PAPER • OPEN ACCESS

Research on the Relations Between Machine Translation and Human Translation

View the article online for updates and enhancements.

Research on the Relations Between Machine Translation and Human Translation

Zhaorong Zong

School of Foreign Languages, Huangshan University, 39 Xihai Road, Tunxi District, Anhui Province, China

Corresponding author e-mail: Zongzhaorong@hsu.edu.cn

Abstract: Machine translation, as a kind of revolutionary technology, ignited a reform and transformation in the field of translation. It has gone through three stages, from early dictionary-matched machine translation to corpus-based statistical computer-aided translation, and then to neural machine translation with artificial intelligence as its core technology in recent years. Although neural machine translation in the age of artificial intelligence has made major breakthroughs in machine translation, some translation problems are far from being completely resolved. Language, as a living being, keeps on shifting and changing with regard to the context and discourse. Due to the various defects in machine translation, in order to make up for its deficiencies, this paper analyzes the problems with examples and puts forward a new model of machine translation plus human translation as well as its applications. In the age of artificial intelligence, computers with deep learning ability work together with human will produce translation works with high efficiency and better quality.

1. Introduction

Artificial intelligence, big data, cloud computing, and the Internet of Things have become new hot topics after the Internet and mobile Internet. Under the drive of related information and communication technologies, artificial intelligence has emerged with new algorithms, new technologies, and new experiences. A large number of artificial intelligence applications such as image recognition, speech recognition, translation, medical care, and even robots have emerged. With the renewal of artificial intelligence technology, it also has a profound impact on the translation industry that also undertakes information exchange and management. At the Boao Forum for Asia in April 2018, the speaker chatted and talked on the stage, while the screen next to him automatically converted the text into words and translated into English in real time. For a time, the artificial intelligence product of "simultaneous machines" quickly became the "web celebrity" in the public view. Following AlphaGo, the application of artificial intelligence in the language field once again caught the public's eye.

Based on the wave of localization triggered by the market-seeking investment of multinational corporations in the 1980s, the translation industry has seen a wave of rapid localization, which laid the foundation for the development of translation industrialization. [1] The foundation is embedding the traditional requirements of the translation industry into the industrialization of translation, so the translation became gradually commercialized. The cost, speed, and quality are regarded as the industry requirements for translation, and these three major guidelines lead the development and practice of the translation industry and become the main factors that measure the innovation of the translation industry model. Research on these issues is currently lacking systematic analysis at home and abroad.

Although at the practical level, various translation service companies headed by artificial intelligence continue to emerge, in addition to the Tencent company that provided simultaneous interpretation for the Boao Forum for Asia, Microsoft, Google, IBM, Youdao, and Sogou, etc. all developed its own intelligent translation tools.

However, there are no clear ideas on how to meet the individual needs and market positioning in the artificial intelligence environment, how to effectively select translation methods, and how to integrate the existing translation industry tools. Therefore, this paper tries to explore a better way to translate efficiently and effectively with a combination of machine translation and human translation, and then puts forward a framework of machine translation blending human translation under the artificial intelligence environment.

2.   Three Stages of Machine Translation: MT, CAT and NMT

Machine translation, sometimes referred to by the abbreviation MT is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one language to another. [2] The development of machine translation has been closely linked with the development of computer technology, information theory, linguistics and other disciplines. It has gone through three stages, from early dictionary-matched machine translation to corpus-based statistical computer-aided translation, and then to neural machine translation with artificial intelligence as its core technology in recent years.

In 1954, cooperated with IBM Corporation, Georgetown University in the United States has completed the first English-Russian machine translation experiment with the IBM-701 computer. It demonstrated the feasibility of machine translation to the public and the scientific field, thus unveiled the prelude of machine translation research. In the 1970s, with the development of science and technology and the increasing frequency of scientific and technological information exchanges among countries, the language barriers between countries have become more serious. The traditional manual translation has been far from meeting the needs and urgently requires computers to perform translation work.

