

## Corpus Linguistics in Language Teaching

*Xonqulova Nilufar Ravshanovna*<sup>1</sup>

**Abstract:** The article deals with the corpus linguistics as one of the most promising areas in modern theoretical and applied linguistics, based on the use of language electronic corpora and corpus technologies. A corpus of a language is a collection of texts in a given language in electronic form, specially processed (“annotated”), so that a language researcher can quickly and completely find information about the language that interests him in the corpus. Depending on the nature of the abstract, it is possible to search by the grammatical characteristics of words and sentences of the language (morphological, syntactic, semantic parameters), as well as by various characteristics of the texts themselves included in the corpus: by author, date of creation of the text, genre, subject, etc.

**Keywords:** corpus linguistics, corpus of a language, electronic corpora, corpus technology, grammatical characteristics of words, characteristics of the text, corpora in teaching.

Over the past two decades, corpora have been used not only in linguistic research, but also in language teaching and learning, although, apparently, the creators of corpora did not even consider their use in this way. To date, there are a significant number of works based entirely on corpora (for example, dictionaries and grammars) available to teachers and students, as well as developments on the use of concordances and exercises based on working with corpora that can be used in the classroom. The number of printed materials and resources for teaching foreign languages, marked with the corpus-based icon (“created using corpus data”), is increasing. It is rather difficult to find a modern dictionary of a well-known publishing house that would not be compiled on the basis of a linguistic corpus. Corpus data forced us to take a fresh look at our understanding of the language and reconsider the approach to compiling grammar and lexical reference books, textbooks and other materials for those who want to master a foreign language. Thus, the contribution of corpus linguistics to the description of the language, as well as to the methodology of teaching foreign languages, seems to be difficult to dispute. Corpus linguistics brings innovative changes to scientific methods and technologies, and perhaps anticipates even more fundamental changes that will challenge our traditional ideas about education, about the role of the teacher, about the cultural context of the provision of educational services.

In the leading universities of the world, it is becoming a daily practice to use corpus data as an empirical component of lecture courses, student assignments, and independent projects. At the same time, it turns out that the corpus approach is optimal for visual representation of such aspects of the language as historical, geographical and social variation and changes in the language system, while simultaneously giving a living opportunity to master the basic principles of corpus methods of linguistic analysis. Significant assistance in this area can be provided by the methods of corpus linguistics: automated information extraction, data-based learning, text searches in large-scale corpora using natural language processing methods.

### Indirect use of corpora in teaching a foreign language

The influence of modern technologies on foreign language learners, the emergence of new technological types of language functioning lead to a rethinking of the modern definition of communicative competence. In terms of such a conceptualization of the language as the context-correct use of various registers of speech [2], the concept of communicative competence should be defined as corresponding to the natural conventions (both grammatically and stylistically) of using the

<sup>1</sup> Teacher, Applied translation department, Translation faculty Uzbekistan state world languages university



language in various communicative situations, including those in which technologies. In this regard, it is inevitable and necessary to introduce technology into all aspects of applied linguistics, not as a separate subject, but as an indispensable research and teaching tool and object of critical analysis [2]. There is numerous experimental evidence that students who are encouraged to independently form their own understanding of the features of the language being studied master language competencies faster and more efficiently than those who are driven into the head with the rules (which, in fact, often turn out to be inadequate to the real state of the language). In addition, the experience of the collective "discovery" of the language in a student group environment using these new techniques introduces a valuable element of discussion and joint generalization.

In the last 10-15 years, researchers in foreign language teaching methods have used extensive corpora of texts to assess the reality of the language in its natural state. These corpora of texts have greatly influenced the improvement of the quality level of the published language manuals. Instead of traditional prescriptivist instructions on how to properly use language, new corpus studies describe, empirically sound analyzes of what people actually say. Special mention deserves new dictionaries created using corpus linguistics techniques, such as Longman, Oxford, Collins, as well as the experience of critical rethinking of the postulates of descriptive grammar of the English language (Longman Grammar of Spoken and Written English, published in 2000).

Experience gained, as well as common practice, suggests that the most consistent results are obtained when these new techniques are introduced step by step into the learning process to ensure effectiveness and consistent motivation. Thus, at the initial stage, carefully selected quotations in the handout with well-designed instructions and tasks should be used, and at subsequent stages, students will be able to adequately cope with the unpredictability of the "live" search for concordances on the Internet and formulate research problems on their own. The most interesting in this aspect are the critical analytical works on modern English grammar obtained in the field of natural spoken language, which, in fact, reveals much more extensive divergences from the standard written language than has ever been indicated in educational and methodological manuals. Now, having been identified with the help of corpus methods, these features can be taken into account in the process of teaching and in the development of modern educational and methodological complexes.

The language corpus is not only a database of empirical data to test our ideas about the language, but also allows us to identify those features of the language that have previously eluded our field of vision (for example, the frequency of using certain set phrases and idioms). However, research has shown that so far, many textbooks have presented a language based on our understanding of its use rather than hard data. Therefore, despite the pedagogically sound use of sprint dialogues, such scientists have disputed their status as a means of improving conversational skills over the past years as R. A. Carter, A. Burns, M. McCarthy, A. O'Keeffe and others. A. Burns notes that sprinted dialogues rarely reflect the unpredictability and dynamism of the conversation, as well as the features and structures of natural conversational discourse [2]. Consequently, students taught only using sprinted oral language will have fewer opportunities to expand their language repertoire for unforeseen situations of communication that go beyond the classroom.

