UDC 004.932.2:004.8:004.58 DOI https://doi.org/10.35546/kntu2078-4481.2025.3.2.15

A. O. HAVRASHENKO
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-8802-0529

O. YU. BARKOVSKA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0001-7496-4353

ASSESSMENT OF THE ACCURACY OF ARTIFICIAL LANGUAGE TRANSLATION METHODS

The article is devoted to evaluating the accuracy of methods for translating artificial languages in an intelligent information system for translating artificial languages using artificial intelligence methods and statistical algorithms, proposing an innovative approach to improving the effectiveness of machine translation in the context of the dynamic development of modern technologies. The analysis showed insufficient translation accuracy, which was proposed to be improved by integrating two methods into one system.

The aim of the work is to evaluate the accuracy of translation methods in the proposed intellectual information system for translating artificial languages. The work examines such translation evaluation metrics as BLEU, ChrF, BLEURT, COMMET, TER, METEOR and human evaluation (as a benchmark). Each of the algorithms (mathematical models, neural network algorithm and combined algorithm) was tested on 12 tests, each of which involved translating one of four languages into another (English, Ukrainian, French and Esperanto) and vice versa. The results show an increase in average translation accuracy of 0.5 % compared to the method based on mathematical models and 0.2 % compared to the method based on neural networks. In some pairs, the result was worse than the individual algorithm, but this was due to the coefficients used when combining the results. Since individual settings are possible for each language pair, the correct result will appear in the worst case with the same accuracy as the best of the two results. However, this requires separate configuration. Based on the results obtained, it can be concluded that the combination of these two methods improves the accuracy of artificial language translation. The results confirm the effectiveness of the developed system, which allows for highly efficient translation of natural and artificial languages.

Key words: *machine learning, artificial intelligence, LSTM, data networks, sequence-to-sequence, translation, language, method, metric, mathematical model.*

А. О. ГАВРАШЕНКО
аспірант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-8802-0529

О. Ю. БАРКОВСЬКА
кандидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0001-7496-4353

ОЦІНКА ТОЧНОСТІ МЕТОДІВ ПЕРЕКЛАДУ ШТУЧНИХ МОВ

Стаття присвячена оцінці точності методів перекладу штучних мов в інтелектуальній інформаційній системі перекладу штучних мов із використанням методів штучного інтелекту та статистичних алгоритмів, пропонуючи інноваційний підхід до підвищення ефективності машинного перекладу в умовах динамічного розвитку сучасних технологій. Проведений аналіз показав недостатню точність перекладу, підвищити яку було запропоновано шляхом інтеграції двох методів в одну систему.

Метою роботи є оцінка точності методів перекладу в запропонованій інтелектуальної інформаційної системи перекладу штучних мов. В роботі досліджено такі метрики оцінки перекладу, як BLEU, ChrF, BLEURT, COMMET, TER, METEOR та людська оцінка (в якості еталону). Кожен з алгоритмів (математичні моделі, нейромережевий алгоритм та комбінований алгоритм) були перевірені на 12 тестах, кожен з яких – переклад *однієї з чотирьох мов в іншу (англійська, українська, французька та есперанто) та в зворотному порядку. Результати демонструють, що дає підвищення середньої точності перекладу на 0.5 % у порівнянні з методом на основі математичних моделей, на 0.2 % у порівнянні з методом на основі нейронних мереж. В деяких парах результат був гірше за окремо взятий алгоритм, але це відбулося через використані коефіцієнти при складені результатів. Так як для кожної окремої пари мов можливо індивідуальне налаштування, то правильний результат буде з'являтися в найгіршому випадку з тією ж точністю, що і найкращий серед 2 результатів. Але це потребує окремого налаштування. Враховуючи отримані результати, можна зробити висновок, що комбінація цих двох методів дозволяє покращити точність перекладу штучних мов. Отримані результати підтверджують ефективність розробленої системи, що дозволяє з високою ефективністю перекладати звичайні та штучні мови.*

Ключові слова: *машинне навчання, штучний інтелект, LSTM, датасети, послідовність у послідовність, переклад, мова, метод, метрика, математична модель.*

Formulation of the problem

Translation quality assessment is a fundamental element in both the practical work of translators and modern research in the field of automatic language processing. [1] It determines the level of compliance of the resulting text with user expectations, communication needs and the norms of the target language. This process is complex in nature, as it involves the analysis of a whole range of parameters covering the accuracy of content transmission, compliance with linguistic and cultural norms, stylistic consistency, and the overall comprehensibility of the text. This makes evaluation a universal tool for quality control and, at the same time, a benchmark for further improvement of translation technologies.

