

The Past, Present and Future of Machine Translation in China: A Visualization Study Based on CNKI Literature (1959–2021)

Li Fan¹ & Meiqi Yun¹

¹ Central University of Finance and Economics, Beijing, China

Correspondence: Li Fan (Ph. D/Professor), School of Foreign Studies, Central University of Finance and Economics (CUFE), Beijing, 102206, China.

Received: August 22, 2022

Accepted: October 15, 2022

Online Published: November 2, 2022

doi:10.5539/ijel.v12n6p98

URL: <https://doi.org/10.5539/ijel.v12n6p98>

Abstract

Machine translation, a product of technical growth, is advancing the field of translation. With the aid of visualization analysis techniques, this study examines the machine translation literature published in the CNKI database from 1959 to 2021 while analyzing its history, key issues, arguments, and future directions of related research. The proposal is then made that, in light of the tendency toward technological advancement, the fields of translation research and education should take on the disciplinary duties associated with the motivation for “collaborations between humans and machine and use of machines for humans.” The study also emphasizes how important it is to improve students’ technological competence, support the localization of language services, and integrate information technology into language and translation instruction.

Keywords: machine translation, human-machine cooperation, translation research and education, literature visualization research

1. Research Background

With the progress of high technology, machine translation has experienced a long history and quick growth. The following phases have been reached in its development, according to Lulu Gao and Wen Zhao (2020): (1) the period of initial development (1949–1961): In 1949, Warren Weaver proposed the use of computers for translation, arguing that in order to avoid “word-for-word” translation, it was essential to consider how context affects lexical meaning, linguistic logic and reasoning, the translation and decoding process, as well as the universal characteristics of the language (Poibeau, 2017). (2) The period of sluggish development (1960–1967) saw the establishment of a number of machine translation models based on syntactic analysis. In 1958, Bar-Hillel drew attention to the semantic barrier that automatic machine translation systems faced. He proposed that they be overcome through machine-assisted translation to give the translator the aid they need during the pre- and post-translation processes (Bar-Hillel, 1958, 1959). The Automatic Language Processing Advisory Committee (ALPAC) was founded around this time to evaluate the financing for the advancement of machine translation. (3) the period of slow development (1967–1990): Although machine translation as a whole developed slowly, there was some significant potential in some nations and regions, including Canada, France, and the European Union. For instance, the European Union collaborated with Systran, one of the first machine translation companies founded by Peter Toma in 1968, to realize automatic machine translation between EU member languages. (4) The period of new development (1990–): The International Conference on Computational Linguistics in Finland in 1990 launched the statistical machine translation era based on massively parallel corpora, which was followed by the emergence of neural machine translation based on “deep learning” (Hinton et al., 2012).

In recent years, continuous breakthroughs in machine translation technology have been making headlines. For example, on September 27, 2016, “Google Brain Group” announced the launch of a new neural machine translation system GNMT (Google Neural Machine Translation system) (Note 1), which can significantly reduce the translation error rate between key language pairs done by the more classic machine translation models significantly down to 55%–85%; on December 22, 2016, the research team rode the wave of success to relaunch the Zero-shot multilingual neural machine translation system, which enables translation of language pairs and phrases that have not been previously trained for direct translation (Note 2). As research on big data, deep learning, and neural network machine translation continues to gain momentum, machine translation is bound to

replace humans to complete a larger number of translation tasks. The gradual transformation of the translation job market will inevitably make foreign language teaching adjust accordingly. The research on machine translation will undoubtedly be the key to efficient integration and significant breakthrough. Therefore, this study combines the trend analysis of literature with the review of representative literature, aiming to present the history, focus, and development prospects of relevant studies in the CNKI database from January 1959 to June 2021.

2. Research Tools and Literature Sources

A literature review is carried out with CiteSpace, a tool for visualizing literature. It is a powerful tool for scientific literature research and is developed by Dr. Chaomei Chen of the School of Information Science and Technology at Redsell University in the United States in conjunction with the WISE Lab at Dalian University of Technology, which can provide visual analysis of literature based on data sources such as Web of Science, PubMed, arXiv, ADS, NSF, CNKI, etc. (Chen, 2004).

First, we searched the CNKI database, the largest research database in China, for journal papers containing the phrase “machine translation” in the titles. We found that the first journal paper included in the database was published in 1959. From January 1959 to June 2021, a total of 1723 papers were published in both English and Chinese journals, and these 1723 papers formed the data pool for this study. Based on a detailed review of representative studies, a total of 649 Chinese journal articles from the fields of “foreign language and literature” and “Chinese language and literature” will be discussed.

3. Overall Research Developments

Among the 1723 Chinese and English journal papers containing “machine translation” in the titles collected in the CNKI database from January 1959 to June 2021, the following research trends of machine translation are shown: In terms of the number of studies, the related studies have been developing continuously since 2001, especially in 2016.

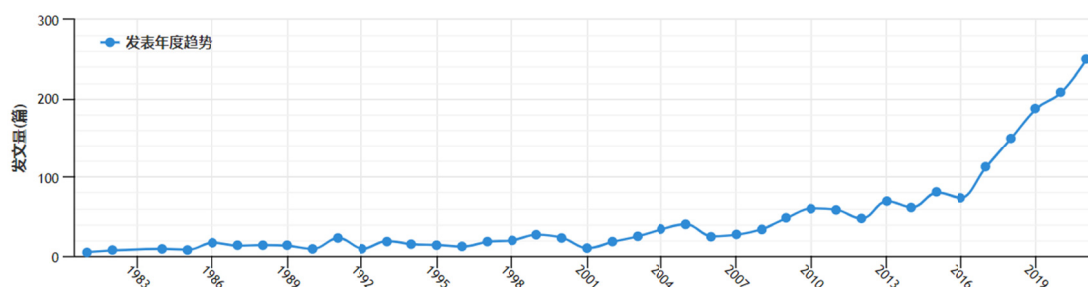


Figure 1. Trend of the number of Chinese and English journal papers with the title containing “machine translation” in CNKI

In terms of topic distribution, in addition to “machine translation,” there are “neural machine translation,” “machine translation systems,” “machine translation research,” “statistical machine translation,” and “statistical machine translation,” which reflect the achievements of technological innovation in machine translation.

