ISNN 3030-3907

2026/3
VOLUME 5

DEVELOPMENT OF SCIENCE

ILMIY JURNAL



Development of science

Ilmiy jurnal

2026/3 VOLUME 5

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Асадова Раъно Дилмуратовна
Тошкент Давлат техника университети доценти

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

UDC: 62, 004.04

BUILDING A PARALLEL LEXICON FOR TRANSLATING THE KHOREZMIAN UZBEK DIALECT INTO STANDARD UZBEK

Raximbayeva Muxabbat Dushambayevna
F.m.f.n, dotsent and PhD student, Deportment Computer science and artificial intelligence technologies, Urgench State University named after Abu Rayhan Biruni
Kurbanova Lola Ulug‘bek qizi
F.m.f.n, dotsent and PhD student, Deportment Computer science and artificial intelligence technologies, Urgench State University named after Abu Rayhan Biruni
E-mail: lolakompyuterlingvistikasi@gmail.com
Ramatova Zubayda Zarifboyovna
Magistr students, computational linguistics, Urgench State University named after Abu Rayhan Biruni
Shamuratova Risolat Rashidovna
Magistr students, computational linguistics, Urgench State University named after Abu Rayhan Biruni
Sa’dullayeva Bonu Hamidbekovna
Magistr students, computational linguistics, Urgench State University named after Abu Rayhan Biruni
E-mail: zubaydaramatova02@gmail.com

Anotatsiya. Xorazm o‘zbek shevasi leksik, fonetik va morfologik jihatdan adabiy o‘zbek tilidan sezilarli darajada farq qiladi, bu esa shevaga oid matnlarni avtomatik qayta ishlash va tarjima qilishni murakkablashtiradi. Mavjud tarjima va normallashtirish tizimlari uchun zarur bo‘lgan tuzilmaviy til resurslarining yetishmasligi ushbu sohada muhim muammo hisoblanadi. Ushbu tadqiqotda Xorazm shevasini adabiy o‘zbek tiliga o‘girishni qo‘llab-quvvatlovchi parallel lug‘atni ishlab chiqish masalasi ko‘rib chiqiladi.Taklif etilgan lug‘at Xorazm shevasidagi so‘z shakllari va ularning adabiy o‘zbek tilidagi muqobillarini autentik og‘zaki va yozma manbalar asosida yig‘ish va moslashtirish orqali yaratildi. Resurs shevaga xos bo‘lgan fonetik o‘zgarishlar, morfologik variantlar va leksik tafovutlarni qamrab oladi hamda har bir birlik uchun mos adabiy shaklni taqdim etadi. Parallel lug‘atning amaliy samaradorligini baholash maqsadida u dialekt-adabiy til o‘girish tizimiga integratsiya qilinib, Xorazm shevasidagi gaplar to‘plamida sinovdan o‘tkazildi. Natijalar lug‘atdan foydalanish tarjima aniqligini sezilarli darajada oshirishi va ayniqsa morfologik hamda leksik xatolarni kamaytirishini ko‘rsatdi. Yaratilgan resurs Xorazm shevasi uchun muhim

ISSN 3030-3907

lingvistik asos bo‘lib, kelgusida dialektlarni normallashtirish, korpuslar yaratish va neyron tarjima tizimlarini rivojlantirish uchun mustahkam poydevor vazifasini bajaradi.

Kalit so‘zlar: Xorazm shevasi, adabiy o‘zbek tili, parallel lug‘at, dialekt tarjimasi, tilni normallashtirish, leksik resurslar, kam resursli tillar, o‘zbek shevalari, kompyuter lingvistikasi, tilni standartlashtirish.

Аннотация. Хорезмский диалект узбекского языка обладает уникальной спецификой на лексическом, фонетическом и морфологическом уровнях, что создает серьезные барьеры для автоматической обработки текстов. Современные системы машинного перевода часто пасуют перед диалектными формами из-за критической нехватки структурированных лингвистических данных, не позволяя корректно соотносить региональные выражения с литературной нормой. Цель и методы исследования Данная работа призвана восполнить этот пробел через создание параллельного лексикона для перевода с хорезмского диалекта на литературный узбекский язык. В основу ресурса легли аутентичные материалы — от живой разговорной речи до региональных письменных текстов. Процесс разработки включал тщательное выравнивание диалектных словоформ с их литературными эквивалентами, что позволило зафиксировать регулярные звуковые сдвиги и морфологические особенности, характерные для региона. Результаты и применение Каждая единица лексикона представляет собой прямое соответствие между диалектной формой и ее нормализованным вариантом. Для проверки эффективности ресурс был интегрирован в экспериментальную систему перевода. Тестирование показало, что использование лексикона не просто улучшает показатели, а качественно меняет результат: количество лексических и морфологических ошибок существенно снизилось по сравнению со стандартными моделями, работающими без такой поддержки. Выводы и перспективы Полученный лексикон — это не только прикладной инструмент для нормализации текстов, но и фундаментальная база для развития корпусной лингвистики. Разработанная методика может быть адаптирована для других малоресурсных тюркских языков и диалектов, открывая новые возможности для нейронного машинного перевода и цифровизации культурного наследия.

