ADVANCEMENTS IN COMPUTER LINGUISTICS IN UZBEKISTAN

Zokirova Khulkar Izzatillayevna
Senior teacher of Angren University
Zakirovakh99@gmail.com

Over the past decade, Uzbekistan has seen significant progress in the field of computer linguistics, marked by the development of language processing technologies, translation tools, and natural language understanding. This progress has been facilitated by the country's strong emphasis on technological development and its recognition of the importance of linguistics in the digital age.

Growth of Computational Linguistics in Uzbekistan

One of the pioneering organizations in the development of computational linguistics in Uzbekistan is the Institute of Information Technology at the National University of Uzbekistan. This institute has been instrumental in researching and implementing various computational linguistic techniques, including machine translation, text analysis, and speech recognition in Uzbek language. Notably, the Institute of Information Technology has been at the forefront of developing Uzbek language corpora and resources for natural language processing. These resources serve as the foundation for the development of language models and various computational linguistic applictions tailored to the unique characteristics of the Uzbek language. The computer linguistics in Uzbek language is created in collaboration with specialists in the field of computer science, linguists, and philologists. Its positive aspect is evident in the collaboration of linguists and computer specialists. In the information age of the 21st century, along with the acceptance of valuable information and communication, paying great attention to its practical aspects would be beneficial. In other words, by combining dry theoretical ideas with comprehensive research, we believe it is essential for our nation to actively participate in creating computer programs that benefit our people, such as text editing, automatic translation, and language learning programs. For example, this work was initiated in the 1930s-1940s by I. Kissen, the founder of Uzbek linguostatistics. In the 1960s-1970s, scholars such as S. Rizaev, S. Muhammadiev, and S. Otamirzaeva actively engaged in linguistic research, leading to the establishment of a special laboratory at the Institute of Language and Literature of the Uzbek SSR. Although the activity of this laboratory continued to decline, it regained attention during the years of independence. Thus, the task is to further develop these works and establish the admission of students in higher education

institutions for the field of "Mathematics and Computer Linguistics", as well as ensuring the preparation of highly qualified personnel through the stages of master's and doctoral programs. Undoubtedly, computer programmers have a great need for linguistic support in carrying out these tasks, which should be provided by philologists, precise linguists, and other leading experts in the field.

Computer linguistics is considered not only as a means of learning Uzbek language but also other languages, especially Russian and English. Additionally, Computer Linguistics can be a valuable assistant in training translators specialized in scientific, literary, business, and advertising translations, as well as creating textbooks and computer programs that teach English in Uzbek and Uzbek in English in higher and secondary special education institutions. In order to solve the aforementioned issues and further develop computer linguistics in new directions, the Computer Linguistics Laboratory was established at the Faculty of Computer Technologies of the Uzbekistan State University in 2001. The laboratory focuses on the training of specialists in computer linguistics based on scientific and practical programs, which include creating mathematical and computer models of the Uzbek language, editing Uzbek texts, and developing English-Uzbek and Uzbek-English computer translation programs.

Establishing such laboratories and departments for "Mathematics and Computer Linguistics" in other higher education institutions in our country is also in line with the goal. This is because such departments are currently included in the curriculum of all philology faculties in higher education institutions. They were responsible for teaching subjects such as "Computer Linguistics", "Higher Mathematics", "Linguostatistical Methods of Text Analysis", "Linguostatistics", "Computer Lexicography", as well as creating lecture materials and textbooks. Since there hasn't been much research conducted in the field of Computer Linguistics in Uzbekistan, there hasn't been much experience in this area. In general, every linguist should be involved in creating websites, producing advertising videos through television and radio channels, and creating banners, poems, and transparencies that are placed on streets, institutions, and organizations.

Collaborative Initiatives and Academic Research

The growth of computational linguistics in Uzbekistan has also been bolstered by collaborative initiatives between universities, research institutions, and industry partners. For instance, the Tashkent University of Information Technologies, in collaboration with the Institute of Computational Linguistics of Heidelberg University, has been involved in joint research projects focused on Uzbek language

processing and sentiment analysis. Furthermore, academic research in the field of computational linguistics has led to the creation of Uzbek language tools for sentiment analysis, part-of-speech tagging, and named entity recognition. These tools have practical implications for various sectors, including social media analytics, market research, and government intelligence.

The promotion of Uzbek language to the level of world languages has remained one of the essential necessities within the ongoing process of integration and globalization worldwide. In achieving this significant and responsible goal, the need for computer technologies, particularly computer linguistics (CL), is increasing. This is because CL provides necessary opportunities for bringing Uzbek language to the global level, its integration with other world languages, and the advancement of learning and teaching tasks.

