of communication are different and have a certain informational content. As it is rightly stated in the theoretical literature, "Any method of expressing content or opinion is a different context, and the meaning of context is also understood under the concept of a method of expressing an opinion through verbal or non-verbal means."

Depending on the text, they can acquire various pragmatic and derived meanings. For example: In a fairy tale, there is talk about a stork and an old man. In the sentence, the words "stork", "old man" are connected with the help of conjunction "and". In this compound, the speech act meaning of the conjunction "and" known to the listener is considered as unity, equality. Let's expand one of the units in this compound (stork and old man) with a determiner. For example: A fairy tale tells about a stork and an old man who hunted it. The speaker's expression of the speech in such a form and its conjunctions can lead to the loss of the above-understood unity and equality act meaning. Now the listener perceives the speech act content not in the form of a stork and an old man (old man with a stork), but in the form of an old man hunting a stork. In order to determine whether or not the conjunction really influenced the meaning of the conjunctions, we will analyze the conjunctions used in the above speech without conjunctions: 1) the stork and the old man 2) the stork and the old man who hunted it. When the first conjunction is used in the style of stork, old man (that is, the story talks about stork, old man), the meaning of equality is weakened, and the meaning of emphasis is strengthened through the tone of counting. The second word compound does not have the meaning of equality when expressed by the speaker in the form of the stork, the old man who hunted it, but it has the meaning of emphasis. In the emergence of the pragmatic meaning of the conjunction "and", not only the speech process, but also the speaker's intention and the listener's awareness of the elements of speech are important.

In conclusion, the requirement of consistent distinction between language and speech in the interpretation of pragmatic meaning requires the separation of stable linguistic and transient speech aspects specific to each word. Since speech phenomena have the nature of infinity and variety, they cannot be the object of lexicographical interpretations.

## REFERENS:

1. Абдуллаева М. Ўзбек тилида боғловчилар таснифига доир айрим мулоҳазалар. – Б. 67.
2. Abdushukurov V. O'zbek tili tarixi. – Toshkent: Nodrabegim nashriyoti, 2021, – Б. 154.
3. Асқарова М. Ўзбек тилида боғловчикламалар ҳақида // Ўзбек тили ва адабиёти, 1962, 5-сон.
4. Ҳожиёв А. Ҳозирги ўзбек адабий тили боғловчилари ҳақида мулоҳазалар // Ўзбек тили ва адабиёти, 1986, 2-сон, – Б. 56.
5. Матғозиев А. XIX аср ўзбек тилининг морфологияси.– Т.: Фан, 1977, – Б. 184–187.
6. Матғозиев А. Ўзбек тилидан эргаштирувчи боғловчиларнинг тарққиёт тарихи. Филол. фан. номз. диссертацияси. Тошкент, 1966.
7. Мыркин И. Я. Текст подтекст и контекст. – М.: 1976. – № 2. – с. 91.
8. Мухтаров Ж. Союзы в современном узбекском языке. Автореф. дисс канд филол. наук. Самарканд, 1953.
9. Раҳматулла Қўнғуровнинг илмий мероси ва Ўзбек тилшунослиги масалалари. Илмий мақолалар тўплами. Самарканд, СамДУ нашри, 2008. – Б. 130–134.
10. Расулов Р. Ўзбек тилда ёрдамчи сўзларнинг семантик-грамматик хусусиятлари. – Т., Фан, 1983.
11. Zakirova N. Ergashtiruvchi boғlovchi vositalarning struktur-semantik tadqiqi: filol fan nomz. diss-ya. – Toshkent, 2011.
12. DN Yuldasheva. BASIC NON-VERBAL COMPONENTS OF SPEECH. CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (2767-3758) 3 (01), 17-24. <https://masterjournals.com/index.php/crjps/article/view/661>
13. Шоабдурахмонов Ш. "Ўзбек тилида ёрдамчи сўзлар", Тошкент, 1953.

## УДК 81-11

## XUSUSIY ATAMALAR NOXUSUSIY DISKURLARDA

*Shaxabitdinova Shoxida Xoshimovna<sup>1</sup>,  
Abduraxmanova Madina Ulugbekovna<sup>2</sup>*

*<sup>1</sup>filologiya fanlari doktori, professor  
Andijon davlat universiteti*

*<sup>2</sup>filologiya fanlari bo'yicha falsafa doktori  
Andijon davlat tibbiyot instituti*

Ushbu maqolada xususiy atamalarning noxususiy diskurlarda qo'llanishi haqida fikrlar keltirilgan. Xususan, "ko'rlik" kasalligining lisoniy tadqiqi amalga oshirilgan. Ko'rlik atamasining tibbiy izohlaridan kelib chiqib, badiiy asarlar, gazeta va jurnallar, korpus lingvistikasi materiallari, shuningdek, og'zaki nutqdagi misollar yordamida tahlili olib borilgan. Maqolada "ko'rlik" kasalligining nomi ko'chma ma'nolarda qo'llanilishi tadqiqi amalga oshirilgan.

**Kalit so'zlar:** ko'rlik, so'qirlik, ko'zi ojizlik, semalar, ko'chma ma'no, kasallik nomi.

## PRIVATE TERMS IN NON-PRIVATE DISCOURSES

*Shakhabitdinova Shokhida Khoshimovna<sup>1</sup>,  
Abdurakhmanova Madina Ulugbekovna<sup>2</sup>*

*<sup>1</sup>Doctor of Philology,*

*Professor Andijan State University*

*<sup>2</sup>Doctor of Philosophy in Philological Sciences  
Andijan State Medical Institute*

This article examines the use of private terms in non-private discourses. In particular, a linguistic study of the disease "blindness" was conducted. Based on medical interpretations of the term "blindness," the analysis was carried out using works of art, newspapers and magazines, corpus linguistics materials, and examples from oral speech. The article conducted a study of the use of the name of the disease "blindness" in a figurative meaning.

**Key words:** blindness, semes, figurative meaning, name of the disease.

## ЧАСТНЫЕ ТЕРМИНЫ В НЕЧАСТНЫХ ДИСКУРСАХ

*Ш. Х. Шахабитдинова<sup>1</sup>, М. У. Абдурахманова<sup>2</sup>*

*<sup>1</sup>доктор филологических наук, профессор*

*Андижанский государственный университет*

*<sup>2</sup>доктора философии по филологическим наукам*

*Андижанский государственный медицинский институт*

В данной статье рассматривается использование частных терминов в нечастных дискурсах. В частности, было проведено лингвистическое исследование болезни «слепота». На основе медицинских трактовок термина «слепота» анализ проводился с помощью произведений искусства, газет и журналов, материалов корпусной лингвистики, а также примеров из устной речи. В статье проведено исследование употребления названия заболевания «слепота» в переносном значении.

**Ключевые слова:** слепота, семы, переносное значение, название заболевания.

Dunyo tilshunosligida ikki til imkoniyatlarini o‘rganish, bir tildan ikkinchi tilga tarjima jarayonida yuzaga keladigan fikriy ifoda, uning shakliy tuzilishi, ichki nutq imkoniyatlarini va aniq struktur xususiyatlarini tahlil qilishga qaratilgan tadqiqotlar yaratilmoqda. Diskurs jarayoni, xususan tibbiy diskurs so‘zlovchi va tinglovchi o‘rtasida nutq vaziyatining o‘rni, adresat va adresant lisoniy ongi, madaniyatlararo munosabat belgilari, baho va talqin muammolari tadqiqi til birliklarining nutqda reallashuv imkoniyatlarini yoritishga xizmat qiladi.

Elektron korpuslarda va lug‘atlarda tibbiy atamalar odatda professional jihatdan ta’riflanadi, misollar ham aynan muayyan tib sohalariga doir bo‘ladi. Vaholanki, tibbiy atamalar tibbiyotdan tashqari diskurslarda ham keng tarqalgan. Ularni tadqiq etish korpus lingvistikasiga oid ma’lumotlarni kengaytirishda salmoqli ahamiyatga ega. Tibbiy atamalarning tibbiyotdan tashqari diskurslarda qo‘llanishini aniqlashda komponent tahlil usulidan foydalanish eng maqbul yo‘llardan biridir.

Jumladan, o‘zbek xalqining og‘zaki nutqida ishlatiluvchi atamalardan *so‘qirlik*, *ko‘zi ojizlik* ya’ni *ko‘rlik* kasalligiga tibbiy lug‘atlarda quyidagicha ta’rif berilgan:

*Ko‘rlik, so‘qirlik, ko‘z ojizligi – ko‘z ko‘rish qobiliyatining butunlay yo‘qolishi, yorug‘likni qorong‘ulikdan ajrata olmasligi. Bunda ko‘rish quvvati “0” bo‘ladi. Ko‘rlikka ko‘pincha glaukoma, traxoma, blennoreya, ba‘zan muguz parda, ko‘ruv nervi va to‘r parda kasalliklari sabab bo‘ladi. Ko‘rlik kelib chiqishiga ko‘ra irsiy, tug‘ma va orttirilgan bo‘lishi mumkin...[1;112].*

Ko‘rlik atamasi hammaga tanish atama bo‘lib, tibbiy izohlarda ko‘rish faoliyatining yo‘qolishi ta’rifi keltirilgan. So‘qirlik (ko‘r, ko‘rlik) atamasini ko‘chma ma’noda ishlatilish holatlarini tadqiq qilishda, avvalo atamaning semalari aniqlanadi [2; 174]:

- 1) ko‘zning ko‘rmasligi;
- 2) qorong‘ulik;
- 3) ojizlik;

4) yon-atrofdagi voqea-hodisalarni sezmaslik;

5) salbiy tus.

Koʻrlik atamasining semalarini aniqlab olgach, gazetalarda, hikoyalarda va boshqa matnlardan koʻchma maʼnolarini aniqlab, tadqiq qilinadi.

**1-misol.** *“Garchi boʻlsada bunga ehtiyoji,  
Zolimga yoqmasdi dinning rivoji.  
Yorugʻlikni koʻrsa gar qalbi koʻrlar,  
Dilga shayton qurgan tuzoqu toʻrlar”.*

Bu sheʼrda *“qalbi koʻrlar”* deb insofi yoʻq, iymonsiz odamlar nazarda tutilmoqda. Demak, aniqlangan semalar tartibida oʻzgarish kuzatilmoqda, salbiy tus maʼnosi yorqinlashuvi sodir boʻlmoqda.

1) salbiy tus;

2) yon-atrofdagi voqea-hodisalarni sezmaslik;

3) ojizlik;

4) qorongʻulik;

5) koʻzning koʻrmasligi.

**2-misol.** *“Koʻrlar bozorida yoʻqdir soʻrovchi surmani*

*Dardi yoʻqlardin topilmas koʻzlaridan tomchi yosh”* [3; 217].

Bu misrlardagi *“Koʻrlar bozori...”* iborasida dardi yoʻq inson, birovning tashvishlarini oʻylamagan inson nazarda tutilyapti, yaʼni beʼtiborlik, befarqlik maʼnolari yuzaga chiqmoqda.

1) yon-atrofdagi voqea-hodisalarga befarqlik;

2) salbiy tus;

3) ojizlik;

4) qorongʻulik;

5) koʻzning koʻrmasligi.

**3-misol.** *“Sodiqlikdan menga vaʼzlar oʻqir edi,  
Chinmi bilmam balki yolgʻon toʻqir edi,  
Bilolmadim kimning koʻzi “soʻqir” edi,  
Yolgʻonlarni oʻziga doʻst bilgan doʻstim.”*

Bu sheʼriy misradagi *soʻqir* soʻzi yordamida haqiqatni koʻrmaydigan, yolgʻon gaplarni gapiruvchi inson nazarda tutilmoqda, 5-sema yorqinlashib qolgan semalar xiralashmoqda.

1) salbiy tus;

2) qorongʻulik;

3) ojizlik;

4) yon-atrofdagi voqea-hodisalarni sezmaslik;

5) koʻzning koʻrmasligi.

**4-misol.** “*So ‘qir kecha. Nurafshon osmon.  
Ko ‘zda yonib yurakda bori  
Tuproq yo ‘ldan qaytmoqda dehqon,  
Yuz-ko ‘zida dala g ‘ubori.*” [4]

Ushbu she’riy misrada “*So ‘qir kecha*” qorong‘u tun ma’nosini izohlamogda, qorong‘ulik ma’nosi yuzaga chiqib, qolgan semalar xiralashmogda.

- 1) qorong‘ulik;
- 2) ojizlik;
- 3) ko‘zning ko‘rmasligi;
- 4) yon-atrofdagi voqea-hodisalarni sezmaslik;
- 5) salbiy tus.

**5-misol.** “... *Baland-baland so ‘qir paxsa devorlar osha ko ‘cha to-monga shaftoli, olhirot, ba ‘zan aymoqi tog ‘uzumlar osilib tushgan...*”

Bu jumlada “*so ‘qir paxsa devorlar*” deb eshik va derazasi yo‘q devorlar nazarda tutilgan. Eshik, derazalari yo‘q devor ortida hech narsa ko‘rinmaydi, demak, ushbu gapda 1-sema yuzaga chiqib, qolganlari xiralashgan [5; 57].

- 1) ko‘zning ko‘rmasligi;
- 2) qorong‘ulik;
- 3) ojizlik;
- 4) yon-atrofdagi voqea-hodisalarni sezmaslik;
- 5) salbiy tus.

O‘zbek tilidagi matnlar tahlilidan so‘ng ko‘rlik, so‘qirlik atamalari o‘zbek tilida ko‘chma ma’noda ishlatiladi, shuningdek, so‘qirlik atamasidagi har beshala sema muayyan holatlarda yorqinlasha oladi deb xulosa qilish mumkin.

Ko‘rlik atamasining semantik tahlili va kontekstual tadqiqi borasida to‘plangan tajriba nafaqat kasallik nomlari hamda tibbiy atamalar, balki barcha ixtisoslashgan so‘zlarning ixtisoslashmagan diskurslarda realizatsiyasining barcha holatlari uchun qo‘llanishi mumkin

#### FOYDALANILGAN ADABIYOTLAR RO‘YHATI

1. To‘raqulov Y. va b. Tibbiyot qomusiy lug‘ati. – T.: “Qomuslar bosh taxiriyati”, 1994. – 288 b. – B. 112.

2. Abduraxmanova M.U. “Ko‘rlik” kasalligining o‘zbek, rus, ingliz tillaridagi lingvistik-semantik tadqiqi // Buxoro davlat universiteti ilmiy axboroti, 1/2023. – Buxoro, 2023. – B. 174–179.

3. Alixonto‘ra Sog‘uniy. Baxtlik nedir? She‘rlar to‘plami. – Toshkent-Istambul, 2021. – 246 b. – В. 217.
4. Davron X. To‘maris ko‘zlari// “Yoz” she‘ri. – Т.: “Yosh gvardiya” nashriyoti, 1984.
5. G‘ulom G‘. Mukammal asarlar to‘plami. Beshinchi tom. Povestlar. Pyesalar. – Т.: Fan, 1986. – 356 b. – В. 57.
6. Юлдашева Д. Н. РОЛЬ МОЛЧАНИЯ В РЕЧИ // ОБРАЗОВАНИЕ И ИННОВАЦИОННЫЕ ИССЛЕДОВАНИЯ. № 1(2020):<https://interscience.uz/index.php/home/article/view/31>



УДК 81-11

**КИНОЛИНГВИСТИКА В КАЧЕСТВЕ ОТРАСЛИ  
ЯЗЫКОЗНАНИЯ*****М. Ш. Ширинова****Бухарский государственный университет*  
mehrigiyoshirinova@gmail.com

В статье идёт речь о том, что кинолингвистика является отраслью, возникшей в точке пересечения кино и лингвистики рассматривается значимость киноязыка и его изучения. Было отмечено, что в мировом языкознании существует ряд лингвистических исследований языка кинофильмов, которые были подвергнуты анализу. Также было отмечено, что данная область актуальна на сегодняшний день, имеет практическую суть и ждёт исследователей узбекского языкознания.

**Ключевые слова:** кино, киноязык, кинолингвистика, языкознание, искусство, кинотекст.

**KINOLINGVISTIKA – TILSHUNOSLIKNING  
BIR TARMOG‘I SIFATIDA*****Shirinova Mexrigiyo Shokirovna****Buxoro davlat universiteti**Buxoro, O‘zbekiston*

mehrigiyoshirinova@gmail.com

Maqolada kinolingvistika tilshunoslikning kino va lingvistika tutashgan nuqtasida yuzaga kelgan tarmoq ekanligi, kino tili va uni o‘rganish ahamiyati haqida so‘z boradi. Jahon tilshunosligida kinofilmlar tili lingvistik tadqiq etilgan tadqiqotlar bir qancha ekanligi ta’kidlanib tahlilga tortilgan. Shuningdek, bu sohaning bugungi kun uchun dolzarbligi, amaliy mohiyatga molikligi, o‘zbek tilshunosligida tadqiqotchilarini kutayotganligi haqida ham aytib o‘tilgan.

**Калит so‘zlar:** kino, kino tili, kinolingvistika, tilshunoslik, san’at, kino matni.

**KINO – LINGUISTICS AS A BRANCH OF LINGUISTICS*****Shirinova Mexrigiyo Shokirovna****Bukhara State University*

mehrigiyoshirinova@gmail.com

The article deals with the fact that film linguistics is a branch that emerged at the point of intersection of film and linguistics, and considers the significance of film language and its study. It was noted that there are a number of linguistic studies of film language in the world linguistics, which have been analysed. It was

also noted that this field is relevant today, has a practical essence and is waiting for researchers of Uzbek linguistics.

**Keywords:** cinema, film language, film linguistics, linguistics, art, film text.

Bugungi kunda tilshunoslik fani ham boshqa fanlar qatori shiddat bilan taraqqiy topib bormoqda. Ayniqsa, sohalar oʻrtasidagi integratsiyaning tilshunoslikka kirib kelishi soha lingvistikasining rivojlanishiga keng yoʻl ochib berdi. Bu esa jamiyat amaliy lingvistikaga qanchalik ehtiyojmand ekanligini isbotladi. Chunonchi, oʻzbek tilshunosligida psixolingvistika, etnolingvistika, kompyuter lingvistika, korpus lingvistika, tibbiyot lingvistika, lingvistik ekspertiza, medialingvistika kabi bir qator yangi yoʻnalishlar oʻz faoliyatini boshladi hamda monografik tadqiqotlari bilan tilshunoslik ilmiga oʻz ulushini qoʻshib kelmoqda.

Kinolingvistika – kino va tilshunoslik sohalarini kesishgan nuqtada yuzaga kelgan tilshunoslik yoʻnalishi. Bunda kinofilmlar tili va ularning lingvistik xususiyatlari oʻrganiladi. Sohada faoliyat yuritish ham kino sanʼatidan, ham tilshunoslik ilmidan xabardor boʻlishni talab etadi.

Sanʼatlar orasida tomoshabinni tezda oʻziga rom qilib, ekran oldidan siljishga yoʻl qoʻymaydigan, koʻngilochar tur kinodir. Kino yunoncha *kineo* “harakatga keltiraman” degan maʼnoni anglatib, oʻzlashma qoʻshma soʻzlarning tarkibiy qismi sifatida kinematografiyaga, kinofilimga aloqadorlik maʼnosini bildiradi [3,368-b.]. Qahramonlar, soʻzlar, rang-tasvir, kadr va boshqalarni “harakatga keltirib”, katta-yu kichikni oʻziga mahliyo qilgan ushbu sanʼat turi oʻziga xos tilda soʻzlaydi.

Dastlab Parijda 1895-yil 28-dekabrda ixtirochi aka-uka Lui va Ogyust Lumerlar oʻzlari kashf qilgan texnika yordamida “Poyezdning kelishi”, “Bolani ovqatlantirish” kabi kinometrajlarni suratga olishgan. Ushbu kun tarixda kino sanʼatining vujudga kelish kuni deya muhrlandi [4, 3-b.]. Shundan buyon kino ommaga, jamiyatga qaratilgan sanʼat turi sifatida baholanib, undagi har bir vosita tarbiya kaliti vazifasini oʻtab kelmoqda.

Bugungi kunda oʻzbek milliy kino sanʼati jamiyatimiz hayotida oʻz oʻrni va mavqeiga ega. Jahon kino sanʼatida ham munosib oʻringa ega boʻlishi uchun hukumatimiz tomonidan munosib sharoitlar yaratilmoqda. Prezidentimiz Sh.M.Mirziyoyev kino sanʼatiga quyidagicha taʼrif bergandilar: “*Dunyo miqyosida turli gʻoya va mafkuralar kurashi, odamlar, ayniqsa, yoshlarning qalbi va ongini egallash uchun turli*

*kuchlar o'rtasida raqobat keskin tus olgan bugungi kunda eng omma-viy san'at bo'lgan kinoning beqiyos imkoniyatlaridan samarali foydalanishimiz kerak*" [1]. Ushbu ta'rifning davomi sifatida yurtboshimiz tomonidan yana bir qator qarorlar qabul qilindi, vazifalar belgilandi. Jumladan, O'zbekiston Respublikasi Prezidentining 2020-yil 24-yanvardagi Oliy Majlisga murojaatnomasini ham aytish mumkin. Murojaatnomada kinematografiyani 2030-yilgacha rivojlantirish konsepsiyasi hamda "Kinematografiya to'g'risida"gi Qonun ishlab chiqilishi muhim masalalar qatorida sanab o'tildi [2]. Bu kabi islohotlarning amalga oshirilishi kino sohasining har tomonlarga keng rivojlanishiga yo'l ochib beradi, deyish mumkin.

Kino tili, eng avvalo, ekran tilidir. Uni o'rganish muammosi tanqili xorijiy olimlarni hamisha qiziqtirgan. Jumladan, S.Eyzenshteyn, L.Zayseva, R.Arnxeym, B.Balash, 3.Krakauer, V.Shklovskiy, A.Bazen, M.Marten, E.Teplits va boshqalarning asarlarida kino tilining o'ziga xos xususiyatlari o'rganildi. Shuningdek, kinoning ekspressivligi, uning ildizi, san'at sohasida tutgan o'rni I.Vaysfeld, A.Kazin, L.Kozlov, V.Mixalkovich, K.Razlogov va boshqa ko'pgina kinoshunolar tomonidan tadqiq etildi. Kino va kino tili tushunchalarining nazariy asoslarini lingvosemiotik tushunchalar asosida R.Bart, Y.Lotman, Chr. Metz, P.Pazolini, B.Eyxenbaum, R.Yakobsonlar; kino tilining falsafiy jihatlarini V.Saltini; kinoning estetika va psixologiya bilan uyg'unligi masalasini J.Mitri; kino tilining psixologiya, estetika va semiotika bilan aloqadorligini R.Bellur kabi olimlar atroflicha o'rgangan [5,6,7].

Dastlabki kino matni bilan bog'liq tadqiqotlar rus formal maktabining nazariyotchilariga tegishli. Ushbu maktab vakili V.Shklovskiy ilk marta "kino tili" va uning me'yorlari tushunchasini olib kirdi ("О киноязыке" 1926) [10,35-b.], uning mazmun-mohiyatini o'lchash uchun *kinoso'zni* tanladi hamda *kinoso'z* deganda "ma'lum bir ahamiyatga ega bo'lgan kinematografik materialning bir qismi"ni tushunadi ("Семантика кино" 1925) [10,33-b.].

O'zbek madaniyati va san'atida ham kino sohasiga oid bir qator manbalar mavjud [4], biroq ushbu manbalarning barchasida kino san'atining bir turi sifatida tadqiq etilgan, kino tili borasida maxsus to'xtalilmagan. Ammo bir necha o'zbek kinofilmlari misolida umumiy planda kino tilining adabiy tilga ta'siri yuqori ekanligi ta'kidlangan maqola mavjud [32,10-b.].

