

A Study on Translation Quality of C-E Scientific Texts Cohesive Devices in Machine Translation

Lin Zu, Zebing Yang

English Department, North China Electric Power University, BaoDing, China
zulin3188@163.com, yangzebingncepu@163.com

Abstract - Studies on Machine Translation have developed rapidly and been put into practicality since 1980s. A great number of machine translation products have already made achievements in commercializing. Nevertheless, difficulties in analyzing complex and ambiguous sentences still have no effective methods so far. This paper will analyze and evaluate the machine translation quality of Chinese-English scientific texts under the guidance of discourse cohesive devices so that to find the strengths and weaknesses of the machine translation in dealing with this typical non-literature texts and present the according strategies in and after the translation process.

Index Terms - translation quality, scientific texts, discourse cohesive devices, machine translation, translation strategy

1. Introduction

Machine translation, which is the product of high advanced technology in modern times, has been widely used in translating various materials so as to promote the international exchanges.[1] However, it still has some distances with the human translation version. In order to achieve a better translation effect, some human intervene methods should be

used. Cohesive devices are a good choice to a large extent.

2. Categories of Cohesive Devices

The cohesive devices have three major categories: grammatical: structural content; lexical: language content of the piece and the logical connectors. [4]

Grammatical Devices include Time Relator, Place Relator, Reference, Substitution and Ellipsis; Lexical Devices include Reiteration, Collocation and Lexical chain; Logical connectors are used as sequencers, contrasters, generalizers, and topic-introducers. [2]

3. Evaluation of Machine Translation Quality

This paper picks one Chinese scientific text as the source text and its professional human translation as well as the according Trados machine translation so as to evaluate the machine translation quality.

3.1 Comparison of Professional HT Version and Trados MT Version

| Chinese Source Text [7] | Professional HT Version [7] | Trados MT Version |
|-------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 周三，科学家称，在完成了对病毒进行的最初试验后，他们已确凿无疑的证实，一种新的冠状病毒系SARS的病原，因为对这种病毒的确认符合所有公认的科学标准。(1) | Scientists said on Wednesday they had proved beyond doubt a new coronavirus was the source of SARS after completing the first tests on it that met all accepted scientific standards. | Scientists said on Wednesday after finishing the first test of viruses, they have conclusively proved a new kind of the pathogen of SARS coronavirus because the virus confirmed met all accepted scientific standards. |
| “这一确认说明，冠状病毒是直接病原。”负责实验的滤过性病原体学者小组领导，阿姆斯特丹Eramus医疗中心的艾伯特.奥斯特豪斯教授如是说。(2) | "This is proof that the coronavirus is the primary cause of the disease," said Professor Albert Osterhaus, head of a team of virologists at the Erasmus Medical Center in Amsterdam that carried out the tests. | "This verification shows that coronavirus is pathogens directly." Scholars' team leader is responsible for the experiment of viral pathogens, Amsterdam Erasmus medical center Professor Albert. Oster newhouse said. |
| “这一发现将大大加快诊断速度.(3)这也将大大加快抗病毒的药物以及疫苗的研制，因为现在我们知道主攻方向。”他在接受一次采访时说。(4) | "It will speed up diagnostics. It will speed up antiviral development and it will speed up vaccine development because now we know what we have to focus on," he said in an interview. | "This discovery will greatly speed up the diagnosis, also will greatly accelerate the antiviral drugs and vaccine development, because now we know that the main attack direction." he said in an interview. |
| 但该小组也表示，其他传染病也可能导致某些SARS患者呼吸道疾病的恶化。(5) | But the team said other infections may have exacerbated the respiratory disease in some SARS patients. | However, the team also said other infectious diseases might lead to the deterioration of some SARS patients with respiratory diseases. |

| | | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>尽管冠状病毒已被确认为是SARS的病因，但是，一种疾病的病原体的最后确认还必须符合德国微生物学家罗伯特·科赫于19世纪末制定的几条科学标准。这几条科学标准被称为“科赫假设”。(6)根据该假设，病毒必须存在于相同疾病的所有病例中；(7)病毒能从被感染者体内分离出来并在实验室进行培养；(8)</p> | <p>Although a coronavirus had been identified as the cause of SARS, certain scientific guidelines laid down in the late 19th century by German microbiologist Robert Koch must be met to establish a pathogen as the cause of a disease. Known as Koch's Postulates, they state that the agent should be found in all cases of the disease. It should be isolated from those infected and cultivated in the laboratory.</p> | <p>Although the coronavirus has been identified as the cause of SARS, the final confirmation of a kind of disease pathogen must conform to a few scientific standards made by German microbiologist Robert Koch in the end of the 19th century. This a few scientific standards is called "Koch hypothesis". According to this hypothesis, the virus must exist in all cases of the same disease; the virus can be isolated from infected body and training in the lab;</p> |
| <p>经培养的病毒在植入相同物种或相关物种体后能导致相同的病症，同时能从新的宿主体内重新分离并能检测出特定的免疫反应。全世界先后有好几个 SARS 科学攻关小组的研究满足了上述三个假设。(9)</p> | <p>The cultivated agent must reproduce the disease when introduced into the same species or a related one, be re-isolated from the new host and a specific immune response must be detected and several scientific groups working on SARS around the world had met three postulates.</p> | <p>The cultivated virus can result in the same condition after the implantation of the same species or related species, at the same time can separate from new host body and can detect the specific immune response. There were several SARS scientific research groups all over the world meet the above three hypotheses.</p> |
| <p>但只有奥斯特豪斯和同事们用 SARS 患者身上的病毒使两只短尾猿受到感染之后，又从受感染者身上分离出了该病毒，对该病原体的确认工作才得以最后确认。(10)</p> | <p>Osterhaus and colleagues completed the final ones when they infected two macaque monkeys with the virus from a SARS patient and isolated it from the animals.</p> | <p>Osterhaus and his colleagues in SARS patients after two macaques infected the virus, and isolated the virus from the infected, for the validation of this pathogen to our final confirmation.</p> |

3.2 Analyses of Machine Translation Quality through Cohesive Devices

3.2.1 Result of Comparison and Contrast of Two Translation Versions through Cohesive Devices

There are ten sentences in this scientific texts and the author has marked each of them with Arabic numerals 1-10. In the first and sixth sentences, the professional HT version, “on Wednesday” and “in the late 19th century” are the representatives of time relator. The equal expressions in MT version are almost the same except that MT version uses “in the end of 19th century.”

The place relator occurs in the second sentence with the professional HT version of “at the Erasmus Medical Center in Amsterdam.” This place expression is translated by machine as “Amsterdam Erasmus medical center”. Compared these two sentences, the MT version has expressed the basic meaning but it loses the prepositions which are quite important for the place expressions, such as “in” and “at” in the HT version. In addition, a lot of the place expressions are the proper nouns or specialized terms which should be translated with the first character capitalized. While the MT version has not paid attention to this so that it wrongly use “medical center” rather than “Medical Center”.

In order to make a better analysis of the translation quality, the reference and substitution, which are quite significant for texts, cannot be ignored at all. In the first sentence of professional HT version, personal reference “they”

is used to refer to scientists, article “a” refers to a kind of coronavirus and “it” to the virus. The MT version does same in the first two as HT version, while MT version does not use “it” to indicate virus, but only use the word “virus”. In the second sentence of HT version, “this” refers to “this verification” and “it” to “this discovery”. In sentence five and nine of HT version, comparative references occur in “other infections” and “the same species”. MT version does the same. Certainly, this has a big relationship with the Chinese version. It also shows the efficiency and accuracy of MT.

Substitution appears in sentence 9 of HT version which uses “one” to infer specie and crucial criterion so as to make the text consistent and concise, while, the detailed content is pointed out in MT which is translated as “species”.

Ellipsis is one of the most obvious differences between the MT and HT version. In sentence 2, the HT version adds “of the disease” after “the primary cause” in order to convey a much clearer meaning. But the MT version emits this important information. In sentence 9, the MT version uses “the above three hypotheses” so that to indicate the previous content. While the HT version only uses “three postulates”, which cannot only conveys the meaning, but also makes the expression clearer and simpler. This is also true in the last sentence, compared with the HT version “Osterhaus and colleagues” and MT version “Osterhaus and his colleagues”, it is easier to find that MT has the tendency to use more words and expressions to convey the same meaning than the HT version. Therefore, MT has the short comings in ellipsis.

Logical connectors exist twice in the discourse. Without the link of logical connectors, the text will be hard for the writers to continue and the readers also have trouble in comprehending it. Therefore, logical connectors can become a pretty useful and practical method to provide help to student in their processing of understanding. In sentence 4, 5, 6, 8 and 9 of the HT version, the logical connectors provide the basic description order and logic pattern of the whole text. No matter the use of “and” inferring the coordinating relation, or “but” adversative relation, the MT version also presents the excellent performance in translating these logical connectors.

