

Sadykova A.K., Aushakhman A.T.

Issues of application of corpus methods in teaching foreign languages

The article discusses the application of the Applied Linguistics methods in the teaching of foreign languages. The authors analyze the concept of "corpus" and the potential of the corpus approach to teaching a foreign language. A brief description of the corpus and the concordance-program is given. The analysis of such methods of corpus linguistics as automated extraction of information, teaching based on data, text searches in large-scale corpuses (concordances) is presented. Moreover, the conditions that must be taken into account for the use of the corpus in the process of teaching a foreign language are explained.

Key words: corpus; concordance; foreign language teaching, corpus methods and tasks, communicative competence.

Садыкова А.К., Аушахман А.Т.

Шетел тілдерін оқытуда корпусстық әдістерді қолдану мәселелері

Мақалада шетел тілін оқытуда қолданбалы лингвистика әдістерінің мүмкіндіктері қарастырылады. «Мәтіндер корпусы» ұғымы мен корпусстық тәсілдің шетел тілін оқытудағы потенциалы кеңінен жан-жақты талданады. Корпус пен конкорданс-бағдарламасының қысқаша сипаттамасы берілген. Ақпаратты автоматтандырылған түрде шығару, деректер негізінде оқыту, ауқымды корпусстарда (конкорданстарда) деректерді мәтінмен іздестіру сияқты корпусстық әдістердің талдауы берілген. Шетел тілін оқу үрдісінде корпуссты қолдану үшін ескерілуі тиіс жағдайларға түсіндірмелер беріледі.

Түйін сөздер: корпус, конкорданс, шетел тілдерін оқыту, корпус-тық әдістер мен тапсырмалар, коммуникативтік компетенция.

Садыкова А.К., Аушахман А.Т.

Вопросы использования корпусных методов в обучении иностранным языкам

В статье рассматриваются возможности методов прикладной лингвистики в преподавании иностранного языка. Анализируется понятие «корпус текстов», рассматривается потенциал корпусного подхода в обучении иностранному языку. Дана краткая характеристика корпуса и программы «конкорданс». Представлен анализ таких корпусных методов, как автоматизированное извлечение информации, обучение на основе данных, текстовые поиски в крупномасштабных корпусах (конкордансы), задания на использование лингвистического корпуса для определения значений слов и др. Поясняются условия, которые необходимо учитывать для применения корпуса в учебном процессе по иностранному языку.

Утверждается, что пользователи национальных корпусов (текстов) являются исследователями различного профиля. При этом, количество пользователей этих корпусов не ограничивается профессиональными исследователями языка. Авторы обосновывают, что корпусный подход может интересовать литературоведов, историков и представителей многих других областей гуманитарного знания.

Ключевые слова: корпус, конкорданс, обучение иностранным языкам, корпусные методы и задания, коммуникативная компетенция.

**ISSUES OF APPLICATION
OF CORPUS METHODS
IN TEACHING FOREIGN
LANGUAGES****Introduction**

In recent years, there has been increasing interest of foreign and domestic researchers to the study of theoretical and practical aspects of using corpus technologies as a factor of improving the quality of foreign language teaching. The research in the field of corpus linguistics is increasingly being reflected in the works on the methodology of teaching foreign languages.

One of the current problems of foreign language teaching is the lack of adequate teaching materials, while the current daily load of teachers in preparation of new teaching materials remains steadily high. In this area, much assistance can attract corpus linguistics methods, which include the automated extraction of information and teaching based on data, text searches in large-scale corpora.

Corpus linguistics, a linguistic direction, which appeared in the 60-ies of XX century in the United States and Western Europe and is engaged in the development of the theory of the corpus and the creation and use of text corpora. With the increasing capabilities of modern computer technology, corpus linguistics gained rapid development, leading to the creation of case projects of various sizes in different languages and for different purposes.

Main part

In the scientific literature can be found different definitions of the corpus, but most of the researchers understand corpus as set organized in a certain way, whose elements are the texts. The organization of the corpus can be different, depending on the pragmatic goals of its creator or the user. The texts, which are the constituent elements of the corpus, may be a complete verbal original product or any part of it. V.V.Rykov considers linguistic corpus as a collection of texts, which is based on a logical plan, the logical idea of combining these elements. The logical idea is embodied in the rules of organization of the texts in the corpus, algorithm and program analysis of text corpus, and conjugate with it ideology and methodology [1]. That is, the corpus is a set of texts collected in accordance with certain principles, marked by a certain standard and provided by specialized search system.

