

**THE USE OF THE UZBEK LANGUAGE IN THE DIGITAL
ENVIRONMENT: OPPORTUNITIES AND CHALLENGES****Mansurova Munavvar Mamadali qizi**

Educator at Preschool No. 22 (AMMC JSC),

Almalyk City, Tashkent Region

3rd-year student of the Preschool Education Department

at Nizami National Pedagogical University

Annotatsiya: Ushbu maqolada o'zbek tilining bugungi kundagi ahamiyati ochib berilgan. Uning raqamlashtirish hamda sun'iy intellekt vositalari bilan birgalikda qo'llash masalalari jahon tajribalari hamda Respublika miqyosidagi tadqiqotlar misolida tahlil qilingan. Bu boradagi mavjud muammolar yechimi uchun muayyan takliflar keltirilgan. Ushbu maqola o'zbek tilining keyingi taraqqiyot yo'lidagi tadqiqotlari uchun asos bo'lib xizmat qiladi.

Kalit so'zlar: o'zbek tili, raqamlashtirish, kompyuter lingvistikasi, korpus, sheva, nutq sintezatori.

Annotation: This article reveals the importance of the Uzbek language today. The issues of its digitization and use in conjunction with artificial intelligence tools are analyzed using world experience and national research. Specific proposals are made to solve existing problems in this regard. This article serves as a basis for further research on the development of the Uzbek language.

Keywords: Uzbek language, digitization, computational linguistics, corpus, dialect, speech synthesizer.

Аннотация: В статье раскрывается значение узбекского языка в современном мире. С использованием мирового опыта и республиканских исследований анализируются вопросы его цифровизации и использования в сочетании с инструментами искусственного интеллекта. Даны конкретные предложения по решению существующих в этой связи проблем. Статья служит основой для дальнейших исследований по развитию узбекского языка.

Ключевые слова: узбекский язык, оцифровка, компьютерная лингвистика, корпус, диалект, синтезатор

INTRODUCTION

In today's world, groundbreaking discoveries are being made in every field. The invention of the computer in science has led to a revolutionary transformation not only in scientific research but also in all areas and layers of society. Humanity has gained the ability to receive information from any corner of the world within minutes. The

opportunity to transmit and receive data online has brought forth important tasks such as developing and digitalizing languages. As a result, new modern disciplines such as computational linguistics, corpus linguistics, machine translation, and natural language processing (NLP) have emerged within the field of linguistics. In particular, these areas have been rapidly developing in Uzbek linguistics as well. One of the key challenges for specialists in this domain is to digitalize the Uzbek language—considering all its features and unique characteristics—and to create Uzbek-language databases for artificial intelligence technologies. These tasks are now recognized as priorities at the level of state policy.

