

Modeling of models and processes that differentiate semantically polyfunctional words in the context of the Uzbek language.

Elov Botir Boltayevich¹ and
Sirojiddinov Shuxrat Samariddinovich²
Axmedova Xolisxon Ilxomovna³
Abdullayeva Oqila Xolmo‘minovna⁴
Xusainova Zilola Yuldashevna⁵

¹ Tashkent State University named after Alisher Navoi University of Uzbek Language and Literature , Tashkent O‘qituvchi street 103, Uzbekistan
`elov@navoiy-uni.uz`

² Tashkent State University named after Alisher Navoi University of Uzbek Language and Literature , Tashkent O‘qituvchi street 103, Uzbekistan
`rector_tsuull@navoiy-uni.uz`

³ Tashkent State University named after Alisher Navoi University of Uzbek Language and Literature , Tashkent O‘qituvchi street 103, Uzbekistan
`xolisa9029@mail.ru`

⁴ Tashkent State University named after Alisher Navoi University of Uzbek Language and Literature , Tashkent O‘qituvchi street 103, Uzbekistan
`abdullayeva.oqila@navoiy-uni.uz`

⁵ Tashkent State University named after Alisher Navoi University of Uzbek Language and Literature , Tashkent O‘qituvchi street 103, Uzbekistan
`xusainovazilola@navoiy-uni.uz`

Abstract. Another urgent issue of applied linguistics is to create a linguistic filter for semantically distinguishing polyfunctional words, and for the field of computer linguistics, it is an urgent issue to develop mathematical models and algorithms for semantically distinguishing them, as well as an information system. This article discusses the models for semantic differentiation of polyfunctional words in the Uzbek language when they occur in the structure of a sentence. We got acquainted with studies of the Turkic language family. A hierarchy of polyfunctional words in the Uzbek language in terms of word groups has been formed. Mathematical models for semantically differentiating polyfunctional words within the categories noun $\vee$ adjective, adjective $\vee$ modal word, adjective $\vee$ adverb, auxiliary verb $\vee$ independent verb and conclusion $\vee$ auxiliary are presented. The business processes distinguishing

Keywords: Polyfunctional words mathematical model set finite set
business processes modeling vocabulary conceptual model
structure and architecture of information system.

1 Introduction

The problem of polyfunctionality of lexical units is one of the urgent problems of modern linguistics. In world linguistics, the issue of polyfunctional words often occurs in Russian, English, Tatar, and Chinese languages. The concept of polyfunctionality appears in some sources as multi-functionality. In many foreign linguistic works, polyfunctionality is considered in connection with the phenomena of homonymy. One of the main issues in the field of natural language processing is the issue of automatic semantic analysis of words, sentences, and texts. Semantic analysis is very important to NLP (Natural Language Processing) because its processes enable the identification of different meanings of words. In addition, these processes help the machine understand the meaning of whole sentences and texts. There are two typical processes of NLP semantics, which in turn are divided into several groups according to the task they perform:

1. Word sense disambiguation
2. Sentiment analysis

Word sense disambiguation. This term describes the automatic process of determining the context of any word. Thus, the process focuses on analyzing a sample of text to find out the meaning of a word. In natural language, one word often has more than one meaning. For example, the word *kul* can mean *kukun* from a fire, but it can also represent the imperative form of the verb *kulmoq*. The task of the computer is to understand the word in a certain context and choose the best meaning. Computers use inference to perform tasks.

Sentiment analysis. A sentence often contains several nouns (words or phrases) that are related to each other. The term *relationship extraction* describes the process of extracting semantic relationships between these objects.

