A Comprehensive Study on Machine Translation: Principles, Accuracy Factors, Improvement Strategies

Huang Wei

Center for Applied Translation Studies, Guangdong Business and Technology University, Guangdong, China

barnett49@163.com

Keywords: Machine Translation; Translation Principles; Accuracy Factors; Improvement Strategies; Interdisciplinary Cooperation

Abstract: In the context of accelerating globalization, computer - aided translation technology is crucial for breaking down language barriers, and its translation accuracy is essential for the quality of information transfer. This paper conducts an in - depth study of machine translation, analyzes its principles, factors affecting accuracy, and proposes improvement strategies. Rule - based machine translation translates based on language rules, but it has difficulties in handling complex language phenomena and has high maintenance costs. Statistical machine translation leverages large - scale corpora and statistical models, yet it is significantly influenced by data quality and quantity. Neural machine translation applies deep - learning technology and has obvious advantages in semantic capture, but it also faces issues such as high data requirements and poor interpretability. The factors affecting translation accuracy include language complexity and similarity, corpus quality and scale, model architecture and training methods, and the choice of evaluation metrics. To improve translation accuracy, strategies such as optimizing the linguistic rule system, enhancing data quality and quantity, innovating model architectures and training methods, strengthening human - machine interaction and post - editing, and promoting interdisciplinary research and collaboration can be adopted. The research shows that although machine translation has made progress, continuous exploration in multiple aspects such as model architecture, data processing, human - machine collaboration, and interdisciplinary cooperation is still needed to promote the development of the technology and better serve global communication.

1. Introduction

1.1 Research Background and Significance

In the current context of accelerating globalization, international exchanges and cooperation are becoming increasingly frequent. Language, as a bridge of communication, is of utmost importance. However, due to the vast number of languages in the world and significant differences in grammar, vocabulary, and cultural backgrounds, language barriers have become a major obstacle to cross - cultural communication. Computer - aided translation technology has emerged as a solution. It can quickly convert one natural language into another, greatly improving the efficiency of information transfer and providing a convenient means of communication for people from different countries and regions. It plays a crucial role in many fields such as business negotiations, academic exchanges, and tourism. The accuracy of computer - aided translation is directly related to the quality of information transfer. Accurate translation ensures that both parties can correctly understand each other's intentions, avoiding communication breakdowns and failed collaborations caused by misunderstandings. In the business field, accurate translation of contracts, agreements, and other documents is the foundation for ensuring the smooth progress of business cooperation (Koehn, 2009) [1]. In the academic field, accurate translation of scientific research achievements facilitates the dissemination and sharing of knowledge, promoting the progress of global academia (Hutchins, 1986) [2]. Therefore, in - depth research on the factors affecting the accuracy of computer - aided translation has important practical significance. From an academic research perspective, computer - aided translation involves multiple disciplines such as computer science,

  DOI: 10.25236/iceesr.2025.070

linguistics, and mathematics. Research on the factors affecting its accuracy helps to promote the interdisciplinary integration and development of these disciplines. By analyzing the problems that occur during the translation process, it can provide a theoretical basis for optimizing computer - aided translation algorithms and improving models, and promoting the continuous improvement of computer - aided translation technology. In addition, this research can also provide references for translation teaching, helping translation learners better understand the characteristics and limitations of machine translation and improving their translation ability and post - editing skills (Bowker, 2002). [3] Research on the accuracy of computer - aided translation plays an important and indispensable role in promoting social development, facilitating cultural exchanges, and enhancing academic standards. It provides strong support for breaking down language barriers and building a closer global communication network.

