



Journal Website:
<https://theusajournals.com/index.php/ijll>

Copyright: Original content from this work may be used under the terms of the creative commons attributes 4.0 licence.

ANALYSIS OF SOME EXCEPTIONS IN HOMONYMY AND HOMONYMIC MODELS IN UZBEK ELECTRONIC CORPUS MANAGER

Submission Date: May 21, 2024, Accepted Date: May 26, 2024,

Published Date: May 31, 2024

Crossref doi: <https://doi.org/10.37547/ijll/Volume04Issue05-10>

Fayruza Alijonova

Lecturer, Tashkent medical academy, Uzbekistan

ABSTRACT

This article discusses the analysis of some homonymous units that create problematic situations in the process of searching for homonyms and dividing them into tokens in the Uzbek electronic corpus manager, and the alternative models that can be developed for them. Addressing the following homonymy problems and developing models for software can help overcome homonymy-related shortcomings for computational linguistics, particularly corpus linguistics.

KEYWORDS

Corpus, model, homonymy, tagging, token, computational linguistics, corpus linguistics, automatic editing.

INTRODUCTION

In World Corpus Linguistics, the issue of homonymy unit tagging and homonymy elimination during automatic text reading has been separately researched. But since Uzbek corpus linguistics is a young field, there are very few studies devoted to the issue of semantic tagging and differentiation of homonyms.

Therefore, we rely on the sources of Turkic languages and Russian linguistics in the study of the research conducted to distinguish lexical homonyms.

According to M. Abjalova, in order to eliminate homonymy, each word should be "classified", that is, it can be compared with a lemma - a phrase and a set of morphological features, which are added to one tag for

convenience. To explore all possible tags, it is sufficient to find relevant references to words in a morphological dictionary, or to use a morphological analyzer such as MyStem, which will help you find word tags. After that, you only need to select the appropriate tag from among several tags. The optimal linguistic method used in the analysis of homonymous word forms is an important factor in the processes of text editing and analysis, machine translation, and text processing.

The Main Findings and Results

As a result of the study of homonyms, it was found that the following cases cause the main problems in searching for homonyms in the corpus and dividing them into tokens.

1. Old Uzbek words that are now out of use.

Sh. Rahmatullayev stated that homonyms are determined primarily within the framework of language. For example, it is not appropriate to say that a lexeme not used in modern Uzbek (belonging to the old Uzbek language) and a lexeme in modern Uzbek are mutually homonymous. According to D. B. Akhmedova, lexemes belonging to the old Uzbek language are often found in the text of works written in the modern Uzbek language, taking this into account, some historical lexemes that are homonyms of the lexemes in the modern Uzbek language have been included in the dictionary. For example:

dol I Noun. In a children's game, the situation that gives the right to hit the gan (kubba) first is that the sokka is closer to the gan. "Who is dol?" - "I am dol."

dol II Noun, old. The name of the eighth letter in the Arabic alphabet. This word is written with zol, not dol.

dol III Adverb, outdated; Exactly. Aim at the center of the target.

As one of the first requirements is to ensure representativeness in the language corpus, it is natural to have texts that reflect the language of different historical periods. Therefore, the problem of distinguishing the phenomenon of homonymy in the lexeme of historicism should also find its solution. Homonymy should be defined separately in relation to literary speech and colloquial speech. Accordingly, the dictionary mainly describes lexemes belonging to the modern Uzbek language, and includes a limited number of colloquial words. But in the corpus of the language, homonyms used in literary speech and colloquial speech are used in the same way. Homonymous forms that are not reflected in the dictionary can be observed in the corpus of the language. It is necessary to take into account that this phenomenon will become a separate object of research only when a corpus is created in the Uzbek language, texts specific to literary speech and colloquial speech find their expression in the corpus, and an array of examples appears.

2. Some common and similar nouns.

In the composition of the texts, there are ambiguities in differentiating certain nouns and adjectives. This situation is mainly observed in words that have moved from adjectives to nouns. For example, adjective words such as beautiful, brave, and brave have been moved to nouns and are considered nouns. Common nouns are always capitalized. We can use the [HW → W.cl → Noun] model to model such homonyms. In this case, HW → homonym, if the word is written with a capital letter (W.cl), then this word is a noun. W.cl is a capital letter. However, if some adjectives that have been

transferred to nouns appear at the beginning of the sentence and are written with a capital letter, then according to this model, the program can read the word of the adjective group as a noun. Such homonyms can only be determined from the context.