Umuman olganda, dunyoda kino va kino tili sohasida olib borilgan tadqiqotlarni yuzdan ortiq sanashimiz mumkin. Bu nimadan dalolat beradi? Izlanishlarimiz shuni ko'rsatdiki, kino tilini o'rganish ko'pgi-

na sohalarning vakolatiga kirar ekan. Masalan, madaniyat, san'at, kinoshunoslik, psixologiya, estetika, tilshunoslik va boshqalar. Kino tilining ko'pgina sohalari bilan aloqadorligi uni 2 ma'noda qo'llanilishini taqozo etadi:

**Keng ma'noda** kino tili deganda kinoni tomosha qilayotganda uni tushunish, his etish uchun ishtirok etadigan barcha unsurlar (ovoz, musiqa, joy tasviri, imo-ishoralar, so'zlar, qahramonlarning tashqi ko'rinishi, kiyimi, o'zini tutishi va b.) majmui tushuniladi.

**Tor ma'noda** esa audiovizual aloqani yuzaga chiqarish uchun xizmat qiladigan so'zlar majmui tushuniladi. Tadqiqotda tor ma'nodagi kino tili tahlil etildi.

Kino, asosan, bo'sh vaqti mavjud auditoriyaga mo'ljallangan. Bu auditoriyaning ko'pchiligini uy bekalari, keksalar va bolalar, bir qismini esa bo'sh vaqti mavjud turli toifadagi tomoshabinlar tashkil etadi. Demak, biz istaymizmi yoki yo'qmi, kino butun jamiyatga, uning kelajagiga ta'sir etadi. Uning ta'siri elning ravnaqiga xizmat qilishini ta'minlash maqsadida kino tilini lingvistik aspektda o'rganish zarur.

Kino tilining ta'sir kuchi badiiy asar tilidan bir necha marta ortiq, chunki kinoni katta-yu kichik qiziqish bilan, ekran yonidan jilmay tomosha qiladi. Kinoni tomosha qilish yosh tanlamaydi. Hattoki savodi chiqmagan yosh bolajonlar ham o'z dunyoqarashlaridan kelib chiqib ibrat oladilar. Agar kino tilida yot unsurlar qanchalik ko'p bo'lsa, tarbiya ham shunchalik salbiy tomonga og'adi.

Kino tili (san'at tili)ni tushunish uchun kinoshunoslik fanidan xabardor bo'lish, ma'naviy salohiyat, bilim talab qilinmaydi. Ushbu jihatta e'tiborsizlik bilan qarash tijorat kinolari, shuningdek, tajovuzkor, buzg'unchi kinolarning g'alabasi bilan tugashiga asos bo'ladi.

Xulosa sifatida shuni aytish mumkinki, bugungi kunda kinolingvistikaning o'zbek tilshunosligiga kirib kelishi hamda mazkur yo'nalishda tadqiqotlarning olib borilishi muhim amaliy ahamiyat kasb etadi. Fan-texnikaning tez sur'atlarda rivojlanayotganligini inobatga olib ta'kidlash lozimki, kinolingvistika o'z vaqtida yuzaga kelgan dolzarb tilshunoslik sohasi bo'lib, o'z tadqiqotchilarini kutmoqda.

#### ADABIYOTLAR:

1. Ўзбекистон Республикаси Президенти Ш.М.Мирзиёевнинг 2017 йил 3 август куни мамлакатимиз ижодкор вакиллари билан учрашувидаги маърузаси [Электрон ресурс] [https://youtube.uz/prezident-shavkatmirziyoyev-3-avgust-kuni-ijod-vakillari-bilan-uchrashuv- o%E2%80%98t-kazdi-toliq-video\\_54f38adf2.html](https://youtube.uz/prezident-shavkatmirziyoyev-3-avgust-kuni-ijod-vakillari-bilan-uchrashuv-%E2%80%98t-kazdi-toliq-video_54f38adf2.html)

2. Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг Олий Мажлисга Мурожаатномаси. 24.01.2020. [Электрон ресурс] <https://president.uz/uz/lists/view/3324>
3. Ўзбек тилининг изоҳли луғати: 80000 дан ортиқ сўз ва сўз бирикмаси. Ж. П. Е-М/ Таҳрир хайъати: Т. Мирзаев (раҳбар) ва бошқ.; ЎзР ФА Тил ва адабиёт ин-ти. – Т.: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2006. – Б. 368.
4. Абулқосимова Х. Кино санъати асослари. – Т.: “ЎЗМЭ” Давлат илмий нашриёти, 2009. – Б. 3.
5. Эйзенштейн С. Э. О чистоте киноязыка // Эйзенштейн С.М. Избранные произведения в 6 томах. Том 2. С. 81 б. Зайцева Л.А. Киноязык: опыт мифотворчества. – М.: ВГИК, 2010. – 671 с.
6. Арнхейм Р. Искусство и визуальное восприятие / сокр. пер. с англ. В.Н. Самохина; общ. ред. и вст. ст. В.П. Шестакова. – М.: Прогресс, 1974. – 392 с.
7. DN Yuldasheva. BASIC NON-VERBAL COMPONENTS OF SPEECH. CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (2767-3758) 3 (01), 17–24. <https://masterjournals.com/index.php/crjps/article/view/661>

УДК 81-11

**НЕКОТОРЫЕ СООБРАЖЕНИЯ ОБ УПОТРЕБЛЕНИИ СЛОВ  
В МЕДИЦИНСКИХ РЕКЛАМНЫХ ТЕКСТАХ**

*Д. Р. Собирова, М. А. Раджабкулиева*  
*Бухарский государственный университет*  
dilnozasobirova@gmail.com, marjona@gmail.com

В данной статье рассматриваются аспекты, связанные с современными текстами медицинской рекламы и употреблением в них слов. Мнения, высказанные автором, обоснованы текстами медицинской рекламы, публикуемыми в средствах массовой информации.

**Ключевые слова:** медицинский рекламный текст, адресат, эффект, лечение.

**TIBBIY REKLAMA MATNLARIDA SO‘Z QO‘LLASH BILAN  
BOG‘LIQ AYRIM MULOHAZALAR**

*Sobirova Dilnoza Rasulovna,*  
*Radjabquliyeva Marjona Anvarovna*  
*Buxoro davlat universiteti*  
*Buxoro, O‘zbekiston*  
dilnozasobirova@gmail.com, marjona@gmail.com

Ushbu maqolada zamonaviy tibbiy reklama matnlari va ulardagi so‘zlardan foydalanish bilan bog‘liq jihatlar ko‘rib chiqiladi. Mualliflar tomonidan bildirilgan fikrlar ommaviy axborot vositalarida e‘lon qilingan tibbiy reklama matnlari misolida izohlanadi.

**Kalit so‘zlar:** tibbiy reklama matni, adresat, effekt, davolash.

**SOME CONSIDERATIONS ABOUT THE USE OF WORDS IN  
MEDICAL ADVERTISING TEXTS**

*Sobirova Dilnoza Rasulovna,*  
*Radjabkuliyeva Marjona Anvarovna*  
*Bukhara State University*  
dilnozasobirova@gmail.com, marjona@gmail.com

This article discusses aspects related to modern medical advertising texts and the use of words in them. The opinions expressed by the author are substantiated by medical advertising texts published by mass media.

**Key words:** medical advertising text, addressee, effect, treatment

Dunyo tilshunosligida reklama tili, reklama matnlari tarjimasida lingvomadaniy va uslubiy xususiyatlarni aniqlash, reklama matnlarida til vositalaridan mohirona foydalanish tamoyillari tadqiqotlar uchun asos boʻlib kelgan. Bugun tadbirkorlik muvaffaqiyatining garovi sifatida tan olingan reklamalarni shaxs omili asosida sotsiolingvistik, lingvo-pragmatik, psixolingvistik, lingvokulturologik jihatdan tadqiq etish orqali samarali reklamalarni yaratishga boʻlgan intilish dunyo kopiraytingining bosh masalasidir. Shunday bir davrda reklamalarni, xususan, tibbiy reklamalarni tilshunoslikning antropotsentrik paradigmasi asosida tadqiq etish masalasi bugungi kun uchun ijtimoiy zaruratdir.

Tibbiy reklama matnida vizual, aniq tasvirni yaratadigan, reklama qiymatini oshiruvchi, katta hissiy kuchga ega soʻzlar qoʻllanishi lozim. Bu borada dunyoning eng mashhur reklamashunoslaridan biri Klod Xopkins 1923-yilda yaratgan «Ilmiy reklama» asarida shunday yozgan edi: «Qisqa yoz, odamlar vaqtini tejaysan! [1, 46]» Demak, reklamadagi qisqalik va soddalik reklama auditoriyasini kengaytirishning muhim omilidir. Shu oʻrinda zamonaviy reklama otasi deya nom qozongan reklamashunos olim Devid Ogilvining: «Agar xaridor ikki ogʻiz soʻz, maqtov bilan biror narsani sotib oladi, deb oʻylasangiz, adashasiz. Unga mahsulot haqidagi batafsil maʼlumot kerak [2, 43]», – degan fikrlarini ham unutmazlik lozim. Reklama matni ixcham, sodda, tushunarli boʻlishi bilan birga, mahsulot haqidagi maʼlumotni yetarlicha qamrab olishi ham juda muhim hisoblanadi.

Reklamashunos A.P. Repev oʻzining reklama tiliga bagʻishlangan asarida qisqa vaqt ichida zarur axborot berishning naqadar mashaqqatli va masʼuliyatli vazifa ekanligini quyidagi hayotiy voqea asosida dalil-laydi:

Bir yozuvchidan bir necha soʻz aytishni soʻrashdi.

– Kechirasiz, – javob berdi ijodkor, – lekin oʻn daqiqalik nutq soʻzlash uchun men kamida uch hafta tayyorgarlik koʻrishim kerak.

– Bir soat davomida maʼruza qilish uchun qancha vaqt tayyorgarlik koʻrish kerak?

– Uch kun.

– Agar uch soat soʻzlashga toʻgʻri kelsa-chi?

– Men hoziroq boshlashim mumkin.

Reklama matnlari muallifi «bir daqiqali nutq»ni mohirona yaratish lozim. Koʻp sonli savdo nuqtalari bilan kichik reklama yaratish uzoq muddatli prospektidan koʻra bir necha barobar qiyinroq. Yaxshi reklama beruvchi yaxshi jarrohga oʻxshaydi – ikkalasi ham u yoki bu vositani ishlatishdan oldin uch marta oʻylashadi [3].

Muhtaram Prezidentimiz Sh.M. Mirziyoyev xorazmlik saylovchilar bilan boʻlgan uchrashuvda

«yomon oʻqigan injener traktorni buzadi, yomon oʻqigan shifokor odamni oʻldiradi», deya taʼkidlagan edilar[4]. Notoʻgʻri axborot bergan tibbiy reklama kopirayteri yomon oʻqigan shifokor bilan barobar. U inson umriga zomin boʻlishi, shu bilan reklama qilinayotgan tibbiy mahsulot iqtisodiyotining inqirozga yuz tutishiga ham sababchi boʻlib qolishi hech gap emas. Tadqiqotchi D.Sh. Mamirova muvaffaqiyatli reklama yaratish uchun reklamachi (kopirayter) yaxshi iqtisodchi hamda yaxshi lingvist boʻlishi kerakligini taʼkidlaydi[5, 75]. Fikrimizcha, tibbiy reklama kopirayterlari samarali tibbiy reklama yarata olishi uchun yaxshi iqtisodchi, yaxshi lingvist boʻlishi bilan birga, yaxshi farmatsevt (tibbiy bilimga ega) ham boʻlishlari lozim. Chunki muvaffaqiyatli tibbiy reklama eʼtiborni jalb qila oladigan, Ogilvi taʼbiri bilan aytganda, «sotadigan» boʻlishi bilan birga tibbiy jihatdan toʻgʻri asoslangan boʻlishi gʻoyat muhim hisoblanadi. Tibbiy reklama matnlarida soʻz qoʻllash bilan bogʻliq vazifaga oʻta ehtiyotkor boʻlish lozim:

– *Tortlarimni moʻjiza kutganday kutishadi. Har kuni bor mahoratimni ishga solaman. Oʻng tarafimda ogʻirlik his qildim va tekshiruvdan oʻtdim. Jigarimda muammo bor ekan.*

*Essensiale Forte N uch tomonlama taʼsirga ega boʻlib, jigar hujayralarini qayta tiklaydi. («Essensiale Forte N» dori vositasi reklama matnidan).*

Mazkur reklama matnida soddalikni, barchaga tushunarli boʻlishni taʼminlash maqsadida *Oʻng tarafimda ogʻirlik his qildim* jumlasini qoʻllanadi. Bir tomondan bu jumla toʻgʻri qoʻllangandek tuyuladi. Biroq reklamani kuzatuvchi adresatda oʻng tomon ogʻrisa, «Essensiale Forte N»ni qabul qilish kerak ekan, degan xulosa paydo boʻladi. Vaholanki, oʻng tarafdagi ogʻirlik jigar bilan emas, balki umuman boshqa tana aʼzosi (masalan, oshqozonosti bezi, oʻt pufagi...) bilan bogʻliq boʻlishi mumkin. Natijada reklamadagi axborot tufayli qabul qilingan dori aks taʼsir koʻrsatishi hech gap emas. Toʻgʻri, reklama matnida *Tekshiruvdan oʻtdim* jumlasini orqali shifokor bilan maslahatlashdim mazmuni implitsit shaklda ifodalangan. Biroq bu jumla ikkinchi planda, adresat uchun ahamiyatsiz axborot sifatida qolib ketadi. Chunki shaxs oʻzini qiynayotgan muammoga, shu jumladan, sogʻligʻi masalasiga oʻzi bilmagan holda ong osti orqali yechim izlay boshlaydi. Reklama matni orqali oʻtkazilayotgan psixologik taʼsir adresatni reklamadagi dorini qabul qilishga undaydi. Fikrimizcha, dori-darmon targʻiboti bilan bogʻliq tibbiy reklama matnida maʼlum bir organizm-



dagi og‘riqqa ishora qilishda ehtiyotkor bo‘lish lozim. Shuningdek, tibbiy reklamalarda psixologik ta‘sirni boshqa reklamalardagiga nisbatan kamaytirish lozim (chunki tibbiy reklama faqat foyda ko‘rishni o‘z oldiga maqsad qilib qo‘ygan «CocaCola»ning san‘at darajasidagi o‘ziga xos targ‘iboti emas). Zero, tibbiyotga oid reklamada inson salomatligi masalasi birinchi o‘rinda turadi.

Tibbiy reklamalarni kuzatar ekanmiz, ayniqsa, dori-darmon reklamalarida targ‘ib qilinayotgan preparat og‘riq, bezovtalikni bar-taraf etuvchi sifatida ko‘rsatiladi:

*Boshog‘riq sizni qo‘rqitmasin endi. «Trimol» sizga yordamga shoshadi («Trimol» dori vositasi reklama matnidan).*

Reklama bilan tanishgan adresat ongida boshda og‘riq sezilsa, «Trimol» qabul qilish kerak, degan axborot shakllanadi. Ayrim adresatlar har safar boshog‘rig‘ini sezganda reklamada targ‘ib qilingan dorini muntazam qabul qilaveradilar. Biroq boshog‘rig‘i shunchaki toliqish, stressning oqibati bo‘lmay, jiddiy xastaliklar (masalan, o‘simta)ning darakchisi bo‘lishi mumkin. Reklamaga ishongan adresat yillar davomida boshidagi og‘riqning asl sababini bilmasdan faqat «Trimol» bilan og‘riqni bartaraf qilib yursa, o‘z vaqtida davolanmagan kasallik uning umriga zomin bo‘lishi hech gap emas. Reklamaga ishonib, o‘z-o‘ziga tashxis qo‘yish yaxshi oqibatlarga olib kelmaydi.

#### ADABIYOTLAR:

1. Хопкинс К. Научная реклама // Телескоп: наблюдения за повседневной жизнью петербуржцев. – Москва, 2006. № 4. – С. 45–57.
2. Огилви Д. Тайны рекламного двора. 2009. С. – 43. www.koob.ru
3. Репев А. П. Язык рекламы. Часть II. [http://www.repiev.ru/articles/ad\\_langII.htm](http://www.repiev.ru/articles/ad_langII.htm)
4. <https://t.me/xushnudbek/11431>
5. Мамирова Д.Ш. Ўзбек реклама матнларининг социолингвистик тадқиқи. Филол. фан. бўй. фалс. доктори дисс... – Самарқанд, 2021 йил, 159 б.
6. Sobirova Dilnoza Rasulovna. Qisqalik–samarali tibbiy reklama yaratishning muhim omili. “Soha lingvistikasi: tilshunoslik va tibbiyot integratsiyasining muammolari, yechimlari hamda istiqbollari” mavzusidagi xalqaro ilmiy-amaliy anjumani. Vuxoro, 2021-yil, 11-noyabr.
7. PUBLITSISTIK MATNLAR VA REKLAMALAR SARLAVHALARINING VAZIFALARI XUSUSIDA. S Dilnoza – Тил, таълим, таржима” халқаро журнали, 2020.

УДК 81-11

**ИЗУЧЕНИЕ ПРОБЛЕМЫ СОВРЕМЕННОЙ УЗБЕКСКОЙ  
ЛИТЕРАТУРЫ И МИРОВОГО  
ОБЩЕСТВЕННО-ЛИТЕРАТУРНОГО ДВИЖЕНИЯ***У. Р. Тураева**Бухарский государственный университет  
niso.2019@mail.ru*

Теоретической и практической основой для работы в этом направлении служат исследования литературоведа Бегали Касимова, который начал изучать сущность джадидизма, возникновение джадидской литературы, деятельность ее первых представителей и достиг больших результатов. С этой точки зрения блестящие работы учёного всегда находились в центре внимания нашего литературоведения. В данной статье освещаются исследования, проведенные литературным критиком, лексикографом Бегали Касымовым по изучению современной узбекской литературы и проблеме мирового общественно-литературного активизма.

**Ключевые слова:** модернист, модернист, национальное возрождение, просветитель, мужество, просвещение, самоотверженность, «мусульманство России».

**O‘ZBEK JADID ADABIYOTI VA JAHON IJTIMOYIY-ADABIY  
HARAKATCHILIGI MASALASINING TADQIQ ETILISHI***To‘rayeva Umriniso Rahmatovna**Buxoro davlat universiteti  
niso.2019@mail.ru*

Jadidchilik mohiyatini, jadid adabiyotining yuzaga kelishi, dastlabki namoyandalari faoliyatini o‘rganishga kirishgan va ulkan natijalarga erishgan adabiyotshunos olim Begali Qosimovning tadqiqotlari shu yo‘nalishdagi ishlar uchun nazariy va amaliy asos vazifasini o‘taydi. Shu nuqtai nazardan olimning serqirra faoliyati hamisha adabiyotshunosligimiz e‘tiborida bo‘lib keldi. Ushbu maqolada adabiyotshunos, qomusiy olim Begali Qosimov o‘zbek jadid adabiyoti va jahon ijtimoiy-adabiy harakatchiligi masalasining tadqiq etilishi borasida olib borgan tadqiqotlari qamrab olindi.

**Калит so‘zlar:** jadid, jadidshunos, milliy uyg‘onish, ma‘rifatparvar, jasorat, ma‘rifat, fidoiylk, “Rusiya musulmonligi”

---

**STUDY OF UZBEK MODERN LITERATURE AND WORLD  
SOCIAL-LITERARY MOVEMENT**

*Torayeva Umriniso Rahmatovna*  
*Bukhara State University*  
niso.2019@mail.ru

The researches of Begali Kasimov, a literary scholar, who began to study the essence of Jadidism, the emergence of Jadid literature, the activities of its first representatives and achieved great results, serve as a theoretical and practical basis for work in this direction. From this point of view, the scholar's brilliant work has always been the focus of our literary studies. This article covers the researches conducted by literary critic, lexicographer Begali Kasimov on the study of modern Uzbek literature and the issue of world socio-literary activism.

**Key words:** modernist, modernist, national revival, enlightener, courage, enlightenment, selflessness, "Muslimness of Russia"

Adabiyotshunos olim Begali Qosimov jadidchilik harakatini keng miqyosda, butun qamrovi, yo'lboshchilari, asosiy g'oyalari, geografik ko'lami, mashhur shaxsiyatlari, qrim va qozon tatarlari, ozarbayjon va qozoq adabiyotidagi jadidlargacha tadqiq etishga ham ilmiy-ma'naviy quvvati yetadigan ulkan bir mutaxassis olim edi. O'zbek jadid ijodkorlarini Jamoliddin Afg'oniydan Ismoil Gacpraligacha, Shahobiddin Marjoniydan Mag'jon Jumaboygacha, Norimon Norimonovdan Zaki Validiy To'g'ongacha bo'lgan oraliqdagi adabiy-ijtimoiy qarashlar bilan bemalol qiyoclay olardi. Dunyo olimlari umumjadidchilik yuzasidan mulohaza yuritadigan, ilmiy tadqiqotlar olib boradigan bo'lsa, aslo B.Qosimov nomini, asarlarini chetlab o'ta bilmaydi. Olim kitoblarida, Ayoz Is'hoqiy, Abdurashid Qozi Ibrohimov, bu kabi nomlar va yozgan asarlari tafsiloti bilan uchraydi. "Ismoilbek Gasprali"ga maxsus risola bag'ishladi.

Qolaversa, Begali Qosimovning 600 sahifani tashkil etgan «O'zbek adabiyoti va adabiy aloqalari tarixi» nomli kitobi, bevosita o'zbek adabiy aloqalari tarixiga oid qimmatli ma'lumotlarni qamrabolganligi bilan nihoyatda ahamiyatli.

Olimning tadqiqotlari XIX asr oxiri XX asrning birinchi yarmiga qadar o'zbek milliy uyg'onish davri adabiyotiga oid bo'lsa-da, olim birgina o'zbek adabiyoti muammolari bilangina cheklanmadi. O'zi tadqiq etayotgan adiblarning ijtimoiy-siyosiy, estetik qarashlari mohiyatini chuqurroq anglash va ochib berish uchun o'sha davr boshqa qardosh adabiyot vakillari ijodini ham o'rgandi, qiyosan tahlilga tortdi. B.Qosimov ular ijodidagi o'xshash tendensiyalar,

g'oya va mavzularga ham to'xtalib, qiyosiy paralellar o'tkazdi, ular asarlaridan iqtiboslar keltirib, tadqiqotlarini faktik jihatdan boyitdi. Jadid ma'rifatparvarlari ijodini o'zbek adabiyotshunosligida ilk marotaba akademik ilmiy tadqiqot mavzusiga aylantirdi. Tadqiqotchi olim jadidchilik harakatining otasi sifatida xalqaro miqyosda nom qozongan Ismoilbek Gasprinskiy hayot yo'li, ilmiy, ijodiy faoliyatiga katta qiziqish va hurmat bilan qaradi. Buning isbotini biz olimning Gasprinskiy shaxsi va ma'rifatparvarlik faoliyatini chuqur o'rganib, u haqdagi qimmatli ma'lumotlarni fanga taqdim etganligida ko'ramiz. Tadqiqotchi olim 2004 yilda, vafotidan biroz avval og'ir xasta bo'lishiga qaramay, Gasprinskiyning bir jildlik asarlar to'plamini o'zbek tilida nashr ettirishga ulgurdi.