1.2 Research Status

Scholars at home and abroad have conducted extensive research on the influencing factors of the accuracy of computer - aided translation and achieved remarkable results in many aspects. Abroad, early machine translation was based on rules and dictionaries for simple translation, with relatively low accuracy. With the development of technology, statistical machine translation has greatly improved translation accuracy by conducting statistical analysis of large - scale parallel corpora to establish translation models. For example, Brown et al. proposed a statistical - based machine translation method that effectively improved translation quality by using word alignment and language models (Brown et al., 1990) [4]. In recent years, the rise of deep learning technology has brought revolutionary breakthroughs to machine translation. Neural machine translation learns the mapping relationship between source and target languages by constructing deep neural networks, which can better handle long sentences and complex language structures, further improving translation accuracy and fluency. Google Translate, which adopts neural machine translation technology, has made significant progress in the translation of multiple language pairs (Wu et al., 2016) [5]. Foreign scholars have analyzed the influencing factors from multiple perspectives. In terms of the characteristics of languages themselves, the complexity of languages and differences in grammatical structures have a significant impact on translation accuracy. Complex grammatical rules, rich vocabulary variations, and special language expressions increase the difficulty of machine translation. Research shows that languages with complex grammatical structures and rich semantics, such as Arabic and Chinese, are more likely to have translation errors (Koehn, 2010) [6]. The degree of difference between languages also affects translation accuracy. Machine translation of language pairs with greater differences, such as English and Japanese, is relatively more difficult. At the data level, the quality and scale of the corpus are of great importance. High - quality corpora contain accurate, rich, and diverse text data, which can provide sufficient learning information for machine translation models. If there are problems such as noisy data, annotation errors, or data skewness in the corpus, the model may learn incorrect language patterns, thereby reducing translation accuracy. Large - scale corpora help models learn more comprehensive language knowledge and improve their adaptability to various language scenarios. However, for some low - resource languages, due to the lack of sufficient corpus data, it is difficult to effectively guarantee the accuracy of machine translation. Regarding model algorithms, different model structures and training algorithms have different impacts on translation results. Recurrent neural networks (RNNs) and their variants, such as long - short - term memory networks (LSTMs) and gated recurrent units (GRUs), have certain advantages in processing sequential data but still have limitations in handling long - distance dependencies. The introduction of the attention mechanism enables the model to dynamically focus on different parts of the source - language sentence during the translation process,improving its ability to capture key information and effectively improving translation accuracy (Bahdanau et al., 2014). [7] In China, the research on computer - aided translation technology started relatively late but has developed rapidly. With the strong support of the country for artificial intelligence and language technology, many domestic scientific research institutions and universities have invested a large amount of research efforts in the field of machine translation.

Many scholars are committed to improving the accuracy of computer - aided translation by combining the characteristics of the Chinese language, such as semantic understanding and grammatical analysis of Chinese. In the research on influencing factors, domestic scholars have some similarities with foreign research, and at the same time, they also pay attention to some factors with Chinese characteristics. In terms of cultural factors, the Chinese language contains rich cultural connotations. Expressions with unique cultural backgrounds, such as idioms, allusions, and poems, pose great challenges to machine translation. Due to the lack of in - depth understanding of cultural background knowledge, machine translation often fails to accurately convey the true meaning of these contents. For example, translating "wangmeizhike" directly as "Looking at plums to quench thirst" fails to reflect the historical and cultural allusion and profound meaning behind this idiom (Mao, 2019) [8]. In terms of terminology, with the rapid development of China in science and technology, economy, and other fields, a large number of professional terms and new words with Chinese characteristics have emerged. The accurate translation of these terms is crucial for information exchange in related fields. Due to the lack of unified terminology standards and specifications and the untimely learning of these new terms by machine translation models, errors or inaccuracies are likely to occur during the translation process (Hu, 2020) [9]. Current research still has some shortcomings. The ability to handle complex language phenomena needs to be further improved. Machine translation models often have difficulty accurately understanding and translating problems such as metaphors, puns, and semantic ambiguities in languages. In terms of multi - modal information fusion, although some research has attempted to combine text with image, voice, and other information, in practical applications, how to effectively integrate these multi - modal information to improve translation accuracy is still in the exploration stage. The professional translation requirements of different fields are diverse and specific. Existing machine translation models are difficult to achieve a perfect balance between generality and professionalism, and there is still much room for improvement in the translation accuracy of specific fields.

