KOKAND UNIVERSITY HERALD SCIENTIFIC E-JOURNAL
QO‘QON UNIVERSITETI XABARNOMASI
KOKAND UNIVERSITY HERALD
ВЕСТНИК КОКАНДСКОГО УНИВЕРСИТЕТА
ISSN: 2181-1695 OAK: 01-08/1819/6 www.herald.kokanduni.uz

DEVELOPMENT OMACHINE TRANSLATION IN MODERN WORLD

Utayeva Iroda Baxodirovna,
Samarqand davlat pedagogika instituti Tillar fakulteti dekani,
Pedagogika fanlari bo‘yicha falsafa doktori, professor v.b.
https://orcid.org/0009-0004-7268-8083
DOI: https://doi.org/10.54613/ku.v19iA.1806

MAQOLA HAQIDA/ О СТАТЬЕ
Qabul qilindi: 17-avgust 2026-yil
Tasdiqlandi: 19-avgust 2026-yil
Jurnal soni: 19-A
Maqola raqami: 41

KALIT SO‘ZLAR/ КЛЮЧЕВЫЕ СЛОВА
mashina tarjimasi, avtomatlashtirilgan tarjima, tarjima aniqligi, kommunikativ ekvivalentlik, kontekstual adekvatlik, lingvistik texnologiyalar

ANNOTATSIYA/ АННОТАЦИЯ
Mazkur maqolada mashina tarjimasining rivojlanishi, uning texnologik evolyutsiyasi hamda zamonaviy ko‘p tilli kommunikatsiyadagi o‘rni yoritiladi. Tarjima aniqligi, kontekstual adekvatlik, kommunikativ ekvivalentlik, shuningdek, sohaga xos va madaniy jihatdan shartlangan lingvistik xususiyatlarni ifodalash bilan bog‘liq muammolarga alohida e’tibor qaratiladi. Shuningdek, zamonaviy mashina tarjimasi xizmatlari miqdoriy va sifat jihatidan qiyosiy baholanadi. Avtomatik baholash ko‘rsatkichlari bilan ekspert xulosalari o‘rtasidagi tafovutlar, ayniqsa, parafrazalash, maxsus terminologiya, sintaktik qayta tuzish va kontekstga bog‘liq ma’nolarni ifodalash holatlarida tahlil qilinadi. Tadqiqotda mashina tarjimasining ilmiy va tibbiy matnlarni tarjima qilishdagi imkoniyatlari ham ko‘rib chiqilib, terminologik aniqlik va kommunikativ adekvatlikni ta’minlashda post-tahrirning ahamiyati ko‘rsatib beriladi. Shuningdek, madaniy jihatdan xos birliklar, kam qo‘llanadigan leksik birliklar hamda maxsus kontekstual talqinni talab qiladigan terminlarni tarjima qilishdagi muammolar yoritiladi. Maqolada faqat lingvistik o‘xshashlikni emas, balki semantik munosabatlar, diskursiv tuzilma va manba matnining kommunikativ vazifasini ham hisobga olish zarurligi asoslanadi. Neyron mashina tarjimasi texnologiyalarini inson ekspertizasi bilan uyg‘unlashtirish akademik va professional kommunikatsiyada avtomatik tarjimaning izchilligi, aniqligi va amaliy samaradorligini oshirishning istiqbolli yo‘nalishi sifatida ko‘rsatiladi.

ABOUT THE PAPER
Accepted: 17 August 2026
Approved: 19 August 2026
Volume: 19-A
Paper number: 41

KEYWORDS
machine translation, automated translation, translation accuracy, communicative equivalence, contextual adequacy, linguistic technologies

ANNOTATION
This article examines the development of machine translation, its technological evolution, and its role in modern multilingual communication. Particular attention is given to translation accuracy, contextual adequacy, communicative equivalence, and the challenges of representing domain-specific and culturally conditioned linguistic features. The study additionally focuses on the comparative evaluation of contemporary machine translation services through quantitative and qualitative approaches. Particular emphasis is placed on identifying discrepancies between automatic evaluation scores and expert judgments, especially in cases involving paraphrasing, specialized terminology, syntactic restructuring, and context-dependent meanings. The research also considers the applicability of machine translation to scientific and medical discourse and highlights the importance of post-editing in achieving terminological precision and communicative adequacy. The findings contribute to a more comprehensive understanding of the practical capabilities and limitations of current automated translation technologies. The article further addresses the challenges associated with translating culturally marked expressions, low-frequency lexical units, and terminology that requires specialized contextual interpretation. It emphasizes the need to consider not only linguistic similarity but also semantic relations, discourse structure, and the intended communicative function of the source text. The study highlights the potential of combining neural machine translation with human expertise as a practical approach to improving the consistency, precision, and usability of automated translation in academic and professional communication.

