

The Current State Of World Corpus Linguistics: National Corpora

Mahmudova Dildora Murodilloyevna

PhD Student of Bukhara State University, ESP Teacher At Bukhara State Pedagogical Institute, Uzbekistan

Received: 27 January 2025; **Accepted:** 16 February 2025; **Published:** 21 March 2025

Abstract: The theme examines the role of national corpora as vital resources for linguistic research and language policy. These large collections of texts reflect the diversity and dynamism of a nation's language, capturing various forms of communication across different contexts. Recent advancements have enhanced the compilation and analysis of national corpora, incorporating digital technologies and a wider range of text types, including spoken and informal language. This development has led to increased collaboration among researchers globally, fostering a better understanding of linguistic variation and change. National corpora are crucial for applications in language education, computational linguistics, and sociolinguistics, providing data that inform teaching practices and natural language processing tools. Looking ahead, the field faces challenges such as ensuring inclusive representation of languages and dialects while embracing new technologies. Overall, national corpora serve as a cornerstone for understanding language in its cultural and social contexts.

Keywords: Corpora linguistics, modern Corpora, first generation Corpora, second generation Corpora, National Corpora, demand, general principle, private principle, representativeness.

Introduction: Corpus is a set of linguistic units that form a set of texts collected for a specific purpose, a collection of texts in natural language, written or spoken, stored electronically, placed on the basis of software in a computerized search system, working in an on-line or off-line system. Language corpora are an indispensable tool of modern linguistics for linguistic research and solving practical tasks. It differs from a simple electronic library. An electronic library aims to cover relatively complete artistic and journalistic works reflecting the socio-political, spiritual, and economic life of the people. From a linguistic point of view, electronic library texts are not processed on the basis of lexical, morphological, grammatical, and semantic signs, which creates inconvenience for linguistic research. Because an electronic library is not created to prepare a base of scientific research material, but rather to collect national and spiritual heritage. A language corpus, unlike an electronic library, involves collecting necessary, useful, and interesting texts for language study. The first factor that distinguishes a corpus from an electronic library is the nature of the text in it and its enrichment with additional information, and this feature constitutes a separate

part of the corpus - a commentary written on the corpus units.

DISCUSSION

If the user needs a word, a regular text editor will find it. However, working with a text that "understands" the grammatical meaning and structure of the word has many advantages and conveniences. If it is necessary to search for a word form, such software, that is, a corpus, can be of great help to the researcher or user. If a researcher spent months, sometimes years, finding examples for his work, copying them into a file (in the era before the development of computer technologies), today the world has the opportunity to find hundreds of examples in a matter of minutes with the help of language corpora and work on them. Special search consists of several programs designed to extract information from the corpus, presenting statistical information and search results to the user in a convenient form. In order to clearly imagine what processes are taking place in the language, it is advisable to further expand the scope of the corpus, using not only written speech, but also oral speech material. With the help of such a corpus, it is possible

to draw clear conclusions about the changes that have occurred and are expected as a result of language development.

A corpus search provides the user with the following results:

- 1) the appearance of a given word in all its forms in different contexts;
- 2) its place and variants in the lexical structure of the language;
- 3) a list of words that can be combined with the given word;
- 4) the frequency or statistics of the use of the same word by a particular writer;
- 5) the original and figurative meaning of the word;
- 6) the hidden model (possibility) of lexical use;
- 7) the state of use in different periods of language development.

The purpose of creating a national corpus. The large-scale nature of the corpus guarantees the uniqueness of the information, ensures the full representation of all aspects of linguistic phenomena. Various types of information are located in the language corpus in their natural contextual form, which becomes the basis for its comprehensive and objective study. A once-compiled and prepared mass of information can be used many times by several researchers for different purposes. The corpus is the most modern, widely available programmed system for language research (monitoring word changes in the language, historicisms, neologisms, expansion and narrowing of meaning; monitoring the emergence of new phraseologisms), language learning, and dictionary creation. It is difficult to imagine today's theoretical and practical philology without a corpus. In linguistic research, when dealing with facts, material must be collected and systematized. In carrying out such a large-scale work, the corpus is a time-saving and labor-saving tool for the linguist. It is not only a tool that accelerates the technical process, but also an information system of a certain modern form of language, a system that can answer unexpected questions and pose a pressing problem that has not been posed before in linguistics.

Type and nature of the material in the corpus. If we look at the proportion of texts in the existing corpora, we see that the contribution of fiction is 40%. This also includes memoirs, which are a genre whose linguistic features are between the artistic and journalistic style, and are more suitable for studying a living language. In the corpora of European languages, fiction material accounts for 20%. For example, although there are more than 20 studies devoted to the study of the linguistic features of modern writers, it is difficult to say

that they have yet fully studied this problem. Because the change in the linguistic features of a particular writer's work cannot yet be considered a new phenomenon in the language. The most convenient tool for observing and studying such processes is a language corpus. Therefore, the content of the text is of great importance. The texts included in the corpus may consist of texts taken from the works of a single author or several writers, texts covering a certain period (several decades or centuries), contemporary texts on a specific topic, and contemporary texts reflecting the current state of language and society.

CONCLUSION

In conclusion, the exploration of national corpora within the field of world corpus linguistics highlights their indispensable role in capturing the complexities of language as it is used in real-world contexts. As linguistic research continues to evolve, national corpora provide essential data that not only reflect the unique characteristics of a nation's language but also facilitate a deeper understanding of linguistic diversity and change. The integration of advanced technologies and collaborative efforts among researchers enhances the richness of these corpora, paving the way for innovative applications in education, computational linguistics, and sociolinguistics. However, the ongoing challenges of inclusivity and representation underscore the need for continued attention and adaptation in corpus compilation practices. Ultimately, national corpora stand as vital resources that enrich our understanding of language and its intricate relationship with culture and society, ensuring that linguistic research remains relevant and impactful in an increasingly interconnected world.

REFERENCES

- Louw, B. (1997). The role of corpora in critical literary appreciation. In A. Wichman, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora*, (pp. 240-251). Harlow: Longman.
- D. Z. Olimova, & M. D. Mahmudova. (2022). POLITICAL DISCOURSE AND TRANSLATION. RESEARCH AND EDUCATION, 1(3), 176–179. 2022
- Saidov Akmal Azimovich, Mahmudova Dildora Murodilloyevna. Anticipation strategy in simultaneous interpretation of political discourse. *Spanish journal of Innovation and integrity* Volume:12,110-116. November-2022.
- Mahlberg, M. (2007). A corpus stylistic perspective on Dickens' *Great Expectations*. In M. Lambrou and P. Stockwell (Eds.), *Contemporary Stylistics*, (pp. 19-31). London: Continuum.
- O'Halloran, K. A. (2007). The subconscious in James

Joyce's 'Eveline': a corpus stylistic analysis which chews on the 'Fish hook'. *Language and Literature*, 16(3), 227-244.

Mahmudova, D. (2023). CORPUS LINGUISTICS. B ACADEMIC RESEARCH IN MODERN SCIENCE (T. 2, Выпуск 23, cc. 104–106). Zenodo. <https://doi.org/10.5281/zenodo.10025001>

Mahmudova, D. (2023). CORPORA AND LITERATURE. Current approaches and new research in modern sciences, 2(10), 63-64. Zenodo. <https://doi.org/10.5281/zenodo.10013233>

Karimov Rustam Abdurasulovich, Mengliev Bakhtiyor Rajabovich (2019). The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation. *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, Volume-8, Issue-5S3 July 2019.