



# TJAS

**Thematic Journal of Applied Sciences**

informing scientific practices around the world  
through research and development

## Thematic Journal of Applied Sciences

Volume 3, No. 4, July 2023

Internet address: <http://ejournals.id/index.php/TJAS/issue/archive>

E-mail: [info@ejournals.id](mailto:info@ejournals.id)

Published by ejournals PVT LTD

Issued Bimonthly

Chief editorS.

**G. Ahmed**

*Professor of Computational Mathematics and Numerical Analysis Faculty of Engineering, Zagazig University, Zagazig, Egypt, P. O. Box 44519*

*Requirements for the authors.*

*The manuscript authors must provide reliable results of the work done, as well as an objective judgment on the significance of the study. The data underlying the work should be presented accurately, without errors. The work should contain enough details and bibliographic references for possible reproduction. False or knowingly erroneous statements are perceived as unethical behavior and unacceptable.*

*Authors should make sure that the original work is submitted and, if other authors' works or claims are used, provide appropriate bibliographic references or citations. Plagiarism can exist in many forms - from representing someone else's work as copyright to copying or paraphrasing significant parts of another's work without attribution, as well as claiming one's rights to the results of another's research. Plagiarism in all forms constitutes unethical acts and is unacceptable. Responsibility for plagiarism is entirely on the shoulders of the authors.*

*Significant errors in published works. If the author detects significant errors or inaccuracies in the publication, the author must inform the editor of the journal or the publisher about this and interact with them in order to remove the publication as soon as possible or correct errors. If the editor or publisher has received information from a third party that the publication contains significant errors, the author must withdraw the work or correct the errors as soon as possible.*

**OPEN ACCESS**

Copyright © 2023 by Thematics Journals of Applied Sciences

## CHIEF EDITOR

**S.G. Ahmed**

*Professor of Computational Mathematics and Numerical Analysis Faculty  
of Engineering, Zagazig University, Zagazig, Egypt, P. O. Box 44519*

## EDITORIAL BOARD

**Yu Li**

*Wuhan University of  
Technology, China*

**Seung Man Yu**

*Seoul National University of Science  
and Technology, South Korea*

**Syed Saeid Rahimian Kolor**

*Universiti Teknologi Malaysia,  
Malaysia*

**Eko Susanto**

*Menegment of journal Indonesia*

**Siti Mazlina Mustapa Kamal**

*Universiti Putra Malaysia, Malaysia*

ELSEVIER



SSRN

Universal  
Impact Factor



## USING CORPUS LINGUISTICS TECHNOLOGIES IN DEVELOPING STUDENTS' LEXICAL SKILLS

**Sharipov Feruzjon Isor o'gli**

Teacher, Department of Methods of Teaching English  
Uzbekistan State University of World Languages  
Tashkent, Republic of Uzbekistan

*Abstract: It is perfectly reasonable that the achievements and developments of corpus linguistics, the emergence of which was a significant event in the history of science, should not remain within the framework of one science of linguistics. One of the first results of the integration of corpus methods was a change in the technology of creating dictionaries, which was followed by a significant transformation of educational materials. The article deals with the issues based on analyzing efficiency of using corpus linguistics technologies in developing students' lexical skills.*

*Keywords: educational materials, lexical skills, corpus linguistics technologies, creating dictionaries.*

### **Introduction**

The rapid development of corpus linguistics in recent decades has led to fundamental changes in the field of lexicography, followed by changes in the field of teaching a foreign language, and also "the contours of a new language model are outlined, which in a number of significant respects differs from the usual models that have developed in the last quarter of the XX century". These changes have become so radical that they are often described as revolutionary.

In this study, we comprehend a linguistic corpus as a set of texts collected in accordance with certain principles, marked up according to a certain standard and provided with a specialized search engine. Today, with the help of the corpus, we can solve problems that were previously impossible. In addition, not only the efficiency and speed of processing language data, but also their reliability have increased many times over. In the "pre-corpus era", the accepted alternative to exact methods was intuition, which, however, was not always sufficient proof. The presence of a corpus and the possibility of collecting and processing large arrays of linguistic data ensure the testability of hypotheses, assumptions, and conclusions.

### **Discussions**

It is now known that large, diversified and representative annotated corpora have been created for languages such as English. However, no corpus is able to serve all purposes at once. Therefore, the creation of specialized methodologically oriented language corpora on the basis of universities is becoming a daily practice. A methodically oriented corpus of the subject area can help in solving the problems of selecting the content of training, including professional lexical minima. In addition, the development of educational materials based on the corpus will allow introducing the world's best practices in teaching a foreign language using corpus technologies. An example of methodically oriented corpora is the LINGVATORIUM resource, created by the staff of the Center for Linguistic Research. A.A. Khudyakov at St. Petersburg State University of Economics, a corpus of engineering texts based on Tomsk Polytechnic University (Shalamova,

Filchenko 2004), a multimedia corpus of oral texts and exercises ELISA (English Language Interview Corpus as a Second-Language Application, ([http:// www.uni-tuebingen.de/elisa\\_index.html](http://www.uni-tuebingen.de/elisa_index.html)), created at the University of Tübingen, Germany, the student error corpus collected at the A.I. Herzen RSPU (Kamshilova, Kolina, Nikolaeva 2008), etc. that corpus projects are still gaining their critical mass, so they often remain on the websites of universities and are not always available to the general public.

