

On the Development of Translation Technology

Ying Wu

Zhenjiang Watercraft College, Zhenjiang, Jiangsu, 212003, China

Qi Pan

Zhenjiang Watercraft College, Zhenjiang, Jiangsu, 212003, China

Abstract—With the rapid development of translation technology, the machine translation system and translation memory system have gradually become popular tools among professional translators. This article made a brief introduction of the machine translation system and the translation memory system, explaining some key terms, sorting the recent development and recent studies in this area, so as to provide some ideas and suggestions for further study.

Index Terms—translation technology, machine translation, translation memory system

I. INTRODUCTION OF TRANSLATION TECHNOLOGY

The development of the translation technology has a close relationship with the significantly increasing demand for translation service in recent years. Nowadays, globalization and localization have become a trend throughout the world, leading to a rapid growth of international demand for translation. According to a survey conducted by the Canadian Translation Industry Sectoral Committee in 1999, the European translation industry was expected to grow by 7% annually over the next five years, and the Canadian industry by 5-10% over the next three years. And according to the estimation of LISA (The Localization Industry Standards Association), in 2001, the global annual expense in localization was expected to be five billion US dollars, and the real expense could be as high as 15 billion US dollars. In 2006, this number could be up to 30 billion US dollars.

In spite of this growing demand of translation service, the supply of qualified translators is on a relatively low increase. In Canada, universities produce only about 400 new translators each year, less than a third of the requirement of the market. And in EU, the lack of translators has become a big problem with the rapid addition of new members and new languages in recent years. In such a condition, the problem of supply and demand becomes increasingly serious in the translation market.

What's more, in such a world of information explosion, the time pressure on translators is just as much as the pressure of the volume and the quality. For example, in the translation of web pages, the translated versions are sometimes required to be updated just as quickly as the original ones. This makes the contradiction in the translation market even worse.

All these conditions and problems bring great challenges to the traditional translation practice mode, and people have to turn to the evolution of translation technology for the solution. During the development of the translation technology, the most familiar concept to the public is perhaps the Machine Translation.

A. Machine Translation System

Machine Translation (MT) is defined by Hutchins and Somers as “computerized systems responsible for the production of translation from one natural language to another, with or without human assistance” (Hutchins & Somers, 1992). The development of MT is marked with a lot of ups and downs. The first idea of MT can be dated back to 1903 when Couturat and Leau put forward the term “ein mechanisches Uebersetzen”, which is German name of “machine translation”. Before 1960, the development of the MT was quite optimistic and researchers believed that the MT could produce “fully automatic high-quality translation (FAHQT)”. However, in 1966, a report of the ALPAC (Automatic Language Processing Advisory Committee) declared that the concept of FAHQT was unrealistic. Although the report was widely considered short-sighted later, it had a great negative influence on the MT researches at that time. After that, some researchers persist in the development of the MT system with a more realistic goal, while others turn to looking for some alternative ways to solve the problem caused by the contradiction between the demand and supply of the translation service.

B. Translation Memory System

One of the most attractive alternatives widely used in recent years is the Translation Memory System (TMS). The TMS is a computer-aided translation tool “designed to increase translation productivity by automating the linguistic transfer from source to target text” (Austermuhl, 2001). It is able to assist human translators by reusing their past translations. There are many differences between the MT and the TMS.

According to Liang Sanyun (2004), differences between the MT and the TMS can be explained in at least three parts. The first and most important one is that they are designed for different purposes. The MT is developed to replace the human translation with the operation of machine, aiming at providing language assistance to the users who are not good at a certain kind of foreign language, while the TMS is a tool devised for professional translators, aiming at improving their working efficiency and the quality of the translated texts. The second difference lies in their scope of applications. It has been generally acknowledged that the MT should only be used for texts from some restricted domains, especially for non-literary texts. In contrast, the application of the TMS does not have such a limitation since it only provides assistance to translators, and translators still play a leading role in the translation procedure. The last but not least, they have different translation engines. As a fully automatic translation device, the MT has a high requirement for both the grammar bank and the knowledge bank of the translation engine. Compared with the TM, however, the TMS is only semi-automatic as it only focuses on the automatic matching between the new segment and the stored segment. Therefore, the storage of large quantity of translated materials is more important than the grammar capability to the translation engine of the TMS.

From the above comparison, it can be seen clearly that the most significant difference between the MT and the TMS is that the former is machine-centered while the latter is human-centered. The strength of the TMS lies in that it provides better cooperation with the human translators and integrates the strengths of computers with that of human beings. By recycling the existing translations, it is able to reduce the cost and time of translation as well as improve the quality and consistency of the translated text. As a result, the TMS gets a better practical usage than the MT and is currently among the most popular CAT tools in use among professional translators.