Automatic implementation of semantic analysis perfects the operation of the information-search system in the national corpus of the Uzbek language. Determining the semantics of the searched lexical units is one of the main tasks in the field of NLP. **Extracting keywords.** It is a branch of semantic analysis and consists of groups of words that can perform different tasks. Keywords include words such as synonym, homonym, antonym, polysemantic, polyfunctional, meronym. Polyfunctional words are one of these elements and are important in semantic analysis. The word *polyfunctional* (poly-many, functional-task) is a word that performs many functions. A question may arise here. What tasks can be performed, what is meant by task? Polyfunctional words are words that have the same form and meaning within the same topic. They can be distinguished only by having answers to questions of different word groups in different sentences.

1. Saodatxon orziqib javob kutar, Tojiboy aka bo'lsa, **aniq** javob bermay, gapni aylantirar edi.
2. Endi kimga og'iz solsang, g'iring demay tegishi **aniq**

In these sentences, the word "aniq" is a polyfunctional word, in the first sentence it belongs to the adjective group, and in the second sentence it functions as a modal word.

2 MATERIALS AND METHODS

The problem of polyfunctionality of lexical units is one of the urgent problems of modern linguistics. The issue of polyfunctional words is often found in Russian, English, Tatar, and Chinese languages in world linguistics. The concept of polyfunctionality appears in some sources in the form of multi-functionality. In many foreign linguistic works, polyfunctionality is considered in connection with the phenomena of homonymy [17]; Kolesnikov 1984; Malakhovsky 1990; Lapteva 1999; Starodumova 2002 and others]. Linguist Gorina Irina Ivanovna in her article “Lekseman slovna polifunksionalnoye slova sovremennogo russkoga yazyka” defines “Multifunctionality is the ability to switch from one fixed part of a sentence to another.” In his research, I.I. Gorina provided detailed information about other polyfunctional usages of the word *slovna* in the Russian language [18]. One of the Chinese linguists Xun Hun found the Chinese equivalent of the polyfunctional words in the Russian language and the Russian-Chinese dictionary in his candidate’s thesis and explained their meanings [19]. Examples of polyfunctional words in English are pronouns. O.D. Vishnyakova proved that the modal word *-able* can be considered as a polyfunctional word with the help of annotations [21]. E.N. Vinogradova separated most of the functions of prepositions in the Russian language [17] and divided them into groups. Tokarchuk, I. N. cited the methods of solving polyfunctionality in the process of POS tagging of the national corpus of the Russian language [25]. Salvadori, J. spoke about the polyfunctionality of adverbs in the French language, based on the lexical-morphological features of adverbs at night [21].

The issue of polyfunctional words has not been fully resolved in world linguistics, and we have not found any research on its automatic detection, elimination in the corpus, linguistic filter and P model. In this article, an attempt was made to summarize the existing theoretical views in the Uzbek language and use them to perform preliminary work on the modeling of polyfunctional words for the semantic analyzer. Filters work to model polyfunctional words. It is important to create mathematical models based on the generated filters and linguistic models. Researcher Sh. Gulyamova says, “Linguistic modeling of polyfunctional words should be defined in terms of categories” [20]. He summarized the existing theoretical views on the Uzbek language and carried out initial work on the modeling of polyfunctional words for the semantic analyzer. The researcher identified a total of 21 groups of polyfunctional words in the section of word groups.

3 MAIN PART

It is necessary to develop a mathematical model of polyfunctional word differentiation for each category. Similar to homonyms, polyfunctional words can be differentiated based on conjunctions. But this method alone is not enough to distinguish all polyfunctional words and create a model. Because some sets of words are determined by the context. Although this is the case, clear laws and models are needed for a system that differentiates polyfunctionality. When Sh. Gulyamova

classified in categories, she developed a linguistic model for the polyfunctional words of each group. Based on the created linguistic models, we will develop legal and mathematical models for the system. We introduce a set of generic polyfunctional words such as Pf. Determining polyfunctionality between noun

