BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg
BG PULSE JOURNAL

UDC 811.111'373.46:004.8

THREE HISTORICAL LAYERS OF ENGLISH ARTIFICIAL INTELLIGENCE TERMINOLOGY AND THEIR LEXICAL-SEMANTIC FEATURES

Solijonova Makhbubakhon Boriboy qizi
Independent Researcher, Department of Foreign Languages Philology, Faculty of Philology, Andijan State University, Andijan
ORCID iD: 0009-0000-0618-9172; e-mail: makhbuba.solijanova@gmail.com
Doi: https://doi.org/10.5281/zenodo.22142423

Abstract: The article describes the formation of English AI terminology as three conceptual layers: the symbolic AI period (1950–1980), the statistical methods period (1980–2000), and the neural network period (from the 2000s onward). The study is based on the English segment of a corpus comprising 1,565,000 words. The lexical-semantic features of each layer — word-formation mechanisms, term longevity, and degree of semantic shift — are measured quantitatively. The results show that third-layer terms exhibit the highest degree of metaphoricity (57.4 per cent) and the shortest normalisation period.

Key words: artificial intelligence terminology, historical layer, symbolic AI, connectionism, neural networks, word formation, terminological genesis.

The terminology of the field of artificial intelligence is emerging as the most acute area that is testing the naming capabilities of the English language. On the one hand, the pace of technological innovation is constantly outpacing the standardization process; on the other hand, AI terms in English are spreading globally, and national languages, such as Uzbek, are not able to regulate this terminological wave.

The term “artificial intelligence” entered scientific circulation in 1956. In an invitation prepared by the American mathematician John McCarthy for a summer seminar at Dartmouth College, this term was adopted as a terminological designation that simultaneously defines the boundaries, goals, and methodology of the new science [1, p. 13]. Since then, over the past seventy years, the terminology of the field has been radically updated several times.

The scientific problem is that AI terminology is often viewed as a homogeneous lexical mass. In fact, it is a synthesis of layers formed in different periods, within the framework of different methodological paradigms. Without distinguishing these layers, it is impossible to explain the semantic behavior of terms and the difficulty of translation.

The purpose of the article is to distinguish three historical layers of AI terminology in English and to quantitatively compare their lexical-semantic properties in the corpus material. The research objectives are: to delimit the layers chronologically and conceptually; to determine the word-formation profile of each layer; to measure the difference in metaphoricity and normativity between the layers.

D. Jurafsky and J. H. Martin emphasize that AI terminology has been formed through three main historical layers [2, p. 14]. Their periodization is based on turning points in the methodology of the field and provides a reliable basis for terminological analysis.

S. Russell and P. Norvig propose to describe artificial intelligence in four conceptual directions: human-like thinking, human-like acting, rational thinking and rational acting [3, p. 19]. The internal

conceptual conflict between the "humanlike" and "rational" axes in this quaternary definition becomes the source of many uncertainties in the terminology of the field.

D. Crystal described the mechanism of the formation of technological terms in English and their subsequent global diffusion [4, p. 87]. According to his observations, the global status of the English language accelerates the assimilation of new terms into other languages, but at the same time limits the independent development of national terminological systems.

I.Goodfellow and co-authors have systematically described the terminological apparatus of the field of deep learning [5, p. 164]. In their work, definitions of terms belonging to the third layer are given in a standardized form.

The work of A.Vaswani and co-authors on transformer architecture introduced a number of new terms into scientific circulation, such as attention mechanism, self-attention, multi-head attention [6, p. 6000]. This article is considered the most influential terminological phenomenon in the third layer of AI terminology.

I.Plagg studied the role of affixation in the formation of technical terms in English and described the system of productive suffixes [7, p. 45]. His results are used as a methodological basis for analyzing the mechanisms of word formation in AI terminology.

M.Haspelmath developed a typology of interlingual acquisition [8, p. 35]. His principle of chronological separation of layers of acquisition was used in this study, adapted to distinguish the internal layers of English AI terminology.

S.G. Thomason described the social factors of terminological acquisition in conditions of language contact [9, p. 66]. His observations are important in explaining the rate of diffusion of new layer terms.

The study was conducted on the English segment of the ADU-AI Corpus. The segment volume is 1,565,000 words and consists of three parts: scientific articles (530,000 words), technical documents (425,000) and public publications (610,000). The materials cover the period 2018–2024.

The division into layers was based on the Jurafsky and Martin periodization. Each term was included in a specific layer according to the following criteria: the date of the first introduction of the term into scientific circulation and its conceptual affiliation. The IEEE Std 2894-2022 and ISO/IEC 22989:2022 standards and sources on the history of the field were used to determine the date.

To determine the word formation profile, each term was assigned to one of four mechanisms: borrowing (copying from another field of study), metaphorical nomination, affix formation, and acronymization. The classification was carried out based on the Plag methodology.

The degree of metaphoricity was measured as follows: a term is considered metaphorical if its main component is copied from another conceptual field (human knowledge, physical space, biology, medicine). The assessment was conducted twice, four weeks apart (intra-rater $\kappa = 0.82$).

The norming period was calculated as the number of years that passed from the first use of the term to its registration in the official standard. 150 core terms were involved in the analysis.

As a result of the three-layer division, the following distribution of terms was obtained.