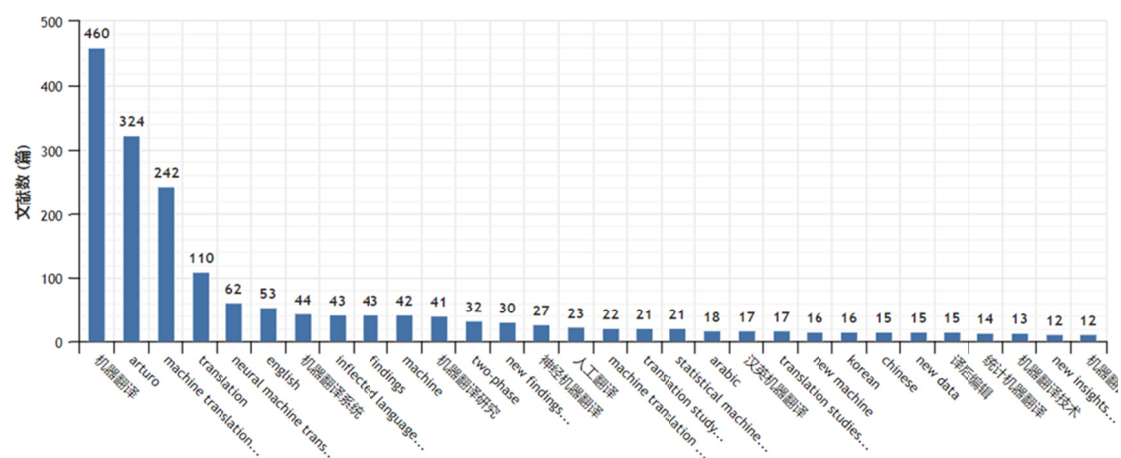


Figure 2. Topic distribution of Chinese and English journal papers with the title containing “machine translation” in CNKI

In terms of journal distribution, *Computational Linguistics*, *Chinese Translation*, and *IEEE/ACM Transactions on Computational Biology and Bioinformatics* are the three academic journals with the most “machine translation”-related papers, two of which are foreign academic journals in the field of computer expertise.

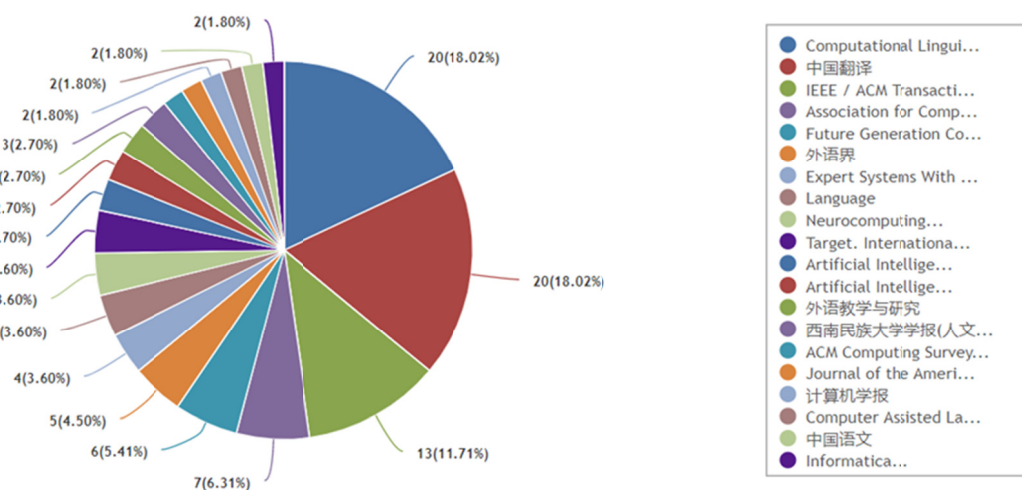


Figure 3. Journal distribution of Chinese and English journal papers with the title containing “machine translation” in CNKI

Concerning disciplinary distribution, Chinese language and literature, computer science and applications, and foreign language and literature are the three disciplines with the most published studies on “machine translation.”

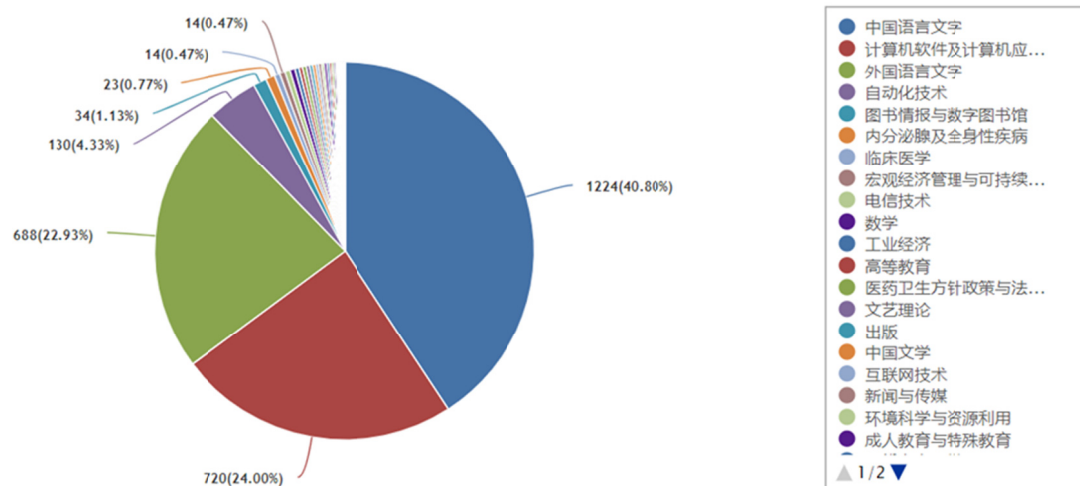


Figure 4. Disciplinary distribution of Chinese and English journal papers with the title containing “machine translation” in CNKI

Regarding authorship, papers on machine translation were mainly finished by scholars working in computer science.

