🔓 **Research Article**

USING CORPUS LINGUISTICS IN TEACHING LEISURE READING

Khursantosh Anvar kizi Tolmasova

Master degree student, Uzbekistan State University of World Languages, Uzbekistan

ABSTRACT

The use of methods of applied linguistics in teaching a foreign language is considered. The potential of learning on the basis of corpus language material is analyzed. An analysis of such corpus linguistics methods as information retrieval (IR), data-driven learning, text searches in large-scale corpora (concordances), natural language processing methods (natural language processing) is presented. NLP)). In addition, the role and importance of the corps in leisure reading was studied.

KEYWORDS

Corpus linguistics, language learning, concordance, leisure time, leisure reading.

INTRODUCTION

One of the most important problems noted in the process of teaching foreign languages is the lack of adequate pedagogical text materials and relevant vocabularies, while the daily workload of teachers in the current preparation of new motivating materials for classes continues to be consistently high. In this area, corpus linguistics methods can be of great help: automated information extraction, data-based learning, text searches in large-scale corpora using natural language processing methods.

The concept of "corpus of texts", on the basis of which corpus linguistics develops, is increasingly included in the scientific circulation of linguists. A corpus of texts

is usually understood as "a unified, structured and marked array of language (speech) data in electronic form, intended for certain philological and, more broadly, humanitarian research" [1.p.52]. A corpus of texts can also be considered "as a complexly organized ontology of speech activity, reflecting all the genre diversity of the kind of literature represented in it" [2.p.18]. As a complex verbal unity, a corpus includes a variety of information not only about the composition and structure of its speech material, but also other formalized methods of its presentation (word indexing, morphological information, etc.). Consequently, it can also be considered as a specially constructed semiotic system [2.p. 21].

MAIN PART

Corpus linguistics in its modern sense as a science dealing with the creation and analysis of text corpora originated in the United States and Western Europe in the late 1960s. With the growth of the capabilities of modern computer technology, since the mid-1980s. corpus linguistics is rapidly developing, corpus projects of various scales in different languages and for various purposes began to appear actively. Achievements in the field of corpus linguistics are widely used in the process of teaching linguistics. In the world's leading universities, 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 [3].

Automated Information Extraction

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 [3], the concept of communicative competence should be defined as corresponding to the natural conventions (both grammatically and stylistically) of using the language in various communicative situations, including those that involve technology. 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 [3]. There is ample 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 hammered into the head with rules (which, in fact, often turn out to be inadequate to the real state of the language).). In addition, the experience of collective "discovery" of the language in the environment of the student group with the help of these new techniques brings a valuable element of discussion and joint generalizations.

In the last 10–15 years, researchers in foreign language teaching methods have used extensive corpora of texts to assess the realities of language in its natural state. These corpora of texts have largely influenced the improvement of the quality level of the published language manuals. Instead of traditional prescriptivist guidelines on how to use language correctly, 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 the 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 stable results are obtained when these new methods are introduced step by step into the learning process to ensure efficiency 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 independently formulate research problems. 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 pointed out in the textbook. methodological aids. Now, having been identified with the help of corpus methods, these features can be taken into account in the teaching process and in the development of modern educational and methodological complexes.

In the last decade, another new and extremely promising direction has emerged 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). It is based on solid empirical evidence that students can learn a language much more effectively when the use of the observe-hypothesize-experiment model is encouraged in the learning process. When they have the opportunity to draw their own conclusions about the meanings of words, phrases, grammatical rules based on authentic language material. This inductive method complements the more common deductive approach, also known as listen-practice-speak, in which students gain knowledge of rules and definitions from instructor explanations and reference books. The process is not necessarily limited to a computer terminal. The results of corpus searches (concordances) in printed form can be easily incorporated into handouts, teaching aids, etc. and used in the process of traditional teaching in the classroom. In addition, the formation of specialized text corpora on a hard disk is quite common. Modern tools make it possible to quickly form a very extensive (several tens of millions of words) corpus of texts on almost any topic, and anyone who has basic skills in working with a personal computer and the Internet can do this.

In computer-assisted language learning (CLA) (computer-assisted language learning (CALL)) there are currently active developments in the field of so-called language technologies using natural language processing (NLP) methods to create "smart methods. In such innovative methods, linguistic analysis and modeling of student reactions are used to analyze and adequately assess the so-called free-form linguistic input by the student, in other words, answers to open questions and even assessment free-form essays. Computer-assisted language learning moves away from the main, traditional, often decontextualized ways of presenting material, and focuses on those activities that stimulate or, one might even say, require an element of creativity. Karol Chapel in his article "Practical application of high technologies in teaching" notes that, for example, "within the course of grammatical analysis, we involve students in the analysis of a computer case, which, firstly, affects their perception of the analysis itself, and secondly, on their ability to independently conduct such an analysis and, thirdly, how they themselves will teach grammar" [4. S. 6]. The content of methodological materials and the

practice of teaching foreign languages and linguistics, both in our country and elsewhere, tend to reflect the division that currently exists between the empirical and rationalist approaches in the humanities, in particular in linguistics. Many textbooks abound with artificial examples, while grammatical and stylistic descriptions are based more on the intuition of their compilers or on secondary sources. However, there are a small number of tutorials that are based on an explicitly empirical approach and use examples and descriptions drawn from corpora of actually used language tools.