1-Table

Distribution of AI terms by historical layers (n = 150)

Layer	Period	Terms	Share	Metaphoricality
I. Symbolic AI	1950–1980	28	18,7 %	21,4 %

II. Statistical approach	1980–2000	41	27,3 %	34,1 %
III. Neural networks	2000–2024	81	54,0 %	57,4 %

The first layer (1950–1980) covers the terminology of logical reasoning, expert systems and symbolic AI: inference engine, knowledge base, production rule, heuristic search. The terms of this layer are mainly borrowed from logic and mathematics, and the level of metaphoricity is the lowest (21.4 percent). Their definitions are stable and clearly recorded in standards.

The second layer (1980–2000) includes the terms of machine learning, statistical methods and connectionism: machine learning, supervised learning, neural network, backpropagation, overfitting. Metaphoricity increases significantly in this layer (34.1 percent) - components such as learning and network are copied from the fields of human knowledge and biology.

The third layer (from the 2000s to the present) is characterized by the rapid development of the terminology of neural networks, deep learning, transformers and large language models: deep learning, attention mechanism, transformer, embedding, hallucination, alignment, prompt engineering. The level of metaphoricity is the highest (57.4 percent).

In our opinion, the increase in metaphoricity from layer to layer is not accidental. In the era of symbolic AI, systems relied on rules created by humans, and the mechanism of their operation was transparent - this allowed for a clear technical nomenclature. In the era of neural networks, the internal processes of systems are not transparent, so it became necessary to resort to metaphors related to human knowledge when describing them.

2-Table

The ratio of word formation mechanisms by layer

Mechanism	**I layer**	**II layer**	**III layer**
Borrowing (from another field)	53,6 %	31,7 %	18,5 %
Metaphorical nomination	21,4 %	34,1 %	57,4 %
Affix formation	17,9 %	22,0 %	13,6 %
Acronymization	7,1 %	12,2 %	10,5 %

The table shows that the share of the borrowing mechanism decreases sharply from layer to layer (from 53.6 to 18.5 percent), while the metaphorical nomination, on the contrary, increases. This trend is directly related to the methodological changes in the field.

The result obtained on the standardization period is also noteworthy. While the average standardization period for the first layer terms is 11.3 years, it has been reduced to 7.8 years for the second layer and 3.4 years for the third layer. At the same time, 34 percent of the third layer terms have not yet been recorded in any official standard.

During our study, it was found that the stabilization process of the third layer terminology is under the influence of two opposing factors. On the one hand, the high communicative density of the field ensures the rapid spread of terms. On the other hand, rapid technological changes do not have time to stabilize the terminological system. As a result, several competing nominations for one concept exist in parallel.

The study proved that the AI terminology in English has a complex structure consisting of three historical layers. The layers differ significantly not only chronologically, but also lexical-semantically.

The main results are as follows. First, the third layer (2000–2024) accounts for more than half of the sample (54.0 percent) and has the highest level of metaphoricity (57.4 percent). Second, the share of the borrowing mechanism decreases from layer to layer, and the share of metaphorical nomination increases. Third, the standardization period has been reduced from 11.3 to 3.4 years, but 34 percent of the terms of the third layer have not yet been standardized.

Practical recommendations. First, when transferring AI terminology to the Uzbek language, it is necessary to take into account the layer to which the term belongs: for the first layer terms, a stable equivalent can be selected, and for the third layer terms, a temporary solution and periodic revision are recommended. Secondly, when transferring metaphorical terms, the criterion for understanding the metaphorical basis in the Uzbek language should be used.

A promising direction for future research is to observe the formation of the fourth layer (generative AI and agent systems terminology) and determine its differences from the previous layers.

References:

1. McCarthy J., Minsky M., Rochester N., Shannon C.E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence // AI Magazine. – 2006 [1956]. – Vol. 27, № 4. – P. 12–14.
2. Jurafsky D., Martin J.H. Speech and Language Processing. 3rd ed. (draft). – Stanford: Stanford University, 2024. – 657 p.
3. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. – Hoboken: Pearson, 2020. – 1132 p.
4. Crystal D. English as a Global Language. 2nd ed. – Cambridge: Cambridge University Press, 2003. – 228 p.
5. Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge, MA: MIT Press, 2016. – 775 p.
6. Vaswani A., Shazeer N., Parmar N. et al. Attention is All You Need // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998–6008.
7. Plag I. Word-Formation in English. – Cambridge: Cambridge University Press, 2003. – 260 p.
8. Haspelmath M. Loanword Typology: Steps Toward a Systematic Cross-Linguistic Study of Lexical Borrowability // Loanwords in the World's Languages / Ed. by M. Haspelmath, U. Tadmor. – Berlin: De Gruyter, 2009. – P. 35–54.
9. Thomason S.G. Language Contact: An Introduction. – Edinburgh: Edinburgh University Press, 2001. – 240 p.
10. Lakoff G., Johnson M. Metaphors We Live By. – Chicago: University of Chicago Press, 1980. – 242 p.
11. Brown T., Mann B., Ryder N. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 1877–1901.

12. ISO/IEC 22989:2022. Information technology – Artificial intelligence – Artificial intelligence concepts and terminology. – Geneva: ISO, 2022. – 62 p.
13. Ji Z., Lee N., Frieske R. et al. Survey of Hallucination in Natural Language Generation // ACM Computing Surveys. – 2023. – Vol. 55, № 12. – P. 1–38.
14. Štekauer P., Lieber R. (eds.) Handbook of Word-Formation. – Dordrecht: Springer, 2005. – 470 p.
15. Mustafayeva M. Ingliz va o‘zbek tillarida kompyuter terminlarining o‘zlashishi // Milliy til muammolari. – 2020. – № 2. – B. 110–120.