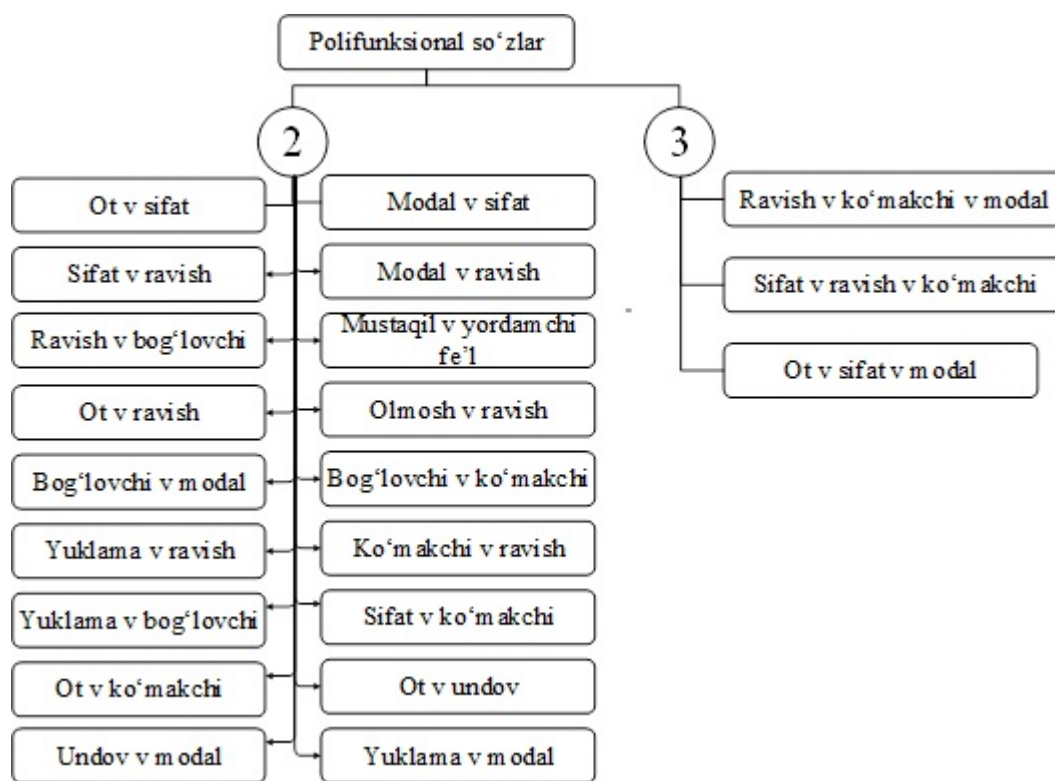


Fig. 1. Cases of meeting polyfunctional words within word groups

√adjectivewordgroups.Wetaketheword”kasal”asanexampleofawordthatcreatespolyfunctionalitybetweenno

“Oʻgʻlingizning kasali maʼlum boʻldi”, – debdi Ibn Sino

Bahonani oshirishmi yoki kasalni yashirish?!

Qoʻshnim kasal sigirini yetaklab qassobxona yoʻl oldi.

Palataga ogʻir kasal bemor joylashtirildi.

In sentences 1-2, the word *kasal* belongs to the noun group. In sentences 3-4, the word *kasal* is an adjective. In these sentences

– *Kasal*+*-i/ni/ga/lar/da/ini/...* aniqlandi/*koʻrmoq/keldi/...*

– *Kasal odam/bemor/hayvon/qush/...*

Based on these analyses, we conclude:

$$Pf_{Adj,N} = \begin{cases} Pf^N, if Pf^{Adj,N} + \downarrow aff^N \oplus V \\ Pf^{Adj}, if Pf^{Adj,N} + \downarrow aff^deg \oplus N \end{cases} \quad (1)$$

Using (1), we propose to determine the polyfunctionality between noun or adjective. Determining polyfunctionality between adjective $\vee$ modal word groups. We define the mathematical model of modal word groups using the exact word and conclude :

Aniq–adjective :

Saodatxon orziqib javob kutar, Tojiboy aka bo'lsa, aniq javob bermay, gapni aylantirar edi

o'pchilik tomonidan ma'qullangan amaliy, aniq takliflar kiritildi

Modal word:

Xolmirza aka ichidan "shu g'ayrating bo'lsa, ikki yilda Xirmontepadan aniq ayrilamiz", deb o'ylaydi.

