Study on the Machine Translation of Chinese Characteristic Terms: A Case Study of China's 2019 Government Work Report

Wu Jinrui[1] and Tian Jianguo[2]

[1]School of Foreign Studies, Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an Shaanxi, 710072, P.R.China
Email: 914650881@qq.com
[2]School of Foreign Studies, Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an Shaanxi, 710072, P.R.China

[1]**Corresponding Author**

Abstract：

Since the 21st century, with the rapid development of information technology and computer industry, Artificial Intelligence and Big Data have gradually penetrated into all areas of society. In the field of translation, the rapid rise of Machine Translation (shortened as MT) with the help of Artificial Intelligence and Big Data has also become the focus of public attention. In the global information era, MT has a strong connection with international cross-cultural communication. It can also clear the communication barriers caused by language problem and promote the cultural exchange among different countries. However, since MT is not yet perfect, whether it can be applied in the translation of Chinese characteristic terms needs further investigation. This paper selects the 2019 Government Work Report as the research material, and chooses Google Translate as the tool. By comparing the translation of MT with the official version of Human Translation (shortened as HT), this paper aims to find out the existing problems in MT and prospect of its further developing tendency.

Keywords: Machine Translation (MT); Chinese Characteristic Terms; Cross-Cultural Communication

1. Introduction

MT has extensive application in the field of scientific and technological documents as well as in legal documents, industrial patents, specifications, etc., and its popularization has produced tremendous social and economic benefits. However, due to the imperfection of MT, it is still unable to be competent for the translation tasks of literature and government documents.

1.1 Machine Translation

MT, also known as Automatic Translation, refers to the use of computers to implement the conversion between two natural languages. As early as the 1930s and 1940s, scientists began to design translation machines. In 1947, Warren Weaver first proposed the concept of MT as well as its related technical difficulties. He believed that translation was an encryption task which could be decrypted by a computer (Hutchins, 1986). Since then, a lot of experiments and research have been carried out in this field.

John Hutchins pointed out that the main linguistic problems encountered in MT systems fall under four main aspects: lexical, structural, contextual, and pragmatic or situational. In each case the problems are primarily caused by the inherent ambiguities in natural languages and the lack of direct equivalences in vocabulary as well as structural differences between one language and another (W. John Hutchins, 2003). In Martin's opinion, he thought the features of a language that make it suitable for fully MT are restricted vocabulary; restricted syntax; restricted meaning for each word in a given category(Martin Kay, 1982).

In China, Research in this field began in 1965(Zhang Wen, 1993). MT is based on the formalization and procedural processing of bilingual descriptions, comparisons, and matching results to achieve automatic translation of different languages. Its characteristics are mainly represented by automation, mechanization, sentence-based translation units, second-degree imitation, and language limited environmental constraints (Hu Kaibao & Li Yi, 2016). Meanwhile, many fruitful results about the basic and applied research in this field have been achieved, such as the studies about the post-editing capabilities, involving its evaluation system

and error recognition (Feng Quangong & Liu Ming, 2018).

In the 21st century, with the help of Artificial Intelligence, Google applied deep learning of Big Data to MT, which changed the traditional translation mode. In 2006, Google Translate further introduced a new type of statistical analysis technology called "Statistical Machine Translation". It applies Big Data analysis technology to the field of learning skills, curriculum design and translation, analyzes the language Big Data being generated and stored, and gains the capabilities required for translation and language learning (Henry R & Venkatraman S, 2015). Besides, with the development of Big Data computing technology in processing language, machine algorithm can effectively analyze huge amount of complex text information, which opens up a new way of language learning (O'Halloran S, 2016).

1.2 MT and Chinese Characteristic Terms

Dai and Yi classified the translation process of MT into three methods: (1) Rule-based, a system conversion process based on large-scale bilingual dictionaries, bilingual conversion rules, and target language generation rules; (2) Statistics-based, which selects the most suitable language probability model and translation probability model, and then estimates the model probability parameters according to the bilingual text; (3) Case-based, by building a massive bilingual corpus to retrieve the most similar source language for matching (Dai Xinyu & Yi Cunyan, 2004). The translation process of Google Translate is to first search the entire sentence in the internet corpus, then quickly count all the relevant translation outcomes, and finally take the one with the highest frequency as the result (Xie Xiaoping & Zhou Limin, 2019). The translation process of Google Translate is basically consistent with the views of Dai and Yi, while its advantages lie in its huge corpus and powerful algorithms benefiting from Big Data technology.

