ISSN 2788-7979 (online)

IRSTI: 16.31.21 DOI: 10.59102/kufil/2026/iss2pp69-82

L. Zadorozhnaya[1], M. Yeskendirova[1], D. Popič[2]
[1]L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan
[2]University of Ljubljana, Ljubljana, 1000, Slovenia

CLASSIFICATION OF TRANSFORMATION-INDUCED TRANSLATION ERRORS: THE CASE OF COMPUTER-ASSISTED TRANSLATION TOOLS

The present paper investigates the characteristics of computer-assisted translation tools from both theoretical and practical perspectives. It addresses current issues related to the development of new typologies and classifications that may be useful in modern computer-assisted translation training and in solving applied tasks arising during the development of digital translation products. Frequency-based errors occurring in machine translation serve as the basis for this study. The authors present the results of a comparative analysis of texts translated into English using digital translation tools.

The classification of transformation-induced translation errors is based on three types of translation transformations: lexical, grammatical, and lexico-grammatical. The analysis identified eight subgroups of frequent machine translation errors within these categories. Based on the extralinguistic analysis, the study proposes the main factors contributing to the occurrence of such errors.

The findings indicate that machine translation errors may be associated with the use of computer-assisted translation tools without adherence to established standards, particularly regarding variations in the transliteration of proper names in Kazakhstani scientific literature. The study contributes to a better understanding of the limitations and challenges of computer-assisted translation tools and provides recommendations for their effective use in both translation practice and research activities.

Key words: computer-assisted translation, translation transformations, translation errors, transliteration, digitalization.

MAIN PROVISIONS

It is becoming increasingly evident that digitalization is a trend that is present in almost every industry in modern Kazakhstan. A range of officially adopted state programs concerning language policy and artificial intelligence have been adopted recently. For example, the Concept for Artificial Intelligence Development for 2024–2029 was adopted [1]. The Concept for the Development of Language Policy in the Republic of Kazakhstan for 2023–2029, in addition to numerous other valuable insights, repeatedly underscores the imperative to advance linguistic and translation technologies, as well as corpus linguistics [2]. In contemporary context, translation practice invariably entails the utilization of diverse technological tools. Consequently, these computer-assisted tools have become the focal point of both applied and theoretical research endeavours.

It is important to note that all contemporary tools are based on the classical theory of translation, and consequently, their existence is contingent on this theoretical framework. It is evident that the field of traditional translation studies is founded upon a range of typologies and classifications that are intrinsic to the processes of translation and interpretation. The foundational principles of the linguistic aspects of contemporary machine translation have been established in seminal works on translation transformations by V.N. Komissarov, L.S. Barkhudarov, E. Nida, P. Newmark and M. Baker. Then N. K. Garbovsky proposed aspects of the typology of translation errors, and Kazakhstani researchers such as S. K. Satenova, D. M. Akizhanova and N. Zhumay made important contributions to this area, as did many other linguists researching translation specifics.

ISSN 2788-7979 (online)

It has been suggested that the incorporation of typical errors that arise during translation or the learning of a foreign language could potentially enhance the effectiveness of tutorials designed to train professionals in these fields. This approach would involve forecasting potential gaps in their skills and then focusing on addressing these areas. The efficacy of this approach was demonstrated using the Learner Corpus, which was first designed in 1990 by Sylviane Granger and given the name "International Corpus of Learner English". By the start of the 20th century, learner corpora that make use of the most typical errors for different types of research had appeared in many top education institutions, including Cambridge University and the Hong Kong University of Science and Technology, amongst others. It is understood that this approach is also becoming increasingly popular in Kazakhstan. In 2024, the Institute of Linguistics named after Akhmet Baitursynuly and Oxford University launched a joint project named "A learning corpus for learning the Kazakh language for English speakers (Anglophones)", which is aimed at English-speaking students learning Kazakh. It is therefore conceivable that the findings of the present study may serve to substantiate its scientific significance.

INTRODUCTION

One of the earliest mentions of computer-assisted translation tools in scientific literature came in 1983 from Alan K. Melby, a researcher in the Linguistics Department at Brigham Young University [3, 174]. The terms "mechanical translation" and later "machine translation" have been recognized and employed by researchers since the mid-20th century [4, 162].

It is imperative to comprehend the distinctions between these two terms that are currently in active use: machine translation and computer-assisted translation. As W. John Hutchins states, the term "machine translation" refers to computerized systems responsible for producing translations with or without human assistance. However, according to the author, computer-assisted translation also involves supporting translators by providing access to remote terminology databases, online dictionaries, and other useful tools, which do not necessarily offer automatic machine translation of entire texts [5, 431].

Although modern digital trends are only just being researched, more descriptive ideas and categorization are needed to increase their effectiveness and create new, efficient technological tools. This is also influenced by the fact that, in the Republic of Kazakhstan, computer-assisted translation is connected with the creation of new Kazakh-language technologies, which are currently gaining popularity among learners and users at home and abroad. For instance, J.A. Tukebayeva's research focuses on Russian-Kazakh translations, and she discusses the direct connection between the development of artificial intelligence and the quality of machine translation. It is asserted by the author that machine translation still faces significant challenges, including the interpretation of culturally specific texts [6, 84].

Researchers N. Zhumay and D.M. Akizhanova also conducted research into the history and features of machine translation, concluding that it is an important competence of modern translators. The authors describe some typical errors that occur in machine translations when using services such as Yandex Translate, DeepL and Google Translate. These include syntactic interference from the source language into the target language, as well as lexical, semantic and morphological errors [7, 95]. The significance of incorporating philosophical, cognitive and linguistic dimensions into the advancement of artificial intelligence technologies has been emphasized by M. Shengelbay, D. Alkebaeva and Zh. Talaspaeva. These scholars underscore the necessity of an interdisciplinary approach that incorporates a philological component [8, 146].

