

Classical Approach-Based 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: 30 October 2025; **Accepted:** 25 November 2025; **Published:** 31 December 2025

Abstract: This article analyses the theoretical and methodological foundations of interlingua-based machine translation and its functional capabilities in the automatic translation process. Within this approach, the mechanisms of transforming the lexical, syntactic, and semantic features of the source text into a language-independent abstract representation and generating the target-language translation text on the basis of this conceptual model are examined. The significance of lexical-conceptual structure, knowledge-based translation, semantic representation, and automatic text generation in improving translation quality is also substantiated. The capabilities of the interlingua approach to neutralize structural and grammatical differences between different languages and to produce semantically adequate translations are revealed.

Keywords: Interlingua, machine translation, intermediate representation, semantic analysis, conceptual model, lexical-conceptual structure, knowledge-based translation, linguistic modelling, source text, target language, automatic generation, translation adequacy.

Introduction: The first generations of machine translation systems were developed within the framework of the classical approach based on the manual formulation of linguistic grammar rules. In this approach, the linguistic basis of the translation process is formed on the basis of bilingual dictionaries of the source and target languages and grammatical rules developed by experts. These rules cover the syntactic, semantic, morphological, and orthographic regularities of languages and constitute the linguistic knowledge base necessary for carrying out automatic translation. Since this methodology is highly dependent on linguistic theory, it requires a considerable amount of time, expert knowledge, and computational resources in terms of creating, systematizing, and continuously improving linguistic resources. Within rule-based machine translation, rule-based translation, transfer-based translation, and interlingua-based translation models are of particular scientific and methodological significance.

Rule-Based Machine Translation (RBMT) is an automatic translation technology that operates on the basis of a knowledge-driven approach relying on linguistic knowledge. The linguistic resources of this system consist of lexical databases, grammatical

models, and formal linguistic rules developed and refined by specialists over a certain period of time. These components perform important functional tasks at different stages of the translation process in analysing source-language units, identifying their equivalents in the target language, and generating a grammatically normative translation text.

LITERATURE REVIEW

In an RBMT system, the translation process is carried out through three closely interconnected main stages: analysis, transfer, and generation. For the effective functioning of these stages, three functional modules are used, namely the source-language parser, the source-to-target transfer module, and the target-language generator. Such a modular architecture makes it possible to process the linguistic structure of the source text step by step and to form it in accordance with the lexical and grammatical norms of the target language.

At the first stage, analysis, the source-language sentence undergoes an in-depth linguistic processing procedure. In this process, important linguistic parameters such as identifying the lexical-semantic meaning of text units and resolving ambiguity (word

sense disambiguation), assigning parts of speech (part-of-speech tagging), identifying morphological features, and recognizing named entities (named entity recognition) are extracted. As a result, a structural and semantic representation of the source sentence necessary for subsequent translation operations is formed.

At the second stage, lexical transfer, bilingual electronic dictionaries and linguistic databases are used to identify a semantically appropriate target-language equivalent for the root or main lexical unit in the source language. After that, corresponding forms in the target language are selected for the affixal components of words, in particular suffixes and other morphological markers. This process is aimed at ensuring lexical and morphological correspondence between the source

and target languages.

At the third stage, generation, the conformity of the translated units with the grammatical system of the target language is ensured. First, in languages that have the grammatical category of gender, the corresponding gender markers are coordinated, after which local and long-distance agreement between the units within the sentence is carried out. These operations are performed through a special word form generation module (word generator), and ultimately a translated text corresponding to the morphological, syntactic, and grammatical norms of the target language is produced [1]; [2]. Thus, the functional mechanism of the RBMT system includes a sequential algorithmic process extending from linguistic analysis to lexical-grammatical transfer and final text generation.

