OPEN ACCESS
Saudi Journal of Engineering and Technology
Abbreviated Key Title: Saudi J Eng Technol
ISSN 2415-6272 (Print) |ISSN 2415-6264 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Review Article

Deficiency Analysis of Machine Translation and Post-editing Based on Computational Linguistics—A Case Study of the Translation of *Government Work Report*

Yan Chu[1*], Zhirong Liu[1]

[1]Department of Foreign Studies, North China Electric Power University, Baoding, Hebei, China

DOI: 10.36348/sjet.2021.v06i11.007 | **Received:** 16.10.2021 | **Accepted:** 20.11.2021 | **Published:** 24.11.2021

***Corresponding author:** Yan Chu

Abstract

Computational linguistics is a subject closely related to contemporary science and technology. The research results of computational linguistics have been applied in machine translation, information retrieval, natural language man-machine interface, and other important fields. At the same time, with the development of machine translation, problems have also emerged. This paper find out some deficiencies of machine translation and post-editing strategies from the comparison between Youdao translation and manual translation of *Report on the Work of the Government 2021*. Machine translation is ultimately a linguistic problem. In post-editing, translators should pay attention to context, logical relationships, four-word phrases, and so on.

Keywords: Computational Linguistics, Machine Translation, Post-editing, Manual Translation.

1. INTRODUCTION

Recently, people pay more and more attention to computational linguistics, which is closely related to science and technology. Machine translation is also developing towards a better future. Computer programming language is as important as natural language. At the same time, there are still some shortcomings in machine translation, and the in-depth translation is not satisfactory. The problem of machine translation is also a linguistic problem. Therefore, no matter how good the software and hardware technology is, if the dictionary grammar is not well compiled, machine translation will not be greatly improved. It is urgent to strengthen language research and summarize the experience of human translation.

2. Computational Linguistics and Machine Translation

Western countries first concentrate on the influence of computational linguistics on machine translation. Since the 1930s, machine translation has been known for its rapid development.

2.1 Computational Linguistics

The earliest linguistics still focused on the analysis of western languages. Some linguists wanted to use linguistics to make machine translation more accurate. Since 1989, the development of machine translation had entered a new era. Machine translation based on large-scale text processing was a revolution in the history of machine translation research. It would make machine translation develop better. After the beginning of the new era of machine translation, computational linguistics had entered its prosperity (Feng Zhiwei, 2019).

Computational linguistics is a subject that uses computer technology to study and process natural language. It is a subject closely related to contemporary science and technology. Computational linguistics is connected with computer science, mathematics and linguistics. Its research results have been applied in important fields such as machine translation, information retrieval, natural language man-machine interface, and so on.

Computational linguistics includes the relationship between computer and language research. One view is that computers are tools for language research. Another view is that the computer is the goal and service object of language research. In recent years, the definition of computational linguistics has gradually become a discipline that uses computers to study languages. Clearly, computational linguistics uses

computers as tools, such as collecting corpus, classifying, distributing statistics, extracting various data, and so on. Scholars believe that the study of language for computers refers to the study of language so that computers can process natural language (Hutchins, 2003). This includes two aspects of work. The first is to summarize the structure and meaning of natural language, extract the syntax convenient for formalization and algorithmic, and establish an appropriate grammatical theoretical model to better organize the syntactic rules of language. The second point is to summarize the research results of linguists on the syntax, semantics, and pragmatics of language, use some formal system to organize and express the structure and meaning rules of language, then find out the appropriate algorithm to describe the steps of sentence structure, and finally use the corresponding computer language to program according to the algorithm.

2.2 Machine Translation

In China, with the development of science and technology, machine translation is also moving forward, but its bottleneck also appears. Chinese and English have different structures and different translation methods. Different from English, Chinese has a more complex grammatical structure and difficult vocabulary expression, which requires Chinese linguists and scientists to spend a long time to solve the phenomenon of word-for-word translation in machine translation. Therefore, it is necessary to solve the problems in machine translation from a new perspective. In fact, the problem of machine translation is still a problem of linguistics. If various phenomena in linguistics can be solved and implemented in machine translation, some problems in machine translation can be solved. China's computational linguistics has made many achievements, but there is still a big gap compared with the international level (Geng Libo, 2021).

