

PLATFORMS AND LEXICAL RESOURCES IN DIGITAL PLATFORMS IN MACHINE TRANSLATION

Utayeva Iroda Baxodirovna
Samarqand davlat pedagogika instituti

Abstract: This article deals with the interpretation and translation of idiomatic and culturally marked expressions by contemporary neural machine translation systems, with particular reference to ChatGPT, DeepL, and Google Translate. The research focuses on the capacity of these systems to reconstruct cognitive, pragmatic, and culturally embedded meanings that lack direct cross-linguistic equivalents. Particular attention is given to semantic, pragmatic, cultural-contextual, and morphosyntactic errors and to the factors contributing to their occurrence. The findings reveal system-specific patterns in the processing of culturally loaded language, including different tendencies toward calquing, adaptation, explicitation, and loss of figurative meaning. The results contribute to current research on neural machine translation, translation equivalence, and cognitive-cultural mediation and provide practical implications for machine translation development, professional post-editing, CAT environments, and translator education.

Keywords: neural machine translation, culturally marked expressions, cultural context, cognitive interpretation, translation equivalence, translation strategies, ChatGPT, DeepL, Google Translate, post-editing, intercultural communication

Introduction. The development of artificial intelligence (AI) technologies has substantially transformed the field of specialized translation. In recent years, neural machine translation (NMT) systems have demonstrated considerable improvements in the processing and translation of general-domain texts. Nevertheless, the translation of legal documentation continues to represent a particularly complex area of machine-assisted linguistic processing. This complexity is primarily determined by the distinctive characteristics of legal discourse, the requirement for a high degree of terminological precision, and the necessity of maintaining legal equivalence between the source text and the target text. Contemporary neural machine translation systems, including Google Translate, DeepL, and ChatGPT, are increasingly employed by both professional translators and legal institutions. Current research indicates that neural network-based models can significantly enhance translation efficiency by reducing processing time and facilitating the translation of large volumes of specialized content. However, their application to legal translation is associated with a number of methodological, linguistic, and professional limitations. Inaccuracies in the translation of legal documents may result in semantic distortion, incorrect interpretation of legal

terminology, failure to preserve relevant legal norms, and potentially serious legal consequences. Consequently, the use of AI-based translation technologies in the legal domain requires systematic quality assessment, careful verification of terminological and contextual equivalence, and, where necessary, professional human post-editing to ensure the accuracy and legal validity of the target text. The aim of the study is to identify the distinctive features of the interpretation and translation of idiomatic and culturally marked expressions characteristic of the Russian-speaking cultural context by neural machine translation systems-ChatGPT, DeepL, and Google Translate-and to determine recurrent adaptation strategies and patterns of translation errors associated with the loss of cultural context and reduced semantic accuracy.

Literature review. The relevance of the present study is determined by the growing importance of language models in an era characterized by global communication and digital interaction, in which such technologies function not merely as instruments of interlingual transfer but increasingly as mechanisms of cultural mediation. The expanding application of language models in cross-linguistic communication, together with the integration of neural systems into professional translation environments, creates a need to examine the extent to which these models are capable of reconstructing the implicit cognitive and cultural meanings embedded in idioms and phraseological units. This issue becomes particularly significant in the context of increasing human-machine interaction, where language models perform not only structural and semantic transformations of texts but also mediate between distinct cultural codes. Investigating the mechanisms involved in the processing of idiomatic expressions makes it possible to distinguish between formal computational interpretation and cognitive reconstruction of meaning, while identifying those dimensions of culturally conditioned semantics that remain particularly vulnerable to machine processing.

The theoretical significance of the study lies in refining current understandings of the mechanisms through which idiomatic and culturally marked expressions are transferred in neural machine translation and in extending theoretical perspectives on the interaction between linguistic, pragmatic, and cultural dimensions in the computational processing of figurative language. The study contributes to a more precise understanding of how contemporary translation models process expressions with substantial cultural and conceptual content and supplements existing theories of translation equivalence with empirical evidence concerning the behavior of neural systems when direct interlingual correspondences are unavailable. The findings further clarify the role of model architecture and training-related factors in shaping translation strategies and provide a theoretical basis for the further development of cognitively oriented approaches within translation studies.