Figure 1. Functional Process of a Rule-Based Machine Translation (RBMT)

System			
No	English Term	Uzbek Translation	Description of the Process
1	Morphological Analyzer	Morfologik analizator	Identifies the morphological structure, root, affixes, and grammatical forms of words.
2	Parts of Speech Tagger & Chunker	Soʻz turkumlarini belgilash va sintaktik guruhlash moduli	Identifies the parts of speech of words and groups them into syntactically related units.
3	Named Entity Recognition (NER)	Nomlangan obyektlarni aniqlash (NER)	Automatically identifies persons, places, organizations, dates, and other named entities in the text.
4	Word Sense Disambiguation (WSD)	Soʻz maʼnosidagi noaniqlikni bartaraf etish (WSD)	Determines which meaning of a polysemous word is used in a particular context.
5	Lexical Transfer	Leksik transfer	Aligns and transfers source-language lexical units with their corresponding equivalents in the target language.
6	Word Generator	Soʻz shakllarini generatsiya qilish	Generates the final word forms in accordance with the morphological and grammatical rules of the target language.

moduli

Although SMT systems provide relatively fluent translation results for low-resource languages, RBMT (Rule-Based Machine Translation) is considered a suitable approach under certain conditions because it takes the semantic and linguistic details of the original text into account more accurately [3]. The first-generation translation systems developed at the initial stage of machine translation were based on the Direct Machine Translation approach. In the scientific literature, this approach is also referred to as literal translation, word-based translation, or dictionary-based translation. In this model, the translation process is carried out mainly on the basis of linguistic rules developed for a specific source and target language pair.

In the machine translation pyramid proposed by Bernard Vauquois [8], the direct translation approach constitutes the initial level of machine translation. This method is relatively simple and effective, especially in developing translation systems between languages that are genetically and structurally close to each other. One important example of rule-based machine translation systems is SYSTRAN (Wheeler, 1984), which was initially developed as a direct translation system adapted for translation from Russian into English. Later, this model was improved, expanded with other language pairs, and had a significant impact on the development of machine translation technologies. PAHO and ANUSAARAKA [7] systems are also noted as important examples of machine translation models based on this approach.

The Direct Machine Translation approach operates mainly at the level of lexical units. In this case, source-language units are translated directly into their target-language equivalents without using any intermediate semantic or abstract representation. Such systems require a large bilingual electronic dictionary, a database of linguistic rules, and integrated software to analyse and generate the target-language translation variant. Therefore, this approach is characterized as a bilingual and unidirectional translation model.

Problems of the Direct Machine Translation approach. One of the main limitations of this approach is that it carries out the translation process mainly at the lexical level and does not sufficiently take into account the syntactic, semantic, and structural relations between units within the sentence. As a result, incorrect lexical equivalents, disrupted word order, syntactic inconsistencies, and semantic ambiguities may occur in the translated text.

Another important disadvantage of the approach is related to its simplified linguistic and computational model. This problem was identified in the earliest studies of machine translation, in which the complex nature of linguistic phenomena was not sufficiently modelled. In addition, a direct translation system is usually developed specifically for a particular language pair. Therefore, the possibility of directly applying linguistic and algorithmic mechanisms created for one language pair to other language pairs is limited.

Interlingua-Based Machine Translation is located at the upper level of the Vauquois pyramid, and its main feature is the use of an intermediate, language-independent representation that expresses the meaning of the source text. In this approach, the translation process is carried out in two main stages.

At the first stage, the source-language text undergoes linguistic analysis, and its lexical, syntactic, and semantic features are transformed into an abstract conceptual representation called an interlingua, which is independent of any particular natural language. In this process, priority is given to modelling the semantic content and conceptual structure of the source text and the relations between its units rather than its surface linguistic form.

At the second stage, the conceptual representation formed in the interlingua is re-encoded on the basis of the lexical, morphological, and syntactic rules of the target language, and the final target-language translation text is generated. Thus, unlike direct lexical substitution between the source and target languages, this approach uses a language-independent semantic-conceptual model as the central component of the translation process.

Numerous scientific research and experimental development works have been carried out at different periods to improve the quality indicators of interlingua-based machine translation. Successful applications of this approach include systems such as CETA, TRANSLATOR [3], and ANGLABHARTI [10].

In addition, the BIAS analyser was developed to provide a unified methodology for logical knowledge processing and inference through the integration of the interlingua model [11]. Subsequent studies investigated the design and development of the KANT system based on the principles of knowledge-based machine translation [12]. Furthermore, possibilities were developed for using dictionary resources formed on the basis of lexical-conceptual structure in

computer-assisted foreign language tutoring (Foreign Language Tutoring — FLT) and interlingua-based machine translation systems [13].

In general, the interlingua-based approach differs from direct machine translation by elevating the translation process from the level of simple lexical correspondence to the level of semantic-conceptual modelling. This feature expands the possibilities of neutralizing structural and grammatical differences between different languages through an intermediate representation, analysing the meaning of the source text more deeply, and generating semantically adequate translation variants in the target language.

