

ACADEMICIA

ISSN (online) : 2249-7137

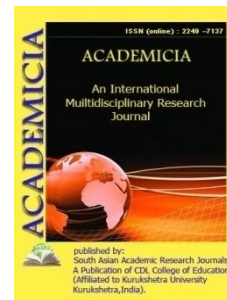
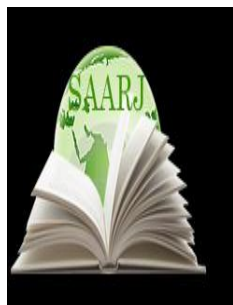
ACADEMICIA

An International
Multidisciplinary Research
Journal



Published by
South Asian Academic Research Journals
A Publication of CDL College of Education, Jagadhri
(Affiliated to Kurukshetra University, Kurukshetra, India)

225.	METHODS FOR DETERMINING HOMONYMS IN HOMONYMY AND LINGUISTIC SYSTEMS Manzura Abdurashetovna Abjalova, Aziz Uygun ogli Yuldashev	10.5958/2249-7137.2021.00522.X
226.	CRAFT AND TRADE RELATIONS IN EASTERN BEKS OF THE BUKHARA EMIRATE Allamuratov Shukhrat Ashurovich	10.5958/2249-7137.2021.00536.X
227.	PECULIARITIES OF THE MORPHOPHENOTYPE AND CHARACTERISTICS OF THE PHYSICAL PERFORMANCE OF YOUNG FOOTBALL PLAYERS AND THEIR RELATIONSHIP WITH THE GAMING AMPLITUDE Mustafaeva Shargiya Axmatovna. Rakhmatova Markhabo Rasulovna, Jalolova Vazira Zamirovna, Mavlyanov Zafar Iskandarovich	10.5958/2249-7137.2021.00537.1
228.	PROGRESS IN THE STUDY OF THE PATHOGENESIS AND DIAGNOSIS OF HYPERTENSION Jumaeva Gulrukh Aliyorovna	10.5958/2249-7137.2021.00538.3
229.	HEALTHCARE SYSTEM IN THE SURKHANDARYA REGION: REFORMS AND RESULTS (ON THE EXAMPLE OF THE KUMKURGAN DISTRICT) Khuzhanazarov Absalom Ziyadullayevich	10.5958/2249-7137.2021.00539.5
230.	ROLE FORMATION OF ECOLOGICAL THINKING AND EDUCATION IN HIGHER EDUCATION INSTITUTIONS Sharafutdinova Rumiya Infarovna, Muratova Gulsara Saitovna	10.5958/2249-7137.2021.00540.1
231.	THE ACQUISITION OF A SUBJECTIVE COLOR AS A RESULT OF THE CONVERSION OF THE MEANING OF THE WORD Mirzakarimova Z.D	10.5958/2249-7137.2021.00523.1
232.	STUDY OF THE COMPLEX-FORMING ABILITY OF POLYGALACTURONIC ACID WITH TIN IONS Mohira Madumarovna Nurmatova, Ilnar Nakipovich Nurgaliev, Sayora Sharafovna Rashidova	10.5958/2249-7137.2021.00524.3
233.	THE PUBLICIST SKILLS OF BEHBUDIY Rustam Sharipov, Diyora Abdurahimova	10.5958/2249-7137.2021.00525.5
234.	THE COURT PROCEEDINGS CONNECTED WITH THE DECISION OF ARBITRATION: IN THE EXAMPLE OF CIVIL CASES Zamira Normurotovna Esanova	10.5958/2249-7137.2021.00526.7
235.	THE EFFECT OF ORGANIZATIONAL FACTORS ON CREATIVITY IN THE IRANIAN ACADEMIC CENTER FOR EDUCATION, CULTURE & RESEARCH UREMIA BRANCH Vali Aghli, Habib Valizadeh, Shiva Habibzadeh, Amin Rostamzadeh	10.5958/2249-7137.2021.00548.6
236.	COMPARATIVE ANALYSIS OF PHRASE LOGICAL FUSIONS IN THE FRENCH AND UZBEK LANGUAGES Kurbonova Gulsara Sodikovna	10.5958/2249-7137.2021.00489.4
237.	THE ROLE OF PARENTS IN INVOLVING CHILDREN WITH DISABILITIES IN GENERAL SECONDARY EDUCATIONAL INSTITUTIONS Gulnoza Salixova	10.5958/2249-7137.2021.00495.X
238.	PATRIOTIC IDEAS IN THE WORKS BY FITRAT: THEORETICAL AND PRACTICAL HARMONY Mavluda Rahmonova	10.5958/2249-7137.2021.00490.0