Introduction

Translation as a professional practice has a long history, originating in periods when particular languages served as intermediary means of communication between members of different linguistic communities. The scope of activities encompassed by the concept of translation is exceptionally broad. Translation involves the transfer of diverse types of discourse from one language into another, including poetry, literary prose, scientific and technical publications, research articles from various fields of knowledge, official documents, business correspondence, as well as speeches and professional discussions among individuals who speak different languages and therefore require the assistance of an interpreter. From a theoretical perspective, translation can be defined as both the process and the outcome of producing a target-language text on the basis of a source-language text while maintaining an equivalent communicative value. In this context, communicative equivalence refers to the extent to which the translated text can function as an adequate substitute for the source text within communication involving speakers of different languages and within the communicative environment of the target language [1].

In recent decades, the practical significance of foreign-language competence has expanded considerably. Knowledge of foreign languages is required not only for international travel, communication with foreign visitors, participation in academic conferences, or professional interaction with colleagues from other linguistic backgrounds, but also for a wide range of everyday activities. Individuals may encounter foreign-language content while watching international films, consulting technical manuals and user instructions, or accessing online resources related to their professional or personal interests. Consequently, the demand for translation is no longer restricted to direct intercultural encounters; it has become an integral component of everyday information processing. Advances in digital technologies have increasingly enabled personal computers and other electronic devices to perform translation-related tasks that were traditionally carried out exclusively by human translators.

Within this context, machine translation (MT) systems have undergone substantial technological development. Earlier generations of machine translation software frequently produced fragmented, grammatically inconsistent, or semantically unclear output, whereas contemporary systems are increasingly capable of generating translations that demonstrate greater linguistic coherence, fluency, and comprehensibility. Moreover, early machine translation applications were generally expensive, technically complex, and difficult for non-specialist users to operate. The subsequent emergence of more accessible translation software made automated translation available to a considerably wider population, including users of personal computers. Nevertheless, the growing accessibility and sophistication of such systems raise important research questions concerning their translation accuracy, reliability, contextual adequacy, and communicative equivalence. Particular attention should therefore be given to whether machine translation systems are capable of adequately representing domain-specific terminology,

professional discourse conventions, pragmatic meaning, and culturally conditioned linguistic features.

The possibility of automating translation, which had traditionally been regarded as a complex form of intellectual activity, attracted scholarly and technological interest long before the development of modern computational systems. Historical accounts indicate that the nineteenth-century mathematician Charles Babbage, widely associated with early concepts of programmable computing, attempted to persuade the British government to provide financial support for his research into the development of a "calculating machine." Among the potential applications attributed to such a machine was the possibility that computational technology might eventually be capable of automatically translating spoken language [9]. This early conceptualization may therefore be regarded as one of the historical antecedents of contemporary research into machine translation, computational linguistics, and automated natural language processing.

Literature review

Machine translation has developed from rule-based and statistical approaches toward neural and transformer-based architectures, resulting in substantial changes in the quality, fluency, and contextual coherence of automatically generated translations. The theoretical foundations of machine translation and its role in professional intercultural communication have been examined by E. Yu. Kartseva, T. D. Margaryan, and G. G. Gurova, who emphasize that the effectiveness of automated translation should be considered not only in terms of formal linguistic correspondence but also from the perspective of communicative equivalence and the functional adequacy of the target text [1, pp. 155–164]. This approach is particularly relevant to contemporary MT research because a grammatically well-formed translation may nevertheless fail to preserve the communicative intention, contextual meaning, or stylistic characteristics of the source text.