As for Chinese characteristic terms, which includes the following characteristics: (1) unique to Chinese culture (2) rich Chinese culture connotation (3) no equivalent words found in target language (Cai Honggai, 2011). In view of the current work process of MT, many problems will appear in the translation process of Chinese characteristic terms. Therefore, this paper will choose Google Translate as the tool and select the official translation version of 2019 Government Work Report as the reference, conducting the analysis in the following aspects: four words idioms, abbreviations, political neologisms and terms with Chinese background.

2. Case Analysis

The Government Work Report is an official government document that puts forward the expected goals and plans for government work in the coming year according to the characteristics and the needs of economic and social development (Xu Zhi, Zhang Jianchao, 2020). It contains a lot of political, economic, cultural and social terms which own distinctive Chinese characteristics as well as the difficult points in translation.

2.1 Four Words Idioms

Example (1) 为政以公，行胜于言。

Official Translation (OT): Government works for the public; words can't compare with actions.

Google Translation (GT): Doing justice for politics is better than words.

Example (2) 文山会海

OT: Mountains of documents and endless meetings

GT: Wenshan Huihai

Example (3) 不忘初心，牢记使命。

OT: Keeping in mind our Party's founding mission.

GT: Don't forget your original intention and keep your mission in mind.

Example (4) 就业是<u>民生之本</u>、<u>财富之源</u>。

OT: Employment is the cornerstone of wellbeing, and the wellspring of wealth.

GT: Employment is the foundation of people's livelihood and the source of wealth.

Modern Chinese Dictionary explains idioms as simple and incisive stereotyped phrases or short sentences that have been used for a long time. Most Chinese idioms are composed of four characters, and they all have their origins. Some idioms are not difficult to understand literally, while the meaning of some idioms can only be obtained through their background or allusions (2005). Each nation has its own unique culture, so the idioms used by people in different nations have their own distinctive features in structure and characteristic. Therefore, when translating Chinese idioms, we should try our best to maintain the national and cultural structure of the original idioms. At this time, the strategy adopted can be literal translation, free translation or literal translation plus interpretation (He Yongbin, 2016). In Example (1), generally speaking, both GT and OT adopt literal translation. On this basis, the OT appropriately adds the subject——"government", while Google can't infer the subject of the sentence according to the context, which makes the main idea of the idiom not conveyed effectively. Besides, Google still has a deviation in the understanding of the word——"公". In the idiom "为政以公", "公"means the broad masses of the people, that is, the "public", not the one in "公平正义", which means "justice". As for the purpose expressed by the idiom, Google is also unable to recognize it correctly. The government's works are for the people, not for the politics. What's more, in this sentence, the two idioms express two main ideas respectively, which should be translated separately like the official version, while Google mistakenly integrates the two idioms, which makes the translation completely distort the original meaning. It is also worth noting that the OT uses negative sentences to highlight the importance of "actions" than "words", and Google obviously can't do this. On the whole, Google Translate failed in Example (1).

Chen Dan once pointed out that idioms are short and meaningful, with profound ideological connotation and delicate emotional meaning. In the process of translating Chinese idioms, we should not only understand the original meaning of idioms, but also deeply understand and interpret their profound cultural significance. We should incorporate the analysis of cultural factors into the translation strategies in order to achieve the vividness and accuracy of translation (Chen Dan, 2014). In Example (2), the idiom uses a metaphor, which means the mountains of documents and seas of meetings. There are many idioms with similar expressions in Chinese, such as, "浩如烟海(There are a tremendous amount of books that are as vast as the misty ocean.)", "汗牛充栋(It is used to describe a large collection of books, which means if the books are carried by the bull, the bull will sweat with fatigue; if the books are put in the room, the whole room will be filled.)". Therefore, on the basis of understanding the connotation of the idiom, the strategy of free translation should be adopted just like the official version, but Google

Translate is totally incapable of recognizing the deep meaning of the original. Accordingly, it is unable to apply some appropriate translation strategies like free translation or amplification. The only thing it can do is to translate Chinese into phonetic transcription of Chinese character that English readers can't understand.

However, there is an interesting phenomenon in Example (3), "不忘初心，牢记使命" was proposed by President Xi Jinping in the report of the 19th National Congress on October 18, 2017, it was subsequently rated as one of the top ten Chinese media buzzwords in 2017 by Beijing Language and Culture University. Therefore, if the original text is widespread, Google Translate is able to translate it into the popular version, but it can't choose the most appropriate one which is suitable for the context, because in 2019 Government Work Report, these idioms are so well known that English readers can understand them without translating them at all, so that the official version makes the translation more economical and efficient on the basis of readers' understanding. Obviously, Google cannot adopt this translation strategy of omission.

