The OSCAR PUBLISHING Services
International Journal Of Literature And Languages

VOLUME Vol.05 Issue12 2025
PAGE NO. 315-321
DOI 10.37547/ijll/Volume05Issue12-83

Types of Word-Based Statistical Machine Translation

Utayeva Iroda Baxodirovna

Doctor of Philosophy (PhD) in Pedagogical Sciences, Acting Professor, Dean of the Faculty of Languages at Samarkand State Pedagogical Institute, Uzbekistan

Received: 02 November 2025; **Accepted:** 25 November 2025; **Published:** 31 December 2025

Abstract: This article analyses the stages of development of machine translation technologies from statistical approaches to modern neural and multilingual modelling paradigms. In particular, the theoretical and technological features of word-based statistical machine translation, phrase-based machine translation, neural machine translation, the attention mechanism, and the Transformer architecture are examined. In addition, such issues as subword segmentation, back-translation, unsupervised machine translation, pivot language-based translation, multilingual models, zero-shot translation, and the "curse of multilinguality" are considered from a scientific perspective. The study also analyses the capabilities of NLLB, Mixture-of-Experts, and Pretrain–Finetune paradigms, including the mBART model, aimed at improving translation quality for low-resource languages.

Keywords: Machine translation, statistical machine translation, phrase-based translation, neural machine translation, Transformer, attention mechanism, parallel corpus, subword segmentation, back-translation, multilingual machine translation, zero-shot translation, unsupervised machine translation, low-resource languages, generative language model.

Introduction: As in the statistical modelling of the learning process for any task, the availability of task-specific sample data is one of the main methodological requirements for effectively training a model in the development of machine translation systems. In the field of machine translation, the most important and widely used type of such resources is parallel data, that is, parallel corpora consisting of mutually translated texts in the source and target languages aligned at the sentence level. The direct modelling of the translation process at the sentence level naturally led the earliest statistical models to adopt an approach based on the principle of direct transfer.

Although the direct transfer method did not always prove to be an optimal solution for rule-based machine translation, it served as an important conceptual basis for the development of early statistical machine translation models. The main reason for this is that the approach was not dependent on predefined rigid linguistic assumptions. In this respect, it was regarded as a universal statistical mechanism that could theoretically be applied to languages with different structural and typological characteristics.

The word-based statistical machine translation approach developed by IBM was one of the methods that brought about a fundamental change in the development of machine translation, interpreting the translation process as a statistical process based on probabilistic regularities. In this approach, by using parallel text corpora in two languages, the system learns to align lexical units in the source language with corresponding units in the target language on the basis of statistical regularities and probability values.

The central mechanism of this model is the process of word alignment. In this process, translation correspondence relations are established between words in the source and target languages, and the probability that particular units are translations of one another is calculated statistically. Such correspondences are refined through an iterative training process based on the Expectation–Maximization (EM) algorithm. By re-estimating alignment probabilities at each iteration, the algorithm makes it possible to identify the most probable lexical pairs between the two languages.

LITERATURE REVIEW

The IBM statistical translation architecture comprises three interconnected main components. The first is the translation model, which determines the correspondence between words in the source and

target languages on a probabilistic basis. The second is the language model, which serves to ensure the grammatical and syntactic naturalness of the generated translation. The third is the alignment model, which models the rearrangement of units while taking into account differences in word order between different languages. Compared with modern machine translation technologies, this approach is relatively simple, mainly modelling word-to-word translation relations and not fully covering complex syntactic-semantic structures. Nevertheless, this method formed the mathematical and conceptual foundation of subsequent statistical and neural machine translation systems.

Phrase-Based Machine Translation. Phrase-Based Machine Translation (PBMT) is an improved form of word-based statistical translation models in which larger translation units are identified and modelled by using word-level correspondences. PBMT first identifies word-level alignments in source and target sentences and subsequently extracts consistent sequences of words, that is, phrases, from these alignments []. As a result, a phrase translation table is formed that reflects many-to-many translation correspondences as well as fixed and typical expressions.

