

LEXICOGRAPHIC FOUNDATIONS OF CREATING AN UZBEK LANGUAGE ELECTRONIC THESAURUS

Xolyorova Gulruksor G'ayratovna

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti tadqiqotchisi

Abstract: This article is dedicated to the lexicographic study of creating an Uzbek language electronic thesaurus, in which the importance of various fields of linguistics, including semantics, ontology, and terminology, is discussed. The article examines the analysis of linguistic thesauri's linguistic models using world experience, and provides a scientific approach to the issues of word classes, semantic classification, and structure. The effectiveness, functional possibilities, and theoretical conclusions of different types of thesauri are analyzed comparatively. Additionally, suggestions are developed for applying and utilizing ontological and terminological principles in creating the Uzbek language electronic thesaurus. The article discusses conceptual approaches aimed at forming a national linguistic model and methods based on international experience.

Keywords: electronic thesaurus, ideographic dictionary, linguistic ontology, terminological thesaurus, linguistic model, semantic field, lexicography, semantics, systemic relationships, corpus linguistics.

Introduction

Many theoretical schools related to corpora are being created in world linguistics. The idea is based on the fact that the corpus can serve as an object of study for all philological research. The analysis of theoretical ideas regarding corpora reveals that this field is significant both as a methodological approach and as a method, as well as an independent science for various philological studies. After independence, many positive changes took place in almost all areas. In particular, a number of government decisions related to the digitization of social fields were adopted. In this regard, positive results have been achieved in fields related to computational linguistics such as machine translation, sentiment analysis, question-answer systems, computational lexicography, computational lexicography, ontology, and semantic technologies. Nowadays, the rapid development of artificial intelligence technologies demands further improvement of computational lexicography in language technologies, emphasizing the relevance of this work. Computational lexicography continues to evolve due to the results achieved through corpus technology. The analysis of the scientific and theoretical foundations of lexicographically corpus-based dictionaries and the productive use of international experiences in corpus-based lexicographic research in Uzbek lexicography demonstrate the relevance of this work.

Creating an electronic thesaurus of the Uzbek language is recognized as one of the most urgent and promising areas of modern linguistic research. The preservation, analysis, and creation of new linguistic resources must become processes reinforced by advanced technologies that meet the demands of the time. An electronic thesaurus will play an important role in systematically reflecting the semantic fields of a language, clearly indicating the interrelationships between words, and facilitating the formation of new terms and concepts. In creating such an electronic resource for the Uzbek language, the harmonious integration of lexicography, semantics, and linguistic approaches is of great importance. This resource is a necessary tool for thoroughly studying the semantic structure of the language and its relationships at various levels, ensuring

the proper usage of the language in speech. The process of creating an electronic thesaurus opens up wide opportunities not only in the lexical structure of the language but also in organizing its terminological systems. Therefore, this article is aimed at studying the theoretical foundations of the Uzbek language electronic thesaurus, its lexicographic principles, and methodology. The article discusses the semantic, ontological, and terminological aspects of the thesaurus, as well as the scientific approaches needed to create its electronic model while considering the specific features of our national language. This work is of significant importance in further developing Uzbek lexicography and taking its scientific research to a new stage.

Materials and Methods

In world linguistics, various theoretical approaches have been created regarding corpora and corpus-based lexicography. Igor Igorevich Saien's works discuss corpus-based lexicography issues in creating corpus dictionaries, while Lynne Bowker emphasizes the role of corpus-based methods in designing terminological systems. D. Prinsloo highlights the peculiarities of corpus-based lexicography for resource-scarce languages. Additionally, Olga Semina's research on the advantages of using national corpora in lexical-semantic analysis is noteworthy. Wolfgang Teubert enriched theoretical views on the possibilities of using corpora in forming multilingual lexicography. The corpus-based approach in thesaurus creation has been analyzed in numerous works. A. Galieva and others presented ideas on using corpora in creating semantic fields in the Tatar language. International and national research in the field of creating the Uzbek language electronic thesaurus is of great significance. In Uzbek linguistics, scholars such as N. Abdurakhmanova, Sh. Hamroyeva, and G. Toirova have made significant contributions to the development of corpus-linguistic research and lexicographic foundations. M. Abjalova conducted research on the creation of ontologies, yet the modeling of the Uzbek language thesaurus has not been studied in depth. This article proposes new methodological approaches for creating the electronic thesaurus of the Uzbek language, taking into account its lexical and semantic characteristics. M. Abjalova carried out research on creating ontologies in Uzbek linguistics, but the modeling of the Uzbek language thesaurus has not been studied in a monographic context. Furthermore, in N. Sabirova's work, theoretical reflections on the dominance of pragmatic components at the periphery, the influence of neighboring paradigms in a particular field, the belonging of some words to multiple lexical-semantic fields, and the expression of synonymy, antonymy, and hyponymy in context have been studied monographically.