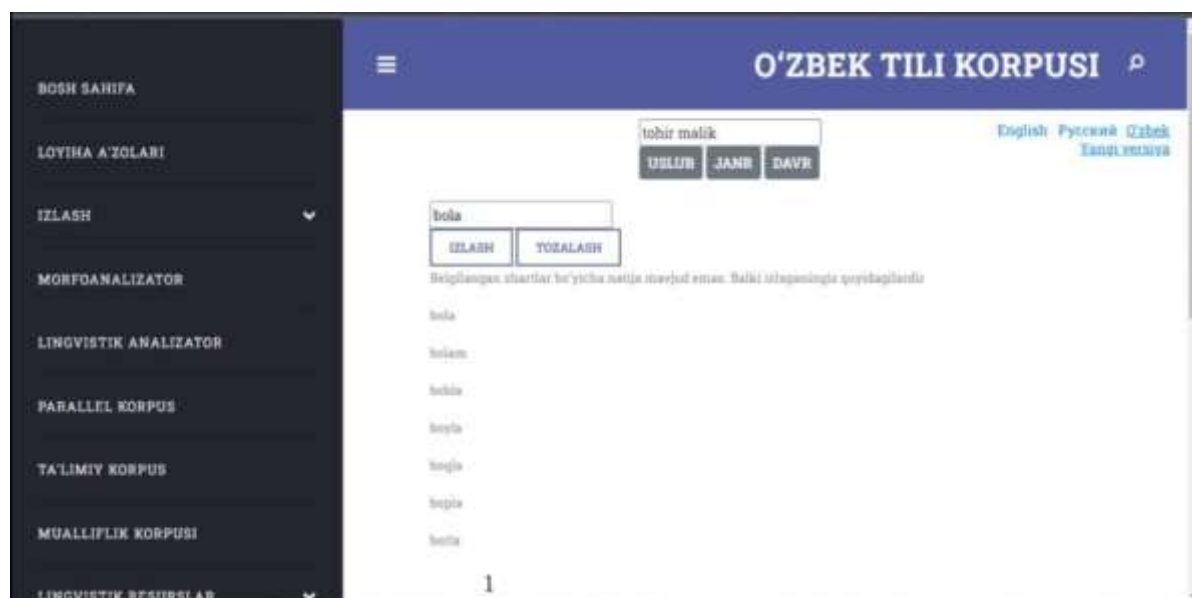
LITERATURE REVIEW

It is well known that, as in all fields of science, linguistics also involves the intersection of several disciplines. In particular, such interdisciplinary areas as sociolinguistics (sociology and linguistics), psycholinguistics (psychology and linguistics), ethnolinguistics (ethnography and linguistics), and neurolinguistics (neurology and linguistics) can be mentioned. Beginning in the 1950s, a new field known as mathematical linguistics emerged at the intersection of mathematics and linguistics. Initially, this discipline was referred to as machine translation or machine linguistics[1]. By the 1960s, computational linguistics had formed as an independent field of study[2]. The main purpose of this field is to develop computer programs capable of solving linguistic problems, to optimize communication between humans and computers, and to process natural language (Natural Language Processing, NLP)[3].

A wide range of tasks and challenges remain to be solved within computational linguistics. Among them are the development of automatic teaching systems, knowledge assessment tools, and systems that provide automatic morphological, syntactic, and semantic analysis of texts. Other important issues awaiting solutions include creating software for statistical analysis of dictionaries and digital texts, designing optimal programs aimed at addressing linguistic problems, and developing computer models of communication. The creation of electronic dictionaries, thesauri, and text corpora not only facilitates linguistic research but also provides significant advantages in many other fields. A number of Uzbek scholars have contributed to the formation and development of computational linguistics in the context of the Uzbek language, including A. Polatov, N. Abdurakhmonova, M. Musayev, M. Hakimov, M. Aripov, Sh. Hamroyeva, H. Arziqulov, S. Rizayev, S. Mukhamedov, S. Mukhamedova, and N. Jo‘rayeva. Their research primarily focused on statistical analysis, algorithmization, the axiomatic theory of the Uzbek language, and the computational analysis and synthesis of verbs.