This approach is supplemented by a distortion model, which performs the task of reorganizing the order of phrases in the target language. Such a model usually uses a distance-based scoring mechanism and restricts excessive displacement of phrases from their original positions during the translation process by means of a certain penalty coefficient. Thus, compared with word-based statistical translation, PBMT makes it possible to take local context into account more effectively, translate idiomatic and fixed expressions relatively accurately, and form a more natural phrase order in the target language. At the same time, the principles of statistical alignment and probability-based translation modelling are retained.

Although PBMT represented an important stage toward more linguistically consistent processing of language units, generalized grammatical conceptualization was not sufficiently developed within it. Since the model did not fully take into account information related to the broader context of the source sentence, it could not effectively model long-distance syntactic and semantic dependencies. As a result, incorrect translation of rare idiomatic expressions, inability to adequately reorganize word order in typologically distant languages, and difficulties in expressing complex syntactic and semantic nuances emerged as the main limitations of PBMT systems.

Neural Machine Translation. The emergence of sequence-to-sequence learning models fundamentally expanded the possibilities of calculating the translation probability of an entire sentence within a single model. The main principle of Neural Machine Translation (NMT) consists in learning distributed representations of words and phrases in vector space and, on the basis of these representations, calculating through neural networks the probability of generating a particular translation sequence.

In an NMT system, each word in the source sentence is initially encoded as a one-hot vector according to its unique identifier in the vocabulary. This discrete representation is then transformed into a dense vector—an embedding—in a continuous vector space. In this way, the model processes lexical units not merely as separate symbols but as multidimensional mathematical representations reflecting their semantic and contextual characteristics.

In early NMT architectures, the encoder was organized on the basis of a Recurrent Neural Network (RNN) and generated a generalized vector representation of the entire sentence by sequentially processing the embeddings in the source sentence. At the next stage, a second RNN functioning as the decoder generated the translation sequence step by step, taking into account the source-sentence representation produced by the encoder and the previously generated target-language units. As a result, the model learned complex statistical and semantic relations between the source and target languages directly from data.

During neural network training, maximizing the log-likelihood function of the parallel corpus is defined as the primary optimization objective. For this purpose, stochastic gradient descent and back-propagation through time algorithms are used.

In early NMT models, all information from the source sentence was compressed into a single fixed context vector. This vector, generated by the encoder RNN, was used as the basis for initializing the decoder and generating the entire translation sequence. However, this mechanism created a serious information bottleneck, particularly when translating long sentences, because the model was required to incorporate all semantic, syntactic, and contextual information from the source sentence into a single vector of fixed dimensionality.

The introduction of the attention mechanism to overcome this problem became an important methodological turning point in the development of NMT. The attention mechanism makes it possible to dynamically calculate a separate context vector at each decoding step. As a result, when generating the next

unit in the target language, the decoder selectively focuses on the parts of the source sentence most relevant to that particular unit. This significantly increased the contextual sensitivity, adaptability, and semantic accuracy of the translation process. Although attention-based NMT demonstrated high efficiency, training recurrent models required substantial computational costs and created a serious obstacle to scaling model size. This problem entered a new stage with the development of the Transformer architecture. The Transformer replaced the sequential computation mechanism characteristic of recurrent neural networks with parallel computation implemented through self-attention. Unlike RNNs, the Transformer does not merely process words in a sentence sequentially but is capable of modelling relations among all units simultaneously. The positions of words in the sequence are encoded through positional embeddings. Such an architecture not only enabled a high degree of parallelization of computational processes but also made it possible to identify long-distance syntactic and semantic dependencies more effectively.

The IBM statistical translation architecture comprises three interconnected main components. The first is the translation model, which determines the correspondence between words in the source and target languages on a probabilistic basis. The second is the language model, which serves to ensure the grammatical and syntactic naturalness of the generated translation. The third is the alignment model, which models the rearrangement of units while taking into account differences in word order between different languages. Compared with modern machine translation technologies, this approach is relatively simple, mainly modelling word-to-word translation relations and not fully covering complex syntactic-semantic structures. Nevertheless, this method formed the mathematical and conceptual foundation of subsequent statistical and neural machine translation systems.

Phrase-Based Machine Translation. Phrase-Based Machine Translation (PBMT) is an improved form of word-based statistical translation models, in which larger translation units are identified and modelled by using correspondences between words. PBMT first identifies word-level alignments in source and target sentences and subsequently extracts consistent sequences of words, that is, phrases, from these alignments. As a result, a phrase translation table is formed that reflects many-to-many translation relations as well as fixed and typical expressions.

