

LINGUISTIC AND MATHEMATICAL MODELING OF GRAMMATICAL HOMONYMS

¹Gulyamova Shahnoza Kahramonovna, ²Akhmedova Kholiskhan Ilhomovna, ³Suyunova
Malika Adil qizi

¹Doctor of philological sciences, Tashkent State University of Uzbek Language and Literature

²Doctoral student of Tashkent State University of Uzbek Language and Literature

³Master's student in Computer Linguistics, Tashkent State University of Uzbek Language and
Literature

<https://doi.org/10.5281/zenodo.7991107>

Abstract. *In this article, homonyms, which are considered one of the problems in corpus linguistics, and their elimination in the corpus, are studied in world and Uzbek linguistics, proposals and solutions, the development of a linguistic filter for lexical homonyms, the importance of grammatical homonyms in automatic text analysis, grammatical homonym identification, collation, linguistic and mathematical modeling are discussed. It has said that grammatical homonyms are one of the important issues to be studied in corpus linguistics.*

Keywords: *language corpus, homonym, grammatical homonym, right and then conjunctions, morphological filter, syntactic filter, N-gram, linguistic and mathematical model.*

Introduction

In corpus linguistics, which is consider a new field in our linguistics, a lot of is doing on automatic text reading and analysis. Corpus is comprehensive set of texts, through which we can learn the statistics of homonyms, morphological, syntactic and semantic properties. It has the most convenient source of information for obtaining information. It is much easier to carry out various scientific studies in the corpus, and the accuracy rate is high. The problem of automatic processing of natural language remains relevant for more than half a century. The complexity of the problem and the lack of the clear idea indicate the difficulty of ways to solve it. All the news systems for recognizing text, speech, paralinguistic tools are being develop. Text processing is one of the oldest and most important researches in this field.

The main part

The problems of identifying homonyms in the corpus and their elimination have been widely studied in world linguistics. In particular, G. I. Kustova, O. N. Lyashevskaya, Ye. V. Paducheva, Ye. V. Rakhilina, B. P. Kobritsov, T. I. Reznikova, V. V. Kukanova, A. A. Kretov have carried out a number of works dedicated to the solution of these issues.

In Uzbek linguistics, the creation of corpora and conducting scientific research on them are widely developed. In Sh. Hamroyeva's doctoral dissertation, she found a solution to the problem of automatic differentiation of the homonymy of predicates in the context of homonymy with other grammatical forms in the word, linguistic form inhabitants and form-forming morphemes in order to determine the homonymy and ambiguity of the word-forming morpheme in the context.

In our linguistics, lexical homonyms are model to eliminate homonymy in the corpus. At firstly, homonyms are divided into groups according to the series of words. For instance, homonyms within nouns only, homonyms within verbs only, homonyms between nouns, nouns and adjectives, homonyms between nouns and nouns. Taking into account that such homonyms occur in a large number and among different word groups, all of them are dividing into 68 chapters.

This classification of homonyms based on their nature is very important in creating and modeling filters for them. Because not all words fit the same filter.

After classification by word groups, morphologically different homonyms are separating. Which words have combined with which suffixes and it determines which category the homonyms belong to word. Such homonyms work well for nouns and verbs, adjectives and adverbs.

A syntactic filter is giving to distinguish morphologically indistinguishable homonyms through right and after conjunctions. Summarizing all the above methods, linguistic and mathematical models are developed for homonyms in each word group.

Classification of lexical homonyms according to their category, meaning, morpho-semantic, syntactic filter, linguistic and mathematical models will be a solution for the semantic analyzer to identify and eliminate homonyms in the automatic reading and analysis of the text in the corpus. Of course, this work was carrying out only within the framework of lexical homonyms, and it serves as a programming guide for further scientific work, including solving the problem of identifying and distinguishing grammatical homonyms in the corpus.

Grammatical homonyms mainly occur between lexeme and word form and between word form and word form. In this case, many grammatical suffixes are adding to noun, adjective and verb word groups, or with a new lexeme and a word that has received a grammatical form, or homonymy of two words in the same word group that has received a grammatical form. Such affixes occur between word-forming and form-forming affixes, forming grammatically homonymous words, and this occurs in three cases:

Table 1.