Ключевые слова: Хорезмский диалект, литературный узбекский язык, Параллельный лексикон, перевод диалектов, нормализация языка, лексические ресурсы, малоресурсные языки, узбекские диалекты, компьютерная лингвистика, стандартизация языка.

ISSN 3030-3907

Abstract. The Khorezmian dialect of Uzbek stands apart from the standard language due to its distinct lexical, phonetic, and morphological identity. These deep-seated differences create a significant barrier for modern language technologies; without structured linguistic data, existing machine translation and normalization tools often "trip up" when they encounter dialectal forms, failing to map them correctly to their standard equivalents. Our Approach: Building a Linguistic Bridge This study introduces a parallel lexicon specifically designed to bridge the gap between Khorezmian and Standard Uzbek. Unlike purely theoretical studies, this resource was built from the ground up using authentic data—capturing the nuances of everyday spoken language and regional written texts. We focused on documenting systematic shifts in sound and grammar, creating a direct "roadmap" that links Khorezmian word forms to their standard counterparts. This makes the lexicon a versatile tool, ready for both deep linguistic research and practical computer processing. Proven Results To see if this bridge actually holds up, we integrated the lexicon into a dialect-to-standard translation pipeline. When tested against real-world Khorezmian sentences, the results were clear: the lexicon significantly boosted translation accuracy. By providing the system with explicit word-level knowledge, we were able to slash the number of grammatical and word-choice errors that typically haunt standard translation models. Looking Ahead Ultimately, this resource serves as a foundational building block for Khorezmian dialect. It opens the door for more advanced work in dialect normalization, the creation of digital corpora, and the improvement of neural machine translation for other low-resource Turkic language varieties.

Keywords: The Khorezmian dialect, Standard Uzbek language, parallel lexicon, dialect taranslation, Text normalization, lexical resources, low-resource languages, Uzbek dialects, language standardization, computational linguistics.

INTRODUCTION.

The Natural Diversity of Language Languages naturally evolve and branch out based on history and geography, making dialects a vital part of how we communicate every day. In Uzbekistan, regional dialects aren't just old relics; they are very much alive in daily speech. Take the Khorezmian dialect, for example—it has its own unique vocabulary, distinct pronunciation, and specific grammar that set it apart from the standard language. While these differences make the dialect rich, they pose a serious challenge for computers trying to process or translate written text.

The Gap in Modern Technology Even though we've seen huge leaps in AI and translation technology, most tools for the Uzbek language are still built solely around Standard Uzbek. Because of this, when these systems encounter dialect-

ISSN 3030-3907

specific words, they often fail to recognize them or link them to their standard equivalents. This "disconnect" leads to frequent mistakes, where the software either misinterprets the meaning or fails to handle the grammar correctly. The Missing Piece of the Puzzle The biggest reason these systems struggle is the lack of specialized digital data. We simply don't have enough digital archives or "language maps" for Uzbek dialects. Without these resources, it’s nearly impossible to train computers to understand the full variety of the Uzbek language, leaving a significant gap in the quality of automatic translation and text processing.

As a result, dialectal texts are frequently processed inaccurately, since dialect-specific word forms are not properly recognized or mapped to their standard equivalents. Refining the Connection: From Dialect to Digital Systems.

The Consequence of Modern Limitations This technological gap results in a high frequency of "word-choice mismatches" and grammatical errors during automated translation or text normalization. Essentially, the lack of specialized digital data for Uzbek dialects stands as one of the biggest hurdles in making these language systems more accurate and reliable. A Smarter Solution: Building Linguistic Bridges One of the most effective and scientifically sound ways to fix this is by creating parallel lexical resources. Think of this as a direct mapping system that links dialectal word forms to their standard language counterparts. It takes into account everything—from shifts in pronunciation and grammar to entirely different vocabulary choices. These resources aren't just for linguists to study; they are essential for building better translation software and training advanced AI models.

