Pindus Journal Of Culture, Literature, and ELT
ISSN: 2792 – 1883 **Volume 9**
https://literature.academicjournal.io

From the Experience of Filtering Polysemy Words and Their Use in the Language Corpus

Gulyamova Shakhnoza Kakhramonovna
Doctoral Student of TSUULL (DSc), Doctor of Philosophy In Philology (PhD)

Abstract: The ambiguity of language forms is one of the features of natural language, which is a natural language that contributes to the qualitative development of vocabulary, thereby "saving" verbal material. The problem of lexical ambiguity has been in the focus of attention of researchers for a very long time, and many scientific papers have been conducted on this topic. An important task is to develop the linguistic foundations of a semantic analyzer. To eliminate multitasking, it is necessary to initially develop filters for them. To develop the linguistic basis of the semantic analyzer of the Uzbek language, it is important to first study the experience of filtering polysemantic words and their use in the language corpus.

Key words: Polysemantic word, overcoming polysemy ("removing polysemy"), lexical meaning of the word, language sign, semantic filter, semantic vocabulary, semantic environment, information search system

In corpus linguistics, a number of studies have appeared on the problem of polysemous vocabulary, touching a polysemous unit, eliminating polysemicity in the process of automatic reading of a text ("removing polysemicity") [6; 8; 9]. When solving the problem of lexical ambiguity, the following is necessary:

1. Determine the meaning of each word associated with the text.

2. Choose the most appropriate meaning based on the word that occurs in the context.

First of all, the lack of a clear definition of these concepts creates difficulties in developing and comparing lexical multivalued solutions. The famous Russian linguist V.V.Vinogradov noted the complexity of the meaning of words in modern linguistics as follows: the term "Lexical or Semantic meaning of a word" is considered not very accurate. When we talk about the lexical meaning of a word, we usually understand its subject – the material developed in accordance with the laws of mathematics of a given language, the semantic system of a given language dictionary. It is very difficult to distinguish and integrate relationships in different contexts. These difficulties can be observed in an Explanatory dictionary in confusion and the use of meanings, in insurmountable confusion, in the number of meanings of words and their definition, in constant disagreements. The formalization of the function of many meanings associated with loss is problematic, since there is no generally accepted way to determine where the meaning of a word begins and ends. Many modern scientific works are based on predefined values: a list of words found in dictionaries, translation into a foreign language, and the use of other similar dictionaries [7].

In all texts of the Russian national corpus, there are three types of language signs: metatextual (author, genre of the text, etc.), grammatical (belonging to a lemma and a grammatical sign) and lexico-semantic (lexico-semantic group, types of word formation). Now the first task is to increase the accuracy of labeling and reduce the level of "noise" in search results. Its solution is due to the fact that many synonyms and homonyms take into account different meanings of words and correctly recognize these meanings in the text. In the Corpus Dictionary, each meaning of a word is

Copyright (c) 2021 Author (s). This is an open-access article distributed under the terms of Creative Commons Attribution License (CC BY).To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

provided with a semantic label indicating that it belongs to a particular taxonomic (semantic) class, for example, *lie, roll* 1) movement, action of the subject (*pig's children are lying in a dirty place*); 2) location (*papers are lying on the ground*). In a dictionary, a polysemous word usually has several semantic meanings; these meanings are distributed in different FAMILIES. But when the program automatically puts a character in the text, it writes down all the characters that are present in the dictionary for each occurrence of a word, since the program cannot determine in what sense this word is used in each particular case. As a result, a multi-valued word in the text will have many characters. This often interferes with a more accurate search in the Corpus, creates "noise", sometimes a morphological stroke, as well as a source of errors in lemmatization. To determine the uncertainty of the semantic sign in the body, a special filtering technology was developed. The semantic filter is based on contextual monotony, that is, in each specific context, this word has one meaning (with the exception of word games). A semantic filter is a rule that defines a certain minimal context in which a certain meaning of a word is fulfilled. Thus, the polysemy is removed up to the semantic class (that is, up to the semantic tag). The filter that reduces the uncertainty of the adjective uses only the semantic property of the noun being defined (since the grammatical properties of the noun are gender, Harmony and the fact that it is in the plural/union – does not affect the semantic class (wire) of the adjective). In addition, verb filters use the noun property associated with the verb, but verb filters are more complex than adjective filters. They can use two parameters – semantic classes of nouns associated with the verb, the verb control model (that is, morphological, syntactic features of the noun; another type of subordination is also taken into account – an adverb).

