10th International Conference on Control Engineering &Information Technology (CEIT-2025)
Proceedings Book Series –PBS- Vol 23, pp.6-17

A Survey of Machine Translation Approaches

Fawzia A. E. Mansur #1, AISHA I H BAZAMA*2 , Abduelbaset Mustafa Goweder *3

#1 Computer Science Department
College of Science
Omar Al-Mukhtar University
Albayda,Libya
@ fawzia.abdalgani@omu.edu.ly

#2 Computer Science Department
College of Science
Omar Al-Mukhtar University
Albayda,Libya
@ aisha.ebrahem@omu.edu.ly

**3Computer Science Department*
School of Basic Sciences
Libyan Academy of Graduate Studies
Tripoli Libya
@agoweder@gmail.com

***Abstract*— This survey explores different machine translation methods utilized in various systems and platforms for commercial and research purposes. These methods play a vital role in enabling global communication, enhancing accessibility, supporting business and trade, fostering intercultural understanding, facilitating travel and tourism, aiding education, delivering fast and efficient translations, contributing to humanitarian aid efforts, promoting research and collaboration, and preserving language and culture. The survey aims to equip software developers and researchers interested in machine translation with valuable insights into these methods. Its objective is to help them improve translation quality with great accuracy by providing them with the necessary knowledge and understanding of these approaches. The papers utilized in this survey were obtained from Open Access Journals and online databases. All these methods are essential and can differ based on the specific context, available resources, and the quality of the translation required. To achieve optimal translation results, researchers and practitioners commonly employ a combination of various methods and techniques.**

***Keywords*— machine translation methods, rule-based, neural machine translation, Statistical machine translation, transfer-based machine translation, hybrid approaches, large language model.**

I. Introduction

Artificial intelligence (AI) is integral to machine translation (MT) systems, enabling automatic translation of text or speech. AI techniques used in MT include statistical machine translation (SMT) and neural machine translation (NMT). Deep learning enhances MT by capturing linguistic structures and semantic meaning. Natural language processing (NLP) handles linguistic phenomena for accurate translations. Attention mechanisms concentrate on related parts of the source sentence. Reinforcement learning (RL) optimizes translation quality through trial and error. Transfer learning pre-trains models on large datasets, improving translation accuracy. These AI techniques collectively advance the field of machine translation. Machine translation involves the automated conversion of text or speech from one language to another using computer algorithms and methods. The process entails developing systems or models that can comprehend the meaning and structure of the source language and produce an equivalent text in the target language. Various approaches are used for machine translation, counting rule-based machine translation (RBMT), statistical machine translation (SMT), and neural machine translation (NMT). These methods employ different methods for processing and analyzing the source text, identifying linguistic patterns, and generating the corresponding translation. Machine translation, also known as MT, has a long history dating back to the 1950s. Initial attempts at MT relied on rule-based systems, where linguistic rules were manually programmed. However, these rule-based approaches often struggled to capture the intricacies of language, resulting in inaccurate translations. Over time, statistical and neural machine translation models emerged, bringing significant advancements. These models revolutionized

