

STRUCTURAL-SEMANTIC ISSUES IN PARALLEL CORPUS

Rahmatullayeva Sayyora Jonzoq qizi

Abstract

This article discusses corpus linguistics, concepts related to the corpus, and its parallel corpus link, its structure, corpus types, tokens, lemmas, stemming. Today, the theoretical and practical significance of the corpus is in the study of the existing possibilities of language in Uzbek linguistics, the identification of problematic aspects of linguistics, the creation of electronic dictionaries, increasing the effectiveness of modern information technology in language learning, automatic translation, search and computer analysis. Possibilities of language in the study of Uzbek linguistics, the recognition of challenging issues areas of linguistics, the development of digital dictionaries, growing the efficiency of modern information technology in learning languages, automatic translating, looking, and computer examination. In problem-solving, it is necessary to Create a collection of words in particular fields.

Keywords: Corpus, corpus linguistics, parallel corpus, translation corpus, comparable corpus, segmentation, machine translation, tokenization, lemmatization, stemization

One of the global challenges of the 21st century is to preserve the national character of natural languages. It has become an urgent task to consistently conduct research on language technologies in the creation and development of electronic Corps of world languages.[1] Scientific, practical research carried out in the field of foreign corpus linguistics proved that the corpus is a necessary and necessary point not only for representatives of the verbal sphere, but also for the prosperity of the nation.

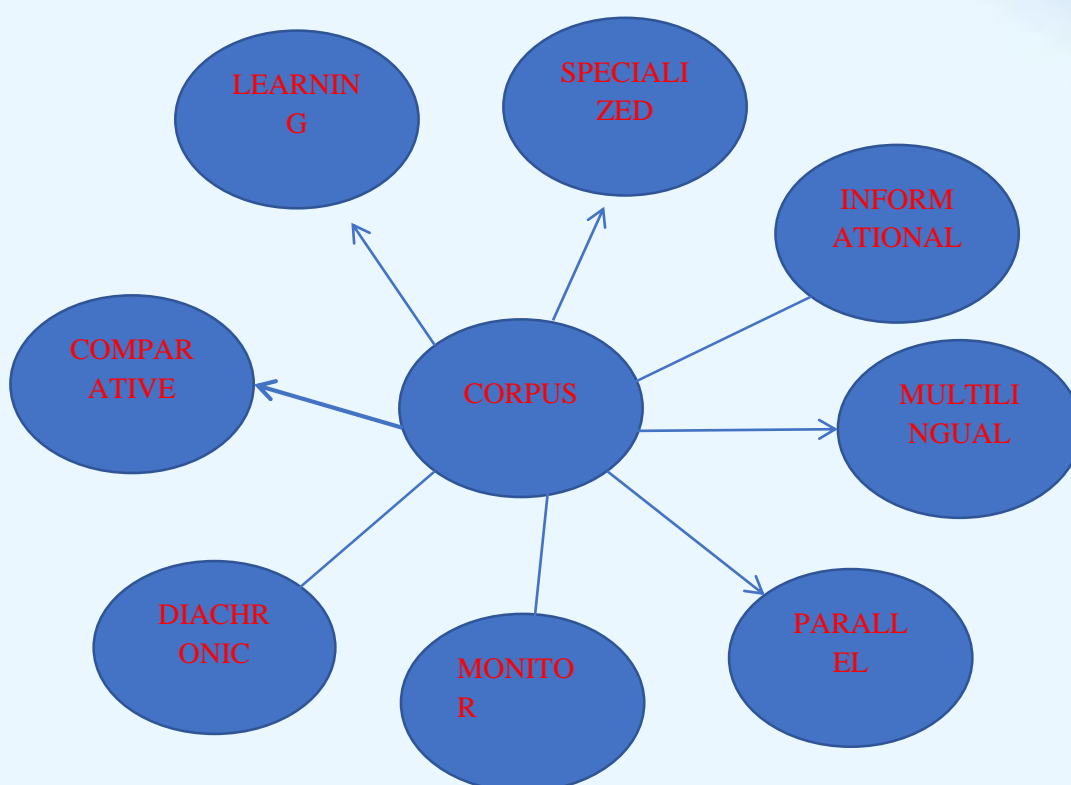
Even in our state, the creation of the National Corps of the Uzbek language, one of the most important issues facing our practical linguistics today, has risen to the level of politics. In particular, in order to increase the prestige of the Uzbek language in the life of society and internationally, the national language of the Uzbek language of electronic appearance, which includes all scientific, theoretical and practical information about the Uzbek language creating a corpus, popularizing the Uzbek language in the Internet world information network, ensuring a worthy place in it,

Uzbek software products the creation of applications, the implementation of computer programs teaching the Uzbek language in large-scale practice, the creation of computer programs for the editing of texts in the Uzbek language have been identified as important tasks facing linguistics retrieved.[2] In practice, a lot of research is being done on this.