2. Analysis of Factors Affecting the Accuracy of machine translation

2.1 Linguistic Factors

2.1.1 Language Complexity

The complexity of a language significantly impacts the accuracy of machine translation. Different languages possess distinct grammatical structures, syntactic rules, and semantic features. For instance, some languages have highly inflected grammatical systems, where words change their forms depending on various grammatical categories such as tense, case, number, and gender. In languages like Latin, nouns, adjectives, and verbs have multiple inflections, which increases the difficulty of machine translation. For example, the Latin verb "amo" (I love) changes to "amas" (you love), "amat" (he/she/it loves) in different persons, and further inflections occur depending on tense and mood. Such complex morphological changes make it challenging for translation systems to accurately map source language forms to target language equivalents, especially when dealing with complex sentences involving multiple inflected words. Semantic ambiguity is another aspect of language complexity. Many words have multiple meanings, and their correct translation depends on the context. Take the English word "bank" as an example, which can mean a financial institution or the side of a river. Without proper context analysis, a machine translation system may wrongly translate it, leading to inaccuracies. Consider the sentence "I went to the bank to deposit money" and "I sat by the bank and watched the river flow." In the first sentence, "bank" should be translated as "yinhang", while in the second, it should be "hean". The presence of idiomatic expressions and cultural connotations also poses challenges. Idioms like "kick the bucket" in English, which means "to die," cannot be translated literally. A machine translation system might produce an incorrect translation if it does not have a comprehensive understanding of such idiomatic expressions. Similarly, cultural-specific phrases, such as "shizhidongyu, shouzhisangyu" in Chinese, carry deep cultural and historical connotations, and their accurate translation requires cultural knowledge that may not be easily captured by a machine translation system.

2.1.2 Language Similarity and Distance

The similarity between the source and target languages affects the translation accuracy. Languages that are phylogenetically related or share similar structures and vocabularies, such as Spanish and Portuguese, may have higher translation accuracy. Their similar grammatical rules and cognate words make it easier for the translation system to establish mapping relationships. For example, many words in Spanish and Portuguese have similar spellings and meanings, and the sentence structures follow similar patterns, facilitating translation. Conversely, languages that are distantly related, such as English and Japanese, present greater difficulties. They have different word orders, different grammatical systems, and distinct semantic expressions. In Japanese, the subject-object-verb (SOV) word order contrasts with the subject-verb-object (SVO) order in English. Translating from English to Japanese or vice versa requires more complex adjustments and transformations, and the translation system must handle these differences carefully to ensure accuracy. For instance, translating the English sentence "I eat an apple" into Japanese requires changing the word order to "私はりんごを食べる", which involves reordering the subject, object, and verb, and adjusting the case markers.

2.2 Data Factors

2.2.1 Corpus Quality

The quality of the corpus used for training machine translation systems is crucial. High-quality corpora should contain accurate, consistent, and representative language data. If the corpus contains errors, such as incorrect translations, misspellings, or incorrect language usage, the translation system will learn incorrect patterns. For example, if a corpus includes an incorrect translation pair like "house" being translated as "ma" instead of "fangzi", the system may produce wrong translations when encountering the word "house" in subsequent translations. The corpus should also cover a wide range of domains and text types. A corpus limited to a specific domain, say, only technical documents, will lead to poor performance when translating texts from other domains like literature or everyday conversation. A translation system trained on a technical corpus may not be able to handle idiomatic and emotional expressions in literary texts effectively. For instance, it may translate the sentence "The moonlight danced on the lake" from a poem literally rather than capturing the poetic and metaphorical meaning.

2.2.2 Corpus Size

The size of the corpus affects the generalization ability of the translation system. A larger corpus generally provides more language examples and translation pairs, enabling the system to learn more language rules and patterns. For languages with abundant resources, a large corpus allows the system to encounter various language phenomena and improve its ability to handle different sentence structures and semantic expressions. However, for low-resource languages, the limited size of the corpus restricts the system's learning. For example, when translating between a low-resource language like Welsh and English, due to the scarcity of Welsh-English parallel corpora, the translation system may have difficulties in handling less common words, phrases, and complex sentence structures, resulting in lower accuracy.

2.3 Model Factors

2.3.1 Model Architecture

The architecture of the translation model plays a vital role in translation accuracy. Different models, such as rule-based, statistical, and neural machine translation models, have their own characteristics. Rule-based models rely on manually crafted rules, which may have limitations in handling complex and flexible language phenomena. Statistical models depend on statistical patterns in the corpus, and their performance may be affected by the quality and quantity of the data. Neural machine translation models, especially those using deep neural networks like recurrent neural networks (RNNs), long-short-term memory networks (LSTMs), and the Transformer architecture, have shown better performance in recent years. The Transformer architecture, for example, uses self-attention mechanisms to handle long-range dependencies and has the advantage of parallel processing. However, its performance depends on the design of its layers, the number of attention heads, and other parameters. A poorly designed Transformer model may not effectively capture the semantic relationships in long sentences, leading to inaccurate translations. For instance, in a complex sentence with multiple clauses and long dependencies, if the attention mechanism does not work properly, the model may fail to correctly associate the elements within the sentence, resulting in translation errors.

