International Journal Of Literature And Languages

VOLUME Vol.05 Issue11 2025
PAGE NO. 108-112
DOI 10.37547/ijll/Volume05Issue11-25

Designing A Lexical Database Of Adjective Thesauri For The Uzbek Wordnet

Pulatova Gulkhayyo Gofurjanovna
Independent Researcher, Namangan State University, Uzbekistan

Received: 24 September 2025; **Accepted:** 16 October 2025; **Published:** 21 November 2025

Abstract: This article discusses the principles of designing a lexical database of adjective thesauri within the framework of the Uzbek WordNet. The research analyzes the semantic structure of adjectives, explores synonymic, antonymic, gradaution and examines methods of integrating these into a digital lexical database. The study contributes to the development of digital linguistic resources and enhances the use of the Uzbek language in artificial intelligence systems.

Keywords: WordNet, lexical database, adjective thesaurus, semantics, antonymy, synonymy, linguistic analysis, computational linguistics.

Introduction: In recent years, the development of digital lexical resources has become a key direction in linguistics. The WordNet model, first created for English, represents a network of semantic relationships among words. Designing a similar structure for the Uzbek language, particularly for adjectives, requires complex linguistic and computational analysis due to the semantic richness and contextual variability of adjectives. The aim of this study is to outline the design principles of a lexical base for adjective thesauri in the Uzbek WordNet. In contemporary lexicographic research on the Uzbek language, the phenomenon of synonymy, particularly within the category of adjectives, holds significant importance for the development of a thesaurus model. The manifestation of adjectives in synonymic series allows for the determination of their degree of semantic closeness and the selection of contextually appropriate variants. Identifying stylistic distinctions between words and grouping them into synsets within a thesaurus system facilitates the creation of a complex lexical-semantic network. Therefore, examining the linguistic nature of synonymous adjectives, as well as their paradigmatic and syntagmatic relationships, constitutes a necessary foundation for the structure of a thesaurus.

Review of literature

Previous studies show that systems such as Princeton WordNet, EuroWordNet, and BabelNet provide solid foundations for building national language WordNets. In the Uzbek context, existing projects like UzWordNet mainly focus on nouns and verbs, while the adjective domain remains underdeveloped. Research by Miller (1995), Fellbaum (1998), and Kilgarriff (2000) outlines universal principles for WordNet development, which can be adapted to the Uzbek linguistic system. In Uzbek linguistics, the systematic study of synonyms and their compilation in dictionary form was first established academically by A.Hojiyev. He developed principles for identifying the semantic properties of synonymous units, determining criteria for distinguishing them in context, and grouping them accordingly. These scholarly foundations have since been refined using contemporary computational linguistics approaches, enabling the compilation of synonyms in electronic form and the creation of automated thesauri. In particular, H.Dadaboyev, M.Abjalova and U.Rashidova compiled a learner's dictionary of Uzbek synonyms, while explanatory dictionaries of Uzbek synonyms were produced under the editorship of N.Mahmudov. M.Abjalova and M.Najmiddinov also expressed their perspectives on creating a lexical database for the category of adjectives in the Uzbek thesaurus, highlighting the synonymic and quasi-synonymic relations within this word class.

METHOD

The methodology for designing the adjective thesaurus includes:

1. Automatic extraction of adjectives from corpus-based data.

2. Semantic mapping of synonymic, antonymic, and gradational relations.

3. Manual curation to verify automatically generated results.

4. Data formatting in JSON/XML to align with WordNet structure.

5. Implementation of Python NLTK tools to automate processing.

Using these methods, semantic networks were built for over 2,500 adjective lexemes, forming the foundation of the adjective module in the Uzbek WordNet. In linguistics, words that differ in form but share the same or slightly different meanings are defined as synonyms. Studying adjectives from a synonymic perspective in a thesaurus involves not only "word similarity" but also a careful consideration of their semantic, stylistic, functional and contextual properties. Indeed, a thesaurus is a linguistic repository that should encompass all features of a language. When organizing adjective synsets in a thesaurus, systematicity, semantic consistency and lexicographic precision play a crucial role. The following seven principles serve as the main methodological foundation for a thesaurus model:

First, the semantic core (primary meaning) must be identified, as each synset (group of synonyms) is structured around a single semantic nucleus. For example: (big), (huge), (large) – (semantic core: size magnitude).