This approach is supplemented by a distortion model, which performs the task of reorganizing the order of phrases in the target language. Such a model usually uses a distance-based scoring mechanism and restricts excessive displacement of phrases from their original positions during the translation process by means of a certain penalty coefficient. Thus, compared with word-based statistical translation, PBMT makes it possible to take local context into account more effectively, translate idiomatic and fixed expressions relatively accurately, and form a more natural phrase order in the target language. At the same time, the principles of statistical alignment and probability-based translation modelling are retained.

Although PBMT represented an important stage toward more linguistically consistent processing of language units, generalized grammatical conceptualization was not sufficiently developed within it. Since the model did not fully take into account information related to the broader context of the source sentence, it could not effectively model long-distance syntactic and semantic dependencies. As a result, incorrect translation of rare idiomatic expressions, inability to adequately reorganize word order in typologically distant languages, and difficulties in expressing complex syntactic and semantic nuances emerged as the main limitations of PBMT systems.

Neural Machine Translation. The emergence of sequence-to-sequence learning models fundamentally expanded the possibilities of calculating the translation probability of an entire sentence within a single model. The main principle of Neural Machine Translation (NMT) consists in learning distributed representations of words and phrases in vector space and, on the basis of these representations, calculating through neural networks the probability of generating a particular translation sequence.

In an NMT system, each word in the source sentence is initially encoded as a one-hot vector according to its unique identifier in the vocabulary. This discrete representation is then transformed into a dense vector—an embedding—in a continuous vector space. In this way, the model processes lexical units not merely as separate symbols but as multidimensional mathematical representations reflecting their semantic and contextual characteristics.

In early NMT architectures, the encoder was organized on the basis of a Recurrent Neural Network (RNN) and generated a generalized vector representation of the entire sentence by sequentially processing the embeddings in the source sentence. At the next stage, a second RNN functioning as the decoder generated the translation sequence step by step, taking into account the source-sentence representation produced by the encoder and the previously generated target-language units. As a result, the model learned complex

statistical and semantic relations between the source and target languages directly from data.

During neural network training, maximizing the log-likelihood function of the parallel corpus is defined as the main optimization objective. For this purpose, stochastic gradient descent and back-propagation through time algorithms are used.

In early NMT models, all information from the source sentence was compressed into a single fixed context vector. This vector, generated by the encoder RNN, was used as the basis for initializing the decoder and generating the entire translation sequence. However, this mechanism created a serious information bottleneck, particularly when translating long sentences, because the model was required to incorporate all semantic, syntactic, and contextual information from the source sentence into a single vector of fixed dimensionality.

The introduction of the attention mechanism to overcome this problem became an important methodological turning point in the development of NMT. The attention mechanism makes it possible to dynamically calculate a separate context vector at each decoding step. As a result, when generating the next unit in the target language, the decoder selectively focuses on the parts of the source sentence most relevant to that particular unit. This significantly increased the contextual sensitivity, adaptability, and semantic accuracy of the translation process. Although attention-based NMT demonstrated high efficiency, training recurrent models required substantial computational costs and created a serious obstacle to scaling model size. This problem entered a new stage with the development of the Transformer architecture. The Transformer replaced the sequential computation mechanism characteristic of recurrent neural networks with parallel computation implemented through self-attention. Unlike RNNs, the Transformer does not merely process words in a sentence sequentially but is capable of modelling relations among all units simultaneously. The positions of words in the sequence are encoded through positional embeddings. Such an architecture not only enabled a high degree of parallelization of computational processes but also made it possible to identify long-distance syntactic and semantic dependencies more effectively.

In early statistical machine translation research, the issue of modelling translation at the word or phrase level was a central problem, whereas the development of NMT created the need to model translation units at the level of even smaller segments—subword units. The main reason for this is that the neural model architecture is based on the requirement to learn embeddings for a vocabulary of fixed size.

Limiting the vocabulary size helps control the computational complexity of the model, but learning high-quality vector representations of words requires observing them sufficiently often in different contexts. Under conditions of high data sparsity, rare words or words not encountered during training may remain outside the model vocabulary. As a result, adequately translating such units becomes an important methodological problem for NMT.