Computer-aided translation originated in machine translation and is a translation technology that has been rapidly developed after the digitized translation materials. Computer-aided translation is different from the previous machine translation software. It does not rely on the automatic translation of machines, for it is the completion of the entire translation process with the participation of human. [3] The working principle of computer-aided translation software is: TM (translation memory) plus MT (machine translation) plus HT (human translation/proofreading). Since the 1980s, with the emergence of corpus and the increasing demand of translation markets for translation efficiency, CAT has borrowed from machine translation methods to strengthen the management of translation memories and terminology databases and entered a new era. As a tool to assist translators, computers can help translators to improve translation efficiency when processing translations with high repetition rates, ensure translation quality, and optimize the translation process. Among them, memory database is the core part of computer-assisted translation. For the exact match sentences, CAT can produce translation simultaneously with the help of the database, while for the partial match ones, CAT needs to resort to manual proofreading. The working mechanism of CAT is shown in Figure 1:
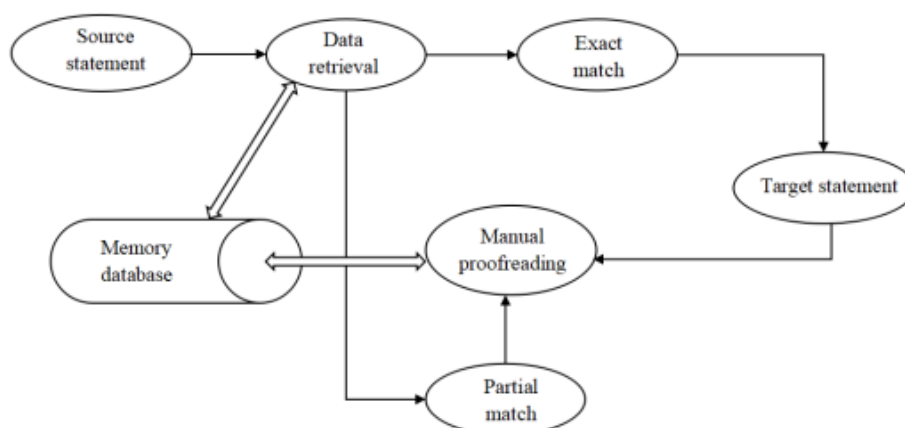
Fig 1: Working Mechanism of Computer-aided Translation

Trados is currently the most widely used CAT software in the world. Based on the principle of translation memory, Trados supports bidirectional translation between 57 languages and supports all popular document formats without the need for users to typeset. (DOC, RTF, HTML, SGML, XML, FrameMaker, RC, AutoCAD DXF, etc.) At the same time, its comprehensive auxiliary functions, such as time, metrics, tables, automatic replacement of fixed formats, etc., can help customers greatly improve work efficiency.

In addition, Google's online computer-assisted translation platform also supports translation from English to more than 50 languages (Online Computer Aided Translation System). The Google Translator Toolkit (GTT) integrates the Google Translate, WYSIWYG editor, open rating system, sharing system, Wikipedia, and Knol. Currently, GTT translations support some common file types, including: HTML file (.html), Word file (.doc), Open office file (.odt), Text file (.txt), Rich text file (. Rtf) format. The translation tools of GTT are mainly Translation memories and Glossaries. The auxiliary tools are Google's original machine translation tools (Google Translate) and dictionaries.

Neural Machine Translation (NMT) is the newest phase of machine translation with the development of artificial intelligence. Its theory is based on the theory and techniques of natural language understanding (NLU), natural language processing (NLP), machine translation (MT), translation memory (TM), and statistics-based machine translation (SMT)as well as deep learning. [4]As a new theory and a new discipline, it broke through the bottleneck of computer translation technology and achieved high-quality machine translation. Its characteristics are using big data and cloud computing backstage as computing platforms based on the new developments in the field of neural networks, continuously receiving various training data to the backstage based on the mobile Internet, and carrying out features data mining and rapid training through deep learning capabilities. [5] In 2016, Google introduced Google's Neural Machine Translation System (GNMT) and put GNMT into the very difficult translation production of Chinese-English language pairs, which has aroused great concern in the industry. Google once again made a major breakthrough in the field of machine translation and launched the GNMT, a neural machine translation system after its introduction of the phrase-based machine translation system Google Translate a decade ago.

3.   Interaction of Machine Translation and Human Translation
The automatic translation techniques that machine translation relies on include speech translation and word translation. The four key technologies are word analysis, grammar analysis, meaning analysis and style analysis. Its working process is: first divide the sentence into each word and clear the meaning of the word through the electronic dictionary stored in the machine database, then analyze the meaning of the sentence according to the grammar rule, and transform it into concept construction, then use the language model to generate the target language. [6] Among them, the language model is

the intermediate language between the original language and the target language, through which various languages can be translated into another desired language. If coupled with bidirectional translation software, the automatic translation system can translate multiple languages.