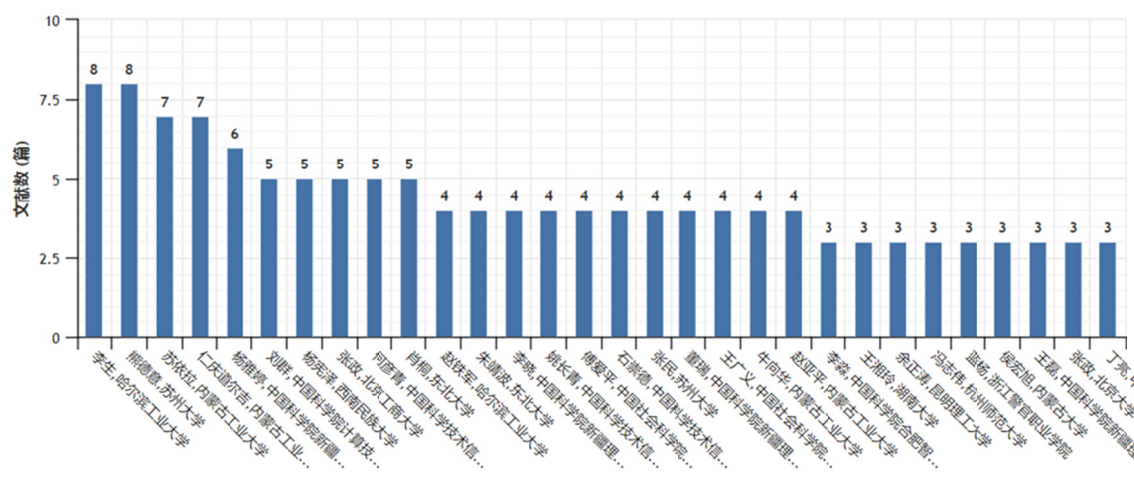


Figure 5. Authorship distribution of Chinese and English journal papers with the title containing “machine translation” in CNKI

To further understand the work in the field of linguistics, literature, and translation on the issue of machine translation, we continue to focus on the visual analysis of literature in CiteSpace for 649 documents belonging to two disciplines, Chinese language and literature, and foreign language and literature. First, the information on the frequency of occurrence, centrality, and year of the first occurrence of keywords is presented. The following table contains the top 30 keywords in the order of frequency.

Table 1. Keywords used in the journals of Chinese and foreign language and literature with the title containing “machine translation” in CNKI

No.	Keyword	Frequency	centrality	Year of the first occurrence
1	机器翻译(machine translation)	531	0.92	1959
2	低资源语言(low-resource language)	64	0.02	1959
3	中朝机器翻译(Chinese and Korean machine translation)	63	0.00	1959
4	回译(translate back)	63	0.00	1959
5	共享参数(shared parameter)	63	0.00	1959
6	多语言神经机器翻译(multilingual neural machine translation)	60	0.00	1959
7	差异化特征(differentiating features)	59	0.00	1959
8	多语言翻译(multilingual translation)	59	0.00	1959
9	强化学习(intensive learning)	59	0.00	1959
10	无监督翻译(unsupervised translation)	58	0.00	1959
11	机器翻译质量评估(quality assessment of machine translation)	53	0.00	1959
12	枢轴语言(pivot language)	50	0.00	1959
13	人工翻译(manual translation)	49	0.14	1984
14	神经网络机器翻译(neural network machine translation)	49	0.01	1959
15	汉-越神经机器翻译(Chinese-Vietnamese Neural Machine Translation)	47	0.00	1959
16	汉语-英语-缅甸语(Chinese-English-Burmese)	45	0.00	1959
17	汉越平行数据(Chinese-Vietnamese parallel corpus)	43	0.00	1959
18	稀缺资源语言(scarce resource language)	38	0.00	1959
19	神经机器翻译(neural machine translation)	37	0.11	2017
20	联合训练(joint training)	34	0.00	1959
21	迁移学习(transfer learning)	31	0.01	1959
22	语义空间映射(semantic space mapping)	29	0.00	1959
23	语言特点(language features)	29	0.00	1959
24	质量估计(quality estimation)	26	0.00	1959
25	机器翻译系统(machine translation systems)	25	0.15	1982
26	人工智能(artificial intelligence)	25	0.05	1997
27	译后编辑(post-translation editing)	20	0.05	2017
28	统计机器翻译(statistical machine translation)	15	0.07	2009
29	机器翻译研究(machine translation research)	15	0.03	1982
30	语料库(corpus)	12	0.02	2000

CiteSpace, v. 5.7.R2 (32-bit)
June 18, 2021 9:28:27 AM CST
WoS: F:\Lily\A_BJ_Research\A_机器翻译与公示语的研究\citespace analysis\data
Timespan: 1959-2021 (Slice Length=1)
Selection Criteria: g-index (k=25), LRF=3.0, LBY=5, e=1.0
Network: N=696, E=1136 (Density=0.0047)
Largest CC: 516 (74%)
Nodes Labeled: 1.0%
Pruning: None



Figure 6. Keyword clustering of the papers with the title containing “machine translation” on Chinese and foreign language and literature in CNKI

It can be seen that, since the first literature in 1959, some traditional keywords (such as “back translation,” “shared parameters,” “continuous differentiation features,” “enhanced learning of multilingual translation,” “assessment of machine translation quality”, etc.) have continued to be hot research concerns. However, discussions on the differences and cooperation between machine translation and human translation have also been persisting, as the keywords like “human translation” and “post-translation editing” arise.

More importantly, with the development of technology and the deepening of technical application, some emerging keywords representing the characteristics of the times have emerged in recent years, such as “neural machine translation,” “artificial intelligence,” “post-translation editing,” “statistical machine translation,” “corpus,” etc. The list in Figure 7 also reflects that the keywords with more prominent citation bursts in different periods are “neural machine translation,” “post-translation editing,” “statistical machine translation,” “machine translation,” “quality assessment of machine translation,” “computational linguistics,” “attention mechanisms,” “deep learning,” etc. In terms of the duration of citation bursts, in 2021, the research in this field will continue to be the main focus, with popular key terms like “neural machine translation,” “neural networks,” “artificial intelligence,” “artificial intelligence,” and “post-translation editing,” among others. The keywords “statistical machine translation,” “deep learning,” and “neural network” fully reflect the trend of machine translation development. “Statistical machine translation” refers to the corpus-based statistical machine translation that automatically constructs machine translation models by word alignment, phrase alignment, and sentence alignment in parallel texts. “Deep Learning,” proposed by Hinton et al. in 2006, is multi-level nonlinear processing. It provides more accurate results and efficient performance in processing problems than traditional linear processing. “Simulated neural network” is to enable the machine to have the ability to automatically learn abstract expressions and thus achieve the best results in machine translation. The neural machine translation based on deep learning simulates how the human brain learns and processes information, thus replacing many types of traditional machine translations.