Olim tadqiqotlaridan Gasprinskiyning Rusiya jadidchilik harakati rahnamosi sifatida amalga oshirgan ishlari, olib borgan siyosiy faoliyati to'g'risida to'la tasavvur hosil qilamiz. Turkiston o'lkasida XIX asrning oxirlaridan boshlab ko'zga tashlana boshlagan yangilanish harakatlari umumiy jihatlari bilan Gasprinskiy nomiga, u tomonidan chop etilgan «Tarjimon» gazetasiga borib tutashadi. Shu davrdan boshlab XX asrning 20-yillariga qadar faoliyat olib borgan biror turkistonli ziyoli yo'qki, bu nomni bilmasa. Shuning uchun Gasprinskiyning ijodiy faoliyati va maslak-intilishlari bilan tanishish Turkiston jadidchiligi va jadid adabiyotining shakllanish hamda rivojlanish bosqichlarinigina emas, umuman mazmun-mohiyatini anglashda, yo'nalishlarini belgilashda ham benihoya muhim ahamiyat kasb etishini olim izlanishlari orqali isbotladi. Olim Gasprinskiy nomini yuz yil oldin turkiy dunyodagi eng mashhur va hurmatga sazovor nom sifatida e'tirof etadiki, uni Qashqardan Londongacha, Sankt-Peterburgdan Bombeygacha o'zini ziyoliman degan barcha bilishi, u tomonidan chiqarilgan «Tarjimon» gazetasining 1908 yildagi 25 yilligi nishonlanganda hatto, Xitoyning «Tarancha»sidan tabrik telegrammasi kelganligi haqidagi ma'lumotni keltiradi. Xuddi shunday 1910 yilda Fransiyaning juda e'tiborli jurnallaridan biri «Revyu dyu monde musulman» Gasprinskiyning millati oldidagi buyuk xizmatlari uchun nomzodini «Tinchlik borasidagi Xalqaro Nobel Mukofoti»ga tavsiya qilganligi va buni xorijiy mamlakatlardagi juda ko'p matbuot organlari qo'llab-quvvatlab chiqqanligini zo'r hurmat bilan e'tirof etadi. Olim tadqiqotlaridan Ismoil Gasprinskiy vafot etgan 1914 yilning 11sentabrida mashhur olim Narimon Narimonovning: «Millat Ismoilbek kabi qahramonlarini unutsa, o'z hayotini barbod etajaqdir», degan ma'lumotni keltiradi. Ismoilbek Gasprinskiy tirikligidayoq millatning otasi, rahnamosi degan

nom oldi. Uning eng katta xizmatlaridan biri: Rusiya sarhadida yashab turgan, biroq siyosiy jarayonlar natijasida bir-biridan uzoqda yashayotgan, natijada begonalashish darajasigacha yetgan turkiy xalqlarni bir-biriga tanitgani, din va til birligiga erishish yo'lida o'z jonini fido qilganligidir.

Olim Gasprinskiy faoliyatini haqiqiy jasorat va o'rnak sifatida baholaydi. Jadid ma'rifatparvari Mahmudxo'ja Behbudiy ta'biri bilan aytganda, «Rusiyadagi butun turk-tatar xalqlarini juda oz tarixiy muddatda yaxlit bir milliy oilaga birlashtirmoqqa noil bo'ldi». Bizga ma'lumki, bu masala o'sha davr chor Rosssiyasi tarkibidagi o'z erku huquqi, asriy urf-odatlar, ma'naviy qadriyatlaridan tortib, tili, dini, e'tiqodigacha unutishga mahkum etilgan turkiy xalqlarning eng katta dardi, so'ngsiz armoni edi. Gasprinskiyning Sharq xalqlari ma'naviy hayoti, xususan, maktab-maorifida chinakam inqilob yasagani, «usuli jadid» nomi bilan mashhur bo'lib tarixga kirgan «usuli savtiya»ni boshlab berganligi fikrimizni dalillaydi. XIX asr oxiri XX asr boshlarida Sharqning eng mashhur, taraqqiyparvar harakatchiligiga – jadidchilikka asos soldi. Gasprinskiy barcha turkiy xalqlarni yagona millat sifatida ko'rishni, ilm-ma'rifatda ham, haq-huquqini bilishda ham dunyoning taraqqiy qilgan millatlari bilan teng ko'rishni orzu qilganligi olim tadqiqotlari orqali ochib berildi. Gasprinskiy bir ma'rifatparvar sifatida butun aqli-vujudi millatning ravnaqi va muhofazasi yo'liga sarf etdi. Buni tadqiqotchi olim shunchaki mutaassibona bir xayol emas, balki zamonasining ilm-ma'rifatli, iqtisod va siyosatdan boxabar, tarixni va taqdirini teran anglagan, Sharq va G'arbni baravar deb bilgan bir ma'rifatparvar kishining aniq va rejali xatti-harakati sifatida baholaydi. Gasparalini ko'rolmaydigan do'stlari, dushmanlari unga siyosiy ayb qo'yib, turkiy xalqlarni Rossiya hukumatiga qarshi birlashtirishda, yagona turk-tatar davlati tuzishga urinishda ayblaganlar. Buning natijasida ma'lum bir muddat xuddi bizning jadid ma'rifatparvar ziyolilarimiz singari uning nomi, xotirasi avlodlar qalbidan yulib olindi. Yoshlarga uni o'z xalqi, millatining dushmani qilib ko'rsatilganligini olim tadqiqotlaridan bilib olamiz. B.Qosimovning «Maslakdoshlar:Behbudiy, Fitrat, Ajziy» nomli kitobida M.Behbudiyning Ismoil Gasprinskiy ijodi, uning faoliyatiga juda qiziqqani va u bilan uchrashish istagida bir necha marta Tataristonga borganini aytadi. Biroq Behbudiy bu baxtga musharraf bo'la olmaydi.

Istiqlol yillarida Toshkentda E.Ablaevning «Ismoil Gasprinskiy – gumanist, prosvetitel, pedagog» risolasi nashr etildi. Shunday tadqiqotlardan yana biri Hoqon Qrimlining «Qrim tatarlarida milliy

o'zlik va milliy harakatlar (1905-1916)» (Anqara, 1996) kitobi bo'lib, 2003-2004 yillarda Turkiyada Gasprinskiy «Tanlangan asarlari» uch jildligining 2 jildi bosmadan chiqqanligi bu ma'rifatparvar siyomoning millat oldidagi xizmatlari e'tirofi sifatida baholanishga loyiq. B. Qosimov izlanishlari natijasida Gasprinskiyning ota-bobolari Yalta-Alupka yo'lida joylashgan Gaspra qishlog'idan ekanligi, Ismoilbek taxallusini shundan olganligini, biroq uni ko'pchilik «Gasprinskiy» nomi bilan tanishi, ilk o'zbek matbuotida uning nomi «Gasprinskiy» deb berilganligini ma'lumotlarida keltiradi. U 1851 yilning 21 martida Gaspradan uncha uzoqda bo'lmagan, to'rt tomoni tog'lar bilan o'ralgan Ajiko'y (ovchi ko'y – ovchilar qishlog'i)da tavallud topganligi o'zbek kitobxonlariga ma'lum bo'ldi. Gasprinskiyning otasi Mustafobey o'z davrida Kavkaz va Qrim general gubernatori knyaz Vorondovga tarjimon bo'lib ishlagan, sadoqatli xizmati uchun 1853 yilda poruchik unvonini olib, dvoryan martabasiga ko'tarilgan. Biroq oradan ko'p vaqt o'tmay, Krim urushi boshlanadi. Natijada ularning oilasi Bokchasaroyga ko'chib keladi. Ismoilbekni sakkiz yoshida boqchasaroylik Hoji Ismoilning eski makgabiga beradilar. O'n yoshga kirganida esa uning otasi onasi Fotimaning e'tirozlariga karamay, Oqmachit ya'ni Simferopoldagi guberniya gimnaziyasiga olib boradi. U yerda ikki yil o'qigandan so'ng, Voronej (Yol Kermon)dagi ofitserlar harbiy o'quv yurtiga o'tadi. O'n uch yoshida esa Moskvadagi Milyutin gimnaziyasiga o'qishini ko'chiradi. Shu yillarda ba'zi rus ziyolilarida avj olgan buyuk rus millatchiligi Milyutin gimnaziyasida ham kirib kelgan edi. O'sha yerda gimnaziya murabbiylaridan biri yozuvchi M. Katkovning qarashlari yosh Ismoilbekda o'zlikni anglash, millatparvarlik hissini uyg'otadi. Ma'lumotlardan uning hatto Kritda yunonlar bilan urush olib borayotgan turk qo'shiniga ko'ngilli bo'lib qo'shilish uchun do'sti litvalik Mustafo Davidovich bilan Istanbulga bormoqchi bo'lgani, ular Odessa bandargohida qo'lga olinib, orqasiga qaytarilganini bilib olamiz. Natijada Ismoilbek Gasprinskiy 1868 yilda Moskvadagi o'qishini tugatib, Boqchasaroyga qaytadi. O'sha joydagi «Zanjirli» madrasasida rus tilidan dars bera boshlaydi. Jarayonda Gasprinskiy o'z ilmini mustahkamlash, yangicha qarashlarga ega bo'lish maqsadida davrning mashhur yozuvchilaridan D.I.Pisarev, N.G.Chernishevskiy asarlari bilan yaqindan tanishadi. O'qish, ilm olishga bo'lgan ishtiyoqi uni 1872 yil Fransiyaga boshlab keladi. U yerdagi Vena, Myunxen, Shtutgardt shaharlarida bo'lib, Parijda mashhur qalamkash I.S.Turgenev bilan uchrashish baxtiga musharraf bo'ladi. Ma'lum bir muddat Turgenevga assistent-kotib bo'lib ishlaydi. Fransuz tilini mukammal

oʻrganib, oʻsha yillarda «Eʻlonlar» idorasida tarjimonlik faoliyatini olib boradi. Shuning barobarida, Gasprinskiy mashhur Sorbonna universitetiga oʻqiydi ham. 1876 yili Boqchasaroyga qaytib, «Zanjirli»da rus tili muallimi sifatida ishini davom ettiradi. Shu yili uylanadi, biroq ular uzoq oila boʻlib yashay olmaydilar, ajrashib ketadilar. Gasprinskiy sidqidildan oʻqishga beriladi. Qunt va sabr bilan turkiy tilni, tarix va adabiyotga oid ilmlarni oʻrganadi. U sharqdagi anʻanaviy taʼlim tizimining zamondan orqada qolayotganini, shuning uchun musulmon xalqlari jahon sivilizatsiyasidan ancha chetga surilib qolayotganini angʻlaydi. Shu jarayonda, u oʻz millati, ilm-fani ravnaqi uchun Milyutin gimnaziyasidagi oʻqitish usulini, Sorbonna universitetida koʻrganlarini musulmon maktablariga koʻchirish, har bir maktab, madrasa uchun, garchi u biroz mashaqqatli boʻlsa ham, dastur ishlab chiqish lozimligi haqida oʻylaydi. Diniy bilimlar bilan birga dunyoviy fanlar ham teng oʻqitilishi kerakligi, ikkinchi tomondan, oʻqishni nazorat qilish shakllari, bilimni baholash mezonlarini ham ishlab chiqish haqida bosh qotiradi. Atrofdagilar Gasprinskiy kofirga chiqaradilar. Tahdidlar, xoʻrlanishlar kuchayib, Ismoilbek yakkalik qiladi va yon berishga majbur boʻladi. Ish yoʻnalishini oʻzgartiradi. 1878-1884 yillarda shahar boshligʻi vazifasiga saylanadi. 1882 yilda Qozon tatarlaridan mashhur millioner Oqchurinning qizi Zuhraxonimga uylanadi. Ular olti farzand tarbiya qiladilar. Olimning taʼkidlashicha, «Yildiz» jurnalining 1990 yildagi maʼlumotiga koʻra:

«Gasprinskiy farzandlaridan Diloraxonim va Ismoil sogʻ-salomatlar. Hozirda Toshkentda yashaydilar», degan maʼlumotni keltiradi.

B.Qosimov Gasprinskiy faoliyati boʻyicha olib borgan izlanishlarida uning 1881 yilda «Rusiya musulmonligi» kitobini eʻlon kilgani, unda koʻtarilgan eng muhim masalalardan biri Rossiyaning oʻz qoʻl ostida yashovchi musulmon fuqarolariga munosabati masalasi ekanligi, bu bevosita qardosh mamlakat farzandining barcha musulmonlar va ularning taraqqiysi kurashganligiga guvoh boʻlamiz. Adabiyotshunos olim jadidchilik harakatining otasi Ismoilbek Gasparali merosini chuqur oʻrganib, bu jarayonda Behbudiy boshlab bergan yoʻlni munosib davom ettirdi. Gasparinskiy oʻzbek kitobxonlariga tanitishda jonbozlik koʻrsatdi.

#### ADABIYOTLAR:

1. Qosimov B. Dolimov U. Ahmedov S va boshqalar. Vatan va millat muqaddasdir. – T.: Oʻqituvchi, 1986. 0,1 b.t.

2. Qosimov B. Jadidchilik, – Milliy uygʻonish va oʻzbek filologiyasi masalalari. – T., 1993 yil, b.t.
3. Qosimov B. Izlay- izlay topganim/ – T.: Adabiyot va sanʼat., 1983, 228 b.
4. Qosimov B. Ismoilbek Gasprali. – T.: Fan, 1992, 0.6 b.t.
5. Qosimov B. Isyonkor qalb. – T.: Oʻzbekiston, 1985, 0.7 b.t.
6. Qosimov B. Milliy uygʻonish: jasorat, maʼrifat, fidoyilik – T.: Maʼnaviyat, 2002, 397 b.
7. Raxmatovna T. U., Zaripovna R. R. Study of Begali Kasimov's Activity in Literary Studies and Scientific Biography of the Scientist //Journal of Survey in Fisheries Sciences. – 2023. – T. 10. – №. 2S. – C. 3454–3459.
8. Toraeva U. R. Definition and research methods of Uzbek modern literature //CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES. – 2021. – T. 2. – №. 10. – C. 104–107.
9. Toraeva U. R. Definition and research methods of Uzbek modern literature //CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES. – 2021. – T. 2. – №. 10. – C. 104–107.

**Internet saytlari:**

10. [www.ziyouz.com](http://www.ziyouz.com)
11. <https://kh-davron.uz/>



УДК 81-11

**ИМЕНА, ОБРАЗОВАННЫЕ НА ОСНОВЕ ИМЕНИ ЧЕЛОВЕКА.  
(БУХАРСКАЯ ОБЛАСТЬ, ВОБКЕНТСКИЙ РАЙОН  
НА ПРИМЕРЕ)**

***М.О. Хайруллаева<sup>1</sup>, Сулайманова Огиллой<sup>2</sup>***

*<sup>1</sup>БухГУ, Бухара, Узбекистан*

*<sup>2</sup>(студентка 3 курса БухГУ)*

Топонимы – народные названия топонимов. Они связаны с многовековой историей и отражают повседневную жизнь, политико-социальные, культурно-духовные, религиозно-философские взгляды наших предков, живших в прошлом. Можно сказать, что топонимы – это историческое богатство узбекского народа и языка, образец национального творчества. Узбекские существительные подвергаются научному изучению со второй половины XX века. На сегодняшний день по узбекской топонимике проведено немало исследований. В области узбекской топонимики многими учеными проведены значительные исследования, заслуживающие внимания. Их авторы – ведущие узбекские номологи Х. Гасанов, Т. Нафасов, З. Досимов, С. Караев. Благодаря усилиям этих ученых и ряда номологов были собраны и углубленно проанализированы материалы топонимов Хорезмской, Кашкадарьинской, Сурхандарьинской, Самаркандской областей. Ниже приведены мысли по поводу некоторых материалов, касающихся Бухарской области.

**Основные понятия:** топоним, топоасос, топофармант, индикатор, название, топоним.

**KISHI NOMI ASOSIDA SHAKLLANGAN TOPONIMLAR  
(BUXORO VILOYAT, VOBKENT TUMAN MISOLIDA)**

***М.О. Хайруллаева<sup>1</sup>, Sulaymonova O'g'iloy<sup>2</sup>***

*<sup>1</sup>Buxoro davlat universiteti*

*Buxoro, O'zbekiston*

*<sup>2</sup>(BuxDU 3-kurs talabasi)*

hayrullayevamarjona1994@gmail.com

Топонимлар joy nomlarining atoqli oti hisoblanadi. Ular uzoq tarixga aloqador bo'lib, o'tmishda yashab o'tgan ajdodlarimizning kundalik turmushi, siyosiy-ijtimoiy, madaniy-ma'naviy, diniy-falsafiy qarashlari kabilarni o'zida aks ettirgan. Aytish mumkinki, joy nomlari o'zbek xalqi va tilining tarixiy boyligi, xalq ijodkorligining namunasidir. O'zbek tilining atoqli otlari yigirmanchi asrning ikkinchi yarmidan boshlab ilmiy asosda o'rganila boshladi. Bugungi kunga qadar o'zbek toponimiyasini tadqiq qilish borasida ko'plab ishlar amalga oshirildi. O'zbek toponimiyasi yo'nalishida ko'plab olimlar e'tiborga loyiq salmoqli tadqiqotlar qildi. Ularning mualliflari etakchi o'zbek nomshunoslari

H.Hasanov, T.Nafasov, Z.Do‘simov, S.Qorayevlardir. Ushbu olimlar va bir qator nomshunoslarning sa‘yi harakati tufayli Xorazm, Qashqadaryo, Surxondaryo, Samarqand viloyati toponimlarining materiallari to‘plandi va chuqur tahlil qilindi. Quyida esa, Buxoro viloyati bilan bog‘liq ayrim materillar haqida fikr yuritilgan.

**Tayanch tushunchalar:** toponim, topoasos, topofarmant, indikator, nom, toponegiz

### NAMES FORMED ON THE BASIS OF A PERSON’S NAME. (IN THE EXAMPLE OF BUKHORA REGION, VOBKENT DISTRICT)

*M. O. Khairullayeva<sup>1</sup>, Sulaymanova Oghiloy<sup>2</sup>*

<sup>1</sup>*BukhSU, Bukhara, Uzbekistan*

<sup>2</sup>*(3rd year student of BukhSU)*

hayrullayevamarjona1994@gmail.com

Toponyms are popular names of place names. They are related to a long history and reflect the daily life, political-social, cultural-spiritual, religious-philosophical views of our ancestors who lived in the past. It can be said that place names are the historical wealth of the Uzbek people and language, an example of national creativity. Uzbek nouns have been scientifically studied since the second half of the 20th century. To date, a lot of research has been done on Uzbek toponymy. In the field of Uzbek toponymy, many scientists have made significant studies worthy of attention. Their authors are leading Uzbek nomologists H. Hasanov, T. Nafasov, Z. Dosimov, S. Karayev. Thanks to the efforts of these scientists and a number of nomologists, the materials of toponyms of Khorezm, Kashkadarya, Surkhandarya, Samarkand region were collected and analyzed in depth. Below are thoughts on some materials related to Bukhara region.

**Basic concepts:** toponym, topoasos, topofarmant, indicator, name, toponym

Ma‘lumki, toponimiya muayyan hududlarda tarqalgan va xalq tilida yashovchi lisoniy material bo‘lib, mohiyatan xalq lahja va shevalarining lug‘aviy bo‘yligidir. Afsuski, shunga qaramasdan, shu vaqtga qadar shevashunoslar o‘z ishlarida onomastik materialni chetlab o‘tishdi.

Qayd qilingan mulohazalardan ma‘lum bo‘ldiki, Buxoro viloyati hududitoponimlaridagi hozirga qadar qoniqarli ilmiy tadqiq qilgan emas. Xususan, viloyatning Vobkent tumani mikrotoponimiyasi bo‘yicha ham alohida tadqiqot olib borilmagan.

Joylarni kishilarning ismi laqabi va familiyasi bilan, ba‘zi hollarda taxallusi bilan atash juda qadimiy tarixga egadir. Bunday rasm-rusunning qachon paydo bo‘lganligi haqida nomshunoslar turlicha fikrlar bildirishadi. Taniqli nomshunos V.A.Nikonov “Vvedenie v toponimiku” kitobida bu jarayon feodalizm davrida, yerga egalikning

kuchayganligi bilan izohlaydi. Ammo ba'zi tadqiqotchilar antropotoponimlarning toponimlarga o'tishini qadimiy hodisa deb biladi:

«Antroponimika va toponimika sohalari tarixan bir-biri bilan bog'liqdir. Chunki bir qator antroponimlar toponimlar asosida paydo bo'lgan, ko'pgina toponimlar esa kishilarning ismlaridan yasalgandir»<sup>12</sup>, deb qayd etgan.

Kishi nomi asosida shakllangan toponim – antropotoponim to'rt xil sababga ko'ra paydo bo'lishi mumkin:

1) obyekt uni bunyod etgan, uning yuzaga kelishi uchun sababchi bo'lgan shaxs nomi bilan ataladi;

2) obyekt unga egalik qiluvchi kishi nomi bilan ataladi;

3) biror shaxs ideallashtiriladi yoki ilohiylashtiriladi hamda ma'lum obyektlar uning nomi bilan atala boshlanadi;

4) shaxs nomi rasmiy ko'rsatma va qaror asosida biror joyga beriladi (memoreal toponimlar). Ayniqsa, keyingi holat sho'rolar davrida, nihoyatda, avj olgan va adolatning toponimik siyosatining tarkibiy qisimga aylangan edi. Antropotoponimlar, ya'ni odamlarning ismlari tarixiy geografiyada keng tarqalgan bo'lib, Sharq va G'arb mamlakatlarning qadimgi va o'rta asr toponimiyasida turli geografik nomlar tarkibida ko'plab uchraydi. Jahonning turli qadimiy madaniyat markazlarini bog'lovchi karvon yo'llarining tutashgan erida joylashgan Buxoro ham bundan istisno emas. Shuning uchun uning tarixiy toponimiyasida turli madaniyatlar va tarixiy davrlarga taaluqli bo'lgan antropotoponimlar o'z aksini topgan.

Tuman toponimiyasida keyingi davrda tarixiy shaxs yoki adabiyot, fan namoyandalarining nomlarini abadiylashtirish maqsadida *Nodirabegim, Abdulla Qodiriy, Sadridin Ayniy, Muqimiy, Sattor Yarashev, Fayzulla Xo'jayev, Sulton Jo'ra, Mirzo Bobur, Hamid Olimjon, Alisher Navoiy, Ibn Sino, Mirzo Ulug'bek, G'afur G'ulom, Furqat, Ahmad Donish, Muhammad Narshaxiy, Hoshimxo'ja, Hamro buvi Rahimova, Beruniy* kabi toponimlar ham paydo bo'lgan.

Vobkent tumani toponimiyasida quyidagi antroponimik toponegizlar mavjud:

1. Shaxsning ismi: *Imomqozixon, Boltabek, Mirsulaymon, Mo'minobod, Niyozxo'ja, Qo'ldosh, Qoryog'di, Shofayoz, Shamsiobod* kabi tarixiy nomlar. Ba'zan kishi ismi obod komponenti bilan birikib toponim hosil qiladi: *Mo'minobod* tipli toponimlar zamirida ham aslida antroponim yotadi.

<sup>1</sup> Никонов В. А. Введение в топонимику. – М., Наука, 1965. – С. 4.

2. Kishining laqabi yoki ismi va laqabidan yasalgan toponimlar: *Latifsobungar* kabi.

3. Kishining ismi va familiyasidan yasalgan toponimlar. Bunday nomlar asosan shoʻrolar davrida keng udum boʻldi va koʻproq koʻchalar, guzarlar, maydonlar, xiyobonlar kishining ismi va familiyasi bilan ataladi: *Sattor Yarashev, Fayzulla Xoʻjayev, Hamro buvi Rahimova* kabi.