The lexical devices can play a significant and necessary part in evaluating the translation quality. The first sentence of HT version uses “source” to show the most important and basic reason to cause SARS, while MT uses “pathogen” to present the same meaning. Both of two words can describe diseases. But, “pathogen”, here is much more suitable because it is a specialized term. Using “pathogen” can form a relatively integrated lexical chain. From this point of view, MT does better in some certain aspects. In sentence 2, HT version uses “primary” to show the prior cause of the disease, while MT uses “directly”. According to the lexical chain under lexical devices which indicate that the words or the expressions should be in the same kind and belong to one category. “Directly” does not match the whole content of this discourse. In addition, another aspect that should be paid attention to is the translation of the specialized terms. Specialized terms are quite critical to constitute the whole lexical chains. In sentence 7, the HT version uses “Koch's Postulates” to refer to some scientific standards, while the MT version translates it as “Koch hypothesis” which seems quite strange for people who are familiar with this term and may mislead others who have no idea about it. As is known to all, MT can offer a relatively accurate and professional translation in most specialized terms. However, it does not mean that it can do it perfect to some extent. Therefore, it is quite significant to focus on the translation of the specialized terms so as to ensure the specialty of the scientific texts.

Sentence 3 and 4 are the typical examples that can be analyzed and evaluated by lexical devices. The HT version uses three times “it will speed up” to constitute the repetition so that to make the text easier to read and full of musical rhymes. Obviously, it is quite difficult for MT to achieve this goal. However, the MT version presents the application of reiteration which is also worth using for reference. The MT version uses “speed up” and “accelerate” so as to express the fast and tremendous influence on other medical aspects. In the last sentence, the HT version uses “the animals” (superordinate) to indicate the infected two macaque monkey, while the MT version only uses “the infected”. Therefore, as to the aspect of reiteration, the MT can achieve the goal to some extent. However, the improvement and proper human intervene is necessary.

3.2.2 *Analyses of the Result and Offering According Recommendations*

From above analysis of the translation quality through the comparison and contrast of the HT and MT version, on one hand, the translation quality of MT has something worth for using as reference; on the other hand, some certain aspects should be paid more attention to so as to improve the translation quality. [5] The above result presents that MT has an excellent performance in translating the time relator, place relator, and logical connectors. The MV version is extremely similar with the HT version in these three aspects. Therefore, after the translation of MT, it has no big necessity to make the revisions to a large extent. However, some certain aspects should be more focused on. The result of MT version shows both the strengths and weaknesses of the in reference, substitution, ellipsis, lexical chain and reiteration. The analysis presents that the MT version can perform well when the original Chinese text has the according pronouns and the relations of nouns and pronouns are clear. The MT version would like to use “this” or “that” +nouns to express the according meaning, while the HT version directly use “this”, “that” or “it” to refer to the relevant contents so as to make the text much simpler and more concise. Lexical chain is one of the most significant aspects that should be paid attention to. Lexical chain can form a sequence of related words as well as expressions so as to integrate the whole text. As in this aspect, the MT version performs well in some certain words and expression. However, the problem cannot be ignored and denied at the same time. The above analysis concerned with lexical chain shows that the MT version has the shortcomings in using the words or expressions according to different context. The later adjustment is quite in need. Among these, the translation of the specialized terms should be improved of the MT version so that to make sure the coherence and specialty. Last but not least, it is quite surprising that the MT operate well in reiteration, especially in verbs. As to the application of superordinate, synonymy, hyponymy and general word, it is the shortage of the MT version. Therefore, the later revision and human intervene is needed so as to make the translation quality of MT more excellent. If people can notice the weaknesses of the MT translation of the scientific texts, they can have the according revisions under this beneficial guidance. Thus, the machine translation can contribute more to the transcultural communication. [6]

4. Conclusion

Under the influence of information age, people have to deal with various kinds of information so as to have a relatively clear understanding about what happening in the world. Translation has become more and more significant at present and thus, machine translation, which has the benefits of quickness and cheapness, has been popular to a large extent. This paper tries to evaluate the machine translation quality under the guidance of cohesive devices so as to seek the strengths and weaknesses of the machine translation of C-E scientific texts and giving the according suggestions concerned

with the human intervention, which may have some help in promoting the translation quality.

References

- [1] B.J.Dorr. 1993. *Machine Translation*. Cambridge, MA: The MIT Press.
- [2] Brown,G. Yule,G.1983.*Discourse Analysis*. Cambridge: Cambridge University Press.
- [3] Carrell, P.L.1982.*Cohesion is not Coherence*. TESOL Quarterly, 16, 4:479-488.
- [4] Hatch, E.1992. *Discourse and Language Education*. Cambridge: Cambridge University Press. eaton,
- [5] J.B.1988.*Writing English Language Testing*. London: Longman.
- [6] J.E.Holmstrom.1951. *Report on the Interlingual Scientific and Technical Dictionaries*. Paris: UNESCO
- [7] Zhiwei Feng. 2002. *The Current Situation and Problems in Machine Translation*. In: colloquium of computer science. KAIST.