https://doi.org/10.52810/TC.2021.100022

### Errors of Machine Translation of terminology in the patent text from English into Chinese

Cheng Ying<sup>1</sup>, Yue Shuyu<sup>2,\*</sup>, Li Jing<sup>3,\*</sup>, Deng Lin<sup>1</sup> and Quan Qi<sup>4</sup>

- <sup>1</sup> School of English Studies, Xi 'an International Studies University 710128 Xi'an, China
- <sup>2</sup> School of Foreign Languages, Yulin University, 719000 Yulin, China
- <sup>3</sup> Shaanxi Business College, The Open University of Shaanxi 71077 Xi'an, China
- <sup>4</sup> School of Translation Studies, Xian International Studies University, 710128 Xian, China
  - \* Corresponding authors: Yue Shuyu (yueshuyu@163.com); Li Jing (1606366625@qq.com)

Manuscript submitted 30 April 2021; Accepted 17 May 2021; Published 19 May 2021

Academic Editor: Jinchao Chen®

Abstract: This paper summarizes eight types of error of terminology in the patent text in the output of Machine Translation from English into Chinese, including term being mistranslated as a verb, term being mistranslated as a common noun, term being redundantly translated, term being mistranslated as a homophone, term being mistranslated as a wrong term, term being mistranslated due to Chinese expression, term being mistranslated without initial, and term being mistranslated due to wrong acronym. These errors can be solved by the translator before Machine Translation and the translator can identify and correct these errors by pre-editing of the source text.

Index Terms: Errors of Terminology Translation, Machine Translation, Pre-editing, Patent Text.

\_\_\_\_\_\_

#### 1. Introduction

In 2019, the number of patent applications in China is over 1 million, ranking the first in the world according to the report of Patent Office of the People's Republic of China in 2017. Moreover, Chinese people also actively apply for the patents in the United States and other nations in the world in the resent years. Patent abstract is a summary of the patent specification, an important element applied to the innovation applications and utility model applications. Accordingly, the demand for Chinese-into-English translation of patent text has increased greatly.

The error of terminology could be identified and corrected in the output of Machine Translation (shorten as MT). It is a good way to identify and correct the error of

terminology in the output of Machine Translation by preediting of the source text. During the pre-editing process, the translator should be a careful proofreader. In this paper, the types of error of terminology in the output of Machine Translation are summarized which can help translators identify and correct when pre-editing of the source text.

# 2. Reviews on pre-editing of Machine Translation and feature of patent text

#### 2.1 pre-editing of Machine Translation

Generally speaking, Machine Translation cannot replace the translation of human beings completely so the output of Machine Translation should be post-edited by human beings. Since the birth of Machine Translation, the human beings

have studied how to improve the quality of Machine Translation. Nowadays, the common MT modes are "MT+post-editing" and "pre-editing+MT+post-editing". Of course the "pre-editing+MT+post-editing" is more efficient.

Some foreign researchers especially have developed the pre-editing systems. Pym proposed an idea of pre-editing on source texts before using MT. He introduced ten pieces of advice of how to simplify the source texts, in which five pieces are meant to keep the sentence short and simple, and the other five pieces are to make the sentence explicit [1]. Babych and Hartley [2] developed the pre-editing system of ACCEPT, and Seretan, Bouillon, and Gerlach [3] developed the pre-editing system of Automatic Text Simplification (ATS). Okazaki [4], Ding, Sakanushi, Touji, and Yamamoto [5] performed the pre-editing systems of MT between Japanese and English.

Some Chinese researchers also have focused on the study of pre-editing strategies. Zeng and Hou [6] proposed five pre-editing strategies for MT of CE texts, including rewriting, deleting redundant information, explanation for proper nouns, adding proper function words and adjusting position of attributive. Based on the experiments on CE translation in the Google MT engine, Hu [7] proposed nine pre-editing strategies, including reordering, adding, omitting, pre-translating, rewriting, segmentation, paraphrasing, simplifying and re-punctuating. Huang [8] proposed four pre-editing strategies: sentence reconstruction, addition, pre-translation of terminology and addition of punctuation. Based on their translation experience, Cao and Ma [9] put forward five pre-editing strategies for the Machine Translation of Tradition of Chinese Medicine news, including omitting words, adding words, adjusting word order, rewriting proper nouns, and explaining and replacing the special concepts in source text. In the light of Machine Translation of Tradition Chinese Medicine, Wang and Yun [10] proposed a pre-editing principle of "maintaining the meaning of the source text, and simplifying the language chunks characteristics of the classical Chinese".

