INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT
eISSN 2394-6334 Impact factor: 8.854 Volume 13, issue 04 (2026)

MODELING THE SEMANTIC INTERPRETATION OF PHRASEOLOGICAL UNITS IN AI-BASED MACHINE TRANSLATION SYSTEMS: A COGNITIVE AND HYBRID APPROACH

Shukurova Yulduz Yaxshimurotovna
Department of Foreign Language and Social Sciences,
Asia International University, English Teacher, Bukhara, Uzbekistan
E-mail: stellashukurova23@gmail.com

Abstract. This article provides a comprehensive analysis of the problem of semantic interpretation of phraseological units (idiomatic expressions) in artificial intelligence (AI)-based machine translation systems, with a particular focus on Uzbek-English language pairs. Phraseological units pose significant challenges due to their morphological complexity, polysemy, and cultural specificity.

The study compares different translation approaches, including rule-based, statistical phrase-based, neural machine translation, and large language models, highlighting their strengths and limitations. It is shown that these systems often struggle with capturing contextual meaning and tend to produce literal translations.

The research emphasizes the importance of integrating cognitive linguistics and conceptual metaphor theory into translation models. A set of six English-Uzbek phraseological units is analyzed contextually to demonstrate translation strategies.

Finally, the paper proposes solutions such as preprocessing, paraphrasing, phraseological dictionaries, and hybrid models to improve translation quality. The findings suggest that combining multiple approaches can significantly enhance the semantic accuracy of idiomatic translation. Future research directions are also outlined.

Key words: Phraseological units, idiomatic expressions, machine translation, artificial intelligence, neural machine translation, cognitive linguistics, semantic interpretation, hybrid translation models

Introduction. Phraseological units are idiomatic expressions whose meanings cannot be derived from the individual meanings of their components. As an integral part of figurative language, they embody cultural traditions, worldview, and cognitive patterns of a speech community.

They function as the “internal memory” of a language, making cultural context a crucial factor in translation. According to recent studies, phraseological units are closely linked to conceptual metaphors and reflect national cognitive models.

Therefore, accurate translation requires not only grammatical correctness but also the preservation of semantic, cultural, and psychological nuances.

Linguistic features of phraseological units. Phraseological units are typically non-compositional, meaning their meanings cannot be inferred directly from their components. For example, the Uzbek expression *“to‘kilgan sut uchun yig‘lama”* corresponds to the English *“don’t cry over spilled milk,”* while its functional Uzbek equivalent is *“o‘tgan ishga salovat.”*

Another important feature is polysemy. For instance, *“kick the bucket”* means “to die,” and its Uzbek equivalent is *“ko‘z yumdi,”* rather than a literal translation.

Phraseological units can be classified into proverbs, sayings, idioms, and fixed expressions. They may be fully fixed or partially flexible structures.

Every language has a rich system of phraseological units. In Uzbek, expressions such as *“tuyaning dumi yerga tekkanda”* illustrate culturally embedded metaphorical thinking.

Thus, their complex semantic nature necessitates specialized translation strategies.

AI-based translation systems and phraseological units

Machine translation has evolved through several stages: rule-based (RBMT), statistical (SMT), neural (NMT), and large language models (LLMs).

Rule-based machine translation (RBMT): This approach relies on linguistic rules and dictionaries. Its strength lies in precision and controllability. However, it requires extensive manual effort, especially for encoding phraseological units.

Statistical machine translation (SMT): SMT uses parallel corpora to generate phrase tables. It can automatically learn phrase correspondences but struggles with context and often produces literal translations.

Neural machine translation (NMT): NMT models analyze entire sentences and capture global context. They show improved performance but require large datasets and still tend to mistranslate idiomatic expressions.

Large language models (LLMs): Modern models demonstrate advanced contextual understanding. However, they may still generate incorrect or overly literal translations when encountering unfamiliar idioms.

