Study on the Importance of Cultural Context Analysis in Machine Translation

Zonghua Wang

Abstract Context of culture affects the specific meaning of the language. So the analysis of cultural context is essential for Machine Translation (MT). If the cultural context analysis of the source language is omitted in MT, ambiguity or mistranslation will be produced. The paper analyzes briefly the present situation of MT first and then, based on the theory of cultural context, it discusses with examples the importance of cultural context analysis in MT. The paper holds that if the cultural context analysis is paid enough attention in MT, the literal meaning and cultural connotation of source language can be exactly and completely translated into target language so as to improve the quality of MT. Therefore, the large-scale context corpus is very necessary for MT.

Keywords Machine translation · Cultural context analysis · Context corpus

Machine Translation (MT) is not only the conversion between natural languages, but the communication between different cultures. Because language is a part of culture while culture is the soil for language to grow up gradually. Context of culture affects the specific meaning of the language. So the analysis of cultural context is essential for MT. If the cultural context analysis of source language is omitted in MT, ambiguity or mistranslation will be produced. The paper discusses with examples the importance of cultural context analysis in MT. The paper holds that if the cultural context analysis is paid enough attention in MT, the literal

Z. Wang (✉)
School of Foreign Languages, Anhui University of Science and Technology,
Huainan, People's Republic of China
e-mail: Wanggmp6@yahoo.com.cn

Z. Wang
School of History and Culture, Shandong University, Jinan, People's Republic of China

meaning and cultural connotation of source language can be exactly and completely translated into target language so as to improve the quality of MT. Therefore, the large-scale context corpus is completely necessary for MT.

1 Machine Translation: A Brief Introduction

Machine Translation (MT) is the translation by computer between natural languages including their texts and sounds, which is the time-honored section of natural language processing by computer and "the most important application field in Computational Linguistics" [1]. Besides, MT is a comprehensive discipline combined with linguistics, mathematics and computer science in the information age. The technique of language analysis is the core of MT [2]. At present, MT system can be classified into Rule-based Machine Translation system, Corpus-based Machine Translation system, Multi-engine Machine Translation system and Spoken Language Machine Translation system, among which Rule-based Machine Translation system is classified into Literal Translation system, Transition system and Interlingua system, and Corpus-based Machine Translation system into Example-based Machine Translation system and Statistical Machine Translation system [3].

In the early 1930s, French engineer G. B. Artsouni put forward the idea of translation by machine and got a patent for "Translation Machine" named "Mechanical Brain". In 1946, the first Electronic Numerical Integrator and Computer (ENIAC) in the world came out. Its fantastic speed at that time indicated the innovation of translation technique. In March, 1947, Warren Weaver, an American Scientist, and Andrew Booth, a British engineer, met together at Rockefeller Center in New York to discuss the computer application problem. In 1949, Warren Weaver issued a memorandum titled *Translation,* in which he presented formally the problem of computer translation. And then some scholars from USA, Soviet Union, Britain and so on took part in the research on machine translation in following several years. Georgetown University and IBM jointly carried out the experiment of machine translation first in the world in 1954 and succeeded in translating several simple Russian sentences into English, which seemingly indicated that the high-qualified MT software would be made very soon and there was no language barrier in the future.

China is one of the countries earlier carrying out the study of MT, beginning with its research work in 1956. The study of MT approximately goes through four periods: the beginning period (from the middle 1950s to the middle 1960s), the stagnant period (from the middle 1960s to the middle 1970s), the developing period (from the middle 1970s to the late 1980s) and the flourishing period (from the early 1990s to the present), which has made some achievements in theory and practice. However, there is a long way for Chinese experts to go to catch up with and even surpass the international level. One reason is that Chinese is a more complicated language than English, one of Indo-European languages, making it

more difficult in MT studies. At present, many experts have realized that the study of Chinese itself is basic for MT from Chinese to other languages and vise versa.

It is almost 60 years since the first MT experiment was successfully made in 1954. In the past 60 years, MT made great progress in many countries. With the development of information technology and the popularization of internet and computer, MT application has been spread increasingly, such as Canadian TAUM-METEO (which can automatically receive, count and translate the data of weather forecast, and finally issue it in English and French.), automatic translation telephone made by UK, USA, Germany, etc., electronic dictionary, net-automatic translation system, GPS for car drivers (based on speech recognition, transformation and synthetic system) and so on. Now anyone can use the MT tools on the internet. MT plays an important role instead of human in the translation of sci-tech, political documents and so on, and increases efficiency greatly.

