

Corpus-based technology in creating the methodological foundations of the development of lexicography and lexicogrammar of the Uzbek language

Khoshimjonova Qizlarkhon Ne'matillo kizi

Annotatsiya: Ushbu maqola soʻz va grammatik strukturalarni bir-biriga bogʻliq holda oʻrganilishining nazariy-metodologik asosini korpus lingvistikasida aks ettirilishi hususiga bagʻishlangan

Kalit soʻzlar: korpus lingvistikasi, leksikogrammatika, meronym, xolonim, statistik tahlillar

Annotation: this article is devoted to the reflection of the theoretical and methodological foundations of the interdependent study of word and grammatical structures in corpus linguistics.

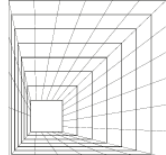
Key words: corpus linguistics, lexicogrammar, meronym, xolonim, statistical analysis

Аннотация: Данная статья посвящена отражению теоретико-методологических основ взаимообусловленного изучения слова и грамматических конструкций в корпусной лингвистике.

Ключевые слова: корпусная лингвистика, лексикogrammatika, мероним, холоним, статистический анализ.

In today's rapidly evolving world, the "corpus language" is effective for all areas. In this regard, it was considered the most alternative way to revise natural language and create a complete electronic database, and many countries have successfully experienced this. Corpus linguistics for Russian and English languages in various areas V. Zakharov, A. Sedov, A. Baranov, R. Potapova, V. Rikov, U. Francis, N. Leontyeva, V. Martin, S. Kubler, A. Laurence, E. Etwell, S. Hunston, L. Boizou, McKenney, J. Grafmiller, J. Grieve, N. Groom, S. Hansson, K. McAuliffe, M. Malberg, P. Milin, A. Murakami, R. Peych, A. Schembri, P. Thompson, B. Winter, G. Leach and other foreign scientists have conducted scientific research in the field of corpus studies (corpus linguistics) in Turkic studies. Aksan, Deniz Zeyrek, Kemal Oflazer, Umut Özge Bular on the corpus of the Turkish language; Yusup Aibaidulla, Kim-Teng Lua on the Uighur language; L.A. Buskunbaeva, Z. Sirazitdinov on the Bashkir language; Sheymovich on the Khakas 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 [2]. Nowadays, a lot of work is being done in certain studies on the corpus of the Uzbek language. In particular, the "Electronic Corpus of the Uzbek Language" has been created over the last five years. The emergence of corpus linguistics began with theoretical research in 2018, and in 2021 the "Educational Corpus of the Uzbek Language" was created.

With the creation of corpora, their functions in linguistics also expand, and this is directly related to the pedagogical sphere - various texts are collected that the teacher can use in any form. For example, the teacher can extract the necessary text from a fiction to prepare the task [3]. In addition, language corpora can also be used in native language tests. It serves as the main source for finding authentic texts that are used to evaluate students' answers in written and oral tests as well as to test reading and listening comprehension [4]. Lexicogram means educational dictionaries that are based on the age characteristics of students, are intended for education, are set to certain standards in terms of scope and are characterized by the fact that they are aimed at a specific subject or study when disclosed the nature of the subjects in the field of study, that is, it has a language didactic character. Finally, experts in this field recognize educational lexicography as an independent linguomethodological direction, a science that



arose at the interface between traditional lexicography and linguomethodology[5].Lexico-grammatical and morphological processing consists of the automatic analysis of the words used in the corpus. As part of this analysis, the word, compound words and phrases in the sentence (text) are processed.

The main goal of morphological analysis is to identify the words in the text, extract them from the context, analyze their morphemic structure, determine their original meaning and syntactic role in the text. The information obtained as a result of word processing can be used for NLP tasks such as determining word meaning, creating vocabulary, analysis and language learning. Greene and Rubin (1971), Karttunen and Wittenburg (1983), Koskenniemi (1983), deHaan (1984), Garside (1987), Church (1991), deRose (1991) on the implementation of morphological analyzes for English and other languages, scientists like Merialdo (1994) conducted scientific research. Morphological analysis performed by natural language specialists is extremely effective. Today, over 95% efficiency is achieved when performing morphological analyzes using machines [6]. Lexico-grammatical analysis in the Uzbek language can be carried out using the following word “childhood”:

1. to make with sentences: innocent childhood, childhood world, childhood threshold, childhood joy, childhood problems, childhood heart, childhood secrets, childhood fantasy, happy childhood, childhood fun... Continuation and these units It is further explained in itself [7]. In the explanatory dictionary of the Uzbek language, childhood is the behavior of young children. The child goes into childhood. Even if a child's face is injured, he will not leave his childhood. The time of a person who has not yet grown up, youth. He loved horses since childhood and always enjoyed watching good horses. It is explained that the adoption involves a child[8].

2. Word semantics: Hyperonym: A part of human life.

Hyponym: baby, baby, guest, small child, baby.

Synonym: youth, adolescence, adolescence.