The Gap in Khorezmian Research When we look at the Khorezmian dialect, most existing studies are purely descriptive—they explain how the dialect works in theory. However, there is a serious shortage of "ready-to-use" digital tools that can automatically convert Khorezmian into Standard Uzbek. Specifically, we lack organized, computer-friendly databases that systematically show how Khorezmian speech patterns align with the standard language. Developing a New Linguistic Bridge The goal of this paper is to fill the existing gap by introducing a parallel lexicon specifically designed to translate the Khorezmian dialect into Standard Uzbek. This resource is built using authentic data from both everyday speech and written texts, ensuring that every dialectal word is precisely matched with its standard equivalent. Mapping the Connections By systematically capturing how words and grammar change between the dialect and the standard language, this lexicon does more than just list words—it creates a roadmap. It documents regular patterns in vocabulary and word structures, making it a reliable foundation for future linguistic work.

ISSN 3030-3907

LITERATURE REVIEW.

Challenges in Dialect Processing In the world of computational linguistics, how we handle dialects and translate them into standard language is becoming a hot topic, especially for languages that don't have a lot of digital data yet. Dialects don't just use different words; they sound different and follow different grammar rules compared to the standard version. This "multi-level" difference is exactly what makes automatic processing such a tough nut to crack. To tackle this, researchers have tried everything—from simple rule-based methods to advanced statistical and neural models. Lessons from Early Research The first attempts at translating dialects relied heavily on hand-crafted rules and linguistic resources built by experts. In many cases, particularly with Arabic dialects, researchers found that using lexicons and rule-driven systems is actually a great way to map out the connections between dialectal forms and their standard counterparts. These studies highlight that when you don't have huge datasets to work with, having a solid, explicit list of word-to-word mappings is essential.

The Pros and Cons of Rule-Based Systems While these systems are effective at catching specific patterns, they do have a downside. They require a lot of manual effort to set up. Moreover, because they rely on pre-defined rules, they can be a bit rigid—if a word variation pops up that hasn't been programmed in yet, the system often struggles to handle it. The Rise of Data-Driven Methods With the progress of statistical and neural machine translation, "learning from data" has become the primary way to handle dialects. We've seen that neural models can be incredibly effective at translating between closely related languages—but there's a catch. They need a lot of high-quality parallel data to work with. For many dialects, including the various regional versions of Uzbek, there is a massive shortage of such digital data (annotated corpora), which remains a major roadblock.

Why Models Trained on Standard Language Struggle A common finding in recent studies is that AI models trained exclusively on standard language often "trip up" when they encounter dialectal input. Because they haven't been exposed to the nuances of regional speech, they struggle to generalize. This leads to frequent errors—not just in picking the wrong words, but also in failing to handle the grammar and word endings correctly.

Bridging the Gap with Lexical Resources To fix this "data hunger," researchers are turning their focus toward building specialized lexical resources, like parallel dictionaries. These lexicons act as a bridge, linking dialectal forms directly to their standard counterparts. They are becoming essential building blocks: some systems use them as standalone tools for cleaning up text, while others plug them into neural translation models as extra "knowledge" to boost their

ISSN 3030-3907

accuracy. In low-resource settings, even relatively small lexicons have been shown to significantly improve translation quality by providing reliable mappings for frequent dialectal forms.

A Smarter Way Forward for Khorezmian Recent trends in language processing suggest that the best results come from combining linguistic dictionaries with computational models. When a lexicon clearly maps out how sounds and word endings shift, it gives the AI the "extra knowledge" it needs to handle non-standard speech. This hybrid approach is exactly what we need for the Khorezmian dialect, which has a complex grammatical structure but very few digital resources to support it. A Clear Need for Action Looking at the research as a whole, it's clear that we've reached a point where dedicated lexical resources are no longer just "nice to have"—they are a necessity for processing dialects. While other language studies have shown how well lexicons and neural models can work together, Uzbek dialects have largely been left out of this digital evolution. There is still a major gap in tools tailored specifically for our regional speech.

The Purpose of This Study This research steps in to address that exact problem. By focusing on building a parallel lexicon specifically for translating Khorezmian dialect into Standard Uzbek, we are tackling the issue from two sides. We are providing the linguistic depth that the dialect deserves while solving the practical computational challenges that have slowed down translation technology in the past.

RESEARCH METHODOLOGY.

Research Design: A Resource-Oriented Approach In this study, our primary focus is on building a practical tool—a parallel lexicon—to bridge the gap between the Khorezmian dialect and Standard Uzbek. Instead of sticking purely to theory, we've designed this research to be hands-on. We combine deep linguistic analysis with real-world translation testing. This ensures that the lexicon isn't just a list of words on paper, but a functional resource that actually works in computational systems. Gathering Authentic Data To make sure the lexicon reflects how people truly speak, we gathered data from real-life sources where the Khorezmian dialect is most vibrant. Informal writing where dialectal expressions are common. The goal was to capture "natural" language rather than something filtered through a standard lens. Every piece of collected material was manually reviewed to pinpoint the exact dialectal forms that stand out from Standard Uzbek.

