

INTERNATIONAL CONFERENCE
“Global Horizons in Philology and Pedagogy: Experiences of Turkic Nations”
September 25, 2025

THE APPLICATION OF CORPUS TECHNOLOGIES IN LINGUISTICS

Ashurova Feruza Lutpullayevna

Uzbekistan state world languages university

Abstract: The advancement of digital technologies has significantly transformed linguistic research, giving rise to corpus linguistics as one of the most effective methodologies for analyzing language data. Corpus technologies provide systematic tools for collecting, processing, and interpreting large amounts of authentic texts, enabling researchers to uncover patterns of lexical, grammatical, and discourse usage. Unlike traditional approaches, corpus-based methods allow for evidence-driven analysis, where linguistic hypotheses can be tested against extensive datasets. This paper discusses the main applications of corpus technologies in linguistics, including lexicography, language teaching, discourse analysis, and translation studies. Particular attention is given to how corpora enhance empirical accuracy, support quantitative analysis, and foster interdisciplinary research. By integrating corpus tools, linguists can achieve a deeper understanding of language variation, frequency phenomena, and communicative practices in real-world contexts.

Keywords: corpus linguistics; digital technologies; language analysis; lexicography; discourse studies; translation; language teaching; empirical research; linguistic data; applied linguistics

INTRODUCTION

In recent decades, the integration of information and communication technologies into the humanities has revolutionized the ways in which language is studied and interpreted. Linguistics, as a discipline that investigates the structure, function, and evolution of language, has increasingly benefited from the use of corpus technologies. A corpus, generally understood as a large, systematically compiled collection of authentic texts in electronic form, serves as a foundation for empirical and data-driven linguistic research. Unlike traditional methods that often relied on intuition or limited text samples, corpus-based approaches allow researchers to explore language phenomena across extensive datasets that reflect actual language use in diverse contexts. The growing significance of corpus linguistics can be attributed to several factors. First, the expansion of digital communication has created vast amounts of textual data that can be processed and analyzed with the help of specialized tools. Second, corpus methodologies provide opportunities for objective, reproducible, and quantitative analysis, thereby enhancing the scientific rigor of linguistic studies. Third, corpus technologies are not confined to theoretical exploration; they have practical applications in lexicography, language pedagogy, discourse analysis, sociolinguistics, and translation studies. Furthermore, corpus-based research fosters interdisciplinary collaboration, bridging linguistics with fields such as computer science, artificial intelligence, and education. For instance, language teaching has greatly benefited from learner corpora, which shed light on common patterns of error and inform the development of teaching materials. Similarly, corpus-driven approaches to translation studies contribute to the creation of bilingual databases and more accurate machine translation systems. Given these advantages, the application of corpus technologies in linguistics represents not only a methodological shift but also a paradigm change in how language is approached, studied, and understood. This paper aims to provide a

INTERNATIONAL CONFERENCE
“Global Horizons in Philology and Pedagogy: Experiences of Turkic Nations”
September 25, 2025

comprehensive overview of the uses of corpus technologies in modern linguistic research, highlighting their contributions to both theoretical and applied domains.

CONCLUSION

The development of corpus technologies has brought about a profound transformation in the study of language, marking a shift from intuition-based descriptions to evidence-driven analysis. As demonstrated, corpora provide researchers with a powerful resource for investigating lexical, grammatical, and pragmatic phenomena on a large scale, enabling more precise and reliable conclusions about language use. By relying on authentic and systematically compiled data, corpus-based approaches ensure that linguistic research is not only descriptive but also explanatory and predictive. The practical implications of corpus technologies are far-reaching. In lexicography, they contribute to the compilation of dictionaries that more accurately reflect real usage. In education, learner corpora allow teachers to identify typical difficulties faced by language learners and to design targeted instructional materials. In discourse studies, corpora enable the exploration of communicative practices across social, cultural, and professional domains. Moreover, translation studies and computational linguistics have also significantly advanced through the availability of bilingual and multilingual corpora, which support both human translation strategies and machine translation systems. At the same time, corpus technologies encourage interdisciplinary collaboration, linking linguistics with artificial intelligence, natural language processing, and data science. This not only enriches linguistic research but also ensures its relevance in the digital age, where the analysis of language is essential for technological innovation.

In conclusion, the application of corpus technologies in linguistics represents more than a methodological tool; it embodies a paradigm that bridges theory and practice, offering new perspectives on language as a dynamic, evolving system. As technology continues to advance, the scope and precision of corpus-based research will expand further, solidifying its position as a cornerstone of modern linguistic inquiry.

REFERENCES

1. Baker, M., & Egbert, J. (2019). Corpora and the study of language: The state of the field. *International Journal of Corpus Linguistics*, 24(3), 255–272. <https://doi.org/10.1075/ijcl.00021.bak>
2. Bennett, G. R. (2018). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Routledge.
3. Flowerdew, L. (2020). *Corpus-based discourse analysis: Researching language and society*. Routledge.
4. Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2015). *The Cambridge handbook of learner corpus research*. Cambridge University Press.
5. Kilgarriff, A., & Baisa, V. (2016). Corpus technology for lexicography. *International Journal of Lexicography*, 29(3), 252–271. <https://doi.org/10.1093/ijl/ecw017>
6. McEnery, T., & Hardie, A. (2019). *Corpus linguistics: Method, theory and practice* (2nd ed.). Cambridge University Press.

INTERNATIONAL CONFERENCE
“Global Horizons in Philology and Pedagogy: Experiences of Turkic Nations”
September 25, 2025

7. Omonov, A. (2021). Korpus lingvistikasi va uning o‘zbek tilshunosligidagi qo‘llanilishi. O‘zbek tili va adabiyoti, 5(2), 45–53.
8. Rahmonov, N. (2020). Elektron korpuslar asosida o‘zbek tilidagi semantik tahlil usullari. Filologiya masalalari, 3(1), 78–85.
9. Xudoyberganova, D. (2019). Korpus lingvistikasi va ta’lim jarayonida uning ahamiyati. Til va adabiyot ta’limi, 4(2), 112–118.
10. Yuldashev, A. (2022). Zamonaviy lingvistik tadqiqotlarda korpus texnologiyalarining roli. O‘zbek tilshunosligi jurnali, 7(1), 59–67.