The evaluation of machine translation system is mainly based on the following seven factors: translation quality, application efficiency, work style, use environment, maintainability and expansibility, price ratio on machine translation system performance, robustness . [7] Among them, the quality of translations has always been regarded as the most important and critical indicator of machine translation evaluation.

Although neural machine translation in the age of artificial intelligence has made major breakthroughs in machine translation, some translation problems are far from being completely resolved. NMT will still make some major mistakes that human translators never make, such as misspellings and mistranslations of proper nouns or rare terms, and the translation of sentences alone without regard to the context of their paragraphs or pages. Therefore, it is difficult to achieve "full automatic high quality" translation.

For example, in the Boao Forum for Asia in April 2018, Tencent simultaneous interpretation based on neural machine translation for the first time made several mistakes, such as vocabulary repetitions, misuse of phrases, and so on. The Tencent translation official report interpreted the technical problems in some of the translation cases and stated that Tencent simultaneous interpretation has indeed experienced a mistake and wrong answered a few questions in the face of the complicated language environment and professional content on the Boao Forum for Asia.

In response to these problems, Tencent has called it the phenomenon of "wordless repetition, capitalization, and massive character confusion." This problem is mainly caused by the frequency of bilingual switching between Chinese and English. When the sound source is continuously converted between the two languages, the Chinese and English recognition engines in the background will start working at the same time, which will cause the two recognition engines to "rack" each other, and the translation result can only be selected in one language for output. In addition, each of the guests' tone words "ahah" will be translated exactly. All these put together led to mistakes. [8]

This problem is caused by the uncertainty of deep learning algorithms including machine translation of neural networks. In certain circumstances, there is a certain probability of triggering translation bias. In the guest speeches, there were common repetitions in oral English, such as "for for for" and "that's that's that", which were equivalent to the Chinese "em em em" and "that and that", and the translation engine just amplified this repetition, which results in translation errors. Finally, for the "一带一路" translation mentioned above, Tencent said that from the live screenshots, the speaker said "the road and belt" while the correct translation of "一带一路" should be "the belt and road". The different order leads the machine to interpret literally. "For fixed phrases, AI interpretation can accurately translate the correct expression of the speaker, but if the speaker's expression is biased, the accuracy of machine translation will be greatly reduced." Tencent official respond. [8]

As far as the current situation is concerned, most of the machine translation software with a relatively high market share is "pure translation". Not only does the thesaurus have limitations, but it also cannot be easier for the most appropriate translation to be accurately selected from the possible meanings of proverbs, sayings, and Internet buzzwords, and it is difficult to break the flexible application of natural language in grammar, rhetoric, and logic. problems or big mistakes will occur easily when it is out of a particular context. It is even more difficult to accurately translate poetry, novels, and so on.

However, the artificial intelligence translation software with deep learning ability does not have powerful translation ability from the beginning, but gradually improves the translation level by deep learning. [9] Due to the various defects in artificial intelligence translation, in order to make up for its deficiencies, many translation companies are trying new models of artificial intelligence translation blending human translation, which utilizes the corpus of computers to do the basic translation, then human proof-readers will correct the errors of the machine translation with feedback to the memory database. And finally the translated text was further revised by human editing to strive for greater efficiency while ensuring quality. The process is illustrated in fig. 2.
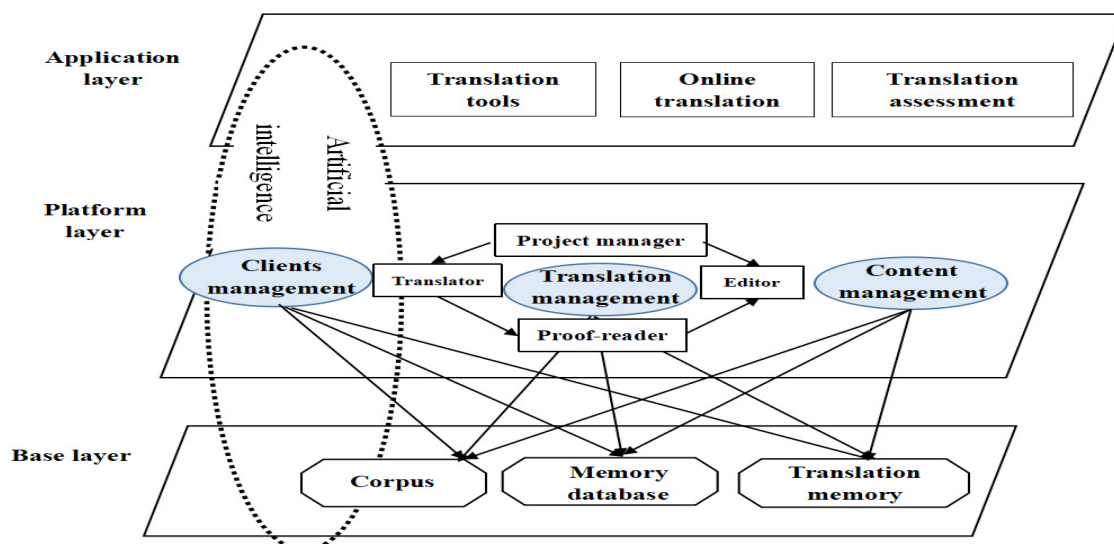