As to Example (4), Google not only retains the structure of the original, but also roughly conveys the connotative meaning of idioms, while in the choice of words, Google is still inferior to the official version. In other words, Google has achieved faithfulness and expressiveness without conveying the elegance of the original text.

2.2 Abbreviations

Example (5) "五位一体"总体布局，"四个全面"战略布局。

OT: Five-sphere integrated plan, Four-pronged comprehensive strategy.

GT: Five in one overall layout, Four comprehensive strategic layout.

Example (6) 一带一路

OT: The Belt and Road Initiative (BRI).

GT: Belt and Road.

Example (7) 三去一降一补

OT: Five priority tasks.

GT: Three go one drop one make up.

Example (8) "四个意识"，"四个自信"

OT: Maintain political integrity, think in big-picture terms, follow the leadership core, and keep in alignment. Confidence in the path, theory, system, and culture of socialism with Chinese characteristics.

GT: Four Consciousness. Four Self-confidences.

Lin Yutang, a remarkable language user, in his masterpiece *My Country and My People*, pointed out that the extreme monosyllabic nature of Chinese has created a very concise style. The skillful use of this style means the perfect choice of words (Shen Yihong & Hao Zhidong, 1994). Therefore, the multiple abbreviations used in Chinese are very common. Abbreviation, a brief form of terms or phrases, has the characteristics of regional, professional, contemporary, informative, concise and convenient. The use of this structure conforms to the economy principle of speech. If a term or phrase is long and complex, and it is often used, an abbreviation will be created to condense the information. Zeng Jianpin believed that most Chinese abbreviations belong to the field of politics and culture. They have strong characteristics of the times and nationality. Moreover, digital abbreviations account for the majority and may become buzzwords in a certain period of time. Zeng also gave four suggestions about the translation of

abbreviations: explanation translation, literal translation, literal translation with notes and free translation (Zeng Jianpin, 2003).

In Example (5) and Example (6), neither the official nor the Google version makes explanation to the abbreviation, both of them choose the literal translation, because these abbreviations are like "一国两制 (one country, two systems)", "和平共处五项原则 (the five principles of peaceful coexistence)" and "三民主义 (three people's principles)", which are often seen in foreign media and familiar to English readers. It is because the target readers have a general understanding of the abbreviation, so explanation translation is not adopted. However, in Example (5), compared with the official one, Google Translate doesn't add a modifier after the quantifier, so that the integrity of this strategy or plan cannot be reflected adequately. In Example (6), Google also has similar flaws. In the official version, "initiative" has been added to indicate the character of the abbreviation, which is an initiative rather than a strategy or plan, this also shows China's attitude of putting forward this view. While Google only translates it literally and does not express this deep meaning.

In Example (7), neither GT nor OT conducts explanation translation. It is because the connotation, "cutting overcapacity, reducing excess inventory, deleveraging, lowing coasts and strengthening areas of weakness", is implied in the context, at this time, it is not necessary to express the specific content of the abbreviation. This strategy is a solution to the recurring problems in a given context, and also conforms to the economy principle of speech. Therefore, free translation is used in the OT, and the connotative meaning of the abbreviation is summarized into "five tasks". As far as the translation provided by Google is concerned, it does not have the ability to analyze and summarize at present, so the result is confusing.

In Example (8), these two abbreviations may be unfamiliar to English readers, because compared with Example (5) and Example (6), they are China's national policies, which often appear in international occasions. While the abbreviations in Example (8) are a proposal for the Chinese people, which rarely appear in international media or international occasions. Therefore, the explanation translation should be adopted to make them easier to understand. Similarly, Google can't identify what its four consciousness and four self-confidences refer to, it can only translate the literal meaning and cannot make an explanation translation due to its functional limitation. For English readers, the Google version is too general and difficult to understand.

2.3 Political Neologisms

Example (9) 结构性去杠杆

OT: Structural deleveraging

GT: Structural deleveraging

Example (10) 供给侧结构性改革

OT: Supply-side structural reform

GT: Supply-side structural reform

Example (11) 社会主义核心价值观

OT: Core socialist values

GT: Core values of socialism

Example (12) 精准脱贫

OT: Precision poverty alleviation

GT: Precision poverty alleviation

A neologism is a relatively recent or isolated term, word, or phrase that may be in the process of entering common use, but that has not yet been fully accepted into mainstream language (Anderson, 2006). As the name suggests, political neologism refers to the new word appearing in the political field, which usually relates to the major political events and conferences in China as well as the major policies formulated by the Party and the state (Song Pingfeng & Zhang Chunyan, 2016). As we all know, most of the contents in the Government Work Report are the latest events or policies, so a large number of political neologisms will appear in it. Therefore, it is obviously important to translate these political neologisms correctly.