An analysis of the differences between speech acts in dialogues in textbooks and real spontaneous conversations reflected in the corpus can be found in the works of D. Boxer and L. Pickering. R. A. Carter also compares data from the Cambridge and Nottingham Corpora of English (CANCODE) with textbook dialogues and finds that the latter do not include enough of the main features of the spoken language, such as discourse markers, ambiguous expressions, ellipsis, etc. [3]. A. Gilmore examines the discourse means of seven dialogues published in textbooks from 1981 to 2000, compares them with similar authentic conversations in the corpus and comes to the conclusion that the former differ significantly from the latter in terms of the length and patterns of statements, lexical richness, the number of false starts and repetitions, pauses, use of means of expressing uncertainty and reciprocal signs. However, after analyzing textbooks published later, the scientist points to the increasing attention of authors to more natural discourse means [4]. Corpora are becoming more and more an indispensable resource for developing curricula with an emphasis on communicative competences and



give students the chance to encounter the linguistic phenomena that await them in real communication when learning a language. In the context of such teaching, the Collins COBUILD English Course (1989), a branch of the COBUILD project devoted to pedagogically oriented lexicography, was probably the pioneer. The content of this new corpus-based course was the most common words and phrases in the English language. The course focused on vocabulary collected from corpus research and on the ideas underlying corpus research, namely that language is made up largely of repeated combinations of words rather than individual words, and the idea of a close relationship between vocabulary and grammar. .

From modern textbooks, for example, the Touchstone series of textbooks is an attempt to show how dialogues in textbooks and all educational material in general can be based on corpus data. In addition to the traditional content, taking into account the four types of speech activity - speaking, listening, reading, writing, the authors of Touchstone include crosscutting sections aimed at teaching conversation strategies based on the most common words and phrases from the North American conversational segment of the Cambridge International Corpus.

Let's consider an example of how strongly the type of a scripted and non-scripted speech utterance can differ (it is necessary to pay attention to discursive markers, fillers, repetitions, etc.)

New English File, Intermediate level  
New Total English, Intermediate level

## 2. Direct application of corpora in teaching a foreign language (data-driven learning)

In the last decade, another new and extremely promising direction has arisen in the organization of the process of teaching foreign languages, in which the student has the opportunity to resort to the use of "raw" language data directly from the corpus. This direction is called data-driven learning, or data-driven learning (DDL). The term Data Driven Learning (DDL; literal Russian translation - learning activated or stimulated by electronic information) was introduced into wide use by Tim Jones (University of Birmingham) in 1991. This is an approach to language learning, where the student is a language researcher using information computer technologies and a corpus presented in electronic form [4]. Naturally, the student cannot, without the help of the teacher, and completely independently explore the corpus. DDL requires a significant preparatory work from the teacher: a careful selection of information must be made, and then the information must be presented in a convenient and methodically effective form. At this stage, the role of the teacher is decisive.

However, with the direct use of prepared materials, both the teacher and the student work together. Their cooperation is to some extent mutually beneficial. And if the help from the teacher is obvious and common, the student can act in a role that is not typical for himself in such training. Often a person studying a language at the first stages may notice such lexical and grammatical features that escape the attention of teachers and philologists (for example, similar features of the native and target languages that are obvious to a person who does not speak a foreign language).

The DDL method was first used in individual teaching (one-to-one tuition) of students in the English for Academic Purposes program. The DDL was then used at the University of Birmingham in teaching grammar and vocabulary to students in the same program - in lectures with a large number of students using specially prepared handouts.

Thus, the role of the corpus approach in such an area as computerized teaching of foreign languages is central. Recent research by Lancaster University on software for teaching grammar and basic grammatical analysis to undergraduate students has shown that programs such as Cytos and similar, created quite easily from a corpus of texts annotated or by parts of speech or by grammatical / syntactic roles, are extremely effective and provide the necessary degree of interactivity along with autonomy [3].

When receiving a task of grammatical analysis of a text with a hidden annotation, students independently parse sentences, having the opportunity to request help from the program in the form of a list indicating information about the frequency of use of a particular ion or lexical unit or the frequency of joint use of examples (collocations).



## Bibliography

1. Corpora and language teaching. Access Mode:  
[http://www.lexically.net/wordsmith/corpus\\_linguistics\\_links/Roemer%202008%20HSK%20CL%20chapter%20final%20print%20version.pdf](http://www.lexically.net/wordsmith/corpus_linguistics_links/Roemer%202008%20HSK%20CL%20chapter%20final%20print%20version.pdf)
2. Information and Communications Technology for Language Teachers. Introduction to the applied meaning of the corpus. Access mode: [http://www.ict4lt.org/en/en\\_mod2-4.htm](http://www.ict4lt.org/en/en_mod2-4.htm)
3. O. V. Nagel, Corpus linguistics and its use in computerized language teaching // Language and Culture. 2008. No. 4. Access mode: <http://cyberleninka.ru/article/n/korpusnaya-lingvistika-i-ee-ispolzovanie-v-kompyuterizirovannom-yazykovom-obuchenii>
4. Raitskaya LK, Didactic possibilities of corpus Internet technologies in teaching a foreign language in higher education. Access mode:  
<http://mgimo.ru/upload/iblock/fa4/fa494cc8fe575018273ad7f2d462124a.do>