2.3.2 Model Training

The training process of the model influences its accuracy. The choice of training algorithms, optimization methods, and the setting of hyperparameters are all important. For neural machine translation models, the choice of optimization algorithms like Stochastic Gradient Descent (SGD) or Adam can affect how quickly the model converges and the final performance. Improper hyperparameter settings, such as learning rate, batch size, and number of epochs, can lead to suboptimal training results. A too-high learning rate may cause the model to overshoot the optimal solution, while a too-low learning rate may result in slow convergence or getting stuck in a local minimum. The amount of training data used during training is also critical. Insufficient training data may lead to underfitting, where the model fails to learn the complex mapping between languages. On the other hand, overfitting may occur if the model is trained excessively on a small dataset, making it perform poorly on unseen data. For example, if a model is trained only on a limited set of business documents and then used to translate literary works, it may produce inaccurate translations due to overfitting to the business domain and lacking generalization ability.

2.4 Evaluation Metrics

The choice of evaluation metrics affects our perception of translation accuracy. Commonly used metrics like BLEU (Bilingual Evaluation Understudy), METEOR (Metric for Evaluation of Translation with Explicit Ordering), and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) have their own strengths and limitations. BLEU measures the n-gram overlap between the machine-generated translation and reference translations, but it may not fully capture semantic equivalence. A high BLEU score does not necessarily mean a good translation in terms of semantic and contextual accuracy. For example, a translation that uses different words but conveys the same meaning may receive a lower BLEU score compared to a more literal translation that overlaps more in terms of word sequences. METEOR considers synonymy and paraphrasing, attempting to capture semantic similarity better than BLEU. However, it still has difficulties in evaluating complex and creative translations, especially those involving cultural and metaphorical language. ROUGE is often used for summarization tasks but is also applied to translation evaluation. It focuses on the recall of important information, yet it may not account for the fluency and naturalness of the translation. The choice of evaluation metric thus affects how we understand and improve the accuracy of machine translation systems, and different metrics should be used in combination depending on the nature of the translation task to obtain a more comprehensive assessment.

3. Strategies for Improving the Accuracy of machine translation

3.1 Enhancing Linguistic Rule Systems

3.1.1 Refining Rule Construction

To improve the accuracy of rule-based machine translation, linguists can refine the construction of linguistic rules. Firstly, more comprehensive and detailed grammar rules should be developed. This involves not only covering common sentence structures but also paying attention to complex and rare structures. For example, in addition to handling simple declarative, interrogative, and imperative sentences, rules should be designed to deal with complex sentence patterns such as inverted sentences, elliptical sentences, and appositive clauses. In the case of an inverted sentence like "Never have I seen such a beautiful scene," the rules should accurately identify the inversion pattern and guide the translation process to ensure that the translated sentence maintains the same emphasis and meaning in the target language. Secondly, semantic rules need to be strengthened. By analyzing the semantic relationships between words, phrases, and sentences more deeply, the translation system can better handle semantic ambiguity. For polysemous words, semantic rules can be refined based on context analysis. For instance, when translating the word "light," different semantic rules can be applied depending on the context. In "The light is bright," "light" is translated as "guangxian" when referring to illumination, while in "The box is light," it should be translated as "qingde" when referring to weight. These semantic rules should be able to distinguish different semantic contexts accurately and provide appropriate translations accordingly.

3.1.2 Rule Integration with Machine Learning

Another strategy is to integrate linguistic rules with machine learning techniques. By incorporating rules into statistical or neural machine translation models, the advantages of both can be combined. For example, in a neural machine translation model, certain linguistic rules can be used as prior knowledge. When the model is generating translations, it can refer to these rules to adjust its predictions. If there is a rule that states that certain verb tenses in the source language should be translated into specific tenses in the target language, the model can take this rule into account during the translation process. This integration can help the model handle structured language information more effectively and avoid errors that may occur due to relying solely on data-driven learning.