To overcome this limitation, optimization methods such as segmenting the model memory allocated for computing embeddings were proposed. In particular, subword segmentation demonstrated effective results in translating rare words, borrowed units, and numerical expressions. However, because purely statistical segmentation lacks explicit knowledge of linguistic structure, in some cases it may cause morphological errors and lead to partial loss of semantic information.

In order to develop more universal and practical solutions, hierarchical character/word architectures were proposed that make it possible to acquire morphological structure through self-supervised learning. Large-scale studies showed that neural models operating entirely at the character level are also capable, to a certain extent, of independently learning morphological regularities. At the same time, achieving high translation quality demonstrated the need to ensure an optimal balance among model architecture, representation level, computational power, and network capacity.

Multilingual Machine Translation and Generative Language Modelling. The rapid development of deep learning methods sharply accelerated the pace of research in machine learning. This paradigm demonstrated that model performance can be systematically improved by expanding model size, training data, and computational resources, that is, through scaling. However, increasing model scale requires increasingly large volumes of training data.

Since creating parallel corpora, verifying them, aligning translations, and controlling their quality are labour- and resource-intensive processes, the scarcity of parallel data became one of the main limitations in scaling NMT models. In PBMT systems, this problem could be partially alleviated by integrating monolingual target-language corpora into the language model. However, directly applying the same approach to NMT did not produce the expected level of performance.

One of the more effective approaches in this regard was the back-translation method. This method is a data augmentation technique aimed at artificially increasing

the amount of data, in which a baseline translation model previously trained on a parallel corpus is used to translate new monolingual sentences available in the target language into the source language. The resulting synthetic or pseudo-translations are subsequently used as new parallel data for additional model training. In this way, it becomes possible to increase the amount of training material for large-scale neural models and improve translation performance.

However, the quality of the result obtained through back-translation directly depends on the accuracy of the synthetic translations generated by the baseline model. The quality of the baseline model, in turn, depends on the quantity and quality of the initial parallel resources. Therefore, although this method provides significant benefits for high-resource language pairs, its capabilities may be relatively limited for languages with extremely scarce parallel data.

Based on the distributional semantics hypothesis and theoretical views concerning certain universal properties of languages, studies aimed at creating statistical language models based on several languages showed that distributed representations of different languages can naturally align within a single semantic space. These results provided a basis for reconsidering the previously established concept of interlingua in computational linguistics within a new neural-statistical paradigm.

DISCUSSION

In subsequent studies, this idea was applied to the development of unsupervised machine translation models. Such models are aimed at learning translation relations between two languages without parallel corpora, using only monolingual resources for each language. They employ reconstruction objectives such as the denoising autoencoder and adversarial refinement mechanisms inspired by the principles of online back-translation to align cross-lingual representations in a shared semantic space. As a result, the vector representations of different languages are gradually brought closer together in a common space during training.

Although unsupervised machine translation offers promising opportunities for reducing the shortage of parallel data in many language pairs, ensuring high-quality translation remains a complex task. In particular, the effectiveness of this approach may decrease for languages that differ sharply in typological, morphological, and syntactic terms or have limited digital resources.

In such cases, the use of a pivot language to align languages within a shared semantic space has emerged as a traditional and practically stable alternative approach [10]. In pivot-based machine translation systems, the appropriate selection of the pivot language and the organization of the system architecture are important factors determining the final translation quality. In this case, the maximum translation quality depends on the quality of the best intermediate translation performed through the selected pivot language. Errors arising at the intermediate stage may be transferred to subsequent translation directions and lead to a decline in overall quality.

Studies demonstrating that representations of different languages can be learned in a single distributed space also stimulated new research into how neural network parameters can be distributed and shared across multiple languages. As a result, multilingual machine translation approaches emerged. These models proposed carrying out multiple translation directions within a single system by using one shared encoder and decoder architecture.

Multilingual MT models are trained on sets of parallel corpora belonging to several languages [4]; [6]. In this case, the training data do not necessarily have to fully cover all possible many-to-many translation combinations among all languages. This feature created the possibility of zero-shot translation, that is, translating between language pairs that the model has not directly observed during training.