ISSN: 2961-6611

machine translation by leveraging statistical patterns or neural networks to improve translation accuracy and fluency[1]. Some studies focus on normalizing machine translation for Indonesian social media content. These studies specifically examine the application of Statistical Machine Translation (SMT) to normalize the colloquial Indonesian language found in social media posts. The research entails constructing a parallel dataset for training and testing, incorporating slang words, standard words, and social media posts from the available torrent dataset. The aim is to address the challenges associated with translating informal and colloquial language used in Indonesian social media [2]. Machine Translation (MT) systems are proposed to convert the interpretation of a source language into a target language while preserving the original context. This is achieved through the application of various Natural Language Processing (NLP) techniques. The goal is to ensure that the translated text maintains the intended meaning and linguistic nuances of the original content [3]. In conventional machine translation (MT), the collaboration between humans and machines is crucial for rectifying inaccuracies in the translations produced by machines. Human translators play an essential role in post-editing and refining the outputs of MT systems. They contribute corrections and improvements to the machine-generated translations, which machines can incorporate to improve their ability to translate. This partnership between machine learning and human skill greatly enhances the dependability and precision of machine-translated text (Qian & Kong). In today's world, effective communication is vital for conducting day-to-day activities and businesses. While sign languages have been developed by individuals with hearing disabilities to overcome the barriers of spoken language, they still face numerous challenges in their daily lives. While two deaf individuals can communicate using sign language, there are significant barriers when a hearing person needs to communicate with a deaf person and vice versa. In such cases, a translation process is necessary to convert spoken language into sign language or vice versa. translators are an option for language translation, but they are expensive and not always readily available. Therefore, there is a need for a practical solution to meet the everyday communication needs of people. Machine translation offers a promising solution by automating the translation process. This advancement is particularly crucial for deaf individuals as it provides a platform for effective communication between deaf and hearing individuals, ensuring equal access to information for the deaf community. By automating translation, machine translation can break the linguistic barrier and provide deaf individuals the same opportunities to access information as everyone else. [4]. Significant progress has been made in machine translation (MT) in recent years, the automated translation of text from one language to another. These advancements have been largely propelled by developing neural machine translation (NMT) models. These models have shown impressive abilities in capturing intricate linguistic patterns and generating high-quality translations. However, it is worth noting that most existing models operate at the sentence level, translating sentences independently without considering the context of the surrounding document. [5]. Machine translation (MT) has made significant strides in recent years, driven by the increasing demand for understanding information in different languages online and the growing global trade. This progress can be attributed to improvements in hardware components, which have led to faster computing speeds, and the easy availability of monolingual and bilingual data. These factors have significantly assisted in the accomplishment and effectiveness of machine translation.[7]. Adaptive machine translation (MT) is a form of MT that leverages user feedback to gradually enhance the quality of translations. The feedback typically consists of corrections to previous translations, terminology, and style guidelines, and ratings assessing translation quality. This approach proves particularly valuable in domain-specific contexts, where standard MT systems may lack sufficient relevant data to translate specific terms or phrases accurately. However, there are still several challenges to effectively integrating user feedback into the translation process, especially during the inference phase. [7]. The MT GenEval benchmark is designed to be comprehensive and representative, encompassing a diverse range of contexts for gender disambiguation. It ensures balance by incorporating human-generated gender counterfactuals. The initial release of MT GenEval focuses on English-to-eight-target-languages translation for two genders. The primary objective of MT GenEval is to address the task of gender in machine translation, particularly in terms of evaluating and assessing gender-related aspects.[8]. Rule-based machine translation (RBMT) is an approach where linguistic knowledge is formalized by linguists into lexicons and sets of grammar rules. This knowledge is then utilized by the system to analyze sentences in the source language and generate translations. One advantage of RBMT is that it does not rely on training corpora and offers control over the output translations. However, a

ISSN: 2961-6611

drawback of this approach is that encoding linguistic knowledge into the system requires a significant investment of expert time and effort.[9]. In the past two decades, Statistical Machine Translation (SMT) has emerged as the dominant paradigm in machine translation (MT). The availability of large parallel corpora, such as Europol, and the introduction of the open-source MT customization toolkit Moses have significantly contributed to the widespread adoption of SMT in both academic and industrial settings. While language pairs like French ↔ English and Spanish ↔ Portuguese have received considerable attention and demonstrated high-quality results, limited research on MT involving Turkish has been conducted. Existing studies have mainly focused on English ↔ Turkish MT or MT between Turkic languages.[10]. This section highlights several papers that have contributed to defining the concept of machine translation. The objective of this study is to provide a valuable resource for researchers and software developers interested in machine translation, helping them understand the various methods employed and facilitating the development and enhancement of translation quality. Subsequent sections will delve into previous research. The methodology used analysis, a general summary of the research findings, and future directions, including open problems that warrant further exploration.

I. RELATED WORK

Machine translation (MT) has made significant progress, and three major approaches have emerged as dominant players: Rule-Based Machine Translation (RBMT), Statistical Machine Translation (SMT), and Neural Machine Translation (NMT). [11] acknowledge the limitations of rule-based methods and highlight the success of SMT in industrial applications. It also mentions the rise of NMT, which has been made possible by advancements in deep learning techniques. However, it points out that the literature survey does not extensively cover other approaches utilized in machine translation. [11]

[12] offer a comprehensive examination and analysis of assessment methods and challenges related to large language models. It highlights the significance of MSC (Master of Science) assessment at mission and societal levels. It proposes a framework encompassing three aspects of evaluation: what, where, and how to evaluate. The paper includes various assessment tasks, including natural language processing, reasoning, medical applications, ethics, education, natural and social sciences, and agent applications [12].

