

Corpus Linguistics as a Branch of Applied Linguistics

Xonqulova Nilufar Ravshanovna

Teacher, Applied Translation Department, Translation Faculty, Uzbekistan State World Languages University

ABSTRACT

The article deals with the basic concepts of corpus linguistics. General characteristics of the national corpus of the English language. The study of lexical categories of polysemy and homonymy, synonymy and antonymy. The use of corpus data in the study of syntax and morphology.

KEYWORDS: *corpus linguistics, applied linguistics, information resources, information potential, language units*

We live in an era of widespread formation of the information society. Informatization processes significantly change and make our life easier. The most important resource in modern society is information. The preservation, multiplication and rational use of this resource has become of great importance for society.

More and more information that we use in everyday life comes in electronic format, moreover, traditional printed publications, works of literature that have become classics, priceless museum manuscripts are being transferred into an information format. This form of information presentation allows you to organize work and access to it at a completely different qualitative level.

Ensuring public access of users to electronic information resources has become one of the priorities of information services for science, education and culture. The current information service system allows us to access virtually any information in unlimited quantities. A huge role in this is played by the Internet system, which since 1991 has managed to turn into a distinctive and practically inexhaustible information environment.

With the advent of the Internet, a fundamentally new situation has been created in the education system. The potential of the network for education and personal development is truly unique. The information potential of the Internet is simply inexhaustible. The use of Internet resources allows diversifying the educational process, as well as organizing it taking into account the interests of students, their level of preparation.

The 21st century requires fundamentally new approaches to education. Education should be developing in terms of developing independent critical and creative thinking. However, for this it is not enough to have in the education system only a teacher and a textbook (understood in a broad sense as a system of teaching aids), reflecting one point of view accepted in society. A wide information field of activity is needed, various sources of information, different views, points of view on the same problem, encouraging the student to think independently, to search for his own reasoned position.

The Internet uses a wide variety of resources for educational purposes - from web pages with interesting educational materials to voluminous projects. One such major project is

the National Corpus of the English Language. Such a branch of linguistics as corpus linguistics carries out the development, creation and use of linguistic corpora. The National Corpus is an information and reference system based on a collection of texts in electronic form. Anyone can use it, but first of all it is designed for those who are engaged in linguistic research, study or teach English. A good scientific apparatus (usually called a markup, or annotation, of a corpus) allows you to quickly and efficiently find examples of words, forms, and structures of interest to the user in texts.

Corpus linguistics is a branch of applied linguistics that deals with the development of general principles for the construction and use of linguistic corpora (corpuses of texts).

One of the main sources of linguistic material necessary for linguistic and philological research is the text, written or oral. Based on the analysis of a set of texts, one can draw a conclusion about the linguistic phenomenon of interest to the researcher, for example, the behavior of a grammatical structure, the use of expressive means in a language, etc. The development of computer technology has contributed to the fact that a large number of texts have become available in electronic form. In order to be able to work with such volumes of texts, extracting the necessary information from them, linguistic corpora began to be created all over the world, i.e. collections of texts, specially selected, marked up according to various linguistic parameters and provided with a search system. The increase in the volume of research material required the use of new methods for analyzing linguistic data, including their statistical processing.

Thus, corpus linguistics includes two aspects: firstly, the creation and marking (annotation) of text corpora and the development of search tools for them, and, secondly, the linguistic aspect itself - experimental research based on corpora.

The ability to use extensive electronic resources has greatly facilitated the process of collecting material in linguistic research. However, such accessibility of language data has radically changed the requirements for the evidence base of linguistic research: in our time, it is not enough to give a few examples to confirm the concept put forward, it is necessary to provide a complete sample of data from a particular language. However, working with megabytes of "raw" texts is no less difficult than looking for examples manually.

The need and possibility of processing a large array of texts to extract linguistic, literary and other data from them led to the rapid growth of electronic resources, including linguistically annotated text corpora.

The use of text corpora makes it possible to observe the behavior of linguistic units (words, phrases, grammatical categories, syntactic constructions, etc.) of interest to the researcher in a natural language environment, i.e. in real-life, not artificially constructed contexts.

In addition, corpus studies allow, using statistical methods, to formulate, confirm or refute some hypothesis about a particular linguistic phenomenon on a large amount of material.

Moreover, if a researcher uses an already existing corpus, he completely bypasses the long and laborious stage of collecting material (surveying informants, working with vocabulary cards or written texts, etc.).