In scientific and engineering practice, particular attention is paid to systematic approaches to evaluation. They allow avoiding subjectivity and form a single basis for comparing different methods or models. On the one hand, translation quality assessment acts as a diagnostic mechanism: it helps to identify weaknesses related to vocabulary choice, syntactic structure, or grammatical norms. On the other hand, it acts as a driving force for progress, as systematic observations and analysis results are used to create new algorithms capable of producing more natural and accurate translations.

Modern approaches to evaluation are largely focused on automation. This is due to the rapid development of neural networks and the growing use of translation technologies, where traditional verification methods are becoming too labour-intensive. Automated evaluation allows for the effective analysis of large amounts of data, while ensuring speed and repeatability of results. However, linguistic and qualitative assessments remain relevant, as they take into account nuances that are difficult to formalise.

In general, translation quality assessment has a dual significance. It is a practical tool for control and standardisation, ensuring that a text complies with certain norms, as well as a scientific method for developing and improving modern translation systems. Thus, it is through evaluation that a balance is achieved between technical capabilities, linguistic requirements and end-user expectations.

This work is devoted to evaluating the accuracy of artificial language translation methods. It examines various automatic methods of assessing the quality of text translation. Particular attention is paid to comparing different types of artificial language translation methods and combining them into a single system.

The scientific novelty lies in evaluating the accuracy of artificial language translation methods in the proposed intellectual information system for artificial language translation. The practical significance of the proposed system is the evaluation of the effectiveness of the method of translating artificial languages based on mathematical models and the method of translating artificial languages based on neural network translation, and their comparison and integration to ensure high accuracy and reliability of results.

Analysis of the latest research and publications

Evaluating the quality of machine translation is a critically important step in the research and practical application of automatic translation systems. Without reliable measurement methods, it is difficult to understand how well the system performs the task at hand, whether its performance has improved after modifications, and how it compares to other approaches. Therefore, over the past decades, a whole range of methods and metrics have been developed to evaluate translations both automatically and with the help of experts.

All metrics for evaluating the accuracy of machine translation systems can be divided into two broad groups (Fig. 1):

1. Automated metrics – algorithms that compare machine translation with one or more reference translations created by humans. They allow for the rapid processing of large amounts of data and are widely used in research.
2. Human evaluation – experts or native speakers evaluate the translation according to criteria such as comprehensibility, grammatical correctness and content adequacy. This approach is considered the most accurate, but it is expensive and labour intensive.

Automated metrics cannot completely replace human evaluation, but they provide a fast and relatively objective way to compare different models.

