

THE DEVELOPMENT OF CORPUS LINGUISTICS IN
UZBEKISTAN

Quliyeva Maxfurat

Student of BSU

***Abstract.** Corpus linguistics, the study of language using large collections of written or spoken texts, has been growing worldwide. While I don't have specific details on the development of corpus linguistics in Uzbekistan, I can provide a general overview. 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 Corps 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.*

***Key words:** Corpus linguistics, Uzbek National Corps , computer technology, mathematical linguistics.*

Introduction.

In many countries, corpus linguistics has gained importance due to its applications in language teaching, translation, lexicography, and linguistic research. It involves compiling, annotating, and analyzing large bodies of texts to uncover patterns of language use. In Uzbekistan, the study of language has likely evolved with advances in technology and a growing interest in linguistic research. Universities and research institutions may have developed their corpora for studying Uzbek and other languages spoken in the region.

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

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).

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.

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.[2:34] 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.[3:198] 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. [4:66]At the end of the conference, specific tasks were identified to create excellent national and educational corpus, drawing on international experience. 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. [5:13]Database technology is used in the process of creating traditional and electronic dictionaries. Dictionary bases of special and terminological dictionaries are being actively developed.

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. [6:65]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. [7:23]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". [8:87]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. [9:65]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. [10:13]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. L. I. Belyaeva, V. A. Chizhakovsky, Thesaurus in automatic text processing systems., Chisinau (1983)
2. L. Bloomfield, Language. Moscow, "Progress" (1968)
3. H. Bongers, The history and principles of Vocabulary control, Woerden: WOCOPI(1947)
4. V. G. Britvin, Applied modeling of syntagmatic semantics of scientific and technical text (by the example of automatic indexing), Moscow State University (1983)
5. M. Charlez, English corpus linguistics: An introduction. Cambridge University Press,UK, 168 (2004)
6. A. Eshmuminov, Synonymous database of the Uzbek language national corpus. Dissertation of PhD in Philology, Tashkent (2019)
7. Ch. C. Fries, The structure of English. An introduction to the construction of English sentences, London (1969)
8. Sh. Hamroeva, Linguistic bases of creation of the author's corpus of the Uzbek language: Author's Abstract of the Dissertation of PhD in Philology, Tashkent (2018)
9. R. Karimov, Linguistics and programming issues of creating a parallel corpus of Uzbek and English, Author's Abstract of dissertation of PhD, Bukhoro, 151 (2021)
10. R. Kuldoshev, 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, 371 (2023)
11. Rakhimov, M. (2024). KORPUS LINGVISTIKASI TARAQQIYOTI VA O'ZBEK TILSHUNOSLIGIDA KORPUS TAHLIL ASOSLARI. CENTER FOR SCIENTIFIC PUBLICATIONS (buxdu.Uz), 45(45). Retrieved from http://journal.buxdu.uz/index.php/journals_buxdu/article/view/11736