At the same time, a problem known as the multilinguality curse, or the “curse of multilinguality,” was identified in multilingual language models. According to this phenomenon, when the parametric capacity of the model remains unchanged, increasing the number of languages in the training data may, after a certain threshold, lead to a decline in model performance for each individual language. Thus, adding more languages to the system does not always improve translation quality for low-resource languages. One of the reasons for this is related to incompatibilities among the morphological, syntactic, and writing systems of different languages.

Some studies have found that separate bilingual systems developed for high- and low-resource languages may provide higher translation quality than shared multilingual models. Therefore, special optimization strategies have been developed for creating effective multilingual MT systems, including how vocabulary units should be distributed across languages, in what proportion training data from each language should be selected, and to what extent network parameters should be shared.

One of the important research directions aimed at increasing the efficiency of multilingual MT models

focused on determining which components of the model should be shared across all languages and which should be retained as language-specific parameters. Research results showed that separating language-specific vocabulary units and neural network parameters for languages whose writing systems, morphosyntactic structures, or typological characteristics differ significantly may improve performance in low-resource translation directions.

In particular, to address scaling problems for low-resource languages, large-scale initiatives such as No Language Left Behind (NLLB) expanded the capabilities of neural machine translation to more than 200 languages. These systems jointly applied strategies such as automatic data mining and extraction, back-translation, and sparsely gated Mixture-of-Experts (MoE). Although Transformer-based NMT systems significantly improved translation quality through the effective use of parallel corpora, their dependence on task-specific supervised data and sensitivity to data scarcity remained important methodological limitations. To overcome these problems, the Pretrain–Finetune (PF) paradigm emerged [5]; [9]. One important example of this paradigm is the mBART model. Within the Transformer encoder–decoder architecture, mBART uses a multilingual denoising pretraining strategy. In this process, by masking word spans, permuting sentence order, and applying language identifier tokens, the model learns to reconstruct the original text structure from corrupted input data.

Pretraining the model on data from 25 languages enabled it to develop the ability to generalize across different translation directions. At the subsequent fine-tuning stage, the model is specialized for specific language pairs through the use of supervised parallel data. During large-scale pretraining, 256 V100 graphics processing units, the Adam optimization algorithm, and scheduled learning-rate decay mechanisms were used [11]; [12]; [13]. Although the PF paradigm significantly improved machine translation quality, the need for supervised parallel data remained at the final adaptation stage. This created the need to develop more universal models capable of adapting to different languages, domains, and tasks with minimal task-specific supervision.

CONCLUSION

In general, although the PF paradigm continued to rely on supervised adaptation mechanisms, it significantly expanded multilingual generalization capabilities and established new methodological foundations for taking linguistic diversity into account as an important factor in neural modelling. These approaches laid the foundation not only for improving translation quality across a larger number of languages but also for creating cross-lingual and cross-domain language models capable of transferring knowledge across different languages and subject domains. As a result, the development vector of machine translation research shifted from specialized systems relying on large-scale parallel corpora for specific language pairs toward broadly pretrained generative language models with strong generalization and knowledge-transfer capabilities. This process has made minimizing the need for task-specific supervised data, supporting low-resource languages, and forming multilingual semantic representations within a single model one of the priority directions of modern machine translation research.

REFERENCES

1. Ha T.-L., Niehues J., Waibel A. Ko‘p tilli neyron mashina tarjimasini universal enkoder va dekoder asosida yaratish yo‘lida // Og‘zaki nutq tarjimasi bo‘yicha 13-xalqaro konferensiya materiallari. – Sietl, AQSh, 2016-yil 8–9-dekabr.
2. Tu Z., Liu Y., Shang L., Liu X., Li H. Rekonstruksiya asosidagi neyron mashina tarjimasi // Sun’iy intellekt bo‘yicha AAAI konferensiyasi materiallari. – San-Fransisko, AQSh, 2017-yil 4–9-fevral. – 31-jild, № 1.
3. Gehring J., Auli M., Grangier D., Dauphin Y. Neyron mashina tarjimasi uchun konvolyutsion enkoder modeli // Hisoblash lingvistikasi assotsiatsiyasining 55-yillik konferensiyasi materiallari. 1-jild: Kengaytirilgan maqolalar. – Vankuver, Kanada, 2017-yil 30-iyul – 4-avgust. – B. 123–135.
4. Gu J., Bradbury J., Xiong C., Li V.O.K., Socher R. Avtoregressiv bo‘lmagan neyron mashina tarjimasi // Reprezentatsiyalarni o‘rganish bo‘yicha xalqaro konferensiya (ICLR) materiallari. – Vankuver, Kanada, 2018-yil 30-aprel – 3-may.
5. Ott M., Edunov S., Grangier D., Auli M. Neyron mashina tarjimasini masshtablash // Mashina tarjimasi bo‘yicha uchinchi konferensiya: ilmiy maqolalar to‘plami. – Bryussel, Belgiya, 2018-yil 31-oktabr – 1-noyabr. – B. 1–9.
6. Irie K., Zeyer A., Schlüter R., Ney H. Chuqur transformerlar asosida tilni modellashtirish // Interspeech 2019 konferensiyasi materiallari. – Grats, Avstriya, 2019-yil 15–19-sentabr. – B. 3905–3909.
7. Han J.M., Babuschkin I., Edwards H., Neelakantan A., Xu T., Polu S., Ray A., Shyam P., Ramesh A., Radford A. va boshq. Faqat generativ til modellariga asoslangan nazoratsiz neyron mashina