Grammatical forms that create grammatical homonymy

		-ingiz			-ma			
		-lar			-ish			
		-siz			-ay			
		-cha			-in			
		-i			-ir			
		-im			-ng			-ish
Noun+	aff ^N	-da	Verb+	aff ^V	-gan	Adject+	aff ^{Adj}	-i
		-dan			-indi			
		-ga			-ar			
		-m			-iq			
		-imiz			-moq			
		-ing			-ing			
					-ingiz			

Grammatical homonyms are the first way words have found

In the first case, if we add suffixes such as *-ingiz*, *-lar*, *-siz*, *-cha*, *-i*, *-im*, *-da*, *-dan*, *-ga*, *-m*, *-imiz*, *-ing* to the noun group, some of the words formed we can see the formation of grammatical homonyms. For example: *teringiz* (noun) – *teringiz* (verb); *bog'lar* (noun) – *bog'lar* (verb); *onasiz* (adjective) – *onasiz* (noun+proposition); *yigitcha* (noun+diminutive form) –

yigitcha (adverb); *changi* (noun) – *changi* (verb); *tilim* (noun) – *tilim* (noun); *boshda* (noun) – *boshda* (adverb); *burundan* (noun) – *burundan* (adverb); *tanga* (noun) – *tanga* (noun); *terim*, *tering*, *teringiz* bring about grammatical homonymy in the form of noun and noun.

Grammatical homonyms are the second way words have found

In the second case, adding suffixes such as *-ma*, *-ish*, *-ay*, *-ir*, *-ng*, *-gan*, *-indi*, *-ar*, *-iq*, *-moq*, *-ing*, *-ingiz*, *-in* some are grammatical homonyms. For example: *tugma* (noun) – *tugma* (verb); *oqish* (adjective) – *oqish* (verb); *aylanay* (exclamation) – *aylanay* (verb); *burang* (adjective) – *burang* (verb); *burgan* (noun) – *burgan* (verb); *yuvindi* (noun) – *yuvindi* (verb); *ochar* (noun) – *ochar* (verb); *ochiq* (adjective) – *ochiq* (verb); *quymoq* (noun) – *quymoq* (verb); *tiling* (noun) – *tiling* (verb); *teringiz* (noun) – *teringiz* (verb); *yig'in* (noun) – *yig'in* (verb) grammatically form homonymy.

A third way in which grammatical homonyms have found

When the diminutive suffix *-ish* is added to the word of the “*oq*” adjective group, the word “*oqish*” is formed, which means that the color is small and when the action noun suffix *-ish* is added to the verb “*oq*”, we can see that a grammatically homogenous word is formed with the verb “*oqish*”. Also, the formative suffix *-i* is added to nouns and adjectives and becomes homonymous with the possessive form *-i*. For example, *boyi* (adjective + possessive) – *boyi* (adjective + verb) and *changi* (noun + possessive) – *changi* (noun + verb).

A filter to detect grammatical anonymity

Grammatical homonyms cannot be detected by morphological filter in automatic analysis. Because homonyms of this type are homonyms only in the process of speech, when they take grammatical forms in a sentence, and this is not a permanent phenomenon. In another sentence, these grammatical forms may change and homonymy may be lost. For example:

- Hikmatli *so'zlar* kishilarning miyasiga singib qoladi, ildiz otib, gullaydi, hosil beradi va hamisha ta'sir ko'rsatib boradi.
- Qalandarov minbarda og'ir, vazmin *so'zlar*.

The words used in these sentences are both nouns and verbs, forming grammatical homonymy. If the words used in the second sentence are using in the verb form, then grammatical homonymy is lost. Morphological additions create grammatical homonymy only in some cases. Therefore, the morphological filter does not work in such cases.

A syntactic filter, i.e. identifying the right and the right conjunctions of a word, and modeling on this basis helps to identify grammatical homonyms. On this, we use bigrams and trigrams. For example, when an “*aralashma*” noun is a group of words, it is preceded by the *quyidagi*, *kimyoviy*, *maxsus*, *sutli*, *sodali*, *spirtli*, *yopishqoq* and similar adjectives; It is combined with demonstrative pronouns such as this, that, those, these. After itself, the verb has combined with a group of words and auxiliary words: *aralashma ishlab chiqardi*, *aralashma hoida tayyorladi*.