4. Kishining taxallusi asosida shakllangan toponimlarga: Tuman toponimiyasida keyingi davrda tarixiy shaxs yoki adabiyot, fan namoyandalarining nomlarini abadiylashtirish maqsadida *Nodirabegim, Abdulla Qodiriy, Sadridin Ayniy, Muqimiy, Sulton Joʻra, Mirzo Bobur, Hamid Olimjon, Alisher Navoiy, Ibn Sino, Mirzo Ulugʻbek, Gʻafur Gʻulom, Furqat, Ahmad Donish, Muhammad Narshaxiy, Hoshimxoʻja, Beruniy* kabilarni kiritish mumkin.

Xulosa qiladigan boʻlsak, bu oʻrganishimiz natijasida, Vobkent tumani toponimiyasi boʻyicha maʼlum ilmiy xulosalar chiqarildi. Mazkur xulosalar Buxoro viloyat toponimlariga oid hozirga qadar mavjud boʻlgan ilmiy tasavvurlarni maʼlum darajada toʻldiradi. Qolaversa, ishdagi tadqiq va undan kelib chiqadigan xulosalar oʻzbek nomshunosligi, dialektologiyasi hamda til tarixi, leksikologiyasi va leksikografiyasi uchun ham manba vazifasini oʻtaydi deb oʻylaymiz.

#### FOYDALANILGAN ADABIYOTLAR

1. Киличев Б. Бир топоформат ҳақида // Филология масалалари. Т.: 2007.

2. Каримова Л. Ўзбек тилида топонимларнинг ўрганилиши. – Т., Фан, 1982, – 96 б.

3. Qorayev S. Toponimika. T.: “Oʻzbekiston faylasuflari milliy jamiyati” nashriyoti. 2006, 46-b.

4. Kili chev B.E., Safarova M.Z. Bukhara region’s typical toponyms transformed by means the names of nations. International Scientific Journal Theoretical & Applied Science, <http://t-science.org/> p-ISSN: 2308-4944 (print) e-ISSN: 2409-0085 (online) 2021.02.05.

5. Kili chev B.E., Safarova M.Z. The Concept of Totemism and Ethnonym. Central Asian Journal of Literature, Philosophy and..., 2022.

УДК 81-11

## ЛИНГВИСТИЧЕСКАЯ СЕМАНТИКА МОДАЛЬНОСТИ

*Г. Ф. Хайруллоева**БухГУ, Бухара, Узбекистан**gulasalxayrulloeva@gmail.com*Corporativ e-mail: *g.f.xayrulloeva@buxdu.uz*ORCID ID: <https://orcid.org/0000-0001-9891-967X>

В данной статье предпринята попытка уточнить лингвокультурную семантику модальности в узбекском языке. Модальность – это языковое явление и его высший уровень как составная часть текста, а также форма реализации культурного существования. Итак, под модальностью обычно понимают грамматико-семантическую категорию, выражающую отношение говорящего к выражаемой мысли или отношение говорящего к содержанию предложения. Говорящий воспринимает объективную реальность как оценку, отражает ее в своем сознании и описывает свое отношение к действительности, используя различные смысловые категории.

**Ключевые слова:** антропоцентрический принцип, модальность, семантика, грамматико-семантическая категория, лингвокультурная семантика, контрастивное исследование, коммуникативная цель, модальность, адресуемость, информативность, ситуативность, интертекстуальность.

## LINGUISTIC SEMANTICS OF MODALITY

*Khairulloeva Gulasal Faizulla kizi**BuxSU, Bukhara, Uzbekistan**gulasalxayrulloeva@gmail.com*Corporativ e-mail: *g.f.xayrulloeva@buxdu.uz*ORCID ID: <https://orcid.org/0000-0001-9891-967X>

This article tries to clarify the linguistic and cultural semantics of modality in the Uzbek language. Modality is a linguistic phenomenon and its highest level as a component of the text, while it is also a form of implementation of cultural existence. So, modality usually means a grammatical-semantic category that expresses the attitude of the speaker to the expressed thought or the attitude of the speaker to the content of the sentence. The speaker perceives the objective reality as an assessment, reflects it in his mind and describes his attitude to reality using various semantic categories.

**Key words:** anthropocentric principle, modality, semantics, grammatical-semantic category, linguocultural semantics, contrastive research, communicative purpose, modality, addressability, informativeness, situationality, intertextuality.

## MODALLIKNING LINGVISTIK SEMANTIKASI

*Xayrulloeva Gulasal Fayzullaning qizi.**BuxDU, Buxoro, O'zbekiston*

gulasalxayrulloeva@gmail.com

Corporativ e-mail: g.f.xayrulloeva@buxdu.uz

ORCID ID: <https://orcid.org/0000-0001-9891-967X>

Ushbu maqolada o'zbek tilidagi modallikning lingvomadaniy semantikasini oydinlashtirishga harakat qilingan. Modallik lisoniy hodisa va uning matnning ajralmas qismi sifatidagi eng yuqori darajasi, shuningdek, madaniy borliqni amalga oshirish shaklidir. Demak, modallik deganda, odatda, so'zlovchining aytilayotgan fikrga munosabatini yoki gap mazmuniga munosabatini ifodalovchi grammatik-semantik kategoriya tushuniladi. So'zlovchi ob'ektiv voqelikni baho sifatida qabul qiladi, uni o'z ongida aks ettiradi va turli semantik kategoriyalar yordamida voqelikka munosabatini tasvirlaydi.

**Tayanch so'zlar:** antropotsentrik tamoyil, modallik, semantika, grammatik-semantik kategoriya, lingvomadaniy semantika, kontrast tadqiqot, kommunikativ maqsad, modallik, manzillilik, informativlik, situatsionlik, intertekstuallik.

In modern linguistics, the anthropocentric principle has become the leading paradigm. At this stage, much attention is paid to the study of the human factor in the language. As a result of this, the interest in researching the conceptual features of reality based on structural-semantic and cognitive approaches, in studying the possibilities of expression of national-ethnic features in the linguistic landscape of the world is increasing. The development of contrastive studies through the analysis of paradigmatic, syntagmatic and variant relations of the communication process, the activation of interest in language and culture problems led to the problem of expanding cognitive, ethnolinguistic research opportunities, studying national-cultural concepts. That is why, on the basis of the principles of the anthropocentric paradigm, it is studied in the interaction of the language with culture, ethnos, national mentality, which was created in the cooperation of the fields of cultural studies, ethnography, psycholinguistics.

Textologists such as V. Dressler, M. Holliday, R. Hasan, I. R. Galperin, M. Kojina, T. M. Dridze, P. Hartman were interested in the problem of categorical symbols necessary for speech structures to acquire the status of text. For example, relying on their theoretical ideas, we can say that the status of the text depends on its following categorical features: communicative purpose, modality, addressability, informativeness, situationality, intertextuality.

The research of the specific verbal means of the category of modality in the Uzbek language based on the conceptual methodology allows to scientifically justify the manifestation of the linguocultural and pragmatic features of this conceptual semantics.

Modality is a set of conditional language signs that serve to express modal meanings, meaning various relationships between objects and sentences. Modality is a phenomenon in the communicative content, which is related to the nature of the text's occurrence in the direct communication process. Indeed, when a speaker or writer thinks about reality, he expresses his attitude towards it. Therefore, in addition to the judgment about whether reality is real or imagined, judgments such as the truth or falsity of an opinion, its connection with desire and opportunity are also important. It is these kinds of judgments and considerations that ensure the meaningful course of speech activity and allow the text to serve as a "transmitting" tool.

So, the content and essence of the category of modality is formed in the process of speech- thinking activity and occurs in a certain text, in a communication situation. A complex characteristic modality is characteristic of an artistic text. After all, the author of a work of art expresses his personal attitude towards the described events. This relationship, in turn, is not a simple statement, the author uses language tools that have a quick and strong impact on the reader. The choice of tools is not simple, because the author should take into account that the reader compares the information presented in the text with reality and gives his own assessment. The pragmatic effect of the text will be as intended only if this ability of the receiver of information is taken into account.

In linguistics, when analyzing modal units from a linguistic and cultural point of view, two points of view are mainly put forward:

1. Modality category represents a certain linguistic and cultural concept. This concept often reflects the national aspects of the conceptual sphere of a particular nation.

2. The cultural semantics of the modality reflects the national thinking and mentality of the people. Because the category of modality is called the category of modality, a number of functional-semantic modality categories are formed and developed, and in the process of forming a wide functional-semantic space, they are combined with various subtleties of meaning, meaning the relationship of the speaker to the objective existence, and the semantic meanings expressed in the implementation of the communicative dialogue between the speaker and the listener set.

To understand the category of modality, it is necessary to distinguish the concepts of proposition, dictum and mode. A proposition

is an invariant meaning that reflects an objective state. In a particular sentence, the proposition is interpreted by the speaker in one way or another. As a result, the expression has a double content: the objective content and its interpretation. To define these two mandatory aspects of the semantic structure of speech, the French linguist Charles Balli introduced the terms *dictum* and *modus*. [1] M. Hakimov describes the pragmatic aspect of the mode as follows: “The substantive analysis related to the concept of the mode is characterized by its internal semantic structure, regardless of the formal aspect of the expression – the open and hidden statement of the participants of the speech about “what they want to say” [2]

When distinguishing communicative types, he distinguishes according to the modality function of the part of the sentence and puts the main emphasis on the part, that is, “signs distinguishing the sentence from other types are determined by the part.” [4] A. Gulomov explains the differentiation of communicative types according to the aspect of modality on the basis of examples. For example, the participle fully expressed with the tense and person-number in the form of “go” means the reality of the event, the participles in the form of “go” and “should” indicate suspicion, “go” means “goal”, “go” means “desire” is expressed. [5]

We can observe the reflection of the specific aspects of the Uzbek culture, people’s thinking, perception of the world in the language in the example of interrogative sentences:

1. Declarative (lat. “*declarativus*” – “explain, make a statement”) – interrogative sentences used in reality to change the status of a person: Will you accept the daughter of Kumushbibi Mirzakarim, the son of Otabek Yusufbek Haji, as your wife?

2. Directive (lat. “*directivus*” – “instruction, management”). Interrogative sentences with command, question, request, warning, anger and sarcasm content: Command: Won’t you walk faster? Can you complete the assignments on time? Q: How many days will you submit the article? Please: Can you attend the event on my behalf? Warning: Will you be late next time? Anger: Why didn’t you say that earlier? Sarcasm: Did they miss us? Nechuk?

3. Prohibitive (lat. “*prohibitio*” – “prohibition”). Interrogative sentences with prohibition content: This is my life. Let’s not return to this topic? (forbids returning to the topic about himself).

4. Espressive (lat. “*espressivus*” – “lively, expressive”) – interrogative sentences with the meaning of condolence: Are you sitting without feeling sad?



5. Predictive (eng. “predict” – “to say in advance, to be expected”) – interrogative sentences with the meaning of estimation: Guess: Will I come?

6. Commissive (lat. “comissio”-”promise”) – interrogative sentences in the meaning of promise, offer: Will you come to my wedding? Do you keep your word?

7. Acknowledgment (“acknowledgment” – “confirm”). Interrogative sentences with the meaning of takalluf, respect: Won’t you try? May I pour you some tea with your permission?

In conclusion, it can be said that modality as a component of the text is a linguistic phenomenon and its highest level, at the same time, it is also a form of implementation of cultural existence. So, modality usually means a grammatical-semantic category that expresses the attitude of the speaker to the expressed thought or the attitude of the speaker to the content of the sentence. The speaker perceives the objective reality as an assessment, reflects it in his mind and describes his attitude to reality using various semantic categories.

#### REFERENCES:

1. Цыбульская М.П. Диктум и модус как текстообразующие категории журналистского текста // Журналистика-2019: стан, праблемы і перспектывы 21-я Міжнародная навукова- практычнай канферэнцыі. Мінск, 2019. // <https://mediana.by/gubriki/soobshchestvo/2115>; Ярыгина Е.С. Модус и модальность – терминологические синонимы? // Вестник ВятГУ 2012, том 2, № 2.

2. Ҳақимов М. Ўзбек прагмалингвистикаси асослари. Т.: Академ-нашр, 2013

3. Махмудов Н., Нурмонов А. Ўзбек тилининг назарий грамматикаси. – Тошкент, 1995. – Б. 47.

4. Фуломов А., Аскарова М. Ҳозирги ўзбек адабий тили. – Т.: Ўқитувчи, 1987. – Б. 55.

5. qizi Xayrulloeva G. F. SO ‘ROQ BELGISINING PRAGMATIK ANAMIYATI //Educational Research in Universal Sciences. – 2023. – Т. 2. – №. 4. – С. 194–199.

6. Xayrulloeva G. Individuality in the Interchangeable Use of Punctuation //ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu. uz). – 2023. – Т. 35. – №. 35.

7. Xayrulloeva G. OZBEK TILIDA TINISH BELGILARINING KETMA-KET QOLLANILISHI VA ULARNING OZARO SINONIMIK MUNOSABATLARI //Центральноазиатский журнал образования и инноваций. – 2023. – Т. 2. – №. 6 Part 2. – С. 163–166.5.

УДК 81-11

**ВЗАИМОСВЯЗЬ ЛИЧНО-ОПИСАТЕЛЬНОЙ ЛЕКСЕМЫ  
И СЛОВСОЧЕТАНИЯ***М. Ю. Ходжиева**Бухарский государственный университет  
hojiyevamaxfirat8@gmail.com*

В данной статье поясняется, что основное отличие лексемы от словосочетания заключается не в задаче называть/называть, а в том, каким образом, каким образом и в какой форме она выражает эту задачу.

**Ключевые слова:** лексема, словосочетание, человек, язык, лингвосинтаксический образец, номинатив.

**SHAXS TAVSIFLI LEKSEMA VA SO‘Z BIRIKMASI MUNOSABATI***Xojiyeva Maxfirat Yusupovna**Filologiya fanlari bo‘yicha falsafa doktori (PhD)  
Buxoro davlat universiteti  
hojiyevamaxfirat8@gmail.com.*

Ushbu maqolada leksema va so‘z birikmasi orasidagi asosiy farq atash/nomlash vazifasida emasligi, balki shu vazifani qaysi yo‘l bilan, qaysi usul va qaysi shakl bilan ifoda etishi izohlangan.

**Kalit so‘zlar:** leksema, so‘z birikmasi, shaxs, lison, lisoniy sintaktik qolip, nominativ birlik.

**RELATIONSHIP OF PERSONAL DESCRIPTIVE LEXEMA  
AND WORD COMBINATION***Khodzhiyeva Makhfirat Yusupovna**Bukhara State University  
hojiyevamaxfirat8@gmail.com*

In this article, it is explained that the main difference between a lexeme and a phrase is not in the task of calling/naming, but in what way, in what way and in what form it expresses this task.

**Key words:** lexeme, phrase, person, language, linguistic syntactic pattern, nominative unit.

So‘z birikmalari mohiyatan, atash, nomlash (nominativ) sathi birliklari hisoblanadi. Boshqacha qilib aytganda, so‘z birikmasi lisoniy qiymat jihatidan so‘z (leksema)ga tengdir. Buni isbotlash hech qanday qiyinchilikni tug‘dirmaydi. H.Nematov buni “o‘zbek tilidagi *ayg‘ir*

soʻzi bilan notanish boʻlgan kishi *aygʻir* leksemasi oʻrnida bemaolol *er-kak ot* soʻz birikmasini ishlataveradi”, deb osongina isbotlaydi [2,19]. Chunki bu leksema va soʻz birikmasi atash qiymatiga koʻra teng qiymatlidir. Shuningdek, *togʻa* leksemasi oʻrnida *onaning inisi*, *onaning akasi* yoki *onaning ukasi* soʻz birikmalarini qoʻllash imkoniyati bor. Ayni paytda, leksemalar oʻrnida qoʻllash bilan bir qatorda leksema va soʻz birikmalarini yonma-yon, parallel ishlatish imkoniyati ham borki, bunda *muallima* va *ayol oʻqituvchi*, *oʻqituvchi ayol* singari shaxsni tavsiflovchi soʻz birikmalari nafaqat bir-birini almashtirib, balki nutqda yonma-yon, parallel holda voqelana oladi. Shunday qilib, tilimizda soʻz birikmalarining leksemalarga atash vazifasi jihatidan teng qiymatlilikiga yuzlab misollarni dalil sifatida keltirish mumkin. Leksema va soʻz birikmasi orasidagi asosiy farq atash/nomlash vazifasida emas, balki shu vazifani qaysi yoʻl bilan, qaysi usul va qaysi shakl bilan ifoda etishdadir. Til sistemasida bu vazifa leksik sathda tub, yasama, sodda, qoʻshma soʻzlar yordamida bevosita bajarilsa, soʻz birikmalari ham atash, nomlash vazifasiga ega boʻlib, ular bu vazifani bevosita amalga oshiradi. Atash, nomlash vazifasiga ega boʻlganligi uchun ham soʻz birikmasi tilning nominativ birliklar tizimiga mansub birlik hisoblanadi. Bu xulosada soʻz birikmalarini nimaga xizmat qilishi, qisqasi, vazifasi aniqlashadi.

Demak, soʻz birikmasining erkin sintaktik aloqa bilan bogʻliqligi haqidagi bahsni davom ettirish lozim. Avvalo, erkin sintaktik aloqa tushunchasiga aniqlik kiritaylik. Erkin aloqa deganda soʻzlarning muvaqqat, faqat nutq ehtiyoji va zaruriy atash vazifasini bajarish uchun tobe-hokimlik munosabatlariga kirishi tushuniladi [2,19]. Chunonchi, shaxsning u yoki bu xususiyatini tavsiflash ehtiyoji tugʻilganda hosil boʻladigan erkin sintaktik aloqa natijasida shaxsni tavsiflovchi soʻz birikmalari yuzaga keladi. Oʻzbek tilida tobe va hokim aʼzolari shaxs otlaridan tuzilgan soʻz birikmalariga eʼtibor qilaylik:

1. *sotuvchi yigit* [shaxsni kasb jihatidan tavsiflovchi ot + shaxsni yosh jihatidan tavsiflovchi ot];

2. *mahalladosh ayol* [shaxsni yaqinlik munosabati jihatidan tavsiflovchi ot + shaxsni jins jihatidan tavsiflovchi ot].

Qiyoslanayotgan har ikkala soʻz birikmasi ham shaxsni atash, nomlash vazifasiga koʻra umumiy oʻxshash boʻlib, soʻz birikmalarining shaxsni tavsiflovchi maʼnaviy tipini shakllantiradi. Biroq birikmalar tarkibidagi kengayuvchi soʻzlar bir-biridan lugʻaviy-manoviy xususiyatlari asosida farqlanadi. Negaki, *sotuvchi yigit* birikmasida kengayuvchi *yigit* leksemasi hokim boʻlak mavqeyida shaxsni yosh jihatidan

tavsiflovchi ot reallashganligidan dalolat beradi. Keyingi birikmadagi *ayol* leksemasi kengayuvchi bo‘lib, hokim bo‘lak mavqeyini shaxsni jins jihatidan tavsiflovchi ot egallaganligini ko‘rsatadi. Demak, mazkur so‘z birikmalarida kengaytiruvchi so‘zlar (*sotuvchi, mahalladosh*) tobe bo‘lak mavqeyida shaxsni kasb jihatidan (*sotuvchi*) va yaqinlik munosabati jihatidan (*mahalladosh*) tavsiflovchi otlar voqelashganligini ko‘rsatish bilan birga, birikish imkoniyatlarining ikki ko‘rinishi reallashganligini bildiradi. Ayni ko‘rinishlar birikuvning o‘ziga xos turi mavjudligidan darak beradi.

Anglashiladiki, so‘z birikmalari nutq jarayonida so‘zlarning ma‘lum sintaktik aloqalari natijasida yuzaga chiqadi. So‘z birikmalaridagi sintaktik aloqa nutq jarayonida so‘zlarning o‘zaro tobe-hokimlik munosabatiga kirishib, atov birliklari hosil qilish yo‘llari ekan, lisoniy-sintaktik qurilish qolip (LSQ)larida xuddi mana shu jihat o‘z aksini topadi. Shu boisdan o‘zbek tili uchun so‘z birikmalari cheksiz turlarining umumiy qolipini, ya’ni til sathida so‘zlarning tobe-hokimlik munosabatini ifodalovchi umumiy qurilishini [T~H]=SB qolipi asosida berish mumkin. Shuni alohida ta’kidlash kerakki, [T~H]=SB qolipi o‘zbek nutqida qo‘llanila oladigan milliard-milliard muayyan so‘z birikmalarining oliy, eng yuksak darajasidagi umumiy qolipidir. Ushbu qolipni o‘zbek tilida so‘z birikmalarining umumiy lisoniy qolipi deb atash mumkin. Negaki, [T~H]=SB qolipi o‘z atrofida so‘z birikmalari hosil qilishning muayyanroq qoliplarini birlashtiruvchi giponimlar qurshoviga ega. O‘z navbatida, bu giponimik qurshovdagi har bir so‘z birikmasi LSQ i xususiy qoliplardan iborat giponimik qurshovni hosil qilishi mumkin. Shunday ekan, ko‘p bosqichli [gipo- giperonimik], ya’ni [tur-jins] munosabatidagi qoliplar [T~H]=SB markazidan boshlanib, muayyan so‘z birikmasi – nutqiy hosilagacha davom ettirilishi ham qonuniydir.

#### ADABIYOTLAR:

1. Nazarova S. A., Khojiev M. Y. Personal description in the word combination and expression in a speech //International Journal on Integrated Education. – T. 3. – №. 2. – C. 15–18.
2. Неъматов Х., Расулов Р. Ўзбек тили систем лексикологияси асо-слари.– Т.: Ўқитувчи, 1995. – 128 б.
3. Xojiyeva M. Y. Shaxsni tavsiflovchi so‘z birikmalari va assesment metodi //Oriental renaissance: Innovative, educational, natural and social sciences. – 2021. – T. 1. – №. 7. – C. 302–307.
4. Yusupovna K.M., Yusufovna Y.S. Language and speech expression of personal descriptions//Thematics Journal of Education. – 2021.

---

УДК 81-11

**МОЛЧАНИЕ – ЕДИНСТВО, РАСКРЫВАЮЩЕЕ ВНУТРЕННЕЕ  
СОСТОЯНИЕ ЧЕЛОВЕКА**

*Д. Н. Юлдашева*

*Бухарский государственный университет,*

*Бухара, Узбекистан*

*yuldasheva21dilorom@gmail.com*

**ORCID ID ORSID 0000-0003-1976-5575**

Учитывая общность национального языка и национального мышления, рассмотрение когнитивных, экспрессивных, аккумулятивных сторон каждого языкового явления считается одной из основных задач этого направления. Ведь важно анализировать каждый язык в совокупности таких факторов, как социальная среда, национальный менталитет, национальный этикет, а также речь и культура носителя этого языка.

**SILENCE IS A UNITY THAT REVEALS THE INNER STATES  
OF A PERSON**

*Dilorom Nigmatovna Yuldasheva*

*Bukhara State University,*

*Bukhara, Uzbekistan*

*yuldasheva21dilorom@gmail.com*

**ORCID ID ORSID 0000-0003-1976-5575**

Based on the commonality of the national language and national thinking, consideration of the cognitive, expressive, accumulative aspects of each linguistic phenomenon is considered one of the main tasks of this direction. After all, it is important to analyze each language in the aggregate of such factors as the social environment, national mentality, national etiquette, as well as the speech and culture of the speaker of this language.