The RusNet-IV resource is being developed for the Russian language under the leadership of I. V.Azarova. Creating such a resource is a very time-consuming process. When it comes to statistical methods, a controlled and unsupervised learning method is used to solve the problem of uncertainty: this method was actively developed using the material of the English language. The technologies of using these systems are actively used in the semantic annotation of the corpus when solving the verb polysemy using many languages, including the control model [5].

The meaning of a polysemous word in the Corpus differs not by a number, as in a simple Explanatory dictionary, but by a semantic sign, for example: sawing (stump) – "physical action", sawing (husband) – "speech". Programs use semantic filters. If each character is put in the corresponding meaning in the dictionary by analogy, then the whole word is put in corpus texts, so it is impossible to distinguish the meaning of the word when automatically placing the symbol. A multi-valued program uses semantic filters. A polysemous word is used in a sentence in a certain sense (without taking into account the correspondence of words), which means that the meaning of the word corresponds to the context. For example, in the dictionary of the corpus verb for indignant (indignant) there are two meanings: "natural phenomenon" and "human behavior". Accordingly, each of them has two characters in the main text. The first meaning is used in the context of nouns – "a natural phenomenon" (the rain is increasing), the second-in the context of nouns – "a person" (a neighbor is angry). The semantic filter includes a context property corresponding to the specified value. As soon as the appropriate context is defined, the program removes the unnecessary icon, leaving the correct one. Thus, many semantics are defined up to semantic classification, that is, semantic touch (of course, not all verb meanings also have a separate sign) [5].

When solving the multiplicity problem, it will be necessary to optimize the semantic dictionary, that is, to establish a hierarchy of values, list them in places. Configuring the search according to the main meaning of the word provides a representation of the most reliable meaning. In the same spirit, the ordered application of the meaning of a word is easy and is an effective tool that predicts sufficient compatibility of the representation [10].

Copyright (c) 2021 Author (s). This is an open-access article distributed under the terms of Creative Commons Attribution License (CC BY).To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

The problems of using polysemous lexemes in the Uzbek language are highlighted in the works of D.Akhmedova. The researcher emphasizes the convenience of the faceted method when describing many significant words, given that a word can express more than one meaning. Also, based on the semantic filter, it will be possible to determine which area the plural word belongs to. The researcher cautiously states that if the semantic stroke is not implemented correctly, if a filter is not developed that generates all the meanings of a multi-valued word, the cyberslabels of the Uzbek language Corpus will be left behind even from traditional dictionaries [3]. In her dissertation, she points out that the markup of the corpus usually acts depending on the phrase of the unit "neighbor" in the definition of many values, a set of rules for creating such an algorithm (all typical and rare cases) should be included by a linguist in the database of linguistic support [2].

Scientist M.Abdurakhmanova and researcher A.Rakhmanov in the article "Polysemy in corpus linguistics", emphasized the importance of computer methods, in particular, in the selection of alternative units in the translation process, in mastering the semantics of polysemous words, using the colloquial, that is, the method of semantic blockade. The meanings of polysemous words are analyzed as part of a syntactic blockade. It was noted that the method of word combinations (right and left collocates) can also be used in the analysis of polysemous words, only in this place the specifics of the Turkic languages, in particular the Uzbek language, based on the subordinate-dominant relationship in these compounds, will be revealed. The article explains that it is more correct to distinguish what is before or after the word, that is, there is no part standing to the right or left of the word [5]. In our opinion, this method can also be called correct aggregation of macros. Because the essence of the word, sema, manifests itself when a series is combined with another word.

