pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

MODELING THE SEMANTIC INTERPRETATION OF PHRASEOLOGICAL UNITS IN AI-BASED TRANSLATION SYSTEMS: A LINGUOCULTURAL APPROACH

Shukurova Yulduz Yaxshimurotovna
Department of Foreign Language and Social Sciences,
Asia International University, English Teacher
Bukhara, Uzbekistan
E-mail: stellashukurova23@gmail.com
https://doi.org/10.5281/zenodo.20285071

Abstract.
This study investigates the semantic interpretation of phraseological units within AI-based machine translation systems from a linguocultural perspective. Phraseological units are considered complex linguocultural entities that integrate semantic, metaphorical, cognitive, and cultural components. Although neural machine translation systems have achieved considerable success in recent years, they continue to face serious difficulties in interpreting idiomatic and culturally marked expressions. The research identifies the major limitations of AI translation systems in processing non-compositional meanings and culturally bound phraseological structures. To address these challenges, the study proposes a hybrid semantic interpretation framework that combines contextual embeddings, phraseological detection mechanisms, cultural annotation, and frame semantics. Comparative analysis of English and Uzbek phraseological units demonstrates that structural, semantic, and ethnocultural asymmetries significantly influence translation adequacy. The findings indicate that integrating linguocultural information into AI translation architectures substantially improves semantic accuracy and minimizes pragmatic distortion. The study contributes to the development of culturally adaptive and semantically sensitive translation models within the field of artificial intelligence and natural language processing.

Keywords: phraseological units, linguocultureme, semantic interpretation, machine translation, artificial intelligence, idioms, cultural semantics, NLP, translation strategies, frame semantics.

1. Introduction

The rapid advancement of artificial intelligence technologies has fundamentally transformed the field of translation studies, particularly through the emergence of neural machine translation systems. Modern AI-based translation tools are capable of processing vast amounts of multilingual data and producing grammatically accurate translations within seconds. Despite these technological achievements, phraseological units remain one of the most problematic linguistic categories for automatic translation systems. Their semantic complexity, figurative nature, and strong cultural associations create significant obstacles for computational interpretation.

Phraseological units, including idioms, proverbs, collocations, and fixed expressions, possess meanings that cannot be understood solely through the analysis of their individual lexical components. In many cases, their meanings are metaphorically motivated and culturally conditioned. Consequently, AI systems often generate literal translations that distort the intended communicative meaning and lead to pragmatic failure. For example, the English idiom “spill the beans” may be translated literally into Uzbek as “loviya to‘kmoq,” which fails to convey the intended meaning of revealing a secret. Such cases demonstrate the inability of purely statistical or neural approaches to capture deeper semantic and cultural dimensions of language.

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

From a linguocultural perspective, phraseological units function not only as linguistic structures but also as carriers of national mentality, historical experience, and cultural values. Therefore, their successful interpretation requires the integration of semantic, cognitive, and cultural knowledge into translation systems. This study aims to analyze the linguocultural nature of phraseological units, identify the primary challenges in their automatic translation, and propose a hybrid semantic interpretation model capable of improving phraseological translation in AI environments.

2. Theoretical Background

Phraseological units are widely recognized as one of the most culturally significant layers of language. According to linguocultural theory, they represent linguoculturemes, which are linguistic units containing both verbal meaning and cultural information. Researchers such as Telia and Vorobyov emphasize that phraseological expressions preserve collective cultural memory and reflect national patterns of thinking. Their semantic structure is often based on metaphorical associations, symbolic imagery, and ethnographic realities specific to a particular linguistic community.

One of the defining features of phraseological units is semantic opacity. Unlike free word combinations, the meaning of a phraseological expression cannot usually be predicted from the meanings of its separate elements. Their interpretation depends heavily on contextual, pragmatic, and cultural knowledge. For instance, the English phrase “break the ice” does not literally refer to physical ice but metaphorically describes initiating social interaction in an uncomfortable situation. Similarly, Uzbek phraseological units often contain culturally specific imagery connected to traditions, family relations, or everyday life.