Antonym: young, young.

Meronym: rock, rock, play, nibble, walk, pack.

Gradually: baby, childhood, adolescence, youth, youth, adulthood, old age.

Example: Even as a child, he enjoyed feeding and watching horses.

Proverbs: A house with children is a market, a house without children is a grave. The child of good in the net, the child of evil in the door.

Grammatical meaning: Related words: verb, noun, adjective. Bola turns into an abstract noun through a specific noun suffix that denotes the characteristics associated with this time. In our research work, we examined in detail the method of using corpus technology in textbooks in the development of the Uzbek lexicon through the uzbekcorpus.uz platform, created at the initiative of Professor N. Abdurahmonova. From theoretical methods: analysis-synthesis; Practical methods were used: observation, experiment and differentiation. Accordingly, the native language textbook for the 9th grade was based on the topic “Simple and compound sentences” and in the experimental test group the students were presented with words based on the table.

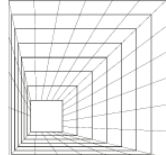
Anger	Joy	Kindness
Grief	Alas	Honor
Happiness	Angel	Confidence

We analyzed these words in the 9th grade, group 1 and studied the grammatical and lexical meaning of these words, referring to the electronic corpus of the Uzbek language. Goal:

1- Use similar words in the written essay and avoid repetition of words

2. Effective use of keywords and proverbs.

Learning the 3rd language levels in relation to each other.



4. The lexical assignment of the words to the corresponding word.
5. Grammatical compatibility of words with related words.
6. The organic connection of the created text to the given topic.

In this method, it is important that the focus is on the method of conscious verbal-cognitive education through the creation of a text in independent and creative research. The combination of reproductive (working with ready-made teaching materials) and problem-oriented teaching methods has become a priority. As can be seen from the above, a hyperonym is a general term for a number of other personal concepts, hyponym (love) = hyponym (anger), hyponym (joy) = these qualities are summarized under the hyperonym of feeling. A meronym is the division of a species into parts. When divided into parts, we can carefully observe the gradualness, for example, if we pay attention to the gradualness of the word anger, it goes from displeasure to anger. Let's look at it like this:

In the phenomenon of gradation, the concept of graduonymic series is taken as the base word and the members of the graduonymic series are arranged in the order of increasing levels of series members in the order of increasing quantity, quality, coloring and expressiveness. The method of teaching using the corpus was effective for the students, and the following results were obtained:

1. Widely used synonyms in words,
- 2- at least 2 proverbs are written according to the content,
3. Correct use of phrases,

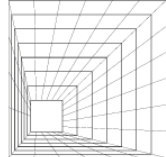
In 4 words, semantic compatibility with the associated word (for example: not to be angry, to be angry, to be angry)

5-Hyponyms and hyperonyms were used effectively. There was no deviation from the scope of the topic.

In conclusion, the lexico-grammatical and lexicographic study of educational dictionaries in the corpus contains fully systematized information about a specific area of knowledge, a set of terms describing a specific subject area and indicating semantic relationships (connections) is formed in the form of a file dictionary with additional information about the relationships between terms, such as: B. Synonyms and general relationships, and is one of the most effective tools for describing specific topics in electronic form. In this context, the further development and perfection of this field is one of the important tasks of linguists and literary scholars today.

References:

1. Xolmonova Z. Kompyuter lingvistikasi. Toshkent: Asian book house, 2020. B. 198.
2. N. Abduraxmonova, U. Tuliyeu, G'. Abdurahobov O'zbek tili korpus lingvistikasining nazariy va amaliy masalalari. Компьютерная лингвистика: проблемы и решения. Т.; 2022
3. Xamrayeva Sh. Korpus lingvistikasi atamalarining qizqacha izohlilug8ati.
4. Toshkent: GlobeEdit is a trademark of International Book Market Service Ltd., member of OmniScriptum Publishing Group, 2020. B. 93.
5. Abdiraimov Sh. Baholash va milliy korpus. <O8zbek milliy va ta9limiy korpuslarini yaratishning nazariy hamda amaliy masalalari=xalqaro ilmiy-amaliy konferensiyasi materiallari (2021, 7 01). B. 226 3228.
6. Морковкин В.В. Основы теории учебной лексикографии: Дисс. ... д-ра филол. наук (в форме научного доклада). – М., 1990.; Денисов П.Н. Учебная лексикография: итоги и перспективы. Проблемы учебной лексикографии. –М.: Изд.-во Моск.ун-та, 1977.
7. Elov, B.B., Hamroyeva, Sh.M., Abdullayeva, O.X., Husainova, Z.Y., Xudoyberganov, N.U. 2023. "Agglutinativ tillar uchun pos teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida)". O'zbekiston: til va madaniyat 2: 6-39.



-
8. <http://uzbekcorpus.uz>
 9. O‘zbek tilining izohli lug‘ati. 2-jild. O‘zbekiston milliy ensklopediyasi. 2006. –B. 347.