ISSN 2616-5783 Vol.6, Issue 24: 98-103, DOI: 10.25236/AJHSS.2023.062417

Accuracy and Fluency of High-frequency COVID-19 Words Translated by Three Machine Translation(MT) Tools

Zhang Li, Chi Lixia, Yang Wenyi

Institute of Disaster Prevention, Sanhe, China, 065201

Abstract: *This study aims to investigate accuracy and fluency of COVID-19 words generated by the three MT tools, namely, Google Translate, Baidu Translate and DeepL Translator. Data analysis is done through human evaluation toward translation texts by using a translation rubric at the word, sentence and paragraph level. To ensure the credibility of translation texts analysis, ten raters are involved to analyse the translation texts at sentence and paragraph levels. It is found that the three machine translation tools can maintain a high accuracy rate in the translation of simple words and short sentences, but the collection of new words is relatively slow and the translation of longer sentences or paragraphs is not adequate and fluent enough. Further, there are language problems such as grammar and logical coherence.*

Keywords: *Accuracy; Fluency; COVID-19 high-frequency words; MT*

1. Introduction

In the wake of the Covid-19 outbreak, timely and accurate translation of public health information has become more critical than ever as it helps keep the public informed of the latest pandemic situation, personal hygiene advice and preventive measures, enabling speakers of different languages in the community to work together in the fight against the epidemic [7]. Previous studies had focused on MT in assisting health communication [3]. In order to meet the needs of the international community to combat COVID-19, linguists, translators and translation companies continue to develop the corpus of the corona crisis at the monolingual, bilingual and multilingual levels, such as TAUS and Systran, EMEA, Sketch Engine, English-Corpora and so on [6]. People around the world can easily get access to the MT tools now.

Despite MT's development during the pandemic, some scholars found that MT errors and omissions were still in words, sentences and paragraphs[1]. Other researchers emphasized interdisciplinary research [8]. The potential of machine translation as a valuable tool in improving health information access for linguistically diverse populations was highlighted [2]. Therefore, it is worthy of evaluating the translation texts of MT tools [4].

2. The Purpose of the Research

This study aims at comparing three Machine translations of high-frequency COVID-19 words generated by Google Translate, Baidu Translate and DeepL Translator through human evaluation. The research questions include:

①How accurate is the meaning of the high-frequency COVID-19 words in the target language at the word level generated by the three machine translation tools?

② How accurate and fluent are the high-frequency COVID-19 words generated by the three machine translation tools in the target language at the sentence level?

③ How accurate and fluent are the high-frequency COVID-19 words generated by the three machine translation tools in the target language at the paragraph level?

ISSN 2616-5783 Vol.6, Issue 24: 98-103, DOI: 10.25236/AJHSS.2023.062417

3. Materials and Methods

3.1 Accuracy criteria by Deni Nugraha and Ratna Dewanti

The word samples will be analyzed based on the accuracy criteria established by Deni Nugraha and Ratna Dewant[6]. The criteria consist of four levels (Table 1). In this paper, the assessment is based on the CNKI Dictionary, which is a powerful tool for students, teachers and translators at all levels in universities and research institutions. As can be noted in Table1, if the word scores 1, it is categorized as "Not recognized or disfluent". The word level is described as Non-Native Chinese if the word is not available in CNKI Dictionary, but it can still be understood under two reasons (loan, calques). The word is categorized as Good Chinese if the word is available in CNKI Dictionary and the words are still attributed to loanwords and calque. The word is categorized as Flawless if it is available in CNKI Dictionary and dismisses the two issues (loan, calque).

Table 1: Scoring Rubric Quality of the word-level

Score	Word availability	Quality
1	Not recognized	Disfluent Chinese
2	Not available in the dictionary	Non-native Chinese(loan, calque)
3	Available in dictionary	Good Chinese(loan, calque)
4	Available in dictionary	Flawless

3.2 Adequacy and fluency criteria by Koehn

The sentence and paragraph samples will be analyzed based on the Adequacy and Fluency translation evaluation criteria of these five levels (Table 2) [5]and the criteria descriptions are made based on the Assessment Specifications of Machine Translation Systems issued by the State Language Commission of China (Table 3).

