

THE EVOLUTION OF CORPORA

M. Ch. Fayzieva

Teacher of English Linguistics Department,
Foreign Languages Faculty,
Bukhara State University, Bukhara
e-mail: m.ch.fayziyeva@buxdu.uz

G. Turobova

Student of the group 11-7 Eng. 20,
Foreign Languages Faculty,
Bukhara State University, Bukhara
e-mail: gulshanturobova@gmail.com

ABSTRACT

The article reveals common corpus secrets and types. It gives some data about COCA, common corpus, sublanguage corpora, and the Oxford English Corpus. One of the most rapidly expanding approaches in modern linguistics is corpus linguistics. In the field of applied linguistics, large collections of written and spoken texts that have been specifically chosen to reflect certain linguistic contexts, such as casual conversation or academic writing, are analyzed.

Keywords: *corpus linguistics, O'Keefe, McCarthy, COCA, common corpus, sublanguage corpora, Oxford English Corpus.*

INTRODUCTION

Initially, it is noteworthy to mention that in spite of the corpus linguistics' current prominence, linguists still don't totally realize what it is and why it is necessary. Both written and spoken variants of the word corpus are derived from the Latin word for body. Textual information in one or more languages can be found in a corpus. For the past 30 years, many linguists have been working on Corpus. Corpus linguistics started to develop quickly in the 1980s. But for many years, Chomsky's influence led linguistics to focus on "potential language," which is linguistics' own internal language but not actual. Many linguists have studied corpus linguistics over the years. For instance, Köding produced 11 million German word complexes in 1897 and examined them using 5,000 word processors concentrating on spelling. Preyer's experiments in 1898, which were based on corpus linguistics acquisition, depended on parents' diaries for first language learning.

METHODS

As some time passes, linguists add new knowledge to corpus linguistics. Corpus linguistics greatly benefited from the invention of the first personal computer in 1976. Word sorting may now be done quickly by researchers. It is critical to state what is and is not related to a corpus while learning corpus linguistics for the first time. One is not able to offer contradictory proof, explain why, or present all potential lengths at once. Corpus linguistics doesn't offer contradictory evidence. For example, it doesn't state what is feasible or impractical, right, or wrong, in a language. It merely indicates whether or not it is included in the corpus. Corpus linguistics doesn't define anything; it simply demonstrates what something is without providing any justification. Language has a close relationship with humans. We utilize language on a daily basis, both verbally and in writing. From a frequency list of 10 million words created by O'Keefe, McCarthy, and Carter, the Corpus was able to extract 80 terms out of a total of 2,000 words. Moreover, Corpus Linguistics offers files on word repetition, pronunciation, and other related topics.

RESULTS

It is vital to note that the Corpus of Contemporary American (COCA) English includes 195,000 spoken and 198,000 written instances. In contrast to COCA, the oral portion is greater than the written portion and appears far more frequently in oral (2,543 times), written (673 times), and overall (3,216 times). One could utilize a few additional words as an illustration of them. For instance, the word "really speaking" has a value of 1.637, while "written" has a value of 150, and "very" has a value of 267, for a total of 417. They demonstrate how corpus linguistics is applied in various contexts today, contributing to the effectiveness and expressiveness of language instruction.

ANALYSIS

There are various types of corpora, such as spoken or written language, contemporary or classical literature, monolingual, multilingual, etc. Books, newspapers, periodicals, speeches are considered as examples of texts. The types of texts that are contained and how these various forms of texts are combined make them different from one another. For instance, a "common corpus" is a collection of unconnected works that fall under the same textual genre, topic, or register. The British National Corpus is one illustration of this. Moreover, some corpora include terms from a range of particular languages, such as a dialect or a certain subject. These kinds of corpora are commonly referred to as "Sublanguage Corpora." The given text is the same in all languages when translated into the Parallel Corpus.

DISCUSSIONS

These days, corpus linguistics has spread throughout the globe, and several nations have produced well-known corpora. For instance, the Oxford English Corpus (OEC) is a corpus of over 2.1 million English words that is only accessible to the Oxford University Press's language research program and those who created the Oxford English dictionary. It includes languages from the UK, USA, Australia, New Zealand, the Caribbean, Canada, and more. It is one of the biggest and most well-known corpora. Nonetheless, researchers at Oxford University Press can normally access this corpus. Other researchers with a significant interest are welcome to apply as well. The vocabulary base had 1 billion words as of April 27, 2006.

In a nutshell, corpus linguistics is seen as a language resource made up of significant amounts of texts. Corpus linguistics is highly helpful for statistically analyzing inside a particular language or a specific area of a language, as well as testing theories.

REFERENCES:

1. Ataboyev, N. (2019). THE RESULTS OF THE CORPUS ANALYSIS AS A QUANTITATIVE REPRESENTATION OF LINGUOCULTURAL CONCEPTS. *Philology Matters*, 2019(2), 97-103.
2. Атабоев, Н. (2020). ИНГЛИЗ ТИЛИ КОРПУСЛАРИДА КОЛЛОКАЦИОН БИРЛИКЛАРНИНГ ЛИНГВОМАДАНИЙ ТАДҚИҚИ. *ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz)*, 1(1).
3. Bobojon o'g'li, N. A. Corpus-Based Research On The Language Features Of Corpus Linguistics: In The Example Of ECOCL. *Language*, 3(2139), 950.
4. Choshovna, F. M. (2019). Forms of integrated-skill instruction. *Проблемы педагогики*, (4 (43)), 31-32.
5. Fayzieva, M. (2021). STUDENTS' KNOWLEDGE OBSERVATION BY THE HELP OF INFORMATIVE PROGRAM TOOLS. *ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz)*, 8(8).
6. Файзиева, М. Ч., & Туробова, Г. Ю. (2022). ВАЖНОСТЬ ИНКЛЮЗИВНОГО ОБРАЗОВАНИЯ В УЗБЕКИСТАНЕ. *INNOVATION IN THE MODERN EDUCATION SYSTEM*, 3(25), 49-52.
7. Fayzieva, M. (2021). THE IMPACT OF BILINGUALISM AND MULTILINGUALISM IN FLT. *ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu.uz)*, 8(8).
8. O'G'Li, A. N. B. (2019). Structure of 'Coca'(corpus of Contemporary American English) and simple queries on it. *Наука, образование и культура*, (10 (44)), 32-34.
9. <https://studylib.net/doc/7308485/1.-history-of-corpus-linguistics>
10. <https://libguides.bodleian.ox.ac.uk/english-language/Corpora>
11. https://en.m.wikipedia.org/wiki/Oxford_English_Corpus