The main objective of CL is to develop computer programs that address linguistic issues and apply them to real life situations. It serves to perform various complex tasks related to processing texts using computers. Today, it is important to emphasize the following problems that require attention in the area of Uzbek linguistics and its computer-related solutions:

1. Developing a unique computer style for the Uzbek language.
2. Creating models for accuracy, conciseness, and clarity in information texts.
3. Establishing standards for creating Uzbek language websites.
4. Producing glossaries and translation dictionaries using computer technology.
5. Enhancing technologies for creating electronic versions of textbooks in Uzbek language and literature.
6. Creating English-Uzbek and Uzbek-English translation software programs.
7. Developing software programs for editing written texts and implementing them in daily life.

Given the importance of these issues, it is imperative to emphasize the creation of a distinctive computer style for the Uzbek language, as it is the foundation for solving all the aforementioned problems. It is worth noting that Uzbek language is proud of its capabilities, rich vocabulary developed over centuries, and its beauty, as emphasized by Alisher Navoi and other outstanding representatives. However, it should not rely solely on literary and descriptive tools, various figurative expressions, speech patterns, or occasionalisms. Although both literary style and computer style are developed and evolved in parallel through artistic (or artistic-literary) means, their application fields differ from each other. Representatives from

different fields use both styles (literary style and computer style). Computer style is based on clear, concise, and display-oriented ideas tailored to meet global requirements. We believe that it is essential to introduce computer style alongside literary style in schools, lyceums, and colleges, in order to effectively promote the complete use of the Uzbek language in our country. This way, students have the opportunity to choose the optimal variant based on their own personal needs: some may focus more on artistic style or literary style, while others can study computer style more deeply. For example, let's consider a comparative fact. In Tashkent, the 1st volume of the book "Dunyoviy o'zbek tili" (World Uzbek Language) was published with the initiative of the scientists from Uzbek State University (O'zMU) in 2003. It contains approximately 100,000 forms (lexico-grammatical paradigms) of a single verb in Uzbek language, while in English, there are about 100 forms per verb. In computer style, only the most necessary and acceptable (optimal) variant of these 100,000 forms is selected, while the rest are stored in the human memory with a certain degree of possibility.

Application of Computational Linguistics in Uzbekistan

The application of computational linguistics in Uzbekistan extends beyond academia and research, as it has found practical applications in the development of language technology products and services. For instance, local tech companies have been leveraging computational linguistics to create Uzbek language interfaces for chatbots, virtual assistants, and automated customer service platforms. These innovations have not only facilitated human-computer interaction in Uzbek but have also driven the digitalization of services across different sectors.

Additionally, the development of machine translation systems for Uzbek and other languages has been a key focus area, aiming to bridge communication gaps and facilitate cross-lingual information exchange. As a result, Uzbekistan has seen the emergence of machine translation solutions that cater to specific domains such as legal, medical, and technical documentation.

Challenges and Future Outlook

Despite the strides made in the field of computer linguistics, Uzbekistan faces certain challenges in the domain. One of the primary challenges is the scarcity of annotated linguistic data in Uzbek, which is essential for training and improving language processing models. Addressing this issue requires concerted efforts to expand language resources and create standardized linguistic datasets for research and development purposes. Furthermore, the integration of computational linguistic technologies into educational curricula and professional training programs

remains an area that requires attention. Training a skilled workforce proficient in computational linguistics is crucial for sustaining the growth and innovation in this field. Looking ahead, the future of computational linguistics in Uzbekistan appears promising as efforts continue to focus on enhancing language technology infrastructure, fostering academic-industry collaborations, and nurturing a robust ecosystem for research and development.

Conclusion

The development and application of computational linguistics in Uzbekistan have brought forth a wave of innovation, offering solutions that are tailored to the linguistic and cultural context of the region. By leveraging emerging technologies and interdisciplinary research, Uzbekistan is paving the way for the integration of language technology into various facets of society, from education and governance to commerce and industry.

As the field continues to evolve, the collaboration between academia, industry, and government will play a pivotal role in charting the course for advancements in computational linguistics, ultimately contributing to the robust digital ecosystem that Uzbekistan aspires to cultivate.

REFERENCES

Khayrullayev, R., "Computational Linguistics in Uzbekistan: Achievements and Perspectives," International Journal of Computational Linguistics (2018).

Ablimit, A., "Developments in Language Technology in Uzbekistan," Proceedings of the National Conference on Computational Linguistics (2020).

Karimov, S., "Applications of Computational Linguistics in Uzbekistan," Journal of Language Technology and Information Processing (2019).

Akhmedov, D., "Challenges and Prospects of Computational Linguistics in Uzbekistan," Conference Proceedings of Uzbekistan Linguistic Society (2021).

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)

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. aftoref." (2018).

Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.

Abdurakhmonova N, Tuliyev U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.

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.

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.

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

Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.