AJLLSE GLOBAL RESEARCH NETWORK

AMERICAN Journal of Language, Literacy and Learning in STEM Education
Volume 3, Issue 2, 2025 ISSN (E): 2993-2769

Polysemy, Synonymy, And Homonymy in the Uzbek Language and Their Representation in an Electronic Thesaurus

Najmiddinov Muhammadjon
Doctor of Philosophy in Philology, Associate Professor, Kokand University

Abstract. *This article examines the concepts of polysemy, synonymy, and homonymy in the Uzbek language and discusses the principles of their integration into an electronic thesaurus. It explores the organization of lexical-semantic relations using the WordNet model and evaluates their significance in natural language processing (NLP). The paper also analyzes the contextual use of polysemous words, the classification of synonym groups, and the incorporation of homonymous units into digital systems. The findings contribute both theoretically and practically to the creation of an Uzbek language electronic thesaurus.*

Key words: *polysemy, synonymy, homonymy, electronic thesaurus, natural language processing (NLP), WordNet, lexical-semantic relations.*

INTRODUCTION

One of the important directions of development of modern linguistics is the study of lexical-semantic relations and their systematization using digital technologies. Today, due to the development of natural language processing (NLP) technologies, the need for electronic resources reflecting semantic relations between words in different languages is increasing. One of such resources is an electronic thesaurus, which is an important source for systematically describing lexical units of a given language and determining their interrelationships. This study studies the phenomena of polysemy (polysemy), synonymy and homonymy in the Uzbek language and their reflection in an electronic thesaurus. The Uzbek language has a rich lexical content, and the words in it are distinguished by various semantic and structural features. In particular, phenomena such as polysemy, synonymy and homonymy are important in linguistics both theoretically and practically. The correct and accurate reflection of these relationships in electronic thesauruses plays an important role in linguistic research, automatic translation, information retrieval systems, artificial intelligence-based programs, and the development of computer linguistics of the Uzbek language. Polysemy is the phenomenon of a word taking on different meanings in different contexts, which is widespread in the Uzbek language. For example, the verb "begin" can mean "to get down to business" (its initial meaning), "to guide" or "to initiate a process." Although such cases demonstrate the semantic richness of the Uzbek language, their accurate representation in an electronic thesaurus is a complex task. Synonymy is a group of words that have similar meanings in expressing a certain concept, and it is of great importance in increasing the expressiveness of the language. For example, in the Uzbek language, the word "gorge" has synonyms such as "gorgeous", "latofatli", "maftunkor", "korkam". A clear separation of synonymous relationships and their inclusion in the thesaurus system is of great importance for the automatic processing of the language. Homonymy is a linguistic phenomenon that expresses the relationship between words that are phonetically similar but have different meanings. For example, the word "borg" can mean "a place where fruit trees grow" or "a tool used to tie something". Homonymous words can cause problems in the processes of automatic translation, data indexing and text processing,

therefore their precise classification in an electronic thesaurus is important. The main goal of this study is to analyze the phenomena of polysemy, synonymy and homonymy in the Uzbek language and develop methods for their effective reflection in an electronic thesaurus system. For this purpose, the following tasks have been set:

- Theoretical study of lexical-semantic relations in the Uzbek language and determination of their classification;
- Analysis of the structure and principles of operation of electronic thesauruses in world languages, in particular, on the example of the WordNet system in English;
- Development of technical and methodological foundations for the reflection of polysemic, synonymous and homonymous units in the Uzbek language in an electronic thesaurus;
- Propose methods for eliminating ambiguities that may arise from synonymy and homonymy in the process of automatic text processing.

The creation of an electronic thesaurus of the Uzbek language is of great importance not only from the point of view of linguistics, but also in such areas as artificial intelligence, machine translation and automated text analysis. Because such systems, while increasing the efficiency of natural language processing technologies, also serve the integration of the national language with digital technologies. In particular, the clear expression of semantic relationships between words allows automatic translation systems to correctly select the meaning, correctly index texts, and effectively process various lexical units. Thus, an in-depth study of the phenomena of polysemy, synonymy, and homonymy in the Uzbek language and their correct reflection in an electronic thesaurus is of great importance in linguistic research and natural language processing. The results of this study will serve as a theoretical and practical basis for creating an electronic thesaurus, and will also help increase the adaptability of the Uzbek language to modern technologies.