Endi kimga og'iz solsang, g'iring demay tegishi aniq.

Aniq – modal so'z, shubhasiz, shaksiz so'zlari bilan ma'nodosh.

It can be seen from the sentences that after a polyfunctional word, a verb can be found in both cases, and a noun word can be found before it. It follows that it is possible to distinguish polyfunctionality between adjective $\vee$ modal words by forming a list of compounds for this type of words

$$Pf_{M,Adj} = \begin{cases} Pf^{Adj}, if Pf^{M,Adj} + \downarrow aff^{Adj} \oplus W^{Pf^{Adj}} \\ Pf^M, if W^M \oplus Pf^{M,Adj} + \oplus W^M \end{cases} \quad (2)$$

Using the model (2), polyfunctionality between an adjective or a modal word can be determined.

Determining the polyfunctionality between Independent verb and auxiliary verb. We bring the word "**boq**" to the polyfunctional words between adjectives or adverbs:

Boq –1.to look. 2. To care for a long time by giving food (verb).

Boq– to the lexical meaning of an independent verb, it adds the grammatical meaning of "execution of an action for the purpose of testing, checking" i (auxiliary verb).

Taking into account the above analysis, we present the following mathematical model

$$Pf_{I,L} = \begin{cases} Pf^I, if W^x + aff^{CS} \oplus Pf^{I,L} \\ Pf^L, if V^L + aff^L \oplus Pf^{I,L} \end{cases} \quad (3)$$

Through this model, it is possible to distinguish between polyfunctional words that can become leading and auxiliary verbs.

Determining the poly-functionality between adjectives and adverbs

Let's consider the analysis of "betartib" words as one of the words that create poly-functionality between adjective $\vee$ adverb word groups:

betartib+roq/-gina odam/xona/uy/ko‘cha/
betartib kiyinmoq/yumoq/kelmoq/...

$$Pf_{Adj,Adv} = \begin{cases} Pf^{Adj}, if Pf^{Adj,Adv} + aff^{deg} \oplus N^{PF_{Adj}} \\ Pf^{Adv}, if Pf^{Adj,Adv} \oplus V \end{cases} \quad (4)$$

Using the model (4), it is possible to determine the polyfunctionality of the adjective Vadverb word group.

Determining the polyfunctionality between the conjunction V auxiliary . The conjunction V is modeled using the polyfunctionality event between the auxiliary and the word as follows (if it becomes a conjunction bilan):

Ota bilan bola/ gul bilan lola/oq bilan qora/kelish bilan ketish/oy bilan quyosh/...

Qunt bilan o‘qi/ aql bilan o‘ylamoq/tuni bilan ishlamoq/...

$$Pf_{Conj,Aux} = \begin{cases} Pf^{Conj}, if W^{Conj} \oplus Pf^{Conj,Aux} \oplus W^{Conj} \\ Pf^{Aux}, if W^{Aux} \oplus Pf^{Conj,Aux} \oplus V \end{cases} \quad (5)$$

With the help of this mathematical model, it is possible to identify words that create polyfunctionality between the conjunction and auxiliary words. Similar mathematical models can be developed to identify other groups of polyfunctional words in the hierarchy of the above classification of polyfunctional words. To correctly define polyfunctionality, it is necessary to distinguish the set of words that can be combined with these words. In short, a large amount of context is necessary. With enough context, models can be tested.