Machine translation is the product of the combination of linguistics, mathematics, computing technology, automation technology, and other scientific departments. Every step of machine translation needs to skillfully combine these different disciplines to make them work together for machine translation. There are many achievements of machine translation, but most of them are due to the progress of software and hardware technology. Due to the complex operation process and the imperfect development of artificial intelligence technology, machine translation is a new technology with a wide application range and insufficient development. Of course, a practical system will not stick to a single theoretical model. From literal translation to language knowledge acquisition, it will absorb many methods that can improve the accuracy and efficiency of machine translation.

3. Deficiencies of Machine Translation

Machine translation still needs improving. Since the establishment of machine translation, it has gone through many stages. However, machine translation has only existed for less than a century. Machine translation has been developing, which continuously reduces the occurrence of errors. Scientists are also improving machine translation in many ways (Song Shizhen, 2019). Nevertheless, no one will completely trust machines. People will always process and modify the text of machine translation. In a word, machine translation is still an underdeveloped technology so far, which still needs technicians to continue in-depth research. There are still some deficiencies in machine translation. The following is a comparison between the Youdao translation and the official translation of the *Report on the Work of the Government 2021*, and then summarizes the defects of machine translation.

3.1 Vocabulary Translation Errors

Vocabulary is the smallest semantic unit and the basic element of a sentence. By analyzing the translation of words, it can be intuitively examined whether the translation has achieved semantic equivalence with the original text. The current machine translation platforms include a cloud database and a large and constantly updated corpus, which makes machine translation contain the ability of word recognition and translation. Theoretically, if the corpus database of machine translation is rich enough to expand and accommodate professional vocabulary from time to time, the translation of terms with high accuracy can be realized in machine translation. But in fact, the translation methods of terms and quasi terms that appear frequently on the Internet are not necessarily the same, but they will lead to errors in word translation. However, due to the imperfection of machine translation and the complexity of different languages, there are still some errors or deficiencies in dealing with vocabulary translation in machine translation, which needs the translator's post-editing.

Example 1. 纵深推进“放管服”改革，加快营造市场化、法治化、国际化营商环境。

Youdao translation: We will deepen reform to delegate power, improve regulation and improve services, and speed up efforts to create a market-based, law-based and internationalized business environment.

Official translation: We will deepen reforms to streamline administration and delegate power, improve regulation, and upgrade services and, move faster to create a market-oriented, law-based, and internationalized business environment.

The translation of words plays a fundamental role. When dealing with some special words, it can not translate accurately by machine translation, which will lead to problems such as missing translation and wrong translation. In the original text, "放管服" refers to streamlining administration and decentralization,

reducing access threshold, innovating supervision, providing efficient services and creating a convenient environment. Youdao translation cannot translate the full meaning of this word. In addition, there is the problem of word accuracy. For the translation of "市场化", Youdao translation is "*market-based*", while the official translation is "*market-oriented*". If the original text is an ordinary article, both translations can be used. However, the original text attaches great importance to accuracy, and the use of "*market-oriented*" reflects the role of the market, which is more accurate.

3.2 Translation Errors of Non-subject Sentences

Chinese language characterizes non-subject sentences. Subject does not appear literally but reflects the subject through context. In contrast, English is a hypotaxis language, which emphasizes the preciseness of the subject-predicate relationship. Therefore, in Chinese-English translation, it is often necessary to restore the subject of the non-subject sentence (Shi Zhaojia, 2021). Chinese emphasizes parataxis. The subject is sometimes not a necessary component, and the relationship between the subject and the predicate is relatively loose. Most importantly, the omission or ambiguity of the subject will not cause obstacles to the understanding of Chinese readers. However, English emphasizes hypotaxis. Most English sentences strictly abide by the subject-predicate structure and pay attention to grammar. In English, the subject and predicate are closely related and are an indispensable part of a sentence. While translating, machine translation tends to restore the subject to the third person without much distinction. Although this inflexible method improves the non-subject sentence translation of machine translation to a certain extent, this method shows mechanical rigidity.