Table 2: Adequacy and Fluency Criteria by Koehn

Score	Adequacy	Score	Fluency
1	None	1	Incomprehensible
2	Little meaning	2	Disfluent Chinese
3	Much meaning	3	Non-native Chinese
4	Most meaning	4	Good Chinese
5	All meaning	5	Flawless Chinese

Table 3: Criteria Descriptions

1	Incomprehensible	Only a few words in the translation correspond to the original text, making the translation obscure and difficult to understand.
2	Disfluent	The translation is partly faithful the original text, and the translation is not fluent.
3	Non-native	The translation basically expresses the information of the original text, and the translation is basically fluent.
4	Good	The translation conveys most of the information in the original text and is fluent but not authentic.
5	Flawless	The translation conveys the message of the original text accurately and completely and is fluent and authentic.

4. Data Collection

The word samples were obtained from Google Trends. We typed the two keywords "COVID-19" and "coronavirus" and then Google Trends showed the words frequently used by most people in the last 12 months, ending on February 27th, 2023. Based on the Google Trends' results, a total of 20 samples were retrieved, including words and phrases. Next, we used three machine translation tools to automatically generate translation texts, and then searched on the CNKI Dictionary with the Chinese translation of each word, and categorized and scored them according to the criteria of the word-level .

The sentence data were drawn from The Coronavirus Corpus of the English-Corpora. Twenty words were selected from the word section to search in this corpus. As these 20 words were relatively

ISSN 2616-5783 Vol.6, Issue 24: 98-103, DOI: 10.25236/AJHSS.2023.062417

common, the search yielded a very large number of results, which were automatically arranged in chronological order. We selected the sentences directly in order and extracted the sentences where the words were marked, finding a total of 6 sentences.

When assessing machine translation at the paragraph level, we still used the data of The Coronavirus Corpus to search for the same keyword in the corpus, which not only displayed the sentence, but also provided the source of the original sentence. We then visited the original web page with the help of link provided by the search results. Then, we selected the context of the sentence provided by the corpus as the paragraph samples. Finally, we find 3 paragraphs. With the help of the three machine tools, we got the translated texts.

As to the assessment of sentences and paragraphs, to ensure the credibility, we manually assessed the translation in the form of a questionnaire and invited 10 raters, who are all English teachers from the university, to rate the 27 MT translated texts. At the sentence level, 18 translated sentences were rated and 9 translated paragraphs were rated. The translation of the 27 sentences and 9 paragraphs were rated as a whole, instead of only part of the original texts rated. As of May 9, 2023, 10 rating results were received and all were valid.

5. Results and Discussion

5.1 Evaluation of the word level

At the word level. It's found that seventeen words are categorized as flawless (85%), two words are qualified as disfluent (10%), and only one word is classified as non-native (5%). The results indicate that the accuracy of Machine Translation in configuring a source text into a target text at the word level is not the issue. However, there are still some other issues. For instance, some translations aren't found in CNKI dictionary. Here is an example:

Reddit: This is a website where users of the site can browse and submit links to content on the Internet or post their original posts, and other users can vote, comment and act like an online community. All three machine translation tools transliterate it and Baidu Translate also explains that it is a website name, but the name of the website still cannot be found in CNKI Dictionary.

Then, there are also some translation errors. For instance,

Moderna: This word is shown in Google Trends as a company name and in Baidu as a biotechnology company based in Massachusetts, USA. Google, Baidu and DeepL, three machine translation tools have their different translated texts, but none of them has something to do with the name of the company.