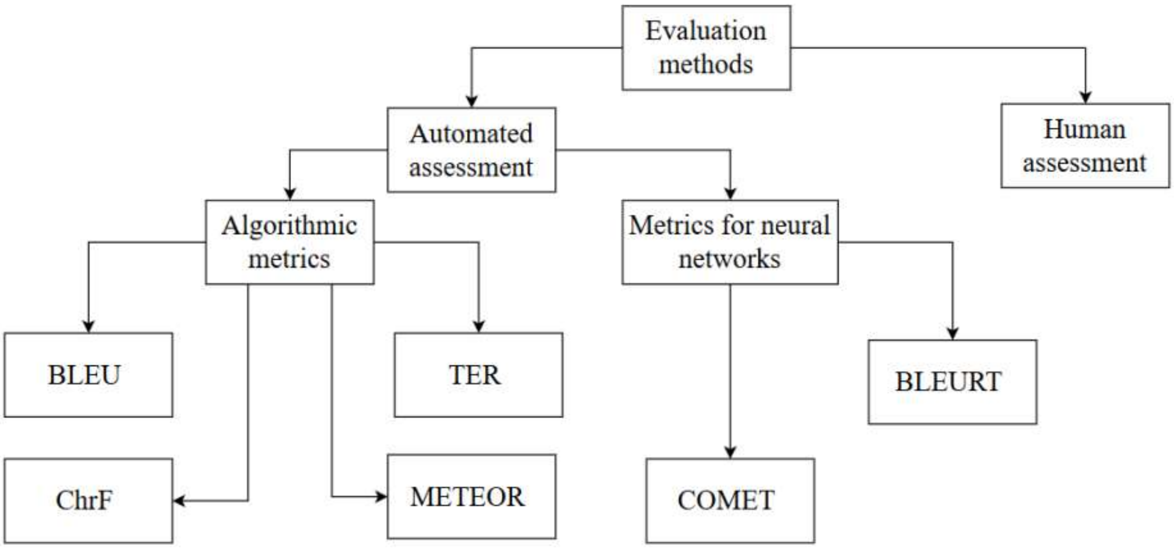

Fig. 1. Classification tree of metrics for evaluating the accuracy of machine translation systems

BLEU is one of the first and most well-known metrics, proposed in 2002 by IBM researchers. It is based on the idea of counting n-grams (sequences of n words) in a machine translation and in a so-called ‘reference translation’ (a benchmark created by a human). If the machine translation contains a large number of matches with the reference, the result is considered high quality.

A distinctive feature of BLEU is the penalty for overly short translations. For example, if the system attempts to translate a long sentence with only a few words that match the reference, it could receive a high score without the penalty. The penalty reduces the metric value in such cases, encouraging the completeness of the translation.

ChrF is a metric that focuses not on words, but on character sequences. [5, 6] It works well for languages with rich morphology, where changes in endings or word formation can significantly alter the appearance of a word.

The principle of operation is to count character matches (n-grams) between the machine translation and the reference. Thanks to this, even if the words differ but have a common root, the metric will consider the similarity.

The METEOR metric appeared in 2005 as a response to the limitations of BLEU. [7] It considers not only exact word matches, but also morphological variations, synonymy, and word order in a sentence. This makes it closer to human perception.

The principle of METEOR is to find correspondences between the words of the machine translation and the reference text. To do this, different levels of comparison are used:

- exact word matches;
- stemming (matching word roots, e.g., ‘translate’ and ‘translated’);
- synonymy (thanks to dictionaries such as WordNet);
- permutation (assessing the impact of incorrect word order).

As a result, METEOR correlates better with human evaluation than BLEU, especially on short texts.

TER was developed in 2006 as a tool to show how many edits a human would need to make to correct a machine translation. These edits may include:

- insertions;
- deletions;
- word replacements;
- fragment rearrangements.

Therefore, TER essentially measures the ‘cost of correction.’ For example, if only one word needs to be changed to edit a sentence, TER will be low, but if most of the text needs to be rewritten, the value will be high.

In the 2020s, a new class of metrics emerged that use contextual representations of words obtained from language models.

BLEURT is a modern metric based on neural networks. It uses pre-trained transformers (BERT and its modifications), which allow evaluating not only the formal similarity of texts, but also their semantic proximity.

BLEURT undergoes preliminary training on a large amount of text and then undergoes additional training on corpora with human translations. This makes it more ‘human-like’ in its evaluation.

BLEURT is a trained model that combines large corpora of synthetic data and human evaluations. Thanks to this, it is able to take into account both lexical matches and overall semantic adequacy.

COMET is another modern metric developed on the basis of transformers. It uses multilingual models that allow comparing texts in different languages without the need for direct word matches [10].

A distinctive feature of COMET is that it takes into account three texts simultaneously:

1. the original (source text);
2. the machine translation;
3. the reference translation.

This enables COMET to better understand whether a translation corresponds to the original content, rather than just formally resembling the reference.

Despite significant progress in the development of automatic metrics, they cannot completely replace human evaluation. Research often uses the MQM (Multidimensional Quality Metrics) framework, which classifies errors by type (lexical, grammatical, stylistic) and severity (minor, serious, critical). MQM provides a detailed picture of the strengths and weaknesses of the system, and automatic metrics are used as a supporting tool. Not only the translation of words, but also the conveyance of meaning, style, and intention of the statement.

Formulation of the purpose of the research

The aim of this work is to evaluate the accuracy of methods for translating artificial languages in the proposed intellectual information system for translating artificial languages. The research component of the proposed system is to evaluate the effectiveness of the method of translating artificial languages based on mathematical models and the method of translating artificial languages based on neural network translation, and to compare and combine them in order to ensure high accuracy and reliability of results.

