

SCIENTIFIC RESEARCHES IN CORPUS LINGUISTICS

Inna Antonenko

PhD, Associate Professor,

Nataliia Chizhova

Lecturer,

National Technical University of Ukraine ‘Igor Sikorsky Kyiv Polytechnic Institute’

Modern corpus linguistics has gained new status in science, it represents real changes in language and the study of linguistic problems has become important in the study of language for professional and scientific purposes. Researchers emphasize the role of the development of corpus linguistics, where they study large collections of texts and the role of words in different uses (Biber et al., 2006). Corpus linguistics studies language in use through corpora which are large parts of text containing different words. This studies based on scientific research of a corpus which is a collection of texts stored in an electronic database.

Scientists consider the process of design and construction of the corpus, which must be resolved the compilation of it and also they study the issues about size of the corpus, its representativeness and balance. They introduce corpus linguistics and its disciplinary status, discuss its impact on university English teaching, and offer suggestions for the application of corpus linguistics to solve the problems associated with the traditional form of teaching English. Based on this, the articles delve into an in-depth discussion of corpus linguistics and English teaching (Li, 2024). John Sinclair, one of the most influential researchers in modern corpus linguistics, noted that a word has no meaning by itself, but is often conveyed through words in a sequence (Sinclair, 1991), and this opinion is fundamental in corpus linguistics.

The application of corpus linguistics has pointed out a new direction for English teaching and has provided new development opportunities for the discipline in education. In addition, analysis results can be generated in easy to read reports, allowing researchers to work interactively and share results among each other. Perspectives of study point to the widespread use of texts as appropriate resources on the basis of which specific linguistic problems and phenomena are studied. The study aims to research issues related to the lexical or grammatical features of specific patterns, as well as how to compare patterns which differ in types and registers.

Technology plays a dominant role in corpus linguistics today, as advances in computer technology have increased the scope of research as well as provided new advantages over previous studies. The software helps to find variants of the word and first create and compile a word list, then save and organize written large texts for linguistic analysis. Among the features of computer analysis of corpora, linguists note the empirical nature of corpus analysis, the study of real structures in texts, which are

the basis for analysis. There is a close connection between the current state of technologies which have brought the field statistical reliability of results and the ability to use a variety of text databases. The program allows researchers to create a corpus, store texts, and use them for corpus analysis. It includes three key computational techniques for linguistic analysis: text comparison, text categorization and data reporting. The research focuses on online language resources, new technologies and specialised translation (Dejica, 2016). The program allows the researchers to design different corpora at the same time, which can be further structured into subcorpora, and they can also organize them into separate groups.

When aiming to study specific complex text patterns, it is important to remember that computer analysis methods can be applied to different types of corpora. It also briefly touches on the design of the corpus and describes in detail the processes of finding texts and their subsequent transformation, recognition and possible correction. The development of corpus linguistics has been motivated by the purpose of linguists to understand:

- what are the most common words that are usually used in certain contexts;
- what is common and what is unusual for certain language variations;
- how words are used in language.

Data analysis tools are used in the field of research, leading to the introduction of methods that rely on qualitative and quantitative methods, and research has provided methods of reliable analysis, where it is important to organize large texts on computers for linguistic analysis that is no longer limited to sentences. Applied to the study of language, corpus-based analysis aims to investigate specific linguistic structures, how they appear in different contexts, and functions they acquire.

Researches try to achieve success by including different methods in the corpus of speech occurrences, which should be represented in proportions that reflect their distribution in the speech community. The scientists consider that appropriate rules should be established to create reliable corpus. They conclude that different approaches can lead to improvement of language theory.

References

Biber, D., Conrad S., & Reppen, R. (1998). *Corpus Linguistics: investigating language structure and use*. Cambridge. Cambridge University Press.

Dejica, D., Hansen, G., Sandrini, P., Para, I. (2016). *Language in the Digital Era. Challenges and Perspectives*. <https://www.degruyter.com/document/doi/10.1515/9783110472059/html>

Li, H. (2024). *Research on Corpus Linguistics and English Language Teaching*. *Education Reform and Development*. 6(3), 30-35.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press. https://www.academia.edu/16757872/SInclair_Corpus_Concordance_Collocat_ion