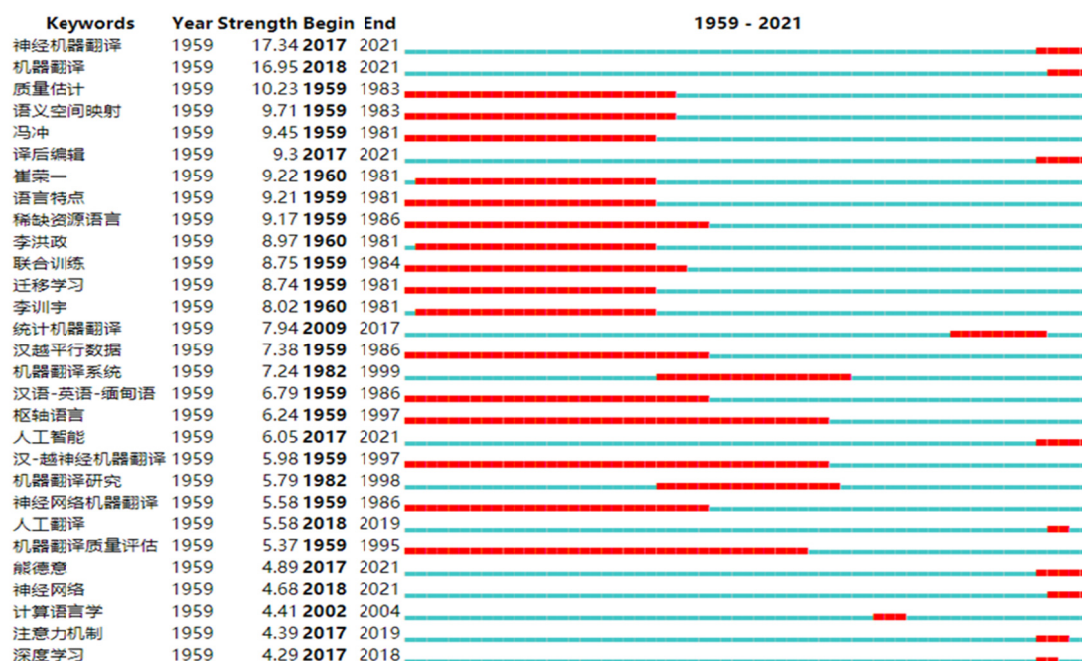


Figure 7. Keyword citation bursts of the papers with the title containing “machine translation” on Chinese and foreign language and literature in CNKI

Table 2 reflects the information of the top 20 authors with the highest frequency of citations. Among them, Xiong Deyi and Renqing Daoji, who have been cited more frequently in recent years, are active in the research on neural machine translation.

Table 2. Top authors of the papers with the title containing “machine translation” on Chinese and foreign language and literature in CNKI

Ranking	Names of the authors	Frequency	Year of the first appearance
1	余正涛(Yu Zhengtao)	128	1959
2	高盛祥(Gao Shengxiang)	127	1959
3	朱俊国(Zhu Junguo)	126	1959
4	冯冲(Feng Chong)	64	1959
5	崔荣一(Cui Rongyi)	60	1959
6	李飞雨(Li Feiyu)	57	1959
7	李洪政(Li Hongzheng)	57	1959
8	李训宇(Li Xinyu)	57	1959
9	杨福岸(Yang Fu'an)	57	1959
10	毛存礼(Mao Cunli)	56	1959
11	满志博(Man Chibo)	56	1959
12	杨飞扬(Yang Feiyang)	56	1959
13	赵亚慧(Zhao Yahui)	48	1959
14	邹翔(Zou Xiang)	46	1959
15	黄河燕(Huang He Yan)	45	1959
16	冯志伟(Feng Zhiwei)	21	1978
17	熊德意(Xiong Deyi)	10	2017
18	张政(Zhang Zheng)	9	2003
19	刘群(Liu Qun)	7	1998
20	仁庆道尔吉(Renqingdao'erji)	7	2018

4. Review of Representative Studies

After outlining the overall trend of machine translation, the following review will be conducted according to the categories based on the perspectives of literature reviews on machine translation, studies on text-based features of translation, comparative studies on human translation and machine translation, interdisciplinary studies on machine translation and linguistics, and challenges and applications of machine translation.

4.1 Literature Reviews on Machine Translation

The studies reviewing the research on machine translation include Zhiwei Feng (2018), Hanji Li (2018), Xiangling Wang (2019), Lulu Gao and Wen Zhao (2020). Zhiwei Feng (2018) demonstrated the close connection between machine translation and human translation. At the same time, analyzing the prevailing situation, the study concluded that although machine translation predates artificial intelligence, both are immature and still in the development stage. Hanji Li (2018) and Xiangling Wang (2019) provided an overview of the relevant research trends with the visualization tool of CiteSpace. Hanji Li (2018) discussed the papers related to “machine translation” published in CSSCI and Chinese core journals in CNKI during the period of 2007–2016, reported the hot areas and keywords of machine translation research in China and proposed that the linguistic community and the computer science community should cooperate to improve machine translation. Reviewing 12 core foreign language journals included in CNKI, Xiangling Wang (2019) compared the history of machine translation research explored by domestic translation scholars in the past 60 years, and agreed with the operation model of “machine translation + machine-assisted tools + post-translation editing,” pointing out that the cooperation and exchange among scholars, as well as those between scholars and industry, should be further strengthened. It also suggested that it should be crucial to integrate the training of translation technology into traditional translation curricula. Lulu Gao and Wen Zhao (2020) outlined the technical characteristics, development trends, advantages and limitations of machine translation at various stages, and the opportunities and challenges facing its development. At the same time, the study pointed out that given the statistics related to machine translation, for literary and philosophical texts with less corpus, a combination of machine translation and human translation can be chosen.

Regarding corpus construction, Zhiwei Feng (2020) proposed that parallel corpora are the basis of the methods for statistical machine translation as well as neural machine translation. He emphasized the importance of large-scale and authentic corpus and pointed out that the lack of linguistic data resources is a challenge for neural machine translation at present. For corpus-based machine translation, Jie Zhu (2019) compared the results of two approaches, rule-based and corpus-based, and concluded that a rule-based approach is a rationalist approach with rigid grammatical rules, low quality and high labor cost; a corpus-based approach is an empiricist approach with high accuracy and easy understanding.

As to the research on statistical machine translation and neural machine translation, Yunlu He et al. (2019) proposed the principles of rule-based machine translation, statistical machine translation and neural machine translation, and analyzed the advantages and disadvantages of machine translation. Jinsong Su et al. (2020) investigated machine translation from discourse and introduced the existing research on neural machine translation from three aspects: context modeling (single encoder, multiple encoder, and multiple decoders), model training, and model analysis. They finally analyzed the main difficulties currently faced by neural machine translation and claimed that future work should focus more on building massively parallel data, designing more reasonable automatic evaluation indexes for translations, and conducting research on simplifying the neural machine translation models of discourse.