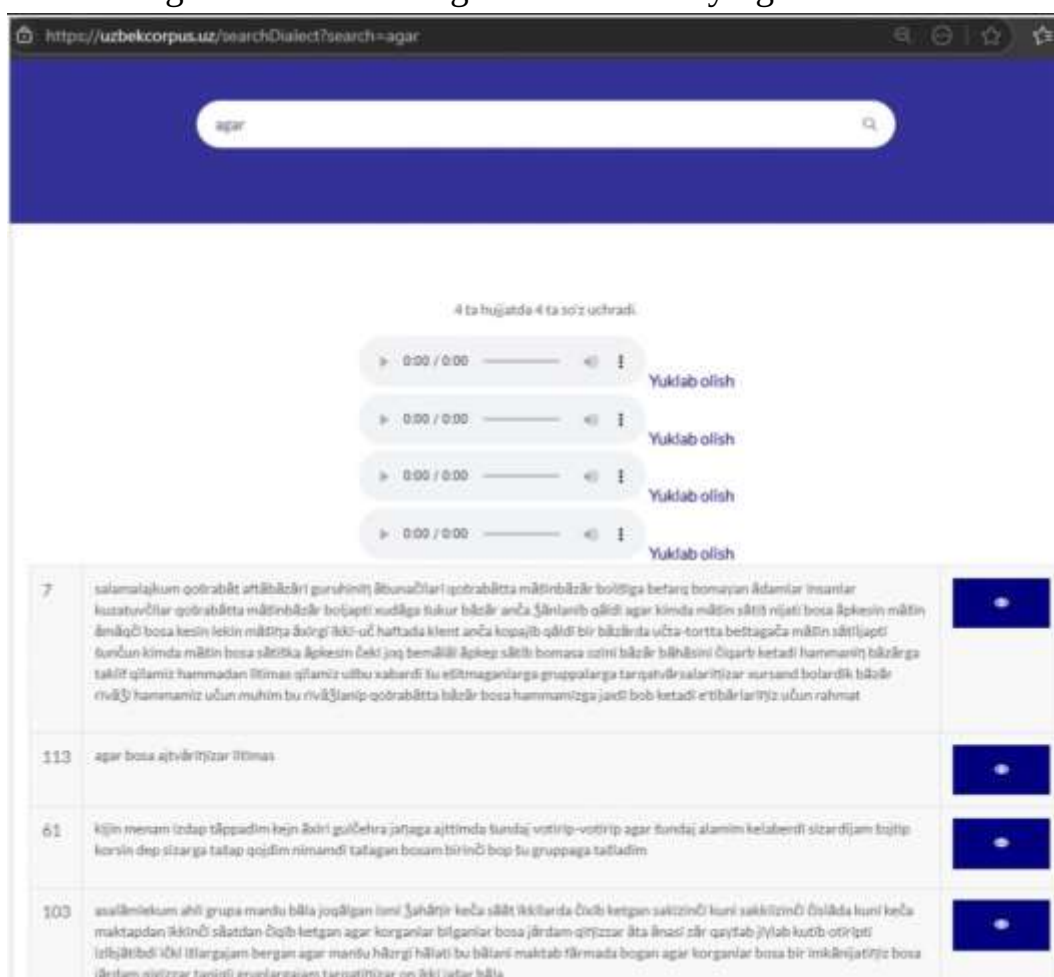
Currently, specialists in this field are engaged in developing explanatory and translation dictionaries of the Uzbek language, creating automated text-editing software, constructing a computational model of Uzbek grammar, and improving the parallel corpus of the Uzbek language. They are also expanding the Uzbek language corpus database and developing software tools. In short, computational linguistics is a modern branch of linguistics that serves to ensure theoretical and scientific interconnection between mathematical and natural sciences and professional linguistic disciplines [4]. The first computer-based text corpus, the Brown Corpus (BC), was created at Brown University in 1961. It contained 500 text fragments, each consisting of 2,000 words. In the 1970s, based on a corpus containing one million words, a frequency dictionary of the Russian language was compiled. As computational lexicography developed, the need for large-scale text corpora grew, leading to the creation of massive text corpora in many countries starting from the 1980s. These corpora were designed for various purposes and applications. The genre and diversity of a corpus depend on the user's field or interests—for example, Wikipedia serves as a large-scale text corpus within the field of science and education.

Digitalizing the Uzbek language ensures its preservation and eternity, enhances its global status, opens new opportunities for users, facilitates new discoveries, and enables it to be passed on as a rich heritage to future generations. Many dedicated Uzbek scholars have devoted their expertise and experience to this noble cause. The Uzbek Language Corpus was created within the framework of the scientific project “Creating an Electronic Corpus of Literary Works on the Themes of Family, Community, and Gender Equality” (Project JHBL-20) of the Institute for Scientific Research on Family and Neighborhood (Mahalla). The project was led by Professor Nilufar Abdurakhmonova, an expert in computational linguistics, machine translation,



and corpus linguistics. As a result of their efforts, the Uzbek Language Corpus (<https://uzbekcorpus.uz/>)[5] was officially introduced to users in 2020.

The corpus includes linguistic databases created for the Uzbek language. In particular, as part of the digitization of the Uzbek language, the peculiarities of homonyms, synonyms, analogies, and phraseological units of the lexical level have been studied from a linguistic point of view, and their electronic dictionaries have been created. The corpus includes linguistic databases created for the Uzbek language. In particular, as part of the digitization of the Uzbek language, the peculiarities of homonyms, synonyms, analogies, and phraseological units of the lexical level have been studied from a linguistic point of view, and their electronic dictionaries have been created. A multimedia corpus of the Uzbek language has also been created, in which the specific qualities of the spoken style can be seen through a variety of video content. In linguistics, the study of dialects is considered as a separate complex area. Computational linguistics has also begun work on studying various dialects of the area and



creating their databases. Through this corpus, you can get acquainted with both the transcription version and the audio text version of the dialect of a particular area. It also provides information about respondents in order to identify differences between dialects.