In the context of the structuring of translation processes that utilize various computer-assisted tools, it becomes evident that the notion of frequency emerges as a pivotal consideration. Understanding common errors and translation models may contribute to the development of new algorithms in various fields of applied linguistics. The feasibility of this occurrence may be attributable to the fact that machine translation tools are designed to select the most frequently used translation equivalents, a process that is informed by the frequency principle. This approach, however,

ISSN 2788-7979 (online)

may result in the algorithm exhibiting biases and errors [9]. In their research on theoretical concepts and models of translation, G.I. Kuldeeva and Zh.N. Kuzar state that, in order to translate a text adequately, it is necessary to consider its linguistic, textual and cultural features [10, 157].

The stylistic richness of machine translation outputs may also present a challenge for computer-assisted translation tools. As posited by the scholars M. Brglez and Š. Vintar, machine-generated translations may exhibit a vocabulary that is quantitively more extensive than that of human-generated translations. Nevertheless, regarding the issue of quality, there is frequently an absence of consistency, with the presence of semantic errors and a lack of clarity in meaning [11]. It is therefore recommended that research be conducted into the most frequent errors occurring during the utilisation of computer-assisted translation, with a view to taking these into consideration when evaluating texts combining different themes and pragmatic purposes.

The objective of this research is to identify the characteristics associated with the utilization of various computer-assisted translation tools, with the aim of classifying the translation errors induced by the use of these tools. It is hypothesized that this will contribute to different purposes, such as the elimination of transformation-induced translation errors in the aforementioned tools, as well as the training of translators to work with these tools. The focal point of this study is the process and outcomes of various computer-assisted translation tools that facilitate machine translation.

MATERIALS AND METHODS

The present study will examine the qualitative and quantitative characteristics of the translation features of texts provided on the official website of the Astana Ballet Theatre. Furthermore, it will explore the potential classification of translation errors induced by transformation, with the analysis being illustrated by the translation of such texts made by various digital tools. Following a thorough examination of the extant literature on the subject, it was determined that the translation experiment would be conducted on the basis of text material that combines the following characteristics: themes that reflect contemporary Kazakhstani society, a rich cultural content, a diverse vocabulary, and full internet access for linguistic research without compromising confidentiality regulations. This description corresponds to the following types of journalistic style texts, as found on the website of a cultural venue: press releases, advertisements, performance schedules, and biographies of public figures in the sphere of art.

Following the selection of materials, the issue arose of choosing the most appropriate computer-assisted instruments from the extensive range of digital software available on the market. For example, the website g2.com, which provides reviews of software products for various sectors, contains 64 listings for computer-assisted translation. In order to ensure the representativeness of the research findings, it was necessary to select the most popular and frequently utilized computer-assisted translation tools, which offer translation services free of charge. In that regard, we conducted a preliminary survey among students at different levels of study, including bachelor’s, master’s and PhD programs of translation studies. The survey was carried out at two universities: Gumilyov Eurasian National University in Kazakhstan and Ljubljana University in Slovenia. The survey results indicated a high level of utilization of computer-assisted translation tools by the respondents, along with a notable degree of confidence in the use of SmartCat (a free, Yandex Translate-integrated version), DeepL, Google Translate and Promt.

The combination of the survey results with the research conducted previously by Kazakhstani scholars resulted in the selection of these four machine translation resources. A total of 200 texts were analyzed, with a word count ranging from 156 to 748 words, thus yielding an approximate total of 113,000 words. All texts are available online at the Astana Ballet website and are used for scientific purposes only. Furthermore, the results of this study, which are intended to validate the research hypothesis, will be disseminated in a manner that adheres to copyright regulations, with no more than 3% of the entire text being disclosed.

The next stage involved downloading the Russian source texts into machine translation tools, after which a comparative analysis of the translation strategies employed in the classical theory of

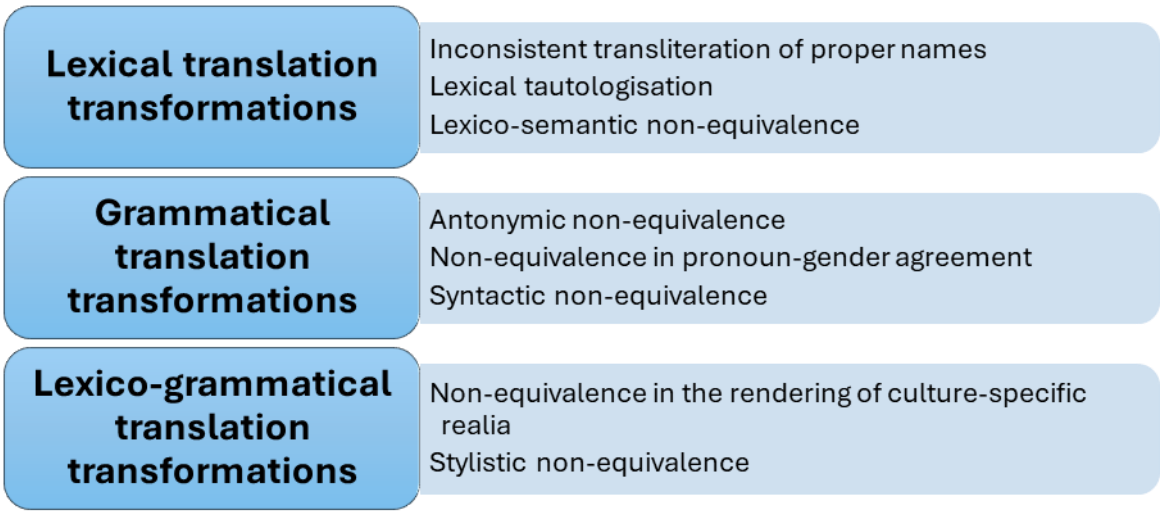
ISSN 2788-7979 (online)

translation was conducted. In fact, the typology of translation errors has already been established. This typology was proposed by N.K. Garbovsky in 2007, and it offers a multilevel classification of typal translation errors made by human translators. Nevertheless, this typology is predicated upon the systematization of the reasons for the occurrence of translation errors, depending on the particular stage of the translation process at which they arise [12, 514]. It is acknowledged that the human factor and cognitive processes play a significant role in this typology. Consequently, it is anticipated that it will be challenging to base the case for computer-assisted translation tools on it. In that regard, it was deemed necessary to establish a correlation between the machine translation errors and the translation transformations classification proposed by V.N. Komissarov. This classification is predicated on the levels of language and thus may be correlated to them. However, after conducting the research and shaping the aforementioned classification, we also discovered a significant new purpose for using it. The classification of transformation-induced translation errors has demonstrated that the utilization of computer-assisted translation tools, in conjunction with their non-controlled usage, can expeditiously engender variations that may cause extralinguistic challenges.