Another important characteristic is structural stability. Phraseological expressions tend to preserve fixed lexical and grammatical forms over time, which distinguishes them from ordinary syntactic combinations. At the same time, they demonstrate strong context dependency because their interpretation may vary according to discourse situation, speaker intention, and sociocultural background. These features create serious challenges for AI systems that primarily rely on statistical probabilities and contextual prediction mechanisms.

In the context of natural language processing, semantic interpretation involves identifying the intended meaning of linguistic structures within a particular communicative environment. However, existing AI translation systems frequently fail to recognize idiomatic expressions as unified semantic entities. Instead, they process them word by word, resulting in literal and semantically inadequate translations. Therefore, the incorporation of linguocultural analysis into computational models becomes essential for achieving semantic equivalence in phraseological translation.

3. Methodology

This study employs a mixed-method research design combining comparative linguistic analysis, semantic interpretation, and error analysis of AI-generated translations. Approximately 150 English and Uzbek phraseological units were selected from literary texts, bilingual dictionaries, phraseological corpora, and authentic discourse materials. The selected expressions represent various semantic categories, including emotional states, social relations, behavioral characteristics, and communicative situations.

The research process consisted of several analytical stages. First, the phraseological units were classified according to their semantic structure, metaphorical basis, and cultural specificity. Second, their translations generated by AI-based systems were examined to identify major translation errors and semantic distortions. Third, a comparative linguocultural analysis was conducted to determine structural and conceptual differences between English and Uzbek phraseological systems.

The analytical framework of the study is based on a five-component model that includes metaphorical foundation, cultural code, ethnographic component, linguistic structure, and frame semantics. The metaphorical foundation examines figurative imagery underlying phraseological

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

meaning. The cultural code identifies national and historical associations embedded in the expression. The ethnographic component focuses on social traditions and cultural practices reflected in language. Linguistic structure analyzes grammatical and lexical stability, while frame semantics investigates the cognitive scenarios activated during interpretation.

This integrated methodological approach allows for a deeper understanding of how phraseological meaning is constructed and why AI systems encounter difficulties in translating culturally marked expressions.

4. Results and Analysis

The analysis revealed that AI-based translation systems produce several recurrent types of errors when processing phraseological units. The most common problem is literal translation. Since neural systems often analyze phraseological expressions compositionally, they generate translations based on direct lexical correspondence rather than figurative meaning. As a result, many idiomatic expressions lose their semantic and pragmatic functions in translation.

Another major issue is cultural mismatch. Certain phraseological units contain culturally specific concepts that have no direct equivalents in the target language. English idioms may reflect historical events, religious symbolism, or social realities unfamiliar to Uzbek linguistic culture, while Uzbek phraseological expressions frequently contain ethnographic imagery connected with local traditions and collective values. AI systems struggle to establish adequate semantic correspondence in such cases because they lack cultural background knowledge.

Pragmatic failure was also identified as a significant challenge. Even when literal meaning is partially preserved, AI-generated translations often fail to convey the emotional tone, stylistic nuance, or communicative intention of the original expression. This problem reduces translation naturalness and affects intercultural communication.

Comparative analysis demonstrated important differences between English and Uzbek phraseological systems. English phraseological units tend to exhibit relatively fixed structures and abstract metaphorical patterns, whereas Uzbek expressions are often more flexible and grounded in concrete ethnographic imagery. These asymmetries complicate the process of semantic mapping between the two languages and increase the probability of translation errors.

The findings confirm that phraseological translation cannot rely exclusively on lexical equivalence or neural prediction mechanisms. Effective interpretation requires deeper semantic modeling supported by cultural and cognitive information.

5. Proposed Hybrid Semantic Interpretation Framework

Based on the identified limitations, this study proposes a hybrid semantic interpretation framework designed to improve phraseological translation in AI systems. The proposed model combines neural processing mechanisms with linguocultural and cognitive analysis.

The first component of the model is the contextual embedding layer, which utilizes transformer-based architectures such as BERT to capture contextual semantic relations within discourse. This layer enables the system to distinguish literal usage from idiomatic usage through contextual analysis. The second component is the phraseological detection module, responsible for identifying idiomatic and fixed expressions before translation occurs. This stage prevents word-by-word processing and treats phraseological units as semantically integrated structures.

