A Comparative Study of Machine Translation and Human Translation—A Case Study of Kingsoft Powerword

Fang-fei WU, Qiu-ping LIANG and Gao-da HE*

College of Foreign Studies, South China Agricultural University, Guangzhou, China

*Corresponding author

Keywords: Machine translation, Human translation, Human involvement.

Abstract. This paper intends to figure out the limitations of machine translation by conducting a comparative study of machine translation and human translation at lexical level and syntactic level. The outcome demonstrates that the most severe defect of machine translation is that it lacks human’s cognitive competence. With this defect, machine translation repeatedly makes errors in the target text, reducing the output quality. In order to address this issue, this paper states that human involvement should be supplemented to machine translation by pre-editing the source text and post-editing the target text. Finally, this paper expects that, with the improvement and advance in machine translation technology, machine translation could obtain a leap in its output quality, promoting the development of translation.

Introduction

Machine translation (MT), is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one language to another [1]. According to its principle, MT mainly consists of rule-based MT, statistics-based MT, example-based MT and multi-engine MT [2]. Human translation (HT) refers to the act of transmitting source text into target text done by human translators who, following certain principles of translation, apply various translation strategies and use various methods to accomplish the task of translation [3].

With the rapid development of globalization, the increasing translation demand is far from satisfactory by HT. Thus, characterized by high efficiency and low cost, MT is gaining ever-increasing popularity in translation market. However, MT’s output quality is not sufficient enough to be accepted. In view of this, this paper is going to conduct a comparative study of MT and HT to figure out limitations of MT and puts forward solutions to this problem.

In translation market, more and more MT tools have been emerging, such as electronic dictionaries, on-line translation tools and modern translation software. Among various MT tools, Kingsoft Powerword, as one of the most frequently-used translation software in China, will be employed in this text.

The translation exercise of A Survey of the Olympic Games from English Translation Practice II [4] will be selected as the source text. Its machine translation text produced by Kingsoft Powerword will be compared with the HT version, which is the reference version produced by editorial board member of this book.

A Comparative Study of Machine Translation and Human Translation

This section will make a comparative research at the lexical level and syntactic level with examples. In the following examples, A is the target text produced by MT while B by HT.

Lexical Level

The errors at lexical level will be analyzed mainly from the angle of lexica ambiguity and named-entity.
**Lexical Ambiguity.** Lexical ambiguity refers to the presence of two or more possible meanings within a single word [5]. To determine the right register of words with multiple meanings, context matters a lot. However, being incapable of analyzing the context of source text, MT system usually selects the most commonly-used meaning of a word to translate in its memory, e.g.

The number of *events* increased to 20.
- A *Shi Jian* De Shu Mu Zeng Jia Dao 20.
- B Ao Yun Hui *Bi Sai Xiang Mu* Zeng Jia Dao 20.

In the examples above, if the words “event” appears on its own, the target texts “Shi Jian” is correct. But in a text regarding Olympics, the translations produced by MT is not in line with the context, failing to reveal the intended meaning of the source text.

**Named-Entity.** Named-entity refers to entities identified with names, including people’s name, organization name, places name etc. In a broad sense, it also includes expressions of time, number. Named entity recognition (NER), also known as entity identification and entity extra, is to searching and judging whether a character string stands for a named entity, and also to determine its category [6]. NER is a significant step in the process of translation. Only by recognizing these named-entities can the MT system accurately translate them into target text. So the authors illustrate different outputs by MT and HT in the translation of people’s name and organization name:

<table>
<thead>
<tr>
<th></th>
<th>MT</th>
<th>HT</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Milo Croton</em></td>
<td>Mi Luo Ba Dou</td>
<td>Mi Le Ke Ru o Dun</td>
</tr>
<tr>
<td><em>Eastern Bloc</em></td>
<td>Dong Bu Lian Meng</td>
<td>Dong Ou Ji Tuan</td>
</tr>
</tbody>
</table>

Apparently MT system fails in the procedure of NER. In the translation of people’s name, MT system doesn’t take “Milo Croton” as a whole, but divides it into two parts to translate. Thus the output doesn’t appear as a person’s name. The organization “Eastern Bloc” refers to Warsaw Pact nations which is widely known as “Dong Ou Ji Tuan” in Chinese. MT output “Dong Bu Lian Meng”, on the other hand, which refers to basketball teams in National Basketball Association.

The expression of time is also difficult for MT system to recognize. In the source text, the most frequently-shown expression of time is a particular year, e.g.

- Until 1992, they were both held in the same year.
  - A Zhi Dao 1992, Ta Men Dou Zai Tong Yi Nian Ju Xing.
  - B Zai 1992 Nian Zhi Qian, Xia Ji He Dong Ji Ao Yun Hui Shi Tong Nian Zhao Kai De.

In the example above, the number “1992” represents “the year of 1992”. However, MT system only regards it as a figure, without adding a corresponding quantifier after the number.

**Syntactic Level**

The errors at syntactic level will be analyzed mainly from the angle of word order, word segmentation, noun phrase, and verb phrase.

**Word Order.** Word order refers to the linear ordering between words in a sentence [7]. MT usually chooses to follow the words order of source text while human translation, more flexible, adopts to different strategies to change the word order in target text, e.g.

A competitor must be the citizen of the country he or she represents.
- A Jing Zheng Zhe Bi Xu Shi Gai Guo De Gong Min, Ta Huo Ta Dai Biao.
- B Bi Sai Zhe Bi Xu Shi Ta (He) Huo Ta (She) Suo Dai Biao De Guo Jia De Gong Min.
The example above is a complex sentence with a postpositive attributive clause modifying the word “country”. But target text A, whose word order is the same as that of source text, fails to display the relationship between “country” and “he or she represents”, because the attribute is usually arranged before the object that need to be modified in Chinese.