There are a number of models of information systems. For example, information model, conceptual model, models representing business processes in the system. In today’s article, we describe the processes that determine polyfunctionality in the system. A number of notations can be used to model business processes. UML, IDEF3, BPM notations are among them. One of the modern notations in business process modeling is BPM notation. The following business processes are modeled using this notation. The information system for semantic analysis of sentences in the Uzbek language performs general processes as shown in Figure 3, and each complex process includes several processes. Initially, the sentence entered by the user is divided into tokens. We make sentences in the Uzbek language using only blank spaces, because commas (,) in the sentence are also important in determining homonymous modal words. The next process is the process of removing unnecessary words (StopWords), which involves removing words from the sentence that do not affect its semantics. Unimportant words in the Uzbek language are found among word groups such as conjunctions, prepositions, pronouns, adverbs. There are also such word combinations that, when encountered in the text, remain insignificant for its semantics. But these words are sometimes important in the composition of texts. Although the list of bigrams and trigrams

of such ambiguous words was separately highlighted by Kh.A.Madatov and Sh.Bekchanov [16], B. Elov and A. Abdullayeva, they formed a list of pure non-essential words in the Uzbek language [22]. A total of 182 purely Stop-Words can be seen in the list of purely stopWords, that is, words that do not affect the semantics of any sentence. In this article, since the processes of semantic differentiation of polyfunctional words in the sentence are considered, we turn to the 4th process in the main fig 3. After removing ambiguous words from the sentence, the main process is to check whether a polyfunctional word is included in the sentence.

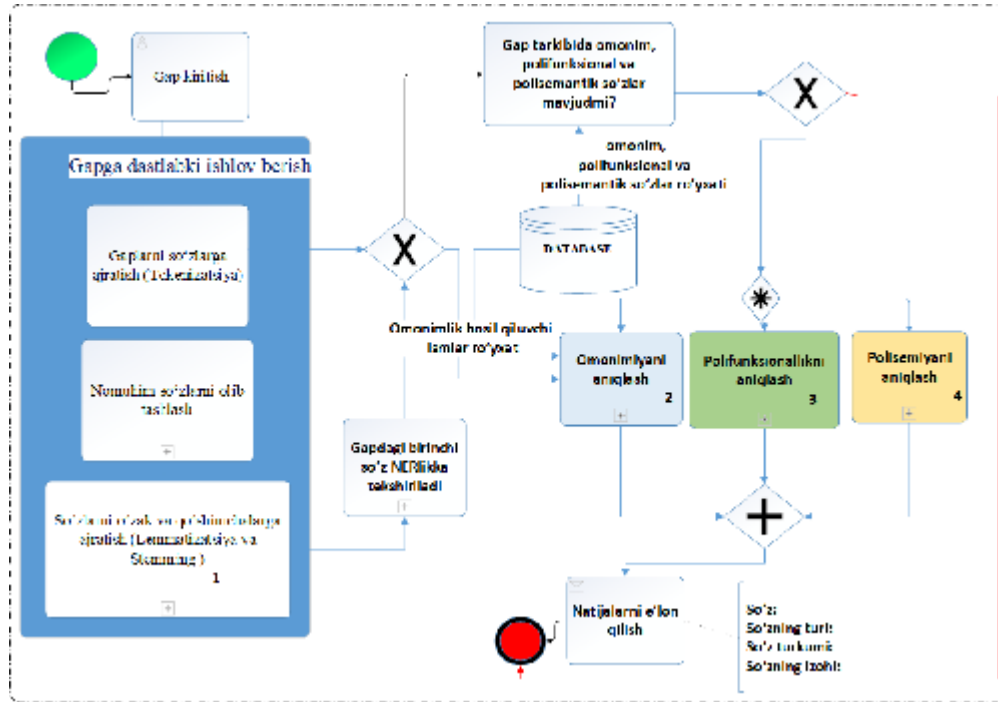


Fig. 2. BPM model of general information system processes.