This article analyzes the issue of approaching the phenomenon of silence, which is a component of discourse, as a unit of action of internal and external speech in the process of verbal communication.

**JIMLIK – INSONNING ICHKI HOVLATINI FOYDA QILGAN  
BIRLIK**

*Dilorom Nigmatovna Yuldasheva*

*Buxoro davlat universiteti,*

*Buxoro, O'zbekiston*

*yuldasheva21dilorom@gmail.com*

**ORCID ID ORSID 0000-0003-1976-5575**

Milliy til va milliy tafakkurning mushtarakligini hisobga olib, har bir til hodisasining kognitiv, ekspressiv, kumulyativ tomonlarini hisobga olish ushbu yoʻnalishning asosiy vazifalaridan biri sanaladi. Zero, har bir tilni ijtimoiy muhit, milliy mentalitet, milliy odob, shu tilda soʻzlashuvchining nutqi va madaniyati kabi omillar bilan uygʻunlikda tahlil qilish muhim ahamiyatga ega.

## Introduction

By the end of the 20th century, linguistic paradigms were identified in Uzbek substantive linguistics – a number of phenomena similar to each other, informational meaning, position in the language system, and phenomena that have the same function, as equal to each other from a purely linguistic point of view, are combined into paradigm or series. On this basis, he managed to interpret and classify the phenomena of the language system on the basis of the thought-language-speech trichotomy.

Since there are multiple units of the same event in the language system, this does not happen automatically. Naturally, these units differ from each other in some respects. Usually this difference is not always connected with the linguistic structure, linguistic meaning, linguistic meaning: for some reason, “in a certain case” in order to convey certain information to the listener, in many cases it is “that” that is used. non-linguistic factors – the speech situation, speech conditions, the subject of speech, the subject of speech, the purpose of the communication process, the contact and distant situation of the speaker and listener, the mental state of the speaker or listener, interaction, as you know, is associated with a number of aspects, such as the national cultural level, social status, age, gender, religious beliefs. Substantial linguistics evaluates all of these as non-linguistic factors and adds them to the list of “non-essential” issues of theoretical linguistics.

Not enough attention was paid to such speech units as tone, silence, which are evaluated as extralinguistic, paralinguistic, nonlinguistic, nonlinguistic phenomena [7]. Naturally, these speech means are more effective than linguistic means (phoneme, morpheme, phrase, model) in terms of efficiency in the communication process, i.e. in practice, speech units (sounds of speech, morphemes, phrases, etc.) in combination or sentence) acquires great importance and value.

## Literature review

The anthropocentric direction began to pay attention to little-studied speech means, which until recent years did not attract the attention

of Uzbek linguists. After all, the need to pay attention to speech, which is a manifestation of language, to choose the process of speech activity as an object of study, was recognized by members of the Prague Linguistic School as early as 1929 [2]. In fact, any language can never live in a vacuum; as a rule, a language retains its essence only when it becomes a tool for communication and interaction between people in a certain social situation. So, the true essence of the language is fully manifested in its speech pragmatics.

We know that there is an internal and external side of linguistic information, the inner side is related to the structure of the language, and the outer side is related to the speech situation. In fact, the external side of linguistic information requires non-verbal means, as well as verbal units in discourse. After all, without them, the discourse would not be complete.

Practical linguistics, pragmalinguistics, is recognized as a new middle direction in Uzbek linguistics, the purpose of which is to analyze the systemic possibilities of a language construction that open up as a result of the study of linguistic elements, methods of rational-structural research. In particular, the foundations of Uzbek pragmalinguistics were created (M. Khakimov). A number of studies were carried out with the help of paralinguistic tools, in particular, A. Nurmonov, M. Saidkhanov, S. Mominov, Sh. Iskandarova, A. Aripova, Z. Akbarova, K. Rasulov, A. Khaidarov, G. Toirova, K. Kakharova, H. Ismailova, B. Ahmedova, M. Gurbanova.

### **Analysis**

Based on the communicative function of the language, Abdulkhamid Nurmonov spoke about the paralinguistic means that accompany the language when expressing thoughts and functionally compensate for it, as well as about the qualitative and quantitative characteristics of tone (voice), as well as gestures that are important. In communication J. emphasizes that the object of study should be linguistics [3]. Indeed, since the human factor is at the center of any language, and the discourse consists of the relationship between the addresser and the addressee, it is appropriate to study the language together with the person who is its regular user. Here I would like to add that silence, which is extremely important in communication, should be studied along with tone and gestures.

Therefore, the study of the processes associated with the stage of speech in the pragmalinguistic direction requires the study of such units

as silence and tone. Because the place of pragmalinguistics among the linguistic sciences, the very concept of the pragmatic system, the relationship between its components, the associativity of the pragmatic system with other related systems, silence in speech, tone, etc. have not yet been fully resolved.

In fact, both silence and tone are parts that are directly related to the main purpose of communication and provide new information when necessary. At this stage, we analyze silence as a phenomenon of conversation and the realization of inner speech in it.

When a person is silent and why is he silent? In what situations does silence reveal the inner speech inherent in people?

Usually, when a person gathers his thoughts, sums up what he has done or should do, when he looks at something, when he thinks about a lifestyle, about what he did, and when he thinks about a situation, neither about himself nor about others, he is silent. According to experts, two years is enough for a person to learn to speak, but sometimes even a whole life is not enough to learn to be silent.

In the process of communication, silence reveals the following internal states that are characteristic of a person:

1) **humiliation** (represents such meanings as hatred, disgust; contempt, insult; humiliating, degrading words or actions):

*Otabek recognized this old broom and noticed its irony. After that, he fully understood what was about to happen to him, and he did not answer these absurd questions, he was **silent**. By his **silence**, he seemed to give Muslimkul a meaningful answer. Muslimkul was very offended by such **silence** and shouted as if smoke rose to the sky:*

*– Why don't you answer?! (A. Kadiri. Past days)*

The hero of the work *Musulmankul* is known to the people for his bloodthirstiness, cruelty and ruthlessness. A Muslim is used to everyone trembling in front of him, shaking hands, kissing the ground, and every question he asks is immediately answered. At that moment, Otabek's silence really offended him and made him nervous. In Otabek's silence there was an expression of contempt, disgust, contempt for the specimen standing before him.

2) **indecisiveness** (sometimes this means that a person cannot come to a firm decision about a certain idea, concept or reality; be in different thoughts; be in a state of doubt):

*"This case should be used as an excuse to lose Asadbek! Only you can rule this city..."*



*There was **silence** after that. Hasilboyvacca appeared before Kilic with a smile on his face. It was incomprehensible to Hasilbayvachcha to hear such a statement from the person who was on Asadbek's wing, whether the head of the plant was sincerely saying this or started to provoke – he was not clear. (T. Malik. Shaitanat. Book 1)*

In this example, one of the interlocutors – the speaker is silent, waiting for the effect of what was said, while the listener is silent in a state of doubt and hesitation.

3) **succumbing to memories** (in such cases, a person remembers an event of his past life or during his lifetime and reveals the meaning of being in a whirlpool of memories):

*– He who has no homeland has no faith... – Ismailbey bowed his head and fell **silent**. What did the old man do, with flesh stuck to his forehead, his long nose even more unsympathetic, and his eyes narrowed, shrouded in this tent of **silence**, what was he thinking? And what about Zelikhan, whom this statement killed? Why is he silent? What is he thinking about? And what about Ismailbey's son Ahadbey, who was pouring tea? Ahadbey is equal to Zelikhan. If not for the cruelty of fate, would one be born and raised in the bosom of the mountains, and the other on the seashore, and would they meet here? They moved to the village one by one. Returning from the war, Ismailbey got a job on the railway and moved his family here. Zelikhan's father did not return from the war, being an orphan, he continued to gather food in the village. (T. Malik. Shaitanat. Book 1. p. 160)*

It's no secret that certain events in a person's life can leave pleasant memories or capture bitter suffering. In this example, the condition that silenced the interlocutors is statelessness. The agony of being forced to live away from home.

4) **internal munodzhot** (Munodzhot, in fact, is an Arabic word, which, in our opinion, was adopted and activated in the Uzbek language after the establishment of Islam in Maverannahr. Munodzhot means prayer, a secret conversation.) Sometimes the silence of a person's heart. It also represents the state of sending deep aspirations, prayers and repentance to God, which cannot be openly told to anyone. For example,

*Ismailbey often prays to Allah when he bows his head and is **silent**: "I begged you to save my life in the war, thank you, you spared me. I advised him to live in my village, close to my family, and see the waves of the sea, if only for one day. Thank you for bringing me back to my village. But you only lasted one day. You can't get enough of the sea waves. If I have many sins, I will be glad if you make me a patriot. But*

*I am glad that you have made my Motherland a patriot. But what if we are all sinners...” These words come to his mind, and he immediately repents of reproach to God. (T. Malik. Shaitanat. Book 1. p. 161)*

## Discussion

In this place, even when silence fully compensates for the speech means, one can feel the trace of the speech tool – articulation. In this case, it seems that the use of silence in the task of expressing an opinion with the requirement of a certain speech situation is temporary – optional.

Naturally, the presence of the interlocutor is an important factor in communication. On the contrary, the speech of the speaker in a solitary situation without any listeners, which does not require a response, in linguistics is called a speech act without an interlocutor or monologue speech. The German scientist D. Vandegveken calls it a non-communicative speech act, and a group of linguists call it egocentric speech and note that this type of speech act is often expressed by exclamations, exclamations, cries, prayers, calls [1].

6) ***unwillingness to continue the conversation*** (sometimes one of the interlocutors in the discourse may not want to continue the conversation for dozens of reasons, such as stuttering, boredom, fatigue, saving time, resentment. The expression of the listed cases is also done by default):

– *Our representatives left for Moscow, – Ahadbey said, – They will unite with the Crimean Tatars and reach the truth, if God wills.*

– *Many Chechens have left. Did I leave you too? Ismailbey said.*

– *My mother is here, my grandmother is here. Where am I going? Zelihan didn't say, “I was in prison.” The old man bowed his head and said nothing. (T. Malik. Shaitanat. Book 1. p. 161)*

Silence, like gestures and tone, has a certain meaning in human thinking. In other words, it is known that everything, events, actions, internal and external situations, relations in objective existence – everything is reflected in a person's thinking. Thought expressions materialize in language and appear in speech through words, phrases, sentences and texts. So, in this respect, silence performs in discourse the function of a word, a phrase, a sentence, a text.

7) ***unspoken answer to an unspoken question*** (In the process of discourse, the relationship between the speaker and the addressee is so tense that it becomes impossible to realize the goal of asking and

answering using speech or paralinguistic means. In such cases, the default question will be returned by default):

*Elchin asked today: "Does he get up and walk around?" he asked and received a simple answer: "Yes, he walks." "If he's coming, why isn't he coming out?" he wanted to ask, but refrained from asking because he knew why.*

*Wise Manzura looked into the eyes of his son-in-law and understood the question that he wanted to ask. "How to come to your presence? John blurted out some nonsense bitterly. Now his face is screwed up. I wonder what your future marriages will be, "she wanted to say, but did not say anything, because her son-in-law did not answer. **The default question was also answered by default.** (T. Malik. Shaitanat. Book 2. p. 76)*

Zainab, who lost her child, blames her husband Elchin for this tragedy. If at that moment the writer had "talked" to the characters (son-in-law Elchin and mother-in-law Manzura), then the impact of this episode would have decreased. The live performance of this episode was ensured that "the answer to the default question is also given by default".

8) **suppression of inner anger** (represents such meanings as curbing anger, trying not to be rude):

*"Things aren't going well for you. If we want, we can take you by car. Don't play with us.*

*This caused the old man's anger to flare up. But in order not to drive the guest away, he kept **silent** and forced himself to put pressure on him. Kesakpolvan decided this **silence** in his favor and said: "You have only one soul, yours." (T. Malik. Shaitanat. Book 3. p. 16)*

*Asadbek was very angry. He remained **silent**, realizing that if the argument continued like this, he might hurt his friend. Jalil considered this **silence** a sign of defeat and waited for his acquaintance to speak again with the air of a winner. (T. Malik. Shaitanat. Book 3. p. 71)*

Another example.

– *May God protect you and forgive your sins. Repent, my dear.*

*Asadbek screamed in pain. He paused for a few moments to drown it out. (T. Malik. Shaitanat. Book 3. p. 152)*

It can be seen from the examples that grammatical and lexical units provide full expressiveness, content and effectiveness of communication. However, a certain result can be achieved due to the communicative abilities of a person, eloquence, psychological factors, the interaction of group representatives in social communication, as well as the correct use of speech in the process of communication, taking into account the situation and conditions of communication. as well as

the effective use of non-verbal means, that is, communication can be effective or ineffective.

9) **confusion** (represents a state of silence due to not knowing where to start):

*Zainab wanted to shout out the next words. However, his lips moved a little, but he did not utter a sound. Finally the **silence** between the two sides ended and the black woman spoke up.*

– *What are you doing, are you going to hang yourself?*

*The black woman was polite this morning, but now she growls and her voice is rough. (T. Malik. Shaitanat. Book 3. p. 18)*

10) **a feeling of falsity of what was said or said** (sometimes, no matter how the addressee tries to convince the addressee, the addressee remains silent, and this situation indicates that he feels the falsity of what was said or said):

*The ascetic did not take his eyes off Jamshid, who was immersed in silence. Jamshid, who is able to teach many people intuition, realized that many things are hidden under this look.*

*Because of this, he felt that the fabrications he had made were in vain, and he was dissatisfied with himself.*

*Zahid's **silence** did not last long.*

– *You watch TV? – He said.*

*“Yes, so and so...” said Jamshid, not understanding the purpose of the question.*

– *They are making an announcement. There are programs called “Good night, kids.” It takes a good storyteller to run this show. Let me know. “This job is for you,” Zahid said with a venomous smile, holding out his hand to say goodbye and adding, “I am a fan of your stories.” We will meet again. (T. Malik. Shaitanat. Book 3. p. 204)*

## Conclusion

Sometimes the speaker or listener can remain silent in the process of reasoning, whether he wants to or not, and in a sense cannot deviate from the norms of the communication culture or violate them. It seems that human speech is an extremely complex, multifaceted phenomenon, and many aspects of it cannot be conveyed in writing. In order to reveal the full picture of linguistic information, it is necessary to comprehensively study all aspects of speech. After all, the complex of all situations involved in the speech process can reveal the full and true meaning of a dialogue fragment [3].

So, during a conversation, the thoughts of the interlocutors, their inner state of mind, often expressed verbally or non-verbally, their attitude is reflected in their mental and physical state, and all the thoughts that they want to express in a speech/speech situation Communication is reflected in their voice, speech can be transmitted to the listener through silence, as well as his posture, look and mental physical state. Indeed, in the process of verbal communication, it is important to approach the phenomenon of silence as a motor unit of internal and external speech.

## REFERENCES:

1. Vanderveken D. *Meaning and Speech Acts*. – Cambridge university press, 2009. – P. 137.
2. De Saussure, F. *Course in General Linguistics*. Columbia: University Press, 2011. – P. 260.
3. Yuldasheva D. N. Silence-is a nonverbal unit of speech communication// *American Journal of Research*”, 2020, 11–12 issue of the Journal. – Pag. 87–95. [https://journalofresearch.us/wp-content/uploads/2020/12/2020\\_11\\_12-AJR\\_11.pdf](https://journalofresearch.us/wp-content/uploads/2020/12/2020_11_12-AJR_11.pdf)
4. DN Yuldasheva. Basic non-verbal components of speech// *Current research journal of philological sciences (2767-3758)* 3 (01), 17–24. <https://masterjournals.com/index.php/crjps/article/view/661>
5. Yuldasheva Dilorom Nigmatovna, Gulchekhra Toshpulot qizi Chullieva. The role of nonverbal means in speech meaning//*International Engineering Journal For Research & Development*, Vol.6, Issue 2 published May 16, 2021.– Pag.1956-1961. [www.iejrd.com](http://www.iejrd.com) <http://www.iejrd.com/index.php/%20/article/view/1956>
6. Chullieva G.T. Intonation. Intonema, its structure. Functions of intonema// *Middle European Scientific Bulletin*. Volume 8, January 2021. ISSN 2694-9970. Pag. 141–148. <https://cejsr.academicjournal.io/index.php/journal/article/view/179>
7. G.T. Chullieva. Intonema and its Types//*Middle European Scientific Bulletin*. Volume 10, March 2021. ISSN 2694-9970. Pag. 91–95. <https://cejsr.academicjournal.io/index.php/journal/article/view/314>
8. D.N. Yuldasheva, G.T. Chullieva. Expression Of Euphemistic Meaning Through Silence And Intonation// *Published/ publié in Res Militaris (resmilitaris.net)*, vol.13, n°2, January Issue 2023.– 5596–5606-p. <https://resmilitaris.net/menu-script/index.php/resmilitaris/article/view/3187>
9. D.N. Yuldasheva, G.I. Toirova, D.Y. Yusupova. Silence is a Pragma linguistic event taking medical Ephemeric significance// *Philosophical Readings XIII.4 (2021)*, pp. 2290–2297. 2290 [https://www.researchgate.net/publication/365726821\\_Silence\\_is\\_a\\_Pragma\\_linguistic\\_event\\_taking\\_medical\\_Ephemeric\\_significance](https://www.researchgate.net/publication/365726821_Silence_is_a_Pragma_linguistic_event_taking_medical_Ephemeric_significance)

УДК 81-11

**ЛИНГВОПОЭТИКА ПОЭЗИИ ХАЛИМЫ ХУДОЙБЕРДИЕВОЙ  
И СПОСОБНОСТЬ ИСПОЛЬЗОВАТЬ ЯЗЫКОВЫЕ  
ВОЗМОЖНОСТИ УЗБЕКСКОГО ЯЗЫКА***Д. Ю. Юсупова**Бухарский государственный университет**Бухара, Узбекистан**dildorayusupova@gmail.com*

Недостаточно лишь теоретически определить различные описания языковых систем. Учитывая, что изучение речевых явлений языковых единиц в системе общения повышает внимание к человеческому фактору, ниже мы рассмотрим особенности эвфемизмов узбекской речи. Потому что национальная, культурная самобытность отражается в поведении, культуре и общении этих людей.

**LINGUOPOETICS OF HALIMA KHUDOIBERDIEVA'S POETRY  
AND THE ABILITY TO USE THE LANGUAGE CAPABILITIES OF  
THE UZBEK LANGUAGE***Yusupova Dildora Yunus kizi**Bukhara State University**Bukhara, Uzbekistan**dildorayusupova@gmail.com*

It is not enough to define the various descriptions of linguistic systems only theoretically. Considering that the study of speech phenomena of linguistic units in the communication system increases attention to the human factor, below we will consider the features of the euphemisms of Uzbek speech. Because national, cultural identity is reflected in the behavior, culture and communication of these people.

To achieve spiritual maturity, it is necessary to correctly, rationally and effectively use the language, which is a cultural heritage, to study it deeply and enjoy it. After all, the interaction of any person with people in the environment in which he lives, his beliefs, adherence to traditions, aspirations, national values ensure his formation as a person. Every aspect of this process is associated with language, or language acts as a vehicle in any situation.

**HALIMA XUDOBERDIEVA SHE'RLARI LINGVOPOEIKASI VA  
O'ZBEK TILINING TIL POTENTIALINDAN FOYDALANISH  
KOBILIGI.**

*Yusupova Dildora Yunus qizi*

*BuxDU*

*Buxoro, O'zbekiston*

*dildorayusupova@gmail.com*

Til tizimlarining turli tavsiflarini nazariy jihatdan aniqlashning o'zi etarli emas. Til birliklarining nutqiy hodisalarni aloqa tizimida o'rganish inson omiliga e'tiborni kuchaytirishini hisobga olib, quyida o'zbek nutqidagi evfemizmlarning xususiyatlariga to'xtalamiz. Chunki milliy-madaniy o'ziga xoslik bu odamlarning xulq-atvori, madaniyati va muloqotida namoyon bo'ladi.

**Introduction** . In Uzbek linguistics – both in traditional linguistics and in substantial linguistics – the grammatical nature of consonants has been studied in more detail. Some theoretical questions concerning the stylistics of the forms of coordination of R. The semantic and stylistic possibilities of affixes in the Uzbek language are revealed by Kolokolov on the basis of a large linguistic material, which makes this work one of the valuable works on the method of affixes. In the Uzbek language, the conjugation category, along with the main grammatical content, can perform a certain stylistic task in the text. This is especially evident when using one form of consent in the meaning of another. It is known that all consonants, except the main one, have grammatical indicators. For example, the transition agreement is expressed by the affixes-ning, the arrival agreement-ni, the departure agreement-ga, the place-time agreement-da, and the exit agreement-dan. Their application varies depending on the situation. But in some cases, the meanings of consonants can be understood even without their use in speech. In Uzbek linguistics, this phenomenon is called the asymptomatic use of matching affixes. The use or loss of consonant affixes is a phenomenon associated with the speech process. On the other hand, their inapplicability is also related to the law of striving for brevity in speech. The textbook “Stylistics of the Uzbek language” provides general information about euphemism and cocofemism (dysphemism) from a team of authors. This guide focuses on euphemisms and their stylistic role. In the manual, a “veiled” word or phrase used in place of a word or phrase that is inconvenient or inappropriate to pronounce is considered a euphemism. The use of euphemisms in fiction as the only stylistic means of getting rid of awkward and awkward situations, as well as

euphemisms used in working with images and characters, have a specific stylistic purpose, such as expressing purpose, worldview, etc., the level of each image and the character is underlined.

**Literature review.** In linguistic literature, dysphemism, cacothemism and strong words are considered the antonyms of euphemism. The first and second terms have an offensive and less humorous connotation, they are used as a pre-planned insult when something is intentional. The latter is mainly used in debates.

In the monograph of the linguist M. Mirtozhiev “Semasiology of the Uzbek language”, special attention is paid to euphemisms. In his monograph, the scientist pays special attention to the relationship between the phenomenon of taboo and euphemism, the history of research. The scientist emphasizes that the euphemistic meaning is a derived meaning, that the derived meaning is formed in a metaphorical, metonymic way, from which, in particular, the metaphor plays an important role in the formation of the euphemistic meaning. For example, the lexical meaning of the word *scorpion* is voiced by the word *donkey*. Because the tail of a scorpion is similar to that of a donkey. This similarity leads to the formation of a euphemistic metaphorical meaning of the word *donkey*. In his monograph, the scientist pays special attention to the relationship between the phenomenon of taboo and euphemism, the history of research. The scientist emphasizes that the euphemistic meaning is a derived meaning, that the derived meaning is formed in a metaphorical, metonymic way, from which, in particular, the metaphor plays an important role in the formation of the euphemistic meaning. For example, the lexical meaning of the word *scorpion* is voiced by the word *donkey*. Because the tail of a *scorpion* is similar to that of a *donkey*. This similarity leads to the formation of a euphemistic metaphorical meaning of the word *donkey*. This is due to the similarity between the metaphor-forming and the derived semantic referent. The scientist argues that the euphemistic meaning does not arise on the basis of synecdoche and function.

**Analysis.** N. Gaibullaeva studied medical euphemisms on the basis of periodicals, Sh. Gulomova drew attention to the gender specificity of euphemisms. It should be noted that in these studies, the euphemistic phenomenon is studied in relation to the human factor.

Euphemisms are manifested in the reflection of the national-mental interconnection of being, which shows the manifestation of cultural values in the language, in euphemism, along with clarity and expressiveness, there is also an attractiveness, abstractness of the worldview.