Regarding research-based teaching, Ying Qin (2018) discussed the quality assessment of machine translation based on neural networks and the impact on translation teaching. She proposed that in the teaching of translation, the division of labor between human and machine should be clarified, and traditional teaching methods should be improved with the introduction of post-translation editing courses.

4.2 Studies on Machine Translation Based on Text Types

The applicability of machine translation to diverse types of texts is also the focus of related research because text types and features directly affect the quality of machine translation. Considering the machine translation of legal texts, Falian Zhang (2020) found that there are some issues with language comprehension and logic judgment in the machine translation of legal texts by comparing the translation capabilities of four machine translators, namely Google, Bing, Oulu, and Tencent Translate. He proposed that machine translation still requires human translators with professional knowledge for post-translation editing and reviewing.

On the machine translation of medical texts, Bin Tang and Shuo Chen (2020) compared the performance of medical texts in seven online machine translation software (Baidu, Sogou, Google, Bing, Jinshan, Youdao, and Tencent) through data processing and quantitative analysis. They found differences in the seven machine translations, and none of them could fully achieve translations that met medical standards. This study also suggested that combining human post-translation editing should be a critical way to ensure the quality of medical text translation at present and in the future.

On the machine translation of subtitles, using TED talks as the research materials, Weiqing Xiao and Jiahui Gao (2020) used the FAR model (functional equivalence, acceptability, reading experience) to compare the quality of English to Chinese subtitles automatically generated by “Wangyi jianwai” of NetEase (网易见外) with those of NetEase Open Class subtitles. This study reported a significant difference between the two, with machine translation subtitles having more mistakes in semantic segmentation and selection. According to it, we should concentrate on advancing translation technology and semantic segmentation in this area of translation while cultivating corresponding post-translation editing skills.

4.3 Comparative Studies on Human Translation and Machine Translation

Machine translation has unique advantages. Jun Yang (2018) proposed two crucial technologies in the revolution of machine translation: neural machine translation and speech recognition technology. Humans and machine have distinctive characteristics and uses, and they can be divided but cooperate. Machine translation has great market prospects in the face of daily communication that does not require exceptionally high quality. Wenhong Huang (2018) proposed that machine translation has three main functions nowadays: query function, speed-up function, and reference function, and the application methods are mainly direct translation (low cost and high efficiency), auxiliary translation (human translation as the main body), and modular translation (translators and translation software form good interaction).

On the other hand, machine translation has inherent defects, the most important of which is that it cannot have human intelligence to understand human language. Starting from the subjectivity of translators, Wei Chen (2020) discussed the essential difference between machine translation and human translation and emphasized that even if the technology of “deep neural network” is advanced, it still lacks the skills of a human brain — the social attributes of human beings.

Quality assessment of machine translation is the core of research. Qing Wang and Xiao Ma (2020) analyzed the linguistic problems (lexical polysemy, differences in syntactic structures, word separation, logic of meaning, context) and the quality assessment in machine translation. They tried to propose a set of quality assessment methods covering techniques such as statistics, cosine similarity, text sentiment analysis, etc., respectively, from macroscopic fidelity and accuracy (vocabulary, sentence, semantics, pragmatics) and microscopic perspectives (word frequency, word collocation, average sentence length, syntactic analysis, text similarity, subject matter,

text sentiment, and tendency) to examine the translation quality. This study discussed three aspects of assessing translation quality: linguistic knowledge as a parameter, computer technology as an aid, and translation fidelity as a criterion; various instruments are discussed in exploring assessment methods: mutual information methods to calculate relevance (with formulas and cases), word frequency statistics, Stanford Parser tool (for syntactic analysis), cosine similarity and implicit Dirichlet distribution models to analyze text similarity and keyword matching at the semantic level. Similar studies are also available, such as Han Na (2018), which selected general texts, tourism texts, science and technology texts as the test data and compared Google and Youdao translation systems with human translation to analyze the differences at the lexical, syntactic, and discourse levels. It also discussed METEOR evaluation criteria based on machine translation, stating that error correction, which is done by humans, provides a corpus for training machines, which helps to optimize the technology of machine translation.

The tug-of-war between humans and machines revolves around the two main topics of whether humans will be replaced by machines and how humans and machines can work together. The prevailing view of current research is that machine translation cannot wholly replace human translation. Research on the future relationship between machine translation and human translation can be found in Jiangbo Tian (2018), Shi-Zhen Song (2019), Ying-Yu Pang (2019), Gulipina-Jiadi (2020), Li Li (2020), Chaowei Zhu (2018), etc. Chaowei Zhu (2018) argued that machine translation is unlikely to replace humans, especially since machine translation will never be competent in areas such as literary translation. The study cited characteristics of literary works such as imagery, emotion, fiction, metaphor, and polysemy, all of which are difficulties that machine translation has to overcome. In addition, the study suggested that in the context of quickly evolving web technology, machine-assisted translation systems, such as Trados, can be used to achieve human-machine cooperation and fully develop well-rounded translators. In terms of translation education and teaching, the study advised that the “whole-person education” of translation should be strengthened, dealing with the relationship between translation ability and humanistic literacy, the relationship between a specialist and a generalist, and the relationship between national standards and talents’ qualities. Since there is no substitute, the best solution is cooperation. Shi-Zhen Song (2019), Benjin Zou (2018), and Rong Li (2019) have provided insights on solutions, such as realizing the combination of technology and humans; implementing human-machine cooperation models for translating different texts, including machine-led (e.g., processing texts such as WeChat and email), human-guided (e.g., processing literary and artistic texts), and human-machine cooperation (e.g., dealing with news and technology texts).

The study of specific models of human-machine cooperation is the key to the implementation of the vision, with discussions such as those by Xiangling Wang, Tingting Wang (2019), and Mei-Lin Yuan (2019). Xiangling Wang and Tingting Wang (2019) conducted an empirical study comparing human translation with machine translation integrated with post-translation editing (human-machine combination). 31 MTI students were recruited for the study to compare the speed, quality, and attitude of translating scientific and technical texts from English to Chinese and the MQM (Multidimensional Quality Metric) method was used to measure the effectiveness of the translation. The findings of this study demonstrate that when human post-editing is used, machine translation is superior to human translation. The ability of the translators, their attitude (positive or negative) toward post-editing, and the caliber of the machine translation all affect the quality of the translation. The best results can be achieved when a translator approaches post-editing with a positive mindset. This study recommended introducing post-editing in translation teaching and cultivating students’ post-editing ability to make them more competitive in the translation market.