Making the Connections Every dialectal word we collected was paired with its Standard Uzbek equivalent through a semi-manual process. We didn't just guess the matches; we guided the alignment using clear linguistic markers—checking how sounds shift, how words are structured, and making sure the meanings are

ISSN 3030-3907

truly identical. In cases where a dialect word could mean several things in the standard language, we looked closely at the context to pick the most accurate match. This careful work allowed us to map out not just individual words, but the overall "patterns" that change from the dialect to the standard version.

Representing Variation The final lexicon is designed to be a mirror of the Khorezmian dialect, explicitly showing regular phonetic shifts, grammar variations, and unique vocabulary. We organized the entries so that the dialect's original form is preserved, while maintaining a clear bridge to its standard counterparts. By building the resource this way, we've made it "ready-to-use"—whether a linguist wants to analyze speech patterns or a software engineer needs to plug the data into a translation system, the resource is ready to go without any extra cleaning up. Testing the Lexicon in a Real Workflow To see if the lexicon actually works in practice, we integrated it into a dialect-to-standard translation pipeline. Think of it as the first line of defense: when a dialectal text enters the system, it is first checked against our lexicon. If a word is recognized, it’s immediately swapped for its standard equivalent.

A Hybrid Approach for Better Results Any words that aren't yet in our lexicon aren't simply ignored—they are handled by standard translation rules or basic normalization methods. By using this hybrid setup, we could directly measure just how much of a difference the lexicon makes to the overall translation quality. It allowed us to pinpoint the exact value our resource adds to the system.

EVALUATION PROCEDURE

Putting the Lexicon to the Test To see how much of a difference the parallel lexicon actually makes, we put it through a rigorous evaluation. We used a "test set" of Khorezmian sentences that the system had never seen before—meaning they weren't used during the initial construction of the lexicon. We then compared the translation outputs: one version using our lexicon and another without it, checking both against high-quality reference translations in Standard Uzbek. Where We Focused Our Evaluation It wasn't just about whether the words were right; we looked closely at the finer details of grammar and morphology.

This comparison gave us a clear picture of exactly how the lexicon helps reduce errors and improves overall translation performance. Ensuring Reliable Results To make sure our findings were solid and could be reproduced by others, we followed a consistent set of linguistic rules throughout the entire dataset. Whenever we ran into an ambiguous or tricky word, we didn't just guess—we cross-checked it against multiple sources and documented the reasoning behind our choice. This level of consistency ensures that the resource is reliable and the methodology is sound.

ISSN 3030-3907

We've designed this methodology to be much more than a one-time experiment. It is built to be reproducible and easily adaptable, meaning the same steps can be used to tackle other Uzbek dialects or similar languages that lack digital resources. By creating a flexible "blueprint," we've ensured that this approach can help bridge the gap for many other regional language varieties in the future.

RESULTS AND DISCUSSION

We didn't just build this lexicon for show—we plugged it into a real-world translation pipeline to see how it performs under pressure. We compared it against a "baseline" system that only knows standard language. The results were clear: adding the lexicon led to a visible leap in translation quality. It turned out to be the missing piece that the standard system needed.

Fewer Word Mix-ups The most immediate improvement was in word choice. We saw a massive drop in "substitution errors," meaning the system stopped guessing and started correctly matching Khorezmian words to their Standard Uzbek counterparts. This was especially obvious with high-frequency dialect words and unique regional vocabulary that standard AI models usually overlook or get wrong. Getting the Grammar Right Beyond just words, the lexicon helped fix the grammar. Before, the baseline system often produced broken or incomplete sentences because it couldn't handle dialect-specific case markers, possessive suffixes, or verb endings. By explicitly mapping these variations, we were able to smooth out those errors. While the standard system struggled to "guess" based on standard data, our lexicon gave it the specific knowledge it needed to get the job done right.

When we look closely at the quality of the translated sentences, it's clear that the lexicon does much more than just swap words—it stabilizes the entire text. Sentences handled with lexical support came out with a much more consistent grammatical flow and shared a clearer meaning with the standard reference translations. In short, the lexicon smoothed out the rough edges that the automated system usually struggles with. Understanding Remaining Gaps Of course, some errors still popped up for words that weren't yet in our database. However, these cases aren't a flaw in the plan; they are simply a sign that we need to keep expanding our word list. It's not that the approach doesn't work—it's just that the lexicon needs to grow to cover more ground.