To achieve this goal, the following tasks must be solved:

– analysis of metrics for evaluating the accuracy of automated translation systems;
– development of experimental methodology;
– creation of a model for an intelligent information system for translating artificial languages;
– evaluation of each method separately, their comparison and combination.

The experiments conducted form the basis for further improvement of methods for translating artificial languages.

The limitations of the system include the need to use a parallel corpus of texts in different languages. Another limitation is the need to train the system on the words that will be used for translation.

Research Results and Discussions

This study was the first to compare the method of translating artificial languages based on mathematical models with the method of translating artificial languages based on neural networks in the proposed artificial language translation system.

To test the effectiveness of the developed approach, an experiment was conducted involving four languages: Ukrainian, English, French, and Esperanto. These languages were chosen because of their different linguistic characteristics. English is

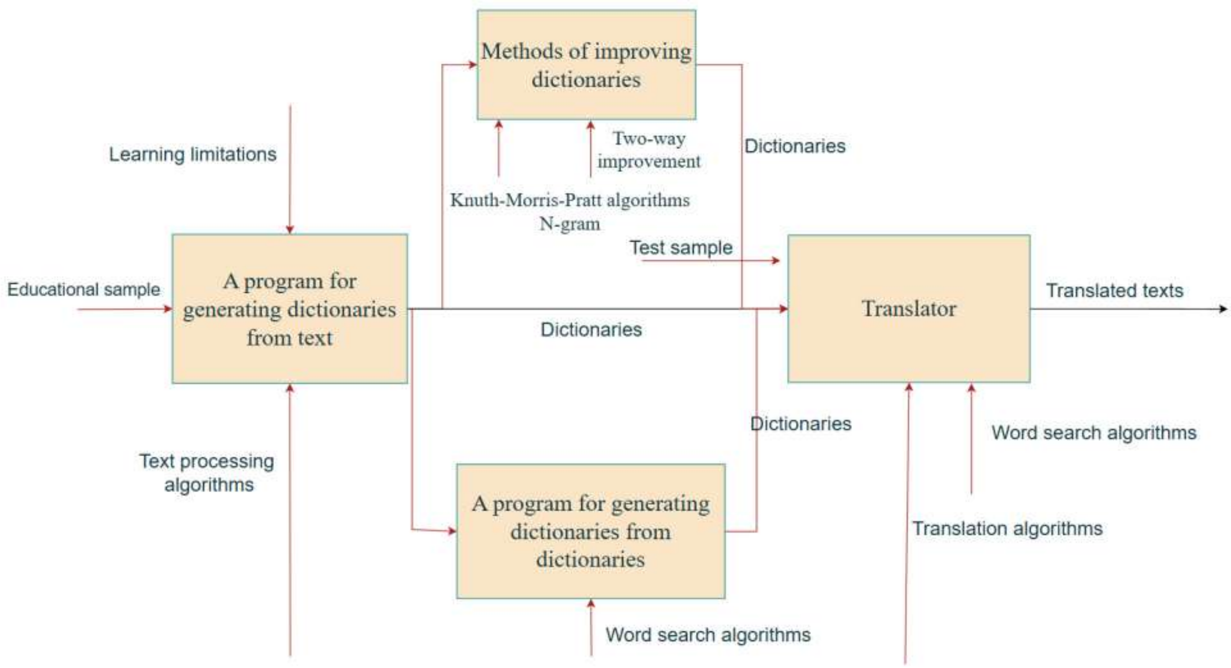

Fig. 2. Method of translating artificial languages based on mathematical models

relatively easy to process by machines, French and Ukrainian have developed morphology, and Esperanto is distinguished by its simplified grammar. This made it possible to evaluate the universality of the system.

The experiment used a developed set of programs consisting of two main components.

A module based on mathematical models (Fig. 2) implements translation using statistical methods. It processes input texts using a set of modules, each of which contains an algorithm based on a mathematical model. After that, the aggregate result is processed, and the resulting statistical translation dictionary is generated. [11, 12]

The method of translating artificial languages based on neural networks (Fig. 3) is built on modern machine learning architectures, capable of taking context into account and identifying hidden patterns.