DISCUSSION

The transfer approach is aimed at carrying out the translation process on the basis of an intermediate representation that models the semantic content of the original sentence. Due to differences between the structural and grammatical systems of the source and target languages, transfer-based machine translation systems are usually divided into three interconnected stages. At the first stage, namely analysis, the source-language parser determines the syntactic structure of the input sentence and forms its formal-syntactic representation. At the second stage, the resulting syntactic representation is transformed into an equivalent structural representation corresponding to the linguistic characteristics of the target language. At the final stage, the target-language translation text is generated on the basis of this representation.

EUROTRA, METAL [15], and the SHAKTI machine translation system [7] can be cited as examples of machine translation systems based on the transfer principle. Developing linguistic rules for rule-based machine translation systems requires a considerable amount of time, specialist labour, and linguistic resources. Therefore, studies have been conducted to automate the process of generating transfer rules. Such rules include inflectional changes of words, patterns of declension and conjugation, as well as interlingual correspondence rules at the level of lexical units [6]; [16]. In particular, since agglutinative languages are characterized by complex morphological structures, sequential combinations of affixes, and the expression of grammatical meanings through morphemes, the formation and systematization of the necessary linguistic data in a bilingual environment require substantial resources. From this perspective, approaches have been developed to generate transfer rules automatically or semi-automatically using small- and medium-sized corpora.

A number of theoretical, linguistic, and technological problems are observed in the development of transfer-

based and rule-based machine translation systems. First, the process of creating electronic and bilingual dictionaries may give rise to copyright-related restrictions. The insufficient availability of high-quality electronic dictionaries creates the need to develop new lexicographic databases. Developing, verifying, and regularly updating such dictionaries requires considerable time and human resources.

Second, rule-based systems have relatively low adaptability to new subject domains and face certain limitations in rapidly acquiring new terms, concepts, and communicative units emerging as a result of technological development. In large-scale systems, managing the interaction of linguistic rules, resolving semantic and syntactic ambiguities, and adequately interpreting idiomatic and phraseological units are complex tasks. As a result, in some cases, the original semantic content of the source sentence may be partially altered or lost during the translation process.

CONCLUSION

In general, the evolutionary development of machine translation technologies is characterized by the transition from rule-based, example-based, and statistical approaches to neural machine translation and modern models based on the Transformer architecture. The analysis of the study showed that while traditional approaches carried out the translation process on the basis of linguistic rules, parallel corpora, statistical probabilities, and separately developed translation templates, neural models made it possible to model complex semantic-syntactic relations between source- and target-language units in an integrated manner. In particular, the introduction of the encoder-decoder architecture, recurrent neural networks, the attention mechanism, and Transformer models increased the efficiency of identifying long-range contextual dependencies, differentiating context-dependent meanings of words and expressions, and automatically selecting translation equivalents. On this basis, the modern development paradigm of machine translation is characterized by the transition from algorithmic processing of linguistic knowledge to intelligent translation systems based on deep learning of contextual-semantic representations and their neural modelling.

REFERENCES

1. Nirenburg S. Knowledge-based machine translation // Machine Translation. 1989. Vol. 4. P. 5–24.
2. Cohen W.W. Fast effective rule induction // Machine Learning Proceedings 1995. Amsterdam : Elsevier, 1995. P. 115–123.