Today, with the help of the corpus you can solve problems that were previously not feasible. The corpus is often seen as a kind of reduced model of language or sublanguage. The validity of the results obtained on its basis is mainly determined by its representativeness, which usually refers to the ability of a limited number of texts selected according to certain criteria, reflect patterns of language use, and characterize the language as a whole. The presence of the corpus and the possibility of collecting and processing large amounts of linguistic data provide verifiable hypotheses, assumptions and conclusions [2].

The users of national corpora are, of course, linguists, researchers of various profiles. However, the number of users is not limited by professional researchers of language. The data on language of a particular period or an author may also interest literary scholars, historians and representatives of many other areas of the humanitarian knowledge.

Most of the major languages of the world already have their own national corpora (which differ in completeness and level of scientific text processing). Today, there are representative national linguistic corpora covering a large number of different genres and registers, such as the British National Corpus (The British National Corpus, BNC, 100 million tokens), replenishing the Bank of the English language (COBUILD Bank of English, 200 million tokens).

However, there were also more specialized corpora, representing both written and oral academic speech, such as the CANCODE (Cambridge and Nottingham Corpus of Discourse in English), MIC-CASE (Michigan Corpus of Academic Spoken English), and LIBEL CASE (LIBEL Corpus of Spoken Academic English). The development of Cambridge International Corpus of North American English (CIC North America, Cambridge International Corpus), whose materials formed the basis of the grammar of oral and written English «Cambridge Grammar of English» is a great achievement of corpus linguistics. One of the first grammar of a new type, based on the corpus data and the comparison of oral and written registers of speech, a «Grammar of spoken and written English language» («Longman Grammar of Spoken and Written English», LG-SWE) publishing house «Longman», which was released in 1999. It is based on materials of Longman corpus (LSWE Corpus, Longman Spoken & Written English Corpus).

Thus, the rapid development of corpus linguistics in recent decades has led to changes in the field of foreign language teaching. The methodology of teaching a foreign language is always looking for

ways to transfer the experience by the most effective way, and is committed to finding simple and easy methods of presentation of complex and diverse linguistic material. Corpus instruments as tools of managing large volumes of data in various ways reduce the chaotic diversity of language to a more organized set of words [3].

There is no doubt that the corpus-oriented approach has great potential in the practice of teaching foreign languages. Methodologically-oriented corpus of the subject field can help in the task of selection of the content of education. Developing teaching materials with the emphasis on corpus will give the opportunity to introduce the world advanced experience of foreign language teaching with the help of corpus methods.

Thus, in the process of learning a foreign language the learner has the opportunity to resort to the use of the language data directly from the corpus. This direction of teaching obtained its name on the basis of data or data – driven learning (DDL). The basic idea is that students master the language much better when the learning process is encouraged the usage of the model of “observe – hypothesize – experiment”. Teaching with the use of the database (DDL) is an approach that provides access of the student to numerous facts of the speech use. During these training students have the opportunity to draw their own conclusions about the meaning of individual words or phrases, grammar rules on the basis of authentic language material, analyze the real facts of speech use collected in the corpus of the language. This inductive method complements the more common deductive approach, also known as “listen – practice – speak,” in which students get the knowledge about the rules and definitions from the explanations of the teacher and reference books. Such exercises are widely used in teaching with the use of computer (computer assisted language learning, CALL) [4].

Electronic corpora allow not only quick and efficient solution of the tasks facing the science, but also to put forward new ones, that virtually used to be impossible to meet because of their complexity [5]. Using electronic corpus not only to global changes in the history of the language, but also the changes less extensive, occupying one or two decades can be traced. Since entering into the texts included in the corpus are dated, it is not difficult to trace the chronology of gradual linguistic change, the emergence or extinction of certain words and designs, the change in their values, etc. [2, p. 4].

The first programs for the presentation of speech samples, programs of drawing up concordances based on linguistic databases have been developed

within the project COBUILT (Collins Birmingham University International Language Database) under the direction of G. Sinclair. Concordance is a program allowing to analyze large amounts of text in order to detect patterns of use in the language of words or expressions. Program-concordance searches the requested word in the corpus and in a new window shows the fragments of sentences from different texts that use the word or phrase. Based on the results of concordance can be understood the

meaning of words out of context, and their use in the language can be analyzed. Nowadays each corpus is equipped with such programs [6].

