

## CORPUS-BASED APPROACH IN TEACHING FOREIGN LANGUAGES

**Radjabova Gulnoza Giyosiddinovna**

Senior Teacher, Uzbekistan State World Languages University, Tashkent, Uzbekistan

E-mail: rad.gulnoza@gmail.com

### Abstract

Corpora of written and oral texts are now successfully used in foreign language teaching and language pedagogy. This article focuses on the Corpus-Based Approach and its application in teaching foreign languages. The Corpus-Based Approach is focused on the empirical study of language, its functioning in real, authentic situations and texts, which is important for language teaching. Additionally, the article identifies the types of corpora that can be used in the process of teaching foreign languages, and provides examples of corpora which are available to foreign language teachers. Besides, this article examines the practical use of parallel corpora in language teaching and translation, as well as educational corpora in studies related to problems of foreign language acquisition. The paper also reveals the effectiveness of using corpus software such as concordance programs in linguistic tasks, including “automatic learning” of foreign languages.

**Key words:** Corpus Linguistics, Corpus-based approach, concordance, teaching languages, parallel corpus.

### Introduction

Corpora of written and spoken texts are now successfully used in foreign language teaching and language pedagogy. Students' active vocabulary, frequency lists of terms are formed with the help of corpora. Dictionary compilers rely on authentic texts of corpora as they reflect the real, authentic functioning of a particular language, and their transfer to computer environments has only intensified their practical and wide use in applied linguistics [1].

Corpus linguistics provides material for various kinds of studies of language and its variants, and defines the main method of corpus-based text analysis (Corpus-Based Approach) [3]. The Corpus-Based Approach, or a method of linguistic research based on corpora of texts, is focused on the empirical study of 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 synonyms (e.g., small/little, big/large), frequency of their usage, their collocability with other words, and their regularity in certain registers.

The main characteristics of Corpus-based approach that determine its reliability and validity are as follows:

<https://conferencea.org>

- It is empirical and analyzes real word usage in naturally-occurred, authentic language environments;
  - It uses a sufficiently large, representative selection of texts;
  - It actively uses computers and special concordance programs for analysis in automatic and interactive modes of operation;
  - It is based on methods of statistical and qualitative text analysis;
  - It is target-oriented, in other words, it should be focused on the real application and results;
- One of the important features of the corpus-based analysis method is the study not only of purely linguistic phenomena (grammatical or lexical functions of words, their relationships with other lexemes), but also of such phenomena as, for example, the frequency of lexemes or grammatical constructions in certain genres or dialects.

## **Electronic corpora in language learning**

Electronic corpora provide rich linguistic material for educational and research purposes. At present there are many classical electronic corpora which are available online. The most famous are the British National Corpus (BNC) and American Corpus of Contemporary English (COCA), The Longman Corpus, LIMAS, Reuters news corpora, electronic archives of major newspapers (for example, The Times) [4, 5, 6, 7, 8]. These corpora can be distinguished by the purpose of its usage, therefore, there are such types of corpora as:

- Exploratory - to study various aspects of the functioning of a language system;
- Illustrative, including instructional (Learner Corpus) - to corroborate and substantiate linguistic facts;
- Monitor - to study the dynamics of linguistic material, content analysis, for example, a corpus on journalism;
- Static - for researching styles, such as author corpora or writers' text corpora;
- Parallel corpus - for comparative analysis of original-translation texts to teach translation methods and techniques. There are two basic forms of organization of such corpora: Unidirectional and Bidirectional or reciprocal [11].

In this article, as an example, we will consider the practical use of learner corpora to investigate its effectiveness in foreign language acquisition.

## **Learner's corpora in language acquisition**

A Learner's Corpus is understood as an electronic corpus of texts by a group of foreign language learners. The main purpose of organizing learner corpora is to analyze them for the purpose of identifying the ways and effectiveness of language acquisition. For example, such corpora can be used for linguistic analysis to identify lexical or syntactical errors in the acquisition of a foreign language. Such an approach helps to identify the frequency of various

<https://conferencea.org>

types of language errors and their characteristics, so that it will be possible for the educator to develop certain teaching approaches and methodological techniques for further correction in language teaching.