The researcher analyzed the method of semantic blockade on the example of the main lexeme with a high frequency of use in the Uzbek language, as well as the potential for the formation of hyphenations. In the Uzbek language, the head is a polysemous word that can be understood with the help of the corpus of this word-forming polysemicity. The information search engine will highlight texts where the word "head" is used. Head – a part of the body above the neck, anterior (in humans, in animals); skull. The corpus shows that this word is used in the 20th place in the text from the explanatory dictionary, it comes from the composition of 60 phrases. He considered the meanings of this word as part of the syntactic environment.

Thus, the problem of the emergence of a polysemous word and its entire meaning (colloquial application) is an urgent problem waiting to be solved in corpus linguistics. An important task is to conduct experiments on the development of a system of semantic filters, which is used to determine the ambiguity of a lexical and semantic nature when creating a national corpus of the Uzbek language. It is necessary to determine to what extent it is possible to use a lexicographic resource that specializes in creating such a filter.

REFERENCE

1. Абдураҳмонова М., Рахманова А. Корпус лингвистикасида полисемия // "Компьютер лингвистикаси: муаммолар, ечим, истиқболлар" Республика I илмий-техникавий конференция. – Vol. 1 №. 01 (2021).
2. Ахмедова Д. Атов бирликларини ўзбек тили корпуслари учун лексик-семантик теглашнинг лингвистик асос ва моделлари. Филология фанлари бўйича фал. док-ри. дисс. – Бухоро, 2020. – 247 б.
3. Ахмедова Д.Б. Семантик разметка тизимида кўп маънолилик ва фильтр. Бердақ номидаги Қорақалпоқ давлат университетининг ахборотномаси. – 2020 йил 4-сон. – Б. 202-205.

Copyright (c) 2021 Author (s). This is an open-access article distributed under the terms of Creative Commons Attribution License (CC BY).To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

4. Кустова Г.И. Семантические фильтры для разрешения многозначности в национальном корпусе русского языка: глаголы // Компьютерная лингвистика и интеллектуальные технологии по материалам ежегодной международной конференции «Диалог» (2008) Периодическое издание, выпуск 7 (14).
5. Кустова Г.И., Толдова С.Ю. Национальный корпус русского языка: семантические фильтры для разрешения многозначности глаголов // tudiorum-ruscorpora.ru/
6. Рахилина Е.В., Ляшевская О.Н., Кобрицов Б.П., Кустова Г.И., Шеманаева О.Ю. Многозначность как прикладная проблема: Лексико-семантическая разметка в Национальном корпусе русского языка // Лауфер Н.И., Нариньяни А.С., Селегей В.П. (ред.). Компьютерная лингвистика и интеллектуальные технологии: Труды между народной конференции “Диалог 2006”, – 2006. – С. 445-450.
7. Селегей В. Лингвистические проблемы автоматического создания интернет-корпуса русского языка / В. П. Селегей // Инновации и высокие технологии: тр. 55-й науч. конф. МФТИ. – М.: Изд-во Московского физико-техн. ин-та, 2012. – С. 53-54.
8. Толдова С.Ю., Кустова Г.И., Ляшевская О.Н. Семантические фильтры для разрешения многозначности в Национальном корпусе русского языка: глаголы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4-8 июня 2008 г.). – Вып. 7 (14). – М.: РГГУ, 2008.
9. Шеманаева О.Ю, Кустова Г.И., Ляшевская О.Н., Рахилина Е.В. Семантические фильтры для разрешения много значности в Национальном корпусе русского языка: прилагательные // Иомдин Л.Л., Лауфер Н.И., Нариньяни А.С., Селегей В.П. (ред.). Компьютерная лингвистика и интеллектуальные технологии: Труды между народной конференции «Диалог 2007». – 2007. – С. 582-587.
10. Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)…диссер. – Қарши, 2019.

Copyright (c) 2021 Author (s). This is an open-access article distributed under the terms of Creative Commons Attribution License (CC BY).To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/