Minaee et al. delve into the development and characteristics of macro language models in natural language processing tasks. It begins by emphasizing the significance of language models in various applications such as speech recognition, machine translation, and information retrieval. It discusses the progression of language models, starting with statistical language models (SLMs) based on n-gram models, followed by early neural language models (NLMs) utilizing embedding vectors. Pre-trained language models are introduced, referring to models trained on vast amounts of unlabelled data and fine-tuned for specific tasks while being agnostic to any particular task. Transformer-based neural models with billions of parameters, such as GT, llama, and palm, are highlighted as touchscreen models offering enhanced language comprehension and generation capabilities compared to traditional language models. It emphasizes the emerging capabilities of the Master's System, including context-based learning, instruction following, and multi-step thinking, enabling the models to learn new tasks and solve complex problems [13].

Furthermore, [4] presents a valuable examination of the current status of research in machine translation of sign language and sign language generation. The analysis is conducted through a systematic review of the available literature, which enhances the reliability and comprehensiveness of the findings. One notable strength of the paper is its extensive coverage of relevant studies. This comprehensive approach allows a thorough investigation of the different techniques and methodologies utilized in sign language machine translation and generation. It is structured into three main categories: conventional machine translation, modern machine translation, and sign language generation. This classification provides a clear framework for comprehending the various technologies and advancements in the field. Additionally, [4] subdivides each category based on the specific type of machine translation or generation method, providing more detailed insights into the analysis. [4]

ISSN: 2961-6611

[14] focus on the advantages of Neural Machine Translation (NMT) over Statistical Machine Translation (SMT). It emphasizes the utilization of end-to-end modelling and distributed representations in NMT. It explores different approaches to enhance translation quality by leveraging multilingual parallel corpora and discusses the strengths and weaknesses of these techniques. However, this survey does not cover the improvement of knowledge transfer from high-resource languages to low-resource languages, the development of resource-efficient models, the exploration of precise language modelling techniques, or the investigation of multi-source translation methods.[14].

[15] focus on machine translation from Arabic to English and examine the key techniques employed in this field, along with their strengths and weaknesses. It highlights popular online translation tools like Google Translator, Microsoft Translator, and Systran that offer support for Arabic translation. It addresses specific challenges related to Arabic language processing, including writing direction, punctuation, gendered nouns, and the shape-changing nature of certain letters. It notes that Arabic exhibits a relatively free word order and is a pro-drop language. Moreover, the morphological ambiguity of Arabic presents a significant hurdle in achieving accurate translation, necessitating the exploration of methods and algorithms that can decipher word meanings based on context. Contextual information, semantic analysis, and extensive linguistic resources are crucial in this process. The survey does not specify a specific methodology employed for the survey itself. It emphasizes the need for further progress in Arabic machine translation to provide more effective and reliable translation solutions for Arabic users, considering Arabic's reputation as one of the most challenging languages.[15].

Despite providing valuable insights into various aspects of machine translation, including the evolution of the techniques and the specific challenges faced by different languages, several criticisms arise regarding the comprehensiveness of coverage, clarity of methodologies, and balance of strengths and weaknesses. They may lack depth in certain areas, reducing the possibility of obtaining actionable insights. Many studies do not specify their methodologies, which may affect the reliability and reproducibility of their results. Table 1 shows the advantages and disadvantages of prior research.

TABLE I
ADVANTAGES AND DISADVANTAGES IN PRIOR RESEARCH.

Study Title	**Authors**	**Advantages**	**Disadvantages**
Neural Machine Translation Survey	[1]	Provides a comprehensive understanding of NMT techniques for researchers and practitioners.	Limited scope, potential selection bias, lack of theoretical analysis, and limited coverage of practical applications.
Knowledge Transfer in MT	[2]	Improves translation quality by leveraging knowledge transfer from multiple languages, generalizes well, facilitates transfer from resource-rich to low-resource languages, and offers compact models.	It involves handling multiple languages simultaneously, adding complexity to training and inference, variability in translation quality, and increased training time.
Arabic MT Advances	[3]	Acknowledges significant achievements in Arabic MT, offering valuable resources due to advancements in SMT and NMT.	It faces linguistic challenges of Arabic's complexity, has a limited focus on Modern Standard Arabic (MSA), restricts coverage of dialects, and acknowledges research gaps.
Enhancements in LLMs	[4]	It offers unprecedented performance and versatility, enhances human-LLM interaction, and ensures safety and reliability through thorough evaluations.	Sensitivity to adversarial prompts, system limitations, safety concerns, and ethical considerations require careful, prompt engineering and ongoing research.