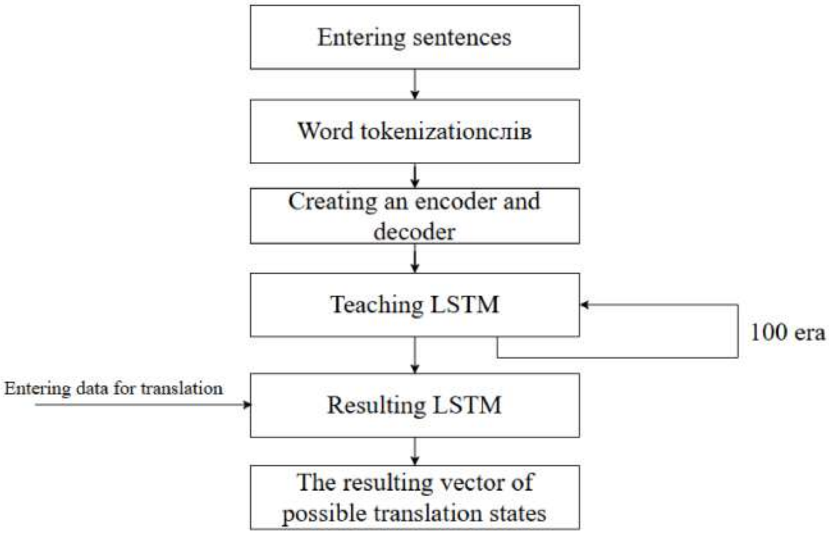

Fig. 3. Method of translating artificial languages based on neural networks

These methods were integrated into a single intelligent information system for translating artificial languages (Fig. 4): pre-processing was performed by an algorithmic block, after which the results were refined by a neural network. This approach combines the rigour of rules with the flexibility of adaptive models.

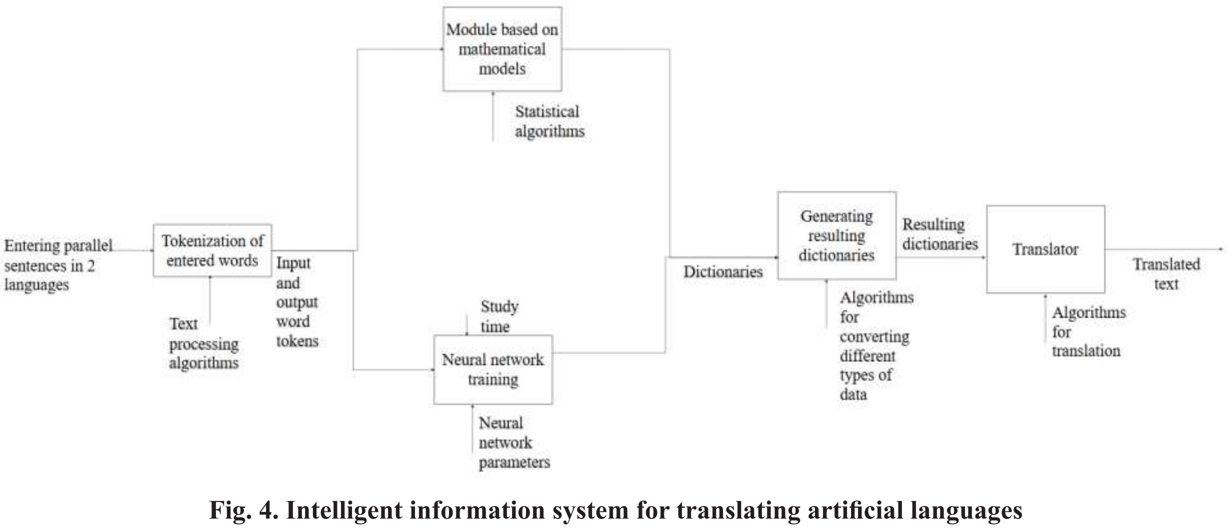

Fig. 4. Intelligent information system for translating artificial languages

For the study, a corpus of sentences on various topics was compiled: everyday dialogues, scientific descriptions, and literary excerpts. Each sentence was translated from Ukrainian into English, French, and Esperanto, as well as in the reverse direction. As a result, a complete set of texts for comparison was obtained.

