

## LINGUISTIC CORPUS AS A METHODOLOGICAL BASIS OF LINGUISTIC RESEARCH

*Qarorova Dilsora O'ktam qizi*

*Termiz State Pedagogical Institute Foreign language and literature: English language 1st year  
master's student , [qarorovadilsora@gmail.com](mailto:qarorovadilsora@gmail.com)*

**Abstract:** Development of Corpus Linguistics today and its future we can find out the issues. A representation of language information for a specific reference study of the dynamics of change processes in the lexical structure of the language in various genres and allows analysis of lexical and grammatical features by different authors. It can be said that corpus linguistics has a wide range of users as its subject linguistic data intended for linguistic research in the interest of theoretical and practical basis of creating and using representative arrays has mechanisms.

**Key words:** features of corpus linguistics, models of corpus linguistics, corpus of texts, corpus linguistics. In modern linguistics as a result of the development of computer technologies semantic analysis of corpus and computational linguistics terms, corpus linguistics, critical discourse analysis, as well as descriptive, comparative and component research based on analytical methods has become more and more important is going Linguistically, a computer The emergence of linguistics is a part of today's science gives an opportunity to find solutions to problems. Computer technologies scientific-theoretical, philosophical problems of linguistics: language and speech, system and structure, semiotics, syntagmatics and paradigmatics, typology, lexicology scientific work on such important and complex issues as corpus based on the positive aspects of its use in research the field of linguistics was created. Computational and corpus linguistics to each other there are misconceptions that they are similar and in some cases the same. XX In the first half of the 90s of the century, corpus linguistics is the science of language was formed as a separate part. At the same time, it is computational linguistics closely cooperates with, uses its achievements and in turn, enriches it. Corpus linguistics - using computer techniques and technologies development of linguistic corpora (corpus of texts) and effective use of them a branch of computational linguistics that develops general principles. Linguistic or a corpus of texts (or usually just a corpus of texts) solves linguistic problems in one sense organized, combined, designed, defined, philologically can be understood as a machine-readable format of qualified language data. That's it because, in order to solve the range of tasks for which this or that corpus was created (sometimes quite wide) taking into account, we can say that the corps always follows a pragmatic direction. Nowadays, there are many definitions of the term "corpus". For example, E. Finegan the definition given in the textbook says: a corpus is a representative set of texts, information about the context in which the text was created, usually in a machine-readable format; for example, information about the speaker, the author is included. Statistical analysis of the Wikipedia corpus and large and used to test hypotheses, confirm or justify language rules defined as a collection of structured texts (now usually in electronic form).

T. McEnery and E. Wilson gives the following definition: a corpus is a model of language is a collection of parts of a language selected according to specific language criteria for use. The above definitions emphasize the main features of the corpus of modern texts: purpose ("logical idea",

pragmatic orientation), machine-readable format, special text selection representativeness, availability of metalinguistic information as a result of the procedure. Oral standardized presentation of the material in the machine from standard programs for its processing allows use. Expediency of creation and use cases meaning is determined by the following conditions:

- sufficiently large (representative) and balanced size of the corpus of data guarantees the typicality and completeness of the presentation of the entire spectrum of language phenomena provides;
- different types of information are in a natural contextual form in the corpus, and they are creates an opportunity for comprehensive and objective learning;
- an array of data created and prepared once by different researchers and can be used repeatedly for different purposes.

The concept of "text corpus" also includes a management system, a corpus manager (or corpus manager) text and linguistic data. This was requested in the corpus data retrieval and presentation in user-friendly form and statistics a specialized search engine that includes software tools for data retrieval. Corpus searching allows you to generate matches for any word, to the source . A list of all uses of a given word in context with references. However, in addition to properties of text or lexical units from the corpus, language units and linguistic categories and metadata (word forms, lexemes, frequency of grammatical categories, various about statistical information about changes in frequencies and contexts during periods). can be used for information. time), information about a joint event is lexical units, genre and stylistic features, etc. These statistics are straight forward can be output (e.g. frequency lists) or they are "internal" calculations and to the corpus can be used to extract new information not directly entered, for example in text quantitative representation of compound stability, paradigmatic (semantic) clustering, lexical units, identifying key words of the text.

Computational linguistics also has its problems and achievements studied from the point of view. A.N. Baranov artificial consciousness in his works stopped to reveal the problem, B. Orekhov semantic analysis, technical if he studied science by translation and automatic morphology I.A.Chesebiyev, the history of development of methods of computer linguistics studied. Here, N. Chomsky's syntactic structures and The theory of universal grammar made a great contribution to the development of science. Ch. Fillmore and as a supplement to this theory, it is like a preposition in languages dissimilarities and their expression through other concepts in the language structure approached from the point of view. W. Woods (Woods) procedural science introduced the idea of semantics. The above-mentioned scientists use language in their work and have set themselves the goal of gathering the concepts of consciousness in technology. After all, one of the most urgent tasks of computer linguistics is human and creating an automatic communication interface between computers. From thi it turns out that the main goal in this field is language learners and researcher provision of special technical devices or programs. A group of scientists such as V. Mamontova and N. Chomsky developed corpus linguistics evaluate as a component of computational linguistics. Theirs most suggest the following definition: corpus linguistics - construction of a corpus of linguistics with the help of computer technology and from them units that develop general principles of use generative computational linguistics. At the same time, usually a computer equipment, programs, computer technologies of the corps and a wide range of data processing mechanism understood in the form of modeling.