In a word, for the pre-editing of machine translation, the researchers abroad mainly focus on the development of preediting systems hoping to develop new machine translation system for pre-editing, while the researches at home are mainly focus on specific pre-editing strategies for dealing with specific texts. The previous studies can be a reference for reducing the error of terminology of Machine Translation in the patent texts.

#### 2.2 Feature of patent text

The patent text is a kind of technical text with the purpose of conveying information. The features of patent text are as followings.

At the lexical level, in the patent text, there are many technical vocabularies, which are terms and abbreviations, some of them being common words, some of them being legal words, some of them technical words.

At the syntactic level, there are a lot of long sentences and compound sentences in the patent text. Also, there are a lot of passive voice sentences in the patent text. The logic of sentences in the patent text is often complex.

For example, the following is a long sentence in a patent text. The present invention is provided with an automatic dispatching machine, which can be added a neutralizing medicinal method to be sprayed with a small round shower head to neutralize the toxic gas in the exhaust gas, and which can maximize the exhaust gas cleaning effect to make the machine a good exhaust gas treatment effect.

At the textual level, patent text has a fixed format, in which the patent specification, claims, drawings and patent abstracts are arranged in a certain order.

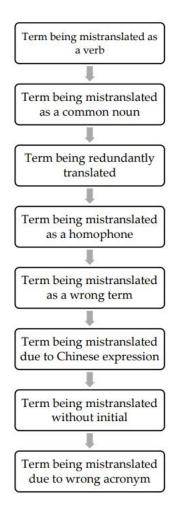
For example, the WIPO's Guidelines for the Translation of PCT Application Abstracts is specified that "an abstract of a patent document is a concise statement of its technical disclosure." [11] In other words, a patent abstract is an overview of the whole invention, a kind of information inspection document by which the readers can learn the contents of invention or utility model quickly.

## 3. Types of error of Machine Translation of terminology

In this section, the types of error of terminology of Machine Translation are illustrated by the comparing the Machine Translation (shortened as MT), and Revised Translation (shortened as RT). Generally speaking, there are eight types

of error of terminology of Machine Translation of terminology, shown in diagram 1.

Diagram 1 Eight types of error of terminology of Machine Translation of terminology



#### 3.1 Term being mistranslated as a verb

In Chinese, some nouns have the same form of the verbs. In the patent abstracts, a few terms would be identified and translated as the nouns in the machine translation because the machine cannot distinguish them in the source text.

Example (1)

MT: The invention relates to an electronic-grade hydrogen fluoride washing and absorbing device.

RT: The invention relates to a washing and absorbing device for preparing electronic-grade hydrogen fluoride.

In the MT, the term "an electronic-grade hydrogen fluoride washing and absorbing device" cannot be understood clearly. After careful examination of the source text, it is found that the verb in front of the device can be used as a noun in Chinese. When the translator tried to use

a synonym to collocate to the device, he can get a suitable expression which can convey the meaning in the source text.

#### 3.2 Term being mistranslated as a common noun

In Chinese, the collocation of a verb and a noun being as a term would be perfect, but the collocation would be mistranslated by the machine translation because the noun being as the term in the collocation is mistranslated as a common noun. The mistranslation is an unsuitable term because it does not fit in the context.

Example (2)

MT: Log in to the PN5/OpenStack background, and perform the following operations.

RT: Log in to the PN5 or OpenStack server, and perform the following operations.

In the source text, "background" refers to a server of a system in the computer science, but it is mistranslated as a common noun by machine. The common noun "background" is a word frequently-used, which is transplanted from Beijing opera, meaning a room where actors make up and wait for going to the stage to perform. The word also is used frequently in Chinese, implying the strong political and economic support for somebody that can be obscured usually but shows great effect at special time. Because "background" is frequently used as a common noun, it is mistranslated in the technique text. Such errors should be corrected by the translator.