In the process of teaching a foreign language, depending on the educational task the students can be asked to perform tasks on the use of linguistic corpus to determine the meanings of words. In order to test this hypothesis, we construct a concordance for the word “education” based on the British National Corpus (Picture 1):

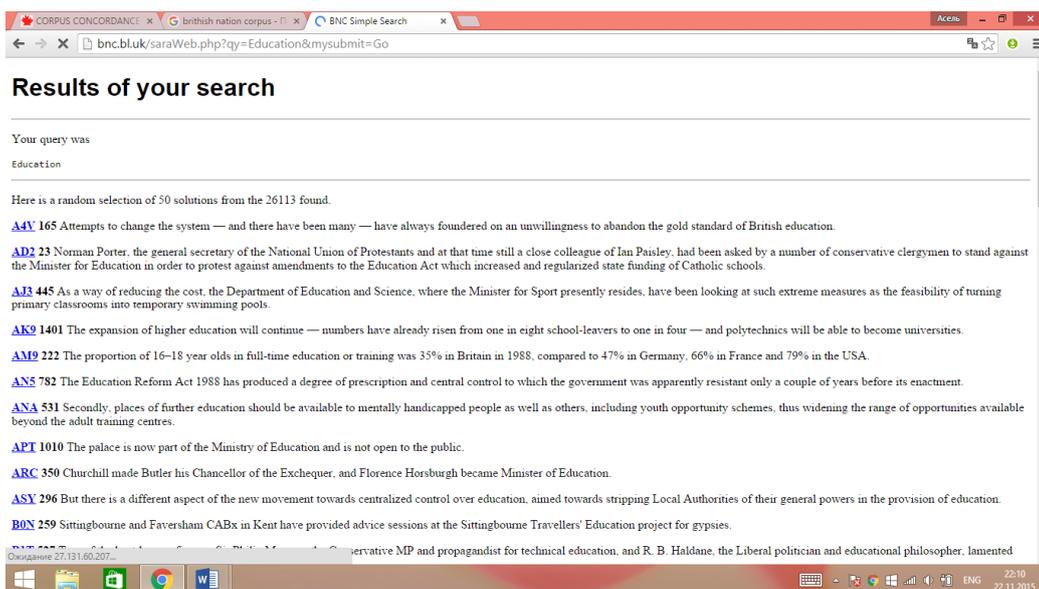


Figure 1



Figure 2 – Excerpt from the search results window of the program Corpus Concordance English (v.6.5) for the word “education”

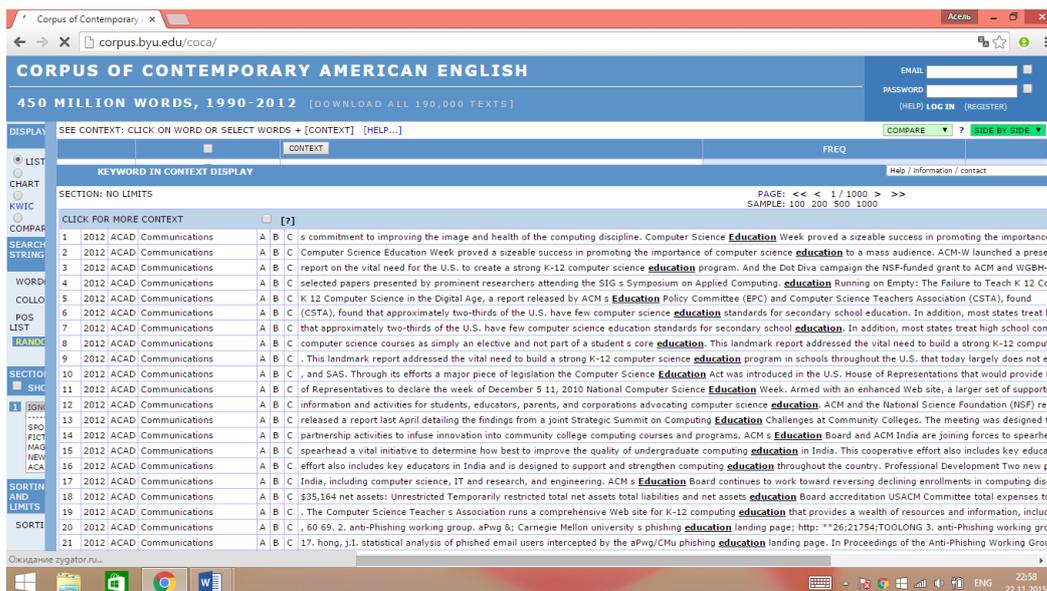


Figure 3 – Excerpt from the search results window of the Corpus of Contemporary American English for the word “education”