Many scholars have put forward their own views on the translation of political neologisms. Wang Xue and Cai Lihua believed that the political neologisms are highly generalized and concise, closely related to the development of the times and involve a wide range (Wang Xue & Cai Lihua, 2013). Wang Fei summed up the five characteristics of political neologisms: relatively new in time; appearing in a specific historical stage; having strong policy nature; paying attention to communication effect; and containing Chinese culture (Wang Fei, 2010).

Therefore, there are several requirements for the translation of political neologisms. First of all, neologisms that need to be translated are often cutting-edge. If the translator lacks the background knowledge of the corresponding discipline, it will be difficult to understand the original text. Secondly, translator needs to master a certain amount of historical and cultural knowledge. Thirdly, translator should have the ability of political sensitivity. Finally, the translation should be as concise as possible.

From Examples (9), (10), (11) and (12), it is not difficult to find that their structures are relatively simple and do not involve too deep cultural background. Therefore, without too much analysis, a strategy of literal translation can be used. With its enormous and constantly updated corpus, it is easier for Google Translate to overcome the above difficulties, and completes the translation accurately and quickly. Therefore, Google and official version are basically the same.

2.4 Terms with Chinese Background

Example (13) 更接地气、更合民意。

OT: Respond to people's desire, and better satisfy public demand.

GT: More grounded and more popular.

Example (14) 摘帽县和脱贫人口

OT: Counties and populations that have recently been lifted out of poverty.

GT: Chimo County and poverty alleviation population.

Example (15) 决不能"新官不理旧账"

OT: New officials must not be allowed to get away with ignoring obligations undertaken by predecessors.

GT: Never "new officials ignore old accounts".

Terms with Chinese background are the linguistic crystallization of Chinese knowledge with local characteristics, and have distinct particularity. In the process of cross-cultural communication, this particularity is further superimposed into complexity due to the heterogeneity of another language and culture. It is precisely because of the special complexity in the three dimensions of concept transplantation, sign conversion and communicative adaptation, so that the practice of cross-cultural knowledge reproduction always faces problems

and challenges (Wei Xiangqing & Yang Ping, 2019). These terms cover various fields in China's history, culture, economy and politics, etc., and have profound cultural connotation. In form, they have no fixed structure, and are more colloquial and vivid. Because of their distinct Chinese cultural background, it is difficult to find the corresponding expressions in English.

The official versions of Examples (13), (14) and (15) all adopt domesticating method, which brings the original text into the target language culture, and makes the terms with Chinese background conform to the reading and understanding habits of the target language. While Google Translate obviously does not have the function of this cultural transformation.

In Example (13), the original meaning of "接地气" is to follow the principles of human beings and follow their nature, then it is applied in the political field, which means that government officials or celebrities have extensive contact with the lives of ordinary people, reflecting the wishes, demands and interests of the public. However, Google only recognizes the meaning of "地" in the original text, and only translates the literal meaning——"grounded".

In Example (14), "摘帽" means "to take off the hat", and it is used as a metaphor which refers to the counties that were once considered poor and whose names of "poverty" were removed and they no longer fall into the scope of poverty. The translation of this word in *China Daily* can also convey this meaning properly——removing the label of an impoverished county. While the transliteration applied by Google not only fails to convey the meaning of the original text, but also causes the confusion from English readers.

As for Example (15), although Google can accurately translate its literal meaning, it cannot convey the connotation. This term means that after the appointment of the new leaders at the same level, they do not admit the contracts, commitments and debts left by the previous leaders. Thus, the connotation of this term is much more complicated than its literal meaning.

3. Conclusion

Nowadays, cultural exchanges between countries are increasingly frequent, meanwhile we are trying our best to tell Chinese stories to the world. Therefore, it is very important to accurately convey the connotation of Chinese characteristic terms and achieve cross-cultural communication. For MT such as Google Translate, Bing Microsoft Translate, Baidu Translate, etc., although they are widely used, they can't complete this task perfectly.