A representational array of language data for a given data period is the lexicon of the language to study the dynamics of change processes in different genres and different authors enables the analysis of lexical and grammatical features. Linguistic researchers are increasingly interested in functions it can be said that there are intelligent processing of corpus data and such programs that they are no longer a corpus manager as a factographic type of information retrieval system, but different software that provides a wide range of linguistic functions, linguistic, mathematical is a complex conglomerate of tools. We name this concept "corpus service" we offer. Today, corpus linguistics is often done through computers and electronic corpora as a new linguistic discipline concerned with learning to use language in real life understood. At least two features that form the basis of corpus linguistics as an independent discipline there are: 1) the nature of oral speech is material, that is, fixed texts; 2) tools specific features of the collection. If the syntax, semantics and While fields such as sociolinguistics focus on describing or evaluating language structure or language use, corpus linguistics is a broader concept that encompasses not only linguistics but also can be applied to many aspects of linguistics. So-called "culturometrics" (culturomics) and humanities based on corpus methods a new science that applies to all areas of research. Corpus linguistics is "linguistic research it is called a collection of methods from different fields. Linguistic analysis method corpus linguistics, as well as between languages, dialects or language variants in the process of comparative study related to contrast studies aimed at identifying general and specific facts. Linguistic many types of analysis thrive on solid and extensive empirical evidence. In general In general, when the question arises about the role of corpus linguistics in linguistics, it seems that it is

it would be correct to say that it is a linguistic research methodology used in practice, to any field of linguistics. However, there is another view: corpus linguistics is, in fact, real is scientific linguistics. In English literature, these approaches are corpus linguistics linguistics methodology and as a separate discipline - corpus-based (corpus-oriented approach) and is called corpus-based (corpus-based approach). The first approach to support, confirm, refute linguistic theories or hypotheses, or implies the use of corpuses for improvement. The second approach declares that the corpus itself is the primary and only source of our theories about language, corpus linguistics here assumes the status of a theory and is considered "linguistics as the most important concept in the theory". So, the corpus indirectly includes the theory of language and you just have to "get" it from there. This concept is from the first third of the 20th century one goes back to the work of the American structuralists. In the depth of corpus linguistics it is .The approach is called neofirthian because it is a concept of collocation introduced by J.R. Firth is strongly related to Perhaps the most famous quote in corpus linguistics. The creation of the first corpus dates back to 1812, when a German scientist Kadling analyzed the distribution of consonants in his German words. However, at that time the term computer did not even exist. Later, Brown as one of the examples of modern English corpus corpus was created in 1960-1961 and it was published for the first time in printed form. Only a year later, this corpus was digitized. Apparently, the first corpora exist even without computer technology and research went Later, the development of science introduced the electronic search method (concordance) required transfer in format texts and with computer linguistics there was a need for inter-sectoral relations. At the same time, computational linguistics is "generally, computer tools (programs, a computer for organizing and processing data technologies) in certain conditions, situations, problem areas and the scope of language models

not only in linguistics, but also in other disciplines also implies the application". A.N. Baranov "computer modeling of language programming theory in linguistics (building computer models and computer graphics science) can be considered as a field of application. Because In recent years, this problem has been recognized only as a linguistic issue is being done," he said. Corpus Linguistics is a computer specific tool uses as and without them, of course, it cannot perform the tasks it performs. However, this applies to almost every field of modern knowledge, and they does not make them integral parts of computer science.

J. R. Furse: "You can tell a word by the company it keeps." of this approach the essence is that the meaning of the word (like other linguistic concepts) is only will be present in the context (text). No data analyzer is installed a priori it is assumed that it does not use theoretical concepts. Another basis for language learning is this the concept of speech. For them, speech is not only a text, the "practice" of language, but the language itself or .The way of doing subtil is not only a way of speaking, but also a way of thinking. And it can be noted that he recognized the corpus-based approach many publications by linguists are actually corpus-oriented studies.Our In our opinion, an important aspect of defining corpus linguistics is that it is simply language not a learning methodology - the object of learning is the resulting system. E. Finegan Corps to study the necessary and natural use of linguistics for the construction and use of a corpus defined as a focused activity. This definition is the creative direction of corpus linguistics emphasizes. Its dual nature (creating a text corpus and using it focused) is determined by the dual nature of its object - the corpus of texts, which,on the one hand, the source for corpus linguistics and other linguistic disciplines is speech material, on the other hand, it is itself a product of corpus linguistics. It can be said that the body .Linguistics as its subject is linguistic in the interests of a wide range of users creation of representative arrays of linguistic data intended for research and has theoretical foundations and practical mechanisms of their use.

Summary: Any emerging science has computers in its practice uses the achievements of its development, but these disciplines (corpus linguistics, computer linguistics, computer mathematics, computer physics) should not be considered as part of computer science. The computer as a tool in researching the object of analysis of these sciences only uses. It can be said that corpus linguistics is of computational linguistics not a separate part or section. They are a direction and an object of research .It is an independent field with different goals and problems.

References:

1. Abdurakhmonov N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation. Journal of Social Sciences and Humanities Research. 2017;5(03):89-100.
2. "Foreign languages" of the Cabinet of Ministers of the Republic of Uzbekistan on measures to effectively organize the popularization of learning" Resolution No. 312 of May 19, 2021, [www.lex.uz](http://www.lex.uz).
3. Ataboev N.B. Chomsky's Critique of Corpus and the Problem of Expertise Speakers' Linguistic Intuition // Issues of Philology. - Tashkent. (4)/2015.- B. 108-111.
4. Abdurakhmanova N.Z. Linguistic support of machine translation. Monograph.Tashkent-2018 pp. 30-31