

THE IMPORTANCE OF CORPUS LINGUISTICS

Bukhara State University Foreign language and literature (English language) faculty

Teacher: Raximov Mubin

The students of 11-21 chi 20 group

Alixanova Zilola

Fayzullayeva Sevara

Fatillayeva Laylo

Abstract: *The types and significance of corpus linguistics are discussed in this article. Corpus linguistics is among the current linguistics techniques that is growing the fastest. Huge sets of spoken and written texts that have been hand-picked to accurately represent certain linguistic situations,¹² such informal conversation or academic writing, are examined in the area of applied linguistics.*

Key words: *corpus, monolingual, parallel corpus, comparable, diachronic and synchronic corpus, specialized, multimedia.*

INTRODUCTION

In the 1960s, the science of corpus linguistics emerged as a branch of computational linguistics. The creation of theoretical frameworks is being done through the use of computer-generated text corpora. The phrase "linguistic (lexical) corpus" describes a large electronic database and a set of tagged electronic texts that are particularly selected to address linguistic issues.¹³ Corpus linguistics is the study, analysis, and investigation of language via the use of computer technology.

METHODS

An enormous collection of text (usually billions of words) produced by real language users is called a text corpus, and it is used to investigate how words, phrases, and language are used generally.¹⁴ It is used in a wide range of fields, including as natural language processing, social science, linguistics, lexicography, and the humanities. Additionally, a corpus may be used to generate a variety of language databases that are utilized in the development of software, such as text-to-speech modules, text-to-speech comprehension systems, spell check, grammar check, predictive keyboards, machine translation, and many other applications. Additionally very helpful is corpus linguistics for scientific study in many different areas of linguistics. The field of corpus linguistics is one that is currently expanding quickly. It is developed by specialists in corpus linguistics for use in academic study and language instruction.

¹² Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef." (2018). Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016

¹³ Rakhimov, M. (2024). KОРPUS LINGVISTIKASI TARAQQIYOTI VA O'ZBEK TILSHUNOSLIGIDA KОРPUS TAHLIL ASOSLARI. ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu. uz), 45(45).

¹⁴ Jon McHardy Sinclair, Tillami o'qitishda korporatsiyani qanday ishlatish kerak, Jon Benjaminlar, 2004-y.

Learning a foreign language and one's mother tongue both benefit from educational corpora.¹⁵ The corpus is useful in language instruction because it provides a variety of examples that illustrate the breadth of vocabulary and clarify how a word could be used through various grammatical constructions. The ability to represent this ongoing updating of examples—which is crucial for language learning—can only be provided by a corpus.¹⁶

RESULTS

It is important to remember that the corpus cannot be properly categorized into a single category. Alternatively, corpora could include features or qualities that could be applied to group them. The same corpus may have one or more of these traits. A monolingual corpus is the most common kind of corpus. It contains literature in a single language.¹⁷ A wide range of users use the corpus, which is typically annotated for parts of speech, for a variety of tasks ranging from highly practical ones like confirming the correct usage of a word or finding the most natural word combinations to scientific ones like spotting frequent patterns or emerging trends in language. There is a parallel corpus with two or more monolingual corpora. For example, the corpora are thought to influence each other's translations. Take fiction and translation.

ANALYSIS

It is crucial to note that Orzubek Anarbayev adheres to the parallel corpus and presents an idea: "In the modern period, when multicultural contact is common, parallel corpus is an essential fact. Real and lacunar units can be distinguished by means of parallel corpora in various linguistic contexts, cultural universals, and certain mental properties of languages.¹⁸ Computer lexicography facilitates the development of automated translation by providing a corpus of parallel texts. A corpus of parallel texts will be used to construct concordant programs, and other specialized dictionaries can be produced." It is necessary to match analogous elements, which are frequently sentences or paragraphs, in order to align the two languages. One corpus among two or more monolingual corpora that were all created using the same techniques and are typically each in a different language is referred to as a similar corpus.

DISCUSSION

It is feasible to compare the results between the corpora even though they are not translations of one another due to the comparable content. A diachronic corpus is a collection of writings spanning many historical periods that is utilized to study how language has evolved or changed over time.¹⁹ Comparatively, a synchronic corpus is made up of works that were authored simultaneously. It captures the vernacular of that particular moment. Texts that are limited to one or more domains, subjects, subject

¹⁵Sobirova, S. (2023). KORPUS LINGVISTIKA MASALALARI. Бюллетень студентов нового Узбекистана, 1(6 Part 2), 26-29.

¹⁶Muradovich, R. M. (2021, December). Linguotherapy as the method of treatment. In Conferences.

¹⁷<https://cyberleninka.ru/article/n/korpus-lingvistikasi>

¹⁸file:///C:/Users/user/Downloads/korpus-lingvistikasi-va-parallel-korpuslar-tavsifi%20(1).pdf

¹⁹<https://uz.eferrit.com/>

areas, etc. make up a specialized corpus. This corpus is used to look at how the specialized language is used. A collection of texts with extra audio, video, or multimedia components is called a multimedia corpus. Photographs of the original manuscript, printed version, or movies with the corpus text read aloud may be found in other corpora.

It is clear from the conclusion that the corpus's mission is enormous and unachievable. Corpus has evolved into a time- and labor-saving tool for workers today. It takes a lot of work and money to update book-form dictionaries with changes in language throughout time.²⁰ Corpus offers the chance to quickly and cost-free fix difficulties.

REFERENCES:

1. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018). Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016
2. Rakhimov, M. (2024). KORPUS LINGVISTIKASI TARAQQIYOTI VA O'ZBEK TILSHUNOSLIGIDA KORPUS TAHLIL ASOSLARI. ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu. uz), 45(45).
3. Jon McHardy Sinclair, Tillarni o'qitishda korporatsiyani qanday ishlatish kerak, Jon Benjaminlar, 2004-y.
4. Sobirova, S. (2023). KORPUS LINGVISTIKA MASALALARI. Бюллетень студентов нового Узбекистана, 1(6 Part 2), 26-29.
5. Muradovich, R. M. (2021, December). Linguotherapy as the method of treatment. In Conferences.
6. <https://cyberleninka.ru/article/n/korpus-lingvistikasi>
7. [file:///C:/Users/user/Downloads/korpus-lingvistikasi-va-parallel-korpuslar-tavsifi%20\(1\).pdf](file:///C:/Users/user/Downloads/korpus-lingvistikasi-va-parallel-korpuslar-tavsifi%20(1).pdf)
8. <https://uz.eferrit.com/>
9. <https://www.sketchengine.eu/corpora-and-languages/corpus-types/>

²⁰ <https://www.sketchengine.eu/corpora-and-languages/corpus-types/>