When the verb is a group of words, it has preceded by a pronoun, followed by nouns in the form of the case of departure. For example, *sen aralashma*, *bu ishga aralashma*, *suhbatimizga aralashma*, *masalaga aralashma*, etc.

The **-ma** form has added to some words when it becomes a noun, an adjective, and an infinitive suffix of a verb, forming grammatical homonymy. The number of such words was 78. The Uzbek language is very diverse, we use such words every day in our conversation, and given that they do not have a dictionary form and do not always adopt the same grammatical form, it is

quite difficult to determine the exact number of grammatical homonyms and collect them. We only found out what grammatical forms that create grammatical homonyms are added to words and which of those words are grammatical homonyms using the morphological analyzer of the Uzbek language. Among all the words with the **-ma** form, we collected those that can be homonyms between nouns, adjectives and verbs.

A linguistic model for grammatical homonyms

A general model for these words would look like this:

$$V + \text{Neg_aff} = V_{gr}$$

Here, V – is a verb, Neg_{aff} – is a form of infinitive, V_{gr} – is a grammatical homonym of a verb.

$$V + f_1^N = N_{gr}$$

Here V – is a verb, f₁^N – is a noun-former, N_{gr} – is a grammatical homonym of a noun.

$$V + f_1^{ADJ} = ADJ_{gr}$$

In this case, V – is a verb, f₁^{ADJ} – -ma is an adjective, ADJ_{gr} – is a grammatical homonym of an adjective group. The patterns for grammatical homonyms between these three word groups are:

Table 2.

Models of grammatical homonyms

V _{gr} – the verb is a grammatical homonym	N _{gr} – the noun is a grammatical homonym	ADJ _{gr} – the adjective is a grammatical homonym
N+Dat_CS_aff + V _{gr}	N _{gr} + N+PS_aff	ADJ _{gr} + N
ADV + V _{gr}	N _{gr} + V	ADJ _{gr} + V
Wp + Acc_CS_aff + V _{gr}	N+Dat_CS_aff+N _{gr}	ADJ + ADJ _{gr} + N
Wp + Abl_CS_aff + V _{gr}	N _{gr} + II	ADJ _{gr} + N+ PL_aff
PR+ V _{gr}	ADJ+ N _{gr}	PR + ADJ _{gr}
ADJ+ V _{gr}	PR+ N _{gr}	ADV + ADJ _{gr} +N
Wp+ II+ V _{gr}	N _{gr} + N+PL_aff + PS_aff	NUM + ADJ _{gr} + N
	V _s + N _{gr}	
	N+Gen_CS_aff + N _{gr} + PS_aff	
	N+II+ N _{gr}	
	Wp + Acc_CS_aff + N _{gr}	
	NUM+ N _{gr}	

	$Wp+d^N + N_{gr}$	
	$N+Loc_CS_aff + N_{gr}$	
	$N_{gr} + N$	

In the modeling of our grammatical homonyms, from the models in the monograph "Linguistic bases of the semantic analyzer of the Uzbek language", model of grammatical forms from Sh. Hamroyeva's doctoral dissertation "Linguistic support of the morphological analyzer of the Uzbek language" and the tags of the morphological analyzer of the Uzbek language has been used:

- V – verb;
- N – noun;
- ADJ – adjective;
- ADV – adverb;
- NUM – number;
- PR – pronounce;
- II – auxiliary words;
- V_{gr} – the verb is a grammatical homonym;
- N_{gr} – the noun is a grammatical homonym;
- ADJ_{gr} – the adjective is a grammatical homonym;
- Neg_aff – infinitive form of the verb;
- V_s – adjective form of the verb;
- Dat_CS_aff – dative case;
- Acc_CS_aff – accusative case;
- Abl_CS_aff – ablative case;
- Gen_CS_aff – genitive case;
- Loc_CS_aff – locative case;
- PS_aff – possessive form;
- PL_aff – plural form;
- Wp – word previous;
- d^N – the locative suffix "-dagi" of the noun.

These models were developing only for grammatical homonyms with **-ma** form.

Separate filters and models are developed for grammatical homonyms of -lar, -da, -dan, -ing and similar forms. Grouping these according to their grammatical form and developing a model for these groupings is more effective in identifying grammatical homonyms.

