JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS
VOLUME 5, ISSUE 02
MOPNTHLY JOURNALS
ISSN NUMBER: 2751-4390
IMPACT FACTOR: 9,08

SEMANTIC INTERPRETATION OF PHRASEOLOGICAL UNITS IN ARTIFICIAL INTELLIGENCE–BASED TRANSLATION SYSTEMS: CHALLENGES AND ENHANCEMENT STRATEGIES

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

Abstract. The rapid development of artificial intelligence–based machine translation systems has significantly improved cross-linguistic communication; however, linguistic errors remain a persistent challenge, particularly for morphologically rich and low-resource languages such as Uzbek. This study examines the theoretical foundations of identifying and correcting linguistic errors in neural machine translation (NMT) systems, with specific attention to morphological, lexical, semantic, syntactic, stylistic, and idiomatic inaccuracies. The agglutinative structure of Uzbek, combined with its rich phraseological inventory and flexible syntax, increases the likelihood of structural misinterpretation and semantic distortion in AI-generated translations.

Keywords: Artificial Intelligence (AI); Neural Machine Translation (NMT); Transformer architecture; linguistic error detection; Automatic Post-Editing (APE); Quality Estimation (QE); Reinforcement Learning from Human Feedback (RLHF); phraseological units; semantic accuracy; morphology-aware modeling; Uzbek language translation; low-resource languages.

The paper analyzes the limitations of Transformer-based architectures in handling non-compositional expressions and culturally bound phraseological units, emphasizing the need for context-sensitive and morphology-aware modeling. Furthermore, it evaluates contemporary error detection and correction mechanisms, including Quality Estimation (QE), Automatic Post-Editing (APE), and Reinforcement Learning from Human Feedback (RLHF). These approaches are examined as complementary strategies for improving translation adequacy and fluency without relying solely on reference-based evaluation metrics.

The findings suggest that a hybrid framework integrating idiom-aware training, annotated corpora development, semantic embedding enhancement, and human-in-the-loop optimization significantly enhances translation reliability. The study concludes that combining linguistic theory with advanced AI architectures is essential for achieving higher semantic precision and cultural adequacy in machine translation systems.

Introduction

The rapid expansion of digital communication has significantly increased the demand for high-quality machine translation systems. Artificial intelligence (AI)-driven translation technologies, particularly Neural Machine Translation (NMT) systems based on Transformer architectures, have demonstrated remarkable performance in general language translation tasks. However, phraseological units-idioms, fixed expressions, and culturally bound metaphorical constructions-remain one of the most complex linguistic phenomena for AI-based translation systems.

Phraseological units are characterized by semantic non-compositionality, structural stability, and strong cultural-connotative associations. Their meaning cannot be derived solely from the literal interpretation of their constituent elements. This poses substantial challenges for statistical and neural translation models, which rely primarily on probabilistic token prediction and large-scale parallel corpora.

Yaxshimurotovna (2025) emphasizes that phraseological units embody cultural memory and collective cognitive patterns, and their interpretation through artificial intelligence often results in semantic neutralization or distortion. Building upon this perspective, the present study examines the semantic challenges of translating phraseological units in AI-based systems and proposes enhancement strategies grounded in computational linguistics and cross-cultural semantics.

Literature Review

Phraseology has long been recognized as a distinct subsystem within linguistic theory. Phraseological units function as semantically integrated expressions whose overall meaning is not reducible to the sum of their parts. Their stability and idiomaticity complicate translation processes, particularly in structurally and culturally distant language pairs such as English and Uzbek.

In her study, Yaxshimurotovna (2025) analyzes the cultural-connotative features of phraseological units across languages and demonstrates that AI-based translation systems frequently fail to preserve:

- Cultural imagery embedded in idiomatic expressions,
- Pragmatic and emotional nuances,
- Metaphorical associations specific to a linguistic community,
- Context-sensitive semantic shifts.

Neural Machine Translation systems, particularly those based on the Transformer architecture, utilize attention mechanisms to model contextual dependencies. Despite these advances, such systems typically process phraseological units as sequences of individual tokens rather than holistic semantic units. Consequently, literal translations frequently occur, leading to semantic distortion.

Methodological challenges in AI-based translation of phraseological units

1. Non-compositional semantics

Phraseological units are inherently non-compositional. For example:

- “to spill the beans” → “to reveal a secret”
- Uzbek: “ko‘z ochib yumguncha” → “in the blink of an eye”

When NMT systems rely on token-level probability distributions, they often generate literal translations such as “pour the beans,” which distort the intended meaning. This reflects the model’s statistical bias toward compositional interpretation.

2. Cultural-connotative dimension

According to Yaxshimurotovna (2025), phraseological units encode national mentality and historical experience. For instance, the Uzbek expression “boshi ko‘kka yetmoq” (literally “one’s head reaches the sky”) expresses extreme happiness. Without cultural mapping, AI systems may translate it literally, thereby eliminating its emotional intensity.

AI models do not possess cultural awareness; they approximate meaning through pattern recognition. Therefore, culturally bound idioms remain vulnerable to semantic reduction.

3. Contextual disambiguation

Certain phraseological units function both literally and idiomatically depending on context. Transformer-based models incorporate contextual embeddings, yet performance declines when training data lacks sufficient idiomatic examples. This limitation is particularly evident in low-resource languages such as Uzbek, where annotated phraseological corpora remain scarce.