#### 3.3 Term being redundantly translated

In Chinese, some nouns in a sentence should be repeated, especially in a definition sentence, which is a practice. However, in English, if the repeated nouns are used in a definition sentence, they are redundant words. The redundant words in source text often lead to constituent redundancy which refers to the occurrence of unnecessary words in a sentence that has a complete structure and clear meaning (Huang & Liao, 2002:148). In the output of machine translation, the repeated nouns in Chinese would be faithfully conveyed into English, which becomes the constituent redundancy in English.

Example (3)

MT: The balloon is a balloon made of a shrinkable material.

RT: The balloon is made of a shrinkable material.

In Chinese, the term "the balloon" is repeated twice, but the meaning is clear. However in English, the term "the balloon" is repeated twice, resulting a non-standard expression, which does not make the sentence structure concise and clear.

#### 3.4 Term being mistranslated as a homophone

In Chinese source text, some terms are misused by wrong terms, which are the homophones. The homophones are words with different meanings which are pronounced in the same way but are spelled differently. In Chinese, there are many homophones. Because a lot of authors write a Chinese text by Pinyin input method, and if they have not carefully proofread the source text, there will be some homophones in the source. However, the homophones in the source text are translated into English with the meanings which are totally different from the context.

Example (4)

MT: A vacuum plate is laid on the slide cover.

RT: An array control plate is laid on the slide cover.

In Chinese, "vacuum" and the abbreviation of "array control" is a pair of homophones. However, the meaning of "vacuum" is totally different from that of "array control". Therefore, the meaning of the machine translation clearly indicates the wrong logic in the context. Sometimes, because the output of the machine translation clearly indicates the wrong logic in the context, the translator can find the mistake easily, but sometimes, the translator can find the mistake easily, because the homophones mean the two different elements or two different theories, which look likely correct in the context. In the light of this type of error, it is suggest that the translation version should be finally proofread by the academic experts.

#### 3.5 Term being mistranslated as a wrong term

Sometimes, a same term are repeated twice by the output of the machine translation. Such clear mistakes can be found during the proofreading, but the proofreader should spend more time on look for the reasons. Such mistakes are happened because of the mistaken terms in Chinese text, which sometimes should be checked by consulting the author of the Chinese text to confirm what on earth the two elements or two devices are.

#### Example (5)

MT: The invention includes a box body, an air duct, a brush, a rotating rod, a control panel, a single-chip microcomputer, a voltage sensor, and a voltage sensor.

RT: The invention includes a box body, an air duct, a brush, a rotating rod, a control panel, a single-chip microcomputer, a voltage sensor, and a current sensor.

Generally speaking, if a device has two voltage sensors, it should be written as "two voltage sensors". In Chinese, the related terms would share a same element which is featured with same property or relationship. For example, many terms are begun with Chinese character "dian" (meaning electricity). The terms beginning with Chinese character "dian" signify the objects which are related to electricity. For the above example, the translator does not find the reason unless he consults to the author of the Chinese text. According to the author of the Chinese text, he wants to express "a voltage sensor (dian ya) and a current sensor (dian liu)", but he writes "a voltage sensor (dian ya) and a voltage sensor (dian ya)".

#### 3.6 Term being mistranslated due to Chinese expression

In Chinese, some expressions are unique, which are popular among common people. Such expressions are often used in order to highlight the advantages of something. Such expressions are used in some patent texts by scientists. Such expressions can cause the terms to be mistranslated by machine translation.

Example (6)

MT: The ceramic device has the advantages of movable stirring blades.

RT: The ceramic device is provided with movable stirring blades.

In the source sentence, an advantage of the device is outstood if the device is added some accessories. However, such unique expressions in Chinese are faithfully translated into English by the machine translation, which become the redundant words, and add the additional meaning. In other words, the movable stirring blades are parts of the invention, not being the advantages of the invention.

On the other hand, In Chinese, some expressions are colloquial, which are often expressed without punctuations, and popular among common people. Such expressions are

used in some patent texts by writers, which would cause the terms to be mistranslated by machine translation.

Example (7)

MT: The invention includes a good upper shaft stirring, a driving shaft, a general lower stirring, and a last coupling.

RT: The invention includes an upper shaft, an upper stirring, a driving shaft, a lower stirring, and a coupling.