Example 2. 对工作难度大的贫困县和贫困村挂牌督战，精准落实各项帮扶措施。

Youdao translation: We will supervise poverty-stricken counties and villages where work is difficult, and take targeted measures to help them.

Official translation: Counties and villages facing difficulty in poverty eradication were placed under special supervision to see they fully implemented all assistance and support policies.

This example clearly shows the rigidity of machine translation. Youdao translation restores the implied subject of the original text to the plural form of the first person. Although this translation method is correct in grammar and semantics, it weakens the textual function of the original text and still fails to realize the functional and value equivalence between the original text and the translation. The subject is often omitted in Chinese, so the processing of official translation is very flexible. Take "*counties and villages*" as the subject and use the passive voice. The whole sentence contains many non-finite clauses to connect each information. In contrast, translators can choose how to translate non-subject sentences flexibly according to the actual situation, while machine translation can only mechanically supplement subjects in general. When dealing with sentences without subject or fuzzy subject, machine translation often adds inappropriate subjects, which affects the accuracy of translation information. Therefore, in post-editing, the translator needs to infer the real subject according to the context. In translation, the translators should ensure the integrity of sentence structure, the standardization of grammar, and the accurate transmission of the original information.

3.3 Lack of Flexibility

Machine translation lacks flexibility. The essence of machine translation is automatic and mechanized translation. Machine translation is a scientific and technological means based on artificial intelligence and big data technology. It can translate the text that scientists input into the translation software. In this case, the machine system can only analyze the grammatical structure of the sentence mechanically, but it is difficult to combine the language environment, customs and habits, and the final result will be inconsistent with reality. Machine translation is still unable to accurately translate complex and multi-meaning sentences. The difference between Chinese and other languages is that for Chinese, a word can contain different meanings in different contexts. This brings some difficulties to machine translation. Of course, human translation can clearly know what the word should be translated into in different environments, but machine translation is quite rigid and can not show the deep meaning. Machine translation is often used to translate some simple texts because of its high efficiency.

Example 3. 不得扰民渔利，让市场主体安心经营、轻装前行。

Youdao translation: We must not disturb the people or benefit the people, and let market entities operate with ease and travel lightly.

Official translation: and no action that seeks to make gains at the expense of our people and businesses will be tolerated. All these efforts will lighten the burden on market entities and enable them to focus on doing business free from undue concern.

Machine translation lacks flexibility. Errors often occur when there are some uncommon expressions or metaphorical expressions in the original text. There are many metaphorical words in the original text. For example, "渔利" refers to profit, not fishing. The result of the Youdao translation is contrary to the original meaning. It is translated as "*not disturb the people or benefit the people*", which is a serious

mistake. In addition, for example, "轻装前行" refers to reducing the pressure on market players, which has no meaning of travel at all. The Youdao translation is "*travel lightly*", which is different from the actual expression. The official translation uses a lot of words to explain it. It can be seen from here that machine translation is inflexible. Translation does not mean translating every word of the original text. When there is a special metaphor, it needs to be translated according to the actual needs.

4. Implications for Post-editing

Machine translation has gradually formed four machine translation systems based on corpora, multi-engines, and oral machine translation systems. The models of each machine translation system are different, but machine translation still contains many problems. It has not really replaced human translation.