3.2 Improving Data Quality and Quantity

3.2.1 Data Cleaning and Enrichment

Data quality is crucial for machine translation. Data cleaning is essential to ensure the accuracy of the corpus. This includes removing noise data, correcting misspellings, and eliminating incorrect translations. For example, if a corpus contains incorrect pairs like "mouses" being translated as "laoshumen" (the correct plural of "mouse" is "mice" and should be translated as "laoshu"), such errors should be corrected. Outliers and inconsistent data should also be identified and removed. Enriching the data involves expanding the diversity of the corpus. Collecting data from different domains, such as adding more literary, legal, medical, and technical texts, can improve the model's ability to handle various language scenarios. Additionally, including data from different language registers, including formal and informal language, helps the model adapt to different communication contexts. By collecting a wide range of data, the translation system can learn different language expressions and semantic understanding, thereby improving its accuracy across various fields.

3.2.2 Utilizing Multilingual Corpora

Multilingual corpora can provide more information for translation models. By utilizing multilingual corpora, the translation system can learn the relationships among multiple languages, not just between a pair of languages. For instance, when translating from English to Chinese, a multilingual corpus that includes Spanish, French, and German can help the system understand common language patterns and semantic structures shared by multiple languages, which may facilitate the translation process. The system can leverage the similarities and differences among languages to generalize translation rules and improve its translation accuracy. For example, if Spanish and Italian have similar expressions for a certain concept, and this similarity is captured through a multilingual corpus, it can be beneficial for translating the same concept from English to Chinese.

3.3 Optimizing Model Architectures and Training Methods

3.3.1 Architecture Innovation

Innovations in model architectures can lead to significant improvements in translation accuracy.

For neural machine translation models, new architectures can be explored. For example, improving the Transformer architecture by introducing more complex attention mechanisms or combining different neural network structures. A possible innovation could be to integrate convolutional neural networks (CNNs) with the Transformer to enhance the model's ability to capture local features. In a sentence, the CNN can capture the local patterns of words, while the Transformer's attention mechanism can handle long-range dependencies, and combining them can lead to more accurate translation results. Another aspect is to develop models that are more suitable for specific tasks or language pairs. If a model is designed specifically for translating legal documents, it can incorporate legal language features, such as legal terminologies and sentence structures, to improve its accuracy in that domain. This requires an in-depth understanding of the characteristics of different tasks and languages and tailoring the model architecture accordingly.

3.3.2 Advanced Training Techniques

Using advanced training techniques can enhance model performance. Adversarial training can be employed, where a generator (the translation model) and a discriminator compete with each other. The generator tries to produce more natural and accurate translations, while the discriminator aims to distinguish between machine-generated translations and human translations. Through this adversarial process, the translation model can be trained to produce translations that are closer to human quality. Fine-tuning techniques can also be used. After pre-training a model on a large general corpus, it can be fine-tuned on a specific domain corpus. For example, a pre-trained neural machine translation model can be fine-tuned on a medical corpus to improve its performance in translating medical documents. This allows the model to adapt to specific domain requirements by adjusting its parameters based on domain-specific data, thus improving accuracy.

3.4 Human-Machine Interaction and Post-Editing

3.4.1 Incorporating Human Feedback

Human feedback can play an important role in improving translation accuracy. Users can provide feedback on the translations generated by the computer, indicating where the translations are inaccurate or unnatural. This feedback can be used to adjust the model's training. For instance, if a user corrects a translation of a technical term, this information can be fed back to the system. The system can then update its knowledge base or adjust its parameters based on this feedback, gradually improving its translation of similar terms in the future.

3.4.2 Post-Editing

Post-editing by human translators is a common practice. After the computer generates a translation, human translators can review and edit it. This post-editing process can correct errors and improve the fluency and accuracy of the translation. Moreover, the post-editing data can be used to train the machine translation model. By learning from the differences between machine-generated translations and human-edited translations, the model can understand its weaknesses and improve. For example, if a machine translation of a legal document is post-edited by a legal translator, the corrected version can be used as training data to help the model better translate similar legal documents in the future.

3.5 Cross-Disciplinary Research and Collaboration

3.5.1 Linguistics and Computer Science Collaboration

Collaboration between linguistics and computer science is essential. Linguists can provide in-depth language knowledge, helping computer scientists design more linguistically informed models. For example, linguists can analyze the semantic and syntactic features of languages, which can guide the design of model architectures and the formulation of semantic rules. Computer scientists, on the other hand, can apply their computational techniques to implement and optimize these models. Their combined efforts can lead to more accurate and reliable translation systems.

