



## **CORPUS LINGUISTICS AND CORPUS-BASED APPROACH IN FOREIGN LANGUAGE TEACHING**

**Jabborova Gulhayo Faxriddinovna**

A teacher of  
English Linguistics department  
Bukhara State University

<b>Article history:</b>	<b>Abstract:</b>
<b>Received:</b> August 6 <sup>th</sup> 2021 <b>Accepted:</b> September 7 <sup>th</sup> 2021 <b>Published:</b> October 6 <sup>th</sup> 2021	The article considers "corpus-based approach" in Foreign Language Teaching and the basic characteristics of this method that determine its reliability. Corpus-based approach or the method of linguistic research based not act entirely on text corpora focuses on the applied study of language, its functioning in natural environment and texts that is important in language teaching.

**Keywords:** Corpus linguistics, national corpus, training with the help of corpus, corpus usage, modern linguistics.

### **I. INTRODUCTION**

The world practice of development of Corpus Linguistics (CL) proves the efficiency of these applications, though in today England the rich opportunities of CL methods have not been thoroughly realized in applied linguistics, linguistic pedagogy, native and foreign language teaching. We specify the types of corpora which can be used in practice of language teaching, and present the examples of corpora accessible to a teacher of foreign languages. The paper considers practical use of Parallel Corpora in language teaching, learning and translation, and Learner Corpora in Language Acquisition research. We prove the efficiency of application of concordance programs in linguistic tasks, e.g. Computer-Assisted Language Learning and Training. In the conclusion we mention the real application of Corpus Linguistics, method of corpus analysis in linguistic research and practice of Foreign Language Teaching.

### **II. LITERATURE REVIEW**

Currently, corporuses of written and oral texts are successfully used in teaching a foreign language and in linguistic pedagogy. Lists of active vocabulary of students, frequency lists of terms for use in professional courses, etc. are formed on the basis of the corpus. Academic dictionaries and textbook developers rely on authentic text arrays (Corpora). In addition, collections, libraries and arrays of texts reflect the real functioning of a particular language, and their transfer to computer environments only intensified their practical and widespread use in applied linguistics.

### **III. ANALYSIS**

Corpus linguistics provides material for various kinds of studies of the language and its variants, and determines the main method of text analysis based on corpus (Corpus-Based Approach). The corpus

approach, or the method of linguistic research based on text corporuses, is focused on the applied study of the language, its functioning in real environments and texts, which is important for language teaching. For example, corpus-based lexicographic analysis clearly helps to reveal the contextual use of certain words, especially synonymous ones (for example, small / little, big / large), their frequency compatibility with other words, their regularity in certain styles, and clearly define their semantics.

The main characteristics of the method, which determine its reliability and validity, are as follows:

- is empirical and analyzes real word usage in a natural language environment,
- uses a sufficiently large, representative selection of texts,
- actively uses computers and special concordance programs for analysis in automatic and interactive modes of operation,
- based on the methods of statistical and qualitative analysis of the text,
- is target, i.e. should be focused on real application and results.

One of the important features of the text corpus-based analysis method is the study not only of purely linguistic phenomena (grammatical or lexical functions of words, their connections with other lexemes), but also of such phenomena as, for example, the frequency of lexemes or grammatical structures in certain genres, dialects.

Electronic Corporuses in Language Teaching

Electronic corpora provide rich linguistic material for teaching and research purposes. Currently, the Internet contains many classic electronic corpora in foreign languages. The most famous of them are the British and American national corporuses of the English language, the German-language corporuses LIMAS, COSMAS. The most accessible for the average user-teacher of foreign languages are Gutenberg



Texts, British National Corpus Sampler, The Longman Corpus, LIMAS, Reuters news corpus, electronic archives of major newspapers (for example, The Times).

As for the typology of text corpora, in applied linguistics it is possible to use such types as:

Research - to study various aspects of the functioning of the language system;

Illustrative, including educational (Learner Corpus) - to confirm and substantiate linguistic facts;

Monitor - to study the dynamics of linguistic material, conduct content analysis, for example, a corpus on journalism;

Static - for exploring styles, for example, author corpora or writers' text corpus;

Multimedia - text + video + audio;

Corpus of parallel texts - for comparative analysis of "original-translation" texts for teaching methods and techniques of translation. There are two basic forms of organization of such corpora: "original-translation/s" (Unidirectional), "original - translation - reverse translation" (Bidirectional or reciprocal), arranged in parallel.

In this article, as an example, we will consider the practical use of parallel corpora in language teaching and educational corpora in research related to the problems of foreign language acquisition.

#### **IV. DISCUSSION**

Parallel Corpora in Language and Translation Teaching

In the methodology of teaching a language (grammar-translation method) and teaching methods in translation, an interesting application is the development of parallel electronic corpora of texts (Parallel Corpora) and the use of concordance programs for parallel texts. Such developments in Russia are under development, although parallel texts have long been used for comparative translation and teaching.