The METEOR methodology was chosen for evaluation, which takes into account the accuracy of the translation of words, their morphemes and permutations, which is suitable for a complete evaluation of both translation options. The result was reduced to a single value in the range from 0 to 1, where 1 is the ideal option with which the automatic translation is compared (human translation). Each incorrect position, permutation, synonym, or incorrect morpheme reduces the score by a certain coefficient. After that, this value was converted into a percentage for convenience. That is, the result will be the percentage of coincidence with the reference translation.

The quality of the translation was checked in pairs of languages. These include English, Ukrainian (local dialect), French and Esperanto.

The tables show the results of the experiments. The row indicates the source language, the column indicates the target language. The number of the experiment is indicated in brackets.

The results of the method of translating artificial languages based on mathematical models are given in (Table 1).

Table 1

Results of the method of translating artificial languages based on mathematical models

	English	Ukrainian	French	Esperanto
English	–	89.4 % (1)	92.9 % (3)	93.5 % (5)
Ukrainian	90.9 % (2)	–	88.7 % (7)	91.1 % (9)
French	92.7 % (4)	89.2 % (8)	–	89.9 % (11)
Esperanto	92.1 % (6)	92.7 % (10)	89.0 % (12)	–

The results show a fairly consistent high level of translation quality across all languages considered, which indicates the consistency and reliability of the mathematical model used. The system works best in directions involving English: from English to other languages (92.9–93.5 %) and from Ukrainian to English (90.9 %). Slightly lower values are observed for translations where English is not involved, in particular between French and Ukrainian (89.2 %) or French and Esperanto (89.0 %). This may indicate more complex interactions between these languages and fewer indirect connections in the model.

Overall, the indicators remain within the range of 88–94 %, which characterises the system as stable, with high translation quality and minor fluctuations depending on the language pair.

Results of the neural network-based artificial language translation method in (Table 2).

Table 2

Results of artificial language translation based on the neural network method

	English	Ukrainian	French	Esperanto
English	–	89.7 % (1)	92.2 % (3)	92.8 % (5)
Ukrainian	90.3 % (2)	–	89.6 % (7)	90.5 % (9)
French	91.1 % (4)	89.2 % (8)	–	92.9 % (11)
Esperanto	92.7 % (6)	92.1 % (10)	93.0 % (12)	–

In the case of the neural network method, the results also remain high and balanced. The best results are observed for translations from Esperanto into other languages, as well as between English and French. The values for Ukrainian-French are slightly lower, but the difference is insignificant. Overall, the system demonstrates stable quality with a tendency to improve in pair combinations directly with Esperanto.

After combining the methods by combining different types of data with certain coefficients, the following results were obtained (Table 3).

Table 3

Results of the intellectual information system for translating artificial languages

	English	Ukrainian	French	Esperanto
English	–	89.6 % (1)	93.0 % (3)	93.8 % (5)
Ukrainian	90.1 % (2)	–	89.6 % (7)	90.7 % (9)
French	92.9 % (4)	88.8 % (8)	–	91.9 % (11)
Esperanto	92.7 % (6)	92.8 % (10)	91.2 % (12)	–

The results of comparing all methods can be seen in (Fig. 5).

As we can see from these results, on average, the combined result shows the best performance. The maximum accuracy achieved was 92.9 %. In some pairs, the result was worse than the separate algorithm, but this was due to the coefficients used when combining the results. Since each individual language pair can be configured separately, the correct result will appear in the worst case with the same accuracy as the best of the two results. However, this requires separate configuration.