The third component is the cultural annotation layer, which incorporates linguocultural metadata into the translation process. This layer contains information about historical references, cultural symbolism, ethnographic associations, and pragmatic functions of phraseological units. The fourth component, semantic mapping, searches for culturally and semantically equivalent expressions in the target language rather than relying on direct lexical substitution. Finally, the adaptive translation generator produces the final output by integrating contextual, semantic, and cultural information.

The operational logic of the model follows several sequential stages. First, the input text is analyzed to detect phraseological expressions. Second, contextual semantic interpretation is performed through neural embeddings. Third, cultural mapping mechanisms identify culturally

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

appropriate equivalents. Fourth, semantic alignment ensures communicative equivalence between source and target expressions. The final stage generates a translation that preserves both semantic meaning and cultural functionality.

The proposed framework demonstrates how linguocultural knowledge can be integrated into AI translation systems to improve phraseological interpretation and reduce semantic distortion.

6. Discussion

The findings of this study confirm that contemporary AI translation systems still lack sufficient mechanisms for deep semantic and cultural interpretation. Although neural architectures achieve high levels of grammatical accuracy, they frequently fail in cases involving figurative language and culturally marked expressions. Phraseological units require more than statistical prediction because their meanings are closely connected with collective cultural knowledge and cognitive models.

The proposed hybrid framework addresses these limitations by combining neural semantic processing with linguocultural analysis. Integrating cultural annotation and frame semantics allows AI systems to move beyond literal translation and achieve greater semantic adequacy. This approach also contributes to the broader development of culturally adaptive natural language processing technologies.

Furthermore, the study demonstrates the importance of interdisciplinary integration between linguistics, artificial intelligence, cognitive science, and translation studies. Phraseological interpretation cannot be solved exclusively through computational methods; it requires linguistic and cultural expertise capable of modeling human semantic understanding. Therefore, future AI translation research should focus on creating multilingual phraseological databases, culturally enriched corpora, and context-sensitive evaluation metrics.

7. Conclusion

Phraseological units represent one of the most challenging areas for AI-based translation systems because of their semantic opacity, metaphorical structure, and cultural specificity. Their successful interpretation requires the integration of linguistic, cognitive, and ethnocultural dimensions into computational translation models.

This study demonstrates that linguocultural analysis significantly improves the semantic adequacy of phraseological translation and helps reduce pragmatic and cultural distortion. The proposed hybrid semantic interpretation framework offers an effective solution by combining contextual embeddings, phraseological detection, cultural annotation, and semantic mapping mechanisms. Comparative analysis of English and Uzbek phraseological units further reveals the importance of accounting for structural and cultural asymmetries in machine translation.

The research contributes to the advancement of context-aware and culturally adaptive AI translation systems and highlights the necessity of integrating cultural semantics into natural language processing. Future investigations should focus on multilingual phraseological corpora, AI training based on culturally annotated datasets, and the development of evaluation metrics that extend beyond traditional statistical measures such as BLEU.

References

1. YaxshimurotYaxshimurotovna, S. Y. (2026). INTEGRATED STUDY OF AUTOMATED TRANSLATION QUALITY AND PHRASEOLOGICAL EQUIVALENCE IN ENGLISH-UZBEK TRANSLATION. Shokh Articles Library, 1(1), 1044-1050.
2. ovna, 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

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

[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
6. https://www.academia.edu/129376429/CHALLENGES_AND_STRATEGIES_IN_TRANSLATING_ENGLISH_IDIOMS_INTO_UZBEK
7. PolyFrame at MWE-2026 AdMIRe 2: When Words Are Not Enough: Multimodal Idiom Disambiguation https://arxiv.org/html/2602.18652v2
8. Kurbonova, M. S. (2025). Linguoculturological Research of Phraseological Units with Several Somatic Components in English and Uzbek Languages. International Journal of Artificial Intelligence, 1(7), 411–421.
9. Isroilov, E. N. (2025). Linguocultural Features of Phraseological Units in English and Uzbek Languages. Journal of Multidisciplinary Sciences and Innovations, 4(3), 646–652.