Fig 2:   A framework of machine translation blending human translation

In fig 2, artificial intelligence with deep learning is the core of translation process with reference to three layers of translation. A translation company receives orders from clients, AI analyses the order and assigns to a matching project manager who is responsible for the specific translation management. Then, with assistance of corpus, memory database and translation memory, the project is done primarily by computers relating to different content. Human participate in the platform layer whereas AI with deep learning interferes with the three layers of translation.

AI will learn accurate word expressions and human language habits based on the results of human translation and revision, thereby enhancing their translation capabilities so that they can bring more natural translation experience to people in the future. As for deep learning, deep-seated neural networks are used to simulate the human brain to interpret and analyze data such as images, speech, and text, which is a hot area in machine learning research. At the same time, the computer will also track the results of human translation and help translators avoid low-level errors. [10]Artificial intelligence translation and human translation create a virtuous circle. In addition, the translation memory can also be stored, improved and shared, and the combination of language technology and systems can play an important role in the translation project management and other modules of translation.

4. Conclusion

From the current point of view, both machine translation and human translation have their own application scenarios and services. For example, machine translation can provide tens of millions of translations per day, and its ability to quickly grasp new terminology customizations is difficult for human translators to do. It is particularly effective in domains where formal or formulaic language is used. While the translators' interpersonal understanding and fuzzy semantics integration is not easy for machine translation to achieve in the short term. [11]At present, with the emergence of large corpora and the maturity of NMT with deep learning ability, more and more companies have already launched relevant machine translation products or services, and the market is very competitive with the updating of the latest technology. From the current trend, if machine translation and human translation can truly complement each other, the efficiency and quality of translation services will be greatly enhanced. The future development direction of translation technology research should be a combination of multiple research methods, complementary advantages, and hybrid modelings. [12].

References:

[1] Wang Chuanying, Cui Qiliang, [J]. Chinese Translations Journal, [J]. 2010 (4).76-79
[2] https://en.wikipedia.org/wiki/Machine_translation
[3] McEnery , T .&Wilson, A.Corpus Linguistics [M] .Edinburgh University Press .1996.34-35
[4] Halliday, Mark Linguistics and machine translation [A]. In J. Webster (ed.)Computational and Quantitative Studies [C]. Beijing: Peking University Press, 2007. 166
[5] Anastasiou D，Gupta R．Comparison of Crowdsourcing Translation with Machine Translation [J] Journal of Information Science，2011，37(6) : 237-238
[6] Allen J．Post-editing［A］．In Somers H (ed.)．Computers and Translation: A Translator's Guide ［C］．Shanghai: Shanghai Foreign Language Education Press，2012.297 -303
[7] Kruger H．Training editors in universities: Considerations，challenges and strategies［A］．In Kearns J (ed.)．Translator and Interpreter Training: Issues，Methods and Debates ［C］．London: Continuum International Publishing Group，2008.39 -65
[8] http://tech.sina.com.cn/roll/2018-04-11/doc-ifyteqtq7855682.shtml
[9] Wang, Wenjie Principles and Applications of Artificial Intelligence. Beijing: People's Post and Telecommunications Press. 2004:123-126
[10] Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR015).Boston, Massachusetts, USA: IEEE, 2015.478-485
[11] Wray, Alison．Formulaic language: Pushing the boundaries ［M］．Oxford: Oxford University Press，2008: 8．
[12] Foster I, Kesselman C, Nick J, et al. Computer Grid Services for Distributed System Integration[J]. IEEE Computer, 2002, 35(6):37-46.