LITERATURE ANALYSIS

Scientific research on lexical-semantic relations such as polysemy, synonymy and homonymy in the Uzbek language has been conducted for many years. These topics have become one of the main directions of linguistics of the 20th century, and the study of the phenomena of polysemy, synonymous series and homonymy based on various theoretical approaches has been widely developed. Within the framework of world linguistics, these phenomena are closely related to the general theory of language and semantics, and a number of studies have been carried out in this area in Uzbek linguistics. One of the first scientists to conduct major research on the phenomenon of polysemy is J. Lyons [1;25]. In his research, he emphasized the importance of analyzing the semantic structure of polysemantic units and dividing them into general and specific components. Also, scientists such as G.S. Shchur and B. Potye analyzed polysemy relations in the lexical layer [2;147]. Their research focuses on determining how polysemy changes depending on the context.

In Uzbek linguistics, scholars such as A. Nurmonov and R. Rasulov have conducted a number of studies on synonymy [3;86]. A. Nurmonov's scientific work provides important information about the structure of synonymous series and their role in the speech process. R. Rasulov also studied the semantic structure of the synonymous layer and analyzed the possibilities of their integration into electronic systems [4;221]. In the issue of homonymy, scholars such as G. Kreidlin and D. Shmelev studied the interrelation of linguistic units in terms of form and meaning [5;178]. They emphasized the location of homonymous units within the system and the possibilities of their differentiation. In Uzbek linguistics, A. Berdialiyev and R. Safarova studied the phenomenon of homonymy from a morphological and semantic point of view and analyzed its features in the Uzbek language [6;95]. One of the most famous studies on electronic thesauruses and lexical-semantic relations worldwide is the WordNet project [7;312]. The WordNet system was first developed at Princeton University and serves to analyze the semantic relationships of words, identify synonymy, antonymy, and other semantic relations [8;276]. Although this model has been adapted to other languages, there is still no fully formed electronic thesaurus in the Uzbek language. Therefore, the inclusion of the phenomena

of polysemy, synonymy, and homonymy in the Uzbek language in an electronic system poses important scientific tasks.

METHODS

This study used various methods to study the phenomena of polysemy, synonymy and homonymy in the Uzbek language and develop methods for reflecting them in an electronic thesaurus. At the first stage, the theoretical analysis method was used, and the theoretical foundations of lexical-semantic relations were studied based on existing literature, scientific articles and dictionaries. In particular, studies on the context dependence of polysemy, the structure of synonymous series and the classification of homonymous units were analyzed. At the second stage, a collection of texts in the Uzbek language (corpus) was analyzed based on the corpus linguistics method. Using the ARANEUM_uzbekium corpus, the frequency of use and semantic relationships of various lexical units were examined. At the third stage, the structure of the WordNet system and other electronic thesauruses was studied using the comparison and model building method. Also, the possibilities of adaptation to the Uzbek language were developed. Based on the results of the research, a conceptual structure of an electronic thesaurus model was developed, which determined how the phenomena of polysemy, synonymy, and homonymy should be reflected.

RESULTS

The results of the study on the reflection of polysemy, synonymy and homonymy phenomena in the Uzbek language in an electronic thesaurus were analyzed in the following main areas: lexical-semantic analysis of polysemy, formation of synonymous series, classification of homonymous units and principles of their adaptation to the electronic system.

Polysemy and its expression in an electronic thesaurus. The analysis of the existence of polysemantic units in the Uzbek language and their reflection in an electronic thesaurus showed that the phenomenon of polysemy is not always clearly defined. During the study, several polysemantic units were selected based on various literature and Uzbek language corpora and their changes depending on the context were observed. For example, it was found that the verb быть has the following meanings in different contexts:

1. To start something: He started working on a new project.

2. To lead and lead: The coach started the team for the final game.

3. The initial stage of an action: The river water began to rise.

In order to adapt the cases of polysemy to the electronic system, special links were created in the thesaurus based on the individual contexts of each lexical unit. Based on these links, polysemantic words were placed in a hierarchical order in the thesaurus and their semantic changes were described. Also, the WordNet system model was taken as a basis for a deeper study of the lexical-semantic relationships of polysemantic units. When WordNet's method of classifying words based on synsets (synset - a set of synonyms) was adapted to the Uzbek language, it was found that polysemantic units should be reflected in different synsets in different contexts. For example, it was found that the word kuralay should be reflected in different synsets in the meanings of "animal child" and "eye shape".