ISSN: 2961-6611

Automatic Evaluation Metrics	[5]	Provides a cost-effective method for quantitative assessment of translation systems, enabling large-scale evaluation and tracking improvements.	Metrics may overlook semantic and grammatical nuances, lacking human judgment's comprehensive assessment.
Natural Language Task Performance	[6]	Demonstrates strong performance in various tasks, general-purpose language understanding, emergent abilities, and can be augmented with external knowledge.	LLMs are often seen as "black boxes," making it difficult to understand how they arrive at specific outputs, which poses challenges for trust and accountability.

II. METHODOLOGY

The collection of papers for this study involved obtaining approximately twenty-two recent papers and a selection of reviews and survey papers from Open Access Journals and online databases. These papers specifically focused on various aspects of machine translation. The survey began with a summary and introduction, which covered the concept of machine translation as discussed in multiple survey papers and previous works. The five most recent surveys were selected and analysed to summarize their findings on machine translation and assess the strengths and weaknesses of each study. The analysis section followed, providing an in-depth examination of the different methods employed in machine translation. It highlighted the most commonly used methods in recent years, contributing to enhancing and refining translation techniques to achieve effective and high-quality results. These methods have gained attention from software developers, researchers, and individuals interested in the field. Finally, the survey concluded by summarizing the key findings and insights, and a list of references used in the study was provided. *Section Headings*

IV .ANALYSIS

This survey examines machine translation methods in recent years, specifically focusing on Rule-Based, Statistical machine translation, Neural Machine Translation, Hybrid approaches, and Example-based machine translation.

A. *The Rule-Based and neural approaches*

Rule-based machine translation (RBMT) is an approach to machine translation that uses predefined linguistic rules and patterns to translate text from one language to another. It applies regulations governing the structure, grammar, and meaning of the source and target languages. RBMT systems typically have three components: a morphological analyser, a syntactic analyser, and a transfer component. The morphological analyser identifies word properties like tense and number, while the syntactic analyser sentence structure and word relationships. Based on the linguistic analysis, the transfer component then applies predefined rules to convert the source language into the target language. RBMT systems require extensive linguistic knowledge and expertise to develop the translation rules and patterns. Linguists and translators are crucial in creating and refining these rules for accurate and high-quality translations. RBMT systems are often specific to particular languages and require separate rule sets for each language pair. One advantage of RBMT is its precise control over the translation process through explicit rule definitions. However, developing RBMT systems can be complex and time-consuming due to the need for comprehensive and accurate linguistic rules. RBMT systems may struggle with ambiguous or idiomatic expressions as predefined rules may not capture all linguistic nuances. [16] introduces a novel evaluation method that incorporates linguistic phenomena and test sets for the purpose of analyzing and comparing the performance of different machine translation systems.[16]. [9] discusse different strategies for integrating the information present in a Rule-Based Machine Translation (RBMT) system, particularly LucyLT, into a corpus-based Neural Machine Translation (NMT) model. It emphasizes using

ISSN: 2961-6611

morphological information and inflection classes derived from the RBMT system and compares its efficacy against sub-word units in the source language. [9].

B. Statistical machine translation

Statistical machine translation (SMT) is an approach to machine translation that utilizes statistical models to convert text from one language into another. A data-oriented method acquires translation patterns and probabilities from extensive parallel corpora consisting of translated sentences in both the source and target languages. The translation process in SMT involves two primary stages: training and decoding. During the training phase, a statistical model is constructed by aligning source and target sentences in the parallel corpus and estimating translation probabilities. This allows the model to understand how words or phrases in the source language correspond to their counterparts in the target language. The trained model generates translations for new input sentences in the decoding phase. The input sentence is analyzed, and the model searches for the most probable translation based on the learned probabilities. This decoding process commonly uses algorithms like the Viterbi algorithm or beam search. SMT models often incorporate linguistic features such as n-gram language models, which capture the likelihood of word sequences, and translation models, which estimate the probability of translating a source phrase to a target phrase. These models are combined using log-linear models, which assign weights to different translation options based on their probabilities. SMT has gained widespread usage and significant advancements in machine translation research and practical applications. However, SMT models typically necessitate a substantial amount of parallel training data and may encounter difficulties when translating rare or unseen words. They also have limitations in handling long-range dependencies and capturing syntactic structures compared to more recent neural machine translation (NMT) models. This approach discussed is Statistical Machine Translation (SMT), which involves the development of bilingual statistical machine translation models for translating English into fifteen low-resource Indian languages and vice versa. The authors conducted experiments using the open-source Moses toolkit and evaluated the translation quality using standard metrics such as BLEU, METEOR, and TER.[3]. Statistical Machine Translation (SMT) was employed to normalize Indonesian text, particularly in the context of social media posts. Lexical normalization aims to convert informal or colloquial text, such as social media posts, into a standardized form for subsequent analysis.[2]. [6] offer a comprehensive overview and analysis of machine translation models developed through classical, statistical, and deep learning approaches. They researched the latest advancements in machine translation and provided a comparative evaluation of various model architectures. Additionally, they delve into the potential future directions of the translation task.[6].