The first research on automatic text processing dates back to the 1950s. Automatic text processing has divided into several stages, one of which is morphological classification. At this stage, morphological descriptions (conjugation, declension, type, etc.) and the initial form of the word called lemma has defined for each wordII. Morphological classification has complicated by the phenomenon of homonymy. For texts in some inflected languages, homonymy detection methods based on the use of probabilistic models are very common, but they provide very high accuracy.

Based on the developed linguistic models, mathematical models and algorithms have developed and the results have analyzed. Mathematical models have based on grammatical rules. Therefore, mathematical models might consider as a rule-based method of determining grammatical homonymy.

It has known that the suffix **-ma** forms the infinitive form of the verb and in addition serves as a noun and an adjective. Grammatical homonymy means identifying the exact function of this suffix. For example, if we analyze “aralashma” word found in a sentence.

Bu ishlarga umuman *aralashmadim*.

Bu *aralashmaga* nimalar qo‘shilgan?

We offer a mathematical model to “*aralashma*” where the suffix **-ma** forms an adjective and where it forms the infinitive form of the verb. If a word contains the suffix **-ma**, it can be determined by looking at the suffixes that follow it. That is, if the syntactic and lexical form of the adverb after the suffix **-ma** consists of adverbs and their combinations, this adverb can consider as an adverb. If the suffix **-ma** has a suffix that has added to a noun phrase, then this suffix is a noun-forming suffix. Based on these rules, the following mathematical model can give.

$$H_{gr} = \begin{cases} W + aff^{neg} + aff, & aff \in aff^V \\ W + aff^{noun} + aff, & aff \in aff^N \\ W + aff^{adj} + aff, & aff \in aff^{Adj} \end{cases}$$

In this case, W – is the root word, aff – is a suffix or a combination of suffixes following the suffix, aff^N – is a suffix added to the noun group and a set of their combinations; aff^V – a set of suffixes added to the verb group and their combinations; aff^{Adj} – a suffix added to an adjective phrase and a set of their combinations.

Here the question arises - what if there is no suffix after the suffix **-ma**? That is

Sen bu ishlarga *aralashma*.

Probirkadagi qanday *aralashma*?

Of course, this mathematical model does not work in such cases. In such cases, it can be determined using bigrams and trigrams of words. The use of word breakers is determined using statistical and machine learning algorithms.

Conclusion

Today, great results are achieving in the field of computer linguistics in our linguistics. A morphological analyzer of the Uzbek language has created, and a full morphological analysis of the text became possible. Identifying grammatical homonyms and collecting them helps solve one of the problems in automatic text analysis. Dividing grammatical homonyms into groups based on their grammatical form and developing separate linguistic and mathematical models for these groups helps to identify and eliminate grammatical homonyms in the morphological analyzer process.

REFERENCES

1. Gulyamova Sh. Linguistic foundations of the Uzbek language semantic analyzer. Monograph - Tashkent: 2021.
2. Khamroeva Sh. Linguistic support of the morphological analyzer of the Uzbek language. Tashkent: 2021.

3. Rahmatullaev Sh. Explanatory dictionary of homonyms of the Uzbek language. - Tashkent: Teacher, 1984. - 108 p.
4. <http://uznatcorpara.uz/uz/POSTag>
5. Park, J. Y.; Shin, H.J.; Lee, J.S. Word Sense Disambiguation Using Clustered Sense Labels. Appl. Sci. 2022, 12, 1857. <https://doi.org/10.3390/app12041857>
6. Elov B.B., Akhmedova K.I. Modeling the business process distinguishing homonymy within three word groups// Journal of the Ministry of Innovative Development of the Republic of Uzbekistan, Scientific Journal of Science and Innovative Development 2022 / 1, pp. 150-162.
7. Akhmedova Kh. I. Mathematical models for identifying homonymy between different word groups// Science and innovation international scientific journal volume 1 issue 7 uif-2022:8.2|issn: 2181-3337. <https://doi.org/10.5281/zenodo.7238546>
8. Akhmedova K.I. Determining homonymy using the frequency method // "PROSPECTS OF UZBEK APPLIED PHILOLOGY" Republican Scientific and Practical Conference Tashkent: 2022.-164-170 p.
9. Elov B.B., Akhmedova K.I. Determining homonymy using statistical methods. //"Computational models and technologies (HMT 2022)" Proceedings of the second Uzbekistan-Malaysia international conference - Tashkent, 2022 September 16-17,-106 p.
10. 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>