Second, contextual interchangeability – it is necessary to consider whether synonyms can be used in place of one another and to analyze this through texts. For example: (beautiful girl) ↔ (pretty girl) – acceptable; but (beautiful situation) ↔ (pretty situation) – stylistically inappropriate.

Third, the compatibility of stylistic properties should be verified, determining their frequency of use in formal, colloquial, literary, and poetic registers. For example: (big, neutral) ↔ (gigantic, expressive) ↔ (grand, in spoken dialect).

Fourth, the phenomenon of gradation among synonyms should be explained. For example: (small) ← (medium) ← (big) ← (huge) ← (gigantic).

Fifth, balance along the antonymic axis – each synset should be linked to an opposing synset. For example: (beautiful) – (ugly), (beautiful) – (ugly, pejorative).

Sixth, semantic shifts related to metaphor and metonymy must be taken into account. For example: (hard, strict) – (strict; e.g., acting firmly in a certain task).

Seventh, the syntactic role and collocation of adjectives. Some synonyms combine only with specific nouns: (green light – correct), (deep blue light – emotional/emphatic), (sky-blue light – stylistically restricted).

These principles demonstrate the necessity of considering stylistic richness and emotional-expressive features when representing adjective synonyms in a thesaurus. In particular, when forming a synonymic thesaurus for adjectives that convey specific semantic and functional loads – such as those denoting color, character and traits, smell, space and time, or state – these approaches serve as an important methodological foundation. Indeed, adjectives within these groups constitute a key layer that determines the expressive power of language due to their connotative potential, metaphorical extensions, and collocational properties. When examining the representation of adjectives in WordNet-type dictionaries, it is evident that adjective lexemes are systematically organized in a thesaurus according to synonymy, antonymy, and collocation. In developing the semantic fields of the adjective category for the Uzbek thesaurus and creating its thesaurus, we studied the principles underlying WordNet and other WordNet-like dictionaries. Based on this research and taking into account the rich potential of the Uzbek language, we attempted to demonstrate the practical inclusion of collocations, connotations, idiomatic expressions, and textual units among adjectives in the system. Furthermore, this approach is crucial for creating adjective thesauri in Uzbek, as it allows for a more precise representation of semantic nuances and for organizing them into a semantically coherent network. Adjectives were extracted from the explanatory dictionary of the Uzbek language and organized into a linguistic database. Out of a total of 12,998 adjective lexemes in Uzbek, approximately 1,000 were analyzed. These adjectives were categorized according to their semantic types (Table 3.1.1). The results show that, in Uzbek, adjectives denoting character and traits are significantly more numerous compared to other semantic groups (Diagram 3.1.1).

This indicator evolves in parallel with the development of the language, as the semantics of adjectives actively participates in processes of meaning expansion and narrowing.