In the source patent sentence, there are some punctuation mistakes made by the writers or typists. However, such punctuation mistakes cannot be judged by the machine translation because the "I translation+" platform lacks of punctuation.

#### 3.7 Term being mistranslated without initial

The machine translation cannot distinguish names of scientists and names of units which are named after those scientists.

Example (8)

MT: It is watt on the steam locomotive improvement that greatly increased the energy use rate.

RT: It is Watt on the steam locomotive improvement that greatly increased the energy use rate.

In the source text, Watt refers to the name of inventor of steam locomotive, not the unit "watt", which should be capitalized the initial letter.

The machine translation also cannot distinguish the initial names which are not capitalized.

Example (9)

MT: The invention also includes a CPLC controller, which can control an automatic liquid injection tool.

RT: The invention also includes a CPLC controller, which can control an automatic liquid injection tool.

In the source text, CPLC is an initial name. However, the machine translation cannot distinguish it, and make it capitalized.

#### 3.8 Term being mistranslated due to wrong acronym

A few acronyms misspelled by the author's careless would be the mistakes by the machine translation. The machine translation conveys the misspelled acronyms in the source text faithfully into the target text, which should be identified by the proofreader carefully.

Example (8)

MT: The concept is consistent with TSL Syslog and Logstash TCP.

RT: The concept is consistent with TLS Syslog and Logstash TCP.

In the source text, TSL is an acronym terminology, which is wrongly spelled. The correct spelling for this acronym should be TLS, shortened from "Transport Layer Security".

#### 4. Summary

The demand for Chinese-English translation of patent texts has increased in China. The machine translation is a good way to translation of patent texts and it is important for a translator to pre-edit the source texts before machine translation. Eight types of error of terminology in the output of Machine Translation include term being mistranslated as a verb, term being mistranslated as a common noun, term being redundantly translated, term being mistranslated as a homophone, term being mistranslated as a wrong term, term being mistranslated due to Chinese expression, term being mistranslated without initial, term being mistranslated due to wrong acronym. These errors can be solved by the translator before machine translation and the translator can identify and correct by pre-editing of the source text.

#### Acknowledgement

The research is supported by research fund from Xi'an International Studies University (Fund No. 16XWA05).

#### References

- [1] Pym, P. J. (1990). Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system. *Translating and the Computer*, 10, 80–96.
- [2] Babych, B., & Hartley, A. (2009). Evaluation-guided preediting of source text: improving mt-tractability of light verb constructions. *Proceedings of the 13th Annual Conference of the* EAMT, Barcelona, 36-43.
- 3] Seretan, V., Roturier, J., Silva, D. & Bouillon, P. (2014). The ACCEPT portal: an online framework for the pre-editing and post-editing of user-generated content. Workshop on Humans and Computer-assisted Translation, 66-71.

- [4] Okazaki, Y. (1995). Pre-processing of punctuation in Japanese-English machine translation. *Ipsj Sig Notes*, 109, 1-8.
- [5] Ding, C. et al. (2016). Inter-, intra-, and extra-chunk preordering for statistical Japanese-to-English machine translation. ACM Transactions on Asian and Low-Resource Language Information Processing, 15(3), 1-28.
- [6] Zeng Liren, & Hou Xiaofang. (2012). Preliminary study on the pretreatment strategy based on the structural differences between English and Chinese in Chinese-English machine translation. *Overseas English* (5x): 18-20.
- [7] Hu Miaodan. (2012). Pre-translation editor of Chinese-English machine translation. Master's thesis of Zhejiang Normal University.
- [8] Huang Yueyue. (2017). On the application of pre-translation editing in machine translation. *Journal of hubei correspondence university* (3): 175-176.
- [9] Cao Yuan & Ma Yongliang. (2018). Pre-translation editor of Google Chinese-English translation — a case study of the English translation of a press release of Beijing information network of traditional Chinese medicine. Overseas English (8): 153-155.
- [10] Wang Nanxing & Yun Hong. (2019). Machine assisted translation of the efficacy instruction of traditional Chinese medicine. Journal of mianyang normal university (4): 93-99.
- [11] WIPO. (1994). General Guidelines for the Preparation of Abstracts of Patent Documents. Retrieved from https://www.wipo.int/export/sites/www/standards/en/pdf/03-12-a.pdf