It is determined how many part of speech meet the polyfunctional word found in the sentence in the business process presented in Figure 4. It is known that polyfunctional words in the Uzbek language are found within two or three parts of speech. A separate rule is developed for each group of polyfunctional words presented in Figure 2. Because polyfunctional words can be defined by their morphological features. In the Uzbek language, polyfunctional words within two part of speech are divided into 15 groups. The main goal of the article is to distinguish polyfunctional words, which are one

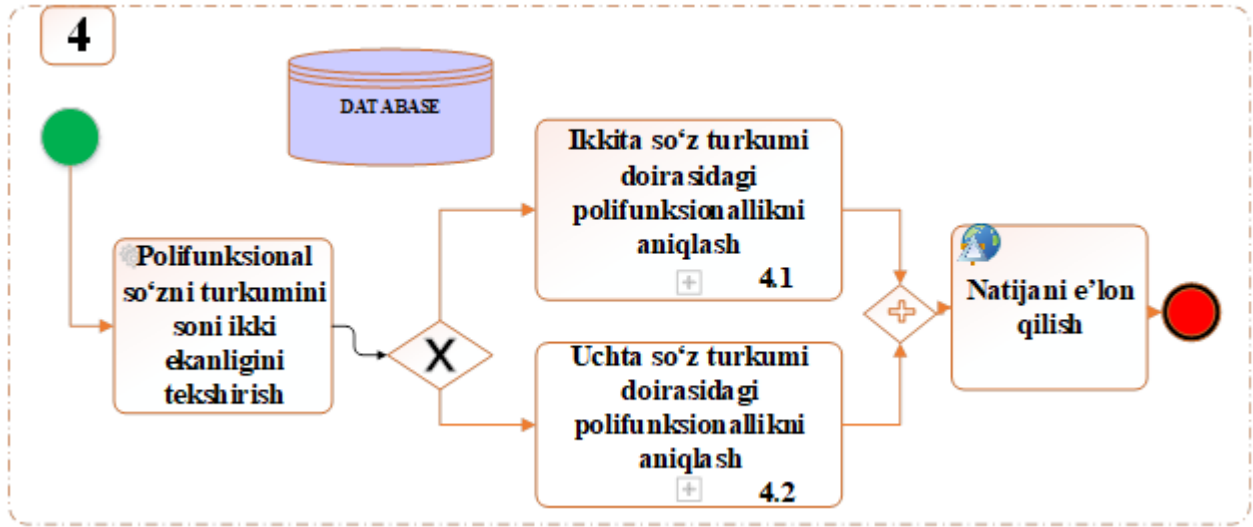


Fig. 3. The process of determining in which part of speech a polyfunctional word can be found

of the elements of the semantic analyzer of the Uzbek language, based on the above-mentioned business processes. Based on the mentioned mathematical models and algorithms, we believe that we have approached the goal. Before developing an information system for semantic analysis of sentences in the Uzbek language, it is required to develop its architecture, structure, and conceptual models. This information system was developed based on the Django MVT (Module View Template) architecture

The capabilities of the Django framework were used in the development of the information system. The structure of the information system for semantic analysis of sentences in the Uzbek language has also been developed. Based on the architecture and structure of the information system that analyzes sentences in the Uzbek language, a conceptual model has also been developed. With the help of a conceptual model, the function of the information system can be understood at a glance. It follows from this that it is possible to see the processes that take place in the identification of polyfunctional words and the methods used in their implementation.

4 Conclusion

The role of polyfunctional words in the development of the information system for the semantic analysis of sentences in the Uzbek language is significant. Polyfunctional words in the Uzbek language can be determined using methods based on basic rules for semantic differentiation. Because they differ in morphological characteristics. In the development of the information system for the semantic analysis of polyfunctional words in the Uzbek