RESULTS

The findings of the present study's comparative analysis of the translated texts demonstrated that the digital tools employed in the study exhibited diverse translation outcomes. However, the analysis revealed the presence of translation errors, which were subsequently categorized into three distinct groups based on the classification of translation transformations proposed by V.N. Komissarov. Within these categories, they are also subdivided into subgroups based on frequency, as determined by statistical analysis. Figure 1 illustrates the classification's categories and subgroups.

Figure 1: Classification of transformation-induced translation errors.



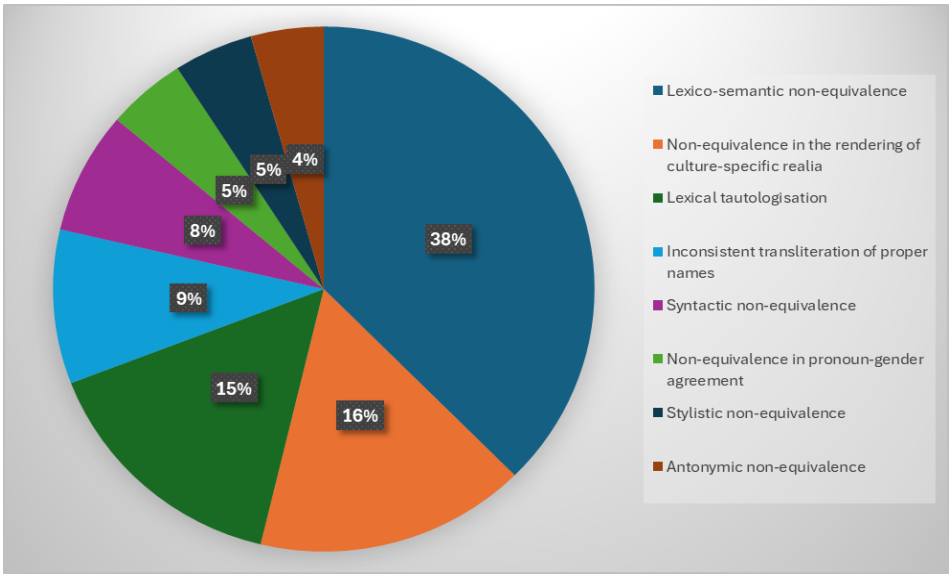
The qualitative representation of the errors, variations and non-equivalent transformations amounts to 1,012 examples, which comprises an average of 5.06 errors per text under research. The category of lexical transformations was found to contain the highest percentage of errors and variations, with 634 examples identified. The corpus contains 380 instances of lexico-semantic non-equivalence, 156 instances of lexical pathologization (redundant lexical repetition), and 96 instances of inconsistent transliteration of proper names. The subsequent qualitative category is that of lexico-grammatical translation transformations, which demonstrate 212 instances. These are divided into two subgroups: 164 instances of non-equivalence in the rendering of culture-specific realia, and 48 instances of stylistic non-equivalence. As will be demonstrated in the following section, this issue, occurring during machine translation, concerns not only Kazakhstani cultural features, which prevail in the source texts under consideration, but also specific cultural features in general. In general,

ISSN 2788-7979 (online)

cultural realia, phraseologisms and other characteristics of literary texts represent a key interest of modern Kazakhstani researchers. This assertion is substantiated by the findings of qualitative analysis conducted by A.K. Zhumabekova in her recent research, wherein the author underscores the necessity for further investigation into the utilization of contemporary digital translation technologies [13].

The category of grammatical transformations demonstrates the lowest percentage of instances, with a total of 168. It is important to note, however, that this category contains three subgroups: 44 examples of antonymic non-equivalence, 48 examples of non-equivalence in pronoun-gender agreement, and 76 examples of syntactic non-equivalence. It is hypothesized that this phenomenon may be associated with the systematic nature of computer-assisted translation tools, which are predicated on rigorous structural algorithms. The percentage ratio of results obtained from the classification of transformation-induced translation errors is hereby presented.

Figure 2. The percentage ratio of results obtained in the classification of transformation-induced translation errors.



In order to provide a more thorough description of each category, the following practical examples will be considered. These examples were obtained from the machine translation of computer-assisted tools, including SmartCat (a free, Yandex Translate-integrated version), DeepL, Google Translate and Promt. The most substantial subgroup in the category of lexical transformations is *lexico-semantic non-equivalence*, which encompasses the following errors:

1. Errors of the lexical non-equivalence type
2. Secondly, the issue of translator's false friends is to be considered.
3. The third point pertains to the literal translation.
4. The fourth point pertains to the process of concretization based on frequency.

Examples of such errors, known as "translator's false friends", include errors based on similar pronunciation in the source and target languages but with different meanings.

Table 1. Examples of "translator's false friends" within the classification.

Source Text	**Machine Translation**	**Edited Equivalent Translation**
… приобретает ореол культовости	an aura of cult/ a halo of cult	…becoming iconic
идеалы пластической красоты (о пластичности движений в танце)	ideals of plastic beauty	ideals of graceful beauty
Лаконичность	Laconicity	Laconicism

ISSN 2788-7979 (online)

It can be posited that the phenomenon of "translator's false friends" frequently manifests in machine translation due not only to an erroneous interpretation of the source text's intended meaning, but also to the absence of regulated linguistic elements within the database of a computer-assisted translation tool. The following example illustrates the issue: the term "laconicity" is rendered as such in the translation, despite the Oxford Dictionary's assertion that the term is not generally recognized in the English language. Nevertheless, it is employed in informal discourse and non-official internet sources. Another voluminous group of examples within the lexico-semantic non-equivalence subgroup is literal translation, which does not take into account semantic and stylistic harmony of words in collocations and sentences.

Table 2. Examples of "literal translation" and lexical non-equivalence within the classification.