5.2 Evaluation of the sentence level

The average score of 3.66 for all the 18 generated sentences by the ten raters suggests that when the three machine translation tools translated six sentences into Chinese, the translation of sentence 1 and 2 conveys much meaning in terms of adequacy and fluency; However, in terms of fluency, part of the translation of sentence 3,4 and 5 by the three tools is disfluent, non-native or even incompressible.

Here are some examples:

Table 4: Evaluation of The Sentence Level

Sentence 3	A recent study showed that Stockholm's congestion pricing cut air pollution by 5-15% and reduced asthma attacks, while also making it easier to move around the city.		
The average rating score by the 10 raters	Google Translate:3	Baidu Translate:3.6	DeepL Translator:4.5

Sentence 3 analysis: As can be seen in table 4, "making it easier to move around the city", Google and Baidu translate it as "making people to walk in the city easily", which is non-native Chinese. DeepL translates it as "The city enjoys convenient communications". It's clear that DeepL's translation is much more targeted to Chinese people with all meanings sent out.

ISSN 2616-5783 Vol.6, Issue 24: 98-103, DOI: 10.25236/AJHSS.2023.062417

Table 5: Evaluation of The Sentence Level

Sentence 4	The unidentified man, who is in his 60s, has been isolated in the same suburban hospital as his wife since Tuesday, when he began exhibiting symptoms consistent with the early stages of the virus, including fever, coughing and shortness of breath, officials said.		
The average rating score by the 10 raters	Google Translate: 3.9	Baidu Translate:3.8	DeepL Translator:2.9

Sentence 4 analysis: In table 5, "The unidentified man, who is in his 60s". In this sentence, the subject is "man" and "unidentified" and "in his 60's" are the two messages to characterize the man. Google Translate and Baidu Translate both take "the man" as the subject. Generally speaking, in Chinese, when multiple adjectives or messages modify a noun, the position of these adjectives or messages is determined by how close they are to the modified noun. The word order should be put like this: "articles, pronouns, possessives" + "characteristic" + "size, height, shape" + "age" +" color" + "nationality" + "material". As can be noted, "unidentified" is the feature of "the man", the message of "age 60" should be put behind that of "unidentified". Google Translate rendered it flawless and more native Chinese. However, DeepL stresses the meaning that "the identified man is over 60 years old", which not only carries little meaning of the original sentence, but also reads disfluently.

Table 6: Evaluation of The Sentence Level

Sentence 5	"Patients who have a negative rapid antigen test for influenza, but in whom the clinical suspicion for influenza infection is high should be treated with antivirals since the sensitivity of these tests may be low," he said.		
The average rating score by the 10 raters	Google Translate: 4.1	Baidu Translate:2.9	DeepL Translator:4.1

Sentence 5 analysis: As can be noted in table 6, "In whom the clinical suspicion for influenza infection is high", The original sentence means patients who are highly suspected of being infected with influenza in clinical practice. The translations of these three machine tools are different, and the texts generated from Google and DeepL are comprehensible and native with slight differences. The translation from Baidu is a bit disfluent and even confusing for Chinese when both messages of "the clinical" and "high" are used to modify the word "suspicion". The word order is changed a little bit but it's much more disfluent.

5.3 Evaluation of the paragraph level

The average score of the three paragraphs is 3.6, which means that in terms of adequacy, the translated texts expressed much meaning; In terms of fluency, the translation could be considered non-native Chinese. The results show that machine translation still needs to improve the adequacy and fluency of translation, especially for long sentences and paragraphs.

