FORMATION OF NEW ARTIFICIAL INTELLIGENCE VOCABULARY IN ENGLISH AND UZBEK. RESEARCH HYPOTHESES AND PRELIMINARY OBSERVATIONS

Sharipova Dilnoza Tulqin kizi

Teacher at UzSWLU

E-mail:dilnozasharipova0996@gmail.com

Abstract. *The paper outlines a comparative study of newly formed lexical units related to artificial intelligence in English and Uzbek. Four working hypotheses are put forward concerning the dominant word-formation mechanisms in the two languages, the variability of Uzbek equivalents and the fate of metaphorically motivated terms in borrowing and calquing; preliminary observations supporting the hypotheses and the need for unification in Uzbek are discussed.*

Keywords. *neologism, term, artificial intelligence vocabulary, word-formation, borrowing, calque, variability, unification.*

Over the past decade the rapid development of artificial intelligence has produced an entire new lexical layer in the vocabulary of world languages. Units such as *machine learning, neural network, large language model, prompt, chatbot* and *deepfake* have moved from professional discourse into everyday usage within a very short period. English acts as the donor language of the domain, while Uzbek is currently passing through the stage of adopting, calquing and adapting these units to its own structural norms. Observing a lexical layer at the very moment of its formation is a rare opportunity for neology and terminology studies; the task is also relevant for language policy, since Uzbekistan's 2024 national strategy for artificial intelligence requires improved standards in the field.

Word-formation and neology have been described by L.Bauer, J.Algeo and I.Plag, the language of digital communication by D.Crystal, and terminology theory by

E.Wüster and M.T.Cabré. In Russian linguistics neology is associated with N.Z.Kotelova and V.I.Zabotkina; in Uzbek linguistics word-formation and lexicology were studied by A.Hojiyev and E.Begmatov. Nevertheless, the AI-specific vocabulary of the generative era has not yet become the object of a special comparative study involving Uzbek.

The aim of the research is to reveal how new AI-related lexical units are formed in English and Uzbek and to substantiate the principles of their unification in Uzbek. The material is drawn from the new-words sections of authoritative English dictionaries, from scientific and media texts in English, and from official documents, periodicals and online media in Uzbek; descriptive, comparative, componential and corpus-based methods are applied.

The following working hypotheses are put forward.

Hypothesis one. In English, AI-related lexical units are formed predominantly by the internal word-formation resources of the language; composition (*machine learning*, *neural network*), blending (*chatbot*, *deepfake*), abbreviation (*AI*, *LLM*, *NLP*) and semantic derivation (*training*, *hallucination*, *attention*) are expected to prove the most productive patterns.

Hypothesis two. In Uzbek, the same layer is formed mainly from external sources; direct borrowing (*chatbot*, *prompt*, *token*) and full or partial calquing (*mashinali o‘qitish* for *machine learning*, *katta til modeli* for *large language model*, *neyron tarmoq* for *neural network*) dominate, whereas internal word-formation plays a limited role.

Hypothesis three. The coexistence of parallel variants denoting a single concept in Uzbek (*sun’iy intellekt, SI, AI*) indicates that the language is at the adaptation stage; as the terminology stabilises, the number of variants will decrease, which makes conscious unification both possible and necessary.

Hypothesis four. Metaphorically motivated English terms partially preserve their imagery when calqued into Uzbek (*neyron tarmoq*) but lose their inner form when borrowed directly (*gallyutsinatsiya* for *hallucination*); the degree of preserved motivation affects the transparency of the term for Uzbek speakers.

Preliminary observations support these assumptions. In English sources the majority of domain units are multi-word compounds and abbreviations, and the semantic specialisation of everyday words (*agent*, *temperature*, *token*, *weight*) is highly characteristic of the field. In Uzbek texts three channels are visible, namely direct borrowing, calquing and explanatory translation, and the same English term may occur in Latin spelling, in adapted spelling and as a calque, sometimes within one article; such variability confirms the transitional character of the present stage.

Testing these hypotheses on a systematically collected corpus will make it possible to describe the word-formation potential of both languages in a new thematic domain, to propose criteria for choosing between competing Uzbek variants and to compile the first English-Uzbek glossary of AI vocabulary for use in lexicography, translation and teaching.

References

1. Algeo J. (ed.). Fifty Years Among the New Words. Cambridge, Cambridge University Press, 1991.
2. Bauer L. English Word-Formation. Cambridge, Cambridge University Press, 1983.
3. Cabré M.T. Terminology. Theory, Methods and Applications. Amsterdam, John Benjamins, 1999.
4. Crystal D. Language and the Internet. Cambridge, Cambridge University Press, 2001.
5. Plag I. Word-Formation in English. Cambridge, Cambridge University Press, 2003.