However, one cannot but agree with the opinion of E.N. Solovova that with a sufficiently broad coverage at the level of theory, the introduction of new information technologies and more efficient use of old techniques is unreasonably slow. The author sees the reason in the lack of professional mobility and flexibility of teachers, the lack of readiness to move away from stereotyped actions (Solovova 2010). In domestic practice, insufficient attention is paid to the development of teachers' skills of independent work with the corpus, evaluation of corpus data, as well as the possibilities of using corpus resources. On the one hand, it is obvious that the case-oriented approach has great potential in teaching a foreign language, however, the difficulties associated with mastering new computer technologies seem insurmountable. Thus, teachers need to share existing experience and some training in order to use the linguodidactic potential of the corpus.

Let us consider how the point of view on the language in general and on teaching a foreign language in particular has been transformed with the advent of corpora and the emergence of the ability to test hypotheses on large arrays of language data.

The main attention of researchers has focused on the knowledge of the phenomena of human speech, the existence of the word in the context, rather than abstract logical reasoning. The priority was not only to study the laws of the language, but also to test them in practice. The compatibility of words began to be studied on the basis of the dialectical unity of collocation and colligation (for details, see Section 1.2.3), since the correct construction of speech is impossible if only grammatical or colligative correctness is observed (Gvishiani 1979). The idea of the use of the word is realized by the speaker under the influence

"the actual functioning of a certain number of phrases with this word." From the point of view of the linguistic community, these phrases correspond to all the criteria of the linguistic norm (including the sociolinguistic aspect), i.e. their linguistic status is recognized (Ibid.: p. 172).

One of the consequences of the transformed views on the language was the emergence of a lexical approach to teaching, which recognizes lexical errors (violation of usage and compatibility) as important as grammatical ones. In this regard, a decisive shift is proposed in teaching towards the study of word combinations, collocations, chunks and a departure from the study of grammatical structure as a formula in which words should be substituted, without taking into account their functioning in speech. The basis for this shift is research in the field of applied linguistics and, especially, in the field of corpus lexicography. The importance and role of the lexical side of speech has increased, lexically-oriented approaches to the study, teaching of the language and to the description of the nature of speech production have been outlined: "...many syntactic phenomena began to be considered as a projection of the general laws of the functioning of the lexicon."

In fact, when teaching the use of LU, it is proposed to change the vector of movement: not from a word to a sentence according to the rules, but from the whole "piece" of the language (chunk) that has developed in the speech community, in the direction of its adjustment to the speech intention according to grammatical rules or the

enlargement of the statement by attracting new chunks and "sticking them together" with the help of grammar rules. In our view, such speech production implies a high degree of automation and flexibility of grammatical skills and rather a virtuoso possession of grammar, rather than its neglect. Such mastery of the language also realizes the dialectical unity of colligation and collocation, i.e. consistency with empirical data on the functioning of LU in speech. It should also be noted the point of view of S.F. Shatilov about the coherence and continuity of the lexical and grammatical skill, the introduction of words in the context, which is reflected in the lexical approach.

Determining the place of the corpus in the system of modern approaches to teaching, it should be noted the humanistic approach, which implies a focus on the personality of the student. Learning using corpus tools is also carried out within the framework of an inductive approach, which implies a direction from the particular to the general, from the functioning of a particular grammatical or lexical phenomenon in various contexts to understanding its meaning and form. Despite the fact that in corpus analysis the initial element is the form, as the only unit of automated search, the organization of exercises on corpus material can be carried out along an inductive path. The inductive approach to learning in the modern methodology is implemented through a series of exercises on understanding the linguistic form of the utterance (consciousness-raising activities, awareness-raising activities, discovery techniques, observation-based activities, noticing) within the framework of the consciously-oriented approach.

Such exercises are widely used in computer-assisted language learning (CALL) as well as in data-driven learning (DDL).

Thus, the undoubted advantage of using the corpus in teaching is the possibility of using several modern approaches to teaching (Tolstova 2009). Within the framework of these approaches, it is assumed that the student acts as a researcher, learns to select "live" language data and systematize them. The shift in emphasis from the deductive approach to the inductive approach significantly changes the role of the teacher, who becomes an assistant, informer, curator of the study, and not only guides, but helps the student in learning the language. The center of corpus language learning is the learner, his needs and preferences.

As is known, the first programs for the presentation of speech patterns, programs for compiling concordance based on a linguistic database, were developed as part of the COBUILD project (Collins Birmingham University International Language Database) under the leadership of J.Sinclair. Now every building is equipped with such programs. They are part of a package of programs called "Case Manager", with the help of which the processing of case data is carried out.