Source Text	**Machine Translation**	**Edited Equivalent Translation**
визитная карточка театра	the theater's business card/ the theater's calling card	the theater's signature performance
образцы народно-сценического танца стали хрестоматийными	The examples of folk stage dance have become textbook	The classical examples of the folk-stage tradition have become truly canonical
Номер (в концертной программе)	room	piece
…искусство, в которое вкладывается душа	…soul of the creative team is invested	…art, into which the creative team puts their heart and soul
Музыкальная обработка	Music processing/ treatment	Music arrangement

The first example, which employed the phrase "theatre's signature performance", was derived from a press release concerning the international tour of a capital's theatre. Consequently, the entirety of the text was dedicated to this notion. In the context of cultural venues, such as theatres, the phrase "визитная карточка" is a commonplace occurrence in written texts. Nonetheless, all of the computer-assisted tools exhibited an aptitude for analyzing context while selecting an appropriate English equivalent, which resulted in literal translation. The second example demonstrates the fact that, on occasion, literal translation occurs on the basis of translating words which have acquired additional meaning due to some cultural phenomena in society. The adjective «хрестоматийный» initially demonstrated its association with a textbook, comprising a collection of literary works that were curated for a specific purpose. It is evident that the adjective has evolved in meaning over time, acquiring the sense of "canonical" or "exemplifying". In contemporary usage, it is employed in this sense in the majority of cases. The final example in the table above employs the musical term "arrangement". It is asserted that this terminology is frequently misinterpreted when translated literally by computer-assisted tools, resulting in a distortion of its intended meaning.

Another interesting example of lexico-semantic non-equivalence is so-called "frequency-based concretization", the transformation error which may occur only in machine translation. The crux of the issue pertains to the profound misinterpretation of the source text's context and cultural nuances by digital translation tools. As previously mentioned, contemporary computer-assisted translation tools operate based on either a statistical approach or a neural network approach. It is evident that both of these approaches are trained on the basis of translation, which has been previously made and edited by human beings. Therefore, in the event of the context being different from that which was predominant in the database, the machine may also offer a concretization equivalent, which would constitute an error in this case. The subsequent translation is a descriptive exemplification of this shift:

1. The following is an excerpt of a source sentence: *"…поэма о любви "Козы Корпеш – Баян Сулу", приуроченная ко* ***дню влюбленных*** *в Казахстане*".

2. The following translation was provided by all four machine translation tools: *"…the poem entitled "Kozy Korpesh - Bayan Sulu", which is dedicated to* ***Valentine's Day*** *in Kazakhstan*".

It is evident that the example of frequency-based concretization, *Valentine's Day*, is clearly induced by the popularity of this holiday in English-speaking countries. Consequently, this example

ISSN 2788-7979 (online)

is the most frequently used in this context. Notwithstanding the reference to *Kazakh Love Day* and the country itself, the programs were unable to provide a culturally appropriate translation.

The subsequent subgroup of the category of lexical transformations is that of *lexical tautologizing*, which is defined as redundant lexical repetition. This phenomenon can be exemplified by the repetition of the same word form in synonymous words of a source language. For instance, the phrase "систематическая и планомерная" is translated as "systematic and systematic". Moreover, it is more frequently represented by the repetition of the same lexemes in different word-forms, providing translations such as "charming charm", "filming of the film", "performance is performed", "active activities", etc. Whilst this is not generally regarded as an error, it has been demonstrated to have a significant impact on the style of the text.

Another subgroup of lexical transformations is the *inconsistent transliteration of proper names*. The findings demonstrated that all four computer-assisted translation tools under review exhibited a propensity to yield disparate transliteration equivalents, giving rise to variations such as "*Abai Qunanbaiuly/ Abay Kunanbayev/ Abai Kunanbaiev*", "*J. Verdi/ G. Verdi*", "*V.A. Mozart/ W.A. Mozart*", "*Kayrasheva/ Kairasheva*", "*Amanzhol/ Amanjol*", and so forth. It is evident that a significant proportion of proper names, such as Wolfgang Mozart, are derived from languages utilizing Latin script. This characteristic renders them amenable to straightforward modification. However, there is still a question of why digital tools provide erroneous transliterations when these names have one definitive variant. Another proportion of proper names, such as Abai Qunanbaiuly, was traditionally written in the Kazakh language. The Kazakh language underwent a shift to Latin script in 2017, and the final alphabet is currently under discussion among linguists. Therefore, inconsistent transliteration of proper names can only cause confusion if it is uncontrolled and stems from non-professional translations. The provision of consistent transliteration variants of proper names is a complex matter that raises numerous questions, including extralinguistic factors. It is therefore impossible to resolve these issues using computer-assisted tools alone.

In comparison with lexical transformations, errors in lexico-grammatical translation are comparatively infrequent, with a total of 212 examples identified. Nevertheless, the non-equivalence subgroup, which concerns the *rendering of culture-specific realia*, accounts for the second highest number of instances, with 164 cases recorded. In this instance, the programs presented two options: simple calque or descriptive translation, with the loss of national zest. All four digital tools translated such variations as "*aksakals/the elders*", "*dzhigits/the horsemen*" and "*shelpek/tortillas/flatbread*". The following example of machine translation, presented in Table 3, will be examined.

Table 3. Illustration of errors relating to culture-specific realia.

Source Text	**Machine Translation**	**Edited Equivalent Translation**
Аксакалы считали, что у мужчины в обязательном порядке должны быть семь богатств.	Aksakals believed that a man must necessarily have seven riches.	In traditional Kazakh culture, respected elders believed that a man was expected to possess the seven blessings that define a fulfilled life.

Although transliteration without explication may seem formally correct when translating culture-specific realia, it is not a transparent solution for the target audience because it often leads to a lack of cultural compensation in the translated text. Therefore, it was decided that "аксакалы" should be translated as "respected elders", which highlights the social status of this group of people while remaining understandable for the English-speaking audience without the need for a footnote. Moreover, the introductory phrase, which is missing from the original text, provides a general cultural context for the entire text.