It is important to note that verbal euphemisms arise depending on a person's involvement in the linguocultural environment, his way of thinking, his worldview, perception of the subtleties of speech and his ability to use language. Euphemistic expressions consist of the main information content and the connotative meaning and attitude superimposed on it. The information is complicated by the euphemistic connotative meaning, when one of the negative or positive attitudes underlying the metaphorical image is in the lead among the connotative meanings. "An increase in connotative meaning occurs with a weakening of denotative and functional meaning" [1; 3].

It can be said that the need to express concepts that are morally and culturally unacceptable or inconvenient, more gently, sometimes in words, with a lot of connections, creates euphemisms, and each social stratum can always be created as a new discovery in speech. The value of euphemisms in speech can be characterized by their uniqueness, relevance, the ability to metaphorically see the harmony between things and events that need to be expressed in human thought, sensitivity to language, the power of intuition. So, a euphemism is a cognitive find.

**Discussion.** Euphemistic units corresponding to popular thinking are rapidly shrinking, gaining popularity, and gaining more and more distribution in public discourse. In fact, euphemism is also distinguished by its incompatibility with traditional semantic-syntactic patterns. Its use in speech is not directly related to an inconvenient reality (object, action, character, features, etc.), but occurs in circular paths. This "workaround" depends on the owner's perception of the euphemistic speech and on the extent to which he can perceive the corresponding quality of the euphemistic phenomenon. For example, the following euphemisms have already become common euphemisms in Uzbek speech:

inside a verb: *haqiqatdan uzoqlashmoq* (to lie; el on gapirmoq), *olamdan o'tmoq* (to die; o'lmoq), *yengil bo'lmoq* (give birth; tug'moq), *begona bo'lmoq* (divorce; ajrashmoq), *nomusiga tegmoq* (rape; zo'rarmoq), *uyli bo'lmoq*, *boshini ikkita qilmoq*, *turmush qurmoq* (to marry; uylanmok), etc.;

in the category of the word noun: *umr yuldoshi*, *jufti halol* (husband or wife), *oliy jazo* (death), *beldan past* (sexual organ), *qo'sh kabutar* (chest), etc.;

in the category of adjectives: *yoshi ulug'* (old), *o'ilmay gapiradigan* (stupid), *aqli noqis* (insane), *imkoniyati cheklangan* (disabled), *ko'zi ojiz* (blind), *aybi bor* (blue), *suyuqoyoq* (prostitute) etc.

Verbal (dynamic) euphemism is not stable and socialized, it is associated with the process of its emergence, in which its formation and application are associated only with the individual, and its non-static character is its main defining feature. The most beautiful examples of such euphemisms can often be found in the speech of masters of words, masters of words, writers, poets, orators. For example,

Original:

*Yel bo 'lib betoat qoqdim-u, lekin*

*Dil xonangni mudloq oholmadim men.* (Khalima Khudoyberdieva)

Transfer:

*I knocked restlessly like the wind, but*

***I couldn't open your heart room at all.***

The euphemistic line of the poem “*I could not open your heart room at all*” was used instead of the unit “*I was not loved*”. Yes, love is really full of love. Lack of love always brings endless longing, tremendous pain, and sometimes endless hatred to the human heart. The euphemism “*I could not open your heart room at all*”, used by the Uzbek poetess Khalima Khudoyberdieva, is very beautiful. Used instead. However, this has not stabilized in the common Uzbek language. It is advisable to refer to eccentric euphemisms as euphemisms inherent in the author’s poetic speech, unstable in general, not socialized.

Original:

*Nomard g'olib kelsa mard diltang-u zor*

*Bosh egib turganda qancha kuch sarflar.*

*... Ayting, yana necha asr solib iz,*

***Bu do'xaxiy so'zdan kuyar tilimiz ?!***

Transfer:

*If the enemy wins, the hero will be sad and hard*

*How much strength does a hero need to be silent ...*

*... Tell me how many more centuries do we need*

*Will our tongues burn again from this **hellish word**?!*

The poetess uses the euphemism “**hellish word**” instead of the word “*separation*”, which means the loss of our loved ones. This euphemism creates originality on the poem.

Or

Original:

*Siz, hey, bu dunyoning hisobdonlari,*

*Taxminan bo'lsa ham qilingiz hisob:*

***Qanchalik umri bor sun'iy shonlarning;***

*Sun'iy ko'rk qanchada bo'lg'usi xarob;  
Qancha davolangach past tushib har kas  
O'zin alloma-yi zamon sanamas...*

Transfer:

*You are the accountants of this world*

*Make an approximate calculation:*

*How long does **artificial glory** last;*

*How dirty the **artificial beauty** is;*

*Everyone needs to understand that a person needs treatment*

*Everyone must understand that no one is the master of this world.*

In his poems, *artificial fame*, *artificial beauty* means artificial pride, arrogance, and conceit inherent in some people. He uses euphemistic units to hide the names of human defects.

**Conclusion.** The Creator commented on the richness, purity of the Uzbek language, beautiful and different from other languages, showing in his poems.

#### REFERENCES:

1. Дилором Юлдашева (Dilorom YULDASHEVA). Профессиональные фразеология (Professional Phraseology). International conference on academic studies in philology (BICOASP) 26-28 September 2019 Bandırma <https://d1wqtxts1xzle7.cloudfront.net/60539392/>

2. Миртожиев М.М. Ўзбек тили семасиологияси. – Тошкент: Mumtoz so‘z, 2010. – 288 б. <https://www.twirpx.com/file/3052388/>

3. Шомаксудов А., Расулов И., Қўнғуров Р., Рустамов А. Ўзбек тили стилистикаси. – Тошкент: Ўқитувчи, 1983. – 248 б. <https://www.twirpx.com/file/1070737/>

4. Гайбуллаева Н.И. Ўзбек тилида тиббий эвфемизмлар (тиббий даврий нашр материаллари асосида): Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси автореферати. – Бухоро, 2019. – 50 б. <http://library.ziyounet.uz/uzc/book/108708>

5. Юлдашева Дилором Нигматовна. ОСОБЕННОСТИ НЕКОТОРЫХ ПРОФЕССИОНАЛЬНЫХ ФРАЗЕОЛОГИЗМОВ УЗБЕКСКОГО И АНГЛИЙСКОГО ЯЗЫКОВ. «Сўз санъати» халқаро журнали (ISSN: 2181-9297 DOI: 10.26739/2181-9297 [www.tadqiqot.uz](http://www.tadqiqot.uz)) –2021, 1-сон, 4-жилд. – В.:119-127. <https://tadqiqot.uz/index.php/art/article/view/446> <https://tadqiqot.uz/index.php/art/issue/view/41>

6. Юлдашева Д. Н. Сукут нуткий мулоқотнинг таркибий қисми сифатида //Сўз санъати халқаро журнали. – 2020. – Т. 3. – №. 3. [www.tadqiqot.uz](http://www.tadqiqot.uz)

7. G.T.Chullieva. Intonema and its Types. Middle European Scientific Bulletin. Volume 10, March 2021. ISSN 2694-9970. Pag. 91–95. <https://cejsr.academicjournal.io/index.php/journal/article/view/314>
8. DY Yusupova. Poetess Khalim Khudoyberdieva and his skill in use of the word. International Scientific Journal ISJ Theoretical & Applied Science Philadelphia, USA issue 01, volume 93 published January 30, 2021.– Pag. 238–241. <https://www.elibrary.ru/item.asp?id=44644284>
9. Dildora Yunus kizi Yusupova. About Polisemantic Words (on the example of the poems of the Uzbek poetess Khalima Khudoyberdiyeva). Middle European Scientific Bulletin. Volume 10, March 2021. ISSN ISSN 2694–9970. Pag. 384–389. <https://cejsr.academicjournal.io/index.php/journal/article/view/367>
10. G.T. Malikova, F.M. Boboxonova. МАЙМУДХО’ЖА БЕҲБУДИЙ: МИЛЛИЙ МА’РИФАТ РАҲНАМОСИ СИФАТИДА. INNOVATIONS IN TECHNOLOGY AND SCIENCE EDUCATION.P. 386M397. <https://humoscience.com/index.php/itse/article/view/370>
11. Худойбердиева Х. Сайланма. – Тошкент: Шарқ, 2000. – 398 б. <http://library.ziyounet.uz/ru/book/119112>

УДК 81-11

**ПЕДАГОГИКО-ПСИХОЛОГИЧЕСКИЕ ДЕТЕРМИНАНТЫ  
ОБЕСПЕЧЕНИЯ ПСИХОЛОГИЧЕСКОГО ЗДОРОВЬЯ  
БУДУЩИХ ИНЖЕНЕРОВ**

*Хайдаров Шахриёра*

*Бухарский государственный университет*

*Бухара, Узбекистан*

*sh.haydarov@gmail.com*

В этой статье представлен научно-теоретический анализ психологического здоровья будущих инженеров. Также раскрывается сущность и значение понятий, затрагивающих данную проблему.

**Ключевые слова:** психологическое здоровье, профессиональная компетентность, оптимизация, моральный дух, эстетический вкус, психическое здоровье

**БЎЛАЖАК МУҲАНДИСЛАРНИ ПСИХОЛОГИК  
САЛОМАТЛИГИНИ ТАЪМИНЛАШНИНГ  
ПЕДАГОГИК-ПСИХОЛОГИК ДЕТЕРМИНАНТЛАРИ**

*Хайдаров Шахриёр Шухрат ўғли*

*Бухоро давлат университети*

*Бухоро Ўзбекистон*

*sh.haydarov@gmail.com*

Ушбу мақолада бўлажак муҳандисларнинг психологик саломатлиги билан боғлиқ илмий-назарий таҳлиллар берилган. Шунингдек, ушбу муаммога дахлдор концепцияларнинг моҳияти ва аҳамияти очиб берилган.

**Калит сўзлар:** психологик саломатлик, касбий компетентлик, оптималлаштириш, ахлокий туйғу, эстетик дид, рухий саломатлик

**PEDAGOGICAL AND PSYCHOLOGICAL DETERMINANTS  
OF ENSURING THE PSYCHOLOGICAL HEALTH OF FUTURE  
ENGINEERS**

*Khaidarov Shakhriyor*

*Bukhara State University, Bukhara, Uzbekistan*

*sh.haydarov@gmail.com*

This article presents a scientific and theoretical analysis of the psychological health of future engineers. The essence and meaning of the concepts affecting this problem are also revealed.

**Key words:** psychological health, professional competence, optimization, morale, aesthetic taste, mental health.

Бугунги кунда бўлажак муҳандисларда психологик саломатликни тадқиқ қилиш ва уни мустҳкамлаш барча давлатларда турлича кўринишларда амалга оширилмоқда. Айнан муҳандислар билан боғлиқ психологик саломатликни шакллантириш тизимининг такомиллаштирилиши тадқиқот методикаларини танлашимизда муҳим аҳамият касб этади. Зеро, ишлаб чиқариш соҳасига ҳукумат томонидан алоҳида эътибор қаратилмоқда. Қолаверса, бугунги кунда ўсиб келаётган ҳар бир ёшнинг ўзига хос психологик имкониятларини тўғри талқин қилиш ва шу асосда уни ҳаётга, турли хил ижтимоий вазифаларни бажаришга йўналтириш барча соҳалар каби психологиянинг ҳам асосий вазифаларидан бири ҳисобланади. Қолаверса, ҳар бир шахснинг ўзига хос табиати, маънавияти, қизиқишлари ва ўзига хос дунёқарашлари мавжудки, ушбу хусусиятлар асосида шахс ҳаётини ҳар жиҳатдан тўғри ва мазмунли тарзда шакллантириш мумкин бўлади. Албатта, ушбу ишни амалга ошириш учун эса олий таълим муассасаларида таълим олаётган бўлажак муҳандисларда психологик саломатликни шакллантириш бўйича амалий тавсияларни ишлаб чиқиш муҳим аҳамият касб этади. Шунинг билан бирга, психологик саломатликни ўлчаш мезонлари, даражалари, кўрсаткичлари, ўзига хос хусусиятлари, тест яратиш ва ундан турли ёшдаги кишилар психологик саломатлигини текширишда фойдаланиш хилма-хил илмий-назарий негизларга қурилгандир. Ана шу боисдан, биз уларни таҳлил қилганда, бир нечта йўналишга ажратиб, талқин моҳиятини осонроқ тушуниш, идрок қилиш, муайян хусусиятга, шаклга, даражага, мантикий изчиллик билан фикрларнинг уйғунлашишига мурожаат қилишга ҳаракат қилдик. Бунда Республикамининг ҳар бир олий таълим муассасаларида таълим олаётган шахслар ўрганилиб муайян методикаларни қўллаш кўзда тутилади. Ушбу ҳолатдан келиб чиқиб тадқиқот методикалари танланди. Шу нуқтаи назардан ҳам тадқиқотимизнинг кейинги босқичида тадқиқот методикалари ва уларни амалиётда қўллаш тартиблари ҳақида маълумот берамиз. Маълумки, ҳар бир илмий тадқиқотнинг муваффақиятли яқунланиши, аввало шу тадқиқот салоҳиятини белгиловчи методикаларга боғлиқдир. Шу нуқтаи назардан тадқиқот мақсадига мос равишда танланган ҳар бир методика мазкур диссертацияда белгиланган вазифаларнинг бажарилишини эмпирик жиҳатдан

асослаб беришга хизмат қилади. Албатта ҳар бир методикани қўллаш ва унинг натижаларини таҳлил қилиш орқали муҳандисларда психологик саломатликнинг педагогик-психологик ўрни ва истикболларига эмпирик жиҳатдан баҳо бериш мумкин, деб ўйлаймиз. Шунинг учун психологик қўллаб-қувватлаш шахснинг касбий шаклланишини ўрганиш, шакллантириш, тузатиш ва ривожлантиришнинг ажралмас жараёнидир. Талаба шахсини психологик қўллаб-қувватлаш психолог фаолиятининг куйидаги турларини ўз ичига олиши мумкин: психопрофилактик иш; ва психологик маслаҳат; ривожланиш психокорректив иш; психологик таълим; психодиагностик иш (Макушкин 2007). Психологик саломатлик психология, назарий ва амалий психотерапия, психиатрияда энг кескин муҳокама қилинадиган муаммолардан биридир. «Психологик саломатлик» тушунчаси қадим замонлардан бери кўплаб олимларнинг тушунчаларида пайдо бўлган. Унга файласуфлар Алкмаеон, Цицерон, Эпикур, каби турли мактаб ва йўналишдаги психологлар З. Фрейд, А. Адлер, К. Г. Жунг, Ф. Перлс, А. Маслоу, В. Франклар кизиқош билдиришди.

Юртимиз психолог тадқиқотчилари орасида бу муаммо, Э.Ф.Ғозиев, Ғ.Б.Шоумаров, В.М.Каримова, А.М.Жабборов, Ш.Р. Баротов, З.Т.Нишонова, Д.Ғ.Мухамедова, Н.С.Сафоев, Р.С.Самаров, З.Г.Гаппаров, М.Маматов, А.А.Собиров томонидан ўрганилган. Бироқ, соғлиқни сақлаш тоифасининг кўплаб назарий ва амалий ишланмаларига қарамай, бугунги кунда унинг ҳолати жуда ноаниқ. Ушбу ҳодисанинг кўп ўлчовлилиги ва уни ўрганишга ёндашувларнинг хилма-хиллиги ҳали ҳам унга тўлиқ таъриф беришга, унинг таркибий қисмларининг етарли ва айни пайтда *minimal* тўпламини аниқлашга, руҳий саломатлик даражасини баҳолашнинг ишончли ва ишончли мезонларини ишлаб чиқишга имкон бермайди. Ушбу концепциянинг мазмуни кўп жиҳатдан тарихий даврга, жамиятнинг ижтимоий-иқтисодий тузилишига, шахснинг миллати ва ижтимоий келиб чиқишига, ахлоққа ва бошқа кўплаб омилларга боғлиқ. Инсон саломатлиги билан бевосита ёки билвосита боғлиқ бўлган ҳар қандай ижтимоий амалиётда мутахассис муқаррар равишда асосий саволларга дуч келиши керак: соғлиқ нима ва касалликнинг энг умумий таърифи нима? Бугунги кунга келиб, биз бу саволга аниқ ва жуда тўғри жавоб топа олмаймиз.

МДХ давлатлари психологлари томонидан ушбу соҳага илмий кизиқишнинг ортиши XX асрнинг сўнгги ўн йиллигига тўғри келади. «Бўшлиқни» тўлдириш ва руҳий саломатликнинг сўнгги

тушунчалари асосида шахсият назариясини қайта кўриб чиқишга аниқ уринишлар бугунги кунда энг йирик рус олимлари томонидан амалга оширилмоқда: Б. С. Братус, В. Я. Дорфман, Е. Р. Калиеевская, Й. М. Орлов, Д. А. Леонтиев ва ҳоказо. Ушбу тадқиқотчиларнинг асарларида инсоннинг рухий саломатлиги муаммосига табиатшунослик ва гуманитар ёндашувлар синтези баён этилган, қадриятлар ва маъно-ҳаёт йўналишлари, унинг фаровонлигини белгиловчи омил сифатида инсоннинг маънавий ва ахлоқий ўлчовлари ўрганилган.

Соғлиқни сақлаш масалаларини ўрганиш (хусусан, шахснинг рухий саломатлиги, ҳаёт сифатининг психологик компоненти) хориж психологиясининг устувор йўналишларидан бири бўлиб, саломатлик психологияси деб номланган мустақил психологик интизомнинг асосини ташкил этади, Хориж психологлари К. С. Счоу, Ж. Хюисон, У. Флик, Ж.Гордон, М. Мюррей, К. Чемберлен, Р. Шварцер, Ҳ. Стам, С. Уилкинсон томонидан талабаларнинг психологик саломатлигини белгиловчи асосий омиллар кўрсатиб ўтилди. Таъкидланишича, ушбу соҳадаги илмий билимларни кенгайтириш ва чуқурлаштириш зарурати соғломлаштириш амалиётининг тиббий ва психологик эҳтиёжларидан келиб чиқади.

Инсоннинг атроф-муҳитга мослашишига имкон берадиган ва шахснинг психологик саломатлигини ифодаловчи фазилатлар, муносабатлар ва функционал қобилиятларнинг умумийлиги. ЖССТ таърифига кўра, бу инсон ҳаётининг стрессларни енгишга, самарали ишлашга, ўз салоҳиятини максимал даражада рўёбга чиқаришга ва жамият ҳаётига ҳисса қўшишга қодир бўлган фаровонлик ҳолатидир. Инсон-бу фаолияти турли даражаларда тартибга солинадиган, бир-бири билан ўзаро боғлиқ бўлган мураккаб тирик тизим бўлиб ҳисобланади. Шундай Қилиб, Г. С. Никифоров (2005) рухий саломатликнинг куйидаги ўзига хос даражаларини кўриб чиқади: ижтимоий, психологик ва биологик. Биологик даражада саломатлик барча ички органлар функцияларининг динамик мувозанатини, шунингдек, атрофдаги дунё таъсирига етарли даражада жавоб беришни назарда тутаяди. Инсоннинг рухий саломатлигининг психологик жиҳати унинг ички дунёсига эътиборни назарда тутаяди: унинг ўзига бўлган ишончи ёки аксинча, ўзига ишончининг йўқлиги, ўз қобилиятлари, қизиқишлари ва қобилиятларини англаши; унинг атрофидаги дунёга, одамларга, ижтимоий воқеаларга, умуман ҳаётга муносабати. Соғлом психика ижобий ҳиссий ҳолат билан ажралиб туради, бу турли фактлар, ходисалар ва хо-



дисаларга жавобнинг барқарор хусусиятини белгилайди. Хукмрон ҳиссий ҳолат (қайғу, ташвиш, кувонч ва бошқалар.)- бу инсон ҳаётининг барча кўринишларида содир бўладиган кайфият ва ҳиссий фон. Ҳиссий ҳолат инсоннинг психологик саломатлиги учун жуда муҳим ва унинг ишончли кўрсаткичидир. Салбий ҳиссий ҳолат умуман соғлиқнинг ёмонлашишини ва хусусан психологик саломатликни кўрсатади. Психологик саломатлик муаммосини шакллантириш ва ривожлантиришдаги тарихий ташаббус Фарбнинг гуманистик йўналишдаги таниқли олимлари – Г. Аллпорт, А. Маслоу, К. Рожерсга тегишли. Илмий ва амалий фаолиятни инсоннинг ўзига хос кўринишлари ва умуминсоний қадриятларга қаратган ҳолда, гуманистик тенденция ўтган асрнинг 50-60 йиллари охирида шаклланди. Гуманистик оқимнинг ўзида келишмовчилик ва унинг чегаралари хиралашганига қарамай, гуманистик психология асосан инсоннинг ўзига хослиги ва ўзини ўзи таъминлашини тарғиб қилувчи янги психологик парадигма сифатида тан олинди. У билан биргаликда профессионал лексиконга индивидуал ҳаёт сифатини белгилайдиган атамалар қиради. Улар орасида психологик саломатлик бор.

Деярли барча муаллифлар «психологик саломатлик» глобал аҳамиятга эга бўлган ижтимоий-маданий, яхлит ҳодиса эканлигини тан олишади, уни ўрганиш ва баҳолашда фанлараро, интегратив ёндашувни талаб қилади. Деярли барча тушунчаларда, бу инсон атроф-муҳит омилларининг таъсирига жавоб беришнинг етарли усулини танлаш учун кўпроқ «эркинлик даражаларига» эга бўлишини англатади. Юқорида айтиб ўтилган муаллифларнинг далилларини умумлаштириб, шуни таъкидлаш мумкинки, инсон табиатига сўзсиз ишонадиган гуманистик ёндашув нуктаи назардан, психологик саломатликнинг умумий принципи-бу шахсий ҳаётнинг қийинчиликларига қарамай, инсоннинг ўзлигини сақлаб қолиш истаги. Биринчи замонавий рус психологларидан бири И. В. Дубровина руҳий ва психологик саломатлик тушунчаларини фарқлашга ҳаракат қилди. У психологик хизматнинг моҳиятини тушуниб, илмий психологик лексикага янги атама – «психологик саломатлик»ни киритиш керак деган хулосага келди. Руҳий саломатликни аниқлаш анъанавий равишда ўрнатилган тиббий ёндашув ва психотерапевтик йўналиш доирасида у соғлом шахснинг қолган хусусиятларини янги «психологик саломатлик» тушунчаси билан бирлаштиради. Агар «руҳий саломатлик» атамаси биринчи навбатда индивидуал психологик жараёнлар ва механизмлар би-

лан боғлиқ бўлса, унда «психологик саломатлик» атамаси умуман шахсни англатади, инсон руҳининг энг юқори кўринишлари билан чамбарчас боғлиқ (Дубровина 2009). «И. В. Дубровина психологик саломатликнинг мазмунли тавсифини беради. Психологик саломатлик шундан иборатки, инсон ўз нуқтаи назаридан ўзи билган дунёда муносиб, жой топади. Бу ҳаётга қизиқиш, фикр ва ташаббус эркинлиги, илмий ёки амалий фаолиятнинг ҳар қандай соҳасига иштиёқ, фаоллик ва мустақиллик, масъулият ва таваккал қилиш қобилияти, ўзига ишониш ва бошқаларни ҳурмат қилиш, мақсадларга эришиш воситаларини идрок этиш, кучли ҳис-туйғулар ва тажрибалар қобилияти, хабардорлик уларнинг ўзига хослиги ва ўзига хослиги ҳақида қувончли ажабланиб атрофдаги барча одамлар, ҳаёт ва фаолиятнинг турли соҳаларида ижодкорликдир» (Репина 2015). Дубровина шахснинг психологик саломатлиги ҳолатига таъсир қилувчи маълум бир таълим макони шароитида шахсни ривожлантириш муаммоларини кўриб чиқди. Ушбу концепция муаммоларнинг пайдо бўлишининг психопрофилактикасига, шу жумладан таълим маконининг параметрларини кузатиш ва тузатиш орқали устувор аҳамият беради.