The practical significance of the study consists in the development of functional recommendations aimed at improving the accuracy and adequacy of rendering idiomatic and culturally conditioned expressions in neural machine translation systems. The identified error patterns and adaptation strategies may be applied to the development and training of machine translation models as well as to the integration of hybrid human-machine translation solutions into computer-assisted translation (CAT) environments. The findings may also support professional translators and editors by providing practical criteria for the post-editing of machine-generated translations and a methodological framework for evaluating texts containing culturally loaded linguistic elements. Furthermore, the research materials may be incorporated into educational programmes in translation studies, machine translation, and intercultural communication, thereby contributing to the development of students' critical competence in assessing the quality and reliability of neural machine translation.

Discussion. Within international translation studies, one of the most influential theoretical frameworks is E. Nida's concept of dynamic equivalence, which prioritizes the response of the target-language recipient and seeks to preserve the functional and communicative properties of the source expression [4]. M. Baker's multilevel model of equivalence [5], together with the discourse-oriented approach developed by B. Hatim and I. Mason [6], provides a methodological basis for conceptualizing idiom translation as a cognitive-interpretive process that requires the reconstruction of pragmatic meaning and culturally embedded context. The work of G. Lakoff and M. Johnson on conceptual metaphor theory [7] further demonstrates that phraseological meaning is structured by culturally conventionalized conceptual schemas. From this perspective, the translation of idiomatic expressions involves not merely lexical substitution but the reconstruction or functional reconfiguration of the underlying metaphorical model.

Recent research on machine translation (MT) indicates that, despite substantial advances in syntactic processing and lexical accuracy, neural systems continue to exhibit significant limitations when processing culturally marked and culturally embedded linguistic units [9; 10]. A particularly important body of research published between 2023 and 2025 has sought to systematically identify recurring error patterns in neural machine translation (NMT) of idiomatic constructions. C. Baziotis et al., for example, developed a method for the automatic evaluation of idiomaticity in translated output and identified a characteristic tendency of neural models to rely on literal translation strategies [11]. In their review, P. Naveen and P. Trojovský emphasize that some of the most persistent challenges for NMT arise from expressions whose interpretation requires the reconstruction of an underlying cultural scenario or conceptual frame [12]. E. Liu et al. demonstrate that retrieval-augmented approaches

can improve the performance of idiom-aware translation systems; nevertheless, a systematic discrepancy between literal renderings and functionally appropriate solutions remains evident [13].

Research focusing on large language models (LLMs) similarly demonstrates that the quality of idiomatic translation is strongly influenced by prompt design, contextual information, and contextual density. A. Castaldo and J. Monti investigate the effects of different prompting patterns on the translation of idiomatic expressions [14]. D. Azizov identifies persistent forms of cultural inadequacy in both NMT systems and LLMs when translating culture-specific references and figurative expressions [15]. Applied studies, including those conducted by N. Darwish et al. [16] and A. S. Aldelaa [17], further confirm that neural translation systems may produce substantial semantic and cultural distortions when processing idiomatic language. Collectively, these findings suggest that improvements in formal linguistic accuracy do not necessarily entail an equivalent capacity for cognitive-cultural interpretation.

The theoretical framework of the present study was based primarily on M. Baker's model of idiom translation strategies (1992) and E. Nida's theory of dynamic equivalence (1964). Drawing on these theoretical perspectives, three principal translation strategies observed in the output of neural systems were identified and subsequently illustrated with empirical examples extracted from the research corpus.

The first strategy, literal translation, was operationalized as the formal reproduction of the lexical components and surface structure of an idiomatic expression without adequate reconstruction of its figurative or culturally conditioned meaning. Such renderings frequently resulted in semantic and cultural distortions because they failed to preserve the idiomatic function of the source-language unit. Representative examples include «как снег на голову» → "like snow on the head," «вешать лапшу на уши» → "to hang noodles on someone's ears," and «дело пахнет керосином» → "the matter smells of kerosene." Although these translations reproduce the lexical constituents of the original expressions, they disrupt their figurative organization and fail to establish pragmatic equivalence in the target language.

The second strategy, functional adaptation (functional equivalence), involved selecting a target-language expression capable of producing a comparable communicative effect for the target recipient. This category included instances in which culturally distinct but functionally corresponding scenarios were successfully matched. For example, «дело пахнет керосином» was rendered as "something's fishy," while «в Тулу со своим самоваром» was translated as "to carry coals to Newcastle." Such solutions demonstrate the capacity of certain models to identify functional phraseological equivalents in the target language and thereby preserve the communicative intention and pragmatic force of the source expression despite differences in their lexical and metaphorical realization.

