



CORPUS LINGUISTICS

Mahmudova Dildora Murodilloyevna

BSMI, the teacher of English language department

E-mail: dildoramahmudova547@gmail.com

<https://doi.org/10.5281/zenodo.10025001>

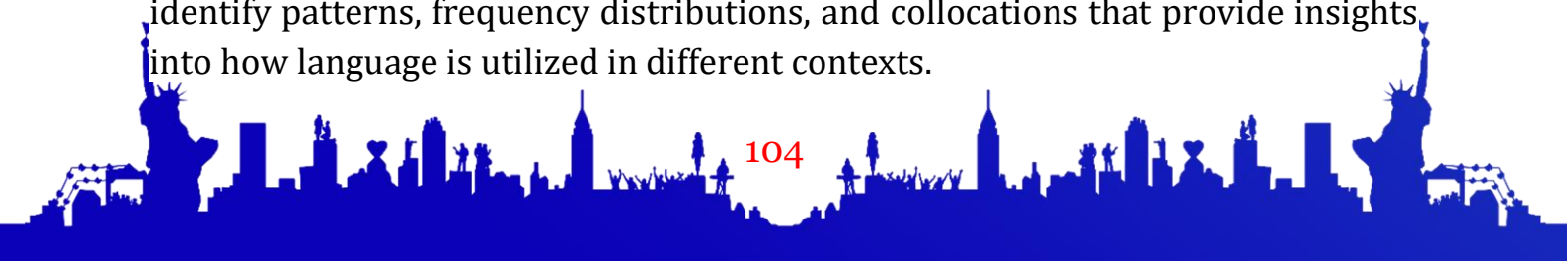
Abstract. Corpus linguistics is the study of a language as that language is expressed in its text corpus (plural corpora), its body of "real world" text. Corpus linguistics proposes that a reliable analysis of a language is more feasible with corpora collected in the field—the natural context ("realia") of that language—with minimal experimental interference. This paper attempts to analyze the relationship of corpus and linguistics.

Keywords: corpus, linguistics, corpora

Introduction. The text-corpus method utilizes a wide range of texts written in any natural language to derive a set of abstract rules that govern that particular language. These derived rules can then be used to explore the relationships between the subject language and other languages that have undergone similar analyses. Initially, creating corpora involved manually extracting information from source texts, but with advancements in technology, this process has become automated.

Corpora have been utilized not only for linguistic research but also for compiling dictionaries, such as The American Heritage Dictionary of the English Language in 1969, and grammar guides like A Comprehensive Grammar of the English Language, published in 1985. Experts in the field have differing views about the annotation of a corpus. These views range from John McHardy Sinclair, who advocates minimal annotation so texts speak for themselves, to the Survey of English Usage team (University College, London), who advocate annotation as allowing greater linguistic understanding through rigorous recording. The purpose of this survey was to conduct a thorough analysis of the English language, with a specific focus on examining its usage and patterns in both written and spoken texts. The work of Quirk served as the foundation for corpus linguistics, which is a methodological approach used to study language.

The text-corpus method has a wide range of applications in linguistic research. It can be utilized to investigate language variation and change, as well as to analyze specific linguistic phenomena such as syntax, semantics, and pragmatics. By examining a large collection of texts, researchers are able to identify patterns, frequency distributions, and collocations that provide insights into how language is utilized in different contexts.



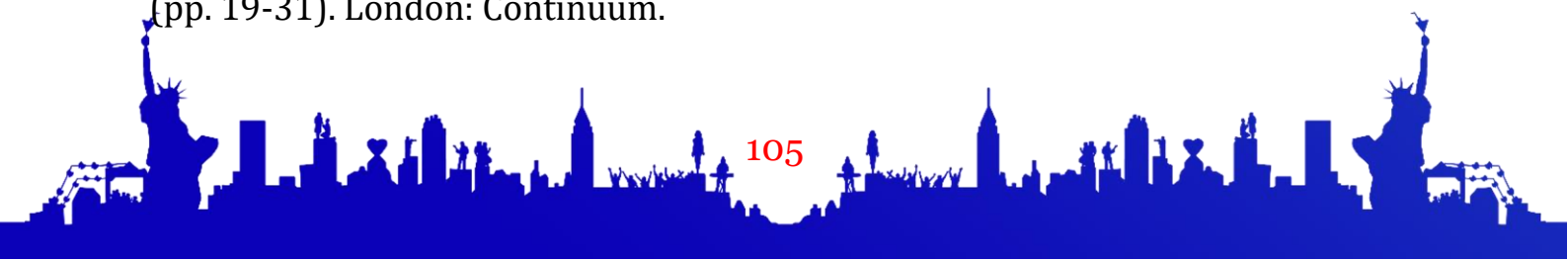


Furthermore, corpus linguistics has played a crucial role in the advancement of natural language processing (NLP) technologies. By training computer algorithms on extensive corpora, researchers have been able to develop language models that possess the ability to understand and generate text that resembles human language. This has led to significant progress in areas such as machine translation, information retrieval, and other NLP applications.

To summarize, the text-corpus method is an immensely powerful tool in linguistic research that enables researchers to gain valuable insights into the rules and patterns of a language by analyzing a vast collection of texts. Its applications range from studying language variation and change to the development of NLP technologies. However, it is important to note that it is not directly related to the specific phrase "run into the sand." One notable milestone in the field of corpus linguistics was the publication of "Computational Analysis of Present-Day American English" in 1967 by Kučera and Francis. This work was based on an analysis of the Brown Corpus, which was a carefully selected compilation of approximately one million American English words from various sources. Kučera and Francis utilized computational analyses along with elements from linguistics, language teaching, psychology, statistics, and sociology to create a comprehensive and diverse body of work. Another significant publication was Quirk's "Towards a description of English Usage" in 1960, which introduced the Survey of English Usage.

References:

1. 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.
2. D. Z. Olimova, & M. D. Mahmudova. (2022). POLITICAL DISCOURSE AND TRANSLATION. RESEARCH AND EDUCATION, 1(3), 176–179. 2022 <http://researchedu.org/index.php/rae/article/view/1263>
3. 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.
4. 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.





5. 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.
6. Kudratova, S. . (2023). 1984: THE MASTERPIECE THAT KILLED GEORGE ORWELL (CHAPTER I). *Евразийский журнал академических исследований*, 3(9), 119–124. извлечено от <https://www.in-academy.uz/index.php/ejar/article/view/20606>
7. Kudratova Sitara Rashidovna. (2023). PRAGMATIC DISCORDANCE OF ENGLISH AND UZBEK PHRASES IN THE EVENT OF SPEECH ACT. MODELS AND METHODS IN MODERN SCIENCE, 2(10), 24–28. <https://doi.org/10.5281/zenodo.8368922>
8. 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
Rustam Abdurasulovich Karimov A senior Lecturer of Department of Foreign Languages In Natural Sciences Foreign Languages Faculty Bukhara State University
9. ELECTRONIC JOURNAL OF ACTUAL PROBLEMS OF MODERN SCIENCE, EDUCATION AND TRAINING. OCTOBER, 2020-V. ISSN 2181-9750 UDC: 158.9(565.2)
ON EXTRALINGUISTIC TIGHTENING OF UZBEK-ENGLISH TEXTS

