

THE MAIN DIRECTIONS OF CORPUS RESEARCH

*Dilnoza Dexkanova**Samarkand state institute of foreign languages
Senior teacher of the department of English Philology*

Abstract. this article gives general information about the main directions of Corpus linguistics and important features of it. In addition, it provides several notions from famous linguists namely V.P. Zakharov, T. McEnery, E. Finegan, M. McCarthy and etc as well as they identified main distinctive charactersitics of Corpus Linguistics.

Keywords: didactic, linguodidactics, methods, corpus, tokens, electronic corpora, cogris-driven, cogris-based.

The training of skilled experts capable of self-realization and progress in all spheres of professional activity is the primary focus of Uzbekistan's contemporary higher professional education. The development of linguistic computer skills, linguistic cultural competences, and the teaching of practical foreign language knowledge are the three most crucial aspects of such an extensive program. The development of a contemporary linguistic personality and enhanced communication abilities that enable the use of all linguistic resources in a variety of communication contexts ought to be the outcomes of such training.

The distinctive didactic qualities of contemporary computer and information technologies enable the complete implementation of contemporary methodological methods to teaching foreign languages. One potential use of these information technologies is corpus linguistics, which is now a promising area of study in linguistics and linguodidactics [V.P. Zakharov, T. McEnery, E. Finegan, M. McCarthy], and others. As of right now, corpus linguistics is recognized as a linguistics method related to the examination of language in authentic conversation. The instruments, processes, and work methods created within the confines of this science are currently being actively employed in linguistic research, dictionary creation, and the process of teaching foreign languages in both their entirety and each of their component parts.

If the goal of the majority of linguistics fields—including semiotics, psycholinguistics, sociolinguistics, and semantics—is to characterize and evaluate linguistic structures, then corpus linguistics is a comprehensive approach that may be used to address a wide range of linguistic issues. In this regard, corpus linguistics may be referred to as “a bundle of methods from different fields of linguistic research” [1] according to V.P. Zakharov and S.Y. Bogdanova. As a relatively new field of linguistic study, corpus linguistics is a subfield of applied (computer) linguistics that aims to provide broad guidelines for the creation and utilization of computer-based linguistic corpora [1].

Conversely, A.N. Baranov proposes that computational linguistics be viewed as “a broader area of using computer tools – programs, computer technologies for organizing and processing data – to model the functioning of a language in certain conditions, situations, and problem areas” [2]. The methodological framework of corpus linguistics generally focuses on achieving the best possible balance between theoretical and empirical knowledge. Researchers do not focus on natural language as a complex structure or the issue of formalization, but rather on natural language

as a complex reflection of meaningful communication in language. An electronic language corpus is the fundamental idea of corpus linguistics.

There are several meanings of this idea in use today. In his textbook, E. Finegan states that “a corpus is a representative collection of texts, usually in a machine-readable format and including information about the situation in which the text was produced, such as information about the speaker, author, addressee or audience” [3]. “A collection of language fragments selected in accordance with clear language criteria for use as a language model” is how T. McEnery and E. Wilson describe the corpus [4]. In the 1960s, N. Francis and G. Kucera built the first electronic language corpus. This building (the Brown Corpus) is a member of the first generation of buildings, as are the Lancaster-Oslo/Bergen corpus, the London-Lund corpus, and the Lancaster Spoken English corpus.

Their little size and the thoughtful placement of materials according to genre set them apart. The British National Corpus, Collins Cobuild, American National Corpus, Corpus of Contemporary American English, National Corpus of the Russian Language, and other corpuses are examples of the second and third generations of the corpus. They are distinguished by better techniques for storing and retrieving tokens from arrays of texts, as well as an increase in volume of up to a hundred and more word usage.

M. McCarthy claims that corpus linguistics innovates scientific methodologies and technology and even foresees the advent of more significant innovations that will challenge conventional notions of education, the teacher's function, and the cultural environment of learning [5].

The genuineness of conversation, the capacity to work with statistical data, and the ability to get beyond the “narrow” perspective of the issue that characterizes traditional study are the main advantages of corpus research. At this point in the field's development, it's crucial to be aware of some of the drawbacks of corpus linguistics, including the lack of material processing skills knowledge among linguists and foreign language instructors, the possibility that research objectives will conflict with the circumstances of the discourses generated, and the challenge of comprehending corpus narratives. However, corpus linguistics has made a significant contribution to both the approach used in teaching foreign languages and the research and description of linguistic phenomena.

As an intermediate in the development of didactic materials based on the corpus, the compiler chooses information that is acceptable and correct for teaching pupils and that he believes is valuable and required (such as phraseological units). The curriculum's prerequisites, favored teaching strategies, etc., might serve as odd selection criterion. This is the reason it's crucial to “leave them to themselves” in addition to demonstrating to instructors how a certain corpus is run and structured while preparing them to work with it. It is crucial that every instructor makes the personal decision to choose whether the corpus guides him or if he guides the corpus in the direction he requires [5].

The originator of the generative direction in linguistics, N. Chomsky, said in an interview conducted in the early 1960s that the corpus approach amounts to nothing more than the repetitive observation of copious amounts of data and that it “is not a method of scientific cognition and cannot provide either a successful solution to practical problems or an increment of knowledge” [6].

But M. McEnery notes—quoting from this conversation with N. Chomsky—that he never

downplayed the use of corpora for language acquisition. Linguistic analysis, also known as corpus-informed analysis or approach, is a current scientific way of examining vast amounts of corpus data. In this method, corpus data is utilized simply as a source of natural language samples.