Corpus Linguistics of Russian and English languages in different fields V. Zakharov, A. Sedov, A. Baranov, R. Potapova, V. Rikov, U. Francis, N. Leontyeva, W. Martin, S. Kubler, A. Lawrence, E. Etwell, S. Hunston, L. Boizou, McKennery, J. Grafmiller, J. Grieva, N. Groom, S. Hansson, K. MMcAulif, M. Malberg, P. Milin, Foreign like A. Murakami, R. Paich, A. Schembri, P. Thompson, B. Winter, G. Lynch scientists by and Corpus Studies (corpus linguistics) in Turkology scientific researches were carried out in the field. Turkish language Aksan, Deniz, Zeyrek, Kemal Oflazer, Umut Ozge Bular; Yusup Aibaidulla, Kim-Teng Lua on the Uighur language; I.A. Buskunbaeva, Z. Sirazitdinov on the Bashkir language; Sheymovich on the Khakaz language, J. Suleymanov, A. Gatiatullin, O. Nevzorova, R. Gilmullin, B. Hakimov on the Tatar language; The works of scientists such as L. Kubedinova on the Crimean Tatar language and Salchak on the Tuva language are noteworthy. From Uzbek scientists B. Mengliyev, Sh. Shahobiddinova, Z. Kholmanova, S. Karimov, N. Abdurakhmonova, L. Raupova, Sh. Hamroyeva, M. Abjalova, G. Toirova, G. Ikromova, J. Djumbayeva, G. Ergasheva, A. The Eshmominovs did scientific work. The conceptology of the national corpus of the Uzbek language is being developed by a team of scientists headed by B. Mengliyev.

A corpus is a set of spoken and written texts stored in a computer database. The time when the materials collected in the corpus were written, which style they belong to, and which source they belong to, will also be explained in detail. Depending on his interests, the user can refer to artistic, scientific, official or journalistic texts. This is especially useful in language learning. In school education, pedagogues are very helpful in quickly giving tasks to students to strengthen their knowledge during the lesson. The corpus is a systematized library with a very wide scope and a high level of importance. Easy to use, saves a lot of time.

It differs from other programs in terms of electronic search engine. Corpus search allows the user to find all forms of the specified word in different contexts. It clearly shows where it is in the dictionary and its options. It can determine the range of words that can be combined, denotative and connotative meanings of the searched word. Describes the frequency or statistics of word usage in a writer's work. It is a sign of modern development that can reflect the state of use of this word in which period.



We observe the information about the structure, composition and possibilities of the parallel corpus in the works of D.O.Dobrovolsky, Yu.Tao, V.Zakharov, A.A.Kokoreva, E.P.Sosnina.

A parallel corpus, that is, a collection of originals and their translations can be used in a variety of ways for the benefit of translation studies, machine translation, linguistics, computational linguistics, or simply the human translator. In computational linguistics, translation corpora have been used for machine translation, as well as term extraction, word semantics, etc., since the early 1980s.

In the late 1980s and early 1990s, the first parallel texts were the avalanche reports collected in Switzerland in German, French, and Italian languages, and the Canadian mass media in English and French. One of the first electronic resources was the Canadian Hansard, which was originally used to perform sentence matching.

(Gale & Church 1991), this task is now a standard feature of applications such as translation memories. In addition, parallel corpora are used as databases for multilingual grammar induction, automatic lexicography, and many other tasks in cross-language information retrieval and language processing. The final goal of all the projects created in these years is machine translation was to create a system.

Parallel corpora are central to translation studies and contrastive linguistics. Many parallel corpora are accessible through easy-to-use concordancers, making the study of cross-linguistic phenomena much easier. Anarbayev Orzubek Rachmanovich gives the following opinion to the parallel corpus: „ ... Parallel corpus is the current era when intercultural communication is widespread.

is an important reality for Through parallel corpora, it will be possible to identify universals in different language environments and cultures, as well as specific mental characteristics of languages, real and lacunar units. The corpus of parallel texts also serves for the development of automatic translation, ensures the development of computer lexicography. Using the corpus of parallel texts, concordance programs will be developed and it will be possible to create dictionaries of various specializations. In fact, the creation of corpora is of great importance, first of all, for the field of lexicography. The reflection of time is reflected in the vocabulary of the language. Constantly "grows". The window that clearly shows the richness of the language is its dictionary. The large number of vocabulary content is considered a factor that increases its value. Such corpora are also a rich source of materials for language teaching. Also, parallel corpora are a tutorial for statistical machine translation systems serves as information.

A parallel corpus is a pair of translated texts. In translation studies, the main focus is on identifying the features that distinguish translations from original texts. These changes may be individual to a given translation task or translation pair, but

they differ from untranslated texts in terms of broad linguistic features of the translated text.

It can clarify the typical features that do. This research is a clear way to empirically identify specific features of corpus translations, and since the 1990s by Baker (1993; 1996), Johansson & Ebeling (1996) and more recently by Hansen (2003); Teich (2003); Used by Mauranen & Kujamäki (2004) and Hansen-Schirra, Neumann & Steiner (2012). In addition, parallel corpora are used as references in translation teaching and professional translation settings, as they provide quick and interactive access to translation solutions (such as translation memories).

REFERENCES

1. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
2. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J FLTAL). 2019;6(1-2019):131-7.
3. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018). Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.