



A LINGUISTIC ANALYSIS OF AI TERMINOLOGY IN ENGLISH AND UZBEK: FOCUS ON NATURAL LANGUAGE PROCESSING

Hayitova Nigora Rahmatillayevna

Xorijiy filologiya kafedrasida katta o'qituvchisi

Renessans ta'lim universiteti

hayitovanigora@gmail.com

DOI: <https://doi.org/10.5281/zenodo.17933772>

Annotatsiya. Ushbu maqolada sun'iy intellekt (SI) sohasidagi ingliz va o'zbek tillaridagi leksik xususiyatlar, ayniqsa, tabiiy tilni qayta ishlash (Natural Language Processing – NLP) terminologiyasiga e'tibor qaratiladi. Tadqiqotning maqsadi ikki tilda SI bilan bog'liq atamalarining shakllanishi va moslashtirilishidagi leksik, morfologik va semantik naqshlarni aniqlashdir. Qiyosiy tahlil orqali asosiy NLP terminlarining o'zbek tilida qanday tarjima qilinishi, qarz olinishi yoki yangi so'z sifatida yaratilishi va bu lingvistik strategiyalarning SI sohasidagi til birligi va izchillikka ta'siri o'rganiladi. Tadqiqot natijalari yangi ilmiy sohalarida standartlashtirilgan terminologiyani rivojlantirishga hissa qo'shadi hamda tez o'zgarib borayotgan texnik lug'atlarni mahalliyashtirishdagi muammolarni yoritadi.

Kalit so'zlar: Sun'iy intellekt, NLP, terminologiya, o'zbek tili, ingliz tili, lingvistik tahlil, tarjima, leksik moslashuv.

Abstract. This paper explores the linguistic features of Artificial Intelligence (AI) vocabulary in English and Uzbek, focusing particularly on terminology used in Natural Language Processing (NLP). The research aims to identify lexical, morphological, and semantic patterns in the formation and adaptation of AI-related terms across the two languages. Through a comparative analysis, the study examines how key NLP terms are translated, borrowed, or newly coined in Uzbek, and how these linguistic strategies affect the clarity and consistency of AI discourse. The findings contribute to the development of standardized terminology in emerging scientific fields and highlight the challenges faced in localizing rapidly evolving technical vocabularies.

Keywords: Artificial Intelligence, NLP, terminology, Uzbek, English, linguistic analysis, translation, lexical adaptation.

Introduction. The rapid development of Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies has significantly impacted not only computer science but also the field of linguistics. As AI systems increasingly interact with human languages, the need for precise and standardized terminologies has become crucial. In English, as the dominant language of science and technology, a vast array of AI-related terms has emerged. These terms are now being adapted into other languages, including Uzbek, through various linguistic processes such as borrowing, calquing, or neologism. However, the adaptation of AI vocabulary into Uzbek presents several linguistic challenges. Given that many of the core concepts in AI and NLP are abstract and technical, finding exact semantic equivalents in Uzbek can be problematic. Moreover, due to the relatively recent introduction of these terms into Uzbek discourse, there is often inconsistency in their usage and translation. The



purpose of this study is to analyze the linguistic features of NLP-related terminology in both English and Uzbek. By comparing selected terms from each language, the paper seeks to identify common patterns, translation strategies, and linguistic issues that arise in the process of localization. Such an analysis can aid in the development of more consistent and understandable technical language in Uzbek, thereby enhancing scientific communication and education in the field.

Theoretical background. Lexical studies in linguistics examine how words are formed, used, and transformed across different contexts and languages. In the context of scientific and technological advancement, especially in fast-growing fields like AI and NLP, terminology evolves rapidly. English serves as the primary source of such vocabulary, and other languages—including Uzbek—tend to import or adapt these terms for local use.

Several translation and adaptation strategies are employed in localizing English technical vocabulary into Uzbek, including:

- Loan translation (*calque*): where the meaning is translated component by component (e.g., *machine learning* → mashinaviy o‘rganish);
- Phonetic borrowing: direct borrowing of the term with phonetic modification (e.g., *token* → token);
- Semantic adaptation: where a native word is assigned a new meaning to reflect a technical concept;
- Neologism creation: inventing a new term based on native roots (e.g., o‘zaro bog‘liqlik for *interdependence*).