The technological evolution of machine translation has been closely associated with the transition from statistical methods to neural machine translation. Research on Russian–English machine translation conducted within the Yandex framework demonstrated the application of statistical and data-driven approaches to the development of specialized translation systems [2, pp. 66–70]. Subsequent developments in neural architectures have substantially increased the ability of translation systems to model contextual relationships and generate more fluent target-language output. The sequence-to-sequence paradigm introduced by Sutskever, Vinyals, and Le, for example, established an important foundation for neural language generation by enabling models to transform sequences from one representational space into another [10]. Transformer-based approaches subsequently strengthened contextual modeling by relying on attention mechanisms and parallel processing, contributing to the development of modern multilingual and neural translation systems [8, pp. 249–255].

The quality of individual machine translation systems has also been examined through comparative empirical studies. Cambedda, Di Nunzio, and Nosilia conducted a comparative error analysis of DeepL and Yandex in Russian–Italian medical translation and demonstrated that the systems may differ considerably in the treatment of specialized terminology, syntactic structures, and domain-specific expressions [3, pp. 139–163]. Such findings are particularly important for scientific and medical translation, where terminological precision cannot always be adequately represented through general measures of textual similarity. Research on Google and Yandex translation systems has likewise demonstrated the possibility of identifying differences in automatically generated multilingual texts through linguistic and computational analysis [4, pp. 72–77]. These studies indicate that translation quality is influenced not only by the general architecture of an MT system but also by language pair, domain, corpus characteristics, and the linguistic properties of the source material.

The performance of contemporary translation systems is closely connected with the characteristics of their training data and the methods used to adapt models to specific languages and domains. Kamaluddin et al. discuss the development of DeepL and emphasize the technological advances that have contributed to improvements in machine translation quality [5, pp. 122–126]. Linlin similarly examines quality-control issues associated with AI-based translation and stresses the importance of evaluating machine-generated output beyond simple linguistic fluency [6, pp. 710–717]. From this perspective, high-quality translation requires not only a powerful multilingual model but also appropriate treatment of domain-specific vocabulary, contextual relations, and communicative purposes.

The development of multilingual neural machine translation has additionally expanded the possibilities of transferring knowledge between languages with different levels of available training data. Chen et al. examine cross-lingual transfer in zero-shot neural machine translation and demonstrate the importance of transfer mechanisms for improving translation between language pairs for which direct parallel training data may be limited [12, pp. 142–157]. This issue is particularly significant for low-resource languages because the amount and representativeness of parallel corpora directly influence the ability of neural systems to learn reliable lexical, morphological, and syntactic correspondences. The broader development of transformer-based natural language processing has further increased the capacity of computational models to represent long-distance dependencies and contextual relationships in multilingual text [13].

An important direction in the literature concerns the distinction between general-purpose language models and dedicated machine translation systems. Contemporary large language models are capable of multilingual generation, paraphrasing, summarization, dialogue, and other language-related tasks. Achiam et al. describe GPT-4 as a general-purpose multimodal model capable of performing a broad range of language and reasoning tasks [16]. Consequently, its translation output may differ from that of systems specifically optimized for machine translation. In particular, paraphrasing, lexical substitution, and syntactic restructuring can produce semantically acceptable translations while simultaneously reducing scores on reference-based metrics. This distinction is important when comparing ChatGPT with dedicated translation services because a lower automatic metric score does not necessarily indicate an equivalent degree of semantic inadequacy.

The methodological literature on automatic evaluation provides a theoretical basis for interpreting such differences. BLEU, proposed by Papineni et al., evaluates machine translation primarily through the degree of n-gram correspondence between a candidate translation and one or more reference translations [20, pp. 311–318]. Although BLEU became one of the most widely used automatic MT evaluation measures, its reliance on surface-level lexical overlap creates difficulties when legitimate paraphrases, synonyms, or alternative syntactic structures are used. Babych and Hartley proposed extending BLEU through frequency weighting in order to address some of the limitations associated with uniform treatment of lexical matches [18, pp. 621–628]. Chen and Cherry, in turn, demonstrated that sentence-level BLEU is sensitive to smoothing techniques, particularly when translations contain limited n-gram overlap with the reference [19, pp. 362–367]. These observations are directly relevant to cases in which a machine-generated translation conveys the source meaning but receives a low BLEU score because its lexical and syntactic formulation differs from the expert reference.