In a practical sense, translation should focus on post-editing, comparison and evaluation of different strategies and interpretations within the context. A translator (especially a beginner) needs resources that could serve as translation standards and translation assessment under certain "standard" conditions. According to some reports, about 50%, and at the initial stage of training, up to 80% of the translation time is spent on referring to abstract information, for example, dictionaries. Electronic parallel corpora and linguistic computer technologies can significantly reduce these time costs and provide samples of professional translation in the study of techniques and methods of translation.

At present, corpora (or parallel texts) of fiction are especially widespread, although for teaching translation at a university it is necessary to develop corpora of different genres and styles and, first of all, focus on scientific, technical, journalistic and business texts.

Academic Corpora in Language Acquisition Research

An educational corpus (Learner Corpus) is understood as an electronic corpus of texts of a group of foreign language learners. The main goal of organizing educational buildings is to analyze them in order to identify ways and effectiveness of mastering the studied language (Language Acquisition).

Corpora of this kind, for example, can be used for linguistic analysis to identify lexical or syntactic errors when mastering a foreign language. This approach helps to establish the frequency of certain types of language errors, characteristic contexts, which is necessary to develop plans and methodological techniques for further correction in language teaching.

Educational buildings are most common in Asia and Europe. The most famous is the International Corpus of Learner English (ICLE) essays by students of the advanced language level. This corpus is mainly used for discourse analysis and statistical analysis of students' vocabulary, comparative research. This corpus is a prime example of the effectiveness of developments in the field of corpus and applied linguistics.

Concordance Programs in Applied Linguistics

In the field of applied linguistics, concordances have received special recognition from linguists due to new opportunities for effective language research and processing of lexical material of texts of various kinds. Recently, computer concordances have been actively used in automated teaching of foreign languages (or CALL - Computer Assisted Language Learning).

A concordance program is a special text processing program that sets some linguistic task to find a particular morpheme, word or phrase in a context. For example, in the case of the English language, find in this group of texts the use cases for the indefinite articles or for all words ending in "-ing". As a result of the work, the concordance program will display all words with the given ending along with the context, as a rule, this is a line of text.

Thus, the teacher receives many examples of both grammatical and lexical forms of the word (in our example, these are verbal nouns, gerund, verb form - participle I, etc.). The student, in turn, receives natural examples of demonstrating certain grammatical or lexical phenomena, can independently conduct



linguistic research, engage in research and development.

In the section of studying the grammar of a foreign language, the student may be asked to find and analyze the forms of expression and use of complex tenses (for example, Perfect), modal verbs and their role in a sentence, the place of adverbs in a sentence, etc. In the vocabulary section - for example, find and explain with examples such often difficult words as MAKE / DO, RISE / RAISE, TELL / SAY, LIE / LAY, etc. In the syntax section - for example, examine the punctuation of a particular language and identify differences in comparison with the native language. Sources for such works can be not only special corpora of electronic texts, but also various electronic publications, electronic libraries (for example, on the Internet).

Currently, concordances are a modern effective tool for text analysis, which should be actively used in language teaching practice and in linguistic tasks.

## V. CONCLUSION

The analysis of text corpora, methods and developments in corpus linguistics are a promising direction in the field of teaching foreign languages. The world practice of the development of this area proves the effectiveness of this kind of applications, although at present the possibilities of the methods of corpus linguistics in Russia have not yet been properly implemented in applied linguistics, linguistic teaching, teaching the native and foreign languages. At the Department of Applied Linguistics, Ulyanovsk State Technical University, research is being carried out related to the development of an educational electronic corpus of written texts for the initial level of foreign language teaching, as well as a parallel corpus (English-Russian) of newspaper texts for teaching translation. For several years now, computer concordances have been used at the department in

the educational process for the preparation and writing of abstracts, term papers in foreign languages, as well as for research work of linguistic students.

## REFERENCES:

1. Baranov A.N. Introduction to Applied Linguistics. M: Editorial URSS, 2001 pp. 112-136.
2. Dobrovolsky D.O. Corpus of Parallel Texts as a Tool for Literary Translation Analysis//<http://www.dialog-21.ru/Archive/2003/Dobrovolskij.htm>
3. D.Biber, S. Conrad, R. Reppen Corpus Linguistics: Investigating language structure and use. - Cambridge University Press, 2001.
4. COSMAS - German corpus. <http://corpora.ids-mannheim.de/~cosmas/>
5. Gutenberg Texts. <http://promo.net/pg/>
6. British National Corpus. <http://www.natcorp.ox.ac.uk/getting/sampler.html>
7. The Longman Corpus. <http://www.longman.com/dictionaries/corpus/lccont.html>
8. LIMAS Corpus. <http://www.ikp.uni-bonn.de/Limas>
9. American National Corpus. <http://americannationalcorpus.org/>
10. International Corpus of Learner English// <http://jupiter.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/cecl.html>
11. Michael Barlow ParaConc: A Parallel Concordancer <http://jupiter.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Events/icamepr.htm#para>
12. Tim Johns Data-driven Learning Page. <http://web.bham.ac.uk/johnstf>
13. Michael Barlow's Corpus Linguistics Page. <http://www.ruf.rice.edu/~barlow/corpus.htm>