For some purposes, it is sufficient to use as a corpus already existing electronic collections of texts, such as virtual libraries, archives of electronic versions of periodicals or news feeds. The entire Internet is a gigantic collection of texts. There is even such a direction of research - the Internet as a corpus. However, since the texts on the web are not systematized and do not have a linguistic annotation (corpus markup) - which also applies to other virtual collections of texts - it is difficult to give an accurate statistical assessment of the entire set or some subset of texts, therefore, there is a need for structured and annotated linguistic corpora.

One of the priority areas of corpus linguistics is the creation and expansion of universal national corpora (corpuses of a particular natural language), representative of the entire language, which can serve to study the most diverse phenomena of this language [2, p. 53]. Most of the world's languages already have their own national corpora. A recognized example is, in particular, the British National Corpus (BNC). Among the corpora of Slavic languages, the Czech National Corpus, created at the Charles University in Prague, stands out. National corpora also exist for German, Chinese, Finnish and other languages. For the English language, such a representative corpus is the National Corpus of the English Language (NCRL).

With the creation of an increasing number of corpora and the development of various text information processing technologies, some universal principles and markup languages of linguistic information in the corpus were developed, special tools were created for working with the corpus in any language, the so-called corpus managers. The development of universal standards and technologies made it possible to create large representative corps in a very short time.

Thus, the subject of corpus linguistics is "the theoretical foundations and practical mechanisms for creating and using representative arrays of linguistic data intended for linguistic research in the interests of a wide range of users" [3, p. 5].

Electronic libraries are also widely used in the educational process. The advantages of the electronic form of a document in comparison with the printed one are the reliability and compactness of information storage, the possibility of its prompt distribution and wide use.

The electronic library contributes to the following main functions:

1. Informational, aimed at meeting the need for information of various categories of users in all branches of knowledge or one of the subject areas.
2. Educational, implemented, including through the popularization of electronic documents related to history and culture.

3. Research, focused on promoting in-depth study of the topic (subject) by scientists and specialists.
4. Educational, within the framework of which support is provided for both basic and additional education by providing not only multimedia educational material, but also the necessary literature.
5. Reference, allowing you to obtain reliable information reflected in documents of a certain type.

Thus, the teacher, using the Internet system, can:

- A. update the material of their own textbooks;
- B. find additional information;
- C. develop training materials;
- D. post educational information.

This paper discusses the widespread use of the Internet in the teaching of the English language. The relevance of this topic is due to the rapid development of information technologies and the need for their implementation in the field of education.

The purpose of this graduation project is to identify the possibilities of using the Internet system in teaching the English language at school.

The tasks of this work are the following:

- A. consider the basic concepts of corpus linguistics and the principles of creating linguistic corpora;
- B. describe how the materials of the National Corpus of the English Language can be used in the study of vocabulary, morphology, syntax;
- C. give a brief description of the largest digital libraries.

Summing up this review, we can highlight the prospects of corpus linguistics and the problems arising from them.

Firstly, a huge amount of information makes it difficult to search for it, which requires the creation of special corpora and additional search tools designed specifically for the purposes of this corpus.

Secondly, the conceptual search, designed to make the user's query more efficient, actually comes down to searching for synonyms and organizing them into semantically nests, i.e. belongs more to the field of semantics than logic, and is difficult to formalize.

Thirdly, corpora are built to provide objective information, so the creators of the corpus are faced with the task of reducing the subjectivity factor in the selection of texts for the corpus and developing strict and clear selection criteria. All this does not reduce the significance and prospects of research in the field of corpus linguistics and points to new areas of applied research.

Reference

- [1] Clive Oxenden, Christina Latham Koenig. New English File. Intermediate Student's Book: Oxford University Press, 2010 - 159 p.
- [2] Rachel Roberts, Antonia Clare, JJ Wilson. New Total English. Intermediate Student's Book: Pearson Longman, 2011 - 165 p.
- [3] Sosnina E.P., Corpus linguistics and corpus approach in teaching a foreign language. Access mode:

- http://ling.ulstu.ru/linguistics/resourses/literature/articles/corpus_linguistics_language_teaching/
- [4] Bakirova Hilola Botiraliyevna DIFFICULTIES IN WORKING WITH TECHNICAL TERMS IN ESP EDUCATION International Conference on Scientific, Educational & Humanitarian Advancements Hosted online from, Samsun, Turkey www.econferenceglobe.com July 15, 2021. 65-67. Retrieved from <https://papers.econferenceglobe.com/index.php/ecg/article/view/605>
- [5] Bakirova H.B. The role of terms of specialty in professionally oriented education. Journal of Hunan University (Natural sciences) Vol 48. No.11.2021. 1430p.
- [6] Bakirova H.B. The content of teaching foreign languages. Eurasian Journal of Learning and Academic Teaching. Vol.2 www.geniusjournals.org. ISSN: 2795-739X. Belgium. 10-14p