The task of the student is the ability to “make discoveries” in a foreign language. The function of the teacher is to provide the context in which these strategies can develop and through which the student will be able to “learn how to learn” [3, p. 173]. Corpus tasks allow the student to try on the role of the experimenter, who spends his unique study. However, the process is not necessarily limited to the computer terminal. Results of corpus searches (concordances) in printed form can be easily incorporated into handouts, manuals, or used in traditional teaching in the classroom.

Thus, the impact of modern technology on the process of teaching a foreign language, the emergence of new technological forms of language functioning lead to the rethinking of the modern definition of communicative competence. No matter what level of resources is used in teaching language, they should be directed to the formation of the communicative competence, which is understood as “the ability to solve by means of a foreign language relevant to the learners problem of communication, the skills of students to use the facts of language and speech for achieving the goals of communication” [7].

In recent years, the methodologist-researchers have used the extensive texts corpora to evaluate the reality of the language in its natural state. New linguistic experience falls on their existing learning schemes dictated by the native language and culture. The scientists have proved that the work with programs-concordances plays an important role in the

teaching a foreign language and formation of communicative competence. This belief is based on the conclusions of the theory of schemes, according to which the acquisition of language is a process associated with the formation of knowledge about reality that are important for the culture of the target language, and their correlation with relevant linguistic and pragmatic schemes.

The collection of case examples allows the student to conduct independent monitoring of real linguistic interaction prevailing in a given language community relations between form and content, on what and how to speak in specific situations of communication and how to manifest the connection between the speech situation and linguistic form[8; 9].

As examples of the work with the corpus speech patterns can be given an exercise to fill the semi-fixed variable part of an expression, which combines well with the search of the studied expression in concordance. This activity gives the students expressions that are variations of those in the text and they have to search the text to find the original expression. For example, the students can be offered the following task: Read the text (“History of education”) and highlight the group of verbs with prepositions or expressions of writing a report. Further testing of such combinations can be performed in a traditional exercise to fill the gaps.

This task gives them practice with reading skills as they will need to skim or scan the text to find the original expression. In order to organize such kind of

task, it is better use a text familiar to the students or a text that they have already been working with, and retrieve a few important semi-fixed expressions (for example, 5 – 10). After that, the variations of these expressions by changing the slot-fillers are written on the board. In order to make it easier for students, it is more appropriate to them in the same order they appear in the text. If the teacher aims to make this kind of task more challenging, he can mix them up. Then, the students are divided into two groups. Each group gets a different coloured marker. The students should look back through the text and find the original expressions. When they finish, one student from each group comes to the board and writes one of the original slot-filler above the changed one in one of the expressions. Then that student returns to his team and gives the marker to another member, who can continue the task and change another slot-filler. When all the expressions have been changed, the teacher counts how many changes have been made in each colour to determine the winning team.

Exercises relying on corpus data are usually ranked tasks on understanding the linguistic forms, namely allowing to appeal to the often incorrect ideas of students of a foreign language. Observing the lexical, grammatical units, the learner seems to read many texts that promote memorization, assimilation of form, meaning and purpose of speech. The student obtains an opportunity to observe the terms of repetition and variation of linguistic units in different contexts, corpus examples enable analyzing specific lexical and grammatical chains with a large number of words, more general ones, and short phrases as well [3, p. 174].

Professionally relevant aspects of language proficiency can also be developed with the help of the corpus. Methodologists, when considering learning through professional-oriented materials, are also turning to the concept of “scaffolding”. The author of this term, as well as the concept of “scaffolded instruction” (educational support) is considered Jerome Bruner, who described this phenomenon in education as a process of successive productions of pedagogical problems, the implementation of which is facilitated by the use of a number of teaching supports (“scaffolds”) which are eliminated as the students master the educational autonomy of action [10]. This strategy is implemented by means of using the language cliché, terminological dictionary, visualization material, mnemonics, supplying many examples of in a foreign language, which allow the student to implement the tasks independently.