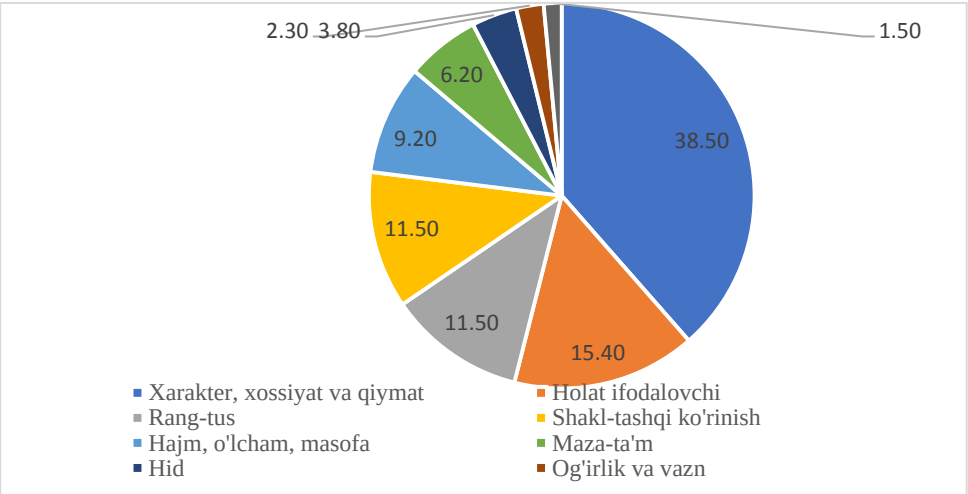

Diagram 1.1.1. Semantic Categories of Adjective Lexemes in the Uzbek Language

Modern computational linguistics offers essential tools and methods for analyzing adjectives and organizing them into a thesaurus. Large-scale text corpora – such as the Educational Corpus of the Uzbek Language and ARANEUM UZBEKIUM – together with advanced algorithms like word embeddings, which identify semantic similarity between words, play a crucial role in this process. Word embeddings are semantic vector models based on the frequency of word usage in context. For instance, while the lexeme “yashil” (green) historically denoted the color of nature and plants, today this adjective forms collocations with various nouns and can express ideological, economic, and ecological meanings in contemporary contexts, such as green economy, green space, and green energy. Word embedding technology allows us to identify these contextual shifts and, although words like ecological, natural, clean, biological are semantically related to “yashil”, they do not form direct synonym sets. Instead, embeddings help group such lexemes into clusters. Reflecting these properties in a thesaurus gives rise to the notion of quasi-synonymy. A quasi-synonym is a relatively new concept compared to traditional synonyms: it refers to words that, while not appearing in the same synonymic set, can merge semantically within a particular domain. Representing such groups of synonyms in a thesaurus provides a broader understanding of word semantics and is highly important for AI, NLP, and linguistic modeling.

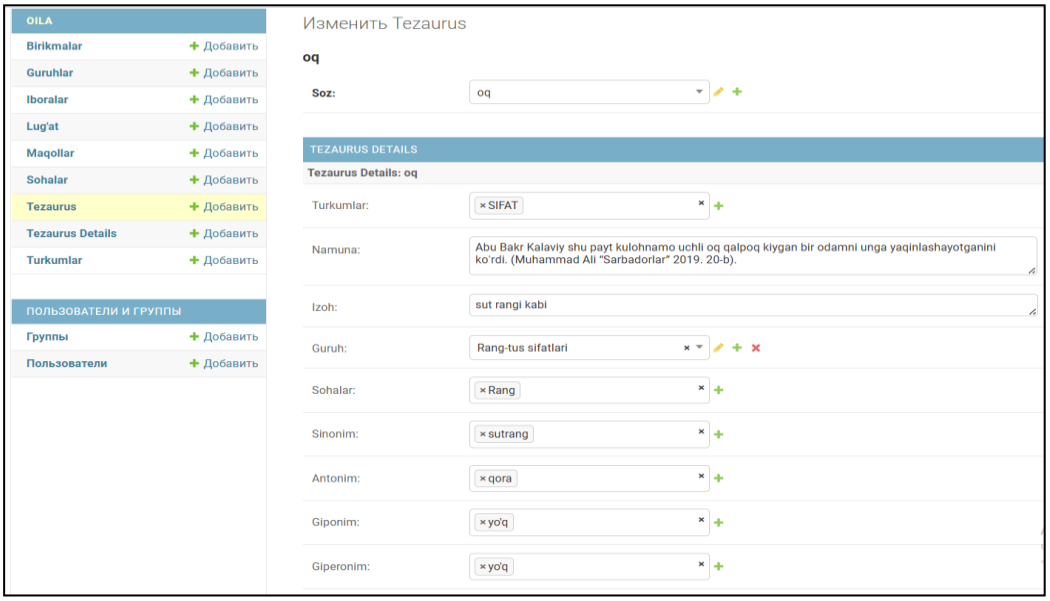

Figure 1.1.2. Display of Adjective Lexemes on the Uzcorpus Website

Based on this approach, the semantic field of the adjective “yashil” can be modeled, presenting its representation in a thesaurus in a form harmonized with modern linguistic technologies.