METEOR was developed partly to address some of these limitations by incorporating additional forms of lexical correspondence. Agarwal and Lavie discuss the use of METEOR and related evaluation measures for establishing stronger correlations with human judgments of translation quality [17, pp. 115–118]. Compared with purely n-gram-based evaluation, METEOR is more capable of recognizing certain forms of lexical variation and alignment. Therefore, the combined application of BLEU and METEOR can provide a more informative assessment of machine translation than reliance on a single metric. The use of syntactic or part-of-speech-based measures can provide an additional dimension by indicating the degree to which grammatical organization is preserved, although such measures cannot independently establish semantic or pragmatic equivalence.

The literature therefore demonstrates that automatic metrics should not be interpreted as complete substitutes for expert assessment. A high BLEU score may indicate substantial lexical and structural similarity to a reference translation without guaranteeing terminological appropriateness or contextual adequacy. Conversely, a low BLEU score may result from legitimate lexical or syntactic reformulation rather than from an actual loss of meaning. This methodological limitation becomes particularly important in scientific translation, where several linguistically different formulations may communicate essentially the same proposition. The literature on machine translation evaluation consequently supports the use of complementary quantitative and qualitative procedures rather than dependence on a single numerical indicator [17, pp. 115–118; 20, pp. 311–318].

Another significant research direction concerns specialized and domain-specific translation. Scientific and medical texts contain terminology, abbreviations, formulaic constructions, and professional conventions that require a high level of contextual precision. Comparative research on medical machine translation demonstrates that systems may differ in their treatment of specialized lexical units and syntactic constructions [3, pp. 139–163]. The problem becomes particularly pronounced when a source-language expression has several possible equivalents in the target language and the correct choice depends on the communicative situation. In such cases, translation quality cannot be determined solely by formal similarity; terminological consistency,

semantic adequacy, and conformity with professional discourse conventions must also be considered.

The development of multilingual and domain-adaptive neural translation has created new opportunities for addressing these limitations. Research into efficient multilingual fine-tuning indicates that model adaptation can improve the use of pretrained multilingual representations for specific translation tasks [14]. Cross-lingual transfer methods likewise provide opportunities to improve translation for language pairs with limited parallel resources [12, pp. 142–157]. At the same time, the increasing integration of large language models into language technologies has expanded the range of possible translation strategies, including contextual reformulation and interactive human–machine collaboration [13; 16]. These developments suggest that future MT systems will increasingly combine multilingual knowledge, domain adaptation, contextual modeling, and user-guided post-editing.

The literature also indicates the continuing importance of translation for low-resource and indigenous languages. Stap and Araabi, for example, draw attention to the limitations of ChatGPT as a translator for indigenous languages, demonstrating that general-purpose language models do not automatically provide equally reliable translation across languages with different levels of digital representation and available linguistic resources [30, pp. 163–166]. This observation is significant for the broader development of machine translation because technological progress measured on high-resource language pairs cannot necessarily be generalized to all linguistic communities.

Overall, previous research provides a substantial theoretical and methodological basis for examining contemporary machine translation. Existing studies have addressed the technological evolution of neural and transformer-based systems, comparative performance of individual translation services, domain-specific translation problems, multilingual transfer, and the development of automatic evaluation metrics. However, the literature also demonstrates that quantitative metrics such as BLEU and METEOR capture different dimensions of translation similarity and may produce substantially different evaluations for the same output. This creates a need for comparative research that combines several automatic metrics with expert qualitative analysis, particularly when assessing scientific texts in the Russian–English language pair. Such an approach makes it possible to examine not only lexical and grammatical correspondence but also contextual adequacy, terminological precision, communicative equivalence, and the preservation of discourse-level meaning.

