Methodological Characteristics of Corpus Technologies in Teaching Foreign Language

Radjabova Gulnoza Giyosiddinovna
Senior lecturer, Uzbek State World Languages University, Uzbekistan

Abstract: Corpus technologies are one of the Internet-based technologies, which help to develop students' communicative competence in foreign language interaction. On the basis of the analysis of the methodological literature on the use of corpus technologies in teaching foreign language, didactic and methodological functions such as multilingualism, conceptuality and authenticity of corpus technologies are distinguished. These functions reveal and ground a number of psychological and pedagogical conditions to develop and enhance the methods of teaching students written and oral communication on the basis of the corpus technologies.

Keywords: concordances; corpus manager; corpus technologies; methodological characteristics; teaching foreign language;

INTRODUCTION

Linguistic corpus texts and corpus technologies have been a subject of research of many and foreign and Russian researchers [1-6]. The review of their works shows that different authors have different approaches to the problem of teaching grammar or vocabulary on the basis of different corpora texts.

However, despite having and implementing different approaches, their views do not contradict each other. On the contrary, each work has highlighted a previously undiscovered aspect of learning, gradually contributing to the development of the theory of corpus-based approach to language learning. In their works, the authors use different terms and give different definitions to corpus and corpus technologies. Rykov V.V. uses the term “corpus of texts” which he defines as “a set of texts organized in a certain way” [7]. Sysoev P.V. interprets the term “linguistic corpus” as “an array of texts collected into a unified system according to certain features (language, genre, time of creation, author, etc.) and provided with a search engine” [4, p. 99]. Chernekova T.A. also uses the term “linguistic corpus” and defines it as a “collection of texts presented as an electronic media” [8, c. 3]. I.V. Deryabina suggests that the term “linguistic corpus” should be understood as “an electronic reference system which allows analyzing texts of various genres and distinguishing the use of grammatical different grammatical patterns” [2, c. 3]. E.A. Ryazanova interprets the notion of a “linguistic corpus” as “a collection of texts on electronic media, selected according to external criteria (genre, purpose, industry, language)” [3, c. 3]. Sysoev P.V. and Kokoreva A.A. introduce the term of “corpus of parallel texts” which “consist of a source text in one language, its translation into other languages” [5, p. 115]. Some authors use the term “corpus technology” in their research, identifying it with the widely spread notion of “linguistic corpus” [6].

Analysis of these definitions of the term “linguistic corpus” reveals the fact that scholars share a common view that a corpus of texts is a collection of texts assembled into a coherent system, classified according to certain characteristics.

1. METHODOLOGY

In this article we will use the term “corpus technologies” as a collection of texts united in a single electronic system according to specific typological features. The first electronic,
linguistic corpus was created at Brown University (the USA) by the scientists W. Francis and H. Kucher. That corpus consisted of half a thousand literary texts of the most popular genres of prose published in the United States [7].

The selection of texts was based on the following features:

A. The author of the text must be a native speaker of an American English;
B. The corpus included works first published in 1961;
C. There should be equal representation of different genres of texts;
D. The use of tags and labels was necessary to classify and retrieve data [4].

The appearance of the first linguistic corpus draws the attention of philologists, linguists, and educators, which spurred the development of corpus linguistics. At the moment there are several types’ electronic linguistic corpora depending on different typological features [4; 5; 7].

According to the type of linguistic data, corpora can include written texts only, spoken texts only, and both written and spoken texts (mixed corpora). While in terms of the language used in corpus, we can distinguish monolingual, bilingual and multilingual corpora which have been developed by different universities and internet platforms and corporations. Considering the didactic and methodological characteristics of corpus technologies, P.V. Sysoyev singled out the following features:

- multilevelness and plurilingualism of information provided;
- Variety of functional types;
- hypertext structure of documents;
- conceptuality of information;
- the ability to automate the processes of informational and methodological support and organizing the management and control of the students’ educational activity;
- building students’ individual educational trajectory;
- implementation of pedagogical technology “cooperative learning”;
- Development of the skills of independent learners; [9].

Aforementioned characteristics of corpus technologies allow implementing them in teaching foreign language. In this article we singled out two corpus technologies which will be described below.

1.1 Concordance as one of corpus technologies

One of the modern and up-to-date trends in foreign language teaching methodology is teaching different aspects of language on the basis of corpora (corpus-based teaching and learning). Corpus-based technologies are used in teaching vocabulary, grammar, translation, etc. In this article we focus on such corpus technologies as concordances and corpus managers, which refer to linguistic retrieval systems.

In the simplest sense, a concordance is a list of all uses of a given linguistic expression in context with references to the source (the latter is not a prerequisite). In this sense, the term is widely used in corpus linguistics. [10]. or in other words, a concordance is a list context, where the searched unit is presented in its lexical environment and characterized by a set of statistic data [11]. Concordances are special programs designed to process text with a particular linguistic task, aiming to search morphemes, words, and word combinations in a context. Concordance will display all words or examples of a given grammatical structure together with the context. The information obtained is useful both for the teacher and student as it provides numerous examples of both grammatical and lexical word forms. In addition, due to representativeness (one of the main advantageous properties of the linguistic corpus), examples presented in the concordance
allow learners to conduct their own research by tracing the lexical and grammatical phenomena in authentic language contexts. Thus, the theoretical knowledge acquired is easier to consolidate in practice, carried out from real life or authentic contexts.