The third strategy, termed cognitive substitution, was manifested through the use of descriptive or explanatory constructions that conveyed the conceptual meaning of an idiom when no direct phraseological equivalent was selected. Examples include «душа болит» → “to feel deeply worried or sad” and «ком в горле» → “to feel unable to speak because of strong emotion.” This strategy enabled the models to preserve the semantic core of the source expression; however, it frequently resulted in the reduction or complete loss of its figurative imagery, metaphorical structure, and culture-specific associations.

Each of these translation strategies was evaluated according to three principal analytical parameters: correspondence to the communicative intention of the source expression, degree of cultural adequacy, and capacity to preserve the cognitive components of the underlying source image. This analytical framework made it possible to examine not only which translation strategy was selected by a particular system but also the linguistic, pragmatic, and contextual conditions under which neural models tended to favor literal, descriptive, or functionally equivalent solutions.

The quantitative error analysis further confirmed substantial differences in the distribution of translation errors across the systems under investigation. Of the 100 idiomatic expressions included in the corpus, ChatGPT produced erroneous or inadequate translations in 27 cases, compared with 46 cases for DeepL and 73 cases for Google Translate. Semantic errors accounted for approximately 44% of all identified failures and were predominantly associated with the literal reproduction of lexical components. Examples included «душа не на месте» → “the soul is not in place” and «свет не мил» → “the light is not nice.” These cases demonstrate that lexical correspondence at the surface level does not necessarily ensure the preservation of phraseological meaning.

Cultural errors constituted approximately 31% of the identified cases and were primarily attributable to the systems’ inability to recognize or adequately reconstruct culture-specific references, conventionalized images, and culturally embedded communicative meanings. For instance, DeepL rendered «ни пуха ни пера» as “not a feather or a hair,” thereby reproducing the lexical structure of the expression while failing to convey its conventional pragmatic function as a formula for wishing someone success.

Pragmatic errors, representing approximately 18% of the total, occurred when the communicative function, discourse value, or genre-specific character of the source expression was not adequately preserved. For example, the proverb «волков бояться - в лес не ходить» was rendered by DeepL as a descriptive or propositional construction, whereas ChatGPT provided the functional equivalent “nothing ventured, nothing gained.” The latter solution preserves the proverbial character and communicative function of the source unit rather than its literal lexical imagery.

Finally, morphosyntactic errors, which accounted for approximately 7% of the total error distribution, were observed predominantly in Google Translate output. These errors involved inappropriate word order, article omission, and structurally ill-formed or unnatural word combinations. Although quantitatively less frequent than semantic and cultural errors, such deviations reduced the grammatical naturalness and overall acceptability of the target-language output.

A comparative analysis of figurative-image preservation demonstrated that ChatGPT more frequently departed from literal reproduction in favor of functional equivalence and pragmatic adaptation. For instance, «дело пахнет керосином» was translated as “something’s fishy,” which preserves the pragmatic connotation of suspicion or an indication that something is wrong, despite replacing the source metaphor with a culturally conventional English expression. DeepL and Google Translate, by contrast, reproduced the original metaphor literally as “the matter smells of kerosene,” resulting in an expression that is formally related to the source text but pragmatically non-idiomatic in English.

A comparable pattern was observed in the translation of «ни пуха ни пера». ChatGPT replaced the source-language expression with the established English pragmatic equivalent “break a leg!”, thereby preserving its communicative function as a culturally conventional formula of encouragement or good luck. DeepL, in contrast, produced a literal rendering that failed to activate an equivalent cultural script in the target language. In several other cases, ChatGPT employed semantic neutralization, replacing the figurative expression with a non-idiomatic lexical unit. Thus, «с гулькин нос» was rendered as “tiny.” Although such a solution entails the loss of the source expression’s imagery and cultural specificity, it successfully preserves its core denotative meaning. DeepL and Google Translate were more likely to reproduce the expression literally, thereby retaining the formal image but producing a construction that remains opaque to an English-speaking recipient.

To provide a quantitative measure of translation equivalence, the study calculated an Index of Cultural Adequacy (ICA) integrating three analytical dimensions: semantic preservation, idiomaticity, and target-language naturalness. The mean ICA values were 0.48 for Google Translate, 0.67 for DeepL, and 0.81 for ChatGPT. The differences among the systems were found to be statistically significant ($p < 0.05$, Wilcoxon test). The most pronounced variation occurred in the category of culture-specific and culture-bound expressions, indicating that successful idiomatic translation depends not only on lexical-semantic correspondence but also on the system’s capacity for context-sensitive processing, pragmatic inference, and cognitive-cultural adaptation. These findings support the conclusion that the reconstruction of culturally embedded meaning constitutes a decisive factor in achieving functional equivalence when idiomatic and phraseological units are processed by contemporary neural translation systems.