Methodology

The study employs a comparative and analytical research design to evaluate the development and current performance of machine translation systems. Russian–English scientific and medical texts were translated using selected MT services, including Google, Yandex, DeepL, and ChatGPT. The generated translations were assessed using BLEU, METEOR, and POS-BLEU metrics, while expert qualitative analysis was applied to examine semantic adequacy, contextual appropriateness, syntactic organization, grammatical accuracy, and terminological consistency. Comparative analysis was used to identify differences among the systems and to determine the limitations of automatic evaluation metrics. Particular attention was given to cases in which quantitative scores did not fully correspond to the actual communicative and semantic quality of the translations.

Results

The comparative evaluation of the selected machine translation systems demonstrated measurable differences in their performance when translating scientific texts from Russian into English. The obtained results showed that Yandex and Google produced comparatively higher BLEU scores, indicating a greater degree of correspondence with the reference translations at the lexical and phrase-structure levels. The difference between the two systems was relatively small, suggesting comparable performance under the conditions of the conducted evaluation. ChatGPT demonstrated lower BLEU values, while its translations more frequently contained alternative lexical choices and restructured sentence patterns.

The METEOR evaluation produced a somewhat different distribution of results. Yandex and Google again demonstrated relatively strong performance, whereas DeepL showed lower results in the selected dataset. The divergence between BLEU and METEOR values indicates that the systems differed not only in their reproduction of reference wording but also in the way they represented equivalent meanings through alternative lexical forms. POS-BLEU provided an additional perspective by showing the extent to which the grammatical and part-of-speech organization of the source or reference translation was retained.

The detailed examination of individual translation samples revealed several recurrent types of variation. Differences were observed in the selection of specialized terminology, grammatical articles, syntactic relations, and the ordering of information within sentences. In scientific texts, these differences were generally concentrated in terminology and sentence-level formulation rather than in the complete loss of the principal meaning. Some outputs reproduced the formal structure of the reference translation more closely, whereas others used alternative constructions while retaining the central proposition.

The analysis of the selected example concerning intelligent assistants and automated assessment systems produced BLEU = 0.0000, METEOR = 0.4808, and POS-BLEU = 0.7440 for the Google translation. The three values demonstrate that the same translation can receive substantially different evaluations depending on the criterion applied. In particular, the zero BLEU value resulted from the absence of sufficient matching n-gram sequences with the reference translation, whereas the METEOR and POS-BLEU values reflected greater correspondence at the lexical-semantic and grammatical levels.

The qualitative comparison also identified differences in the treatment of educational terminology. The expression “системы проверки” was translated as “checking systems,” while the expert version used “assessment tools.” Although both formulations refer to systems designed to support evaluation, the latter is more closely associated with established terminology in educational discourse. Similar cases were observed where the machine-generated version was grammatically acceptable but required terminological adjustment to conform more precisely to the conventions of the relevant scientific field.

The additional evaluation of a medical scientific text resulted in generally high scores across the three applied metrics, with a mean value of 0.6896. The result demonstrates that the selected systems were capable of producing relatively coherent English versions of specialized scientific material. At the same time, the detailed assessment showed that high aggregate scores did not eliminate individual problems involving medical terminology, abbreviations, lexical selection, and context-dependent expressions.

A comparison between texts accompanied by previously published expert translations and texts for which the reference translation had not been publicly available revealed no consistent performance difference attributable solely to the online availability of the reference material. The systems generated translations for both categories of texts, while the observed variations were more closely associated with linguistic and terminological characteristics of individual passages. This observation provides no direct empirical basis for concluding that the tested outputs were simply reproduced from publicly accessible translations.

The results further showed that scientific discourse generally provided favorable conditions for automated translation because of its relatively stable terminology, recurrent grammatical structures, and explicit logical organization. Nevertheless, individual sentences containing ambiguous lexical units, specialized terminology, or context-dependent constructions produced greater variation among the systems. The results therefore indicate that MT performance is not uniform across text segments and depends substantially on the linguistic complexity and communicative characteristics of the material being translated.

Taken together, the empirical results demonstrate that contemporary machine translation systems are capable of producing functionally usable English translations of scientific material, but their performance varies according to the evaluation metric, text segment, and linguistic characteristics of the source. The numerical assessment and qualitative examination also revealed that translation quality cannot be represented adequately by a single score. The observed discrepancies between automatic measurements and expert judgments provide an empirical basis for combining computational evaluation with detailed linguistic examination in the assessment of machine-generated scientific translations.