In addition, concordances can be used in foreign language teaching for self-monitoring as well as in translation. For example, if a student has doubts about the correctness of an expression or phrase, due to the data obtained from corpora, he or she can determine whether it is valid for the use in a certain context. In the case of incorrect usage of a word, there will not be examples of contexts and, on the contrary, their occurrence in large quantities indicates the correctness of the expression or word combination searched for. For example, while teaching grammar of a foreign language, the student may be asked to find and analyze the use of complex tenses (e.g., Perfect tense), modal verbs and their role in a sentence, the place of adverbs in a sentence, etc. While teaching vocabulary, for example, it will be beneficial for students to find and explain some examples of words that often cause difficulties in their usage such as MAKE / DO, RISE / RAISE, TELL / SAY, LIE / LAY, etc.

Corpus technologies make it possible to research syntax, for example, to investigate the punctuation of a particular language and identify the differences compared to the native language. In general, the following tasks which are based on the usage of corpus technologies may be offered to learners:

- comparison of different usage of a single word;
- keyword analysis;
- analysis of the frequency of words and word combinations;
- research of phrases and idioms;
- search of terminology translation;
- Creating word lists (which can be used in academic writing; [10].

B. P. Zakharov offers complete and specific list of functions performed by such corpus technologies as concordance and corpus manager:

- searching for particular word forms and outputting the results in the form of concordance;
- searching for word forms in lemmas or key words searching;
- searching for a group of word forms as a separable or inseparable unity;
- searching for word form by a set of morphological features;
- searching for information about the origin of word forms;
- searching for statistical data of the frequency usage of word within the context;

Concordance based on a corpus of texts is called concordance program or concordance maker [11]. It allows obtaining the frequency of a particular linguistic unit within the context. It also allows to sort contexts by keywords (in their original form) or by the word combinations within their nearest context.

1.2 Corpus manager

As mentioned above, concordances refer to linguistic search engines, similar function is fulfilled by corpus managers. In this section, we discuss the definition of the term “corpus manager”. So, according to V. P. Zakharov’s definition, a corpus manager is “a specialized text management system for search and retrieval of texts and includes software tools to search texts and obtain statistical information which is presented in user convenient form [11].

Corpus manager has the following characteristics:
Search for word sketches – provides the outline of word’s grammatical and collocation behavior;

Two-word comparison – provides comparison and contrast of two word by analyzing their collocations;

Search for concordance – reveals examples of a word form, lemma, word phrase, tag or complex structure;

Search for collocation – analyzes the co-occurrence of words showing their frequency usage;

Extraction of key words – extracts automatically multi-word terms and key words from texts;

Diachronic analysis – displays changes in frequency of the usage which words underwent throughout their development;

Parallel bilingual facilities – shows translation examples;

Based on the abovementioned, we can state that the corpus manager has more options and opportunities than the concordance. For example, the tasks of a simple concordance include building concordances of individual words, phrases, parts of words, punctuation marks, etc. in a contextual environment. Whereas more sophisticated programs, called corpus managers, are capable of constructing complete concordances that include not only words but other corpus elements as well. Corpus managers are designed to handle phenomena such as lemma and morphological characteristics of words, position of a word in a sentence and in the structure of marked text (HTML, XML), bibliographic and typological characteristics of a document from which the context is selected (author, title, source, year of publication, text type, etc.), statistical data and many others. Some of the best known “corpus manager” are Bonito, CQP, DDC, WebCorp, Xaira.

2. DISCUSSION

After dealing with the theoretical aspect of the notions discussed in this article, we suggest that we move on to the issue of practical part of using concordance programs and corpus managers in practice. To illustrate the examples of using concordances we will concentrate on the program that presents corpus-based concordances (Corpus-based concordances) as well as Multi_Concordance [12]. Corpus-based concordances select parameters such as frequency of usage (from the most numerous to the most restricted), page configuration (choosing the number of lines, context and keyword location). For example, while searching for the key word “punishment” with some parameters, we got an illustration of its usage in various contexts (Table 1).