4.1 Relationship between Computational Linguistics and Machine Translation

Machine translation is actually a linguistic problem. No matter how good the software and hardware technology is, if the dictionary grammar is not well compiled, the level of machine translation will not be high. Summarizing the experience of human translation is vital for the improvement of machine translation. There are some difficult linguistic problems in machine translation, such as polysemy processing, preposition processing, and conjunction demarcation. Solving linguistic problems contributes to the development of machine translation.

4.2 The Reasons of Selection and Characteristics of the Original Text

The reason for taking the translation of *Report on the Work of the Government 2021* as an example is that it is clear to find the shortcomings of machine translation and have enlightenment to post-editors. Its English version is translated by the Central Compilation and Translation Bureau, with a high level and authority. This report has distinct Chinese characteristics and contains many Chinese culture-loaded words. How to avoid ambiguity is the difficulty of translation. By analyzing its official translation and Youdao translation, the results can be clearer, which can be used as a reference for researchers of political text translation. With the continuous improvement of China's international status, the need for political text translation is also increasing. As an official government document, this report covers all aspects of China. Its language is characterized by formality and seriousness. It is one of the most authoritative foreign publicity materials in China and an important way for foreign countries to understand China's principles and policies. The language of the Chinese government's political documents is usually concise, clear, and objective. In post-editing, the translator should adopt different translation methods according to the language differences between Chinese and English.

4.3 Implications of Computational Linguistics for Post-editing

Based on the relationship between computational linguistics and machine translation, translators can pay attention to the following aspects in post-editing.

4.3.1 Focus on Four-word Phrases

Four-word phrases are commonly used in Chinese, composed of only four Chinese characters, but they cover a wide range of contents. Some phrases can be broken up and recombined. Some are fixed phrases and Chinese idioms, which have special meaning and rich historical and cultural connotations. From the perspective of grammatical relationships, the Chinese four-character structure can be divided into subject-predicate, verb-object, coordinating relation, and so on. The use of four-character phrases can not only show the cadence of Chinese, but also have rich meaning, which is very rich in rhetoric and artistic beauty. In contrast, English is more hypotaxis, its expression requires clarity, and usually does not allow the repetition of phrases. Therefore, the translation of Chinese four-character structure is often difficult in Chinese-English translation, and it is difficult to choose translation strategies. It needs to be translated flexibly and appropriately according to the specific situation and needs.

Example 4. 特别是保就业保民生保市场主体，以保促稳、稳中求进。

Youdao translation: In particular, we need to ensure employment, people's wellbeing and market entities, and promote stability while ensuring progress.

Official translation: particularly job security, basic living needs, and the operations of market entities. By maintaining security, we were able to deliver stability while also pursuing progress.

Chinese is used to using four-word phrases one after another, so as to highlight a rhythmic beauty. It is necessary to deal with the translation of repeated words in the English translation of "以保促稳、稳中求进". The result of Youdao translation is to use paratactic short sentences. The original text highlights the logical relationship, which requires manual translation to integrate meaning and reprocessing. The official translation connects "*maintaining security*" with a preposition to highlight the simultaneity. In terms of word selection, the official translation uses "*deliver*", which also includes the meaning of fulfilling commitments and fulfilling expectations. In short, it should be paid attention to the translation of four-word phrases in post-editing, because machine translation is inflexible and machine translation often makes mistakes.

4.3.2 Enhance Context Awareness

As a parataxis language, Chinese is highly dependent on context. It makes it common for machine translation to translate only the literal meaning of the sentence but not the connotation of the sentence. The meaning of a sentence is inseparable from context, including textual context, linguistic context, and cultural context. The machine translation platform realizes automatic translation based on language generation and transformation rules and big data corpus, which is highly mechanical (Shao Nan, 2020). Therefore, when the machine translation platform processes the context of the original text, if the associated context is far from the target text, it will affect the translation effect.

Example 5. 以“一老一小”为重点完善人口服务体系，优化生育政策，推动实现适度生育水平

Youdao translation: We will improve the population service system, optimize the family planning policy, and work to achieve an appropriate level of birth.

