

CORPUS LINGUISTICS AS A MODERN METHOD IN THE STUDY OF FOREIGN LANGUAGES

Umarova Barchinoy Botirovna – student, Bukhara state university

**Rasulova Nigora Nematovna – Senior lecturer, Bukhara Institute of
Engineering and technology**

Abstract

Currently, corpus linguistics is widely used to study foreign languages. Interest in this topic is constantly reflected not only in various books and textbooks, but also in reports at conferences devoted to methods of teaching foreign languages. One example is modern dictionaries, which are formed on the basis of constantly updated databases - linguistic corpora. The article discusses the potential of including corpus linguistics materials in foreign language teaching methods. It provides the analyzed examples of using the information base of corpus Internet technologies in teaching foreign languages.

Key words: Internet technologies, corpus linguistics, electronic corpus, teaching a foreign language.

We live in an age of the widespread formation of an information society. The process of informatization is significantly changing and making our lives easier. Information has become the most important resource in modern society, and its preservation, multiplication, and rational use have become essential to society. More and more of the information we use in everyday life is presented in electronic format. Traditional printed publications, such as works of literature that have become classics and priceless museum manuscripts, are being converted to an electronic format. This format allows us to organize information and access it in a completely new and more efficient way.

Internet technologies include a variety of information resources, such as websites, online forums, libraries, and electronic encyclopedias. These technologies are now an integral part of our daily lives, and their influence can be seen in all aspects of society, including education. As a result, teachers are increasingly turning

to the internet for new texts, audio-visual materials, and relevant vocabulary to enhance their teaching methods.

Undoubtedly, one of the most effective approaches to teaching a foreign language is through the use of corpus linguistics. The success of interlingual communication directly depends on the proficiency in correctly using the units of a foreign language in accordance with the cultural norms of the target language. The use of a national corpus in teaching allows not only to effectively teach students how to search for, study, compare and analyze linguistic features of language independently in order to improve their communication skills, but also contributes to successful mastery of new vocabulary and grammar. This is achieved by providing examples from different contexts in real-life language sources, which helps students to better understand and apply new material.

The concept of “corpus of texts” is increasingly integrated into the scientific discourse of linguists, and this undoubtedly enhances the quality of their work. However, corpus linguistics has not yet received adequate attention from educators. Therefore, the significance of this study is evident. Corpus linguistics can be defined as a branch of linguistics that focuses on the creation and annotation of textual corpora, the development of search tools for these databases, and research based on corpora [2]. There are various definitions of a corpus, but most scholars agree that a corpus is a collection of machine-readable, authentic texts that are collected to characterize or represent a language’s state or diversity [5].

The first electronic corpus was developed by Nelson Francis and Henry Coachman at Brown University in the early 1960s. This was a corpus of the American English language, consisting of 1 million words (the Brown Corpus). The prototypes consisted of dictionary card files, which were collections of text fragments containing a specific word, organized alphabetically according to that word. At present, most major languages in the world have their own national corpora. These corpora are collections of specially selected texts that are marked up according to various linguistic criteria and are equipped with a search system. Corpora allow users to work with large volumes of text simultaneously, extracting relevant information from them. By analyzing the contexts extracted from the

corpora, researchers can draw conclusions about linguistic phenomena that are of interest, such as the behavior of a particular language feature. With regard to grammatical structure and the use of expressive language, etc.,

In corpus linguistics, corpus types can be categorized based on several criteria. These criteria include the language of the text (for example, English or Chinese), the access mode (such as open, closed, or commercial), and the genre of the source material (which could be artistic, documentary, academic, or journalistic).

Advanced technologies help to replenish the database not only with written, but also with oral texts. The generation of materials representing oral speech is carried out in an interesting way: during the day, a random volunteer communicates with people around him who cannot know that all conversations in which he participates are being recorded. The received audio recordings are placed in a database and accompanied by a printed text in the form of a transcript [5]. This is how the markup of oral everyday speech is created in the corpus.

To work with the corpus, it is enough for the user to enter a word or a combination of words in the information database search bar.

An expanded version is available in many national corpora, which allows you to find textual information according to almost any linguistic criteria: belonging to a certain group of parts of speech; grammatical features; stylistic and emotional coloring.

From the point of view of pedagogy, the use of corpora in the curriculum will help to form the student's skills of independent work on collecting and analyzing data from a living language. Such tasks allow the student to be in the role of a researcher who creates something unique, thereby increasing his motivation.

The use of a set of texts of the national corpus makes it possible to observe various cases of the use of linguistic units and their behavior in the natural language environment. Such contexts have an advantage over artificially created textbook authors based on intuitive guesswork. In addition, national corpora allow for case studies based on statistical data. Thus, students can formulate, and then confirm or refute hypotheses about a particular linguistic phenomenon. At the same time, the

researcher does not have to go through a long and painstaking stage of collecting material (interviewing, working with dictionaries, written texts, etc.).

Some words or phrases can also help to expand students' background knowledge. For instance, the phrase "true blue" may lead students to consider historical events related to the conflict between the Conservative and Labor parties (referred to as "blue" and "red"), and identify groups of words and stable phrases characteristic of election campaigns and political propaganda. Through such activities, under the overall guidance of the teacher, students can progress further, conducting new research and forming their own educational pathways. Academic buildings can be defined as electronic repositories of authentic materials created by students learning a foreign or additional language. These buildings are a relatively new development. Early academic building projects differed significantly from modern ones, as they primarily served as repositories for erroneous materials, and were smaller in terms of both the volume and variety of texts. The criteria for designing modern academic buildings are much more stringent, making them more suitable for the analysis of various languages. This is essential for the development of effective learning plans and the further refinement of language learning methodologies.

