

CORPUS-BASED COMPARATIVE ANALYSIS OF AI TERMINOLOGY IN ENGLISH AND UZBEK

Nazarova Sevinchoy Umarbekovna
Urgench RANCH University of Technology

Abstract: The fast pace of the Artificial Intelligence (AI) development has resulted in the appearance of an extremely large amount of specialized vocabulary, most of which is in English. The paper is a corpus-based comparative analysis of AI terminology in English and Uzbek that analyzes the usage patterns, translation strategies, and semantic adaptation. The study based on two similar corpora that are compiled using scholarly articles, news outlets, and online editions, finds main distinctions between frequency, lexical organization, and conceptual equivalence. The results show that English AI terms have a high standardization and frequency consistency, whereas Uzbek show a hybrid model of direct translation, transliteration, and borrowing. The paper mentions the difficulties of standardization of terminology and offers suggestions in the further linguistic and pedagogical evolution of Uzbek AI discourse.

Keywords: corpus linguistics, AI lexicon, language Uzbek, comparative analysis, translation strategies.

1. Introduction

The growth of Artificial Intelligence has had a very strong impact on the communication of the whole world, especially in the sphere of the specialized vocabulary. Since English is still the most widespread language of technological advancement, most of the terms related to AI are of English origin and are later translated or borrowed to other languages.

The Corpus linguistics has offered a practical basis of studying the actual usage of language and detecting trends in actual texts [1: 24]. The growing use of digital technologies in Uzbekistan has made the assimilation of AI terms into the Uzbek language faster. Nevertheless, the terminological adaptation process is not consistent and, as a result, it causes a difference in usage, meaning, and standardization.

This paper seeks to engage in a corpus-based comparative study of AI terms in English and Uzbek, with the aim of studying their usage patterns, translation approaches, and semantic equivalence.

2. Literature Review

The application of corpus-based studies has been extensively used in linguistic researches to examine frequency, collocation, and contextual meaning. Sinclair

underlines that the use of corpus analysis enables researchers to see the real usage of language instead of its description according to the theoretical framework [1: 30].

The English language has been popularly accepted as the universal language of science and technology. According to Crystal, English is used in international communication especially in areas that are specialized like information technology and artificial intelligence [2: 5].

In terms of translation strategies, the direct translation, borrowing, and hybridization are most common in the sphere of the study of terminology. Bowker and Pearson emphasize that specialized language is not always easily translated because of the differences in concepts and cultures, and loanwords are used [3: 78].

In spite of such developments, little corpus-based research on AI terms in Uzbek has been done. This research aims to fill this gap by offering a comparative analysis in a systematic manner.

3. Methodology

3.1 Research Design

The study adopts a mixed-method corpus-based methodology that uses both quantitative and qualitative analyses.

3.2 Data Collection

Two similar corpora were put together:

Corpus of English: Around 500,000 words in scholarly journals, blogs on AI and news on technology.

Uzbek Corpus: About 300,000 words of online sources, educational texts and local news sources.

3.3 Data Analysis

Corpus analysis tools like AntConc were used to conduct the analysis. The following techniques were used:

Frequency analysis to determine frequently used AI words.

Concordance to investigate contextual use.

Strategy comparison of translations.

The most important terms to be analyzed are Artificial Intelligence, Machine Learning, Neural Network, and Deep Learning.

4. Results

4.1 Frequency Distribution

Analysis shows that English AI terminology is more regularly and more commonly used compared to its Uzbek counterparts. As an example, Artificial Intelligence is more frequently found than sun'iy intellekt, which means that it is more standardized in the English corpus.

4.2 Translation Strategies

Uzbek AI terminology has found three main strategies:

Direct Translation

Artificial intelligence -suniy intellekt.

Transliteration

Algorithm → algoritm

Borrowing / Hybrid Forms

Machine Learning- mashinaviy o'rganish / machine learning.

This fact of borrowing is consistent with the results of terminology research, where the rapid technical advancement does not allow creating native equivalents [3: 80].

4.3 Semantic Variation

There are also inconsistencies in semantic equivalence in the study. There are words with no consistent translation, like Deep Learning, which result in the differences in meaning and perception among Uzbek texts.

5. Discussion

The results validate the predominant position of English in influencing AI terms in Uzbek. The prevalence of borrowed and transliterated words suggests that Uzbek is yet to become accustomed to the fast-changing sphere of artificial intelligence.

The absence of standard terminology poses difficulties to education and professional communication. According to Crystal, linguistic dependency in the emerging fields is usually caused by the dominance of English [2: 12].

Moreover, corpus data indicate that Uzbek equivalents are still to be developed to the full extent, which can influence its understanding and availability to non-English speakers.

6. Conclusion

This paper will offer corpus-based comparative analysis of AI terminology in English and Uzbek, pointing out the difference in frequency, translation strategies, and semantic adaptation.

According to the results, the English terminology is quite standardized, whereas Uzbek is variable and has to rely on borrowing. These results reinforce the fact that standardization and development of terminology in Uzbek need to be systematic.

Future studies ought to widen the sample and examine the pedagogical effects of AI terminologies in language studies.

References

1. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
2. Crystal, D. (2003). *English as a Global Language*. Cambridge: Cambridge University Press.
3. Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.