The machine translation also ignores the symbolic meaning of the phrase "семь богатств" in this context. Originally representing a cultural concept, it is reduced to a material nomination in translation. This happens because, although modern and precise, the algorithms of a computer program are primarily governed by the priority of a vocabulary equivalent, often without considering

ISSN 2788-7979 (online)

the cultural background. We offer the following translation of this culture-specific item of realia: “the seven blessings”. This does not only imply a material nomination; it also represents the axiological nature of this realia item. While editing the machine translation, we considered not only culture-specific realia, but also the pleonastic phrase “must necessarily have”, changing it to “was expected to possess”.

It was determined that all of these variations were either not understandable to the foreign audience or too neutral for a cultural background. The translation challenge can be addressed to a certain extent by providing a transliterated version of the cultural realia, with the sentence structure being modified grammatically to enable the harmonious incorporation of a descriptive translation of the cultural realia. It is evident that in the contemporary era, this can only be achieved through the intervention of human post-editing of machine translation.

The grammatical transformations constituted the smallest proportion of errors, with 168 instances recorded. At the outset of the study, our hypothesis was that there would be fewer instances, due to the highly structured nature of the grammatical category. We reasoned that this would facilitate effective description and translation without compromising meaning. However, this category exhibited a substantial number of errors in subgroups such as antonymic non-equivalence, non-equivalence in pronoun-gender agreement and syntactic non-equivalence. Antonymic non-equivalence has been demonstrated in the change of acting roles in a sentence, as illustrated by the following examples.

Table 4. Examples of antonymic non-equivalence within the classification.

Source Text	**Machine Translation**	**Edited Equivalent Translation**
В программу гастролей вошли спектакли, которые уже полюбились казахстанской публике	The tour program included performances that have already fallen in love with the Kazakh public	The tour program included performances that have already become beloved by the Kazakh audience
Столичную публику ждёт буйство эмоций	The metropolitan audience is waiting for a riot of emotions	The capital’s public can expect an outburst of emotions

The first example demonstrates the challenges faced by machine translation modules when translating Russian reflexive constructions into English. The second example illustrates the challenges associated with sentences that exhibit a relatively free word order. The machine translation module automatically attempts to preserve the word order, which occasionally leads to a distortion of meaning when translating between languages with different word order systems. The translation results in a complete transformation of the original meaning by altering the roles of the actants in this sentence.

Furthermore, an analysis of digital translation tools reveals a tendency to exhibit frequent errors in the application of *pronoun-gender agreement*, particularly in relation to pronouns such as “he”, “she” and “it”, as well as their possessive forms. This occurrence is notable despite the context providing clear indication as to the intended pronoun usage. The machine translation modules installed in computer-assisted tools frequently exhibit inaccuracies in the segmentation of a text into parts, as well as the phenomenon of language interference and deictic shift. As illustrated in Table 5, the error could be identified by a user with no professional translation background. However, it should be noted that the four machine translation modules utilised in this study were unable to produce a suitable equivalent. In this example, an error in pronominal reference is evident. It has been caused by a deictic shift, which has consequently resulted in a change in narrative perspective. The edited equivalent translation, performed by a human translator, aims to avoid repetition of a pronoun in order to maintain the referent structure of the sentence.

Table 5. Illustration of an error in pronominal reference.

Source Text	**Machine Translation**	**Edited Equivalent Translation**
Как-то на вопрос, не хотите ли вы ставить классический балет, он ответил: “нет, я предпочитаю писать свою собственную поэзию”.	Once, when asked if you would like to stage a classical ballet, he replied: “No, I prefer to write my own poetry”.	When asked about the possibility of staging a classical ballet, he replied, “No, I prefer to write my own poetry”.

Errors in grammatical transformations, including *syntactic non-equivalence*, also serve as a representative illustration of language interference, exerting an influence on the quality of translated results. This happens because the functional, semantic and stylistic structure of a sentence is ruined by the calquing of the source language’s syntactic structure.

As illustrated in Table 6, the following example presents several transformation-induced translation errors simultaneously: syntactic non-equivalence, commonization of a proper name, and an incorrect order of subject-predicate relations. Firstly, let us consider the issue of coreference failure, illustrated by a double relative clause construction that results in syntactic overload and semantic ambiguity: “who loved her, whom she leaves”. In addition, machine translation modules have incorrectly converted the proper name “Torero” into the common noun “bullfighter”, mixing this word with the Spanish language, which was not adjusted in the settings of these programs. Another major error is the incorrect order of subject-predicate relations, with the subject preceding the predicate. This demonstrates a more Russified word order and is an example of language interference. This error could easily have been avoided by a human translator but remains challenging for a machine.

Table 6. Illustration of an error in syntactic non-equivalence.

Source Text	**Machine Translation**	**Edited Equivalent Translation**
В центре балета – трагическая судьба Кармен и полюбившего ее солдата Хозе, которого она покидает ради молодого Тореро.	At the center of the ballet is the tragic fate of Carmen and the soldier Jose who loved her, whom she leaves for the sake of a young Bullfighter.	The ballet centres on the tragic fate of the gypsy Carmen and the soldier José, who falls in love with her, only for her to leave him for the young Torero.

The issue arises from the fact that these errors visually appear correct in the text. Linguistically non-qualified users may not be aware of these errors and may not edit them correctly. They may, however, be more likely to notice and edit visually more obvious errors, such as lexical tautologizing. The systematic nature of this language level facilitates the enhancement of its algorithms for translation in the event of the emergence of more advanced linguistic specifications. All the data obtained in this research demonstrates that computer-assisted tools with machine translation modules still cannot perform error-free qualitative translation. Issues in translation studies such as terminology, culture-specific realia, semantic and syntactic consistency, and idioms and other stylistic devices in a text, still pose the greatest challenges to machine translation algorithms.

DISCUSSION

A comparative analysis of 200 open access texts, retrieved from the Astana Ballet website and translated from Russian into English with the help of four different computer-assisted translation tools, showed that machine translation showcases common trends and particular characteristics. It is important to acknowledge that the typologies, classifications and approaches of the classical theory of translation can be effectively employed in conducting research in contemporary fields of study, such as machine translation.