Table 1. Lexicographic Resources for Digital Translation and Machine Translation Platforms

No.	Russian-language resources	No.	English-language resources
1	SlovoEd for Pocket PC; SlovoEd Compact	1	Advanced English Dictionary and Thesaurus
2	SlovoEd Deluxe, Classic, Compact	2	Merriam-Webster Unabridged Dictionary
3	PROMT Electronic Dictionary	3	Lexigoo
4	Compact English-Russian Dictionary	4	WordBook Dictionary & Thesaurus
5	ABBYY Lingvo x6	5	MSDictionaries
6	Oxford Russian Dictionary for KIT	6	Oxford American Desk Dictionary & Thesaurus
7	Haali Reader: Dictionaries	7	PocketLingo American Heritage College Dictionary
8	Context English-Russian Dictionary	8	Concise Oxford Thesaurus
		9	Word Seeker

The data presented in Table 1 demonstrate the diversity of lexicographic resources available for digital and machine-assisted translation. The identified resources can be classified into several functional categories, including bilingual translation dictionaries, monolingual explanatory dictionaries, thesauri, and specialized reference resources. Such differentiation indicates that digital lexicographic support extends beyond direct lexical equivalence and provides translators with semantic, terminological, contextual, and stylistic information required for accurate translation. The Russian-language group is predominantly represented by bilingual English-Russian and Russian-English resources, such as SlovoEd, PROMT Electronic Dictionary, ABBYY Lingvo x6, and Oxford Russian Dictionary. Their primary function is to facilitate the identification of translation equivalents and provide lexical support during bilingual text processing. The availability of mobile and compact versions also reflects the increasing accessibility and technological flexibility of electronic lexicography in professional translation practice.

In contrast, the English-language resources are largely monolingual explanatory and thesaurus-based dictionaries, including Advanced English Dictionary and Thesaurus, Merriam-Webster Unabridged Dictionary, Oxford American Desk Dictionary & Thesaurus, and Concise Oxford Thesaurus. These resources are particularly important for examining definitions, semantic distinctions, synonymic relations, collocations, and contextual usage. From a translation research perspective,

such tools contribute to the selection of contextually appropriate equivalents rather than relying exclusively on direct dictionary correspondences.

Overall, the comparative analysis suggests that effective digital translation requires the integrated use of bilingual, monolingual, and specialized lexicographic resources. Bilingual dictionaries support cross-linguistic equivalence, whereas explanatory dictionaries and thesauri facilitate semantic disambiguation and contextual adaptation. Therefore, combining different types of electronic lexicographic resources can enhance terminological accuracy, semantic adequacy, lexical variation, and overall translation quality, particularly when machine-generated translations require professional verification and post-editing.

Conclusion. Thus, study have a dual orientation, encompassing both professional translation practice and translation pedagogy. In professional contexts, a hybrid workflow appears to be the most effective approach: a neural language system produces an initial draft with an acceptable level of lexical and grammatical accuracy, while the human translator undertakes cognitive and cultural revision to restore idiomatic naturalness, pragmatic meaning, and sociocultural implications. The integration of generative AI plugins into CAT environments further demonstrates the effectiveness of this collaborative model. The model's capacity to infer discourse-level meaning during the initial solution-generation stage, combined with human verification and refinement during post-editing, contributes to measurable improvements in both qualitative and quantitative translation outcomes.

References

1. Макаревич, И. И. Отражение структуры и функционирования отраслевой терминосистемы дискурса цифровой трансформации в переводном специализированном словаре-справочнике / И. И. Макаревич // Новый мир. Новый язык. Новое мышление : сб. материалов VI Междунар. науч.-практ. конф. (филология, педагогика и межкультурная коммуникация), Москва, 3 февр. 2023 г. / Дипломат. акад. М-ва иностранных дел Рос. Федерации. - М., 2023. - С. 701-705.
2. Швейцер А. Д. Теория перевода: статус, проблемы, аспекты. - М.: Наука, 1988.
3. Рецкер Ю. И. Теория перевода и переводческая практика. - М.: Международные отношения, 1974.
4. Найда Ю. Прагматическая и лингвистическая эквивалентность. - М.: Прогресс, 1980.
5. Бейкер М. Перевод и лексика: учебное пособие In Other Words. - Лондон: Routledge, 1992.
6. Хэтим Б., Мейсон И. Дискурс и перевод. - Лондон: Longman, 1990.