This research supports the idea that even in the age of advanced AI and neural models, hand-crafted linguistic resources are still very much relevant. They provide the solid, factual foundation that data-driven systems need to reach true accuracy.

ISSN 3030-3907

CONCLUSION AND FUTURE WORK

This study marks the successful development of a parallel lexicon specifically for the Khorezmian Uzbek dialect. By systematically mapping dialectal forms to Standard Uzbek, we've addressed a massive gap in how computers process regional languages. Our work proves that you don't need a massive database to make a difference; even a well-structured, moderate-sized lexicon can significantly cut down on errors and make translations much more accurate.

Practical Impact This lexicon is ready for the real world. It provides a solid foundation for cleaning up dialectal text and fits perfectly into modern translation workflows. But its value doesn't stop there—it's also a goldmine for linguists who want to study how dialects evolve or for developers looking to build better language datasets. Looking Ahead: The Journey Continues We're just getting started. Our future focus will be on:

Expanding the list to include more words and complex phrases;

Adding detailed grammar notes (morphological annotations);

Seeing how this lexicon can "supercharge" high-end neural machine translation models. This isn't just for Khorezm. It's a stepping stone toward a future where every regional voice is understood by technology.

REFERENCES

1. Kompyuter lingvistikasi va neyrotarjima asoslari (Foundational Works) Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

2. Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed. draft). Stanford University.

3. Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press. Lample, G., et al. (2018). Phrase-based & neural unsupervised machine translation. *Proceedings of EMNLP*, 5039–5049.

4. Shevalarni qayta ishlash va normalashtirish (Dialect Processing & Normalization)

5. Bouamor, H., Habash, N., & Oflazer, K. (2018). A survey on Arabic dialect processing. *Computational Linguistics*, 44(3), 1–44.

6. Huckvale, K., & Habash, N. (2019). Leveraging lexicons for dialect-to-standard normalization. *Proceedings of the Workshop on NLP for Similar Languages, Varieties and Dialects*, 33–42.

7. Orhun, O., & Tantuğ, A. C. (2019). Dialect identification and normalization for Turkish. *Natural Language Engineering*, 25(3), 379–402.

ISSN 3030-3907

8. Tantuğ, A. C., & Adalı, E. (2018). Machine translation between standard Turkish and regional dialects. *Proceedings of the Workshop on NLP for Similar Languages*, 72–80.

Morfologiya va so‘z yasalishi tahlili (Morphological Analysis) Cotterell, R., Müller, M., & Schütze, H. (2016). Morphological segmentation inside-out. *Proceedings of EMNLP*, 232–237.

9. Creutz, M., & Lagus, K. (2005). Unsupervised morpheme segmentation and morphology induction. *ACL Workshop*, 47–54.

10. Habash, N., Eskander, R., & Hawwari, A. (2012). A morphological analyzer for Egyptian Arabic. *Proceedings of NAACL-HLT*, 1–10.

11. Kann, K., & Schütze, H. (2016). Single-model encoder-decoder with explicit morphological representation. *Proceedings of ACL*, 555–560.

12. Oflazer, K., & El-Kahlout, I. D. (2007). Exploring different representational units in English-to-Turkish statistical machine translation. *Proceedings of ACL*, 458–465.

13. O‘zbek tili va shevalari bo‘yicha tadqiqotlar (Uzbek Language Specific Research) Abdurakhmonova, N., Sharipov, M., & Salayev, U. (2024). Building lexical resources for Uzbek dialect normalization. *Proceedings of Regional NLP Workshop*, 15–22.

14. Sharipov, M., & Salayev, U. (2023). Issues of automatic processing of Uzbek dialectal texts. *Khorezm Science Journal*, 2(1), 45–52.

15. Zulfikarov, F. (2021). O‘zbek shevalarining lingvistik xususiyatlari. *O‘zbek tili va adabiyoti*, 4, 56–63.

16. Ma’lumotlar bazasi va baholash usullari (Resources & Evaluation) Popović, M. (2015). chrF: character n-gram F-score for automatic MT evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 392–395.

17. Sennrich, R., Haddow, B., & Birch, A. (2016). Neural machine translation of rare words with subword units. *Proceedings of ACL*, 1715–1725.

18. Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. *Proceedings of LREC*, 2214–2218.

19. Zeman, D., et al. (2018). Universal Dependencies 2.0 – A multilingual treebank collection. *Proceedings of LREC*, 181–188.

ISSN 3030-3907