ISSN 2788-7979 (online)

The present study demonstrated that machine translation results, pertaining to diverse translation transformations, frequently exhibited common errors, indicative of typical issues in machine translation. The highest number of errors was found in the category of lexical transformations, which demonstrates that even digital tools with extensive databases are unable to fully address the challenges posed by lexical non-equivalence errors, translator's false friends and literal translation. The issues inherent to literal translation pertain even to such a precise domain as terminology. Furthermore, the unique characteristics of the algorithms that underpin machine translation can result in the emergence of novel errors that are challenging to identify in translations produced by humans. This phenomenon is exemplified by the occurrence of so-called frequency-based concretization.

The results demonstrated an even simpler solution to the challenge of lexical tautologizing, which can easily be found automatically because it uses the same digital characters in the stems of words. Nevertheless, this subgroup accounted for a significant proportion of errors in the lexical translation transformation category. Subgroups such as "inconsistent transliteration of proper names" and "non-equivalence in the rendering of culture-specific realia" have revealed hidden potential within the proposed classification of errors. This has in turn helped to unveil new research directions that are important for the Kazakhstani scientific community. It seems that, in general, all findings can be harmoniously grouped into the Classification of Transformation-Induced Translation Errors. This could contribute to the solution of the problem of facilitating learning how to work with machine translation and improving its software products.

CONCLUSION

As demonstrated in the preceding analysis, contemporary technological advancements and the rapid development of the sphere of computer-assisted translation have not yet resulted in the elimination of errors in digital tools. The digital tools under consideration, namely SmartCat (a free, Yandex Translate-integrated version), DeepL, Google Translate and Promt, still demonstrate a considerable number of errors. This fact suggests that the translation process necessitates the involvement of professional human translators, whose expertise lies in meticulous editing, to ensure the delivery of a translation that meets the requisite standards of quality.

Nevertheless, computer-assisted translation tools have been shown to enhance translators' productivity by allowing them to consider and structure all the nuances of a source text. Furthermore, these tools connect translators with professional dictionaries and terminology databases.

The classification of transformation-induced translation errors has the potential to enhance the efficacy of algorithms employed within computer-assisted translation tools. Furthermore, it can be employed as a typology in training tutorials to assist professional translators in utilizing translation software. This may serve as a useful indication to technological companies involved in the development of translation software that the involvement of professional translators in the process of creating new digital projects is a significant factor to consider. Consequently, new areas of professional activities may emerge, allowing translators to acquire new competencies and develop their potential.

The most significant aspect is that it can assist in identifying the roots of such contentious issues as the transliteration of proper names and the variation of English. Furthermore, it can promote the necessity of a unified transliteration standard for English translations of Kazakhstan-related proper names.

INFORMATION ON FINANCING

The research has been carried out under the project AP23490448 "Consistency of Kazakh Onyms in English-Language Texts: Translatory Research" within the framework of subjects of scientific and/or scientific and technical activities grant funding by the Ministry of Science and Higher Education of the Republic of Kazakhstan.

ISSN 2788-7979 (online)

REFERENCES

1 Prime Minister of the Republic of Kazakhstan (2024), “Concept for artificial intelligence development for 2024–2029 adopted by government”, available at: https://primeminister.kz/en/news/concept-for-artificial-intelligence-development-for-2024-2029-adopted-by-government-28786 (Accessed 06.12.2025).

2 Pravitel'stvo Respubliki Kazakhstan (2023), Kontseptsiya razvitiya yazykovoi politiki v Respublike Kazakhstan na 2023–2029 gody (Postanovlenie Pravitel’stva Respubliki Kazakhstan № 914) [Concept for the development of language policy in the Republic of Kazakhstan for 2023–2029 (Resolution of the Government of the Republic of Kazakhstan No. 914)], available at: https://www.gov.kz/uploads/2023/4/13/caa19f0131a736066966b9314734bfe1_original.674816.doc (Accessed 06.12.2025). [In Russian]

3 Melby, A.K. (1983), “Computer-assisted translation systems: The standard design and a multi-level design”, in First Conference on Applied Natural Language Processing, Association for Computational Linguistics, pp. 174–177.

4 Schwartz, L. (2018), “The history and promise of machine translation”, in Lacruz, I. and Jääskeläinen, R. (eds.), Innovation and expansion in translation process research (American Translators Association Scholarly Monograph Series, Vol. 18), John Benjamins, Amsterdam, pp. 161–190.

5 Hutchins, W.J. (1995), “Machine translation: A brief history”, in Koerner, E.F.K. and Asher, R.E. (eds.), Concise history of the language sciences: From the Sumerians to the cognitivists, Pergamon, Oxford, pp. 431–444.

6 Tukebayeva, J.A. (2025), “Current problems and trends in machine translation from Russian into Kazakh”, Russian Insights: Literature, Culture and Linguistic, no. 2, pp. 83–87.

7 Zhumay, N. and Akizhanova, D. (2023), Machine Translation. Oqu quraly [Machine Translation. Study guide], L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, 139 p.

8 Shengelbay, M.K., Alkebaeva, D.A. and Talaspaeva, Zh.S. (2025), Zhasandy intellekt tili: lingvistikanyn zhana kokzhiekteri [Artificial intelligence language: New horizons of linguistics], Bulletin of S. Ualikhanov KU. Philological Series, no. 3, pp. 138–149, doi: 10.59102/kufil/2025/iss3pp138-149. [In Kazakh]

9 Kuzar, Zh.N. and Kuldeeva, G.I. (2022), “Theoretical concepts and models of translation: essence, functions, types”, Vestnik Evraziiskogo natsional'nogo universiteta im. L.N. Gumileva. Seriya filologicheskaya, no. 2(139), pp. 151–159. [In Russian]

10 Sholikhah, N.F.M. and Indah, R.N. (2021), “Common lexical errors made by machine translation on cultural text”, Edulingua: Jurnal Linguistiks Terapan dan Pendidikan Bahasa Inggris, vol. 8, no. 1, pp. 40–51.

11 Brglez, M. and Vintar, Š. (2022), “Lexical diversity in statistical and neural machine translation”, Information, vol. 13, no. 93, 14 p.

