DOI: https://doi.org/10.37547/mesmj-V7-I9-39 Pages: 414-423

MENTAL ENLIGHTENMENT SCIENTIFIC – METHODOLOGICAL JOURNAL



MENTAL ENLIGHTENMENT SCIENTIFIC – METHODOLOGICAL JOURNAL

http://mentaljournal-jspu.uz/index.php/mesmj/index



NEURAL MACHINE TRANSLATION OF RUSSIAN PHRASEOLOGICAL UNITS INTO UZBEK: A LINGUOCULTURAL ANALYSIS OF TRANSLATION CHALLENGES AND METHODS FOR OVERCOMING THEM

Kamila Azimovna Nabieva
Independent Researcher
Chirchik State Pedagogical University
Chirchik, Uzbekistan

ABOUT ARTICLE

Key words: neural machine translation; phraseology; Russian language; Uzbek language; artificial intelligence; equivalence; linguoculture; teaching methodology.

Received: 11.11.26
Accepted: 12.11.26
Published: 13.11.26

Abstract: The article examines linguistic and methodological aspects of neural machine translation (NMT) of Russian phraseological units (PUs) into Uzbek. The relevance of the study is determined by the growing use of NMT systems and large language models in language education, while idiomatic expressions remain an area of increased risk for such systems. The aim is to identify typical sources of translation difficulties and to substantiate a method that makes it possible to use neural tools in teaching Russian without loss of linguistic and cultural accuracy. Descriptive, contrastive, translation-studies and linguocultural methods were applied. On the basis of a comparison of Russian PUs with their Uzbek counterparts, a typology of correspondences (full equivalents, partial equivalents, non-equivalent units) is proposed, and four groups of typical errors are described: literalism, levelling of the expressive-stylistic component, wrong choice between literal and figurative meaning, and failure to recognise transformed units. A three-stage "prediction — comparison — editing" model for classroom work is substantiated. The results can be used in teaching Russian, in translation practice and in designing didactic materials.

Introduction. The rapid development of neural machine translation systems and large language models has significantly transformed the practice of interlingual communication and language education. Philology students, translators, and teachers increasingly turn to neural network-based services as everyday tools for translating texts, checking their own translations, and searching for equivalents. The Transformer architecture [17], which replaced recurrent models with attention mechanisms [11], and the large language models that followed [13] have made it possible to achieve translation quality that is acceptable for everyday and educational purposes across many language pairs and text genres.

Nevertheless, there remains a category of linguistic units for which translation quality noticeably decreases: idiomatic expressions, particularly phraseological units. In the Russian–Uzbek language pair, the situation is further complicated by structural differences. Russian is a fusional language, whereas Uzbek is an agglutinative Turkic language with a different word order and different ways of expressing grammatical meanings. Furthermore, the volume of parallel data available for this language pair is generally smaller than that available for language pairs involving English. These factors make phraseology a useful testing ground for evaluating the capabilities and limitations of neural systems.

A phraseological unit is a fixed, reproducible combination of words whose meaning cannot be derived from the meanings of its individual components [3; 10; 8]. It is precisely this semantic non-compositionality that creates difficulties for models inclined to process expressions component by component. Studies indicate that neural models often process idioms too compositionally, that is, literally [15]. This characteristic poses a double risk for language learners: an erroneous translation, once accepted as correct, may become established in their speech and translation practices.

The aim of this article is to identify the typical sources of difficulties in the neural machine translation of Russian phraseological units into Uzbek and to substantiate a methodology that makes it possible to transform these difficulties into a didactic resource. To achieve this aim, the following objectives are addressed: 1) to characterize phraseological units as objects of machine processing; 2) to propose a typology of correspondences between Russian phraseological units and their Uzbek equivalents; 3) to describe typical translation errors; and 4) to substantiate a model of classroom activities involving neural machine translation systems.