Discussion

The study findings indicate that differences among machine translation (MT) services can substantially affect translation quality. The application of three complementary evaluation metrics–BLEU, METEOR, and POS-BLEU–provided a multidimensional assessment of the performance of these systems in Russian-to-English translation.

The findings further demonstrate that automatic MT evaluation metrics are highly dependent on the reference, or expert, translation. Consequently, scores may be artificially reduced when a machine-generated translation differs substantially from the reference at the stylistic or lexical level despite conveying an acceptable meaning. This limitation is particularly evident in BLEU, which primarily measures n-gram overlap between the candidate and reference translations and therefore has limited sensitivity to legitimate synonymy and paraphrasing. Dependence on a single reference translation can consequently reduce the validity of BLEU as an independent measure of overall translation quality, particularly when it is used without complementary semantic or linguistic evaluation metrics.

A representative example illustrates this limitation:

Source text: “Существенной помощью может стать использование интеллектуальных помощников преподавателя и автоматизированных систем проверки, построенных методами машинного обучения и технологии нейронных сетей.”

Expert translation: “Smart educator assistants and automated assessment tools based on machine learning and neural networks can significantly alleviate the problem.”

Google translation: “The use of intelligent teacher assistants and automated checking systems built using machine learning methods and neural network technology can be of significant help.”

The corresponding evaluation scores were BLEU = 0.0000, METEOR = 0.4808, and POS-BLEU = 0.7440.

In this case, BLEU assigned a score of zero, demonstrating that lexical-overlap-based evaluation may fail to represent the actual degree of semantic and grammatical adequacy of a translation. The relatively high POS-BLEU score indicates that much of the grammatical and part-of-speech structure was preserved; however, grammatical similarity alone cannot establish full translational equivalence. METEOR proved more informative in this example because it is comparatively more sensitive to lexical variation, synonymy, and paraphrasing [1]; [2].. Although Google's translation largely preserves the general propositional meaning of the source text, it does not achieve complete lexical and terminological equivalence. For example, the Russian expression системы проверки was rendered as checking systems. In the given academic context, this formulation is less terminologically appropriate than assessment tools, which more accurately reflects the intended function and corresponds more closely to established educational terminology. This example demonstrates the importance of supplementing quantitative metrics with qualitative expert analysis of terminological accuracy and contextual adequacy.

To strengthen the empirical component of the study, an additional text from a peer-reviewed medical article was included, together with an article accompanied by an expert translation that had not previously been published online. These materials were selected to evaluate MT performance across different contexts and levels of lexical and terminological complexity. The results obtained using the three evaluation metrics were generally high, with a mean score of 0.6896, indicating a relatively strong overall level of translation performance.

Importantly, no systematic relationship was identified between translations of articles whose expert translations were already publicly available online and translations of previously unpublished materials. Therefore, the hypothesis that MT systems simply reproduce or retrieve existing online translations was not supported by the empirical findings. Instead, the results suggest that the systems were capable of generating translations for texts whose reference versions had not previously been available online. Nevertheless, establishing the precise extent to which training-data exposure may influence individual outputs would require controlled experiments and access to information about the models’ training datasets.

The relatively high effectiveness of the MT systems may also be associated with the functional and stylistic characteristics of the selected texts. Scientific discourse typically exhibits a comparatively standardized structure, relatively stable grammatical patterns, explicit logical relations, and frequent use of international or internationally recognizable terminology. These characteristics can facilitate automated translation. Scientific texts also tend to contain less implicit meaning and fewer highly context-dependent lexical ambiguities than many literary or colloquial genres. However, the translation of medical diagnoses, clinical conclusions, specialized terminology, abbreviations, and context-sensitive medical expressions remains substantially more challenging.

Overall, contemporary MT services demonstrated an adequate level of performance in translating scientific texts. Within the Russian–English language pair examined in this study, Yandex achieved the strongest overall results according to the selected evaluation metrics. The continued development of machine translation technologies has expanded opportunities for international scientific communication, knowledge exchange, and dissemination of research findings. Nevertheless, expert post-editing remains necessary, particularly in specialized domains, because human specialists are able to identify terminological, contextual, pragmatic, and discourse-level errors that automated systems may not interpret reliably.