12 Garbovskiy, N.K. (2007), Teoriya perevoda [Theory of translation], 2nd ed., Izdatel’stvo Moskovskogo universiteta, Moscow, Russia, 544 p., ISBN 978-5-211-05333-5. [In Russian]

13 Zhumabekova, A.K. (2022), Nauchnye issledovaniya po perevodovedeniyu v Kazakhstane: novejshij period [Scientific research on translation in Kazakhstan: the modern period], Filologiya. Seriya: Filologicheskie nauki, vol. 66, no. 3, pp. 218–233, doi: 10.48371/PHILS.2022.66.3.018. [In Russian]

14 Astana Ballet (2025), “Official website of Astana Ballet”, available at: [https://astanaballet.com] (https://astanaballet.com) (Accessed 08.12.2025).

Received: 28.03.2026

ISSN 2788-7979 (online)

Трансформациялық-қате аударма нәтижелерінің жіктелуі (автоматтандырылған аударма бағдарламаларының мысалы негізінде)

Л.А. Задорожная [1], М.Ж. Ескиндирова [1], Д. Попич[2]
[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, 010000, Қазақстан
[2]Любляна Университеті, Любляна, 1000, Словения

Бұл мақала аударма теориясы мен практикасы контекстінде автоматтандырылған аударма бағдарламаларының ерекшеліктерін зерттейді. Автоматтандырылған аударма үшін заманауи оқу құралдарына, сондай-ақ жаңа цифрлық өнімдерді әзірлеу барысында туындайтын қолданбалы мәселелерді шешуге пайдалы болуы мүмкін жаңа типологиялар мен жіктеулерді құрудың өзекті мәселелері талқыланады. Мәліметтер базасы ретінде машиналық аудармамен жұмыс істеу кезінде пайда болатын жиілік қателіктеріне назар аудару ұсынылады. Авторлар бағдарламалар орындаған аудармалардың ағылшын тіліндегі мәтіндерін салыстырмалы талдау нәтижелерін ұсынады. Автоматтандырылған аударманың трансформациялық-қате нәтижелерінің ұсынылған жіктелуі аударма трансформациясының үш түріне негізделген: лексикалық, грамматикалық және лексикалық-грамматикалық. Жоғарыда берілген аударма трансформацияларының негізінде машиналық аударманың жиілік қателіктерінің 8 түрі ажыратылады. Экстралингвистикалық талдау нәтижелері бойынша мұндай қателіктердің пайда болуының алғышарттары анықталады. Машиналық аударманың анықталған қателіктері, сондай-ақ автоматтандырылған аударма үшін бағдарламаларды пайдалануды стандарттаудың жоқтығының және қазақстандық ғылымның өзекті мәселелерінің, мысалы, өз есімдерінің транслитерациясының вариативтілігі байланысын болжауға мүмкіндік береді. Зерттеу нәтижесінде автоматтандырылған аударма бағдарламаларымен жұмыс істеу ерекшеліктері туралы қорытынды жасалады және оларды аударма тәжірибесі аясында да, ғылыми тәсіл аясында да тиімді қолдану бойынша ұсыныстар беріледі.

Кілт сөздер: автоматтандырылған аударма, аударма трансформациясы, аударма қателері, транслитерация, цифрландыру

ӘДЕБИЕТТЕР ТІЗІМІ

1 Қазақстан Республикасы Премьер-Министрінің ресми сайты (2024), «Үкімет жасанды интеллектіні дамытудың 2024–2029 жылдарға арналған тұжырымдамасын қабылдады», [Электрон. ресурс]: https://primeminister.kz/news/ukimet-zhasandy-intellektini-damytudyn-2024-2029-zhyldarga-arnalgan-tuzhyrymdamasyn-kabyldady-28786 [кол жеткізу күні: 06.12.2025].

2 Қазақстан Республикасының Үкіметі (2023), «Қазақстан Республикасында тіл саясатын дамытудың 2023–2029 жылдарға арналған тұжырымдамасы (Қазақстан Республикасы Үкіметінің № 914 қаулысы)», [Электрон. ресурс]: https://www.gov.kz/uploads/2023/4/13/caa19f0131a736066966b9314734bfe1_original.674816.doc [кол жеткізу күні: 06.12.2025].

3 Melby, A.K. (1983), “Computer-assisted translation systems: The standard design and a multi-level design”, in First Conference on Applied Natural Language Processing, Association for Computational Linguistics, pp. 174–177.

4 Schwartz, L. (2018), “The history and promise of machine translation”, in Lacruz, I. and Jääskeläinen, R. (eds.), Innovation and expansion in translation process research (American Translators Association Scholarly Monograph Series, Vol. 18), John Benjamins, Amsterdam, pp. 161–190.

5 Hutchins, W.J. (1995), “Machine translation: A brief history”, in Koerner, E.F.K. and Asher, R.E. (eds.), Concise history of the language sciences: From the Sumerians to the cognitivists, Pergamon, Oxford, pp. 431–444.

6 Tukebayeva, J.A. (2025), “Current problems and trends in machine translation from Russian into Kazakh”, Russian Insights: Literature, Culture and Linguistic, no. 2, pp. 83–87.
7 Zhumay, N. and Akizhanova, D. (2023), Machine Translation. Oqu quraly [Machine Translation. Study guide], L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, 139 p.
8 Шенгелбай М. Қ., Алкебаева Д.А., Таласпаева Ж.С. (2025), «Жасанды интеллект тілі: лингвистиканың жаңа көкжиектері», Ш. Уәлиханов атындағы Көкшетау университетінің хабаршысы. Филология сериясы, № 3, 138–149-бб., doi: 10.59102/kufil/2025/iss3pp138-149.
9 Құзар Ж.Н., Кульдеева Г.И. (2022), «Аударманың теориялық түсініктері мен модельдері: мәні, қызметі, түрлері», Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Хабаршысы. Филология сериясы, № 2(139), 151–159-бб., doi: https://doi.org/10.32523/2616-678X-2022-139-2-151-159
10 Sholikhah, N.F.M. and Indah, R.N. (2021), “Common lexical errors made by machine translation on cultural text”, Edulingua: Jurnal Linguistiks Terapan dan Pendidikan Bahasa Inggris, vol. 8, no. 1, pp. 40–51.
11 Brglez, M. and Vintar, Š. (2022), “Lexical diversity in statistical and neural machine translation”, Information, vol. 13, no. 93, 14 p.
12 Гарбовский Н.К. (2007), Аударма теориясы, 2-бас., Мәскеу университетінің баспасы, Мәскеу, Ресей, 544 б., ISBN 978-5-211-05333-5.
13 Жұмабекова А.Қ. (2022), «Қазақстандағы аударма туралы ғылыми зерттеу: қазіргі кезең», Филология. Серия: Филологиялық ғылымдар, т. 66, № 3, 218–233-бб., doi: 10.48371/PHILS.2022.66.3.018.
14 Астана Балет театры (2025), «Астана Балет театрының ресми сайты», [Электрон. ресурс]: https://astanaballet.com [кол жеткізу күні: 06.12.2025].