DOI: **10.5958/2249-7137.2021.00522.X**

METHODS FOR DETERMINING HOMONYMS IN HOMONYMY AND LINGUISTIC SYSTEMS

Manzura Abdurashetovna Abjalova*; **Aziz Uygun ogli Yuldashev****

*Doctor of Philosophy (PhD) in Philology,
Tashkent State University of Uzbek Language and Literature named after Alisher Navoi,
UZBEKISTAN

Email id: abjalova.manzura@gmail.com

**Candidate for a Masters Degree in Computer Linguistics,
Tashkent State University of Uzbek Language and Literature named after Alisher Navoi,
UZBEKISTAN

Email id: yaziz.90@mail.ru

ABSTRACT

As you know, homonymous lexemes differ from polysemous words. Polysemantic words have semantic or formal connections, and homonymous tokens are monosyllabic words with different meanings. Both phenomena are considered an urgent problem in the linguistic processing of texts. In particular, the definition of homonymy is important for translation and analysis in Natural Language Processing (NLP). This article discusses several ways to identify and analyze homonymous forms in Uzbek texts using word combination models.

KEYWORDS: *Homonym Form, Method, Hidden Markov Model, N-Gramma, Model Of Word Combinations.*

INTRODUCTION

Automatic reprocessing of texts is divided into several stages, and one of them is the stage of morphological analysis. At this stage, each word undergoes a morphological description: lemma [1] (base), agreement, government, number, level, ratio, in person, case, gender, and etc. The task of morphological subdivision of words in morphological analysis is complicated by homonyms.

As it is known, homonyms are lexical units in which the forms of words are the same, but their meanings, semantics are different. The phenomenon of survival in automatic reprocessing is found in the following units:

1. Homonymous words are based on the same forms of words, that is, a certain word possesses meanings belonging to one category of words or several categories. For example:

<i>ot(noun)</i>	<i>modal(modal)</i>
<i>Avval birliklarni, keyin o'nliklarni qo'shamiz. – (first we add singular numbers, then tenth numbers, units)</i>	<i>Suv bor joyda hayot bor. Ruchkang bormi? (there is, have/has) where there is water their is life/ Do you have a pen?</i>
<i>ot(noun)</i>	<i>hisob so'z(word used to count)</i>
<i>Kuch – birlikda. (Power is in unity)</i>	<i>Bir necha bor taklif yubordim.(times) – I've sent invitations several times</i>
<i>ot (noun)</i>	<i>ot(noun)</i>
<i>Og'irlik birliklari.Til birliklari Weight unit; language units</i>	<i>Bor – kimyoviy element. – Bor is a chemical element</i>
	<i>fe'l (verb)</i>
	<i>Ishga bormoq. Maktabga bormoq. to go to work, to go to school</i>

In the first column, the word “birlik” (unit) is a homonym within the framework of a series of words: 1) a whole number up to ten; grammar plurals (noun); 2) communion, harmony (noun); 3) a measure taken for the mutual evaluation of quantities of one kind; a terminology specific to the construction of the language (noun).

In the second column, the word “bor” has formed a form belonging to two different categories: 1) mavjud– exist (modal); 2) marta, dafa, bora– times, folds (accounting word)

2. Suffix homonym-belongs to a different group of additives according to a certain additional function. For example:

-ki	<i>ko'chki, tepki, turtki ustki, ichki, kechki</i>	(ot yasaydi) - forms a noun (sifat yasaydi) – forms an adjective
-(i)ng	<i>uying, kitobing, ishing ko'ring, boring, tayyyorlang</i>	(shakl yasaydi: sintaktik mun.shakli – egalik qo'shimchasi) – makes a form: a form of case (genative case, possessive case) (shakl yasaydi: sintaktik mun.shakli –

shaxs-son qo‘shimchasi)

3. Phrase homonym-phraseological units in which the form, that is, the structure is the same, but the semantics are different.