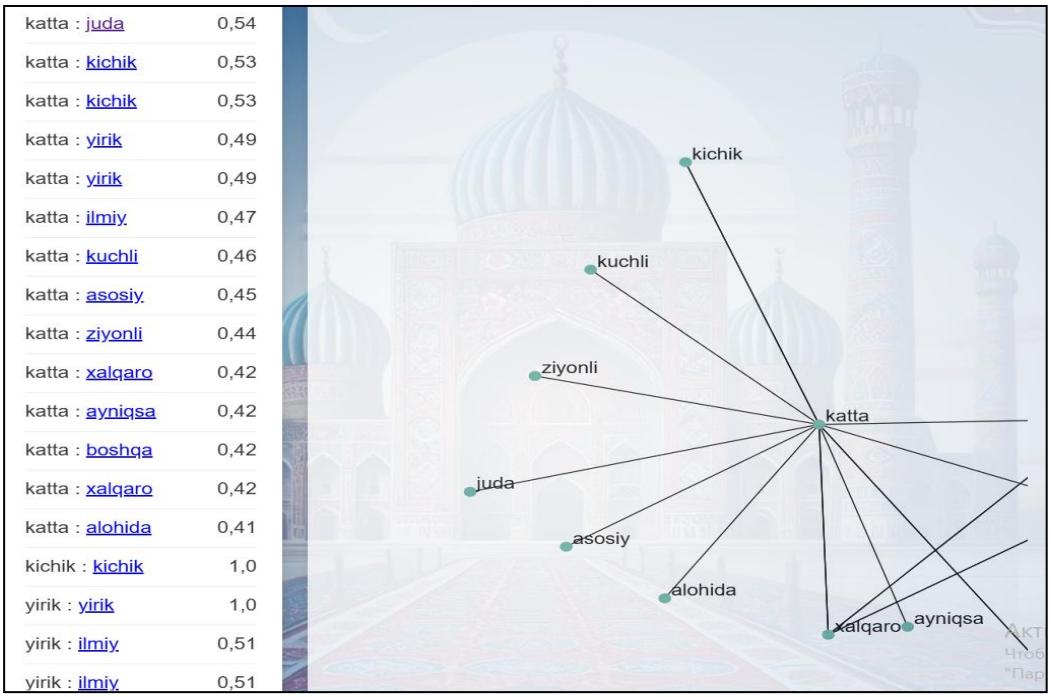

Figure 1.1.3. Display of Adjective Lexemes on the Paratranslator Website

Adjective lexemes denoting character, traits, and properties form a separate semantic field in a thesaurus. In the traditional lexical-semantic system, such adjectives primarily described intrinsic features of objects or phenomena, such as hardness, softness, or dryness. Today, however, lexemes like “shaffof” (transparent), “yuvosh” (gentle), “tiniq” (clear) and “ochiq” (open) have expanded their classical meanings, creating new semantic domains within the thesaurus. For example, shaffof originally denoted the “ability to transmit light” but in social-communicative contexts it conveys concepts such as “transparency” or “transparent governance” (Figure 2.1.1).