Through the analysis, we find that in terms of the translation including political neologisms or professional words, MT has some unparalleled advantages, such as accuracy, efficiency and professionalism, which are also reflected in some well-known Chinese characteristic terms that can be understood by the target readers only by translating the literal meaning.

When MT meets the terms with background knowledge, it is a little difficult to convey the meaning of its connotation, but not all such terms cannot be translated, for some terms which are often used in international occasions, their meaning can be translated roughly.

However, as to abbreviations and four word idioms, the translation versions are not so satisfactory. Due to the limitations of the translation mechanism, MT can only search for the most matching translation results in the corpus relying on the rules of search and match, which leads to its inability to carry out deep analysis and interpretation of terms, and thus cannot express the main ideas completely.

We should admit that the development of Big Data and Artificial Intelligence is extremely rapid, and it has been put on the agenda to apply this technology to MT like Google Translate.

Although it is still in its infancy, its potential is huge. When this technology matures, MT may have the ability to analyze and transform the cultural connotation of Chinese and English. But we don't have a definite answer when that day will come. Therefore, in the recent future, the development of MT may move towards specialization, MT and HT may have differentiation trend, the translation tasks with simpler sentence and more professional vocabularies can be handled by MT; HT would deal with those tasks with more complex sentence and deeper cultural background words, but this definite division of work is not absolute, because some translation tasks still require human and machine to complete together. Only in this way can we give full play to the advantages of MT and enhance the quality of translation under the condition of saving cost and improving efficiency.

Reference

[1] Anderson, James M. (2006). Malmkjær, Kirsten (ed.). The Linguistics encyclopedia (EBook ed.). London: Routledge. p. 601.

[2] Henry, R., & Venkatraman, S. (2015). Big Date Analytics the Next Big Learning Opportunity. Journal of Management Information and Decision Sciences, 18(2), 17-29.

[3] Hutchins, W. J. (1986). Machine translation: past, present, future. Modern Language Journal, 71(4), 437.

[4] Martin Kay. (1982). Machine translation. Comput Linguist. 8(2), 74–78.

[5] O'Halloran S, et al. (2016). Data Science and Political Economy: Application to Financial Regulatory Structure. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2 (7):87-109.

[6] W. John Hutchins. (2003) Machine Translation. Encyclopedia of Computer Science. John Wiley and Sons Ltd., GBR, 1059–1066.

[7] 蔡红改. (2011). 中国特色词语的英译策略. 新闻爱好者, 000(008), 97-98.

[8] 陈丹. (2014). 文化视域下的汉语成语翻译探析. 语文建设(32), 72-73.

[9] 戴新宇, 尹存燕, 陈家骏, & 郑国梁. (2004). 机器翻译研究现状与展望. 计算机科学, 31(011), 176-179.

[10] 冯全功, & 刘明. (2018). 译后编辑能力三维模型构建. 外语界, No.186(03), 57-63.

[11] 何勇斌. (2016). 从文化传播视角探析汉语成语翻译. 语文建设(6X), 73-74.

[12] 胡开宝, & 李翼. (2016). 机器翻译特征及其与人工翻译关系的研究. 中国翻译, 037(005), 10-14.

[13] 林语堂著, 郝志东, & 沈益洪译. (1994). 中国人 (全译本). 学林出版社.

[14] 宋平锋, & 张春艳. (2016). 中国时事热词及其外宣翻译——以 2015 年中国年度时事热词为例. 辽宁医学院学报(社会科学版), 14(03), 130-132.

[15] 王雪, & 蔡丽华. (2013). 中国时事政治术语的语言特点及翻译方法. 辽宁科技大学学报, 36(03), 298-301.

[16] 王菲. (2010). 文化观下的中国时政汉语新词汇英译. 湖北经济学院学报(人文社会科学版), 7(08), 148-149.

[17] 魏向清, & 杨平. (2019). 中国特色话语对外传播与术语翻译标准化. 中国翻译, 040(001), 91-97.

[18] 谢小平, & 周利敏. (2019). 大数据时代的语言学习研究——兼论 Google 翻译学习案例. 教育学术月刊, (11), 95-103.

[19] 许治, & 张建超. (2020). 新中国成立以来政府对科技人才注意力研究——基于国务院政府工作报告 (1954—2019 年) 文本分析. 科学学与科学技术管理, (03), 1-18.

[20] 张文. (1993). 机器翻译研究最新进展——介绍"863 智能型英汉机译系统". 中国翻译, (05), 58-59.

[21] 曾剑平. (2003). 汉语缩略语及其翻译. 中国科技翻译, 16(002), 58-61.