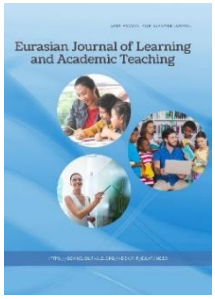

Development Of Computational Linguistics In Uzbekistan

Shahnoza Nasimova Ziyodullayevna

A teacher of English
The department of "Uzbek and Foreign languages" No 1
Tashkent State of Medical University

ABSTRACT

Computer linguistics or computational linguistics is an interdisciplinary field of linguistics, that analyzes and learns human language using computer methods. This field is gaining rapid popularity in advance with artificial intelligence in developing countries, such as Uzbekistan. This article traces the advances of development of computational linguistics in Uzbekistan and explores main methods, theoretical foundations and the problems it faces today in the world of linguistics.

Keywords: Uzbekistan, computational linguistics, AI, linguistics, methods, language, language model, language model, corpus linguistics

Introduction. Computational linguistics is a major area of interest within the field of linguistics. It has emerged as a critical interdisciplinary field among Uzbek linguists recently. Today Uzbek scientists are creating different tools and resources integrating natural language processing with computational methods. They are doing research not only in Uzbek but other world languages, like German, English and Russian too. Key areas of focus of their research include morphological analysis, semantic evaluation and the development of language models such as Uzbek WordNet (Kuriyozov et al., 2023).

A human language is a complex, structured and natural process. It is considered as context-dependent and relies on linguistic, social environment. In different situations the words may change their meanings and the same word may have multiple meanings. Therefore, modeling a language in computer requires not only knowledge on linguistics but knowing algorithmic techniques. The main goal of computer linguistics is to develop computer systems that can comprehend both spoken and written languages. In the past, researchers mainly pay attention to grammar and syntax, but today machine translation and speech recognition is the area of interest among linguists.

Until recently, a lot of research in the field has mainly focused on high resource world languages, like English (Cruz & Cheng, 2020). There is a significant gap in NLP research in low resource languages, like Uzbek (Kuriyozov et al., 2023). Therefore, Uzbek scientists are actively developing computational linguistics in the country, creating different language models to digitalize the language.

This article explores early stages of development of computational linguistics, modern research directions, challenges and future prospects of the field in Uzbekistan.

Methods. The basis of this research is formed by analyzing published scientific papers and linguistic datasets using analytical and qualitative methods.

Data collected from:

- Scholarly, published articles
- Research in corpus linguistics
- Open-source language technologies

Results. According to the data collected from the above sources shows that
computer linguistics has developed in Uzbekistan in different main areas, including linguistic corpora, NLP tools, language models for different functions and machine translation.

1. Uzbek national corpus**:** theoretical research began in 2010s in Uzbekistan. It

led to the creation of different corpora. One of them is Uzbek national corpus. It was developed in 2020 and consists of different functions like thesaurus, educational context, linguistic analyzer, morpho analyzer, dictionaries and etc.

2. Uzbek Speech Corpus. It consists of 105 hours of speech by 958 speakers in

Uzbek. It automatically recognizes speech and translates it. The project is ongoing and different researchers are improving and adding different functions to make it better.

3. NLP Tools in Uzbekistan.

Uzbek POS-tagging system is one of the results of researches on the field. It identifies Uzbek grammatical categories in the texts accurately.
Main areas of interest in Uzbek computational linguistics.

№	Research area	Description
1	Corpus linguistics	Digitalized collection of texts
2	Machine translation	Translation between languages
3	Speech recognition tools	Audio to text; text to audio automatically
4	Terminology	Collection of digital scientific terms

Discussion

Although it began recently in 2010, computational linguistics has made significant progress in Uzbekistan. Different institutions, such as National University of Uzbekistan, Uzbekistan World languages university, Tashkent State of Information technologies and other institutions have already developed different tools, corpora, language models and machine translation tools.

However, researchers are facing different challenges to make further progress in the area. The major challenges are as follows:

- Limited data
- Lack of corpora
- Deficiency of specialists
- High expenses to develop digital tools
- Difficulty in getting grants for research

The Uzbek language is one of the complex languages. One word may have different meanings. Besides that, complex morphological structure of the language makes the development of machine translation apps and the creation of AI more complicated.

Despite having these difficulties, researchers have made significant progress in

the development of the field. Furthermore, they are collaborating with international experts and joining Uzbek language digital linguistic tools to international linguistic datasets.

Conclusion. Computer linguistics is a new field of research in Uzbekistan that
began developing in the last decade. Despite this fact, it developed rapidly and researchers developed NLP tools, corpora, speech recognition tools, machine translation apps and etc to combine linguistics and technology.
The results of the study demonstrate that researchers are still facing different challenges to make research using computational methods. They should develop advanced language models, expand digital datasets and collaborate with international experts to ease these difficulties make further progress.

References:

1. Cruz, J. C. B., & Cheng, C. (2020). *Establishing baselines for text classification in low-resource languages.*
2. Kuriyozov, E et al., (2023). *Text classisfication dataset and analysis for Uzbek language.*
3. Nasimova, Sh.Z. (2026). The importance of choosing the right texts for learning English*. Journal of Til va adabiyot.*

4. Tabossi P. (1988). *Effects of context on the immediate interpretation of unambiguous nouns. Journal of Experimental Psychology*. 14(1). pp: 153–162.