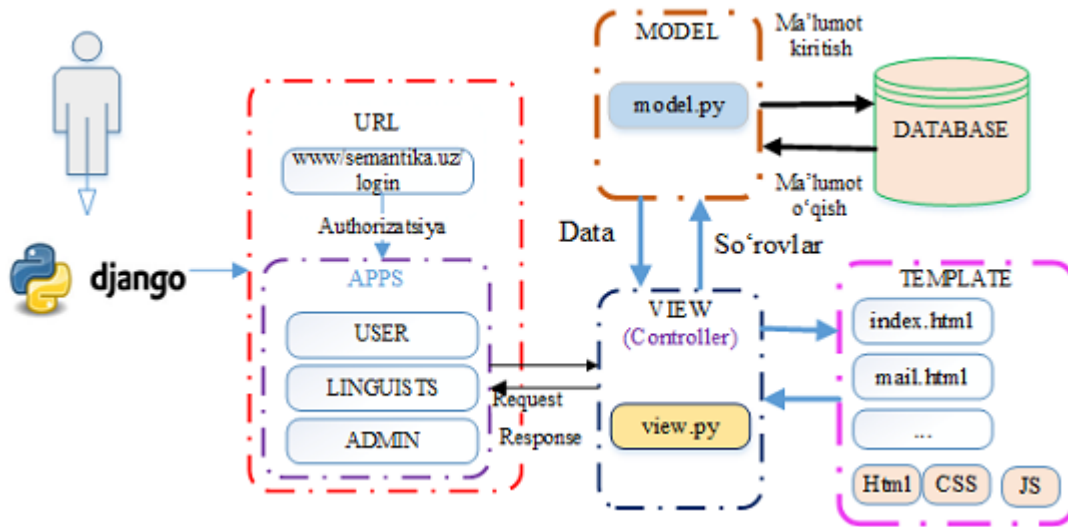


Fig. 4. Information system architecture

language, a number of rules have been developed [1][2][3], each functional process in the information system has been defined and their BPMN models have been developed. In order to differentiate polyfunctional words in the Uzbek language in the context, connections between words are of course of great importance. 101 sentences were entered into the developed information system as a test. The number of exact results obtained from the entered sentences was 71. Overall 71.7 percentage accuracy was achieved. It is advisable to use Machine learning and Artificial intelligence methods to further increase efficiency. These connections are called Grammatical connection in Turkic languages. Semantic analyzers are developed using the interaction of words in a sentence.

References

1. Boltagevich, E. B., Ilxomovna, A. X. (2022). *Business Process Modeling That Distinguishes Homonymy Within Three Parts of Speeches in The Uzbek Language*. In *Proceedings - 7th International Conference on Computer Science and Engineering, UBMK 2022* (pp. 278–283). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/UBMK55850.2022.9919453>
2. Elov B.B., Azmedova X.I. *Uchta so'z turkumi doirasidagi omonimiyani farqlovchi biznes jarayonni modellashtirish*// O'zbekiston respublikasi innovatsion rivojlanish vazirligining, *Ilm-fan va m innovasion rivojlanish ilmiy jurnal* 2022 / 1, 150-162-b.
3. Azmedova X. I. *Turli so'z turkumlari orasidagi omonimiyani aniqlovchi matematik modellar*// *Science and innovation international scientific journal volume 1 issue 7 uif-2022: 8.2 — issn: 2181-3337*. <https://doi.org/10.5281/zenodo.7238546>

