

CORPUS LINGUISTIC TECHNOLOGY IN STUDYING ENGLISH

Olena Skorykova¹

DOI: <https://doi.org/10.30525/978-9934-26-164-0-19>

The particular interest in the English language has sharpened the need for quickly and efficiently obtained information concerning grammar and vocabulary. Taking advantage of such modern technology as concordancer or KWIC (key words in context) can help people to find the necessary information in the shortest term. French and Sweden scientists Seretan and Wehrli (2013) present «an enhanced type of concordancer that integrates syntactic information on sentence structure as well as statistical information on word co-occurrence».

The purpose of this research is to develop students' critical attitude to the English language and to clarify their understanding of finding the correct

¹ Petro Mogila the Black Sea National University, Ukraine

information about grammar and expand the students' vocabulary by using the means of corpus linguistics.

Although considerable research is devoted to the usage of corpus technology for solving linguistic issues, less attention is paid to studying the lexical collocations and linguistic evidence, which either supports or contradicts the prescriptive grammar rules (Derybina, 2012). Given the fact that corpus linguistics emerged as a separate branch of science only at the end of the XX century, the studies on it are insufficient.

Analytical and synthetic methods are used in the search and analysis of information on a selected topic; experimental research method is used to determine the capabilities of the hull technology; modeling method is used to describe instrumental competence. With the help of text corpora, a wide variety of tasks can be solved, so working with them has become one of the leading methods of linguistic research. Based on the analysis of a set of texts, it is possible to draw a conclusion about the linguistic phenomenon of interest to the researcher, for example, about the behavior of a grammatical or syntactic structure, the use of expressive means in a natural language environment, in real-life, and not artificially constructed contexts.

In addition, corpus studies allow, using statistical methods, to formulate, confirm or refute a 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 (interviewing informants, working with dictionary indexes or written texts). Thus, the possibility of widespread use of electronic resources greatly facilitated the process of collecting information.

In English lessons, with the help of a computer, you can solve a number of didactic tasks: to form skills and reading skills using materials from the global network; improve writing skills; replenish the vocabulary of students; to form motivation for learning English.

A corpus is one such database. Corpora are the plural of the corpus (it is many such databases). Concordancer is a computer program that works with corpora (texts) that makes a list of words in a minimal context, or a list of all contexts in which a word or phrase occurs in the text under study. Concordancers allow you to search text in a similar way to how you search information on the web. In addition, students can increase their vocabulary with idioms through concordancer.

The Corpus of Contemporary American English is the largest and the only freely available corpus of American English. It contains 220 thousand texts, 520 million words of the period from 1990 to 2017. In recent years, researchers have become increasingly interested in spoken corpus. Most students are

interested in being able to speak a foreign language fluently, so the spoken corpus provides many opportunities.

A speech corpus (or spoken corpus) is a database of speech audio files and text transcriptions. In linguistics, spoken corpora are used to do research into phonetic, conversation analysis, dialectology and other fields.

There are two types of Spoken Corpora:

Read Speech – which includes:

- Book excerpts
- Broadcast news
- Lists of words

Spontaneous Speech – which includes:

- Dialogues – between two or more people (includes meetings);
- Narratives – a person telling a story (one such corpus is the Buckeye Corpus);
- Map-tasks – one person explains a route on a map to another;
- Appointment-tasks – two people try to find a common meeting time based on individual schedules.

Some of the concordancer's functions are very useful for students, you can see a list of some of them:

KWIC – Key Word in Context. Hans Peter Lohn first coined the term KWIC. The system was based on a concept called the keyword in headings, which was first proposed for Manchester libraries in 1864 by Andrea Crestadoro.

COLLOCATE (of words and phrases) a phrase that has signs of a syntactically and semantically integral unit, in which the choice of one of the components is carried out by meaning, and the choice of the second depends on the choice of the first.

LIST a number of connected items printed consecutively, typically one below the other.

Computer concordancers are introduced into the educational process for the preparation and writing of abstracts, term papers in foreign languages, as well as for research work of students. The possibilities of concordancer in comparison with a text editor are much wider. Concordance analyzes not one, but several texts or electronic text corpora at once and provides information about the context of the use of language units.

References:

1. Barker, Paul. *A Glossary of Corpus Linguistics*. Finland.WS Bookwell, 2006: 9–15. Print.
2. Baumgart, Joanna. *Corpus Linguistics and Cross-Disciplinary Action Research*. London. Routledge, 2021: 14–21. Print.

3. Cambridge Dictionary. Cambridge University Press, 2021
dictionary.cambridge.org

4. Jones, Christian. *Successful Spoken English: Findings from Learner Corpora*. London. Routledge, 2017: 5–17. Print.

5. Zakharov, Vadim. *Corpus Linguistics*. Irkutsk. Author's edition, 2011: 7–20. Print.