There is a significant difference between simple help of the teacher and alignment of supports. Instead of providing a ready answer the teacher can support the student in finding an answer (for example, on corpus) and thus creates supports using which the student in the future solves such tasks by himself.

The next task is to find the arguments “for” (PROs) and “against” (CONs) the Internet-based education. In this case, the students can be entrusted with the choice of reading texts on the subject, which they find themselves. The texts can also be selected by the teacher. In order to ensure the information gap texts can be different. This activity teaches the students to approach the topics in a carefully “weighted” way and to avoid jumping into hasty and biased conclusions. Moreover, students can be provided the following assignment: *Read the text and write arguments, which were given by the experts on the possibilities of Internet-based education. Write down the expressions and fill in the table.*

Table 1 – The sample for the information gap filling activity

Arguments the Internet-based education	
Arguments «for»	Arguments «against»

One of the most important types of corpus for the methodology of teaching foreign languages is the so-called corpus of the learner (learner corpus), in which the foreign written and oral speech of the students is stored in a computer in the form of a database and can be analyzed in the learning process by both the learners and their teachers. This type of corpus makes it possible to track the progress in language learning, to identify common errors and gaps in the knowledge of students.

The international learner corpus of English (The International Corpus of Learner English) includes around 2 million of words of written speech of English language learners who are carriers of 19 different languages. Written speech in the corpus is presented in the form of students essays, learning English as a foreign language at an advanced level. The corpus can also be used to control the level of mastering the material. With the help of the corpus not only lexical and grammatical assignments, but also tests and projects can be designed.

Conclusion

Concluding the above, it should be emphasized that special place in the teaching of foreign languages belongs to different linguistic (specifically marked) corpora. Exactly the corpus technologies are today considered as the basis of innovation in the teaching of foreign languages that is proven in numerous publications. Corpora approach allows to put into practice a differentiated approach to teaching a foreign language. This is the main trend in the frames of professionally-oriented teaching. It is needed to change the approach of using the corpus in the teaching of foreign languages: the corpus must be considered not

only as a reference, available for the students, but also as a means of learning the language.

In this regard, a special role in the implementation of information and computer technologies in teaching foreign languages play multi-purpose and multi-national resources of National corpora. However, for the complete application of the potential of the corpus instruments in teaching a foreign language, and in particular professionally-oriented language it is necessary to create a special, themed buildings, to build a dialogue between the creators of corpora and teachers, to develop teachers' skills of using corpora and help them develop corpora that meet their special needs.

References

- 1 Rykov V.V. Pragmatically oriented corpus//Tverskoy linguistic meridian. Theoretical compilation / Ed. by Susov I.P. Tver. – 1999. – Vol. 3. – P. 89-96.
- 2 Kolokolnikova M. Y. Discourse analysis and corpus analysis in research in the field of historical lexicology (Proceedings of the University of Saratov. The new series. – A series of Philology. Journalism. – T. 10. – Vol. 2. – P. 3-6)
- 3 Gorina O.G. Psychological principles of foreign language teaching profession on the basis of language corpus / Vestnik of LSU named after A.S. Pushkin. – 2013. – № 2. – P. 172- 179.
- 4 Nagel O. Corpus Linguistics and its use in computerized language teaching / Journal “Language and culture”. – 2008. – №4. – P. 53- 5.
- 5 Plungyan V.A. The corpus as an instrument and as an ideology: Some lessons of the modern corpus linguistics / Russian language in a scientific light. – 2008. – № 2 (16). – P. 7-20.
- 6 Sysoev P.V. Linguistic corpus in the methodology of teaching foreign languages / Journal “Language and Culture”. – 2010. – № 1. – P. 99-111.
- 7 Azimov E.G., Shchukin A.N. New Dictionary of methodological terms and concepts (the theory and practice of language teaching) / E.G. Azimov, A.N. Shchukin. – M.: Publishing house IKAR, 2009. – 448 p.
- 8 Bernadini S. Corpora in the classroom: An overview and some reflections on future developments // Sinclair J.McH. (ed.) How to use corpora in language teaching. – Amsterdam: Benjamins, 2004. – 308 p.
- 9 Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating language structure and use. – Cambridge: Cambridge Univ. Press, 1998. – 300 p.
- 10 Bruner J. S. The ontogenesis of speech acts // Journal of Child Language. – 2. – S. 1-10. – 1975.