The most crucial role of human translators in the human-machine cooperation model is in the post-translation editing session of machine translation. Other major studies on machine post-translation editing include Huijun Ren et al. (2020), Huifen Zhu et al. (2020), Yuwan Yang (2020), Yating Lian (2020), Xinjie Cai, and Bing Wen (2021). Yuyuan Yang (2020), Yating Lian (2020), Xinjie Cai, Bing Wen (2021) et al. summarized the types of errors in machine translation and their distribution. These studies demonstrated that fully automated, high-quality machine translation is still not achieved, and the introduction of post-translation editing after machine translation is crucial. Huijian Ren et al. (2020) explored the phenomenon of “translation with emotion, a machine without coldness.” They exemplified the importance of post-translation editing in improving the quality of translations and conveying human distinctions. Based on Newmark’s translation theory, Huifen Zhu et al. (2020) discussed the post-translation editing rules of translators at the levels of text, referentiality, coherence, and naturalness while reviewing many types of mistranslation in machine translation.

4.4 Interdisciplinary Studies on Machine Translation and Linguistics

The development of linguistic research is essential for machine translation breakthroughs. Huijun Zhao and

Guobin Lin (2020) explored the linguistic path of machine translation intelligence, pointing out that the current machine translation lacks the guidance of linguistic principles and entirely relies on mathematical model algorithms. The study further proposed that the fully intelligent machine translation needs sufficient corpus and scientific learning methods, i.e., solutions in the direction of linguistic studies.

In terms of specific linguistic theory application, Xi Wang and Yang Chen (2019) examined machine-translated texts from a functional linguistic perspective. The study compared the translations supplied by Google Translate, Baidu Translate, and the official translations. It focused on evaluating the translations in terms of formal equivalence, meaning equivalence, and functional equivalence in systematic functional linguistics. According to this study, humaneness is one of the apparent factors why machine translation cannot replace human translation. As a result, it is crucial in specific translation research and teaching to focus on using AI research and machine translation to aid in translation research and education as well as to help students build their skills in evaluating, deciphering, editing, and amending machine-translation output.

In terms of translation processing of specific language difficulties, several studies also explored lexical, semantic, and syntactic perspectives. First, at the lexical and semantic level, Chi Yang and Xianze Yang (2020) combined linguistics and computer science research to discuss the processing of words. For example, polysemous words can be adequately treated in different specialized contexts, and a machine translation system can be used to build a dictionary of synonyms. Also, for polysemous words, Shuying Zhu (2019) examined the mistranslation brought on by multiple meanings in Chinese-English online machine translation, and five sentences with varying treatments in “Google Translate” and “Youdao Translation” were analyzed with the aid of NLP-IR-ICTCLAS, a tool developed by the Chinese Academy of Sciences. The study concluded that the error in translation is due to machine translation’s inability to effectively segment words and the misjudgment of the original text as ungrammatical; as a result, it suggested enlarging the corpus and enhancing the treatment of polysemous words as a fix.

Second, at the syntactic level, long sentences have always been a focus of translation research. Jing An (2018) proposed a method of English long sentence segmentation based on dependent syntactic analysis and sequence annotation, which showed that machine translation based on long sentence segmentation could achieve better results.

The processing of chunks is also an important topic at the syntactic level. Fumao Hu and Keliang Zhang (2018) studied bilingual chunk correspondence for machine translation. They proposed a theoretical model for chunk composition analysis (i.e., triangular graphs with constructions, chunks, and words) and an analysis of chunk characteristics within the framework of constructional grammar. In addition, the study presented a theoretical basis and empirical support for the establishment of an English-Chinese construction database of business letters, which explored a new way to improve the translation accuracy of machine translation.

The subject-verb-object structure is a core category in syntactic structure. Using 50 English news articles as the corpus, Yuan Chen (2020b) analyzed the lexical and syntactic errors produced by machine translation in the subject-verb-object structure, concluding that grammatical rules must be added to the machine translation system’s foundation in order to improve the quality.

Chinese is a topic-prominent language, and arguments of verbs can be omitted when the context is clear, thus also posing difficulties for linguistic analysis and cross-linguistic translation. Taking Google online translation as an example, Le Chang and Jun Gao (2020) summarized two types of subject mistranslation in Chinese-English machine translation, i.e., linguistic mistranslation and pragmatic mistranslation. They proposed post-translation editing strategies of complementary subject translation and subject adjustment.

Special and complex sentences have cross-linguistic characteristics and are naturally an important topic in linguistic research. The differences in the use of English-Chinese passive sentences are significant. Yiwei Ji (2019) took the translations of Google Translate and Youdao as an example to evaluate the quality of machine translations of Chinese passive sentences and found that Youdao is better than Google in terms of translation quality, thus illustrating the importance of building a sufficient text database to improve the translation. Relative clauses have also traditionally been the focus of linguistic research. Taking Google machine translation as an example, Wenzhao He and Defeng Li (2019) used a combination of quantitative analysis and qualitative research to evaluate Chinese translations of English relative clauses. They discovered significant differences between machine-translated relative clauses and human-translated relative clauses. According to this study, features in forms are strongly correlated with the quality of translation.

4.5 Studies on Challenges and Applications of Machine Translation

The role of machine translation in cross-cultural communication is also a selected topic in previous studies. Luo Huifang and Caiqi Ren (2018) discussed localization and machine translation technologies as significant development drivers. This paper argued for the advancement of AI and the integration of machine translation with human translation, stressing the importance of machine translation and localization in enhancing the nation's cultural reputation while connecting with foreign audiences. Chunfang Zhou (2019) examined the issue of avoiding or depending too heavily on machine translation, which causes people to overlook the cultural components of "intercultural" conversation. The study emphasized that culture and thought are human attributes that cannot be substituted by machines. Hence machine translation cannot resolve every issue with cross-cultural communication.