However, in world linguistics, research on creating thesauri is sufficiently developed, and various approaches have been formed based on the results obtained. S. Juan [1], in his research, notes that a thesaurus, dictionary, and encyclopedia are distinct from one another, and they can be considered as three separate methods of structuring and organizing observed phenomena. These studies, by combining international and national experience, help to identify the theoretical and practical aspects of creating an Uzbek language electronic thesaurus.

Methodologically, rational-typological and comparative-contrastive methods were employed in this article. The rational-typological method was used to compare thesauri and their structures in different languages, while the comparative-contrastive method analyzed the differences and similarities between the lexical units and semantic fields of the Uzbek language. Through these methods, theoretical foundations and practical approaches necessary for thesaurus creation were developed.

Results and discussion



The main results of the article are as follows:

1. The electronic thesaurus of the Uzbek language systematically reflects its semantic and lexical structures. The electronic thesaurus expresses hierarchical, equivalent, and associative relationships among lexical units.
2. The electronic thesaurus can be used not only as a terminological dictionary but also as an ideographic dictionary. This, in turn, allows for a better understanding of the conceptual structure of the language during the study of the language.
3. In the process of creating the thesaurus, it is crucial to consider the unique semantic fields of the Uzbek language and develop approaches that correspond to them[2].

One type of thesaurus is the terminological thesaurus. It is important to define its classificatory characteristics and differentiate related signs when forming the terminological thesaurus of the Uzbek language and especially in designing its model. The concept of a terminological thesaurus, proposed by the International Organization for Standardization, is widely used in special languages and terminology-related works in the following decades. According to sources[3], the term "terminological thesaurus" is considered preferable and more complete than existing terms such as "classified glossary," "classified dictionary," "systematic dictionary," and "systematic vocabulary." This proposal is based on the argument that the difference between these four terms should not be resolved in any way and that the distinction between "dictionary" and "lexicon" should be clarified. The terms "systematic" or "classified" should not be interpreted beyond their narrow application, and a descriptive characteristic that can be formally identified with similar concepts is more appropriate. Therefore, a special language or terminological thesaurus is not just a dictionary, glossary, or lexicon, but is expressed in dictionary form aimed at systematically organizing the knowledge structure of a specific subject or subjects. A terminological thesaurus must be distinguished from general language thesauri, such as Roget's Thesaurus, in terms of its scope and application. A thesaurus is recognized as a tool in lexicography that systematically presents numerous synonyms and antonyms. However, its interpretation as an ideographic dictionary can have a much broader and deeper theoretical context. The term ideography refers to a system based on representing thoughts or meanings.

From this perspective, interpreting the thesaurus as an ideographic dictionary means analyzing not just a list of synonyms and antonyms of words, but the linguistic and semantic expressions of a particular idea or concept. The main function of a thesaurus is to show the semantic relationships between various words. To achieve this, it groups words based on shared meaning or topic. An ideographic dictionary, on the other hand, specifically illustrates these meanings - that is, it sheds light on the meanings of words from different perspectives. When considering a thesaurus as an ideographic dictionary, it is not only the formal relationships between words that matter, but also the differences in their actual meanings and content. This is especially important for understanding the meanings of words in different languages and their contextual relationships in greater depth. Understanding a thesaurus as an ideographic dictionary allows us to distinguish its two main aspects - semantics and context. Semantics, or the meaning of words, is at the heart of the thesaurus. However, the ideographic interpretation goes beyond the surface meanings of words and considers their contextual and cultural differences. Therefore, using a thesaurus as an ideographic dictionary reflects not only the static structures of semantic patterns but also their dynamic changes.

