REPRESENTATION OF POLYSEMY, SYNONYMY, AND HOMONYMY IN THE UZBEK ELECTRONIC THESAURUS

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

Abstract: This thesis explores the representation of polysemy, synonymy, and homonymy in the Uzbek electronic thesaurus, emphasizing their role in structuring lexical resources for natural language processing (NLP) and information retrieval. The study examines methods for distinguishing word meanings, organizing synsets, and integrating computational approaches such as AI and word embedding models. It highlights the importance of linguistic validation and hierarchical organization in thesaurus development. By improving semantic representation, the Uzbek electronic thesaurus enhances machine translation, search systems, and AI-driven text processing applications.
Keywords: Uzbek language, electronic thesaurus, polysemy, synonymy, homonymy, NLP, computational linguistics, information retrieval, AI

The structure of an electronic thesaurus is based on lexical-semantic relationships, particularly focusing on polysemy, synonymy, and homonymy in the Uzbek language. Electronic thesauri, like WordNet, organize words into synsets, which are collections of synonyms that share a common meaning. The Uzbek electronic thesaurus must account for various semantic relations, including hypernymy, hyponymy, and meronymy, to create a more comprehensive linguistic resource. Polysemy, the phenomenon where a single word has multiple meanings, is a crucial aspect of the Uzbek language. For example, the verb *boshla* can mean "to start an action," "to lead," or "to initiate," depending on context. Each meaning requires distinct synsets to ensure accurate lexical representation. Analyzing polysemy involves identifying context-specific meanings and organizing them hierarchically within the thesaurus. Synonymy, where multiple words share similar meanings, also plays a key role in thesaurus development. Uzbek synonyms such as *go‘zal* (beautiful), *ajoyib* (wonderful), and *betakror* (unique) require careful differentiation to capture subtle semantic nuances. The electronic thesaurus must structure these synonyms to reflect their degree of interchangeability while preserving semantic distinctions. Homonymy, where words share the same form but have different meanings, poses another challenge. For instance, the word *bog‘* in Uzbek can mean "garden" or "knot," depending on context. These homonyms must be clearly separated in the thesaurus to prevent ambiguity. The

integration of lexical resources, ontologies, and computational methods is essential for refining semantic distinctions in the thesaurus. The development of electronic thesauri also involves leveraging NLP techniques and AI models to analyze large text corpora and detect semantic patterns. Word embedding models, such as Word2Vec and BERT, help identify related words by examining their usage in vast linguistic datasets. These computational tools enable the automatic extraction of synonyms, hypernyms, and other lexical relations, improving the efficiency of thesaurus construction.

The process of thesaurus development requires expert linguistic validation to ensure accuracy. While machine learning models can identify word associations, human expertise is needed to interpret semantic nuances and refine the structure of synsets. Linguists contribute to defining relationships between words, verifying meanings, and organizing lexical data systematically. In addition to lexical relations, the thesaurus must consider grammatical and morphological aspects. Uzbek, as an agglutinative language, features extensive word derivation through suffixation, requiring the thesaurus to include morphological variations of words. This ensures that related word forms, such as *yangi* (new) and *yangilik* (novelty), are appropriately linked within the lexical database. The thesaurus also plays a critical role in computational applications such as machine translation, information retrieval, and language modeling. By structuring words into well-defined semantic groups, it enhances the performance of AI-driven language processing tools. Machine translation systems, for example, benefit from thesaurus-based word sense disambiguation, leading to more accurate translations. In information retrieval, thesauri improve search precision by expanding queries to include synonyms and related terms. A search for *kitob* (book) can retrieve results for *asar* (work), *qo‘llanma* (manual), and *monografiya* (monograph) by recognizing their semantic connections. This enhances search effectiveness and ensures comprehensive retrieval of relevant information. The application of electronic thesauri in AI-driven text analysis further extends their utility. Sentiment analysis, topic modeling, and document classification rely on thesauri to interpret word meanings and establish contextual relationships. By incorporating thesaurus-based semantic knowledge, these AI models achieve higher accuracy in processing and understanding natural language texts. The visualization of semantic networks within the thesaurus aids in comprehending word associations. Graph-based representations, such as knowledge graphs, enable intuitive exploration of lexical relationships, facilitating linguistic research and AI development.

The integration of electronic thesauri with search engines further enhances their functionality. By embedding thesaurus-based semantic analysis into search algorithms, systems can deliver more relevant results by recognizing conceptual similarities

between words. This enables more effective information retrieval and knowledge discovery. The construction of an Uzbek electronic thesaurus requires collaboration between linguists, computer scientists, and AI researchers. The combination of linguistic expertise and computational methodologies ensures the creation of a robust lexical resource that serves both academic and technological applications. Ultimately, the development of electronic thesauri represents a significant advancement in computational linguistics, bridging the gap between traditional lexicography and modern AI-driven language processing. By refining semantic structures and leveraging AI technologies, thesauri contribute to more accurate, context-aware natural language understanding.

References:

1. Jurafsky, D., & Martin, J. H. (2021). Speech and language processing (3rd ed.). Pearson.
2. Fellbaum, C. (1998). WordNet: An electronic lexical database. MIT Press.
3. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.
4. Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation, and application of a wide-coverage multilingual semantic network. Artificial Intelligence, 193, 217-250.