However, people are not satisfactory with the quality of MT at the present time. Reasons for this are the complexity of natural languages and their translation and the limitations of computer, e.g., smaller storage capacity and less complicated than the human brain [4]. Nida, a famous American translation theorist, points out that in addition to problems of size and sophistication, there is a need of more appropriate intermediate technical language which can be automatically adapted to the special limitations of computer and has more comprehensive function than the present machine languages, e.g. Comit, Fortran and Mimic [5]. Among the reasons, the failure of cultural context analysis in MT is a very important factor for producing low-qualified versions.

2 The Theory of Context of Culture

Context of culture is one of the most important linguistic categories in the study of language use and function, which is first proposed by British anthropologist Bronislaw Malinowski. It refers to the specific social regulations and idioms of certain speech community. The recent research of cultural context shows that each speech community has its own history, culture, customs, social regulations, mode of thinking, moral ideas and value orientation. These specific ways and factors of certain speech community constitute the full content of cultural context [6].

Language is a very important part and the carrier of a culture, which is indispensable for both the functioning and the perpetuation of the culture; culture is the surroundings for language to grow up and determine its specific linguistic meaning. "Language also constitutes the most distinctive feature of a culture, which may be described in a simplistic manner as the totality of the beliefs and practices of a society" [7]. They are closely connected and indivisible. Juri Lotman said, "No language can exist unless it is steeped in the context of culture" [8]. Words only have meaning in terms of the corresponding culture. Susan Bassnett also said, "The translator treats the text in isolation from the culture at his peril" [8]. For example, he was in the seventh heaven last night. If the sentence is literally translated into Chinese as "他昨天晚上在第七天堂", no Chinese reader

can understand the version because the phrase "the seventh heaven" does not mean itself literally but refers to the God's Palace, meaning here "the state of great happiness", and it means nothing practically in Chinese. Another Example, at the banquet, Mr. Li jokingly said to a foreign guest, pointing at Mr. Zhang next to him, "他有气管炎". All Chinese guests burst into laughing. The young interpreter translated it literally as follows: "He suffers from tracheitis". The foreign guest was in a fog and did not laugh, and asked confusedly, "Why are you laughing so hard when he is suffering from tracheitis?" Obviously, the interpreter makes a pragmatic mistake without considering Chinese cultural context. "气管炎" and "tracheitis" mean the same. But "妻管严" and "气管炎" are homophones of each other in Chinese. "妻管严" with Chinese cultural connotation means hen-pecked.

It is thus clear that "translation is the communication between two cultures. For the successful translation, to know two cultures well is more important than to know their languages" [9]. The meaning of a text is finally determined by its cultural context, not by its external language form. Therefore, while translating, the translator should take the cultural context of source language into consideration in order to understand fully the source text and translate it exactly into the target language; at the same time, readers' cultural context should be considered before final version is produced so that they can understand the target text fully [10]. It is also true for machine translation.

3 Cultural Context Analysis and Machine Translation

It is easily seen from the analysis mentioned above that context of culture plays a very important role in translation. Similarly, MT can not be isolated from the cultural context analysis. Otherwise, the source language text can not be understood exactly and fully and mistranslation will be produced in MT. That is to say, the cultural context analysis is the extremely important condition in MT for computer to make a choice of the given versions. It affects MT's quality and even determines whether the version is right or wrong. The following are four examples (two in English and two in Chinese) to show us the importance of study on cultural context analysis and its application in MT. Versions come from English to Chinese and Chinese to English MT soft systems such as Google online translation system, Baidu online translation system and Kingsoft Fast system which are very popular and easily available in China.

Example 1: He is by no means courageous. He is actually a hare.

Google Version: 他绝不是勇敢的。他其实是一只野兔。
Baidu Version: 他绝不是勇敢的。他实际上是一个兔子。
Kingsoft Version: 他绝不是勇敢的。他实际上是一个兔子。

The word "hare" in English refers figuratively to a person who is very timid. However, the word "mouse" is used in the same way instead of "hare" in Chinese. Three Chinese visions from MT systems are similar. They are all literal translations with English cultural connotation missed and do not express exactly the

meaning of the source text. If the word "hare" is translated into "胆小如兔" instead of "一只兔子", Chinese readers can not understand the version either because Chinese cultural connotation is missed here. So reference translation is "他绝不是勇敢的。他其实胆小如鼠"。

Example 2: Our history teacher is a good talker. I'd like to ask him, "where is the beef?" because we sure don't learn much from him.