3. Words with accent.

As a result of the introduction of borrowed words from a foreign language into our language, homonymous words with different meanings have appeared in the Uzbek dictionary. For example, the word soya has different meanings depending on the accent.

Accent I. If the accent falls on the last syllable of the word soya, it is used in the sense of a cool place, where the sun does not shine.

Accent II. if it falls on the first syllable of the word soya, it is used in the sense of a consumer product. When analyzing words with such emphasis in the texts, it is advisable to collect the base of compounds that come before and after them.

4. Some word forms resulting from the addition of adverbs.

Homonymy is formed as a result of adding some formative and grammatical additions to the base of the word. Let's see some examples and their models.

changi

I. Changi: Noun+i (possessive suff.).
[Noun+N.gsh=HW1.N]

II. Changi: Noun+I (verb yas). [Noun+i (yas)=HW2.V].

terim

I. Terim: Noun+im (possessive suff.).
[Noun+N.gsh=HW1.N];

II. Terim: Noun+m (possessive suff.).
[Noun+N.N.gsh=HW2.N];

III. Terim: Verb+im (maker). [Verb+im=HW3.N].

ko'zlar

I. Ko'zlar: Noun+s (plural). [Noun+N.gsh=HW1.N]

II. Ko'zlar: Noun+la (maker)+r (sh.yas).
[Noun+la+V.gsh=HW2.V]

sana

I. Sana: Noun+a (verb yas). [Noun+a (yas)=HW1.V]

II. Sana: Noun. [HW2.N]

ko'tara

I. Ko'tara: Verb+a (adjective yas). [Verb+a (yes)=HW1.Adj]

II. Ko'tara: Verb + (verb. sh). [Verb+V.gsh=HW2.V].

yuvindi

I. Yuvindi: Verb + in (verb verb) +di (verb tense).
[Verb+V.ns+V.zm=HW1.V]

II. Yuvindi: Verb+ indi (noun). [Verb+indi(yes).
[Verb+now(age)= HW2.N]

suvchi

I. Suvchi: Noun+chi (noun yas). A personal noun is made from a relative noun. [Noun+chi=HW1.N]

II. Suv-chi: Noun+ -chi (reinforcement-emphasis load). [Noun+ -chi.part=HW2.N]

To distinguish the word suvchi, we use the above models, that is, if the suffix -chi is added to the word, it is a formative affix. If the suffix -chi is written with a dash after the noun, it is a preposition. The short form part represents a charge.

senla

I. Senla: Pron+la (verb) = Verb. [Pron+la(yas)=HW1.V]

II. Sen-la: Pron+ -la (abbreviated form of auxiliary bilan). It means with you. [Pron+ -la (Pr.)= HW1.Pron.Pr]

We use the above model to distinguish senlamoq and sen-la forms of the verb group. Here, if -la is added to the word, the formative affix and the verb form a lexeme.

If -la is hyphenated in a word, it represents the abbreviated form of the auxiliary.

5. Words in adjective and noun group.

Due to the fact that the words of the adjective noun group have the same grammatical forms, errors in categorization arise during the analysis process. Adjectives take possessive, accusative, and plural suffixes specific to nouns. In order to differentiate the words of this category, it is advisable to collect the base of combinations with which they are combined. But even then, these models do not completely

eliminate errors in classification. It is only through the context of the sentence that the group of words can be determined.

REFERENCES

1. Abjalova M. A. Homonymy and methods of identifying homonyms in linguistic systems./ Oriental Renaissance: Innovative, educational, natural and social sciences ISSN 2181-1784 Scientific Journal Impact Factor SJIF 2021: 5.423. 2021. -B. 1020. www.oriens.uz
2. Akhmedova D.B. Problems of tagging homonyms in the Uzbek language. // International scientific-practical remote conference on "Application of digital systems in modern education: modern trends and development factors in the field of philology and pedagogy". Tashkent, (May 1) 2020. - B.344-346.
3. Rahmatullayev Sh. An explanatory dictionary of homonyms of the Uzbek language. -Tashkent: Teacher, 1984.
4. An explanatory dictionary of the Uzbek language. - Tashkent: National Encyclopedia of Uzbekistan, 2008. -5 vol. -592 p.
5. Gulyamova Sh. Linguistic foundations of the Uzbek language semantic analyzer: Doctor of Philological Sciences (DSc) dis. autoref. -Fergana, 2022. -77 p.
6. Rachmanova. A. Computer methods in creating the national corpus of the Uzbek language: philology. science. Doctor of Philosophy (PhD) ...diss. autoref. -Tashkent, 2021. -52 p.