4. B.B. Elov, X.I. Azmedova “So‘z ma’nosini aniqlashda Naive Bayes algoritmi-dan foydalanish”, *ILM-FAN VA INNOVATION RIVOJLANISH ilmiy jurnali*, Toshkent, 3/2023,44-54-b
5. Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. *SemEval-2007 Task-17: English Lexical Sample, SRL and All Words*. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 87–92,
6. Nancy Ide and Jean Véronis. *Word Sense Disambiguation: The State of the Art*. *Computational Linguistics*, 1998, 24(1), 1-41-p.
7. Navigli 2009 — Navigli, Roberto. *Word sense disambiguation: A survey*. *ACM Computing Surveys (CSUR)* 41.2: 10. 2009
8. Tripathi, P., Mukherjee, P., Hendre, M., Godse, M., Chakraborty, B. (2021). *Word Sense Disambiguation in Hindi Language Using Score Based Modified Lesk Algorithm*. *International Journal of Computing and Digital Systems*, 10(1), 939–954. <https://doi.org/10.12785/IJCDS/100185>
9. Zouaghi, A., Merhbene, L., Zrigui, M. (2012). *Combination of information retrieval methods with LESK algorithm for Arabic word sense disambiguation*. *Artificial Intelligence Review*, 38(4), 257–269. <https://doi.org/10.1007/s10462-011-9249-3>
10. Basuki, S., Kholimi, A. S., Minarno, A. E., Sumadi, F. D. S., Effendy, M. R. A. (2019). *Word Sense Disambiguation (WSD) for Indonesian homograph word meaning determination by LESK Algorithm Application*. In *Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019* (pp. 8–15). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICTS.2019.8850957>
11. Elov B.B., Azmedova X.I. *Determining homonymy using statistical methods.*// “Hisoblash modellari va texnologiyalari (HMT 2022)” O‘zbekiston-Malayziya ikkinchi xalqaro konferensiyasi materiallari- Toshkent, 2022 16-17 sentabr,-106 b.
12. Uri Roll, Ricardo A. Correia, Oded Berger-Tal// *Using machine learning to disentangle homonyms in large text corpora- Conservation Biology* 31 October 2017 <https://doi.org/10.1111/cobi.13044>
13. J.Y. Park, Shin, H.J.; Lee, J.S. *Word Sense Disambiguation Using Clustered Sense Labels*. *Appl. Sci.* 2022, 12, 1857 <https://doi.org/10.3390/app12041857>
14. Gulyamova Sh.k. *Polifunksionallik xususida* // —O‘zbekistonda ilmiy-amaliz tadqiqotlar//mavzusidagi online konferensiya. 15-avgust. № 19. – Toshkent, 2020. – B. 46-48;
15. Syun Xun *Polifunksionalnye slova v russkom yazyke kak problema russko-kitayskix slovarey : Dis. ... kand. filol. nauk: 10.02.20: Moskva, 2003 196 s. RGB OD, 61:04-10/225-9*
16. Khabibulla Madatov, Shukurla Bekchanov, Jernej Vičič. (2021). *Lists of Uzbek Stopwords (Version 1) [Data set]*. Zenodo. <https://doi.org/10.5281/zenodo.5659638>
17. Vinogradova Ye.N. *Vplotnuyu k narechnym predlogam (kvoprosu o polifunksionalnyx edinitasak)*// *Russkiy yazyk za rubejom* № 1/2021
18. Gorina I.I. *Leksema slovna polifunksionalnoye slova sovremennogo russkogo yazyke*// *Innobatsii i investitsii* №-4, 2014, S-178-180
19. B. B. Elov, *Text generation in Uzbek using N-gram language models*, *Computational linguistics: problems, solutions and perspectives*, *Collection of international scientific and practical conference. Electronic publication, ebook, Tashkent (2022)*

20. Z. Y. Xusainova, *NLP: tokenizatsiya, stemming, lemmatizatsiya va nutq qismlarini teglash. O'zbek amaliy filologiyasi istiqbollari, Respublika ilmiy-amaliy konferensiya to'plami. Elektron nashr, Toshkent: ToshDOTAU, 159-163 (2022)*
21. Salvadori, J., Huyghe, R. (2023). Affix polyfunctionality in French deverbal nominalizations. *Morphology*, 33(1), 1–39. <https://doi.org/10.1007/s11525-022-09401-4>
22. Xu, X., Tokarchuk, I. N. (2022). The main types of usage of a multifunctional word “literally” in modern Russian. *Litera*, (8), 236–249. <https://doi.org/10.25136/2409-8698.2022.8.38594>