ARTIFICIAL INTELLIGENCE IN BILINGUAL LEXICOGRAPHY: DEVELOPING AN AI-ENHANCED UZBEK-ENGLISH THEMATIC LEARNER'S DICTIONARY FOR EDUCATION

Adakhamova Mukaddaskhon Nuriddin kizi
Teacher, Sharda University Uzbekistan
muqaddasxonadahamova@gmail.com

Annotation. *The advancement of artificial intelligence (AI) and natural language processing (NLP) has significantly influenced lexicographic research worldwide. However, low-resource languages such as Uzbek still lack digital corpora and computational tools for dictionary creation. This paper presents a conceptual framework for developing an AI-enhanced Uzbek-English thematic learner's dictionary focused on education-related vocabulary. The research emphasizes the integration of AI technologies – such as semantic clustering, machine learning, and bilingual embeddings – to automate corpus analysis and thematic grouping. This model aims to improve vocabulary acquisition, translation accuracy, and pedagogical usability while contributing to the digital development of the Uzbek language.*

Keywords: *artificial intelligence, bilingual lexicography, thematic dictionary, NLP, Uzbek-English, educational vocabulary, corpus linguistics.*

Introduction

Lexicography is undergoing a transformation due to the application of artificial intelligence (AI) and natural language processing (NLP). Traditional dictionaries, though linguistically rich, often fail to reflect modern pedagogical and technological requirements. Particularly in bilingual education, learners need contextualized and thematic vocabulary resources to support academic communication.

In Uzbekistan, the creation of a high-quality Uzbek-English dictionary remains challenging because of limited digital corpora and the lack of computational lexicographic tools. A thematic learner's dictionary – organized by conceptual domains such as "Classroom Language," "Assessment," or "Educational Technology" – can better serve the needs of teachers, students, and translators.

The proposed AI-based lexicographic model involves several steps:

1. Corpus compilation – collecting bilingual educational materials including textbooks, lectures, and research texts;
2. Lexical extraction – applying NLP algorithms and statistical measures (e.g., TF-IDF) to identify key vocabulary;
3. Semantic clustering – using AI models such as *word2vec* or *FastText* to group words by thematic proximity;

4. Cross-lingual alignment – aligning Uzbek and English equivalents through bilingual embeddings;

5. Manual validation – reviewing automatically generated clusters to ensure semantic and pedagogical accuracy.

The resulting dictionary will contain approximately 12 thematic modules, each with Uzbek entries, English equivalents, grammatical categories, pronunciation, and bilingual examples. For example:

Uzbek word	English equivalent	POS	Example (Uzbek)	Example (English)	Theme
Baholash	Assessment	Noun	Talabalar baholash jarayonida faol ishtirok etadi.	Students actively participate in the assessment process.	Assessment & Evaluation

This human-AI hybrid approach enhances both the linguistic depth and technical scalability of dictionary creation.

Pedagogically, thematic dictionaries provide context-based vocabulary that strengthens students' communicative competence. Learners acquire and retain words more effectively when they are linked conceptually rather than alphabetically. For example, under the "Assessment" theme, users can learn interrelated words such as *grade*, *test*, *exam*, *rubric*, and *criteria*, which improves associative memory and contextual usage.

Despite its potential, several challenges exist: the scarcity of Uzbek-language corpora, structural differences between Uzbek and English, and the need for technical infrastructure for AI implementation. Nevertheless, integrating AI in lexicography offers a sustainable solution for developing modern, accessible, and pedagogically relevant bilingual dictionaries.

Conclusion

The development of an AI-enhanced Uzbek-English thematic learner's dictionary represents a significant step toward modernizing bilingual lexicography in Uzbekistan. It combines linguistic theory, corpus linguistics, and artificial intelligence to create a dynamic and learner-oriented resource. The project contributes not only to educational technology but also to the digital preservation and advancement of the Uzbek language.

Future research may extend this model to other language pairs such as – Uzbek-Russian or Uzbek-Korean – thereby strengthening multilingual education and intercultural communication in the region.

References

1. Atkins, S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford University Press.
2. Kilgarriff, A. (2013). Corpora and Machine Learning in Lexicography. *International Journal of Lexicography*, 26(4), 446–461.
3. Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The Automatic Construction of a Multilingual Semantic Network. *Artificial Intelligence*, 193, 217–250.
4. Nazarov, A. (2020). Developing Digital Resources for Uzbek Lexicography. *Uzbek Linguistics Review*, 12(2), 56–65.
5. Tarp, S. (2021). Learner's Lexicography in the Age of Artificial Intelligence. *Lexikos*, 31, 356–377.