The significance of the thesaurus as an ideographic dictionary lies in its potential to create new, advanced approaches in linguistics. Here, words and expressions are not only shown in their



syntactic form but also take into account their scope, possible meanings, and changes within a cultural context. In this way, the thesaurus can be applied not only in language learning but also in fields such as philosophy, psychology, and cultural studies. Interpreting the thesaurus as an ideographic dictionary enriches its linguistic capabilities, allowing for deeper study of the meanings of words, beyond their formal interpretations. This helps to better understand the subtle semantic connections between words and ideas. Creating a terminological thesaurus, especially identifying its classificatory characteristics and distinguishing between signs, plays a crucial role in the lexicographical development of the Uzbek language. According to the recommendations of the International Organization for Standardization, the terminological thesaurus is widely used in special languages and terminology-related works. Such a thesaurus not only serves as a glossary or dictionary but also expresses a set of lexical units aimed at systematically organizing the structure of knowledge in a particular field of study. Scholars hold various views on thesauri. For example, S.Osokina[4] develops the theoretical foundations and methodological principles of thesaurus creation. She emphasizes that a thesaurus, considered as a system, is capable not only of processing information and reception but also of generating new knowledge. In A.Proshina's research, ideographic terminological dictionaries - thesauri - are viewed as tools not only for describing but also for systematizing and organizing terminological systems. The scholar demonstrates the integration of anthropological linguistics, terminology, and ideography based on the Russian and German languages. This approach allows for the creation of an ideographic representation of knowledge and enables precise modeling of field-specific terms and the structure of knowledge.

It is worth noting that all definitions of a thesaurus emphasize its systemic nature. This can be understood as follows: every phenomenon or concept exists in a certain consciousness and mind. In the communication process, the systematization helps individuals to quickly and accurately structure texts. Thus, S. Osokina[5] emphasizes that the thesaurus, as a knowledge system, consists of elements and their relationships, based on the overall essence of the system. The elements hold semiotic significance, and their relationships at both material and semantic levels are supported. She acknowledges that studying the connection between elements as part of an existing system structure is a valid approach. Moreover, in her research on the thesaurus, she shows that all the general features of the open system, which is its structure, relate to how it interacts with the external environment through the storage of precise structures in the memory system.

When creating terminological thesauri, a structure that reflects the hierarchical and logical relationships of field-specific terminology is necessary to systematize the field's terminology. M.Kovyazina[6], when creating a thesaurus of ecology terms, states that "the terminological thesaurus reflects the logical and conceptual relationships among terminological units and expresses the conceptual structure of the field". This definition presents essential approaches in improving the scientific and practical structure of the field when creating a terminological thesaurus. These findings serve as the fundamental methodological foundation for forming an electronic thesaurus of the Uzbek language. Moreover, these studies contribute to providing an accurate and effective systematized form of field-specific terminology.

Numerous studies have been conducted on the thesaurus worldwide. Today, a thesaurus is considered a structured lexicon of concepts unified by terms, especially with the development of computer technologies and artificial intelligence. The model of a lexicon with a specific structure of lexical units and the semantic space of the language being studied allows for a clear representation of systemic relationships. This model also allows the observation of the



hierarchical organization of the lexicon and its organization at different levels. The thesaurus method, as a way of modeling the lexical-semantic structure of a subject area, provides an opportunity to study semantic fields. Moreover, the author refines the concept of a thesaurus, not as merely a collection of lexicons (a general definition commonly found), but as an effective tool for lexical-semantic modeling that can be applied to almost any subject area.

The creation of electronic thesauruses and linguistic modeling, in line with the technological progress of linguistics, is advancing and demanding new methodologies. This process not only involves analyzing the semantic and syntactic features of a language but also uses computers to model language and create quick and efficient information retrieval in different languages. Below, we will review the primary approaches and methodologies used in the creation of electronic thesauruses and linguistic modeling. Vedernikova[7] presents the modeling of cognitive-linguistic terminology in English, analyzing the creation of thesauruses based on Blending theory. She explores the semantic and cognitive aspects of thesauruses, the interrelations of words, and their organization within semantic fields. Vedernikova emphasizes that "thesauruses play a crucial role in organizing semantic structures within cognitive-linguistic processes." Additionally, Gorbunov[8], in his research, investigates the role of thesauruses in modeling the semantic field of the term *Lexicology* using a semiotic-synergetic approach. He evaluates thesauruses not only as a tool for classifying information but as a significant means for structuring knowledge systems. "A thesaurus helps not only in creating semantic structures but also in generating new information"[9]. These works provide essential methodological foundations for creating an electronic thesaurus of the Uzbek language.