**Word Segmentation.** Word segmentation refers to dividing a sentence into different parts, or translation units. In the process of translation, MT system, firstly, segments the source text into different translation units; secondly, indexes the text by translation unit; and finally, translates those units separately, e.g.

The Olympics *were of* fundamental religious importance.
A Ao Yun Hui *Shi* Zhong Yao De Zong Jiao Yi Yi.
B Ao Yun Hui *You Zhe* Fei Chang Zhong Yao De Zong Jiao Yi Yi.

In the example above, “were of” ought to be regarded as a translation unit, meaning “have”, while “were of” in version A should be segmented into “were // of” and translated into “Shi”, leading to semantic ambiguity.

**Noun Phrase.** A noun phrase is word group with a noun or pronoun as its head [8]. The head of a noun phrase can be accompanied by various components, such as complements, modifiers and determiners.

In 1980 and 1984, the *Cold War opponents* boycotted each other’s games.

“The Cold War opponents” is a polarization phrase, consisting of the modifier, “Cold War”, and the head, “opponents”. In target text A, however, the polarization phrase is translated into a subject-predicate structure due to MT system’s incapability of recognizing the noun phrase’s composition.

**Verb Phrase.** A verb phrase is a syntactic unit composed of at least one verb and its dependents—objects, complements and other modifiers—but not always including the subject [9].

The boycott *reduced the number of nations participating to only 81.*
A Di Zhi Huo Dong *Jian Shao Le 81 Ge Guo Jia Can Jia De Ren Shu.*
B Zhe Ci Di Zhi *Shi Can Jia Ao Yun Hui De Guo Jia Jian Shao Dao 81 Ge.*

“The verb phrase “reduce the number of nations participating to only 81” consists of a verb, a noun phrase and a preposition phrase. The noun phrase is the object of the verb, and the preposition phrase serves as the object complement. MT is unable to conduct an analysis of source text’s grammatical structure, resulting in both semantic and grammatical errors in target text.

**Limitation of Machine Translation**

The comparative analysis shows that HT outperforms MT in many respects: the ability to recognize context, flexible choice of translation unit, flexible arrangement of word order and so on. The most obvious limitation of MT lies in the lack of human’s cognitive competence, because most of the modern MT systems are corpus-based. Corpus-based approaches mainly include statistics-based approach and example-based approach.
Statistics-Based Approach

Statistics-based approach is of no linguistic rules, but depends on large text corpus. Due to the lack of linguistic rules, MT system doesn’t conduct the analysis of source text’s linguistic structure, bringing out an target text that is linguistically ambiguous.

Example-Based Approach

Known as memory-based approach, example-based approach depends on those previously-translated materials. A database of analyzed texts are stored in the Translation Memory of MT system. In the process of translation, MT system can select similar expressions from its database to replace the source text. Its problem lies in the reconstruction of selected examples.

Translation is a complicated process consisting of comprehension, analysis, selection, and reproduction [10]. A translator, with a source text and a given context, is capable of reassembling the target text on the basis of such background information as his bilingual knowledge, common sense, culture, history, geography and social customs [11]. MT, on the other hand, without the abilities of thinking, reasoning and judging, is unable to get a full picture of the source text, and thus generates a target text that is absurd.

Suggestions for Application of Machine Translation in Translation

Full dependence on MT still has a long way to go. In order to obtain high quality of MT translation, MT should be cooperated with HT. Human involvement mainly takes place before and after the translation process, namely, the task of pre-editing and post-editing [12].

Pre-Editing

Pre-editing refers to the selection of the source text. The quality of MT translation is closely connected to how MT-friendly the input is [13]. So the material to be translated with MT system should meet the requirements as below:

- Use short sentences. As it is mentioned above, MT system is incapable of analyzing the grammatical structure of the source text. Long sentences and complicated sentences may bring confusion to MT system. So the sentence structure should be kept simple and clear to avoid ambiguity in the target text.
- Avoid using literary works. Many culture-loaded words are employed in literary works. MT system, unable to figure out those words’ cultural connotations, usually fails to reveal the intended meaning of source text.

Post-Editing

Post-editing refers to the correction of the target text, i.e. the raw translation generated by the MT system [13]. Dr. Falko Schafer [14] pointed out that the process of post-editing consists of three stages:

- General output checking. The translator should go through the whole target and identify the mistakes produced by MT.
- The MT output editing. In this stage, the translator should focus on correcting the translation errors, for example, translating the words or sentences which MT system fails to translate correctly, removing ambiguity, and so forth.
- Proofreading. Proofreading is the last stage where translator reads and checks the target text again to ensure that every single mistake has been corrected.

Conclusion

Although MT is efficient enough to be well accepted by some translators, it has a fatal defect that it lacks human’s cognitive competence, therefore, the output by MT should never be considered as the
final version of target text. In order to improve the quality of translation as well as the efficiency of translation, a corporation of MT and human involvement is encouraged. At present, it is impossible for MT to substitute for HT. But it is still expected that, with the development of artificial intelligence and the advance in MT system, MT system will be constantly developed and upgraded, becoming a useful tool for translators and promoting the development of translation.

Acknowledgment

This work was supported by Guangdong Provincial Postgraduate Educational Creative Project (2014JGXM-ZD05) and Guangdong Provincial Higher Education Reform Project (GDJC20142107).

References

[8] Information on http://grammar.about.com/od/mo/g/nounphraseterm.htm