VOLUME-3, ISSUE-1

SEMANTIC ANALYSIS OF ADJECTIVES IN UZBEK AND COMPUTATIONAL THESAURUS APPROACH

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

Abstract: This thesis examines the semantic analysis of adjectives in the Uzbek language and their representation in a computational thesaurus. Adjectives, essential for expressing qualities and states, are analyzed through tools like word embeddings and corpus-based methods. The study highlights the semantic clustering of synonyms and the disambiguation of polysemous adjectives, ensuring their accurate organization in a thesaurus. Challenges such as contextual ambiguity and incomplete resources are addressed using computational techniques complemented by expert linguistic input. The research underscores the value of visualization and computational tools in creating intuitive and effective linguistic resources. This approach enhances applications in natural language processing, education, and linguistic research while preserving the richness of the Uzbek lexicon.
Keywords: Uzbek adjectives, semantic analysis, computational thesaurus, word embeddings, polysemy, synonyms, NLP, lexical relationships, linguistic research..

Adjectives in the Uzbek language are integral to expressing qualities, states, and conditions, forming an essential component of its grammar and lexicon. These descriptive words serve as critical tools for communication, providing nuanced meanings and enriching the language. However, the analysis and organization of adjectives, especially in a computational thesaurus, present challenges due to their semantic complexity. This **thesis** delves into the semantic analysis of Uzbek adjectives and examines a computational approach to constructing an effective thesaurus, leveraging modern linguistic technologies. Adjectives, as a part of speech, can be categorized based on their semantic and grammatical features. They often exhibit relationships such as synonymy, antonymy, and polysemy. For example, the adjective *ingichka* (thin) can refer to physical thinness, sharpness of sound, or a figurative subtlety, depending on its context. Such semantic variations complicate the task of categorizing adjectives, requiring precise analysis to define their contextual meanings.

Semantic analysis provides insights into how adjectives function within the Uzbek language. By mapping their meanings and relationships, we can identify patterns that are critical for tasks such as machine translation, information retrieval, and language teaching. The analysis also facilitates the development of linguistic resources like thesauruses, which are essential for preserving and enhancing the understanding of the Uzbek language. Modern computational linguistics offers tools and techniques to analyze adjectives and structure them into a thesaurus. The use of large-scale text corpora, such as the Uzbek Language Educational Corpus and ARANEUM UZBEKIUM, combined with advanced algorithms like word embeddings, enables the automatic extraction of semantic relationships. Word embeddings are vector representations of words that capture semantic similarities based on their co-occurrence in large text datasets. For instance, the adjective *mazali* (tasty) shows a high degree of semantic similarity with words like *shirin* (sweet) and *qaymoqli* (creamy). By applying clustering techniques, these relationships can be visualized, providing a clear representation of their semantic fields.

For example, a cluster for *mazali* might include related adjectives such as:

- *Shirin* (sweet)

- *Qaymoqli* (creamy)
- *Ta'mli* (flavorful)

These clusters serve as the foundation for thesaurus entries, grouping semantically related words under a single heading.

Synonymy and polysemy are central to the semantic analysis of adjectives. Synonyms like *ingichka* (thin) and *yupqa* (slender) can be grouped together in a thesaurus, while polysemous words are divided into separate entries based on their distinct meanings. For example:

- *Inchigka* (thin) as a physical quality: Synonyms include *ensiz* (narrow) and *yupqa* (slender).
- *Inchigka* (sharp) as a sound quality: Related to *chiyildoq* (piercing).
- *Inchigka* (subtle) as a figurative expression: Synonyms include *nozik* (delicate) and *latif* (graceful).

This method ensures that polysemous words are comprehensively represented without ambiguity, making the thesaurus more effective.

Challenges in Computational Thesaurus Construction

Developing a computational thesaurus involves several challenges, particularly in a language like Uzbek, where resources and computational tools are still evolving.

1. **Ambiguity in Context**: Polysemous adjectives require precise disambiguation to ensure correct semantic clustering.
2. **Incomplete Lexical Resources**: Existing dictionaries and corpora may lack exhaustive lists of synonyms and antonyms.
3. **Cultural and Contextual Nuances**: Adjectives often carry cultural connotations that need to be preserved in their representation.

To overcome these challenges, computational methods must be supplemented with expert linguistic analysis. For instance, combining automated word embeddings with manual verification can improve accuracy.

The Role of Visualization in Semantic Analysis

Visualization techniques play a crucial role in representing the semantic relationships of adjectives. Graph-based models can depict synonyms, antonyms, and other relationships, making it easier to understand complex networks of meaning. Tools like t-SNE and PCA can project high-dimensional word embeddings into two or three dimensions, creating intuitive visualizations for linguists and developers.

For example, a graph for *nozik* (delicate) might show:

- Connections to synonyms (*latif*, *nazokatli*).
- Contextual associations with words like *ipakdek* (like silk) or *sezgir* (sensitive).

These visualizations enhance the usability of the thesaurus, particularly in educational and computational applications.

Applications of a Computational Thesaurus

A well-constructed computational thesaurus has wide-ranging applications:

1. **Natural Language Processing (NLP)**: Enhances tasks like sentiment analysis, machine translation, and speech recognition by providing nuanced lexical data.
2. **Language Education**: Supports the teaching of Uzbek by offering detailed explanations of adjectives and their usage.
3. **Linguistic Research**: Facilitates studies in semantics, syntax, and lexical relationships within Uzbek and comparative linguistics.

For instance, an NLP application could use thesaurus data to identify sentiment-laden adjectives like *mazali* (tasty) or *achchiq* (bitter), improving the accuracy of sentiment classification.

Conclusion

The semantic analysis of adjectives in Uzbek and their organization in a computational thesaurus represent significant advancements in linguistic research and resource development. By leveraging tools like word embeddings, corpus analysis, and visualization techniques, it is possible to systematically map the semantic fields of adjectives and create a user-friendly thesaurus.

Despite the challenges, the integration of computational methods with expert linguistic input ensures the creation of a reliable and versatile resource. This approach not only preserves the richness of the Uzbek language but also opens new avenues for research, education, and technology-driven applications. As computational linguistics continues to evolve, so too will the opportunities for further enriching and expanding the Uzbek lexicon.

REFERENCES:

1. Бойков, В. Н., Захаров, В. Е., Каряева, М. С., & Соколов, В. А. (2013). Тезаурус по поэтологии как инструмент для информационного поиска и коллекции знаний. Моделирование и анализ информационных систем, 20(4), 125-135.
2. Семенов, М. Ю., Кичерова, М. Н., & Трифонова, И. С. (2024). Экосистема образования взрослых: конструирование терминологического поля и междисциплинарного тезауруса. Образование и наука, 26(3), 12-39.
3. Погорский, Э. К. (2024). Цифровой тезаурус: Роль языковых моделей в интерпретации индивидуальных картин мира. Знание. Понимание. Умение, (1), 243-251.
4. Лукашевич, Н. В., Лашевич, Г. Э., Герасимова, А. А., Иванов, В. В., & Добров, Б. В. (2016). Порождение тезауруса типа WordNet для русского языка. In Пятнадцатая национальная конференция по искусственному интеллекту с международным участием (pp. 89-97).
5. Лесников, С. В. (2011). Тезаурус как отражение системности языка. Вестник Челябинского государственного университета, (28), 52-61.
6. Баранов, А. Н., Добровольский, Д. О., Киселева, К. Л., Козеренко, А. Д., Вознесенская, М. М., & Коробова, М. М. (2007). Словарь-тезаурус современной русской идиоматики. Издательство" Аванта+".