C. Neural Machine Translation

Neural Machine Translation (NMT) is an approach to machine translation that employs artificial neural networks to convert text from one language to another. Unlike traditional methods such as rule-based or statistical machine translation, NMT models are holistic models that directly learn the relationship between source and target languages. In NMT, the translation process involves training a neural network model on a substantial parallel corpus containing pairs of source and target sentences. The network learns to encode the source sentence into a continuous representation, often called a "hidden state" or "thought vector," and then decode this representation to generate the target sentence. Typically, an NMT model consists of an encoder and a decoder. The encoder processes the source sentence and produces the hidden state, capturing the meaning and context of the sentence. The decoder takes the hidden state as input and generates the target sentence word by word, considering the previously generated words. Training an NMT model requires a large amount of parallel data to optimize the model's parameters. This optimization is typically achieved through "back propagation," where the model's predictions are compared to the target translations, and the gradients are computed to update the model's weights. This iterative process allows the model to improve its translation performance over time. This approach is Neural Machine Translation (NMT), which utilizes a novel algorithm called Increasing the Perceived Data of Interest (IADA) to enhance the performance of neural machine translation systems at the document level (DokNMT). [17] focuses on the challenge of data dispersion in DokNMT caused by lengthy inputs and limited availability of training data. To overcome this challenge, the paper proposes strategies to increase the perceived relevance and importance of the data, aiming to generate coherent and consistent translations at the document level.[17].

ISSN: 2961-6611

This approach explores the utilization of transfer learning techniques to enhance translation accuracy in the context of low-resource languages. The study concentrates explicitly on translating English into Khasi, a low-resource Austroasiatic language spoken in Meghalaya, India. The aim is to leverage transfer learning methodologies to improve the quality and performance of translations in such linguistic scenarios.[18]. [19] discussed the challenges encountered when translating user interface (UI) texts using Neural Machine Translation (NMT) techniques. They emphasize that although NMT has succeeded in various translation tasks, it faces difficulties in handling the distinctive characteristics of UI texts, including their conciseness, ambiguity, and the requirement for additional context to resolve ambiguities. [19].

D. *Hybrid approaches*

The fourth approach is a hybrid method used to tackle the task of text simplification in Arabic. The authors put forward a hybrid approach that combines simplification techniques at both the word and sentence levels. The approach incorporates three models: a neural machine translation model, an Arabic-based lexical model, and a hybrid model that integrates both methods.[20]. Hybrid approaches in machine translation involve combining various translation techniques or models to take advantage of their respective strengths and address their limitations. These approaches aim to enhance translation quality by integrating multiple methodologies. One example of a hybrid approach is the combination of rule-based machine translation (RBMT) and statistical machine translation (SMT). RBMT systems utilize explicit language rules and patterns to generate translations, while SMT systems learn translation patterns and probabilities from large parallel corpora. By merging RBMT and SMT, the hybrid approach can provide better control over the output and handle specific linguistic phenomena while offering broader coverage and the ability to hold a wider range of language pairs. Another hybrid approach involves integrating statistical machine translation (SMT) and neural machine translation (NMT) models. Statistical models can generate initial translations, refined or post-edited, using NMT models. This combination leverages the strengths of both approaches, with statistical models providing a solid starting point and NMT models capturing more nuanced language patterns and context. Hybrid approaches can also integrate different NMT models, such as ensemble models. Ensemble models combine multiple NMT models trained with different architectures or hyperparameters. The outputs of these models are combined to generate a final translation, which can improve translation quality by reducing individual model biases and errors. Additionally, hybrid approaches may incorporate techniques like rule-based post-editing. In this method, RBMT rules are applied to post-edit the output of an NMT system, aiming to enhance fluency and accuracy. Hybrid approaches aim to leverage the strengths of diverse translation methodologies while mitigating their weaknesses, ultimately resulting in improved translation quality, coverage, and fluency. By combining complementary techniques, hybrid approaches strive to achieve more precise and contextually appropriate translations, ultimately enhancing the overall performance of machine translation systems.