The theoretical framework of the study is based on research in Russian phraseology [1; 3; 8; 10], translation theory and practice [4; 7; 12], linguoculturology [2; 5; 9], and neural machine translation [11; 13; 15; 17]. The study employs descriptive and comparative methods,

elements of translation studies and linguocultural analysis, as well as the method of modelling educational activities.

The illustrative material consists of Russian phraseological units selected from the Phraseological Dictionary of the Russian Language [6], representing different degrees of figurativeness, stylistic colouring, and types of correspondence in Uzbek. The Uzbek equivalents were verified against the Explanatory Phraseological Dictionary of the Uzbek Language [16]. The literal translations presented below (marked with an asterisk "*") are illustrative models: they have been constructed to demonstrate types of translation errors and do not represent the results of statistical testing of any specific system.

In Russian linguistics, phraseological units are defined through a combination of characteristic features, including structural stability, reproducibility, semantic integrity, semantic reinterpretation, and figurativeness [10; 8]. Based on the degree of semantic cohesion, V. V. Vinogradov distinguished phraseological fusions, phraseological unities, and phraseological combinations [3]. This classification is also important for translation: the greater the semantic cohesion of an expression, the less likely it is that its meaning can be conveyed through the translation of its individual components.

Translation studies have identified several principal methods of rendering phraseological units: selecting a phraseological equivalent, selecting a phraseological analogue, calquing, and descriptive translation [4; 7]. M. Baker identifies strategies that are essentially similar: using an idiom with a similar meaning and form, an idiom with a similar meaning but a different form, paraphrasing, and omission [12]. A qualified translator selects a strategy based on the meaning of the unit, its stylistic colouring, and the context, whereas a neural machine translation system makes its choice statistically, relying on patterns identified in training data.

Two practical conclusions follow from this. First, the probability of accurate translation is generally higher for frequent phraseological units that are widely represented in parallel texts and lower for rare, obsolete, and contextually modified expressions. Second, the system does not inherently "know" that it is dealing with an idiom: it operates on the basis of contextual sequence probabilities rather than the dictionary status of a linguistic unit. Therefore, teachers need a methodology in which students act not as passive consumers of machine translation but as its critical editors.

A comparison of Russian phraseological units with their Uzbek counterparts makes it possible to identify three types of correspondence (Table 1). Full equivalents coincide in meaning, imagery, and stylistic colouring. They arise from shared cultural experience, borrowing, calquing, or the typological similarity of images. Partial equivalents convey the

same meaning but are based on different images. Non-equivalent phraseological units have no established figurative counterpart in Uzbek and are therefore rendered descriptively or through the selection of a neutral lexical item.

Table 1. Types of Correspondence between Russian Phraseological Units and Their Uzbek Equivalents

Russian phraseological unit	Meaning	Uzbek equivalent	Type of correspondence
делать из мухи слона	to exaggerate	pashshadan fil yasamoq	Full equivalent
ловить рыбу в мутной воде	to take advantage of confusion or disorder	loyqa suvda baliq tutmoq	Full equivalent
семь раз отмерь, один раз отрежь	act carefully and think before acting	yetti o‘lchab, bir kes	Full equivalent
золотые руки	a highly skilled person	qo‘li gul	Partial equivalent (different image)
бить баклуши	to idle; to do nothing	qo‘l qovushtirib o‘tirmoq; bekorchilik qilmoq	Partial equivalent (different image)
как снег на голову	completely unexpectedly	to‘satdan; kutilmaganda	Non-equivalent phraseological unit (descriptive translation)
ни свет ни заря	very early in the morning	tong sahardan; erta tongdan	Non-equivalent phraseological unit (descriptive translation)

The linguocultural significance of partial equivalents is particularly evident. The image of “golden hands” in Russian is associated with the value of craftsmanship and skill, whereas the Uzbek idiom qo‘li gul is based on the image of a flower, reflecting a different aesthetic perception of a skilful person. In the expression бить баклуши, the component баклуши (wooden blanks used for making wooden utensils) has long fallen out of independent everyday use, making the image opaque even to native speakers of Russian. By contrast, the Uzbek expression qo‘l qovushtirib o‘tirmoq (“to sit with one's hands folded”) is based on an entirely different motivation [8; 5]. Such pairs are particularly valuable in educational activities because they encourage students to understand that an idiom is not merely a combination of words but a culturally conditioned sign.