tarjimasi // arXiv. – 2021. – arXiv:2110.05448.

8. Zhang B., Ghorbani B., Bapna A., Cheng Y., Garcia X., Shen J., Firat O. Mashina tarjimasi uchun til modeli arxitekturalarining masshtablanishi va transferini tadqiq etish // Mashinaviy o‘rganish bo‘yicha xalqaro konferensiya materiallari. – Baltimor, AQSh, 2022-yil 17–23-iyul. – B. 26176–26192.
9. Guo S., Zhang S., Feng Y. Sinxron tarjima uchun faqat dekoderga asoslangan oqimli transformer // Hisoblash lingvistikasi assotsiatsiyasining 62-yillik konferensiyasi materiallari. 1-jild: Kengaytirilgan maqolalar. – Bangkok, Tailand, 2024-yil 11–16-avgust. – B. 8851–8864.
10. Sun Y., Dong L., Zhu Y., Huang S., Wang W., Ma S., Zhang Q., Wang J., Wei F. Faqat bir marta keshlash: til modellari uchun dekoder–dekoder arxitekturalari // Neyron axborotni qayta ishlash tizimlari sohasidagi yutuqlar. – 2024. – 37-jild. – B. 7339–7361.
11. Shiwen Y., Xiaojing B. Qoidalarga asoslangan mashina tarjimasi // Tarjima texnologiyalari bo‘yicha Routledge ensiklopediyasi. – Abingdon, Buyuk Britaniya: Routledge, 2014. – B. 186–200.
12. ALPAC. Til va mashinalar: tarjima va tilshunoslikda kompyuterlar: hisobot. – № 1416. – Vashington, AQSh: Milliy fanlar akademiyasi, Milliy tadqiqot kengashi, 1966.
13. Chomsky N. Sintaktik strukturalar. – Berlin, Germaniya: Mouton de Gruyter, 2002.
14. Tanaka H. Kelajakda ko‘p tilli mashina tarjimasi tizimlari // Mashina tarjimasi taraqqiyoti. – Vashington, AQSh: IOS Press, 1993.
15. Dong D., Wu H., He W., Yu D., Wang H. Ko‘p tilli tarjima uchun ko‘p vazifali o‘rganish // Hisoblash lingvistikasi assotsiatsiyasining 53-yillik konferensiyasi va tabiiy tilni qayta ishlash bo‘yicha 7-xalqaro qo‘shma konferensiya materiallari. 1-jild: Kengaytirilgan maqolalar. – Pekin, Xitoy, 2015-yil 26–31-iyul. – B. 1723–1732.
16. 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.
17. Сарсенбаева З. Модернизм в узбекской литературе и интерпретация образов //Зарубежная лингвистика и лингводидактика. – 2024. – Т. 2. – №. 1. – С. 193-199.
18. 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.
19. 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.
20. Sarsenbaeva Z. MODERN APPROACHES TO TEACHING LITERARY TERMS IN COMPARATIVE LITERATURE //Mental Enlightenment Scientific-Methodological Journal. – 2026. – Т. 7. – №. 02. – С. 198-209.
21. 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.
22. 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.
23. 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.