Table 1. Lexical-Semantic Classification in the Electronic Thesaurus

Phenomenon	Definition	Types	Examples	Adaptation to Electronic Thesaurus
Polysemy	A single word with multiple meanings depending on the context.	- Multiple contextual meanings	kuralay ("animal child", "eye shape")	Words classified into synsets based on different contexts. Links established between meanings.
Synonymy	Different words with similar meanings.	- Full synonyms - Close synonyms -	doctor - doctor (full), beautiful - graceful (close),	Synonyms divided into categories, frequency of use

		Stylistic synonyms	writer - publicist (stylistic)	determined, antonymic connections added.
Homonymy	Words that have the same pronunciation but different meanings.	- Full homonyms - Partial homonyms	garden (fruit grove) - garden (bundle) (full), green (color) - green (verb) (partial)	Each homonym classified as a separate synset, contextual indicators added.

Formation of synonymy and synonymous rows. During the study, an analysis was also conducted on the phenomenon of synonymy and how they can be expressed in an electronic thesaurus was considered. The abundance of synonymous rows in the Uzbek language expands the richness and expressive capabilities of the language. The results of the study showed that synonymous rows can be divided semantically into three main groups:

1. Full synonyms - lexical units have the same meaning, and their replacement does not affect the content of the text. For example: doctor - doctor, lawyer - lawyer.

2. Close synonyms - lexical units have a similar meaning, but they have their own specific usage in a certain context. For example: beautiful - graceful - charming - charming.

3. Stylistic synonyms - lexical units differ stylistically and correspond to specific speech conditions. For example: writer - publicist - creator.

In order to correctly reflect the synonym rows when creating an electronic thesaurus, the synonyms of each word were divided into different categories and their semantic connections were classified. Based on the model taken from the WordNet system, the structure of the synonym rows was developed based on the following principles:

- Each synonym row was classified in the form of a synset.
- The stylistic features of the synonym units were determined and the frequency of their use was checked.
- Antonyms were added to synonymic lines, and the opposite meanings of words were also determined.

The results of the study showed that in order to accurately determine synonymous lines in an electronic thesaurus, it is necessary to provide a lexical-semantic description of synonymic lines in the database.