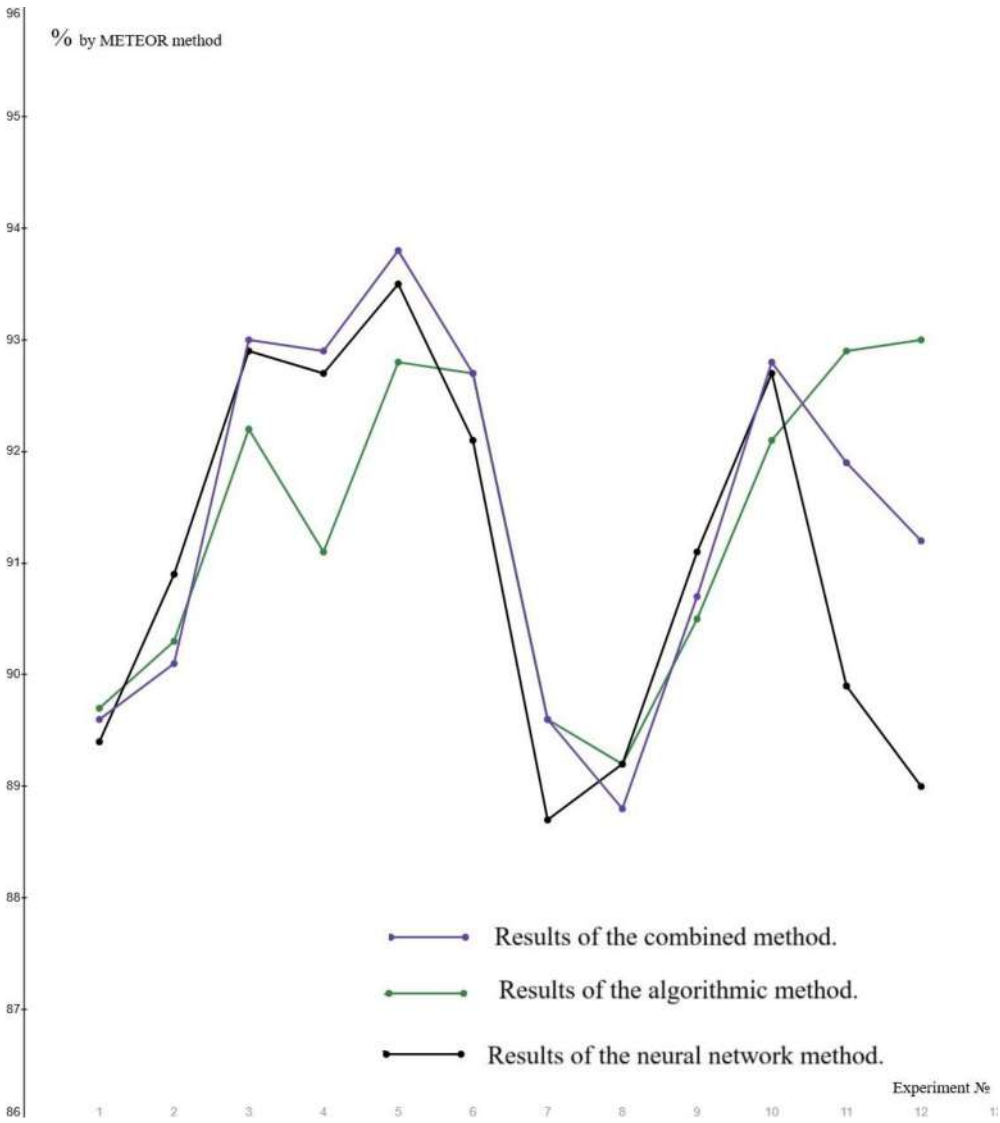

Fig. 5. results of comparison of all methods in the form of a graph

Based on the results obtained, we can conclude that the combination of these two methods allows us to create the best translation method for artificial languages.

Conclusions

The study was the first to evaluate the accuracy of artificial language translation methods in the proposed intellectual information system for translating artificial languages. The study achieved its objectives and solved the tasks set.

A detailed study and comparison of translation quality assessment methods was conducted. This made it possible to determine which translation quality assessment methods are best suited for artificial languages.

All modules of the system were studied, namely neural network translation, algorithmic translation, and combined translation.

As a result, we found that all methods work on all types of selected languages and show results of up to 92.9 % translation accuracy on a randomly selected dataset.

Based on the results obtained, we can conclude that the combination of these two methods allows us to create the best translation method for artificial languages. However, in some cases, the combination may worsen the result of a single method.

The scientific significance of the work lies in deepening the understanding of the mechanisms underlying neural networks and algorithmic methods. The conclusions presented can serve as a basis for further research in the field of machine learning and the development of intelligent systems, which will contribute to progress in the field of artificial intelligence and machine translation.

The practical value of this research lies in the creation of an intelligent translation information system that is independent of the source languages. This will allow automatic translation to be used for common, artificial and slang languages without additional costs.