Figure 2.

In addition, research on the digitalization of the Uzbek language is being carried out at the Alisher Navoiy Tashkent State University of Uzbek Language and Literature, the Tashkent University of Information Technologies, as well as by specialists from Mohirdev LLC. The Mohirdev team is engaged in projects such as developing voice technologies in the Uzbek language and creating speech synthesizers based on both standard literary Uzbek and regional dialects. One of their notable projects, UzbekVoiceAI, is a clear example of such research initiatives[1]. At the same time, alongside these significant advancements in Uzbek language digitalization, there remain unresolved issues. For instance, there is a need for a platform that can perform accurate syntactic and grammatical sentence analysis for Uzbek in digital form. During translation processes, difficulties and ambiguities often arise with words that convey national and cultural nuances. Therefore, the creation of electronic translation dictionaries capable of distinguishing between the literal and idiomatic meanings of words has become a new priority for specialists in this field. Today, programs that can convert literary Uzbek text to speech and speech to text have already been developed. The next stage should focus on software capable of processing dialect-based texts and spoken messages, thereby expanding the practical applications of Uzbek in digital communication.

CONCLUSION

Because the Uzbek language is rich and diverse, revealing every aspect of it through digital means requires significant intellectual effort and time from specialists. The aforementioned research projects reflect years of dedicated work by experts in the field. Further advancement of the Uzbek language, its representation on digital platforms, and the creation of databases covering different linguistic styles and registers remain among the most important ongoing objectives.

LIST OF REFERENCES:

1. O‘zbekiston Respublikasi Prezidentining Farmoni, 28.01.2022 yildagi PF-60-son. PF-60-coH 28.01.2022. 2022-2026-yillarga mo‘ljallangan Yangi O‘zbekistonning taraqqiyot strategiyasi to‘g‘risida <https://lex.uz/ru/docs/-5841063>
2. O‘zbekiston Respublikasi Prezidentining Farmoni, 29.10.2020 yildagi PF-6097-son. PF-6097-coH 29.10.2020. Ilm-fanni 2030-yilgacha rivojlantirish konsepsiyasini tasdiqlash to‘g‘risida <https://lex.uz/ru/docs/-5073447>
3. O‘zbekiston Respublikasi Prezidentining qarori, 04.10.2019 yildagi PQ-4479-son. PQ-4479-coH 04.10.2019. O‘zbekiston Respublikasining “Davlat tili haqida”gi Qonuni qabul qilinganining o‘ttiz yilligini keng nishonlash to‘g‘risida <https://lex.uz/ru/docs/-4664611>

4. Мирзиёев Ш. М. Миллий ўзлигимиз ва мустақил давлатчилигимиз тимсоли // Халқ сўзи, 2019 йил 21 октябрь.
5. Ўзбекистон Республикаси Президенти Ш. М. Мирзиёевнинг 2019 йил 21 октябрдаги ПФ - 5850 - Фармони // Халқ сўзи, 2019 йил 22 октябрь.
6. Ўзбекистон Республикаси Президенти Ш. М. Мирзиёевнинг 2020 йил 20 октябрдаги ПФ-6084 - Фармони <https://lex.uz/docs/5058351>
7. Grishman R. Computational linguistics // Cambridge University Press. 1994. -P.4.
8. Jurafsky D., Martin J.H. Speech and Language Processing. - New Jersey, 2000. - P.2 -3.
9. Rahimov A. Kompyuter lingvistikasi asoslari / A. Rahimov; mas'ul muharrir A. Nurmonov; O'zR Oliy va o'rta maxsus ta'lim vazirligi, Andijon davlat universiteti. - T.: Akademnashr, 2011. - 160 b.
10. Umirzaqov E.B. Kompyuter lingvistikasi (uslubiy qo'llanma). Samarqand, 2011.
11. Uzbekcorpus.uz <https://uzbekcorpus.uz/>
12. <https://uzbekvoice.ai>