Hybrid approaches: Hybrid systems combine rule-based, statistical, and neural methods. They are particularly effective for low-resource languages like Uzbek, though they are complex to implement.

Comparative overview.

- **RBMT:** High precision, but resource-intensive
- **SMT:** Learns from data, but weak contextual understanding
- **NMT:** Strong context handling, but data-dependent
- **LLMs:** Advanced reasoning, but inconsistent accuracy
- **Hybrid:** Most promising, but technically complex

Cognitive approach and modeling.

Cognitive linguistics plays a crucial role in understanding phraseological meaning. Conceptual metaphor theory explains how idiomatic expressions encode abstract concepts through metaphor.

In translation, differences between conceptual systems must be addressed through **re-conceptualization**, where meaning is adapted to the target language's cognitive framework.

This approach allows translators and AI systems to go beyond literal meaning and capture deeper semantic structures.

Examples and contextual analysis

- *"Don't cry over spilled milk" → "o'tgan ishga salovat"*
- *"Let the cat out of the bag" → "sirni fosh qilmoq"*
- *"When pigs fly" → "tuyaning dumi yerga tekkanda"*
- *"Go Dutch" → "har kim o'z ulushini to'laydi"*
- *"Hit the sack" → "uxlagani yotmoq"*
- *"Break a leg" → "omad tilayman"*

These examples demonstrate that preserving meaning is more important than literal translation.

Recommendations and solutions

- **Preprocessing and paraphrasing:** Simplify idioms before translation
- **Phraseological dictionaries:** Integrate idiom databases
- **Context-aware modules:** Improve semantic understanding
- **Hybrid systems:** Combine multiple approaches

- **Evaluation metrics:** Develop idiom-specific evaluation methods

Conclusion

This study demonstrates that phraseological units present a significant challenge for AI-based translation systems due to their semantic complexity and cultural specificity.

While modern models show progress, they still require deeper contextual and cognitive integration. The combination of linguistic theory, AI methods, and hybrid architectures offers the most promising direction.

Future research should focus on building specialized corpora, improving evaluation metrics, and enhancing model training for idiomatic language.

References.

1. Yaxshimurotovna, S. Y. (2026). INTEGRATED STUDY OF AUTOMATED TRANSLATION QUALITY AND PHRASEOLOGICAL EQUIVALENCE IN ENGLISH-UZBEK TRANSLATION. *Shokh Articles Library*, *1*(1), 1044-1050.
2. Yaxshimurotovna, S. Y. (2025). CULTURAL-CONNOTATIVE FEATURES OF PHRASEOLOGICAL UNITS IN DIFFERENT LANGUAGES AND THEIR INTERPRETATION THROUGH ARTIFICIAL INTELLIGENCE. SHOKH LIBRARY, 1(13). Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. 14.
3. Shukurova Yulduz Yaxshimurotovna. (2026). SEMANTIC INTERPRETATION OF PHRASEOLOGICAL UNITS IN ARTIFICIAL INTELLIGENCE–BASED TRANSLATION SYSTEMS: CHALLENGES AND ENHANCEMENT STRATEGIES [Data set]. In Journal of Multidisciplinary Sciences and Innovations (Vol. 5, Number 02, pp. 1805–1808). Zenodo. https://doi.org/10.5281/zenodo.18788395
4. Djalilova Zarnigor Obidovna, & Shukurova Yulduz Yaxshimurotovna. (2025). LinguoculturalAnalysis of Phraseological Units with Onomastic Components. Zenodo. https://doi.org/10.5281/zenodo.14781399
5. CHALLENGES AND STRATEGIES IN TRANSLATING ENGLISH IDIOMS INTO UZBEK https://www.academia.edu/129376429/CHALLENGES_AND_STRATEGIES_IN_TRANSLATING_ENGLISH_IDIOMS_INTO_UZBEK
6. PolyFrame at MWE-2026 AdMIRe 2: When Words Are Not Enough: Multimodal Idiom Disambiguation https://arxiv.org/html/2602.18652v2