Official translation: Improve the population services system with a focus on elderly care and child care. We will refine the childbirth policy, work to achieve an appropriate birth rate.

Machine translation cannot complete translation according to context, so it should be paid attention to the translation of special words and the expansion or reduction of the meaning according to context. The policy of "一老一小" in the original text refers to the inclusion of urban unsecured elderly, school students and preschool infants into the basic medical insurance system. Youdao translation directly chooses to omit translation when encountering this special word. In the actual translation, the policy content cannot be completely written out, and the official translation chooses to be translated as "*elderly care and child care*". This kind of translation not only covers the policy content, but also condenses the content. In the translation of "生育政策", Youdao is translated as "family planning policy", while the official translation is "*childbirth policy*". In the translation of childbirth policy, post-editing is very important. In particular, the *Report on the Work of the Government* is a document of political significance and plays an important role in foreigners' understanding of China's policies. Foreigners do not have the so-called family planning, because different countries have different policies and views, so when translating, the translators should avoid making foreigners have wrong cognition and disgust. In this example, the official translation is more accurate. In post-editing, the translators should pay attention to the flexible translation of some special words through context, and avoid deviating from the original meaning of the article.

4.3.3 Follow Logical Connections

When there are logical problems in machine translation of simple sentences, it usually does not need to make major adjustments, but only needs to supplement relational words. However, sentences with rich content and complex structure need to be adjusted when there is logical confusion in machine translation. A morpheme is the smallest semantic unit of a word. A lexical chunk refers to the collocation of words with complete semantics. The confusion of structure and logic in machine translation often comes from the confusion of morphemes or lexical chunks. In post-editing, it is necessary to split and reorganize the sentence structure to a certain extent based on the logic of the original text to make the sentence meaning clear. In post-editing, a Chinese sentence is often divided into several English sentences. This method not only does not be contrary to the original meaning, but also makes the logic of the original text more clear.

Example 6. 加快数字化发展，打造数字经济新优势，协同推进数字产业化和产业数字化转型

Youdao translation: We will accelerate digital development, build new advantages of the digital economy, and jointly promote digital industrialization and digital transformation of industries.

Official translation: Digitalization will be sped up to create new strengths for the digital economy. We will both develop digital industry and transform traditional industries with digital technologies.

In terms of sentence structure, Youdao translation follows the original text and translates in one sentence. The result is the redundancy of sentences. Many conjunctions are used to connect each clause. However, the official translation makes the Chinese original sentence into two English sentences, making the logic smooth. In post-editing, pay more attention to whether sentences can be split or merged to avoid redundancy.

4.3.4 Adopt Free Translation Strategy

Free translation is one of the commonly used translation methods. In political text translation, the translators should find and analyze the subtle differences and characteristics between Chinese and foreign cultures, and translate according to the thinking habits of the audience language. The translation of political texts is not to mechanically convert Chinese into foreign languages, but to make it acceptable to foreign readers as much as possible. In translation, the original text can be properly processed, sometimes deleted, sometimes added background content. The translator acts as a communicator between two languages and cultures, and the translation conforms to the language norms and expression characteristics of the target language. The purpose is to improve the

acceptance of the target audience and help the dissemination of Chinese in the world.

Example 7. 加快数字社会建设步伐，提高数字政府建设水平，营造良好数字生态，建设数字中国。

Youdao translation: Speed up the pace of building a digital society, improve the level of building a digital government, create a good digital ecosystem, and build a digital China.

Official translation: We will work faster to develop a digital society, digital government, and healthy digital ecosystem as we pursue the Digital China initiative.