Google Version: 我们的历史老师是一个很健谈。我问他，"是牛肉在哪里?"因为我们肯定不会学到很多东西从他身上。
Baidu Version: 我们的历史老师是个很健谈的人。我想问他，"牛肉在哪里?"因为我们当然不从他的身上学到很多。
Kingsoft Version: 我们的历史老师是个很健谈的人。我想问他，"牛肉在哪里?"因为我们当然不从他的身上学到很多。

The short sentence "where is the beef?" in source text is an American idiom, meaning that people complain there is much too little beef in hamburger. Now it refers to something that is short of substantiality. Computer does not know of the cultural background of the idiom. As a result, versions in the Example 2 apparently shows us the literal meaning of the idiom while neglecting its cultural connotation and that there is no the cultural context analysis of source text in three MT systems. Therefore, Chinese versions are very hard for readers to understand. The following version is better with the cultural context analysis of both source text and target text:

"我们的历史老师很能夸夸其谈。我真想问问他，"管用的货色在哪里?因为我们真的没有从他那里学到多少东西" [11]。

Example 3: 三个臭皮匠抵个诸葛亮。

Google Version: Three Stooges arrived Zhuge Liang.
Baidu Version: The Three Stooges against Zhu Geliang.
Kingsoft Version: The Three Stooges against Zhu Geliang.

In Chinese culture, "诸葛亮" (Zhuge Liang) is a person of great wisdom and resourcefulness; "皮匠" (cobbler or shoemaker) refer to less educated people who make a living on repairing shoes. Sometimes,people call sneeringly such less educated people as "臭皮匠". The cultural connotation of source text is "more people, more wisdom". All three MT versions are unacceptable because they all contain no cultural connotation of source text. Second and third MT systems even wrongly translate "诸葛亮", whose family name is "Zhuge", not "Zhu". By analyzing the cultural context of source text, the acceptable version is "Three cobblers with their wits combined equal Zhuge Liang, the mastermind". At the same time, there is the same expression in English as "Two heads are better than one". Considering the cultural context of target language, this is better version for target text readers.

Example 4: 他们俩真是一龙一猪!

Google Version: The two of them is really a dragon, a pig.

Baidu Version: Both of them are Yilongyizhu!
Kingsoft Version: Both of them are Yilongyizhu!

Both "龙" (dragon) and "猪" (pig) in English are derogatory terms while the word "龙" (dragon) in Chinese is commendatory term. The set phrase "一龙一猪" in Chinese figuratively refers to two persons, one of whom is bright, capable and promising, and the other is extremely foolish and hopeless. "一龙一猪" is literally translated in Google version and it is just expressed again in another Chinese form —Pinyin in both Baidu and Kingsoft versions, not translated. With the cultural context lost, three MT versions do not express the cultural connation of source text and make it very difficult for target text readers to know what is really said. The similar or same English expression as "一龙一猪" is not available as well. Maybe, the following version is acceptable: "They are sharp contrastive individuals: one is very capable, while the other is extremely incompetent".

4 Conclusion

In conclusion, cultural context analysis plays an extremely important role in MT. The main reason for the low quality of MT in the examples mentioned above is the failure of cultural context analysis of source language. A computer does not know of the cultural differences between languages and naturally it can not understand the context of source text. Therefore, the large-scale bilingual or multilingual context corpus is very necessary for MT to improve its version quality. Just as Newmark said, "context is an overriding factor in all translation, and has primacy over any rule, theory or primary meaning" [12]. So context corpus should contain an extremely huge number of versions related to the context of utterance and culture as basic language sources for computer. With such context corpus, computer can "analyze" like a human translator such contexts of source text, compare the different versions input and choose the closest correspondence with the source text. However, to set up such context corpus will be an enormous system engineering. With the further development of informationalization and popularization of network in the 21st century, people need more information by computer from different nations, which will certainly promote the research on MT and the really full-automatic and high qualified MT soft system will surely be generated in the future so as to overcome the barrier of languages in the cross-cultural communication for information.

References

1. Zhiwei F (2011) Computational linguistics: its past and present. J Foreign Lang 1:11
2. Shanghui K, Sumin Q (2012) An investigation on machine translation. Intell Comput Appl 2 (42)

3. Zhiwei F (2004) Machine translation studies. China Translation and Publishing Corporation, Beijing
4. Zheng Z (2005) Difficulties for machine translation. Foreign Lang Res 5:59–62
5. Nida EA (2004) Toward a science of translation. Shanghai Foreign Language Education Press, Shanghai
6. Lan H (2008) Cultural context: the hedge of version choice. J Guangxi Soc Sci 6:175
7. Nida EA (2001) Language and culture: contexts in translating. Shanghai Foreign Language Education Press, Shanghai
8. Bassnett S (2004) Translation studies. Shanghai Foreign Language Education Press, Shanghai
9. Liu W (2012) On cultural context factors affecting translation. J Liaoning Univ Technol (Soc Sci Ed) 1(52)
10. Wang J (2006) Cultural context and translation. Shanghai J Translator 2:52–54
11. Xueyun E, Chen X (2007) Context analysis: the indispensable factor in machine translation. Foreign Lang Teach 4(53)
12. Newmark P (2001) Approaches to translation. Shanghai Foreign Language Education Press, Shanghai