7. Лакофф Дж., Джонсон М. Метафоры, которыми мы живём. - Чикаго: University of Chicago Press, 1980.
8. Бархударов Л. С. Язык и перевод. - М.: Международные отношения, 1975.
9. Toral A., Way A. What Level of Quality Can Neural Machine Translation Attain? // Machine Translation. - 2018. - Vol. 32, No. 2. - P. 179-193.
10. Freitag M., Bapna A., Firat O. et al. Experts, Errors, and Context in Neural Machine Translation // Transactions of the Association for Computational Linguistics. - 2022. - Vol. 10. - P. 20-37.
11. Базиотис К., Матор П., Хаслер Э. Automatic Evaluation and Analysis of Idioms in Neural Machine Translation // Proceedings of EACL. - 2023.
12. Навин П., Троёвский П. Overview and Challenges of Machine Translation for Contextually Appropriate Translations // iScience. - 2024. - Vol. 27, No. 10.
13. Лю Э., Чодхари А., Ньюбиг Г. Crossing the Threshold: Idiomatic Machine Translation through Retrieval Augmentation and Loss Weighting // Proceedings of EMNLP. - 2023.
14. Кастальдо А., Монти Д. Prompting Large Language Models for Idiomatic Translation // Creative-text Translation and Technology Workshop. - 2024.
15. Азизов Д. From Idioms to Algorithms: Translating Culture-Specific Expressions in AI Systems // IRE Journals. - 2024. - Vol. 7, No. 10. - P. 543-551.
16. Дарвиш Н. и др. A Reception Study of AI-Translated Idioms and Proverbs Between Arabic and English // Research Journal in Advanced Humanities. - 2025. - Vol. 6, No. 3.
17. Альделаа А. С. Investigating Problems Related to the Translation of Idiomatic Expressions in the Arabic Novels Using Neural Machine Translation // Theory and Practice in Language Studies. - 2024. - Vol. 14, No. 1. - P. 71-78.
18. Сарсенбаева З. Analysis of images and symbols in english non-realistic works //Ренессанс в парадигме новаций образования и технологий в XXI веке. - 2023. - Т. 1. - №. 1. - С. 229-232.
19. Сарсенбаева З. Модернизм в узбекской литературе и интерпретация образов //Зарубежная лингвистика и лингводидактика. - 2024. - Т. 2. - №. 1. - С. 193-199.
20. Zoya S. LITERARY ANALYSIS OF DAVID MITCHELL'S WORKS IN ENGLISH LITERATURE //Talqin va tadqiqotlar ilmiy-uslubiy jurnali. - 2024. - Т. 2. - №. 41. - С. 11-19.
21. Zoya S. LITERARY ANALYSIS OF ULUGBEK HAMDAMOV'S WORKS IN UZBEK LITERATURE //Talqin va tadqiqotlar ilmiy-uslubiy jurnali. - 2024. - Т. 2. - №. 41. - С. 20-25.

22. Sarsenbaeva Z. MODERN APPROACHES TO TEACHING LITERARY TERMS IN COMPARATIVE LITERATURE //Mental Enlightenment Scientific-Methodological Journal. - 2026. - T. 7. - №. 02. - C. 198-209.

23. Belgibayeva G., Sarsenbaeva Z., Zhumagulova K. MAG ‘JAN JUMABAYEV IJODIDA OZODLIK G ‘OYASI //Ijtimoiy-gumanitar sohada ilmiy-innovatsion tadqiqotlar. - 2026. - T. 3. - №. 1. - C. 159-165.

24. Sarsenbayeva Z., Zillolova G. COMPARATIVE ANALYSIS OF TWO TRANSLATED VERSIONS OF A LITERARY TEXT //Ijtimoiy-gumanitar sohada ilmiy-innovatsion tadqiqotlar. - 2025. - T. 2. - №. 4. - C. 183-188.

25. Sarsenbayeva Z., Zillolova G. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДВУХ ПЕРЕВОДНЫХ ВЕРСИЙ ХУДОЖЕСТВЕННОГО ТЕКСТА //Scientific and innovative research in the social and humanitarian sphere. - 2025. - T. 2. - №. 4. - C. 183-188.