Machine translation shows a continuing upward trend, and ethical issues will become increasingly prominent in translation practice. Wen Ren (2019) examined the changes that machine translation activities bring to the field of translation today and proposed that machine translation challenges traditional translation ethics, covering these main aspects: the principle of importance (translated vs. original), the principle of responsibility (who is responsible for the translation), the principle of loyalty (interpersonal, client, translator), the principle of fairness and justice (the interests of the translator), the principle of harmonious ethical relations (stakeholders), the potential risks of other non-ethical behavior (linguistic diversity and creative expression). The study offered solutions to unlock related issues, advocating the adherence to traditional ethical norms of translation (e.g., fidelity, loyalty, etc.) and the optimization of ethical statutes of creative responsibility as guidelines for machine translation activities in the technological age.

5. Concluding Remarks

The current research on machine translation reflects the trend towards human-machine cooperation and language services. The advantage of machines lies in their speed, cost, and uniformity, while the disadvantage lies in their inability to think like humans. In the face of these technological trends and advancements, what the study and teaching of languages can do is to provide language services in the context of "collaborations between humans and machine and use of machines for humans," mainly through the following initiatives:

First, it is crucial to strengthen interdisciplinary research in linguistics, literature, and culture in the service of machine translation.

In terms of the overall research status, according to the distribution of the keywords of the current relevant journal papers, as shown in Figure 2, in the studies involving the keywords of "neural machine translation," "machine translation system," "machine translation research," "human translation," "statistical machine translation," "post-translation editorial" and other related keywords, all must return to the study of human cognition and expression of emotions in language, and all must be built on the basis of the large-scale corpus. It follows naturally that linguistic research is essential. The overall trends of the published results, as shown in Figures 3–5, indicate that the scholarly community is still relatively underrepresented in language research and that more needs to be done to promote interdisciplinary study between language research and other disciplines.

In terms of the difficulties in breakthroughs, the discussion in Part 4 shows that most of the errors found in machine translation in current research are traditional topics of linguistic analysis, and some are areas where linguistic research needs to be expanded, such as missing subjects, confusion between singular and plural, misuse of articles, tense errors, lexical ambiguity, misuse of passive voice, misuse of relative clauses, implied meaning, etc. Therefore, topics that have not yet been systematically investigated need to be targeted toward cross-linguistic differences, human cognition, cultural characteristics, etc. All of these require close interdisciplinary cooperation between linguists and computer scientists to realize the combination of literature and science.

Secondly, the training of localized foreign language services should be strengthened to enhance students' ability to use technology and renovate it. As seen in the discussion in Part 4, there have already been some studies on translation education under the guidelines of human-computer cooperation, such as Ying Qin (2018), Hanji Li (2018), Chaowei Zhu (2018), Xi Wang, Yang Chen (2019), Xiangling Wang (2019), Xiangling Wang, Tingting Wang (2019), etc. However, in response to the need of the new era, more focused and organized studies are required for the educational reform of translation as well as the development of language services. Under the premise of consolidating the fundamental language abilities, and literary and cultural skills, the technical course contents need to be increased, and teachers and students need to be guided to have the consciousness of language service, the quality of technical learning, and the demand of market.

The advancement of machine translation is certain to make strides as a result of groundbreaking developments in artificial intelligence. The daily lives, jobs, and thoughts of individuals are being transformed by the technological revolution. Everyone working in the field of language study and translation is truly a part of the revolutionary wave that is propelling human civilization forward. Moving forward or retreating are our only options.

Acknowledgments

The 11th China Foreign Language Education Fund is funding this project, “Research on the international communication ability of college students to tell China’s stories in foreign languages in the era of digital intelligence” (ZGWYJYJJ11Z006). The project’s principal investigator is Prof. Li Fan.

References

- An, J. (2018). Machine translation of long English sentence based on dependency parsing and sequence labeling. *Journal of Lanzhou University of Technology*, 44(01), 100–103.
- Bar-Hillel, Y. (1958). Some linguistic obstacles to machine translation. In *Proceedings of the Second International Congress on Cybernetics* (pp. 197–207). Namur.
- Bar-Hillel, Y. (1959). *Report on the state of machine translation in the United States and Great Britain*. Jerusalem: Hebrew University.
- Cai, X. J., & Wen, B. (2021). Statistical analysis on types of C-E machine translation errors: A case study on C-E translation of publicity text. *Journal of Zhejiang Sci-Tech University*, 46(02), 162–169.
- Chang, L., & Gao, J. (2022). A study of subject mistranslation in Chinese-English machine translation: A case study on Google online translation. *Overseas English*, 21, 157–158, 168.
- Chen, C. (2004). Searching for intellectual turning points: progressive knowledge domain visualization. In *Proceedings of the National Academy of Sciences of the United States of America* (101, Suppl 1). <https://doi.org/10.1073/pnas.0307513100>
- Chen, W. (2020). Deconstruction of translators’ subjectivity by machine translation: The future of human translation. *Foreign Languages Research*, 37(02), 76–83.
- Chen, Y. (2020a). The lexical errors of subject-verb-object structure in English-Chinese machine translation. *Journal of Henan University of Science and Technology (Social Science)*, 38(02), 31–36.
- Chen, Y. (2020b). On syntactic errors in English-Chinese machine translation of SVO structure. *Journal of Luoyang Normal University*, 39(09), 85–91.
- Feng, Z. W. (2018). Parallel development of machine translation and artificial intelligence. *Journal of Foreign Languages*, 41(06), 35–48.
- Feng, Z. W. (2020). Rosetta stone and machine translation. *Foreign Language Research*, 01, 1–17.
- Gao, L. L., & Zhao, W. (2020). An overview study on machine translation. *Foreign Languages in China*, 17(06), 97–103.
- Gulipina, J. (2020). Machine translation and human translation in the context of artificial intelligence: competition and integration. *International Public Relations*, 05, 176, 178.
- Han, N. (2018). Evaluation of Google and Youdao Chinese-English neural machine translation systems. *Journal of Shanxi Institute of Energy*, 31(05), 123–124.
- Hao, Y. L., Hu, Q., Li, G. L., & Liu, J. (2019). Analysis on the application and prospect of machine translation. *English Square*, 09, 20–21.
- He, W. Z., & Li, D. F. (2019). Evaluation of English-Chinese machine translation of relative clauses: A case study on Google machine translation. *Chinese Science & Technology Translators*, 32(03), 30–34.
- Hinton, G. E., Deng, L., Yu, D. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 6, 82–97. <https://doi.org/10.1109/MSP.2012.2205597>
- Hinton, G. E., Osindero, S., & Yee-Whye, T. (2006). A Fast-Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7). <https://doi.org/10.1162/neco.2006.18.7.1527>
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786). <https://doi.org/10.1126/science.1127647>