The nature of phraseological correspondences is largely determined by the sources of figurativeness. Russian phraseological expressions originate in everyday life, crafts, folk customs, folklore, and biblical and classical traditions. Uzbek phraseology, in turn, is closely associated with agricultural and artisanal lifestyles, as well as with cultural perceptions of home, hospitality, and kinship. In general, the greatest number of correspondences can be observed among somatic phraseological units based on universal human bodily experience (“eyes,” “hands,” and “heart” – ko‘z, qo‘l, yurak) and among expressions of international origin,

whereas differences are more frequently associated with everyday and historical realia. From an educational perspective, this means that the selection of teaching materials should begin with groups in which students can rely on their native language, gradually progressing to expressions containing culturally specific imagery, where the risk of literal translation is highest [2; 9].

In addition to lexical and figurative differences, grammatical factors also exert a significant influence. In Russian, many phraseological units have a verbal structure consisting of a variable verb and dependent words in fixed case forms. In Uzbek, however, a considerable proportion of verbal phraseological units follow an analytical pattern consisting of a nominal component and an auxiliary verb (for example, ko‘z-ko‘z qilmoq – “to show off; to display something ostentatiously”). Word order also differs: Uzbek is characterized by the placement of the predicate at the end of the sentence, whereas Russian has a relatively free word order. During translation, the components of a phraseological unit inevitably undergo reorganization, meaning that the system must not only recognize the expression but also reconstruct it within a different syntactic framework. The greater the differences between the two structural models, the higher the probability of violating phraseological integrity.

An analysis of the difficulties encountered in translating Russian phraseological units into Uzbek makes it possible to identify four groups of errors.

Literal Translation. The phraseological unit is translated component by component, and its imagery is transferred without preserving its actual meaning. For example, the model translation “baklusha urmoq” for the Russian expression «бить баклуши» has neither figurative nor communicative meaning in Uzbek. The risk is particularly high for expressions containing obsolete components, since the system cannot identify familiar lexical patterns to support an appropriate translation.

Neutralization of the Expressive and Stylistic Component. The meaning is conveyed correctly, but its expressive effect is lost. For example, «ни свет ни заря» is reduced to the neutral Uzbek word “erta” (“early”), while «как снег на голову» is translated as “to‘satdan” (“suddenly”), without preserving the evaluative connotation of suddenness and unpleasant surprise. This represents a significant loss in literary or journalistic texts, since conversational imagery serves as a vehicle for the author's evaluation.

Incorrect Selection between Literal and Figurative Meanings. A number of phraseological units are homonymous with free word combinations. For instance, «сесть в лужу» can refer either to the literal action of sitting in a puddle or to “finding oneself in an awkward or embarrassing situation.” Selecting the correct meaning is possible only through an

analysis of the broader context. In a short or ambiguous fragment, the system is highly likely to select the more frequent literal interpretation.

Failure to Recognize Modified Phraseological Units. In everyday speech and literary texts, phraseological units frequently undergo occasional modifications through component substitution, expansion, or the deliberate exploitation of their literal meanings. For readers, such transformations create an aesthetic effect, whereas for machine translation systems, they constitute "noise" that disrupts familiar sequences. As a result, the expression is either translated literally or remains entirely unrecognized.