E. *Example-based machine translation (EBMT)*

The fifth approach is Example-Based Machine Translation (EBMT), which involves utilizing a database of previously translated sentences or phrases. In this method, when confronted with a new translation task, the system searches for similar examples in the database and leverages them as a reference for generating translations. EBMT proves to be advantageous, especially when dealing with specialized domains or domain-specific terminology. Example-based machine translation (EBMT) is a machine translation approach that relies on a database of bilingual sentence pairs as the leading resource for translation. Instead of using explicit linguistic rules or statistical models, EBMT utilizes a database of previously translated sentences to generate translations for new sentences. In EBMT, the translation process involves searching the bilingual sentence database for similar or matching source sentences to the input sentence. Once a match is found, the corresponding target sentence is retrieved and used as the translation. EBMT systems often employ techniques to handle variations and differences between the input sentence and the retrieved example, such as word reordering or substitution. The performance of EBMT heavily depends on the quality and coverage of the bilingual sentence database. The larger and more diverse the database is, the better the chances of finding relevant examples. EBMT systems may also incorporate techniques like chunking or phrase-based matching to improve the search process and match larger segments of the input sentence. One advantage of EBMT is its ability to handle specific or domain-specific language patterns and

ISSN: 2961-6611

terminology, as it relies on real translation examples. It can also be more robust to out-of-vocabulary words or rare phrases than other machine translation approaches. Additionally, EBMT systems can be easily updated or expanded by adding new sentence pairs to the database However, EBMT can face challenges when encountering sentences or language patterns that have not been previously translated and are not present in the database. It may struggle with generating fluent and natural-sounding translations for complex or long sentences. EBMT is also heavily reliant on the quality and coverage of the bilingual sentence database, meaning that the system's performance is limited by the available data. Overall, EBMT is a machine translation approach that relies on a database of previously translated sentences to generate translations. While it has advantages in handling specific language patterns and terminology, it also has limitations in handling unseen sentences and producing fluent translations for complex structures.

F. Transfer-based machine translation

The sixth approach is Transfer-Based Machine Translation, which involves translating text by transferring the meaning from the source language to an intermediate representation and then generating the target language output. These systems often rely on linguistic analysis and semantic representations to facilitate the transfer process. A notable method within this category is using large language models and advanced artificial intelligence systems trained on diverse textual data. Qian and Kong introduce a novel approach to machine translation by integrating large language models with concept-based motivation. It highlights the challenges conventional machine translation systems face in human-computer interaction contexts. Generative Artificial Intelligence (GAI) and touchscreen, specifically GT-4, are proposed as a more adaptable and user-friendly option for machine translation.[21]. This approach involved utilizing large-scale language models for real-time adaptive machine translation. The primary objective was to leverage the learning capabilities within the context of the Master of Science (MOS) to enhance translation quality and adaptability.[7]. One advantage of TBMT is its capability to handle complex linguistic phenomena and maintain the structure and meaning of the source sentence during translation. It is beneficial for language pairs with significant structural differences. TBMT systems also provide fine-grained control over the translation process since rules can be tailored or modified to handle specific linguistic features. However, developing TBMT systems can be labor-intensive, requiring manual rule creation or extensive linguistic knowledge. They may encounter challenges when dealing with ambiguous source sentences or rare and unseen linguistic patterns. The performance of TBMT systems heavily relies on the quality and coverage of transfer rules, and the availability of comprehensive linguistic resources can limit their effectiveness. In summary, TBMT is an approach to machine translation that transfers the structure and meaning of a source sentence to produce a target sentence. It excels in handling complex linguistic phenomena and allows for precise control over the translation process. Still, it may demand significant manual effort and face difficulties with ambiguity and limited linguistic resources.

All these methods are essential and can vary depending on the specific context, available resources, and the quality of the translation required. Researchers and practitioners often combine multiple methods and techniques to achieve the best translation results. Recently, there has been a growing inclination towards employing the extensive language model approach in machine translation due to its exceptional performance. Such as numerous studies have demonstrated its effectiveness with remarkably high success rates. The term "large language model" refers to a specific artificial intelligence model that undergoes extensive training on vast quantities of text data to comprehend and generate language closely resembling human expression. [22]. These models are typically built upon deep learning architectures, such as transformers, and are trained using techniques like unsupervised learning. Transfer-based machine translation (TBMT) is an approach to machine translation that involves transferring the structure or meaning of a source sentence to generate a target sentence. TBMT systems typically undergo three stages: analysis, transfer, and generation. The source sentence is analyzed during the analysis stage to extract linguistic information like part-of-speech tags, syntactic structure, and semantic representations. This analysis helps capture the structure and meaning of the source sentence. The analyzed source sentence is mapped to the target language in the transfer stage. This mapping process involves applying language-specific rules or patterns to convert the structure and meaning of the source sentence into their equivalents in the target language. Transfer rules define how different linguistic elements are transformed from the source to the target language, considering language-specific characteristics. Finally, the transformed