N.Abdurakhmonova and Abdulhay Sobirov, in their work "Conceptual Importance of Creating a Thesaurus Using a Corpus," analyze the conceptual aspects of the thesaurus creation process using the corpus method[10]. They discuss the role of the corpus in identifying semantic and syntactic relations between words and how it helps clarify the natural use and relationships of words. Abdurakhmonova and Sobirov emphasize the significance of the corpus methodology in thesaurus creation, stating that "a thesaurus created using a corpus more accurately and reliably reflects the natural usage of words and their relationships." This approach allows for a deeper analysis of not only terminology systematization but also the semantic and pragmatic contexts of words. This methodology is also crucial in the creation of an electronic thesaurus for the Uzbek language because it requires taking into account not only linguistic but also cultural and pragmatic contexts. Thus, the works of Abdurakhmonova and Sobirov highlight the importance of conceptual approaches in creating the Uzbek thesaurus.

Conclusion

The lexicographic study of modeling the electronic thesaurus of the Uzbek language holds an important position not only in linguistics but also in the fields of terminology and semantics. The process of creating an electronic thesaurus requires a deep study of the language's lexical structure, which, in turn, involves systematic analysis of the meanings and functions of words. A thesaurus serves as a convenient tool for identifying synonyms, antonyms, and various semantic relations between words, playing a significant role in automatic language analysis, lexical system organization, and the development of specialized terminologies. In the modeling of the Uzbek electronic thesaurus, it is crucial to identify the semantic aspects of lexical units, such as their meaning structure, contextual relations, and interrelationships. Through semantic analysis, the accurate classification and grouping of words can be achieved. Each word, based on its meaning and function, interacts with other words through mechanisms such as synonyms, antonyms, or hypernyms (higher-level categories). In this way, the semantic relationships



between lexical units are analyzed in-depth, leading to the formation of complex semantic networks. During the thesaurus creation process, the polysemy of each word and its use in various contexts are taken into account. This is related to the dynamic and multi-layered nature of lexical units in the Uzbek language, and each meaning and its interrelationships are identified through distinct models. In this process, the different meanings, layers, and contextual characteristics of words are systematically grouped, ensuring precise and reliable classification. When creating the Uzbek electronic thesaurus, it is essential to apply conceptual approaches that consider the lexical features of the Uzbek language and identify the semantic perspectives of each word. This allows for the inclusion of the language's unique social, cultural, and historical contexts. Through conceptual approaches, the meaning of each word in the thesaurus is classified not only grammatically but also in terms of its societal usage and contextual content.

REFERENCES

1. Sager, Juan C. "Terminological Thesaurus. A more appropriate designation or a deprecated synonym?" *Lebende Sprachen*, vol. 27, no. 1, 1982, pp. 6-7. <https://doi.org/10.1515/les.1982.27.1.6>
2. Gorbunova, N. N. Nekotory'e rezul'taty' tezaurusnogo modelirovaniya angloyazy'chnoj terminosistemy' menedzhmenta (Some results of thesaurus modeling of English management terminological system) // *Uchenyye zapiski natsional'nogo obshchestva prikladnoy lingvistiki*. – 2014. – P. 7-17.
3. Proshina A. Modeling of a Bilingual Dictionary-Thesaurus on Economics: Abstract of Ph.D. in Philology, Yekaterinburg, 2008. – 24 p.
4. Osokina S. Foundations of the Linguistic Theory of Thesaurus: Abstract of Ph.D. in Philology, Chelyabinsk, 2015. – P. 25.
5. Osokina S. Foundations of the Linguistic Theory of Thesaurus: Abstract of Ph.D. in Philology, Chelyabinsk, 2015. – P. 25.
6. Kovyazina M. Functional Model of a Bilingual Ecological Dictionary-Thesaurus: Abstract of Ph.D. in Philology, Tyumen, 2006. – 24 p.
7. Vedernikova, Yu. V. "The Thesaurus Approach to Modeling English Cognitive-Linguistic Terminology on the Example of the Blending Theory Term Field" *Philological Sciences. Theory and Practice Issues* - 2015. - No 12 (54). - Pp. 47-50.
8. Gorbunova, N. N. "Some Results of Thesaurus Modeling of the English Management Terminological System" *Uchenyye zapiski natsional'nogo obshchestva prikladnoy lingvistiki*. – 2014. – P. 7-17.
9. International Organization for Standardization 2011, 2013
10. Abdurakhmonova N. and Abdulhay Sobirov. "Conceptual Peculiarities of Creating Thesaurus by Corpus". *Proceedings of the International Conference on Translation, Information, Communication: Political and Social Bridge*, 2019, pp. 36–39.