- Hu, F. M., & Zhang, K. L. (2018). An MT-oriented study of corresponding bilingual lexical chunks. *Technology Enhanced Foreign Language*, 03, 83–87.
- Huang, W. H. (2018). Application of machine translation in modern translation. *Journal of Jiamusi Vocational Institute*, 09, 333, 335.
- Hutchins, J. (2010). Machine translation: A concise history. *Journal of Translation Studies*, 13, 29–70.
- Ji, Y. W. (2019). Evaluation of Chinese passive voice machine translation: A case study on Google and Youdao. *English Square*, 03, 55–57.
- Jing, S. (1959). Linguistic Translation Series, No. 2, 1959, Special Issue on Machine Translation. *Foreign Language Teaching and Research*, 03, 190–156.
- Li, H. Z., & Chen, H. Q. (2018). Analysis of dynamic scientific knowledge map of machine translation research in China (2007–2016): Based on word frequency of linguistics and computer science journals. *Journal of Xi'an International Studies University*, 26(02), 99–104.
- Li, L. (2020). Relationship between machine translation and human translation. *International Public Relations*, 09, 276–277.
- Li, R. (2019). Pros and cons and solutions of machine translation and computer aided translation. *Popular Literature and Art*, 03, 178–179.
- Lian, Y. T. (2020). Analysis of machine translation errors in software localization translation. *Technology and Economic Guide*, 28(17), 130, 129.
- Luo, H. F. (2018). Foreign cultural communication from the perspective of localization and machine translation. *Chinese Science & Technology Translators Journal*, 31(02), 24–26, 54.
- Osindero, S., Welling, M., & Hinton, G. E. (2006). Topographic Product Models Applied to Natural Scene Statistics. *Neural Computation*, 18(2). <https://doi.org/10.1162/089976606775093936>
- Pang, Y. Y. (2019). Relationship between machine translation and human translation: from the perspective of machine translation and computer-aided translation. *Popular Science*, 11, 164–165.
- Poibeau, T. (2017). *Machine Translation*. Boston: The MIT Press. <https://doi.org/10.7551/mitpress/11043.001.0001>
- Qin, Y. (2018). An analytical study of neural network machine translation and its impacts on translation teaching. *Technology Enhanced Foreign Language*, 02, 51–56.
- Ren, H. J., Ma, X. L., & Shi, J. H. (2020). Exploring the humanistic factors in the post-editing of machine translation. *China High and New Technology*, 20, 124–125.
- Ren, W. (2019). Ethical challenges posed by and ethical principles proposed for machine translation. *Shanghai Journal of Translators*, 05, 46–52, 95.
- Song, S. Z. (2019). On the future relationship between machine translation and human translation. *Future and Development*, 43(02), 25–30.
- Su, J. S., Chen, J. S., Lu, Z. Y., Dong, Y. F., Kang, L. Y., & Zhang, H. Y. (2020). A survey of document-level neural machine translation. *Technology Intelligence Engineering*, 6(05), 4–14.
- Tang, B., & Chen, S. (2020). A comparative analysis of translation of medical texts produced by online machine translation software. *Chinese Science & Technology Translators Journal*, 33(03), 23–26, 49.
- Tian, J. B. (2018). Exploring whether machine translation can replace human translation from the standard of translation. *Overseas English*, 13, 126–127.
- Wang, Q., & Ma, X. (2020). A Study on the Quality Assessment of Machine Translation from the Perspective of Problem Awareness. *Social Sciences in Hunan*, 06, 144–151.
- Wang, X., & Chen, Y. (2019). Functional linguistic analyses of machine-translated texts. *Technology Enhanced Foreign Language*, 06, 73–77, 84.
- Wang, X. L., & Wang, T. T. (2019). A comparative study of human translation and machine translation post-editing in E-C translation: Translation speed, quality and translators' attitude. *Foreign Language and Culture*, 3(04), 83–93.
- Wang, X. L., & Yang, Y. X. (2019). A study of domestic machine translation within 60 years based on the papers in Chinese foreign language core journals. *Journal of Hunan University (Social Sciences)*, 33(04), 90–96.

- Xiao, W. Q., & Gao, J. H. (2020). Assessing quality in machine-generated subtitles: A case study of “netease sight”. *Foreign Language and Culture*, 4(03), 95–105.
- Yang, C., & Yang, X. Z. (2020). Analysis of processing multi-category and polysemic words in the Chinese – English machine translation. *Journal of Southwest Minzu University* (Natural Science Edition), 46(06), 619–622.
- Yang, J. (2018). The dashing waves of machine translation revolution. *China Newspaper Industry*, 19, 107.
- Yang, Y. W. (2020). Post-editing of neural machine translation – Taking “Submarine Hydrodynamics” as an example. *Chinese Science & Technology Translators Journal*, 33(04), 21–23, 42.
- Zhang, F. L. (2020). On machine translation technology in legal translation. *Technology Enhanced Foreign Language*, 01, 53–58.
- Zhao, H. J., & Lin, G. B. (2020). A research on the linguistic solution of intelligent machine translation. *Technology Enhanced Foreign Language*, 02, 42–47.
- Zhou, B. J. (2018). Research on problems and solutions of machine translation. *Comparative Study of Cultural Innovation*, 2(19), 149–150.
- Zhou, C. F. (2019). Machine translation and intercultural communication: Development, problems and reflections. *Journal of Hubei University of Economics* (Humanities and Social Sciences), 16(11), 114–116.
- Zhu, C. W. (2018). The machine to replace human beings as a translator? A discussion of the relationship between science and humanity in the cultivation of translators and interpreters. *Foreign Language and Literature* (bimonthly), 34(03), 101–109.
- Zhu, H. F., Zhao, W. J., & Zhu, Y. F. (2020). On post-editing principles for online machine translation – A case study of “Double eight strategies”. *Chinese Science & Technology Translators Journal*, 33(02), 24–27.
- Zhu, J., & Gu, M. (2019). Corpus-based machine translation. *Modern Communication*, 17, 100–101.
- Zhu, S. Y. (2019). Mistranslation of Chinese polysemy from Chinese to English through online machine translation. *Journal of Nanjing Institute of Technology* (Social Science Edition), 19(03), 10–14.

Notes

Note 1. Related links: <https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>, Quoc V. Le & Mike Schuster.

Note 2. Related links: <https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html>, Mike Schuster, Melvin Johnson & Nikhil Thorat.

Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).