3.5.2 Domain Expertise Integration

Incorporating domain expertise is another strategy. For translations in specialized fields such as medicine, law, and engineering, domain experts can work with machine translation researchers. Domain experts can provide terminologies, language expressions, and typical sentence structures specific to their fields. This information can be used to build domain-specific corpora, adjust model training, and optimize the translation process. For example, medical experts can help in annotating medical corpora and validating the accuracy of medical translations, thus improving the performance of the translation system in the medical domain.

4. Conclusion

4.1 Summary of Findings

In this study, we have explored various aspects of machine translation, from its principles to the factors influencing its accuracy and strategies for improvement. Through the analysis of different machine translation methods, including rule-based, statistical, and neural machine translation, we have gained a comprehensive understanding of the evolution and current state of machine translation technology. We have found that rule-based machine translation, although providing a structured approach based on linguistics, has limitations in handling complex and flexible language phenomena due to the difficulty of exhaustively covering all language rules. Statistical machine translation has made significant progress by leveraging large-scale corpora and statistical models, but its dependence on data quality and quantity poses challenges, especially for low-resource languages. Neural machine translation, with the application of deep learning techniques, has demonstrated remarkable advantages in capturing semantic and syntactic information, yet it still faces issues such as data requirements and model interpretability. The factors affecting the accuracy of machine translation are diverse. Linguistic factors, such as language complexity, semantic ambiguity, and language similarity, play a fundamental role. Data factors, including corpus quality and size, significantly impact the performance of translation models. Model factors, like model architecture and training methods, determine the model's ability to generalize and handle different language scenarios. Evaluation metrics, although helpful in assessing translation quality, have their own limitations and need to be used in combination to provide a more accurate evaluation.

4.2 Implications for Future Research

The research findings have several implications for future research directions. Firstly, there is still much room for improvement in model architectures. Further innovation is needed to develop more powerful and adaptable models that can handle complex language structures, semantic nuances, and diverse language pairs, especially for low-resource languages. For example, exploring new neural network architectures or hybrid models that combine the strengths of different approaches could be a promising direction. Secondly, data collection and processing remain crucial. More efforts should be devoted to collecting high-quality, diverse, and large-scale data, not only for common language pairs but also for less common languages. Techniques for data augmentation and cleaning should be further developed to ensure that the data used for training translation models is of the highest quality. Additionally, the utilization of multilingual corpora and the exploration of cross-lingual transfer learning could provide new opportunities for improving translation accuracy. Human-machine interaction and post-editing offer practical solutions for improving translation quality in the short term. Incorporating human feedback and leveraging post-editing data for model training can bridge the gap between machine and human translation, leading to more accurate and natural translations. Future research could focus on how to integrate human expertise more effectively into the machine translation process, such as developing more intelligent feedback mechanisms and user-friendly post-editing interfaces. Finally, cross-disciplinary collaboration is essential. Collaboration between linguistics, computer science, and domain experts is necessary to address the complex challenges in machine translation. By integrating domain knowledge and language expertise, more accurate and domain-specific translation systems can be developed. This interdisciplinary approach will enable us to build translation systems that are not only linguistically accurate but also suitable for different professional fields, thereby facilitating global communication, academic exchanges, and business cooperation. In conclusion, machine translation has made significant progress, but there is still a long way to go to achieve perfect accuracy. Continued research and innovation in multiple aspects, including technology, data, human interaction, and interdisciplinary collaboration, will drive the development of machine translation technology, making it a more powerful tool for breaking language barriers and promoting cross-cultural communication.

References

[1] Koehn, P. (2009). Statistical machine translation. Cambridge University Press.

[2] Hutchins, W. J. (1986). Machine translation: Past, present, future. Ellis Horwood.

[3] Bowker, L. (2002). Computer - aided translation technology: A practical introduction. Ottawa: University of Ottawa Press.

[4] Brown, P. F., Cocke, J., Della Pietra, et al.. (1990). A statistical approach to machine translation. Computational Linguistics, 16(2), 79 - 85.

[5] Wu, Y., Schuster, M., Chen, Z., et al(2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.

[6] Koehn, P. (2010). New challenges in machine translation. In Proceedings of the 23rd International Conference on Computational Linguistics (pp. 26 - 33). Association for Computational Linguistics.

[7] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[8] Mao, J. (2019). Cultural factors in machine translation of Chinese idioms. Journal of Translation Studies, 12(3), 67 - 74.

[9] Hu, X. (2020). Research on the translation of Chinese - characteristic terms in machine translation. Journal of Language and Culture Research, 8(2), 45 - 50.