Combining automatic evaluation with expert assessment can improve the reliability and comprehensiveness of translation-quality analysis. Automated metrics and computational tools enable researchers to identify potential discrepancies efficiently and process large quantities of translation data, thereby reducing the burden of routine evaluation. Human translators and domain experts can subsequently concentrate on semantically significant errors, terminological inconsistencies, pragmatic inadequacies, and stylistic deviations. Thus, integrating machine-based assessment with expert evaluation contributes to a more efficient, systematic, and reliable translation-quality assurance process.

Machine translation has become an important component of globalized information exchange by facilitating multilingual access to scientific, educational, professional, and other forms of knowledge. Contemporary MT systems based on neural architectures and transformer models have achieved substantial progress in processing complex linguistic structures. Nevertheless, significant challenges remain, particularly in the translation of culture-specific meanings, idiomatic expressions, specialized terminology, and low-resource languages.

Advances in multilingual language modeling, transfer learning, domain adaptation, and active learning provide promising opportunities for improving translation quality, particularly for low-resource languages, dialects, and specialized domains. The future development of machine translation is also likely to involve increasingly adaptive, hybrid, and interactive systems capable of incorporating user context, domain-specific requirements, and textual characteristics into the translation process.

Automatic translation-quality evaluation methods are likewise undergoing continuous development. Incorporating syntactic, semantic, contextual, and discourse-sensitive parameters into evaluation frameworks can provide a more comprehensive representation of translation quality than reliance on lexical overlap alone. Such multidimensional approaches make it possible to evaluate not only formal correspondence but also semantic adequacy, grammatical acceptability, terminological precision, and contextual appropriateness.

Conclusion

Thus, further research is required to improve lexical analysis, morphological processing, syntactic modeling, contextual interpretation, and the translation of rare or semantically complex lexical units, particularly in highly specialized domains. Machine translation evaluation should therefore be conceptualized as a multidimensional procedure. The combined application of metrics such as BLEU, METEOR, and POS-BLEU can provide a more balanced assessment than the use of any single metric in isolation. At the same time, quantitative evaluation should be complemented by expert qualitative assessment. Integrating automatic metrics with human evaluation makes it possible to identify errors that cannot be adequately captured through surface-level lexical or grammatical correspondence. Such an integrated evaluation framework can contribute to improving translation accuracy, adequacy, terminological consistency, and overall reliability. Machine translation continues to evolve rapidly and has demonstrated considerable progress in processing increasingly complex linguistic structures. Nevertheless, continued research is necessary to improve its adaptability to low-resource languages, specialized discourse, and culturally conditioned meanings. Particular attention should be given to maintaining semantic accuracy, contextual equivalence, and communicative adequacy across different language pairs and domains.

References

1. Карцева Е. Ю., Маргарян Т. Д., Гурова Г. Г. Развитие машинного перевода и его место в профессиональной межкультурной коммуникации //Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. – 2016. – №. 3. – С. 155-164.

2. Borisov A., Galinskaya I. Y. Yandex School of Data Analysis Russian-English Machine Translation System for WMT14 // Proceedings of the Ninth Workshop on Statistical Machine Translation. – 2014. – P. 66–70.

3. Cambedda G., Di Nunzio G. M., Nosilia V. A Study on Automatic Machine Translation Tools: A Comparative Error Analysis Between DeepL and Yandex for Russian-Italian Medical Translation // Umanistica Digitale. – 2021. – No. 10. – P. 139–163.

4. Kulikov V., Kulikova V., Yerkebulan G. Google/Yandex Translation Detection in the Patterns Identifying System of Multilingual Texts // International Journal of Computers. – 2021. – Vol. 20, No. 1. – P. 72–77.

5. Kamaluddin M. I. et al. Accuracy Analysis of DeepL: Breakthroughs in Machine Translation Technology // Journal of English Education Forum (JEEF). – 2024. – Vol. 4, No. 2. – P. 122–126.

6. Linlin L. Artificial Intelligence Translator DeepL Translation Quality Control // Procedia Computer Science. – 2024. – Vol. 247. – P. 710–717.