However, the popularity of the TMS does not necessarily mean that no problem exists in the application of the TMS. Although "the TMS does not conduct real translation" and is just a bank of pre-translated materials, it is possible that the translator and the translated text are negatively influenced by the operation of the TMS. One possible problem is that the translator relies too much on the TMS and believes whatever is suggested by the TMS, even if the translation suggestion is wrong itself or inaccurate in some aspects. As a result, the wrong or inaccurate translation may repeat and even recycle in the translated text. At the same time, the creativity of the translator may also be impaired.

Researches on the limitation of corpus-based translation studies may provide some theoretical explanations to the possibility of this problem.

Liu Kanglong (2006) cites the opinion of Mona Baker in his article, saying that researchers tend to be attracted by the large amount of translation information and data from a large-scale corpus by simply clicking the mouse. Therefore, they may focus more on the Translation Norm but neglect the creativity of translators, and even regard a creative expression as a mistake. Liu Jingguo (2006) also mentions the weakness of corpus-based studies. He points out that corpus-based studies have a clear tendency of scientism, focusing too much on the objectivity and tending to constrain the creativity of persons. Therefore, it may lead translation studies into a lifeless condition and even to a stop.

The above ideas of Liu Kanglong and Liu Jingguo are helpful in understanding the potential problem between the translator and the TMS. From the above introduction, it can be seen clearly that the TMS is in its nature an aligned corpus. Therefore, in using the TMS, the translator may also have a tendency to be attracted by the information stored in it and neglect his/her own creativity and even judgment. That is to say, they may tend to trust and rely on all the translation suggestions provided by the TMS without or with little further thinking. If some mistakes exist in the TM database, it is also possible that they may fail to find them out and just adopt them in the translated text as usual. If such a problem does exist, there may be a danger that these mistakes get repeated and recycled in the TM and the translated text, and hence causing damages on a larger scale.

In view of the above assumptions, the author of this thesis attempts to make use of a self-designed experiment and a questionnaire to prove their existence, aiming at finding out whether the mistakes in the TMS will lead to the repetition of the mistakes in the translated text, and what the attitudes of the translators are during this process. The result of this research may have some significance either to the translation education, the translation service, or to the development of the TMS. As for the translation education, the result may help improve the self-awareness of student translators in the translation procedure with the TMS, so as to help them avoid complete trust in the information stored in the TM and to maintain independence in the analysis and judgment. As for the translation service, the result of this research may help stress the importance of the management of the TM database, either to the translation companies or to the client of the translation service. And for the development of the TMS, the result of this research may provide some new perspective of improvement especially in the aspect of the TM management.

II. DEVELOPMENT OF TRANSLATION TECHNOLOGY

A. *Key Terms*

Just as introduced in above section, the TMS is a type of CAT tool that is able to store the previously translated text and the corresponding source text in a database, and allows the translators to reuse them or part of them in a new translation task. A typical TMS consists of two parts: a Translation Memory (TM) and a search engine. The TM is in its nature a database that store translated texts together with the corresponding original texts. The user can make use of the previously translated text and the source text to set up one or more TM database(s). Then, in a new translation task, the system can use the search engine to automatically search in the existing TM database for the same or similar translation

material (such as a sentence, a paragraph, etc.) and provide a translation suggestion to the translator. As a result, the translator does not have to do any repetitive work and can focus his/her energy and time on the translation of the new materials. At the same time, the TM database is going on storing newly translated materials and expanding the volume of the database. As for the translation suggestion, the translator can choose to follow it, edit it or discard it.

Some basic concepts in the technology of the TMS are briefly explained as follows.

Segmentation

When texts are added into a TM database, they have to be first divided into some small chunks that are called segments. A segment is the smallest unit that can be processed by the TMS, and is also the base of the alignment. Generally speaking, a segment is a sentence or a sentence-like structure and is usually ended with a punctuation mark (such as a period or a question mark). However, according to Bowker (2002a), it is also possible for the user to specify what should or should not represent the end of a segment.

Alignment

The alignment is the second step in storing and organizing the texts. Aligning texts means matching the source text segment with the corresponding target text segment, and storing them in the database as an integrated unit for future processing. The correct alignment is the key to the successful reuse of the stored information.

Retrieval

Once the aligned texts have been stored in the TM database, they can be reused by the translator in a new translation task via the automatic search and retrieval of the system. When a new text is opened in the translation editor of a TMS, the system first divides it into segments as was introduced before. Then the system takes the first segment and compares it with those segments that have been stored in the TM database. The comparison is carried out on the basis of some preset rules. The establishment of these rules is also an important topic of researches and studies on the TMS. Once a match is found out in the comparison, the system retrieves the relevant information (a unit of the aligned texts) and presents it as a translation suggestion to the translator.