Bibliography

1. House, Juliane. Translation quality assessment: Past and present. Routledge, 2014.
2. Han, Lifeng, Gareth JF Jones, and Alan F. Smeaton. “Translation quality assessment: A brief survey on manual and automatic meth-ods”. arXiv preprint arXiv:2105.03311 (2021).
3. Rivera-Trigueros, Irene. “Machine translation systems and quality assessment: a systematic review”. Language Resources and Evaluation 56.2 (2022): 593–619.
4. Papineni, Kishore, et al. “Bleu: a method for automatic evaluation of machine translation”. Proceedings of the 40th annual meeting of the Asso-ciation for Computational Linguistics. 2002.
5. Popović, Maja. “chrF: character n-gram F-score for automatic MT evaluation”. Proceedings of the tenth workshop on statistical machine transla-tion. 2015.
6. Popović, Maja. “chrF++: words helping character n-grams”. Proceedings of the second conference on machine translation. 2017.
7. Lavie, Alon, and Michael J. Denkowski. “The METEOR metric for automatic evaluation of machine translation”. Machine translation 23.2 (2009): 105–115.
8. Agarwal, Abhaya, and Alon Lavie. “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.
9. Mukherjee, Aniruddha, et al. “A Detailed Comparative Analysis of Automatic Neural Metrics for Machine Translation: BLEURT & BERTScore”. IEEE Open Journal of the Computer Society (2025).
10. Rei, Ricardo, et al. “COMET: A neural framework for MT eval-uation”. arXiv preprint arXiv:2009.09025 (2020).
11. Гаврашенко, А.., Барковська, О.. (2023). Аналіз алгоритмів аугментації тексту в системах машинного перекладу штучних мов. *Сучасні інформаційні системи*, *7*(1), 47–53. https://doi.org/10.20998/2522-9052.2023.1.08
12. Барковська, О., Гаврашенко, А.. (2023). АНАЛІЗ АЛГОРИТМІВ ПОШУКУ СЛІВ У СЛОВНИКАХ СИСТЕМ МАШИННОГО ПЕРЕКЛАДУ ДЛЯ ШТУЧНИХ МОВ. Комп’ютерні системи та інформаційні технології, (2), 17–24. https://doi.org/10.31891/csit-2023-2-2

References

1. House, Juliane. *Translation quality assessment: Past and present*. Routledge, 2014.
2. Han, Lifeng, & Gareth JF Jones, and Alan F. Smeaton. «Translation quality assessment: A brief survey on manual and automatic methods». *arXiv preprint arXiv:2105.03311* (2021).
3. Rivera-Trigueros, &Irene. «Machine translation systems and quality assessment: a systematic review». *Language Resources and Evaluation* 56.2 (2022): 593–619.
4. Papineni, Kishore, et al. «Bleu: a method for automatic evaluation of machine translation». *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 2002.
5. Popović, Maja. «chrF: character n-gram F-score for automatic MT evaluation». *Proceedings of the tenth workshop on statistical machine translation*. 2015.
6. Popović, Maja. «chrF++: words helping character n-grams». *Proceedings of the second conference on machine translation*. 2017.
7. Lavie, Alon, &Michael J. Denkowski. «The METEOR metric for automatic evaluation of machine translation». *Machine translation* 23.2 (2009): 105–115.
8. Agarwal, Abhaya, &Alon Lavie. «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.
9. Mukherjee, Aniruddha, et al. «A Detailed Comparative Analysis of Automatic Neural Metrics for Machine Translation: BLEURT & BERTScore». *IEEE Open Journal of the Computer Society* (2025).
10. Rei, Ricardo, et al. «COMET: A neural framework for MT evaluation». *arXiv preprint arXiv:2009.09025* (2020).
11. Havrashenko, A.., & Barkovska, O.. (2023). ANALYSIS OF TEXT AUGMENTATION ALGORITHMS IN ARTIFICIAL LANGUAGE MACHINE TRANSLATION SYSTEMS. *Advanced Information Systems*, *7*(1), 47–53. https://doi.org/10.20998/2522-9052.2023.1.08
12. Barkovska, O., & Havrashenko, A. (2023). ANALYSIS OF WORD SEARCH ALGORITHMS IN THE DICTIONARIES OF MACHINE TRANSLATION SYSTEMS FOR ARTIFICIAL LANGUAGES. Computer Systems and Information Technologies, (2), 17–24. https://doi.org/10.31891/csit-2023-2-2

Дата першого надходження рукопису до видання: 30.09.2025
Дата прийнятого до друку рукопису після рецензування: 28.10.2025
Дата публікації: 28.11.2025