Материал 28.03.2026 баспаға түсті

Классификация трансформационно-ошибочных результатов перевода (на примере программ для автоматизированного перевода)

Л.А. Задорожная [1], М.Ж. Ескиндирова[1], Д. Попич[2]
[1]Евразийский национальный университет имени Л.Н. Гумилева, Астана, 010000, Казахстан
[2]Университет Любляны, Любляна, 1000, Словения

Данная работа исследует особенности программ для автоматизированного перевода в контексте теории и практики перевода. Обсуждаются актуальные вопросы создания новых типологий и классификаций, которые могут быть полезны для современных учебных пособий для автоматизированного перевода, а также для решения прикладных задач, возникающих в ходе разработки новых цифровых продуктов. В качестве базы данных предлагается обратить внимание на частотные ошибки, возникающие при работе с машинным переводом. Авторы представляют результаты сравнительного анализа англоязычных текстов переводов, выполненных программами. Предлагаемая классификация трансформационно-ошибочных результатов автоматизированного перевода основана на трех видах переводческих трансформации: лексических, грамматических и лексико-грамматических. На основе указанных переводческих трансформаций выделяются 8 типов частотных ошибок машинного перевода. По итогам экстралингвистического анализа выявляются предпосылки возникновения таких ошибок. Выявленные ошибки машинного перевода также позволяют предположить связь отсутствия стандартизации использования программ для автоматизированного перевода и таких актуальных вопросов казахстанской науки, как вариативность транслитерации имен собственных. В результате исследования делаются выводы об особенностях работы с программами для автоматизированного перевода и даются рекомендации по их эффективному применению как в контексте практики перевода, так и в рамках научного подхода.

ISSN 2788-7979 (online)

Ключевые слова: автоматизированный перевод, переводческие трансформации, ошибки перевода, транслитерация, цифровизация

СПИСОК ЛИТЕРАТУРЫ

1 Правительство Республики Казахстан (2024), «Правительством принята Концепция по развитию искусственного интеллекта на 2024–2029 годы», [Электрон. ресурс]: https://primeminister.kz/ru/news/pravitelstvom-prinyata-kontseptsiya-po-razvitiyu-iskusstvennogo-intellekta-na-2024-2029-gody-28786 [дата обращения: 06.12.2025].

2 Правительство Республики Казахстан (2023), «Концепция развития языковой политики в Республике Казахстан на 2023–2029 годы (Постановление Правительства Республики Казахстан № 914)», [Электрон. ресурс]: https://www.gov.kz/uploads/2023/4/13/caa19f0131a736066966b9314734bfe1_original.674816.doc [дата обращения: 06.12.2025].

3 Melby, A.K. (1983), “Computer-assisted translation systems: The standard design and a multi-level design”, in First Conference on Applied Natural Language Processing, Association for Computational Linguistics, pp. 174–177.

4 Schwartz, L. (2018), “The history and promise of machine translation”, in Lacruz, I. and Jääskeläinen, R. (eds.), Innovation and expansion in translation process research (American Translators Association Scholarly Monograph Series, Vol. 18), John Benjamins, Amsterdam, pp. 161–190.

5 Hutchins, W.J. (1995), “Machine translation: A brief history”, in Koerner, E.F.K. and Asher, R.E. (eds.), Concise history of the language sciences: From the Sumerians to the cognitivists, Pergamon, Oxford, pp. 431–444.

6 Tukebayeva, J.A. (2025), “Current problems and trends in machine translation from Russian into Kazakh”, Russian Insights: Literature, Culture and Linguistic, no. 2, pp. 83–87.

7 Zhumay, N. and Akizhanova, D. (2023), Machine Translation. Oqu quraly [Machine Translation. Study guide], L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, 139 p.

8 Шенгелбай М.К., Алкебаева Д.А., Таласпаева Ж.С. (2025), «Язык искусственного интеллекта: новые горизонты лингвистики», Вестник КУ имени Ш. Уалиханова. Серия филологическая, № 3, с. 138–149, doi: 10.59102/kufil/2025/iss3pp138-149.

9 Кузар Ж.Н., Кульдеева Г.И. (2022), «Теоретические концепции и модели перевода: сущность, функции, виды», Вестник Евразийского национального университета имени Л.Н. Гумилёва. Серия Филология, № 2(139), с. 151–159, doi: https://doi.org/10.32523/2616-678X-2022-139-2-151-159

10 Sholikhah, N.F.M. and Indah, R.N. (2021), “Common lexical errors made by machine translation on cultural text”, Edulingua: Jurnal Linguistiks Terapan dan Pendidikan Bahasa Inggris, vol. 8, no. 1, pp. 40–51.

11 Brglez, M. and Vintar, Š. (2022), “Lexical diversity in statistical and neural machine translation”, Information, vol. 13, no. 93, 14 p.

12 Гарбовский Н.К. (2007), Теория перевода, 2-е изд., Издательство Московского университета, Москва, Россия, 544 с., ISBN 978-5-211-05333-5.

13 Жумабекова А.К. (2022), «Научные исследования по переводоведению в Казахстане: новейший период», Филология. Серия: Филологические науки, т. 66, № 3, с. 218–233, doi: 10.48371/PHILS.2022.66.3.018.

14 Театр «Астана Балет» (2025), «Официальный сайт театра Астана Балет» [Электрон. ресурс]: https://astanaballet.com [дата обращения: 06.12.2025].

Материал поступил в редакцию журнала 28.03.2026