There are a number of types of matches that a TMS can retrieve, so as to find not only *identical* translation units but also *similar* ones.

Perfect or Exact Match

A perfect or exact match means that the new segment in the source text is exactly the same as a segment that is already stored in the TM database (100 percent match). According to Austerlitz (2001), there is another type of match that is similar to the exact match called a *full match*, which means the new segment is identical to a stored segment in all the other aspects except for some variable elements such as dates, numbers, time or measurements. In most TM tools such as the Trados, these elements can be automatically treated by the system by replacing the old dates, numbers, time or measurements with new ones.

Fuzzy Match

Fuzzy match is an important and powerful function of the TMS, meaning that the new segment is not identical to but similar to a stored segment. It is more often encountered by translators in the real practice than the exact match. Translators can edit the fuzzy match so as to produce a valid translation for the new source segment. In most TM tools, different parts between the new segment and the stored segment are highlighted with a different color to indicate which elements in current translation need to be changed or adjusted to meet the source text.

Another important thing in using fuzzy match is the degree of fuzziness, or on the other hand, the degree of similarity, which respectively means to what degree a stored segment found by the searching engine is different from or similar to the new source segment that needs to be translated. In most TM tools, there is a user-defined minimum match value which can be adjusted by the translator before translation according to the desired degree of fuzziness or similarity. Generally speaking, the higher the match value is, the stricter the retrieval will be, and vice versa. However, according to Bowker (2002b), the translator must be very cautious in selecting the degree of fuzziness so as to avoid the problem of silence and noise. That is to say, if the match value is set too high (e.g., 95%), the system may not retrieve matches that could be potentially helpful to a translator. In information retrieval, this type of problem is known as “silence.” In contrast, if the match value is set too low (e.g., 10%), then too many segments, or segments that are not useful, could appear leading to a problem known as “noise,” which may actually hinder the efficiency of the translation process.

B. Recent Development

With the developing popularity of the TMS among translators, it also becomes the subject of a growing number of researches in recent years. However, compared with researches and discussions on the technological problems in the TMS, such as the segmentation, the alignment and retrieval of information, not many studies have been carried out on the application of the TMS. Generally speaking, present studies on the application of the TMS are mainly around three topics, namely, the speed of the translation, the quality of the translated text and the attitude of the translators. Since the problem of speed is not the major concern of this article, this section only focuses on the introduction of the previous researches related to the quality of the translated text and the attitude of the translators.

a. Quality of the Translated text

It is widely accepted that the application of the TMS can greatly improve the quality of the translated text, especially the consistency of terminologies and phrases in a large project. According to O'Brien (1998), if the translation memory

is on a network, several translators working on a single project can achieve greater consistency through instant access to each other's work.

Bowker carries out some experimental studies on the quality of translation with the TMS. She (Bowker, 2003) examines the advantages and disadvantages of using a single TM database in both language directions, showing that when searching from L1 to L2, translators are likely to retrieve a greater number of matches, but these are of a lower quality, while translators searching from L2 to L1 are likely to find fewer matches, but the ones they do find are of a higher quality. She also conducts another study focusing on both the productivity and the quality (Bowker, 2005). The result shows that it is often difficult for translators to achieve a balance between the productivity and the quality because one often comes at the expense of the other when using a TMS.

Moreover, some researchers focus especially on the negative impact of the TMS on the quality of the translated text. Mogensen (2000) points out that the application of the TMS may make the resulting text less readable because of the relative absence of cohesive devices between sentences. She also explains that using the TMS may lead to an oversimplified translation because the sentences are often short. What's more, she indicated that instead of translation tools being modified to accommodate language, language is being modified to accommodate translation tools. This may result in a target text whose syntax closely models that of the source text and is therefore less readable in the target language.

Another important threat in the application of the TMS is the possibility of mistranslation. According to Austermuhl (2001), the recycling of past translations may not only help increase the consistency and improve the quality, but also lead to the repetition and reproduction of mistakes. Therefore, the maintenance of the TM database is of great importance. If this is neglected, the errors may be propagated quickly throughout a document.

However, Austermuhl does not carry out an empirical study to prove this assumption. Therefore, this thesis attempts to test this assumption through a self-designed experiment and to find out more about the problem.

b. Attitudes of the Translators

The development of the TMS technology not only brings improvement to the translation itself, but also brings some changes in the way translators see their work. Although the changes of translators' attitudes do not attract much attention as compared with other aspects in the application of the TMS technology, several scholars still make some observations in this area.