Table 2. Types of Neural Machine Translation Errors in Phraseological Units and Methods of Correction

Error type	Illustrative example	Method of identification and correction
Literal translation	«бить баклуши» → "baklusha urmoq"	Consulting a dictionary; checking whether the expression is conventionally used in Uzbek
Neutralization of expressiveness	«ни свет ни заря» → neutral "erta"	Comparing the stylistic labels of the source expression and its translation; compensating through lexical means
Incorrect selection of meaning	«сесть в лужу» → "ko'lmakka o'tirmoq" (literal interpretation)	Analysing the extended context; checking the translation by substituting the explanatory meaning
Failure to recognize a modified phraseological unit	«ловить рыбу в мутных водах социальных сетей» → component-by-component translation	Reconstructing the original form of the phraseological unit; consulting its dictionary form

It is practically important to note that the errors identified above are not always obvious to learners: the translation may remain grammatically correct and appear fluent. The fluency of the text creates an illusion of accuracy; therefore, the result must be evaluated with reference to dictionaries and linguistic reflection rather than merely relying on a subjective "sense of naturalness."

Taking these characteristics into account, a three-stage model of classroom and independent work is proposed for students enrolled in philological programmes. Its central idea is to shift the learner's role from that of a consumer to that of a critical editor of machine translation, which is consistent with the contemporary understanding of mediation as a form of language activity [14].

Stage 1. Prediction. Before using a machine translation system, students independently identify phraseological units in the given text, determine their meanings with the help of a dictionary [6], produce their own translations, and make written predictions about the difficulties that a machine translation system may encounter. This stage develops linguistic reflection and prevents thoughtless copying.

Stage 2. Comparison. Students receive translations generated by one or more systems, compare them with their own versions and the dictionary equivalents [16], and classify the discrepancies according to the four types of errors described above. The results are recorded in a comparative table. Comparing several versions develops critical thinking and demonstrates the variability of machine-generated translation solutions.

Stage 3. Editing. Students post-edit the translation, justify their choice of strategy (equivalent, analogue, or descriptive translation), create their own contexts illustrating the use of the phraseological units in Russian, and provide a brief linguocultural commentary on the imagery underlying each expression. The final product is an edited translation accompanied by a reasoned explanation of the decisions made.

Assessment should be based on four criteria: recognition of phraseological units and identification of their meanings; accuracy in selecting an equivalent or translation strategy; stylistic adequacy; and the validity of the accompanying commentary. Academic integrity is also an essential requirement: students must indicate which fragments were generated with the assistance of a neural network-based tool and which were produced independently.

Let us consider an example of a classroom assignment: “Всю неделю Рустам бил баклуши, и экзамен стал для него как снег на голову” (“Rustam had been idling all week, and the exam came as a complete surprise to him”). At the first stage, the student identifies two phraseological units, determines their meanings (“to idle” and “completely unexpectedly”), and predicts that the first expression is particularly susceptible to literal translation because of the opaque component «баклуши», while the second is likely to lose its expressive force. At the second stage, the student compares the machine-generated translation with the dictionary equivalents (qo‘l qovushtirib o‘tirmoq, bekorchilik qilmoq; to‘satdan, kutilmaganda) and classifies the identified discrepancies according to the corresponding error types. At the third stage, the student selects an Uzbek phraseological equivalent for the first expression and compensates for the loss of imagery in the second through lexical means, strengthening the connotation of unexpectedness and the unpleasant nature of the situation. The student also explains the cultural motivations underlying the Russian and Uzbek images.

Various types of exercises can be used to consolidate the acquired knowledge: “find and classify the error” in an existing machine translation; “reconstruct the original phraseological unit” based on its translation; “modify and test,” in which students create an occasional transformation of a phraseological unit and analyse the system's response; and “construct a context,” aimed at ensuring the unambiguous identification of the idiomatic meaning. Such activities develop not only students' phraseological competence but also their ability to

formulate prompts for neural systems that take into account the stylistic and pragmatic parameters of a text.