Олиб борилган назарий-илмий таҳлиллар шуни кўрсатадики, бўлажак муҳандисларни психологик саломатлигини ривожлантиришни ҳар томонлама қўллаб – қувватлашни ташкил этиш, касбий компетентликни шакллантиришга ҳамда унинг атроф-муҳит билан ўзаро таъсири жараёнини оптималлаштиришга ёрдам беради.

#### АДАБИЕТЛАР:

1. Баротов Ш.Р ва бошқ. Экспериментал психология. – Магистрлар учун ўқув қўлланма. – Т.: Фан нашриёти, 2008. – 108 б.
2. Василенко, Е. В. Формирование профессионально-психологической устойчивости будущих сотрудников ОВД к экстремальным ситуациям : дис. канд. психол. наук : 19.00.07 / Василенко Елена Викторовна. – Ставрополь, 2008. – 229 с.
3. Ганзен, В.А. Описание психических состояний человека / В.А. Ганзен // Психические состояния / Сост. и общая редакция Л. В. Куликов. – СПб. : Питер, 2001. – С. 60–72. 296
4. Ғозиев Э.Ғ. Психология методологияси. Ўқув қўлланма. – Тошкент: Ношир, 2013.
5. Жабборов А.М. Ўзбек мактаби ўқитувчисининг психологик ва этник хусусиятлари: Псих. фан. док дис. – Т., 1999. – 317 б.
6. Каримова В.М. Социальные представления об узбекской семье юношей и девушек: Автореф. Дисс. Д-ра психол.наук. – Т.; 1994 – 40 с.

УДК 81-11

**ТВОРЧЕСКОЕ МЫШЛЕНИЕ У СТУДЕНТОВ  
УЗНАТЬ ВОЗМОЖНОСТИ  
ЭМПИРИЧЕСКИЕ ОСНОВЫ**

***У. Ш. Яхеева***

*Бухарский государственный университет*

*Бухара, Узбекистан*

*umidayaxyoyeva@gmail.com*

Процветание общества в мире во многом определяется его интеллектуальными возможностями. Потому что страна с высоким уровнем научного потенциала постоянно развивается во всех областях. В настоящее время, в современных государственных образовательных стандартах, становится важным воспитывать у учащихся креативность, критическое мышление, умение находить четкое решение проблем, возникающих в образовательном процессе.

**Ключевые слова:** творческое мышление, креативность, критическое мышление, любознательность, психологическое развитие, способности, стереотип, педагогическая импровизация.

**CREATIVE THINKING IN STUDENTS  
DISCOVER OPPORTUNITIES  
EMPIRICAL FOUNDATIONS**

***Yakheeva Umida Sharifovna***

*Bukhara State University, Bukhara, Uzbekistan*

*umidayaxyoyeva@gmail.com*

The prosperity of a society in the world is largely determined by its intellectual capabilities. Because a country with a high level of scientific potential is constantly developing in all areas. An important place in the historical development of states is the emphasis on youth. Currently, in modern state educational standards, it is becoming important to educate students in creativity, critical thinking, the ability to find a clear solution to problems arising in the educational process.

**Key words:** creative thinking, creativity, critical thinking, curiosity, psychological development, abilities, stereotype, pedagogical improvisation.

**O'QUVCHILARDA KREATIV TAFAKKUR  
XUSUSIYATLARINI O'RGANISHNING  
EMPIRIK ASOSLARI**

*Yahyoieva Umida Sharifovna*  
*Buxoro davlat universiteti*  
*Buxoro, O'zbekiston*  
umidayaxyoyeva@gmail.com

Axborotni tahlil qilish va sintez qilishda, axborotni birlashtirishda o'quvchilarda individuallik qobiliyatini yetarli darajada shakllantirish muhim ahamiyatga ega. Bu esa o'z navbatida boshlang'ich sinf o'quvchilarida kreativ tafakkurni shakllantirish uchun muhim asos bo'lib xizmat qilishi mumkin ekan. Qolaversa, bugungi kunda ilm-fan va texnologiyalarning tez sur'atlar bilan rivojlanishi sharoitida uzluksiz ta'limning barcha bosqichlari, jumladan, oliy ta'limning samaradorlik, natijadorlik va sifatini ta'minlashda ta'lim oluvchilarning kreativlik, mantiqiylik, mazmundorlik, ijodiy qobiliyatlarining rivojlanganlik darajalariga alohida ahamiyat qaratilmoqda.

**Kalit so'zlar:** kreativ tafakkur, ijodkorlik, tanqidiy fikrlash, qiziquvchanlik, psixologik rivojlanish, qobiliyat, stereotip, pedagogik improvizatsiya.

Mamalakatimizda rivojlanib borayotgan ta'lim islohotlari jarayonida o'qituvchi va o'quvchiga qo'yilayotgan talablar kun tartibidagi asosiy masalalardan biriga aylanib qoldi. Jamiyatdagi bu talablarga javoban ta'lim tizimi va sifati ham tubdan ham o'zgardi. Bugungi kunda o'qituvchining vazifasi faqatgina ta'lim berishgina bo'lib qolmay, balki har tomonloma keng tafakkurga ega, ham ruhiy ham jismoniy sog'lom shaxsni tarbiyalash kun tartibidagi kechiktirib bo'lmas vazifalardan biriga aylanib ulgurdi. bo'lib qoldi. Zamonaviy ta'lim tizimi qiziquvchanlik qobiliyati va odatini takomillashtirish va ijodiy faoliyatda tajriba orttirishga qaratilgan. Muhim jihati shundaki, bu tizimlarda har bir muammoni alohida hal qilish o'qitish amalga oshirmaydi, balki muayyan muammolar guruhlarini hal qilishning umumiy yo'lini, ularni hal qilish uchun umumlashtirilgan ishlarni bajarish usuli topiladi, ya'ni mavjud bilimlar ijodiy jarayonga xos bo'lgan yangi sharoitga o'tkaziladi [7]. Agar o'qituvchining ishlash tizimi o'quvchilarning aqliy rivojlanishida eng muhim omil bo'lsa, ijodkorlikning faoliyatda namoyon bo'lishi ularning ijodiy qobiliyatlarini rivojlantirishning sharti hisoblanadi.

Kreativ tafakkurni mezonlari ma'naviy va emotsional, anglash kabi xususiyatlardan tarkib topadi. Shuningdek, kreativ tafakkur jarayoni faoliyat jarayoni uchun muhim bo'lgan kreativ tafakkur darajalari,

ta'lim jarayonining nostandart ko'rinishlari, metodlarida namoyon bo'ladi.

Kreativ tafakkurni omillari: hayrat, faollik, emotsionallik, diqqatni jamlash, qiziqishlarni oshirish, taassurot chuqurligi, tonus va kayfiyatni oshirish, hamdardlik chuqurligi.

Mazkur ilmiy tadqiqotimiz maqsadidan kelib chiqib o'quvchilar kreativ tafakkurni quyidagi funktsiyalarga bo'lish mumkin:

- ijodiy, kognitiv jarayonlarning mavjudligi;

- ekstremal vaziyatlarda kreativ tajriba;

- umuminsoniy va pedagogik-psixologik bilim, malaka va ko'nikmalar. Pedagogik faoliyatning standart usullarining cheksiz xilma-xilligiga pedagogik vaziyatlarning noadekvatligi o'qituvchini obyektiv tarzda ijodkorlikka nisbatan rag'batlantiradi. Pedagogik faoliyatda bu diagnostik va metodik ijodkorlik (o'quvchilarni o'rganishning yangi yo'llarini qidirish va topish, yangi diagnostik metodlarni qo'llash va yaratish, o'qitish metodlarining yangi kombinatsiyalarini qidirish) deb aytiladi.

Shaxsdagi kreativlik xususiyatlarini rivojlanganlik darajalarini aniqlashga mo'ljallangan bir nechta usullar mavjud bo'lib, ularning ichida Vilyams metodikasi nisbatan valid (ishonchli) hisoblanadi. Sababi, mazkur metodika testlari aniq, turli yosh davrlariga mos bo'lib, shaxslardagi kreativ komponentlarni namoyon ettiradi.

Agar bolalar yozishni yoki juda sekin yozishni bilmasalar, eksperimentator va uning yordamchilari ularga rasm chizishda yordam berishlari kerak. Bunday holda, bolalar tasavvurining imkoniyatlari doirasida tasvirlarni to'liq kuzatish kerak.

O'quvchilarni matematika darslari orqali kreativ tafakkurini shakllanishiga ta'sir etuvchi xususiyatlar darajalarini tadqiq qilish zamonaviy ta'lim tizimidagi muhim omillardan biri sanaladi. Chunki, o'quvchilar nafaqat bilimlarni egallashi, balki o'sha bilimlar asosida yangi g'oyalarni ham yaratishi hozirgi kunda dolzarb maslala bo'lib qolmoqda.

O'quvchilarni matematika darslari orqali kreativ tafakkurini shakllanishiga ta'sir etuvchi xususiyatlar darajalarini tadqiq qilish zamonaviy ta'lim tizimidagi muhim omillardan biri sanaladi. Chunki, o'quvchilar nafaqat bilimlarni egallashi, balki o'sha bilimlar asosida yangi g'oyalarni ham yaratishi hozirgi kunda dolzarb maslala bo'lib qolmoqda. Bundan ko'rinib turibdiki o'quvchilarda faoliyat samaradorligidan qat'iy nazar, ularda fikrlashning ravonligida biroz qiyinchiliklarga duch kelishlari mumkin ekan. Bu esa ularning samarali

faoliyat yuritishlarida biroz qiyinchilik tugʻdirishi mumkin ekan.

Shuningdek, axborotni tahlil qilish va sintez qilishda, axborotni birlashtirishda oʻquvchilarda individuallik qobiliyatini yetarli darajada shakllantirish muhim ahamiyatga ega. Bu esa oʻz navbatida boshlangʻich sinf oʻquvchilarida kreativ tafakkurni shakllantirish uchun muhim asos boʻlib xizmat qilishi mumkin ekan.

Oʻquvchilarda kreativ tafakkur xususiyatlari darajasini aniqlash va shu asosida uni shakllantirish, ulardagi kreativ tafakkur sifatlarini rivojlanishiga toʻsiq boʻlayotgan salbiy psixologik taʼsirlarini oldini olish, uni korrektsiya qilish muammosi bugungi kunda kechiktirib boʻlmas muammolardan biri hisoblanadi.

#### ADABIYOTLAR:

1. Айзенк Г.Ю. Понятие и определение интеллекта// Вопросы психологии. – М.: 1995, – № 1. – С. 111–129.
2. Айзенк Г.Ю. Интеллект: новый взгляд// Вопросы психологии. – М.: 1995. – № 1. – С. 111–131.
3. Гилфорд Дж. Три стороны интеллекта // Психология мышления / Под ред. А.М. Матюшкина. – М.: 1965. – С 433–456.
4. Дружинин В.Н. Психологическая диагностика способностей: теоретические основы: в 2 ч. – Саратов : Изд-во Саратов. ун-та, 1990 – Ч.1 –137 с; Ч. 2-С. 139–292.
5. Rogers, C. Towards a theory of creativity // ETS: A Review of General Semantics. 1954. 11 P. 249–260.
6. Sternberg R.J. Creativity for the new Millennium. «-», 2001.
7. Ғозиев Э.Ғ. Усмонова Ш. Интеллект психологияси. – Т.: ТошДУ, Университет, 1996, – 48 б.
8. Ғозиев Э.Ғ. Онтогенез психологияси. – Т.: “Ўқитувчи”, 2010, – 360 б.

---

## СОДЕРЖАНИЕ

|   |     |
|---|-----|
| ПРЕДИСЛОВИЕ .....   | 13  |
| ЗАПРЕТЫ НА СОЧЕТАЕМОСТЬ МОРФЕМ В ХАКАССКОМ АВТОМАТИЧЕСКОМ МОРФОЛОГИЧЕСКОМ ПАРСЕРЕ. <i>А. В. Дыбо, В. С. Мальцева, Э. В. Султрекова, А. В. Шеймович, Ф. С. Крылов</i> .....  | 15  |
| ПРОБЛЕМА МОДЕЛИРОВАНИЯ ЛИНГВИСТИЧЕСКИХ СИНТАКСИЧЕСКИХ СЛОВСОЧЕТАНИЙ В ПРЕДЛОЖЕНИИ. <i>О. Х. Абдуллаева</i> .....  | 27  |
| К ФОРМАЛЬНОЙ МОДЕЛИ ТЮРКСКОГО ИМЕННОГО СОГЛАСОВАНИЯ: ДАННЫЕ КУМЫКСКОГО ЯЗЫКА <i>О. В. Федорова, С. Г. Татевосов</i> .....   | 39  |
| О СКАЛЯРНОМ ХАРАКТЕРЕ ЧИСЛОВОЙ НЕЙТРАЛЬНОСТИ. <i>С. Г. Татевосов</i> .....  | 47  |
| СЛОВОИЗМЕНТЕЛЬНЫЕ ПАРАДИГМЫ РУССКИХ ЗАИМСТВОВАНИЙ В ТУВИНСКОМ ЯЗЫКЕ, ОКАНЧИВАЮЩИЕСЯ НА СТЕЧЕНИИ СОГЛАСНЫХ -РТ. <i>Э. К. Аннай., Б. Ч. Ооржак, Ч. Г. Ондар, Н. М. Монгуш</i> .....   | 56  |
| ЛЕКСИЧЕСКО-ГРАММАТИЧЕСКИЕ СТРУКТУРЫ АГГЛЮТИНАТИВНЫХ ЯЗЫКОВ ДЛЯ РАЗРАБОТКИ НОВЫХ ТЕХНОЛОГИЙ ОБРАБОТКИ ИНФОРМАЦИИ (НА ПРИМЕРЕ ТАТАРСКОГО ЯЗЫКА). <i>Дж. Ш. Сулейманов, Р. А. Гильмуллин, И. Р. Мухаметзянов, А. Я. Фридман</i> .. | 64  |
| КОМБИНАТОРНЫЕ СВОЙСТВА ЛЕКСИЧЕСКИХ ЕДИНИЦ. <i>Б. А. Юнусова</i> .....   | 73  |
| НЕВЕРБАЛЬНЫЕ КОМПОНЕНТЫ КОММУНИКАЦИИ В ПРОИЗВЕДЕНИЯХ ФОЛЬКЛОРА. <i>Д. Б. Уринбаева</i> .....  | 77  |
| ОБЗОР: «ИНТОНАЦИОННАЯ МОДЕЛЬ ДЛЯ ПРОСТЫХ ПРЕДЛОЖЕНИЙ НА КАЗАХСКОМ ЯЗЫКЕ ДЛЯ КАЗАХСКОГО РЕЧЕВОГО СИНТЕЗАТОРА». <i>Енлик Кадыр, Бибигуль Разахова, Айжан Назырова</i> ..  | 82  |
| НЕКОТОРЫЕ СООБРАЖЕНИЯ О НЕВЕРБАЛЬНЫХ КОММУНИКАЦИЯХ. <i>Г. Х. Хасанова</i> .....   | 89  |
| ДИАЛЕКОМЕТРИЯ И АЗЕРБАЙДЖАНСКИЙ ЯЗЫК: ПРОБЛЕМЫ, РЕШЕНИЯ И ПЕРСПЕКТИВЫ <i>Афруз Гурбанова, Мехрибан Багирова</i> ..  | 98  |
| ПОСЛЕДСТВИЯ ФЕНОМЕНА СИНКРЕТИЗМА ДЛЯ СИНСЕТОВ ЛИНГВИСТИЧЕСКИХ ОНТОЛОГИЙ. <i>М.А. Абжалова</i> .....   | 113 |
| ВСПОМОГАТЕЛЬНЫЙ МЕТОД ОЦЕНКИ СТЕПЕНИ ИСТИННОСТИ ЭТИМОЛОГИЙ ОГУЗСКИХ ЭТНОНИМОВ СПИСКА М. КАШГАРИ, <i>И. А. Исмаилов</i> .....  | 122 |

## ЛИНГВОПРОЦЕССОРЫ

|   |     |
|---|-----|
| МОРФОЛОГИЧЕСКИЙ ПРЕОБРАЗОВАТЕЛЬ (АНАЛИЗ И СИНТЕЗ) ДЛЯ ЯКУТСКОГО ЯЗЫКА <i>В. Н. Кортесосо, В. П. Захаров</i> ..... | 132 |
|---|-----|

|  |     |
|--|-----|
| МОРФОЛОГИЧЕСКИЙ АНАЛИЗ СЛОВОФОРМ В УЗБЕКСКОМ, КАРАКАЛПАКСКОМ И КЫРГЫЗСКОМ ЯЗЫКАХ, ПРИНАДЛЕЖАЩИХ К ТЮРКСКОЙ СЕМЬЕ ЯЗЫКОВ. <i>Эльмира Назирова, Нилуфар Абдурахмонова, Шахноза Абидова, Мамура Узакова</i> . . . . . | 152 |
| МОРФОНОЛОГИЧЕСКИЙ АНАЛИЗАТОР ЯЗЫКА САХА: ОПЫТ РАЗРАБОТКИ И АПРОБАЦИИ. <i>Г. Г. Торотов, В. В. Бочкарев</i> . . . . .   | 158 |
| РАЗРАБОТКА СИСТЕМЫ ПРОВЕРКИ ОРФОГРАФИИ ТУВИНСКОГО ЯЗЫКА НА ОСНОВЕ HUNSPELL. <i>Ч. Г. Ондар, А. В. Чемышев, Ч. Б. Хуурак</i> . . . . .  | 166 |
| МОРФОЛОГИЧЕСКИЙ АНАЛИЗ ДЛЯ СБОРА ТЕКСТОВЫХ ДАННЫХ В НАЦИОНАЛЬНОМ КОРПУСЕ КЫРГЫЗСКОГО ЯЗЫКА. <i>Т. С. Садыков, Т. Туратали, А. Б. Турдубаев</i> . . . . .   | 178 |

### **КОРПУСНАЯ ЛИНГВИСТИКА И КОРПУСНЫЕ ИССЛЕДОВАНИЯ**

|   |     |
|---|-----|
| ПРОБЛЕМЫ КЫРГЫЗСКОЙ СИНТАКСИЧЕСКОЙ АННОТАЦИИ В ФРЕЙМВОРКЕ UNIVERSAL DEPENDENCIES. <i>Касиева Аида, Джумалиева Гульнара, Томпсон Анна, Юмашев Мурат, Чонтаева Бермет, Джонатан Вашингтон</i> . . . . . | 189 |
| АННОТАЦИЯ ГРАММАТИЧЕСКИХ ОШИБОК В ТЕКСТАХ НА ТАТАРСКОМ ЯЗЫКЕ С ПОМОЩЬЮ ИНСТРУМЕНТОВ КОРПУС-МЕНЕДЖЕРА. <i>Б.Э. Хакимов, Д. Р. Мухамедшин, З. И. Садыкова</i> . . . . .                                 | 216 |
| ИССЛЕДОВАНИЕ АВТОМАТИЧЕСКОГО ФОРМИРОВАНИЯ СИНТЕТИЧЕСКИХ КОРПУСОВ РЕЧИ ТЮРКСКИХ ЯЗЫКОВ. <i>У. А. Тукеев, Жандос Толеубеков, Толганай Балабекова, Жандос Жуманов, Бигелди Темирханов</i> . . . . .      | 224 |
| ВОЗМОЖНОСТИ СИСТЕМЫ УПРАВЛЕНИЯ КОРПУСНЫМИ ДАННЫМИ ДЛЯ РАБОТЫ С КОРПУСОМ КРЫМСКОТАТАРСКОГО ЯЗЫКА. <i>Д. Р. Мухамедшин, Б. Э. Хакимов, Л. Ш. Кубединова</i> . . . . .                                   | 236 |
| TATSC: ПЕРВЫЙ БОЛЬШОЙ ОТКРЫТЫЙ РЕЧЕВОЙ КОРПУС ТАТАРСКОГО ЯЗЫКА. <i>Р. А. Гильмуллин, Б.Э. Хакимов, М. Р. Галимов</i> . . . . .  | 249 |
| ОСОБЕННОСТИ ДИАЛЕКТНОГО СИНТАКСИСА БАШКИРСКОГО ЯЗЫКА (НА МАТЕРИАЛЕ ТЕКСТОЛОГИЧЕСКОЙ БАЗЫ ДИАЛЕКТОЛОГИЧЕСКОГО ПОДФОНДА МАШИННОГО ФОНДА БАШКИРСКОГО ЯЗЫКА). <i>Л. А. Бускунбаева</i> . . . . .          | 263 |
| АНАЛИТИЧЕСКИЙ ОБЗОР ПО СМЫСЛОВОМУ РАСПРЕДЕЛЕНИЮ СЛОВ В УЗБЕКСКОМ КОРПУСЕ. <i>Исроилов Жасур, Абдурахмонова Нилуфар</i> . . . . .  | 269 |
| ОБРАБОТКА НЕСТРУКТУРИРОВАННЫХ ДАННЫХ ДЛЯ КОРПУСА КАЗАХСКОГО ЯЗЫКА. <i>А. Н. Шормакова, Д. Р. Рахимова</i> . . . . .   | 285 |



|  |     |
|--|-----|
| О СТРУКТУРНО-СЕМАНТИЧЕСКОМ МОДЕЛИРОВАНИИ НА МАТЕРИАЛЕ КОРПУСНЫХ ПРОЕКТОВ БАШКИРСКОГО ЯЗЫКА. З. А. Сирзитдинов . . . . .  | 291 |
| ФУНКЦИОНИРОВАНИЕ ИНОЯЗЫЧНЫХ ФАРМАКОФИТОНИМОВ В БАЗЕ ДАННЫХ МФБЯ И В УСТНОМ ПОДКОРПУСЕ СМИ. А. Ш. Ишмухаметова . . . . .  | 298 |
| СОЗДАНИЕ АВТОРСКОГО КОРПУСА ЗАХИРИДДИНА МУХАММАД БАБУР – ТРЕБОВАНИЕ ПЕРИОДА. М. А. Абжалова . . . . .  | 305 |
| ВОЗМОЖНОСТИ АНАЛИЗА КЛАССИФИКАЦИИ ДИАЛЕКТОВ И ЯЗЫКОВ НА ЛИНГВОДОКЕ (на примере говоров северо-западного наречия башкирского языка). Ю. В. Норманская . . . . . | 311 |
| СОЗДАНИЕ И ЗНАЧЕНИЕ ЯЗЫКОВОГО КОРПУСА В УЗБЕКИСТАНЕ. Г. И. Тоирова . . . . .   | 325 |
| UZBEKCORPORA.UZ: СОЗДАНИЕ КОНКОРДАНСА И ЕГО АНАЛИЗ. А. Б. Каршиев, С. А. Каримов, М. С. Турсунов . . . . .   | 342 |
| РЕФЛЕКСИЯ В ЯЗЫКОВОМ КОРПУСЕ АНТРОПОНИМИЧЕСКИХ ЕДИНИЦ. Г. И. Тоирова . . . . .   | 352 |