Enhancement strategies for AI-based translation systems

1. Development of annotated phraseological corpora

The creation of specialized parallel corpora with explicit idiom annotation is essential. Such corpora should include:

- Clear identification of idiomatic boundaries,
- Distinction between literal and figurative usage,
- Cultural and pragmatic annotation layers.

This enables idiom-aware training, allowing models to learn phraseological units as semantic wholes rather than independent tokens.

2. Enriched semantic embeddings

Contextual embedding models must be optimized to encode phraseological expressions holistically. Approaches may include:

- Multi-task learning frameworks integrating idiom detection tasks,
- Contrastive learning methods distinguishing literal and figurative usage,
- Cultural-semantic tagging systems embedded into training pipelines.

These strategies enhance the representational depth of phraseological units within vector space models.

3. Idiom-aware architectural modifications

Incorporating an idiom identification layer into transformer-based systems can significantly improve performance. This component would:

1. Detect potential idiomatic sequences,
2. Encode them as unified semantic vectors,
3. Generate culturally appropriate equivalents in the target language.

Such architectural refinement aligns computational processing with phraseological theory.

4. Reinforcement learning from human feedback (RLHF)

Automatic post-editing (APE) combined with Reinforcement Learning from Human Feedback can iteratively refine translation quality. Human translators provide corrections specifically targeting phraseological inaccuracies. These corrections serve as feedback signals, guiding model optimization toward culturally accurate outputs.

This hybrid human-AI interaction reduces semantic distortion and enhances pragmatic adequacy.

Proposed conceptual framework

An integrative framework for phraseological translation in AI systems should include:

1. Phraseological detection module,
2. Contextual-semantic encoding layer,
3. Cultural adaptation component,
4. Iterative Human-in-the-Loop Training Mechanism.

This model operationalizes the cultural-connotative principles articulated by Yaxshimurotovna (2025) within computational translation architectures.

Conclusion

Phraseological units represent one of the most challenging domains in AI-based machine translation. Their semantic opacity, cultural embeddedness, and contextual variability expose the limitations of purely statistical and neural approaches.

The findings indicate that effective translation of phraseological units requires:

- Dedicated annotated corpora,
- Idiom-aware training strategies,
- Enriched semantic embedding techniques,
- Human-guided reinforcement learning processes.

Integrating cultural-connotative linguistic theory with advanced neural architectures offers a promising pathway toward improving phraseological translation accuracy. As AI technologies continue to evolve, the inclusion of culturally informed semantic modeling will be essential for achieving translation systems capable of handling the full complexity of human language.

References

1. Yaxshimurotovna, S. Y. (2025). *Cultural-connotative features of phraseological units in different languages and their interpretation through artificial intelligence*. SHOKH LIBRARY, 1(13).
2. Shukurova, Y. Y. (2025). *Ai-baseddetectionandcorrectionoflinguisticerrorsin Machinetranslationsystemstheoreticalfoundations.* Journal of Multidisciplinary Sciences and Innovations, 4(10) P 1507 1511
3. Shukurova, Y. Y. (2025). *Applied translation strategies and challenges in rendering english phrasal verbs into uzbek.* Adaptive Learning Technologies 2(9). P 97 105.
4. Shukurova, Y. Y. (2025) *Phraseological units in translation: role and linguistic characteristics.* Journal of Multidisciplinary Sciences and Innovations: 4(11) P 953 954.
5. Zayniddinovna, T. N. (2021). The Image of the Eastern Ruler in the Works of Christopher Marlowe. *Central Asian Journal Of Social Sciences And History*, *2*(10), 10-14.
6. Zayniddinovna, T. N. (2022). The Problem of "A Strong Personality" in Shakespeare'Dramas: Richard III and Macbeth. *Middle European Scientific Bulletin*, *20*, 7-10.
7. Zaynitdinovna, T. N. (2022). Lyrical Dialogue in Shakespeare's Poems as a Reflection of Renaissance Anthropocentrism and a Strong Personality. *Middle European Scientific Bulletin*, *21*, 120-125.
8. Ташева, Н. З. (2022). КРИСТОФЕР МАРЛОУ (ЛИТЕРАТУРНЫЙ АНАЛИЗ): ТАМЕРЛЕН ВЕЛИКИЙ КАК ТИП ЛИЧНОСТИ ВОСТОЧНОГО ПРАВИТЕЛЯ. *Eurasian Journal of Academic Research*, *2*(2), 234-239.
9. Zayniddinovna, T. N. (2022). THE CHARACTER OF STRONG PERSONALITY ACCORDINGLY WITH EASTERN THEMATICS IN CHRISTOPHER MARLOWE'S PLAY" TAMBURLAINE THE GREAT". *International Journal Of Literature And Languages*, *2*(08), 9-14.
10. Zayniddinovna, T. N. (2022). STRONG PERSONALITY'S ARTISTIC AND POETIC REFLECTION IN THE RENAISSANCE EPOCH. *International Journal Of Literature And Languages*, *2*(11), 43-51.
11. nafisa Zayniddinovna, T. (2022). Lexico-Semantic Word Production as a Way of Forming Theater Terminology of the English Language. *American Journal of Social and Humanitarian Research*, *3*(10), 144-150.