The original text consists of four Chinese verb-object phrases. Youdao translation is a mechanical word-to-word translation, which is not in line with the stylistic characteristics of English. The official translation simplified these verb-object phrases and directly connected the object with the verb "*develop*", highlighting brevity. Sometimes the result of literal translation is not in line with English habits, so free translation should be used. "*Digital China*" in the original text refers to a goal, so the official translation does not juxtapose "*Digital China*" with other words but as a summary. In order to make the international community understand China better and convey information accurately, translators should analyze the characteristics of Chinese and English and choose different translation strategies when translating political texts. In translation, the translator should accurately and finely process the text, so that the target readers can fully understand the meaning of the text and avoid ambiguity. In post-editing, translators often have to find out the inappropriate parts of machine translation. They are not errors but can be translated into more appropriate results through free translation.

5. CONCLUSION

Machine translation has developed for many years since its birth. Machine translation can be developed independently as an artificial intelligence technology. Machine translation is actually a linguistic problem. The translators should not only improve the software and hardware technology, but also pay attention to the summary of linguistic problems. Computational linguistics is related to computer science, mathematics, and linguistics. The translators should continue to develop computational linguistics to solve difficult problems such as polysemy, preposition problems, and conjunction problems, so as to guide post-editing. Because of the development of computer technology, machine translation has made progress and improved the translator's translation efficiency. But machines still can't completely replace people. In order to ensure the accuracy and fluency of the translation, the translators also need to have subjective initiative and make post-editing. Machine translation will definitely get rapid development in the future. Everyone should maintain a rational and objective attitude towards machine translation and correctly handle the relationship between it and human translation.

REFERENCES

- Kunchukuttan, A., & Bhattacharyya, P. (2021). Machine Translation and Transliteration Involving Related and Low-resource Languages. CRC Press.
- Hutchins, J. (2003). Why computers do not translate better [J]. *Aslib Proceedings*, 44(10), 351-359.
- Zhang, J., Utiyama, M., Sumita, E., Neubig, G., & Nakamura, S. (2020). Improving neural machine translation through phrase-based soft forced decoding. *Machine Translation*, *34*(1), 21-39.
- Koponen, M., Mossop, B., Robert, I. S., & Scocchera, G. (Eds.). (2020). *Translation Revision and Post-editing: Industry Practices and Cognitive Processes*. Routledge.
- Koponen, M., Salmi, L., & Nikulin, M. (2019). A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation*, *33*(1), 61-90.
- Wilks, Y. (2008). Machine Translation: Its Scope and Limits [M]. New York: Springer Science & Business Media.
- Jia, Y., Carl, M., & Wang, X. (2019). Post-editing neural machine translation versus phrase-based machine translation for English–Chinese. *Machine Translation*, *33*(1), 9-29.
- 包树南. 基于深度学习的自然语言到机器语言的翻译解析研究[D]. 西安电子科技大学, 2020.
- 冯志伟. 我国计算语言学研究 70 年[J]. 语言教育, 2019, 7(04): 19-29+42.
- 耿立波, 酆格斐, 詹卫东, 杨亦鸣. 中国计算语言学研究现状与展望[J]. 语言科学, 2021, 20(05): 491-499.
- 刘伟伟. 人工智能的语义学基础[J]. 科学技术哲学研究. 2015 , 32 (3): 27-32.
- 王淼. 基于回复式神经网络的机器翻译技术研究及应用[D]. 电子科技大学, 2018.
- 徐英瑾. 心智、语言和机器: 维特根斯坦哲学和人工智能. 科学的对话[M].北京: 人民出版社, 2013: 75-76.
- 宋仕振. 试论机器翻译与人工翻译的未来关系[J]. 未来与发展, 2019, 43(02): 25-30.
- 邵南. 机器翻译错误实例分析及相应译前、译后编辑策略[D]. 北京第二外国语学院, 2020.
- 石赵佳. 机器翻译的错误类型分析与译后编辑策略[D]. 上海外国语大学, 2021.
- 赵小曼. 英汉平行语料库句子级对齐研究及其在机器翻译中的应用[D]. 安徽大学, 2010.
- 张政. 机器翻译难点所在[J]. 外语研究, 2005 (5): 59-62.

Supported by “Teaching Master Cultivation Project of North China Electric Power University”