The proposed typology and model are not intended to be exhaustive. The examples provided are illustrative, while the conclusions regarding the tendency of neural machine translation systems towards literal translation are based on published research [15] and require further verification through a dedicated experiment involving quantitative analysis of results obtained from a representative sample of phraseological units and taking into account the versions of systems, which are subject to rapid change. One promising direction for future research is the development of a small Russian–Uzbek parallel corpus of phraseological units with expert annotation of correspondence types. Such a corpus could serve simultaneously as an educational resource and a testing dataset.

It is also necessary to consider the risk of excessive reliance on machine translation. Working with idiomatic expressions demonstrates that neural tools are useful as a source of alternative translation options but should not be regarded as the ultimate authority on translation accuracy. Under these conditions, the role of the teacher does not diminish; rather, it shifts towards organizing and guiding critical evaluation.

The practical significance of the study lies in the possibility of applying its findings in courses on the Russian language, translation theory and practice, and linguoculturology. The proposed model does not require specialized software: it can be implemented using publicly available online services and printed or electronic dictionaries, making it applicable in higher education institutions with varying levels of technological infrastructure. The combination of dictionary-based work, comparative analysis, and machine-generated text editing enables students to develop skills that are in demand in the professional activities of philologists and translators: evaluating translation quality, justifying translation decisions, and taking into account the cultural specificity of linguistic units.

The analysis conducted demonstrates that rendering Russian phraseological units into Uzbek through neural machine translation remains a challenging task due to the holistic, figurative, and culturally conditioned nature of phraseological units. A typology of correspondences has been proposed, comprising full equivalents, partial equivalents, and non-equivalent phraseological units. Four groups of typical translation errors have also been identified: literal translation, neutralization of the expressive and stylistic component, incorrect selection between literal and figurative meanings, and failure to recognize modified phraseological units.

The proposed “prediction–comparison–editing” model provides a framework for using neural machine translation systems in Russian language teaching as a means of developing linguistic reflection, linguocultural competence, and post-editing skills. Further research should focus on the experimental validation of the proposed model and the development of a parallel corpus of Russian and Uzbek phraseological units.

References:

[1]. Баранов А. Н., Добровольский Д. О. Аспекты теории фразеологии. М.: Знак, 2008.

[2]. Верещагин Е. М., Костомаров В. Г. Язык и культура. М.: Русский язык, 1990.

[3]. Виноградов В. В. Об основных типах фразеологических единиц в русском языке // Виноградов В. В. Избранные труды. Лексикология и лексикография. М.: Наука, 1977.

[4]. Комиссаров В. Н. Теория перевода (лингвистические аспекты). М.: Высшая школа, 1990.

[5]. Маслова В. А. Лингвокультурология. М.: Академия, 2001.

[6]. Фразеологический словарь русского языка / под ред. А. И. Молоткова. 4-е изд. М.: Русский язык, 1986.

[7]. Рецкер Я. И. Теория перевода и переводческая практика. М.: Международные отношения, 1974.

[8]. Телия В. Н. Русская фразеология: семантический, прагматический и лингвокультурологический аспекты. М.: Школа «Языки русской культуры», 1996.

[9]. Тер-Минасова С. Г. Язык и межкультурная коммуникация. М.: Слово, 2000.

[10]. Шанский Н. М. Фразеология современного русского языка. М.: Высшая школа, 1985.

[11]. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate // International Conference on Learning Representations (ICLR). 2015. arXiv:1409.0473.

[12]. Baker M. In Other Words: A Coursebook on Translation. London; New York: Routledge, 1992.

[13]. Brown T. B. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 1877–1901.

[14]. Council of Europe. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume. Strasbourg: Council of Europe Publishing, 2020.

[15]. Dankers V., Lucas C., Titov I. Can Transformer be Too Compositional? Analysing Idiom Processing in Neural Machine Translation // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin, 2022. P. 3608–3626.

[16]. Rahmatullayev Sh. O‘zbek tilining izohli frazeologik lug‘ati. Toshkent: O‘qituvchi, 1978.

[17]. Vaswani A. et al. Attention Is All You Need // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 5998–6008.