Natural language means (Naturally-occurring Language) are extremely important in the process of teaching foreign languages, because. provide students with the opportunity to deal with the sentences that they will meet in a real situation of communication in a foreign language. Students who study on the basis of conservative educational materials with traditional descriptions of written syntax such as Mary puts her book on the table are sometimes not ready to perceive and correctly analyze natural speech, replete with complex sentences such as The government has welcomed a report by an Australian royal commission on the effects of Britain's atomic bomb testing program in the Australian desert in the fifties and early sixties (from Corpus of Spoken English) [5]. In addition to direct application in the process of teaching a foreign language based on a natural empirical approach, the corpus as a method can be used to critically evaluate existing teaching materials. Thus, Kennedy, Holmes, Mindt analyzed the coverage of various aspects of English grammar in existing traditional textbooks, using the method of comparative analysis of the corresponding constructions and vocabulary in textbooks and in the standard English corpus. Most of these studies have found that there are significant discrepancies between

what textbooks prescribe and how the language is actually used by speakers, as evidenced by the colloquial corpus. Sometimes, in some textbooks, certain aspects of the use of language and its stylistic features are brought to the fore, which turn out to be peripheral and less typical, while the more central ones are ignored. The general conclusion of these studies is that traditional prescriptivist teaching materials, not based on empirical methods of selecting and analyzing language material, are inadequate to the real natural state of the language and the realities of its typical use, and that corpus linguistics methods should be mandatory in the development and evaluation the effectiveness of teaching materials and teaching aids so that the most common uses receive priority attention, and peripheral uses take their appropriate place [6].

Thus, the role of the corpus approach in such a field as computerized teaching of foreign languages is central. Recent research by Lancaster University on software for teaching undergraduate students grammar and basic grammatical analysis has shown that programs such as Cytor and similar programs that are created quite easily on the basis of a corpus of texts annotated either by parts of speech or by grammatical - skim / syntactic roles, are extremely effective and provide the right degree of interactivity along with autonomy [3]. When receiving the 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 of information about the frequency of use of a particular lexical unit or the frequency of joint use of examples (collocation).

In the course of experiments conducted by McEnry, Baker, Wilson in order to determine the effectiveness of this technique, the degree of assimilation of

knowledge of the parts of speech was established among students trained according to the new corpus method, and in the control group of students trained according to the traditional lecture method [8]. In general, students who studied using corpus software consistently demonstrated higher performance than students in the control group. Considering the questions that students face daily at various stages of mastering a foreign language, it can be noted that the methods of corpus linguistics, such as the search for concordances, are optimally adapted to provide an interesting and independent search for answers, during which the student has the opportunity to obtain the required information. on particular issues of the language being studied, and along the way to acquire an idea of the real natural state of the language, its historical, geographical and social variation, speech registers, and genre diversity. It is important to note that text corpora, like modern corpus linguistics, as a scientific methodology and a branch of linguistics, are very young and unevenly developed. For some languages, such as English, German, Finnish or Japanese, extensive and representative annotated corpora have been created, while for other languages, including Russian, the process of creating full-fledged corpora that meet the basic requirements is going through a period becoming. The situation with multilingual (parallel or comparative) corpora is even more complex, which significantly limits the incredible potential of corpus methods in the field of linguistic research, language technology and teaching foreign languages.

CONCLUSION

Of the various mono and multilingual corpus projects being implemented at the moment, most of them exist on the basis of universities. For example, the Comprehensive Development Program for 2004–2005 was successfully implemented on the basis of Tomsk Polytechnic University. "Corpus of texts as a kind of language resources", developed by associate professors N.N. Shalamova and A.Yu. Filchenko [9]. In the course of this work in 2004, three final qualifying works were completed within the framework of a single thematic area of "Corpus of engineering texts in English" for its subsequent use in the practice of teaching English and research work at TPU. Corpus methods have proven themselves in the world practice of linguistic research and teaching of foreign languages, including profile-oriented, as highly effective innovative additions to traditional educational technologies. These methods combine such aspects as interdisciplinarity, empirical adequacy, authenticity, flexibility and adaptation to specific tasks and target groups, the possibility of independent work of the student, the use of the "discovery" method in teaching.

REFERENCES

1. Захаров В.П. Поисковые системы Интернета как инструмент лингвистических исследований // Русский язык в Интернете. Казань, 2003.

2. Рыков В.В. Корпус текстов как новый тип словесного единства // Труды Междунар. семинара «Диалог-2003». М.: Наука, 2003. С. 15–23.

3. Information and Communications Technology for Language Teachers. Введение в прикладное значение корпуса. Режим доступа: http://www.ict4lt.org/en/en_mod2-4.htm

4. Carol A. Chapelle // Essential teacher. 2003. Vol. 9. P. 5–11.

5. British National Corpus. Режим доступа: http://sara.natcorp.ox.ac.uk/lookup.html

6. National Foreign Language Resource Center. Режим доступа: http://www.nflrc.hawaii.edu/project/2006EmoreInfo.html

7. Простой пример конкорданса на сайте корпуса Cobuild. Режим доступа: http://titania.cobuild.collins.co.uk/form.html

8. McEnery T., Wilson A. Corpus Linguistics. Edinburgh, 1997.

9. Шаламова Н.Н., Фильченко А.Ю. Корпусная лингвистика и её использование в профильно-ориентированном преподавании иностранных языков. Томск: ТПУ, 2004.