ISSN: 2961-6611

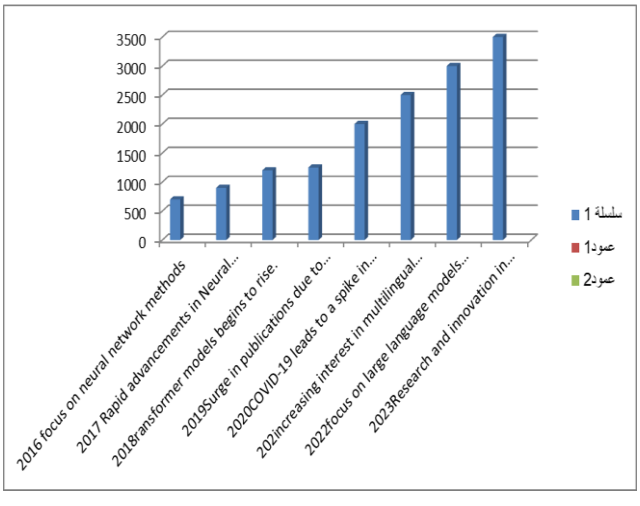
linguistic structure and meaning are used to generate the target sentence in the generation stage. This can involve applying target language generation rules or patterns to produce a fluent and natural-sounding translation. TBMT systems can be rule-based or hybrid systems, combining rule-based and statistical approaches. Rule-based TBMT relies on explicit linguistic rules and patterns to guide the transfer process, which are manually created by linguists or language experts. Hybrid TBMT systems may use statistical models or machine learning algorithms to learn transfer patterns from parallel corpora and enhance the transfer process. Based on the information provided, I have organized the different approaches to machine translation and their associated papers into a table 3, divided by the type of approach and the papers that discuss them:

TABLE 2.

TYPE OF APPROACH AND THE PAPERS.

Aproach	**Papers**
Rule-Based Machine Translation (RBMT)	[7]. [8].
Statistical Machine Translation (SMT)	[9]. [10]. [11].
Neural Machine Translation (NMT)	[12]. [13]. [14].
Hybrid Approaches	[15].
Transfer-Based Machine Translation	[16] . [17].
Large Language Model(LLM)	[13]

The chart below figure1, illustrates the annual number of publications for different types of machine translation. The vertical axis represents the publication count, while the horizontal axis indicates the publication year.



ISSN: 2961-6611

Fig. 1 . The number of publications per year.

V. CONCLUSIONS

This survey examined various machine translation methods employed in diverse systems and platforms for commercial and research purposes. These methods have played a crucial role in enabling global communication, enhancing accessibility, supporting business and trade, fostering intercultural understanding, facilitating travel and tourism, aiding in education, delivering fast and efficient translations, contributing to humanitarian efforts, promoting research and collaboration, and preserving language and culture. The objective of the survey was to equip software developers and researchers interested in machine translation with knowledge of these methods to enhance translation quality with a high degree of accuracy. The papers included in the survey were sourced from open-access journals and online databases. These significant methods could vary depending on the context, available resources, and desired translation quality. Researchers and practitioners often combine multiple methods and techniques to achieve optimal translation results. In recent years, there has been a growing trend towards using the macro-language model approach in machine translation, which has demonstrated exceptional performance with remarkably high success rates. This approach is commonly called "large language model." due to its exceptional performance. such as numerous studies have demonstrated its effectiveness with remarkably high success rates. This approach is commonly referred to as "large language model.

VI. Open problem

1- Enhancing Translation Quality: Focus on developing techniques to improve translation quality, particularly in complex and low-resource language pairs.
2- Domain-Specific Translation: Investigate methods to enhance translation quality in specific domains such as legal, medical, technical, or scientific translations.
3- Multimodal Translation: Develop models that effectively utilize multimodal data for more accurate and context-aware translations.
4- Low-Resource Languages: Develop techniques to improve translation quality for languages with limited parallel data, including leveraging transfer learning, unsupervised learning, or active learning strategies.

References

[1] M. Benmansour , & Y. Hdouch, . (2023). The Role of The Latest Technologies in the Translation Industry. *Emirati Journal of Education and Literature, 1*(2), 31-36.