O'Brien (1998) declares that in comparison with the idea of the MT, translators are more satisfied with the TMS because they feel they are still in control of the whole process. Merkel (1998) also carries out a study to examine the attitude of a group of technical translators toward terminological consistency and variation when working with the TMS. The result shows that translators can become frustrated by the way that consistency is sometimes inappropriately imposed on the texts by these systems. That is to say translators may not be very satisfied with the way that the previously translated texts are reused, since the system may sometimes make improper application.

However, no present researches on the attitudes of translators have mentioned the problem of the trust and reliance of translators on the TMS, even though such a phenomenon is possible to exist as discussed above. Nevertheless, researchers focusing on the assistance of the Internet to the translation practice have investigated a similar problem.

Ni Chuanbin (2003) points out that translators must be very careful in using the Internet search engine to assist their translation. He says that some information provided by the Internet search engine is far from the standard or authentic English due to the variety of sources of the information. Therefore, translators must be very careful, and should think about and analyze the authenticity and correctness of the information before adopting it in their translation.

Ni's opinion is enlightening for understanding the attitudes of translators on the TM database. It is true that a TM database is different from a search engine in many ways, but they are similar in their working mechanism of information storage and extraction. As for the trustworthiness, the TM database may perform better than the Internet search engine, but is still far from being 100% correct. In this case, translators may also need to be very careful in using the stored translation in the TM database instead of completely trusting and relying on the TMS.

On the basis of these observations, it is necessary for us to make a further investigation on the attitudes of translators to the application of the TMS, especially to the translation suggestions provided by the TMS, so as to find out whether the trust and reliance of translators to the TMS may influence the repetition of mistranslations in the translated text.

III. CONCLUSION

From the above discussion, it can be seen clearly that it is quite important to carry out an empirical study on the influence of the quality of the TM to the quality of the translated text, especially on the existence of the repetition of mistranslations and other types of mistakes. Meanwhile, it is also helpful to make a survey on translators' attitudes towards the application of the TMS, trying to find out how translators' attitudes may affect the quality of the translated text.

REFERENCES

- [1] Austermuhl, F. (2001). *Electronic Tools for Translators*. Beijing: Foreign Language Teaching and Research Press.
- [2] Bowker, L. (2002a). *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.

- [3] Bowker, L. (2002b). Information Retrieval in Translation Memory Systems: Assessment of Current Limitation and Possibilities for Future Development. *Knowledge Organization Journal*, 29, 200-201.
- [4] Bowker, L. (2003). Investigating 'reversible' translation resources: Are they equally useful in both directions? In L. Pérez-González (Ed.), *Speaking in Tongues: Language across contexts and users*. Valencia: Universidad de Valencia.
- [5] Bowker, L. (2005). Productivity vs Quality? A Pilot Study on the Impact of Translation Memory Systems. *Localisation Focus*, 4, 13-20.
- [6] Hutchins, J., & Somers, H. (1992). An Introduction to Machine Translation. London: Academic Press Ltd.
- [7] Liang, S. Y. (2004). A Comparative Analysis of Machine Translation and Computer-assisted Translation. *Computer-assisted Foreign Language Education*, 100, 42-43.
- [8] Liu, J. G. (2006). History and Progress of Corpus-based Translation Studies. *Journal of Foreign Language*, 2, 66-71.
- [9] Liu, K. L. (2006). Corpus Linguistics and Translation Studies. *Chinese Translators Journal*, 1, 61-62.
- [10] Merkel, M. (1998). Consistency and Variation in Technical Translation: A Study of Translators' Attitudes. In Bowker, L. et al. (Ed.), *Unity in Diversity? Current Trends in Translation Studies*. Manchester: St. Jerome Publishing, 137-149.
- [11] Mogensen, E. Orwellian Linguistics: How using computer-aided translation tools impacts the target language. *Language International*, 12, 28-31.
- [12] Ni, C. B. (2003). On the Scientificity and Feasibility of Translation Aided by Internet Search Engine. *Shanghai Journal of Translators for Science and Technology*, 4, 53-55.
- [13] O'Brien, S. (1998). Practical Experience of Computer-aided Translation Tools in the Software Localization Industry. In Bowker, L. et al. (Ed.), *Unity in Diversity? Current Trends in Translation Studies*, Manchester: St. Jerome Publishing, 115-122.

Ying Wu was born in Zhenjiang, China in 1979. She received his M.A. degree in Translation Study from Nanjing University, China in 2009.

She is currently a lecturer in the English Department of Zhenjiang Watercraft College of PLA, Zhenjiang, China. Her research interests include Translation and Foreign Language Teaching.

Qi Pan was born in Zhenjiang, China in 1987. She received her B.A. degree in English language and literature from Jiangsu University of Science and Technology, China in 2009.

She is currently a tutor in English Department, Zhenjiang Watercraft College of PLA, Zhenjiang, China. Her research interests include Linguistics and Foreign Language Teaching.