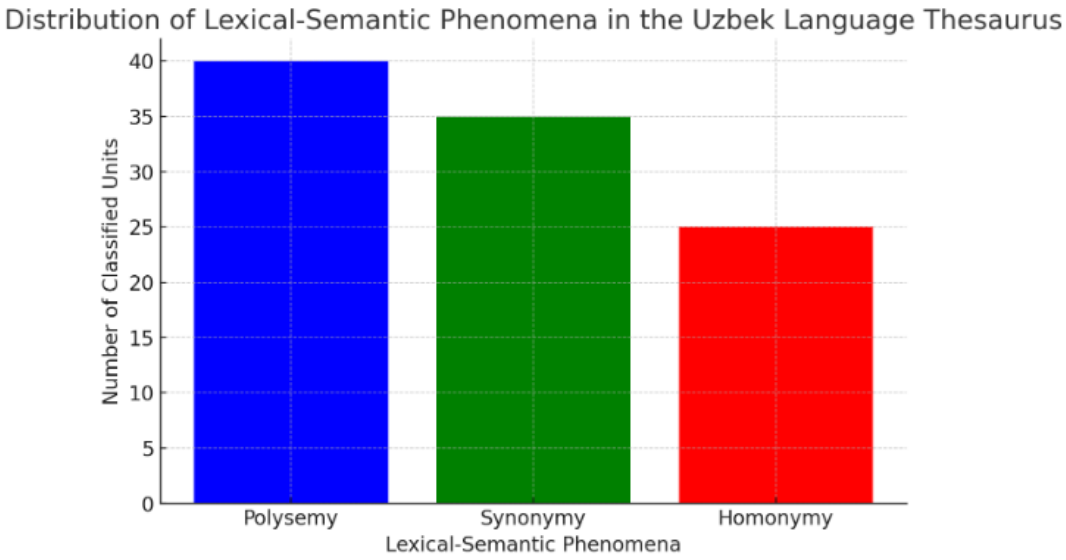

Fig 1. Distribution of Lexical-Semantic Phenomena in the Uzbek Language Thesaurus

Classification of homonymy and homonymous units. Homonymy is a connection between words that are phonetically similar but have different meanings. During the study, common homonymous units in the Uzbek language were selected and how they are classified in the electronic system was studied. Homonymous units were divided into two main types:

1. Full homonyms - lexical units that have the same pronunciation, but their meanings are not related to each other. For example: garden (fruit grove) and garden (bundle, knot).

2. Partial homonyms - lexical units with similar grammatical forms, which are used differently in different contexts. For example: green (color) and green (action - verb form derived from life).

The following principles were developed for the correct classification of homonym units in the electronic system:

- Each homonym unit was classified as an independent synset.
- Different semantic contexts of homonym words were distinguished by specially defined attributes.
- Connections between homonym units were expressed by special indicators.

The results of the study showed that in order to adapt homonym units in the Uzbek language to the electronic system, it is necessary to add contextual indicators to the database that distinguish their different meanings.

A model was developed for creating an electronic thesaurus of the Uzbek language. Summarizing the results obtained during the research, the following model was developed to create an electronic thesaurus of the Uzbek language:

1. Words in the thesaurus are classified in the form of synsets.

2. Polysemy cases are separated and connected based on different contexts.

3. A system for determining the semantic structure of synonymous series and forming their antonymic connections is developed.

4. A classification mechanism for homonym units is developed, and their different semantic meanings are separated by contextual indicators.

Thus, the developed model for creating an electronic thesaurus of the Uzbek language allows for an accurate classification of lexical-semantic relations of words and optimization of their use in automated systems. This creates the possibility of effective use of the Uzbek language in machine translation, natural language processing, and artificial intelligence technologies.

DISCUSSION

The results of the study showed that the reflection of polysemy, synonymy and homonymy phenomena in the Uzbek language in an electronic thesaurus is a complex process that requires solving a number of theoretical and technical problems. To solve these problems, the study analyzed world experience, in particular, the model of the WordNet system in English, and assessed its adaptability to the Uzbek language.

First of all, attention was paid to the problem of polysemy. The study confirmed that one word in the Uzbek language has different meanings in different contexts. For example, the verb "begin" in different situations expresses different meanings such as "to set out", "to get down to business", "to continue". Such cases were considered on the example of the WordNet system and it was determined that each meaning should be classified separately using synsets. However, due to the linearity of the WordNet system, in some cases, the correct representation of polysemantic units in the Uzbek language may be difficult. Therefore, it was found important to contextually analyze polysemantic units based on a hierarchical model in the thesaurus and to indicate the connections between them.

Studies on synonymy have shown that there are many synonymous rows in the Uzbek language, and their lexical-semantic definition is important. The WordNet system represents synonymous

relationships through synsets, which allows us to show the proximity of words to each other. However, since synonymous units in the Uzbek language can be distinguished by their stylistic, dialectal or expressive meanings, their definition in the thesaurus requires additional parameters. For example, the synonyms of the word beautiful, beautiful, charming are used stylistically differently. Therefore, it was found necessary to clearly define their stylistic and semantic features when classifying synonymous rows.