T. Jones, who stood at the origins of corpus learning, noted that the student's task is to "make discoveries" (discovery learning - a heuristic method) in a foreign language, and the teacher's task is to provide a context in which these strategies can develop. With their help, the trainee will be able to "learn how to learn" (learn how to learn). At the center of the process is the use of the computer, not as a surrogate teacher, but as a special type of informant. The difference between a teacher and an informer is best presented in a stream of questions and answers. In traditional methods, the teacher asks a question to which he already knows the answer. The question is asked to test the assimilation of the acquired knowledge. The learner gives an answer and receives the reaction of the teacher. The scheme works: Initiation - Answer - Reaction (of the teacher). If the teacher plays the role of an informer, he is passive and remains silent

until the question comes from the student. The informant answers the question as best he can, after which additional questions may follow from the learner, who will integrate the new information into the schema or knowledge system he already has. The earlier attempt by these scientists to turn the computer into an informer and describe the rule system in computer terms only resulted in the recording of the fact that the machine's ability could not approach that of a human in interpreting what was said or written, or in expressing meaning. As a result, an alternative model was proposed, based on the ability to encapsulate, briefly demonstrate linguistic competence on a set of corpus examples. This method has been called "database learning" - an approach that gives the learner access to numerous facts of speech use and leaves the work of analyzing these facts and finding answers to the learner. The main research tool is word concordance, i.e. word matching strings, ordered relative to the right and left query contexts.

There is a significant difference between help of a teacher and scaffolding. The authors who develop the idea of language learning based on professional content (CLIL, Content and Language Integrated Learning) note that building supports in this area is especially necessary. Instead of providing a ready-made answer, the teacher can support the student in the search for an answer (for example, in the corpus) and, thus, create supports with the help of which the student will solve a similar problem on his own next time.

### Conclusions

One of the main properties of the corpus is representativeness, i.e. the corpus should adequately reflect the phenomenon under study. At the corpus planning stage, ensuring representativeness is a time-consuming task for many reasons. From a methodological point of view, an attempt to fit the corpus planning process into the learning system leads us to the solution of two basic measures.

Problems of selecting the content of training and diagnosing the needs of students. After all, in fact, the selection of learning content is the representation of language and linguistic phenomena (a huge array of corpus texts representing the entire language) into a "significantly smaller" corpus of educational texts. to the least, to the questions: "What to teach?", "How to teach?" and, as a consequence, "What exactly - the formation of what skills and abilities - to control? That is why hull planning is such a thought-intensive task. The linguodidactic potential of a large collection of texts and the possibility of their simultaneous linguostatistical processing in order to create a professionally oriented course is enormous and has not yet been fully mastered. Compiling such a corpus can become a mandatory procedure at the stage of developing a professionally oriented foreign language course.

## References:

1. Zhinkin N.I. On code transitions in inner speech // Questions of linguistics. - M., 1964. - No. 6
2. Zabortkina V.I. New vocabulary of the English language. - M.: Higher School, 1989
3. Zalevskaya A.A. Psycholinguistic research. Word. Text: Fav. works. - M.: Gnosis, 2005.
4. Isenina E.I. Humanization of learning at school: Ideas. Values. Methods: Proc. allowance. - Ivanovo: Ivan. state un-t, 2007.
5. Caprara J., Servon D. Psychology of personality. - St. Petersburg: Peter, 2003.
6. Kamshilova O.N., Kolina M.V., Nikolaeva E.A. Development of the corpus of texts of St. Petersburg schoolchildren: tasks and prospects / O.N. Kamshilov, M.V. Kolina, E.A. Nikolaeva // Applied Linguistics in Science and Education. - St. Petersburg, 2008.
7. ЁРИНБОЕВА, ?. М. (2018). THE IMPORTANCE OF CONSCIOUSNESS AND INTUITION IN STUDYING FOREIGN LANGUAGES. Иностраннные языки в Узбекистане, (3), 67-73.
8. Уринбаева Х. М. МАСОФАВИЙ ТАЪЛИМ ВОСИТАСИДА ОЛИЙ ТАЪЛИМ МУАССАСАЛАРИ ТАЛАБАЛАРИНИНГ ТИЛ КОМПЕТЕНЦИЯЛАРИНИ РИВОЖЛАНТИРИШ //ИННОВАЦИИ В ПЕДАГОГИКЕ И ПСИХОЛОГИИ. - 2020. - №. SI-2№ 9.
9. Makhamadinovna U. K. The Use of an Integrative Approach in Teaching Foreign Languages in the Process of Distance Learning //International Journal on Integrated Education. - Т. 4. - №. 12. - С. 113-116.
10. Abbasovna, A. S., & Mahamadinovna, U. K. (2020). Developing Professional Skills Of Students In Foreign Language Education (In Non-Phylological Universities). European Journal of Molecular & Clinical Medicine, 7(03), 2020.