There is no quantitative analysis done, and the researcher's own language intuition plays a far less part [7]. N. Chomsky's opponents are corpus linguistics representatives who are confident in the huge cognitive potential of this science. According to J. Sinclair, there are two ways to use electronic corpora in linguistic research: corpus-based and corpus-driven [8]. The problem of confirming the accuracy of a certain theory that is provided to the researcher is resolved in the study that uses corpus-based data. Therefore, using the corpus in this case is a research approach. But, there's a chance the researcher may object to anything that doesn't "obey" his theory [9].

In this language analysis, the beginning theoretical positions are predetermined and remain unchanged throughout the investigation. Corpus material is studied both qualitatively and statistically. For instance, the study of the corpus's frequency of third-declension nouns does not challenge the reality of the Russian language's declension system [7]. The study that is corpus-driven operates rigorously empirically and builds a theory exclusively using the corpus's data. Since the hypotheses originate from the corpus observations itself, none are developed before the investigation. In this case, the linguistic corpus is regarded as a separate study subject.

Scientists using this method describe usage by depending only on the corpus material [10]. Since such an analysis makes limited use of preexisting theoretical viewpoints regarding the language, it may be used to uncover patterns and exceptions in the language. The definition of inflectional classes, which are taken from the corpus rather than pre-defined by the classification of parts of speech, is an illustration of this methodology [7]. It is important to acknowledge that the differentiation between these two methodologies is not entirely unconditional.

T. McEnery and A. Hardy remark that the distinctions between the two primary methods for studying language phenomena—corpus-driven and corpus-based—are being eliminated, representing the entirety of contemporary corpus research [4]. Thus, the necessity and significance of using linguistic corpora while undertaking a particular study are undeniable in science today [11]. Nevertheless, corpus linguistics methodologies are not commonly employed in contemporary foreign language instruction methods. The way this material is approached is more of the issue. However, there is no denying the usefulness of corpora in the instruction of lexicophraseological aspects of foreign languages.

M. McCarthy and associates devoted specific portions of their book "From Corpus to Classroom"—titled "Lessons from the analysis of chunks" and "Idioms in everyday use and in language teaching"—to the study of phraseological units using corpus linguistics techniques [12]. According to K.M. Shilixina, this is because the corpus nature of this research boosts their dependability and reliability, enables you to test theories on lexical changes, and eliminates subjectivity and incompleteness of descriptions. Furthermore, as noted by the author [13], the electronic corpus represents a significant amount of data and represents an effective effort to bring "the vast (for example, oral discourse) (discourse presented in marked-up texts that can be studied)" to the public eye.

Furthermore, resolving some methodological issues necessitates handling many structures concurrently. By comparing and analyzing the collected data, lexical variability may be established and usage that deviate from the standard language norm can be found. In addition, the advent of

corpus technologies has greatly simplified the process of gathering and analyzing lexical data and allowed for the observation of uncommon lexicophraseological units as well as the tracking of the dynamics of such changes over brief time periods.

In order to satisfy the demands of contemporary society and be in line with a new quality of linguistic reality, it is thus necessary to include the accomplishments of corpus linguistics into the technique of teaching the lexical and phraseological aspects of foreign languages. The electronic corpus, which is a novel method of acquiring linguistic data, is essential to current linguistic description and is required for modern scientists' scientific work.

LIST OF USED LITERATURE

1. Захаров В. П. Корпусная лингвистика: учебник для студентов гуманитарных вузов / В. П. Захаров, С. Ю. Богданова. – Иркутск: ИГЛУ, 2011. – 161 п.
2. Баранов А. Н. Введение в прикладную лингвистику: учебное пособие / А. Н. Баранов. – Москва: Эдиториал УРСС, 2001. – 360 п.
3. Finegan E. Language: Its Structure and Use / E. Finegan. – Thomson Wadsworth. – 2007. – 611 п.
4. McEnery T. Corpus Linguistics: An Introduction / T. McEnery, A. Wilson. – Edinburgh University Press, 2001. – 235 п.
5. McCarthy M. J. Touchstone. Level 1. Student's Book / M. J. McCarthy, J. McCarten, H. Sandiford. – Cambridge: Cambridge University Press, 2005. – 152 п.
6. Andor J. The master and his performance: An interview with Noam Chomsky / J. Andor // Intercultural Pragmatics – 2004. – P. 93-111.
7. Копотев М. В. Введение в корпусную лингвистику / М. В. Копотев. – Прага, 2014. – 230 п.
8. Sinclair J. Corpus, concordance, collocation / J. Sinclair. – Oxford University Press, 1991. – 170 п.
9. . Борискина О. О. Корпусное исследование языка: мода или необходимость? [Электронный ресурс] / О. О. Борискина // Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация. – 2015. – №3. – Access mode: <https://cyberleninka.ru/article/n/korpusnoe-issledovanie-yazyka-moda-ilineobhodimost> (26.06.2024).
10. Голубкова Е. Е. Использование лингвистических корпусов при решении семантических проблем / Е. Е. Голубкова // Методы когнитивного анализа семантики слова: компьютерно-корпусный подход. – Москва: Языки славянской культуры, 2015. – P. 34-75.
11. Boulton A. Corpus use in language learning: A meta-analysis / A. Boulton, T. Cobb // Lang. Learn. – 2017. – Vol. 67. – P. 348-393.
12. O'Keeffe A. From Corpus to Classroom: Language Use and Language Teaching / A. O'Keeffe, M. McCarthy, R. Carter. – Cambridge University Press, 2007. – 315 п
13. Шилихина К. М. Использование корпусов в исследованиях дискурса / К. М. Шилихина // Вестник Воронеж. гос. ун-та. Сер. : Лингвистика и межкультурная коммуникация. – 2014. – № 3. – P. 21-26.