However, without a standardized approach, the result is often inconsistency. For example, the English term *tokenization* may appear in Uzbek as tokenlash, bo‘laklash,

or so‘zlarga ajratish, depending on the context or translator. Such variation complicates comprehension and academic writing in Uzbek.

Previous studies (e.g., Jurafsky & Martin, 2020; Karimov, 2022) have addressed terminology in AI, but few have conducted detailed linguistic comparisons between English and Uzbek, particularly in the context of NLP.

This paper builds on existing lexical theory, terminology science, and translation studies to highlight the unique challenges and opportunities in aligning AI vocabulary across English and Uzbek.

Comparative analysis of NLP terminology. The terminology used in Natural Language Processing (NLP) reflects the complexity and interdisciplinary nature of the field. This section provides a comparative linguistic analysis of selected key terms



in English and their equivalents or adaptations in Uzbek. The focus is on morphological structure, semantic nuances, and translation strategies.

Morphological Features. Many English NLP terms are compound words or phrases formed by combining roots with affixes, such as *tokenization*, *embedding*, and *language model*. These terms often contain Latin or Greek roots, which are not native to Uzbek. Consequently, the Uzbek language tends to use several strategies to render these concepts:

- Borrowing with phonetic adaptation: Some terms are borrowed almost directly with phonological adjustment. For example, *token* becomes *token*, *embedding* may be borrowed as *embedding*.
- Calquing (loan translation): Complex terms are translated literally by components, for example, *machine learning* → *mashinaviy o'rganish*, *language model* → *tiliy model*.
- Neologism formation: Sometimes new native terms are coined to capture the meaning more intuitively, such as *o'zaro bog'liqlik* (interdependence).

Semantic Features. Semantic equivalence between English and Uzbek NLP terms is not always straightforward. Some English terms have multiple related meanings depending on context, which complicates direct translation.

For example, *tokenization* in English refers to segmenting text into tokens (words, subwords, or characters). In Uzbek, the equivalents vary: *tokenlash* (phonetic borrowing), *bo'laklash* (segmentation), or *so'zlarga ajratish* (separation into words). The choice depends on the translator's background and intended audience, causing inconsistency.

Another example is *embedding*, which in English refers to a vector representation of words or phrases. In Uzbek, *embedding* is commonly borrowed, but alternative descriptions like *o'rinlashtirish* (placement) are sometimes used to better reflect the conceptual meaning.

Translation Challenges and Strategies. The main challenges in localizing NLP terminology into Uzbek include:

- Lack of standardized equivalents, leading to multiple variants for a single concept.
- Morphological constraints of Uzbek, which is an agglutinative language, unlike English.
- Semantic gaps, where some English technical terms do not have direct Uzbek counterparts.

To address these challenges, experts recommend:

- Establishing a standardized glossary of AI and NLP terms in Uzbek.
- Promoting corpus-based research to study term usage and acceptance.



- Encouraging collaboration between linguists and AI specialists to develop terminology that is both linguistically appropriate and scientifically accurate.

Conclusion. This study highlights the linguistic features and translation challenges of AI and NLP terminology in English and Uzbek. The comparative analysis reveals a dynamic interaction of borrowing, calquing, and neologism in Uzbek lexical adaptation. While borrowing is prevalent due to the rapid influx of new terms, there is an increasing need for standardized native equivalents to ensure clarity and consistency. Developing a unified terminology will support the growth of AI research and education in Uzbekistan and facilitate international collaboration. Future research should focus on corpus analysis of Uzbek scientific texts and user studies to assess term comprehension and preference.

References

1. Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd ed.). Pearson.
2. Karimov, S. (2022). Challenges in Translating AI Terminology into Uzbek. *Journal of Linguistics and AI*, 8(1), 45–60.
3. Smith, J. (2019). Terminology Management in Natural Language Processing. *International Conference on Computational Linguistics*, 123–130.
4. Uzbek National Corpus. (2023). Retrieved from <http://uzcorpus.uz>
5. Vasheeva, D. (2021). Neologisms and Borrowings in Uzbek Scientific Language. *Uzbek Linguistics Review*, 5(2), 20–35.