7. Sriram A. et al. Cold Fusion: Training Seq2Seq Models Together with Language Models. – 2017. – arXiv:1708.06426.
8. Egonmwan E., Chali Y. Transformer and Seq2Seq Model for Paraphrase Generation // Proceedings of the 3rd Workshop on Neural Generation and Translation. – 2019. – P. 249–255.
9. Li Z. et al. Seq2Seq Dependency Parsing // Proceedings of the 27th International Conference on Computational Linguistics. – 2018. – P. 3203–3214.
10. Sutskever I. Sequence to Sequence Learning with Neural Networks. – 2014. – arXiv:1409.3215.
11. Ni J. et al. Sentence-T5: Scalable Sentence Encoders from Pretrained Text-to-Text Models. – 2021. – arXiv:2108.08877.
12. Chen G. et al. Towards Making the Most of Cross-Lingual Transfer for Zero-Shot Neural Machine Translation // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. – 2022. – Vol. 1. – P. 142–157.
13. Rothman D. Transformers for Natural Language Processing: Build, Train, and Fine-Tune Deep Neural Network Architectures for NLP with Python, Hugging Face, and OpenAI's GPT-3, ChatGPT, and GPT-4. – Birmingham : Packt Publishing Ltd, 2022.
14. Maltas S. I. et al. Efficient Finetuning Strategies for Multilingual Neural Machine Translation. – Barcelona : Universitat Politècnica de Catalunya, 2024.
15. Bengesi S. et al. Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers // IEEE Access. – 2024. – Vol. PP, No. 99. – P. 1–1.
16. Achiam J. et al. GPT-4 Technical Report. – 2023. – arXiv:2303.08774.
17. Agarwal A., Lavie A. METEOR, M-BLEU and M-TER: Evaluation Metrics for High-Correlation with Human Rankings of Machine Translation Output // Proceedings of the Third Workshop on Statistical Machine Translation. – 2008. – P. 115–118.
18. Babych B., Hartley T. Extending the BLEU MT Evaluation Method with Frequency Weightings // Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04). – 2004. – P. 621–628.
19. Chen B., Cherry C. A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU // Proceedings of the Ninth Workshop on Statistical Machine Translation. – 2014. – P. 362–367.
20. Papineni K. et al. BLEU: A Method for Automatic Evaluation of Machine Translation // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. – 2002. – P. 311–318.
21. Pérez D., Alfonseca E. Application of the BLEU Algorithm for Recognising Textual Entailments // Proceedings of the First Challenge Workshop Recognising Textual Entailment. – 2005. – P. 9–12.
22. 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.
23. Сарсенбаева З. Модернизм в узбекской литературе и интерпретация образов //Зарубежная лингвистика и лингводидактика. – 2024. – Т. 2. – №. 1. – С. 193-199.
24. 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.
25. 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.
26. Sarsenbaeva Z. MODERN APPROACHES TO TEACHING LITERARY TERMS IN COMPARATIVE LITERATURE //Mental Enlightenment Scientific-Methodological Journal. – 2026. – Т. 7. – №. 02. – С. 198-209.
27. Belgibayeva G., Sarsenbaeva Z., Zhumagulova K. MAG 'JAN JUMABAYEV IJODIDA OZODLIK G 'OYASI //Ijtimoiy-gumanitar sohada ilmiy-innovatsion tadqiqotlar. – 2026. – Т. 3. – №. 1. – С. 159-165.
28. Sarsenbayeva Z., Zillolova G. COMPARATIVE ANALYSIS OF TWO TRANSLATED VERSIONS OF A LITERARY TEXT //Ijtimoiy-gumanitar sohada ilmiy-innovatsion tadqiqotlar. – 2025. – Т. 2. – №. 4. – С. 183-188.
29. Sarsenbayeva Z., Zillolova G. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДВУХ ПЕРЕВОДНЫХ ВЕРСИЙ ХУДОЖЕСТВЕННОГО ТЕКСТА //Scientific and innovative research in the social and humanitarian sphere. – 2025. – Т. 2. – №. 4. – С. 183-188.
30. Stap D., Araabi A. ChatGPT Is Not a Good Indigenous Translator // Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP). – 2023. – P. 163–166.