THE LEXICAL AND IDEOGRAPHIC FIELDS OF ADJECTIVE WORDS IN UZBEK AND REPRESENTATION IN A THESAURUS

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

Abstract: This thesis explores the lexical and ideographic fields of adjectives in the Uzbek language and their systematic representation in a thesaurus. Adjectives, classified as monosemantic and polysemous, play a significant role in Uzbek linguistics. Monosemantic adjectives exhibit a single, straightforward meaning, while polysemous adjectives possess multiple meanings depending on context. Using computational tools like the Uzbek Language Educational Corpus and ARANEUM UZBEKIUM, this study identifies semantic relationships among adjectives through word embeddings and contextual analysis.
Keywords: Uzbek adjectives, thesaurus, semantic analysis, word embeddings, monosemantic, polysemous, computational linguistics, lexical fields.

Adjectives are integral to the Uzbek language, providing a vivid means of expressing qualities, characteristics, and descriptive features. These words span various categories, including single-meaning (monosemantic) and multiple-meaning (polysemous) adjectives, each with unique semantic and ideographic fields. The representation of such adjectives in a thesaurus not only aids linguistic research but also supports practical applications like computational linguistics, machine translation, and language teaching. This thesis delves into the lexical and ideographic fields of Uzbek adjectives and their systematic organization in a thesaurus, based on the provided data. In the Uzbek language, adjectives include words that denote static and dynamic qualities, with related forms such as participles and adverbs often complementing them. Adjectives express static qualities (e.g., *ingichka* – thin) and dynamic qualities associated with temporal changes (e.g., *yashil* – green as it relates to ripening). The analysis of these adjectives reveals their significant contribution to the richness of the Uzbek lexicon.

Uzbek adjectives can be broadly classified into monosemantic and polysemous types. Monosemantic adjectives have a single meaning, making their lexical and ideographic categorization more straightforward. For example, the adjective *inja* (delicate) explicitly conveys its meaning without ambiguity. In contrast, polysemous adjectives like *ingichka* (thin, sharp, subtle) exhibit a range of meanings depending on context,

requiring a more nuanced approach for representation in a thesaurus. The primary challenge in representing adjectives in a thesaurus lies in capturing their semantic relationships, including synonyms, antonyms, and contextual associations. The inclusion of computational tools, such as the Uzbek Language Educational Corpus and ARANEUM UZBEKIUM, facilitates the construction of accurate semantic fields through word embeddings and contextual analysis.

Monosemantic adjectives, due to their singular meaning, allow for a straightforward representation. For instance, the word *inja* can be enriched by identifying semantically related terms using corpus analysis. Tools like word embeddings provide lists of related terms, such as *badiiyat* (artistry), *rangin* (colorful), and *latif* (graceful). These words form a synset that reflects *inja*'s semantic neighborhood, as shown in visualizations and tables generated from corpus analysis.

A thesaurus entry for *inja* might include:

- **Synonyms**: *nazokatli* (elegant), *mayin* (gentle)
- **Related Words**: *rangin* (colorful), *latif* (graceful)
- **Contextual Associations**: Poetry, art, refinement

Representing polysemous adjectives in a thesaurus is more complex due to their varied meanings. For instance, *ingichka* can denote physical thinness, auditory sharpness, or figurative subtlety. Corpus analysis helps to group these meanings into distinct semantic clusters. For example:

1. **Physical Thinness**: Synonyms include *yupqa* (thin), *ensiz* (narrow).
2. **Auditory Sharpness**: Synonyms include *chiyildoq* (piercing), *qattiq* (loud).
3. **Figurative Subtlety**: Synonyms include *nozik* (delicate), *latif* (graceful).

By creating separate entries for each meaning, the thesaurus ensures clarity and usability for various linguistic and computational purposes.

The advent of computational linguistics has revolutionized thesaurus creation. Tools like word embeddings, corpus-based analysis, and machine learning algorithms enable precise semantic mapping of adjectives. In this study, the Uzbek Language Educational Corpus and ARANEUM UZBEKIUM corpus provided valuable data for identifying semantic relations. Word embeddings algorithms analyze large text corpora to find semantically similar words. For instance, the embeddings for *mazali* (tasty) reveal related words such as *qaymoqli* (creamy) and *shirin* (sweet). These relationships are visualized using tools like t-SNE or PCA, aiding in the creation of intuitive thesaurus entries.

Semantic relationships can be visually represented using graphs and diagrams. For example, a graph for *mazali* may display its connections to synonyms, antonyms, and

related concepts. This visual approach helps linguists and developers understand complex networks of meaning. Thesauruses designed with computational methods can be integrated into natural language processing (NLP) applications. For instance, machine translation systems can utilize thesaurus data to provide nuanced translations, while educational platforms can use it to teach vocabulary effectively.

Despite advances in computational tools, several challenges remain:

- **Ambiguity in Polysemous Words**: Determining the exact meaning in context requires advanced disambiguation techniques.
- **Incomplete Synonym Lists**: Existing resources may lack comprehensive synonym data, necessitating manual enrichment.
- **Cultural Nuances**: Adjectives like *mazali* carry cultural connotations that must be considered in thesaurus design.

Addressing these challenges involves combining computational approaches with expert linguistic input, ensuring that the thesaurus accurately reflects the complexities of the Uzbek language.

The representation of Uzbek adjectives in a thesaurus involves careful consideration of their lexical and ideographic characteristics. Monosemantic adjectives provide a simpler entry point, while polysemous adjectives require advanced computational and linguistic approaches. The integration of corpus analysis, word embeddings, and visualization tools enhances the thesaurus's accuracy and usability. By addressing challenges such as ambiguity and cultural nuances, a well-designed thesaurus can serve as a powerful resource for linguistic research, language education, and NLP applications. This study demonstrates the potential of computational methods in thesaurus construction, paving the way for further innovations in Uzbek linguistics.

REFERENCES:

1. 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.
2. Погорский, Э. К. (2024). Цифровой тезаурус: Роль языковых моделей в интерпретации индивидуальных картин мира. Знание. Понимание. Умение, (1), 243-251.
3. 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).

4. Лесников, С. В. (2011). Тезаурус как отражение системности языка. Вестник Челябинского государственного университета, (28), 52-61.
5. 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+".
6. Азарова, И. В., Митрофанова, О. А., & Синопальникова, А. А. (2003, June). Компьютерный тезаурус русского языка типа WordNet. In Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог (pp. 43-50).