Teaching corpora are most common in Asia and Europe. The most famous is the International Corpus of Learner English (ICLE) of essays of advanced language students [10]. This corpus is mainly used for discourse analysis and statistical analysis of students' vocabulary, comparative studies. This corpus is an example of the effectiveness of developments in corpus and applied linguistics.

### **Concordance Programs in Applied Linguistics**

Recently, in the field of applied linguistics, concordances have gained special recognition among linguists because of their new possibilities for effective language research and processing of lexical material of texts of various kinds. Therefore, computerized concordances have become actively used in computer aided language learning (or CALL - Computer Assisted Language Learning).

A concordance program is a special text processing program, which has some linguistic tasks and goals to find a particular morpheme, word or phrase in any context. For example, concordance programs can find variants of indefinite articles or all words ending in “-ing” in a given group of texts. As a result, the concordance program will produce all words with the given ending within the context. It is worth mentioning that concordance is usually presented in the form of a line of text which can later be analyzed by disclosing bigger context.

Thus, the teacher gets a lot of examples of both grammatical and lexical forms of a word (in our example these are verbal nouns, gerunds, verb form - participle I, etc.). The student in turn receives natural examples of demonstration of certain grammatical or lexical phenomena, can independently conduct linguistic research, being engaged in research work.

Regarding teaching L2 grammar, the student may be asked to find and analyze forms of expression and use of complex tenses (e.g., Perfect), modal verbs and their role in a sentence, the place of adverbs in a sentence, etc. In vocabulary, for example, learners can be asked to find and explain, using examples, words that often cause difficulty in use, such as MAKE/DO, RISE/RAISE, TELL/SAY, LIE/LAY, etc. In the syntax section, for example, investigate the punctuation of a particular language and determine the differences compared to the native language. Sources for such works can be not only special corpus of electronic texts, but also various electronic publications, electronic libraries (for example, on the Internet).

As abovementioned, concordances are modern effective tools for text analysis, which should be actively used in not only language teaching but also in fulfilling different linguistic tasks.

## Conclusion

Analysis of corpus texts, methods and developments of corpus linguistics is a promising direction in the field of foreign language teaching [12, 13]. The development and introduction of corpus linguistics into teaching L2 has already proved the effectiveness of such applications, though at present the possibilities of corpus linguistics methods in Uzbekistan are not yet adequately implemented in applied linguistics, linguistic education, and teaching of native and foreign languages. However, there is the research conducted by some linguists in Uzbekistan State World Languages University related to the involvement of corpus technologies and tools in teaching writing as well as the development of an educational electronic corpus of written texts of the initial level of foreign language teaching, and a parallel corpus (English-Russian) of newspaper texts for teaching translation. For several years the computerized corpus has been used in the educational process to prepare and write essays and term papers on foreign languages as well as for research works of students of linguistics.

## LITERATURE

1. Baranov A.N. Introduction to Applied Linguistics. M: Editorial URSS, 2001 P.112-136.
2. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating language structure and use. - Cambridge University Press, 2001.
3. Dobrovolsky D.O. The corpus of parallel texts as a tool for the analysis of literary translation //http://www.dialog-21.ru/Archive/2003/Dobrovolskij
4. COCA – Corpus of Contemporary American English - <https://www.english-corpora.org/coca/x.asp>
5. COSMAS - German corpus. <http://corpora.ids-mannheim.de/~cosmas/>.
6. British National Corpus. <http://www.natcorp.ox.ac.uk/getting/sampler>
7. The Longman Corpus. <http://www.longman.com/dictionaries/corpus/lccont>
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>
11. Michael Barlow ParaConc: A Parallel Concordancer <http://jupiter.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Events/icamepr>
12. Tim Johns. Data-driven Learning Page. <http://web.bham.ac.uk/johnstf>
13. Michael Barlow's Corpus Linguistics. <http://www.ruf.rice.edu/~barlow/corpus.html>