[2] A. Kurnia, & E. Yulianti, (2020). *Statistical Machine Translation Approach for Lexical Normalization on Indonesian Text.* Paper presented at the 2020 International Conference on Asian Language Processing (IALP).

[3] S. B. Das, D. Panda, T. K. Mishra, & B. K. Patra, (2023). Statistical machine translation for indic languages. *arXiv preprint arXiv:2301.00539*.

[4] N. K. Kahlon, & W. Singh, (2023). Machine translation from text to sign language: a systematic review. *Universal Access in the Information Society, 22*(1), 1-35.

[5] F. Petrick, C. Herold, P. Petrushkov, S. Khadivi, & H. Ney, (2023). Document-Level Language Models for Machine Translation. *arXiv preprint arXiv:2310.12303*.

[6] S. Sharma, M. Diwakar, P. Singh, V. Singh, S. Kadry, & J. Kim,(2023). Machine Translation Systems Based on Classical-Statistical-Deep-Learning Approaches. *Electronics, 12*(7), 1716.

[7] Y. Moslem, R. Haque, J. D. Kelleher, & A. Way, (2023). Adaptive machine translation with large language models. *arXiv preprint arXiv:2301.13294*.

ISSN: 2961-6611

[8] A. Currey, M. Nădejde, R. Pappagari, M. Mayer, S. Lauly, X. Niu, G. Dinu, (2022). MT-GenEval: A counterfactual and contextual dataset for evaluating gender accuracy in machine translation. *arXiv preprint arXiv:2211.01355*.

[9] D. Torregrosa, N. Pasricha, M. Masoud, B. R. Chakravarthi, J. Alonso, N. Casas, & M. Arcan, (2019). *Leveraging rule-based machine translation knowledge for under-resourced neural machine translation models.* Paper presented at the Proceedings of Machine Translation Summit XVII: Translator, Project and User Tracks.

[10] G. DOĞRU, (2020). Statistical Machine Translation Customization between Turkish and 11 Languages. *transLogos Translation Studies Journal, 3*(1), 98-121.

[11] S. Yang, Y. Wang, & X. Chu, (2020). A survey of deep learning techniques for neural machine translation. *arXiv preprint arXiv:2002.07526*.

[12] Y. Chang, X. Wang, J. Wang, Y. Wu, L.Yang, K. Zhu, Y. Wang, (2023). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*.

[13] S. chang, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, & J. Gao, (2024). Large language models: A survey. *arXiv preprint arXiv:2402.06196*.

[14] R. Dabre, C. Chu, & A. Kunchukuttan , (2020). A survey of multilingual neural machine translation. *ACM Computing Surveys (CSUR), 53*(5), 1-38.

[15] M. S. H. Ameur, F. Meziane, & A. Guessoum, (2020). Arabic machine translation: A survey of the latest trends and challenges. *Computer Science Review, 38*, 100305.

[16] V. Macketanz, E. Avramidis, A. Burchardt , J. Helcl, & A. Srivastava, (2017). Machine translation: Phrase-based, rule-based and neural approaches with linguistic evaluation. *Cybernetics and Information Technologies, 17*(2), 28-43.

[17] M. Wu, Y. Wang, G. Foster, L. Qu ,& G. Haffari, (2024). Importance-Aware Data Augmentation for Document-Level Neural Machine Translation. *arXiv preprint arXiv:2401.15360*.

[18] A. V. Hujon, T. D. Singh, & K. Amitab, (2023). Transfer learning based neural machine translation of english-khasi on low-resource settings. *Procedia Computer Science, 218*, 1-8.

[19] S. Koneru, M. Huck, M. Exel, & J. Niehues, (2023). *Analyzing Challenges in Neural Machine Translation for Software Localization.* Paper presented at the Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics.

[20] S. S. Al-Thanyyan, & A. M. Azmi, (2023). Simplification of Arabic text: A hybrid approach integrating machine translation and transformer-based lexical model. *Journal of King Saud University-Computer and Information Sciences, 35*(8), 101662.

[21] M. Qian, & C. Kong, Enabling Human-centered Machine Translation Using Concept-based Large Language Model Prompting and Translation Memory.

[22] T. Kocmi, E. Avramidis, R. Bawden, O. Bojar, A. Dvorkovich, C. Federmann, M. Fishel, M. Freitag, T. Gowda, and R. Grundkiewicz, "Findings of the 2023 conference on machine translation (WMT23): LLMs are here but not quite there yet," in *WMT23-Eighth Conference on Machine Translation*, 2023, pp. 198--216.

ISSN: 2961-6611