qattiq shovqin soldi
cried loudly, made a loud noise

– yuksak darajada izzat-hurmat qildi – respect highly

uyni boshiga ko‘tarmoq

onasini boshiga ko‘tarmoq – carry his mother
on his head (respect highly)

4. Sentence homonymy is a specific sentence which differs according to the purpose or content of the expression.

darak gap – declarative sentence

so‘roq gap – interrogative sentence

*Ishni bajarmadim – I haven’t done the work**Ishni bajarmadim? – Haven’t I done the work?*

Analysis of Source

In the types of homonyms in the Russian language, homonyms, morphological homonyms and lexical homonyms are distinguished by the category of words [2]. It is noteworthy that the grouping of words by morphological and lexical homonyms plays an important role in transformation of natural language.

In morphological homonymy, the words Lemma (base), which belong to one category, are different, but only in certain forms they become the words which make up the homonymy, and are taken into account. For example:

lemmasi *ter*lemmasi *terim*

terim – mening terim(skin – my skin
possessive inflexion)

terim – hosil(havesting)

(-im egalik qo‘shimchasi: I shaxs, birlik)

In lexical homonymy, the word belonging to one Lemma gives different meanings:

ot

bog‘ – to‘dalab bog‘langan holat – pile –
pile of things

*Bog‘lamoq, bog‘lam. Bir bog‘ piyoz – to tie,
to pile. A bunch of onion.*

ot

bog‘ – o‘simlik va daraxtlar ko‘p ekilgan
joy – garden where trees and
pant grow

bog‘-rog‘, bog‘-bo‘ston;

Uzumini ye, bog‘ini surishtirma.

- *Garden- prospering garden, flowering
garden*

Morphological and lexical grouping of homonyms is important for lemmatization and stemming processes in text reprocessing for software and systems [3]. Both technologies are focused on finding the basis of a word or phrase,

It is worth noting that the phenomenon of survival is the most urgent issue in the process of automatic reprocessing. Therefore, in NLP, the identification of homonyms and their analysis are specially studied, even several methods have been developed.

Methods

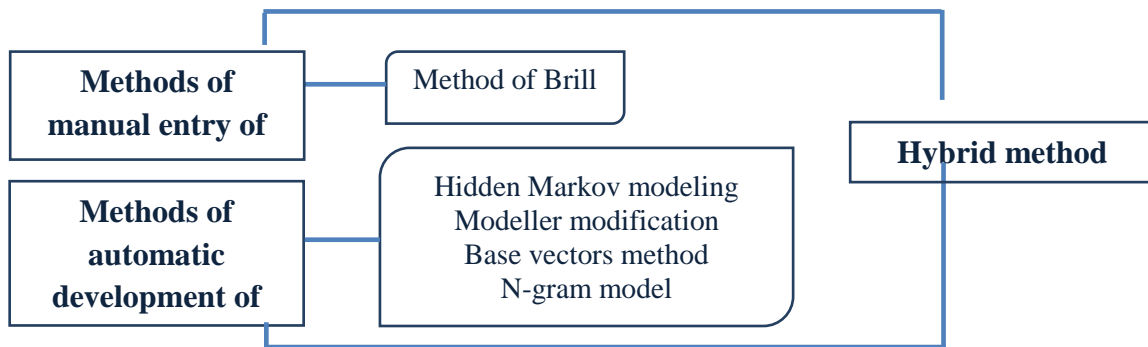
All of the methods for determining homonymy are divided into two groups:

1. Rules-based methods. In turn, they are divided into:

- a) manual input methods of rules.
- b) methods of automatic production of rules.

2. Methods based on statistics.

Each of these groups has its own advantages and disadvantages. As often happens in such situations, combining the characteristics (and advantages) of both groups in one way can show a better result than the results achieved before. Such kind of method is called a hybrid method.



1-chart. Determination of homonymy methods.

Systems that rely on these methods, in turn, make the following groups:

1. Systems based on rules created manually.
2. Systems based on probabilistic models and based on the described corpses.
3. Hybrid systems based on probability models and rules.

To determine the homonymy, each pronoun needs to be "classified", that is, it is connected by its Lemma, a series of words and a set of morphological features that combine into one tag.