3. Abercrombie G. A rule-based shallow-transfer machine translation system for Scots and English // Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). Portoroz, Slovenia, 23–28 May 2016. P. 578–584.
4. Vauquois B. A survey of formal grammars and algorithms for recognition and transformation in mechanical translation // Proceedings of the IFIP Congress. Edinburgh, UK, 5–10 August 1968. Vol. 68. P. 1114–1122.
5. Wheeler P.J. Changes and improvements to the European Commission's SYSTRAN MT system 1976/84 // Terminologie. 1984. Vol. 1. P. 25–37.
6. Vasconcellos M., León M. SPANAM and ENGSPAN: Machine translation at the Pan American Health Organization // Computational Linguistics. 1985. Vol. 11. P. 122–136.
7. Bharati A., Chaitanya V., Kulkarni A.P., Sangal R. Anusaaraka: Machine translation in stages // Vivek-Bombay. 1997. Vol. 10. P. 22–25.
8. Vauquois B. La Traduction Automatique à Grenoble. No. 24. Dunod, 1975.
9. Gale W.A., Church K.W., Yarowsky D. Using bilingual materials to develop word sense disambiguation methods // Proceedings of the International Conference on Theoretical and Methodological Issues in Machine Translation. Montreal, QC, Canada, 25–27 June 1992. P. 101–112.
10. Sinha R., Sivaraman K., Agrawal A., Jain R., Srivastava R., Jain A. ANGLABHARTI: A multilingual machine aided translation project on translation from English to Indian languages // Proceedings of the 1995 IEEE International Conference on Systems, Man and Cybernetics: Intelligent Systems for the 21st Century. Vancouver, BC, Canada, 22–25 October 1995. Vol. 2. P. 1609–1614.
11. Yusuf H.R. An analysis of Indonesian language for interlingual machine-translation system // Proceedings of COLING 1992. Vol. 4: The 14th International Conference on Computational Linguistics. Nantes, France, 23–28 August 1992.
12. Lonsdale D., Mitamura T., Nyberg E. Acquisition of large lexicons for practical knowledge-based MT // Machine Translation. 1994. Vol. 9. P. 251–283.
13. Dorr B.J., Marti A., Castellon I. Spanish EuroWordNet and LCS-based interlingual MT // Proceedings of the MT Summit Workshop on Interlinguas in MT. San Diego, CA, USA, 28 October 1997.
14. Varile G.B., Lau P. Eurotra practical experience with a multilingual machine translation system under development // Proceedings of the Second Conference on Applied Natural Language Processing. Austin, TX, USA, 21–22 August 1988. P. 160–167.
15. Bennett W.S. The LRC machine translation system: An overview of the linguistic component of METAL // Proceedings of the 9th Conference on Computational Linguistics. Vol. 2. Prague, Czech Republic, 5–10 July 1982. P. 29–31.
16. Bharati R.M., Reddy P., Sankar B., Sharma D., Sangal R. Machine translation: The SHAKTI approach. Pre-conference tutorial // ICON. 2003. Vol. 1. P. 1–7.
17. Varga I., Yokoyama S. Transfer rule generation for a Japanese-Hungarian machine translation system // Proceedings of the Machine Translation Summit XII. Ottawa, ON, Canada, 26–30 August 2009.
18. Sumita E., Hitoshi H. Experiments and prospects of example-based machine translation // Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics. Berkeley, CA, USA, 18–21 June 1991. P. 185–192.
19. Somers H. Example-based machine translation // Machine Translation. 1999. Vol. 14. P. 113–157.
20. Mima H., Iida H., Furuse O. Simultaneous interpretation utilizing example-based incremental transfer // Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Vol. 2. Montreal, QC, Canada, 10–14 August 1998. P. 855–861.
21. 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.
22. Сарсенбаева З. Модернизм в узбекской литературе и интерпретация образов //Зарубежная лингвистика и лингводидактика. – 2024. – Т. 2. – №. 1. – С. 193-199.
23. 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.
24. 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.
25. Sarsenbaeva Z. MODERN APPROACHES TO TEACHING LITERARY TERMS IN COMPARATIVE LITERATURE //Mental Enlightenment Scientific-

Methodological Journal. – 2026. – T. 7. – №. 02. – C. 198-209.

- 26.** Belgibayeva G., Sarsenbaeva Z., Zhumagulova K. MAG ‘JAN JUMABAYEV IJODIDA OZODLIK G ‘OYASI //Ijtimoiy-gumanitar sohada ilmiy-innovatsion tadqiqotlar. – 2026. – T. 3. – №. 1. – C. 159-165.
- 27.** Sarsenbayeva Z., Zillolova G. COMPARATIVE ANALYSIS OF TWO TRANSLATED VERSIONS OF A LITERARY TEXT //Ijtimoiy-gumanitar sohada ilmiy-innovatsion tadqiqotlar. – 2025. – T. 2. – №. 4. – C. 183-188.
- 28.** Sarsenbayeva Z., Zillolova G. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДВУХ ПЕРЕВОДНЫХ ВЕРСИЙ ХУДОЖЕСТВЕННОГО ТЕКСТА //Scientific and innovative research in the social and humanitarian sphere. – 2025. – T. 2. – №. 4. – C. 183-188.