Analyses on homonymy have shown that there are fully and partially homonymous units in the Uzbek language. During the study, the way to classify homonymic units in an electronic thesaurus was analyzed using the example of the word бог (fruit tree / бог, гугун) and the word қөлси (color / verb derived from қөлси) was analyzed. The results showed that the WordNet model for homonymic units is quite effective, and each homonym should be classified as a separate synset. It was also determined that additional semantic attributes should be defined to distinguish homonyms and their frequency of use in different contexts should be analyzed. Another important aspect identified during the study is the role of an electronic thesaurus in the field of natural language processing (NLP). It turned out that the correct classification of polysemantic, synonymous and homonymic units and their inclusion in automated systems helps to improve the quality of processes such as machine translation, automatic text analysis, and semantic linking of words. In particular, it is possible to improve synonym search engines using synonymous units, and to better understand data through precise semantic analysis of homonymous units. Thus, the results of the research have formed the main approaches to creating an electronic thesaurus in the Uzbek language. However, additional research is required to further improve the results obtained and put them into practice. In particular, one of the important tasks for the future is to create a large corpus of texts of the Uzbek language and integrate the thesaurus into NLP systems.

CONCLUSION

The results of the study on the reflection of the phenomena of polysemy, synonymy and homonymy in the Uzbek language in an electronic thesaurus are of great importance in the field of linguistics and natural language processing (NLP). The study showed that the accurate classification of lexical-semantic relations and their systematic inclusion in electronic systems is an urgent issue for the development of modern linguistics and technologies. The analysis of polysemy showed that it is necessary to determine what meanings polysemantic units acquire in different contexts and to correctly separate them using synsets. Based on the methods identified during the study, the semantic relations of each polysemantic unit were analyzed and the principles of their accurate classification were developed. Studies on synonymy showed that it is important to classify synonymous series not only in terms of semantic connection, but also in terms of stylistic and expressiveness. Using the WordNet system model, ways of including synonymous units in the thesaurus were developed and how they can be used in NLP systems were determined. The analysis of homonymy showed how homonymous units should be classified in electronic systems. According to the results of the study, it was determined that each homonymous unit should be classified as an independent synset and indicators of their use in different contexts should be added. In general, the results of this study are of great importance in the process of creating an electronic thesaurus in the Uzbek language, which will create new opportunities in the field of natural language processing (NLP). In the future, in order to further develop the electronic thesaurus of the Uzbek language, it is necessary to create a large-scale corpus of texts, integrate the thesaurus into machine learning algorithms and study their application in various fields in more depth. Thus, the results of this study are of great practical importance in the fields of linguistics, computational linguistics and natural language processing, and will serve to adapt the Uzbek language to the processes of global technological development.

References:

1. Lyons, J. (1977). *Semantics* (Vol. 1 & 2). Cambridge University Press.
2. Shchur, G. S., & Potier, B. (1983). *Lexical Relations in Linguistics: A Comparative Study*. Moscow: Nauka.
3. Nurmonov, A. (2008). *O‘zbek tilining leksik-semantik xususiyatlari*. Toshkent: Fan.

4. Rasulov, R. (2010). *Leksikologiya va semasiologiya masalalari*. Toshkent: O‘zbekiston Milliy Ensiklopediyasi.
5. Kreydlin, G. E., & Shmelev, D. N. (1996). *Semantics and Pragmatics of Homonyms in Natural Language Processing*. Russian Academy of Sciences.
6. Berdialiyev, A., & Safarova, R. (2015). *O‘zbek tilida omonimiya va uning lingvistik o‘rni*. Toshkent: Universitet nashriyoti.
7. Miller, G. A., & Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
8. Princeton University. (2020). *WordNet: A Lexical Database for English*. Retrieved from https://wordnet.princeton.edu
9. Ermatov, I. R. (2012). *Morfem va so‘z yasash satҳlari terminologik tizimi*. Toshkent: O‘zbekiston Milliy Universiteti Nashriyoti.
10. Kuzmenko, N. V. (2004). *Meronymy in Modern English in the Light of Grammaticalization Theory*. Moscow State University.
11. Mustafayeva, S. (2017). *Mero-nomik munosabatlar o‘zbek va ingliz tillari misolida*. Toshkent: Fan va texnologiya nashriyoti.
12. Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.
13. Oxford University Press. (2021). *Oxford Thesaurus of English* (4th ed.). Oxford: Oxford University Press.
14. Uzbek National Corpus (2022). *ARANEUM_uzbekium: Annotated Corpus of the Uzbek Language*. Retrieved from https://araneum.uz
15. Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.