Hidden Markov model Baum L.E. and this model, developed by his colleagues [6], helps to take into account the probability of all variants occurring in the statistical process, the statistical method of determining homonymy. For example, if in a certain text words related to the category of the noun meet more often and more often than the connecting, then the homonym present in the same context will be a word related to the category of the noun, not the connecting one with a greater probability, in the next case it will be considered as a connecting. N-Gramm is used to describe the context. N -mama is a mathematical calculation tool that is widely used in automatic

reprocessing of texts. Uzbek computer linguistics S.Rizayev represented letter combinations with the terms bigram, trigram [5].

N – gramma-represents a sequence of N-identifying elements, such as words or labels. The sequence of two elements-bigramma, while the sequence of three elements is called the trigramma. For example, the **prefix+noun** status is an example of a bigram.

To bring a description of the simple statistical method for determining homonymy and other similar techniques, the following methods are used:

- w_i – a word in the sentence in the i -place, t_i – a pronoun identifier (tag) of this word.
- $(D_{(w)} = \{t_1^w, t_2^w, \dots, t_k^w\})$ all possible set of characters in the word w . This information can be obtained with the help of a morphological dictionary. If the word is not in the dictionary, it can be considered as a noun phrase, as done in the Brill method, but it is necessary to put all possible tags so that the linguistic supply is reliable.
- The number of specific cases in the corpora (n-grams). In there $C(t)$ – the number of T tags; and $C(t_1, t_2)$ – the number of bigrams (t_1, t_2) .
- $C_i(w, t)$ – w the number of words a t tag.
- $F(w, t)$ – The probability that the word w has a tag t . Descriptions are calculated using the following formula:

$$F(w, t) = \frac{C_t(w, t)}{C(t)}$$

- $P(t_i|t_{i-1})$ – this is the state of the probability of the arrival of the t_i tag after the t_{i-1} tag. In this case, when $i = 1$, the t_i tag is the first tag in the sentence. The calculation formula is as follows

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})}$$

Based on the Hidden Markov model, as a result of working with the help of the statistical method of determining homonymy, in the sentence with the length n , when $T_i \in D(w_i)$ is $T = \{T_1, T_2, \dots, T_n\}$ the probable sequence of labels is found [7].

Results

In conclusion, when the methods of elimination of homonymy in the world computer linguistics were studied, the appropriate verification formula of vocabulary in Uzbek texts was created using the experience in this regard. To eliminate homonymy, it is necessary to “classify” each word, that is, it can be compared with a lemma – sentence fragment and a set of morphological features, they are added to one tag for convenience. To study all possible tags, it is enough to find related links to words in the morphological dictionary or use a morphological analyzer such as My Stem, which will help in finding the word tags. After that, it will be necessary to choose only the appropriate tag among several tags.

The optimal linguistic method used in analyzing homonyms will be an important factor in the processes of editing and analyzing texts, machine translation, text processing.

REFERENCES:

1. Bolshakova E. I. Automatic text processing in natural language and computer linguistics. Textbook / Bolshakova E. I., Klyshinsky E. S., Lande D. V., Noskov A. A., Peskova O. V., Yagunova E. V. – Moscow: MIEM, 2011. – 272 p.
2. Porokhnin A. A. Analysis of statistical methods for removing homonymy in texts in Russian. Bulletin of AGTU. Ser.: Management, computing technology and Computer Science. – 2013. No. 2. – pp. 168-174.
3. Rakhmatullayev Sh. Explanatory dictionary of homonyms of the Uzbek language. – Tashkent: O'qituvchi, 1984. – P. 5.
4. Abjalova M. Linguistic modules of analysis and editing programs: Monograph. – Tashkent, 2020. – P. 25-27.
5. Rizayev S. Fundamentals of linguostatistics in Uzbek linguistics. – Tashkent: Fan, 2006. – P. 18.
6. Baum, L. E.; Sell, G. R. Growth transformations for functions on manifolds. Pacific Journal of Mathematics. 27(2) 1968. – P. 211–227.; https://en.wikipedia.org/wiki/Hidden_Markov_model.
7. http://www.academia.edu/15517740/Analysis_of_statistical_algorithms_of_morphological_homonymy_in_Russian_language.
8. https://www.researchgate.net/publication/320729376_Using_machine_learning_to_disentangle_homonyms_in_large_text_corpora