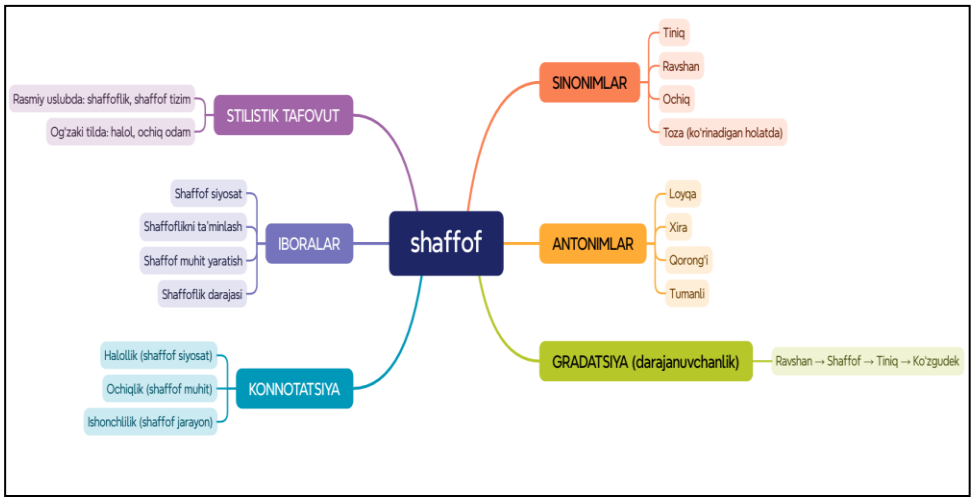

Figure 2.1.1. Thesaurus Structure of the Lexeme *“Shaffof”* (Transparent)

The adjective quruq (dry) is semantically polysemous, as it carries different meanings depending on the context. Its primary sense is the absence of moisture, i.e., the opposite of wet. It forms a synonymic set with qurg‘oq (arid), qoq (very dry), qovjiroq (drier, used for plants or body parts), and sur (used for meat or fish from which water has been removed). In the word qoq, the degree of dryness is stronger than in qurg‘oq. Qurg‘oq is actively used in dialects, particularly in reference to extremely hot summers or deserts. Qovjiroq mainly applies to plants or human body parts, while sur refers to dried meat or fish. Thus, while the classical meaning of quruq emphasizes waterlessness or the absence of moisture (e.g., quruq yer – dry land, quruq havo – dry air), the lexeme now creates new semantic fields, which is a key factor in its precise representation within a thesaurus.

CONCLUSION

The creation of an adjective thesaurus for the Uzbek WordNet is a vital step toward the digitalization of the Uzbek language. This resource can serve artificial intelligence, natural language processing, and translation technologies. Future work includes expanding the database, integrating machine-learning-based clustering algorithms, and ensuring full coverage of the Uzbek lexical system.

In the above analysis of adjective lexemes according to types such as color, smell, state, and character-trait or property, we observed that each unit clearly expresses specific concepts within semantically distinct collocational and connotational structures. The results indicate that each lexeme in the Uzbek language actively participates in semantic expansion and in generating new concepts through combination with various noun classes. This characteristic is particularly pronounced in adjective lexemes, making it a crucial factor in the creation of adjective synsets for the Uzbek thesaurus.

REFERENCES

1. Hojiyev, A. (1975). [Synonyms and their Systematic Study]. Tashkent: Fan.
2. Dadaboyev, H., Abjalova, M., & Rashidova, U. (2002). O‘zbek tili sinonimlarining o‘quv lug‘ati [Learner’s Dictionary of Uzbek Synonyms]. Tashkent: Sharq.
3. Mahmudov, N. (Ed.). (2010). [Explanatory Dictionary of Uzbek Synonyms]. Tashkent: O‘zbekiston Milliy Ensiklopediyasi.
4. Abjalova, M., & Najmiddinov, M. (2018). [Creating a Lexical Database of Adjectives for the Uzbek Thesaurus]. Tashkent: Fan va Texnologiya.
5. Hojiyev, A. (1980). [Lexical-Semantic System of the Uzbek Language]. Tashkent: Fan.
6. Abjalova, M. (2015). [Synonymic Analysis of Adjectives in Uzbek]. Tashkent: 2015
7. Najmiddinov, M. (2017). O‘zbek tili leksikasi va tezaurus tuzilishi [Lexicon and Thesaurus Structure of the Uzbek Language]. Tashkent: Fan va Ta’lim.