The main use of corpora in the classroom nowadays is to help students learn the vocabulary of a language they are studying. This involves not only learning new words, but also discovering new meanings for words that they have already studied. Corpus-based Internet technologies can be used for more than just teaching vocabulary. They can also be used to teach grammar. For example, by searching for the Chinese National Corpus, teachers can show students the differences between grammatical structures such as 被 and 把. This gives students access to a vast database of examples beyond those found in textbooks, allowing them to more accurately understand the differences based on real-life contexts.

Parallel corpora will play an irreplaceable role in language teaching and translation. A parallel corpus is a collection of texts in two or more languages that are related in some way, such as translations of the same source text into different

languages. Words in a parallel corpus are often translated using synonyms (e.g., "beautiful," "amazing," "brilliant"), which can lead to the use of these words in unusual contexts. By analyzing examples from a parallel corpus, we can clearly see the differences between these synonyms and understand their meanings better.

Parallel cases are the most effective way to demonstrate, for example, correct use of prepositions. Italian textbooks, for instance, do not clearly distinguish between the prepositions "a" and "in" with regard to the direction of an action, and cases where they should be used with certain words must be memorized separately. By using the national corpus, the process of immersing oneself in the linguistic environment and memorizing such cases can be accelerated. Linguists acknowledge the importance of intuition in language learning, but knowledge based on independent analysis and intuition differs significantly from that based on explicit evidence provided by the body.

The learner corpus is a promising new approach to the application of corpora in foreign language learning. S. Granger, a pioneer in this field, defines a student corpus as an electronic database of authentic texts written by students learning a language as their second or foreign language [6,7]. The International Corpus of Learner English, created by the Center for Corpus Linguistics in 2009, contains 2 million words of essays written by students from 16 different countries. This corpus consists of 19 subsets, each representing a specific language pair (English and French, English and Russian, etc.). Therefore, the student corpus is an essential tool for studying common student errors [8]. Various studies have demonstrated that corpora are a useful resource for learning a foreign language, both as a reference and as the primary material for students. Increased motivation among students to learn a language through independent use of these language corpora has also been observed.

Thus, the use of linguistic corpora in teaching a foreign language is an innovative and advanced approach to working in the classroom. At present, the use of corpora is no longer limited to a small group of linguists, as it has become an essential tool for enhancing the quality of the learning process. Working with corpora undoubtedly improves the quality of education, increasing the interest and engagement of both students and educators. It is reasonable to believe that corpus

linguistics will continue to evolve and, in the near future, will influence every aspect of language teaching, learning, and research.

References

1. Arto M. The role of corpora in linguistic research and language teaching /Mustajoki Arto // National Corpus of the Russian language and problems of humanitarian education.: collection of materials of the International Scientific Conference (Moscow, April 19-20 2007). – M., 2007. – pp. 58-60.
2. Baranov A.N. Introduction to applied linguistics / A.N. Baranov. – M.: Editorial URSS, 2001. – 360 p.
3. Nagel O.V. Corpus linguistics and its use in computerized language teaching / O.V. Nagel. – URL: [https://cyberleninka.ru/article/n/korpusnaya lingvistika-i-ee-ispolzovanie-v-kompyuterizirovannom-yazykovom-obuchenii/](https://cyberleninka.ru/article/n/korpusnaya_lingvistika-i-ee-ispolzovanie-v-kompyuterizirovannom-yazykovom-obuchenii/) (date of issue: 03/31/2020).
4. Sysoev P.V. Linguistic corpus in the methodology of teaching foreign languages /P.V. Sysoev. – URL: <https://cyberleninka.ru/article/n/lingvisticheskiy-korpus-v-metodike-obuchenia-inostrannym-yazykam> (accessed 03/31/2020).
5. Grigaliuniene J. Corpora in the classroom / J. Grigaliuniene. – Vilnius university, 2013. –P. 10-15.
6. Granger, S., Dagneaux, E., Meunier, F., Paquot, M., 2009. International Corpus of Learner English v2. Handbook + CD-Rom. Presses universitaires de Louvain, Louvain-la-Neuve.
7. Granger, S., 2003. The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. TESOL Quarterly, 37 (3), pp: 538–546.
8. Leńko-Szymańska, A., 2014. Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. ReCALL, available on CJO2014. Doi: 10.1017/S095834401400010X.
9. Raximov, M., & Kholova, Z. (2023). INTERCULTURAL COMPETENCE IN FOREIGN LANGUAGE CLASSROOM. Центральноеазиатский журнал



образования и инноваций, 2(11 Part 2), 150–156. – URL: <https://in-academy.uz/index.php/cajei/article/view/23396>.

10. Raximov , M. ., & Akramova , S. . (2023). THE FEATURE OF RENAISSANCE DRAMATURGY OF THE ENGLISH LITERATURE. Центральноеазиатский журнал образования и инноваций, 2(11 Part 2), 145–149. – URL: <https://www.in-academy.uz/index.php/cajei/article/view/23395>

11. Жумаева З. К., Расулова Н. Н. Инновационный путь развития экономики Узбекистана //Теория и практика современной науки. – 2019. – №. 5. – С. 224-226.