#### ЛИНГВИСТИЧЕСКИЕ РЕСУРСЫ

|   |     |
|---|-----|
| БАЗА СОЦИОЛИНГВИСТИЧЕСКИХ И ЯЗЫКОВЫХ ДАННЫХ ПО ТЮРКСКИМ ЭТНОСАМ РЕСПУБЛИКИ КАЗАХСТАН. И. А. Невская . . . . .   | 357 |
| БАЗЫ ЗНАНИЙ ПОРТАЛА «ТЮРКСКАЯ МОРФЕМА»: СОСТОЯНИЕ, ПЕРСПЕКТИВЫ. А. Р. Гатиатуллин, Н. А. Прокопьев, Дж. Ш. Сулейманов . . . . .   | 370 |
| СОЗДАНИЕ БАЗЫ ДАННЫХ ПОЛИТИЧЕСКОГО ДИСКУРСА НА КАЗАХСКОМ ЯЗЫКЕ. А. Д. Сайранбекова, Л. О. Орынбай, А. Ж. Укенова, А. А. Шарипбаев, Б. Ш. Разахова . . . . .                                 | 389 |
| РУССКО-ТАТАРСКИЙ МАШИННЫЙ ПЕРЕВОДЧИК: ПОДГОТОВКА ДАННЫХ ДЛЯ ЗАПОЛНЕНИЯ БД РУССКО-ТАТАРСКИХ ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ. Мадехур Аюпов . . . . .  | 398 |
| ТЕЗАУРУС ПО МАТЕМАТИКЕ СРЕДНИХ ШКОЛ НА КАЗАХСКОМ ЯЗЫКЕ. А. А. Шарипбаев, А. К. Альжанов, С. А. Нариман, Г. Ж. Ахметова . . . . .  | 404 |
| РЕСУРСЫ СИСТЕМ АВТОМАТИЧЕСКОГО АНАЛИЗА ТЕКСТА. М. А. Абжалова, М. А. Адилова . . . . .  | 413 |
| ПРОЕКТИРОВАНИЕ ЭКСПЕРТНОЙ СИСТЕМЫ «DIALECTEXPERT» С ИСПОЛЬЗОВАНИЕМ ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ. Р. А. Бурнашев, М. Р. Галимов . . . . .  | 423 |
| ПОРТАЛ ИТ-РЕСУРСОВ ПО РАСШИРЕНИЮ ФУНКЦИЙ И ПОВЫШЕНИЮ КУЛЬТУРЫ КАЗАХСКОГО ЯЗЫКА. А. А. Шарипбай, Г. Т. Бекманова, Б. Ж. Ергеш, Алтынбек Зулхажав, А. С. Омарбекова, А. С. Муқанова . . . . . | 428 |

|  |     |
|--|-----|
| СОЗДАНИЕ БАЗЫ ДАННЫХ БУДДИЙСКИХ ТЕКСТОВ ХРАМА «ЦЕЧЕНЛИНГ». <i>А. Я. Салчак, А. Б. Хертек</i> ..... | 436 |
| О ВЫРАЖЕНИИ ЯЗЫКОВЫХ ТЕРМИНОВ В ЭНЦИКЛОПЕДИЧЕСКОМ СЛОВАРЕ. <i>Дилрабохан Рустамова</i> .....       | 441 |

### МАШИННОЕ ОБУЧЕНИЕ

|   |     |
|---|-----|
| РАЗРАБОТКА МОДЕЛЕЙ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА ДЛЯ КАЗАХСКО-РУССКОЙ ПАРЫ ЯЗЫКОВ. <i>В. И. Карюкин, Н. З. Абдурахмонова</i> .....                  | 450 |
| ПРИМЕНЕНИЕ МЕТОДА TRANSFER LEARNING К ЗАДАЧЕ МАШИННОГО ПЕРЕВОДА ДЛЯ ПАРЫ РУССКИЙ-ХАКАССКИЙ. <i>А. Ю. Лебедева</i> .....                             | 460 |
| РАСПОЗНАВАНИЕ ИМЕНОВАННЫХ ОБЪЕКТОВ НА ОСНОВЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ТАТАРСКОГО ЯЗЫКА. <i>В. Р. Гафарова, Ф. М. Гафаров</i> .....                | 472 |
| ПОСТРОЕНИЕ АЛГОРИТМА ПОЛУЧЕНИЯ СИНОНИМА ИЗ МЕДИЦИНСКИХ ТЕКСТОВ ДЛЯ КАЗАХСКОГО ЯЗЫКА. <i>Д. Р. Рахимова, А. С. Карибаева, Е. Р. Сулейменов</i> ..... | 486 |

### ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

|   |     |
|---|-----|
| ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ ДЛЯ ОБУЧЕНИЯ АЛТАЙСКОМУ ЯЗЫКУ. <i>Ч. П. Сабина</i> .....                              | 495 |
| МОБИЛЬНОЕ ПРИЛОЖЕНИЕ «ТЫВАЛАП ЧУГААЛАЖЬЫЛ» ‘ПОГОВОРИМ ПО-ТУВИНСКИ’. <i>С. А. Мылдыргыновна</i> .....                        | 500 |
| ПРИМЕНЕНИЕ STEAM-ОБРАЗОВАНИЯ В ПРОЦЕССЕ ОБУЧЕНИЯ БУДУЩИХ УЧИТЕЛЕЙ ТУВИНСКОГО ЯЗЫКА. <i>А. К. Тарыма, А. М. Чулдум</i> ..... | 505 |

### ЛИНГВИСТИЧЕСКИЕ ТЕХНОЛОГИИ

|  |     |
|--|-----|
| МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСЧЕТА НОРМ ЯЗЫКА ОРТАТЮРК КАК МОДЕЛЬ ЯЗЫКОВОГО РАЗВИТИЯ СИСТЕМЫ ТЮРКСКИХ ЯЗЫКОВ. <i>Б. Р. Каримов, Ш. Ш. Муталов</i> .....           | 510 |
| УЗБЕКСКИЙ НЕЙРО МАШИННЫЙ ПЕРЕВОДЧИК НА БАЗЕ VART: ИСПОЛЬЗОВАНИЕ МНОГОИСТОЧНИКОВЫХ ДАННЫХ. <i>А. И. Зохиоров, Н. З. Абдурахмонова, А. М. Нарзуллаев</i> ..... | 518 |
| МОДЕЛИРОВАНИЕ КАРАКАЛПАКСКОЙ ГЛАГОЛЬНОЙ ГРУППЫ ДЛЯ ЭТАПА МОРФОАНАЛИЗА. <i>А. З. Отемисов, Шарбаев Жарас</i> .....  | 527 |
| ЛЕКСИКО-СЕМАНТИЧЕСКАЯ РЕАЛИЗАЦИЯ КОНЦЕПТА “КРАСОТА” В УЗБЕКСКИХ И АНГЛИЙСКИХ ЭЛЕКТРОННЫХ КОРПУСАХ. <i>Т. Р. Яндашева</i> .....                               | 534 |
| КОМБИНАТОРНЫЕ СВОЙСТВА ЛЕКСИЧЕСКИХ ЕДИНИЦ. <i>Б. А. Юнусова</i> .....  | 542 |

|   |     |
|---|-----|
| МОРФОЛОГИЧЕСКИЕ СОЧЕТАНИЯ ОТ-ЛЕММ ДЛЯ ЯЗЫКОВЫХ КОРПУСОВ И ИХ ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ. С. А. Каримов, С. М. Умирова, Б. Ф. Холмухамедов, Дж. У. Туркашев . . . . .   | 546 |
| ОСНОВАННЫЙ НА ЗНАНИЯХ WSD ПОДХОД В УЗБЕКСКОМ ЯЗЫКЕ. Н. З. Абдурахмонова, Ж. Б. Исроилов . . . . .   | 559 |
| РАЗРАБОТКА УЗБЕКСКО-АНГЛИЙСКОЙ ДВУЯЗЫЧНОЙ ПРОГРАММЫ НА ОСНОВЕ МОДЕЛИРОВАНИЯ ГРАММАТИЧЕСКИХ КАТЕГОРИЙ ГЛАГОЛОВ И МОРФОЛОГИЧЕСКОГО АНАЛИЗА ДЛЯ МАШИННОГО ПЕРЕВОДА. Э. Ш. Назирова, Н. З. Абдурахмонова, Усмонова Камола . . . . . | 563 |
| МОРФОЛОГИЧЕСКИЙ АНАЛИЗ ПРИЛАГАТЕЛЬНЫХ В УЗБЕКСКО-АНГЛИЙСКИХ ЯЗЫКАХ ДЛЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ «ALIGNER». Ш. М. Хамроева, Н. Ш. Матъякубова, А. Ю. Даулетов . . . . .  | 571 |
| КЛАССИФИКАЦИЯ ФЕЙК-НОВОСТЕЙ С ИСПОЛЬЗОВАНИЕМ МОРФОЛОГИЧЕСКИХ ТЕГОВ И N-ГРАММ. Б. Б. Элов, Н. У. Худайбергенов, З. Ю. Хусаинова . . . . .  | 580 |
| ОБРАБОТКА КОРПУСНЫХ ТЕКСТОВ УЗБЕКСКОГО ЯЗЫКА МЕТОДАМИ WORD2VEC, GLOVE, ELMO, BERT. Б. Б. Элов, Р. Х. Алаев, З. Ю. Хусаинова, А. У. Юлдашев . . . . .  | 594 |
| ТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ SVD И NMF. Б. Б. Элов, А. У. Юлдашев, Н. Р. Алоев . . . . .   | 602 |
| ПРАКТИКА МАШИННОГО ПЕРЕВОДА В СОВРЕМЕННОМ МИРЕ: ОБЗОР. М. М. Кадирова . . . . .   | 612 |
| ВАЖНОСТЬ ПРОГРАММЫ ALIGNER ДЛЯ ПАРАЛЛЕЛЬНОГО КОРПУСА. И. А. Холмонова . . . . .   | 627 |
| МЕТОДЫ ONE-NOT КОДИРОВАНИЯ И МЕШКА СЛОВ ПРИ ОБРАБОТКЕ КОРПУСА ТЕКСТОВ УЗБЕКСКОГО ЯЗЫКА. Б. Б. Элов, Ш. М. Хамроева, Н. Ш. Матъякубова, У. С. Йодгоров . . . . .   | 635 |
| ОПРЕДЕЛЕНИЕ ОМОНИМОВ В УЗБЕКСКОМ ЯЗЫКЕ С ПОМОЩЬЮ АЛГОРИТМА ЛЕСКА. Б. Б. Элов, Х. И. Ахмедова . . . . .  | 642 |
| ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕКСТОВЫХ ДАННЫХ. М. Х. Примова . . . . .   | 653 |
| СОЗДАНИЕ ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ ДЛЯ АВТОРСКОГО КОРПУСА АЛИШЕРА НАВОИ. М. А. Абжалова, Н. С. Гуломова . . . . .   | 659 |
| ФОРМИРОВАНИЕ РЕЧЕВОЙ БАЗЫ УЗБЕКСКОГО ЯЗЫКА. С. Н. Ибрагимова, М. И. Абдуллаева . . . . .  | 670 |
| МЕТОД СЕНТИМЕНТ АНАЛИЗА, ОСНОВАННЫЙ НА ЛЕКСИКЕ. С. Ю. Алланазарова . . . . .  | 679 |
| NER: МЕТОДЫ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ТОПОНИМОВ В ТЕКСТАХ НА УЗБЕКСКОМ ЯЗЫКЕ. Б. Б. Элов, М. Т. Саматбоева . . . . .  | 685 |

---

**ФИЛОЛОГИЯ**

|  |     |
|--|-----|
| ИНДИВИДУАЛЬНОСТЬ И КОММУНИКАЦИЯ . . . . .  | 696 |
| ЛЕКСИКА ОРХОНСКИХ ПАМЯТНИКОВ И ЕЁ ОТРАЖЕНИЕ В КАРА-<br>КАЛПАКСКОМ ЭПОСЕ «ҚЫРҚ ҚЫЗ» . . . . .                 | 705 |
| ЛИНГВИСТИЧЕСКАЯ ХАРАКТЕРИСТИКА УЧЕБНОГО ДИСКУРСА И<br>ПРОБЛЕМЫ ПЕРЕВОДА . . . . .                            | 711 |
| МЕСТО ЛИНГВИСКОКУЛЬТУРНЫХ ЕДИНИЦ, РЕПРЕЗЕНТИРУЮ-<br>ЩИХ УЗБЕКСКИЙ МЕНТАЛИТЕТ, В ХУДОЖЕСТВЕННОМ ТЕКСТЕ . .    | 718 |
| КОММЕНТАРИИ К ЖАНРАМ, СВЯЗАННЫМ СО СКАЗКАМИ . . . . .  | 725 |
| ЗНАЧЕНИЕ ПАМЯТИ В ХУДОЖЕСТВЕННОМ ПОРТРЕТЕ . . . . .  | 730 |
| ЭТНО-ТЕРРИТОРИАЛЬНАЯ И КУЛЬТУРНАЯ ХАРАКТЕРИСТИКА ПЕРИ-<br>ФИРАСОВ . . . . .                                  | 734 |
| ДЕТСКИЕ ПЕСНИ, СОЗДАННЫЕ НА МОТИВЫ НАРОДНЫХ ПЕСЕН . .  | 741 |
| ВЫРАЖЕНИЕ СЛОВ В СЛОВАРЯХ С НАРЕЧНЫМИ -ОНА АФФИК-<br>САМИ . . . . .  | 745 |
| ИЗУЧЕНИЕ АРАБСКИХ СОГЛАСНЫХ ГРУППАМИ . . . . .   | 750 |
| ОКАНЬЕ В СИСТЕМЕ ГЛАСНЫХ УЗБЕКСКОГО ЯЗЫКА . . . . .  | 756 |
| ПЕРЕЁНИМЫ – ОДИН ВИД ОНОМАСТИЧЕСКИХ ЕДИНИЦ . . . . .   | 761 |
| СЕМАНТИКА ОБРАЗОВ В ПРОИЗВЕДЕНИИ «МЫШЛЕННЫЕ ДНИ»<br>АБДУЛЛЫ КАДИРИ . . . . .                                 | 765 |
| ГЕНДЕРНОЕ НЕРАВЕНСТВО И НАСИЛИЕ: СРАВНИТЕЛЬНЫЙ<br>АНАЛИЗ КОРЕЙСКОЙ И УЗБЕКСКОЙ КУЛЬТУРЫ . . . . .            | 771 |
| СОСТАВ УЗБЕКСКОЙ МЕДИЦИНСКОЙ ТЕРМИНОЛОГИИ . . . . .  | 782 |
| TALMENNING ERKIN VOHIDOV SHE'RLARIDAGI BADIY<br>VAZIFALARI . . . . .   | 787 |
| SOME THOUGHTS ON TEACHING VISUAL EXPRESSIVENESS OF THE<br>UZBEK LANGUAGE IN RUSSIAN-SPEAKING GROUP . . . . . | 791 |
| REALIYALAR MILLIY MADANIYATNI ANGLATUVCHI BIRLIK<br>SIFATIDA . . . . .                                       | 797 |
| INTONEMES ARE REPRESENTATIVES OF THE INTERNAL STATE OF<br>THE COMMICANT . . . . .                            | 802 |
| THE IMPORTANCE OF PRAGMALINGUISTIC FEATURES OF<br>LANGUAGE TASKS IN INTERNATIONAL COMMUNICATION . . . . .    | 809 |
| ХУДОЖЕСТВЕННО-КОМПОЗИЦИОННЫЕ ЗАДАЧИ ЭПИСТОЛАРНОЙ<br>ПУБЛИЦИСТИКИ И ЛЕТРИСТА В ДЕТСКИХ ПРОИЗВЕДЕНИЯХ . . . .  | 816 |
| LINGUISTIC CULTURE UZBEK CULTURE . . . . .   | 822 |

|   |     |
|---|-----|
| ПРАВОВОЕ ОБРАЗОВАНИЕ И ЕГО ФОРМИРОВАНИЕ НА ТЕРРИТОРИИ УЗБЕКИСТАНА .....   | 828 |
| ИЗУЧЕНИЕ КАДРОВ В ПОСЛОВИЦАХ С УЧАСТИЕМ «ВЕРБЛЮДА» ..   | 833 |
| ZAMONAVIY LUG‘ATCHILIK TARAQQIYIT BOSQICHLARI .....   | 839 |
| ИЗУЧЕНИЕ ЯЗЫКА КАРАКАЛПАКСКОГО ГЕРОИЧЕСКОГО ЭПОСА .....   | 843 |
| ETHNOONYMS OF BUKHARA .....   | 848 |
| “TEMUR MALIK” ASARIDAGI GIDRONIMLAR .....   | 855 |
| MAIN GENRES AND FEATURES OF VIRTUAL COMMUNICATION ..  | 860 |
| ПРОБЛЕМЫ ЯЗЫКА И СТИЛЯ В ГРАЖДАНСКОЙ ЖУРНАЛИСТИКЕ ..  | 864 |
| О ВЫРАЖЕНИИ НЕКОТОРЫХ АВТОРСКИХ РЕЧЕВЫХ НЕОЛОГИЗМОВ В ПОЭЗИИ МУХАММАДА ЮСУФА .....                              | 870 |
| XALQ OG‘ZAKI IJODIDA MAKTUBLAR .....  | 877 |
| ПРОБЛЕМА КОНЦЕПТУАЛЬНОГО И НОВЕРБАЛЬНОГО МЫШЛЕНИЯ   | 881 |
| XALQ SHE‘RIYATIDA GUL SO‘ZI ASOSIDA OBRAZLILIKNING VUJUDGA KELTIRILISHI .....                                   | 889 |
| ХУДОЖЕСТВЕННАЯ ИНТЕРПРЕТАЦИЯ ОБРАЗА ЖИВОТНЫХ В ДЕТСКОЙ ПОЭЗИИ (НА ПРИМЕРЕ СТИХОВ АНВАРА ОБИДЖАНА) ..            | 893 |
| ЖИВОТНЫЕ СИМВОЛЫ В ТВОРЕНИЯХ РАУФ ПАРФИ .....   | 898 |
| ЧЎЛПОН ШЕЪРИЯТИДА ТАБИАТ РАМЗЛАРИ .....   | 902 |
| ETYMOLOGICAL EXPLANATION OF THE PHENOMENON OF SONG AND THE PROFESSION THAT DENOTES IT .....                     | 906 |
| СОЦИАЛЬНО-ПСИХОЛОГИЧЕСКИЕ ФАКТОРЫ ФОРМИРОВАНИЯ ПРЕДСТАВЛЕНИЙ СТУДЕНТОВ ОБ УЗБЕКСКОМ МУЗЫКАЛЬНОМ ФОЛЬКЛОРЕ ..... | 911 |
| GENESIS AND DEVELOPMENT OF THE CONCEPTION AND THE CONCEPT “ROAD” .....  | 915 |
| ОБРАЗНЫХ ПРИЕМОВ И ПОЭТИЧЕСКИХ ФИГУР, ХАРАКТЕРНЫХ ДЛЯ ТЕКСТОВ ЭССЕ .....  | 919 |
| ПРО УЗБЕКСКИХ НАРОДНЫХ АНЕКДОТОВ .....  | 924 |
| О НЕКОТОРЫХ ЭТНОНИМАХ В «ДЕВОН» М. КОШГАРИЙ .....   | 931 |
| YOZUVLAR TARAQQIYOTIDA BOBURNING O‘RNI .....  | 936 |
| ТЕОРЕТИЧЕСКОЕ ИЗУЧЕНИЕ УЗБЕКСКИХ НАРОДНЫХ ПАРЕМ, СОЗДАННЫХ НА ОСНОВЕ ЛСГ «НАСЕКОМЫХ» .....                      | 942 |
| J. K. ROULINGNING FANTASTIK ASARLARIDAGI ANTROPONIMLARNING LINGVO-PERSPEKTIV MUAMMOLARI .....                   | 948 |

---

|   |      |
|---|------|
| THE PRAGMALINGUISTIC FEATURES OF COORDINATING CONJUNCTIONS .....  | 959  |
| ЧАСТНЫЕ ТЕРМИНЫ В НЕЧАСТНЫХ ДИСКУРСАХ .....   | 964  |
| КИНОЛИНГВИСТИКА В КАЧЕСТВЕ ОТРАСЛИ ЯЗЫКОЗНАНИЯ ...  | 969  |
| НЕКОТОРЫЕ СООБРАЖЕНИЯ ОБ УПОТРЕБЛЕНИИ СЛОВ В МЕДИЦИНСКИХ РЕКЛАМНЫХ ТЕКСТАХ .....                                  | 974  |
| ИЗУЧЕНИЕ ПРОБЛЕМЫ СОВРЕМЕННОЙ УЗБЕКСКОЙ ЛИТЕРАТУРЫ И МИРОВОГО ОБЩЕСТВЕННО-ЛИТЕРАТУРНОГО ДВИЖЕНИЯ ...              | 978  |
| ИМЕНА, ОБРАЗОВАННЫЕ НА ОСНОВЕ ИМЕНИ ЧЕЛОВЕКА. (БУХАРСКАЯ ОБЛАСТЬ, ВОБКЕНТСКИЙ РАЙОН НА ПРИМЕРЕ) .....             | 985  |
| ЛИНГВИСТИЧЕСКАЯ СЕМАНТИКА МОДАЛЬНОСТИ .....   | 989  |
| ВЗАИМОСВЯЗЬ ЛИЧНО-ОПИСАТЕЛЬНОЙ ЛЕКСЕМЫ И СЛОВСОЧЕТАНИЯ .....  | 994  |
| МОЛЧАНИЕ – ЕДИНСТВО, РАСКРЫВАЮЩЕЕ ВНУТРЕННЕЕ СОСТОЯНИЕ ЧЕЛОВЕКА .....   | 997  |
| ЛИНГВОПОЭТИКА ПОЭЗИИ ХАЛИМЫ ХУДОЙБЕРДИЕВОЙ И СПОСОБНОСТЬ ИСПОЛЬЗОВАТЬ ЯЗЫКОВЫЕ ВОЗМОЖНОСТИ УЗБЕКСКОГО ЯЗЫКА ..... | 1006 |
| ПЕДАГОГИКО-ПСИХОЛОГИЧЕСКИЕ ДЕТЕРМИНАНТЫ ОБЕСПЕЧЕНИЯ ПСИХОЛОГИЧЕСКОГО ЗДОРОВЬЯ БУДУЩИХ ИНЖЕНЕРОВ ...               | 1013 |
| ТВОРЧЕСКОЕ МЫШЛЕНИЕ У СТУДЕНТОВ УЗНАТЬ ВОЗМОЖНОСТИ ЭМПИРИЧЕСКИЕ ОСНОВЫ .....                                      | 1019 |

ОДИННАДЦАТАЯ  
МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ  
ПО КОМПЬЮТЕРНОЙ ОБРАБОТКЕ  
ТЮРКСКИХ ЯЗЫКОВ  
«TURKLANG 2023»

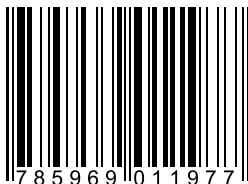
Труды конференции

*В авторской редакции*

Подписано в печать: 01.11.2023.  
Печать офсетная. Гарнитура «Times».  
Формат: 60×84 1/16. Усл.-печ. л. 59,87.  
Тираж 100 экз. Заказ № 1.11/23

Издательство Академии наук РТ.  
420111, г. Казань, ул. Баумана, 20.  
Тел./факс: (843) 292-49-14  
e-mail: izdat.anrt@yandex.ru

ISBN 978-5-9690-1197-7



9 785969 011977