This public interest: # Because the damages award?? is imposed as a punishment for the violation of a public law, the Legislature may adjust its amount to

# Is it not completely ironic that those who really are insane reach for psychological punishment of others who are not insane? The Nazis did this. The Soviets did

left -- literally -- on his or her own. # Rewards go further than punishment in building human cooperation and benefiting the common good, according to research published in

no one can argue against him. However, they both should get the same punishment for deliberately wrecking a driver. # Last year, Kyle Busch deliberately wrecked Ron

Or unto governors, as unto them that are sent by him for the punishment of evildoers, and for the praise of them that do well. 15.

and the vacations we went on, I wonder if my parents were gluttons for punishment. My father, the high school math teacher, and my mother would put

Some of them are now able to express themselves without fear, especially now
corporal punishment has been banned in Burkina Faso's primary schools. Children's rights are taught among other charges, and fined. He was also given an order of punishment for embezzlement. They would have had a fourth child three years ago, but

Table 1. The word “punishment” in the context

Multicondense also offers an excellent, in our opinion, possibility of automatic task-builder option (interactive worksheet). By entering a word or phrase (it is also possible to enter a whole block of lexical units), the researcher will get ready-made tasks where some key words can be skipped and students can insert them themselves, practicing their use (e.g., using a word or phrase in the pronunciation of a word or phrase). (Table 2).

Table 2. Worksheet for students to insert appropriate word in the context. (key word “punishment”).

So, as mentioned above, if the researchers choose the simple mode for creating a concordance based on corpus, there will be the number of lines, each line containing a word, phrase, or structure educators are searching for. For example, when teaching English language vocabulary to law students, a concordance can be used both to illustrate examples of the use of the lexical units being studied (Table 1), and to practice their consolidation by means of exercises and tests, in this case - by means of an interactive quiz, which can be done either online or in the form of multiple-choice test, compiled by the teacher, using the tasks automatically generated by the program - concordance (Table 2). Regular assignments with missing lexical units can also be obtained by means of a multi-concordance (Table 2).

3. RESULTS

The above examples illustrate how researcher or educator can implement corpus technologies in teaching foreign language. It should be noted that in addition to the analysis of the context according to key words, the concordances also help students to identify the usage of words with other words, determine the pre-vocabulary with which they are most frequently used. It is also possible to introduce such a parameter as single-nouns, which would make it easier for students to find and master not only the key word, but also all its cognates.

In the case of grammatical constructions, corpus technologies are similarly helpful not only to
analyze the usage of different grammatical constructions, but also draw practical considerations from the theoretical point of view. In addition, corpus technologies allow to work with contextual environment of the lexical units searching for the definitions of the given terms. The following didactic and methodological features of corpus technologies will form the basis for identifying and justifying the psychological and pedagogical conditions of teaching foreign language. (Table 3.)

<table>
<thead>
<tr>
<th><strong>Multilingualism</strong></th>
<th>Modern electronic corpora contain texts in several languages and are often called “corpora of parallel texts”. They can be used to develop lexical skills of learners.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contextuality</strong></td>
<td>Unlike online dictionaries, corpora show the requested word or collocation in a compact (a few words before and after the requested word) or extended context (up to several sentences before and after the requested word) Corpus technology allows to more accurately determining the meaning of a requested word from multiple contexts of use. Besides, search results with the queried words in context will allow to use materials for creation of exercises to control of formation of students’ lexical competence</td>
</tr>
<tr>
<td><strong>Dynamism</strong></td>
<td>Modern electronic corpora are constantly updated and supplemented with new texts, which allows determining the frequency of use of the searched word more precisely and, if necessary, revealing more up-to-date variants of its usage and mastering the most modern meanings of words and the latest word usage in speech contexts.</td>
</tr>
<tr>
<td><strong>Diversity of functional types</strong></td>
<td>Modern electronic corpora contain texts of different genres: articles of popular magazines, articles from academic journals, literary works, interviews, legal documents, texts of advertisements and announcements, etc.</td>
</tr>
<tr>
<td><strong>Specific pattern search</strong></td>
<td>There is the possibility of conducting a search in a corpus According to a preset pattern (a specific word, a phrase, a lemma (all possible word forms), words with wildcards (un*ly), adjective/verb + searched word, etc.). Corpus technologies allow carrying out point wise search for a particular word, phrase, lemma, collocation for correct determination of its meaning. In addition, corpus technologies can serve as a means and a resource of student’ language research work.</td>
</tr>
</tbody>
</table>

Table 3. Didactic and methodological characteristics of corpus technologies.

4. CONCLUSION AND RECOMMENDATIONS

To summarize all of the above, it should be noted that corpus technologies (concordances, corpus managers), which have been recently developed, offer both simple as well as complex functions used for linguistic analysis. The use of corpus technologies reveals new perspectives in foreign language teaching, contributing to a more effective process of teaching and learning process because it facilitates the work of both teachers and students. In particular, both concordances and corpus managers can be used effectively in foreign language teaching by developing students’ ability to independently search for specific word forms, groups of word forms as broken or unbroken Syntegra, word forms by a set of morphological features. In addition, corpora display information about origin, text type, register; speakers’ background (social, gender, age).

Corpus technologies are indispensable in students’ implementation of linguistic analysis, as they provide the most objective, representative and comprehensive meta-textual information about the frequency of various linguistic elements, such as stable word combinations, the frequency of word collocations, grammatical and lexical functions of words, and the connection of words with other lexemes. All the above-mentioned show the prospects and effectiveness of using corpus technologies in the educational framework.
REFERENCES:


10. Corpus-based Concordancers. URL: http://www.lextutor.ca/concordancers/concord_e.html