Among all the 3 paragraphs, some problems are found in the three passages, and it can be assumed that the more words there are in the paragraph, the more problems are likely to arise. Take Passage 3 as an example:

Table 7: Evaluation of The Paragraph Level

Passage 3	Three-quarters of people with seasonal and pandemic flu have no symptoms. Around 1 in 5 of the population were infected in both recent outbreaks of seasonal flu and the 2009 H1N1 influenza pandemic, but just 23% of these infections caused symptoms, and only 17% of people were ill enough to consult their doctor. These findings come from a major new community-based study comparing the burden and severity of seasonal and pandemic influenza in England over 5 years, published in The Lancet Respiratory Medicine journal. "Reported cases of influenza represent the tip of a large clinical and subclinical iceberg that is mainly invisible to national surveillance systems that only record cases seeking medical attention", explains lead author Dr Andrew Hayward from University College London, UK. " Most people don't go to the doctor when they have flu. Even when they do consult, they are often not recognized as having influenza.		
The average rating score by the 10 raters	Google Translate: 3.8	Baidu Translate:3.6	DeepL Translator:3.5

Passage 3 analysis: As showed in table 7, the three MT translated texts can be seen as somewhere between "non-fluent" or "good", and the three MT generated texts convey much meaning of the

ISSN 2616-5783 Vol.6, Issue 24: 98-103, DOI: 10.25236/AJHSS.2023.062417

original one. However, some problems still exist. Here are some examples:

As to the translation of "Dr Andrew Hayward", the problems arise. "Andrew Hayward", When translating personal names, Baidu adopts the original English text. Google and DeepL transliterate it, although the Chinese characters are the same, there is an error of using "-": DeepL Translator misuses "-" between his last name and first name, while Google Translate doesn't put any mark like "a dot" between his last name and first name which is in line with Chinese expressions. What's more, as to the translation of "lead author", the three generated texts are comprehensible with all the meanings rendered. The word "lead" in Chinese has different meanings, for instance, it means "a position of leadership", "an advantage held by a competitor in a race" or "an actor who plays a principal role". Clearly, here "lead author" means "first author" who plays leading role in conducting the research and/or writing the paper. However, when the two messages of "lead author" and "from University College London" are used to modify "Dr. Andrew Hayward", the translated texts by DeepL seem to be fluent, but more like English expressions instead of native Chinese.

"The Lancet respiratory medicine", "The Lancet" is currently one of the most authoritative academic journals in the world's medical community. The official Chinese translation of "The Lancet" is "Liu Ye Dao", and "The Lancet Respiratory Medicine" is one of "The Lancet" series. All three translation tools translate it as "Liu Ye Dao Hu Xi Yi Xue Za Zhi", but the position of the book title mark in Chinese is different. Google and Baidu treat "The Lancet Respiratory Medicine" as a whole while DeepL Translator only renders "The Lancet" as a magazine title, which isn't faithful to the original text. Therefore, in this translation, DeepL's text can be described as "disfluent" with little meaning.

6. Conclusion

This study described the quality of translated products by the three machine translation tools, which is expected to pave the way for the research and development of machine translation. This study also investigated some new technical terms. It is presumed that Google Translate, Baidu Translate and DeepL Translator could be regarded as alternative ways of translating Covid-19 related terms. However, there are some problems arising at the sentence and paragraph levels such as mistranslated people's names, disfluent or even incomprehensible translation etc. It suggests that each machine tool has its weaknesses and strengths. In other words, machine translation still needs to be improved, and post-editing work should be equally important.

In the end, three suggestions are made: (1) The incoming new terms should be collected in time and their translation should be standardized, which echoes Nugraha& Dewanti(2022)[6] that the government needs to release a policy to support the authority together with linguists in developing a word-sized database; (2) The translation of proper names like personal names, names of places and names of organizations should be standardized;(3) Developers of translation software can establish a user feedback platform. When users use the tool and the quality of the generated translation products is assessed, they can provide feedback based on their experience, and the platform can collect the feedback for further improvements of the tool.

Acknowledgements

This study is supported by the Institute of Disaster Prevention through the project of Teaching Research and Reform (No. JY2022B26) and the project of the New Liberal Arts Research and Reform by the Ministry of Education, China. (No.2021050023)

References

[1] Chon, Y.V., Shin D. and Kim G. E. (2021). Comparing L2 Learners' Writing against Parallel Machine-Translated Texts: Raters' Assessment, Linguistic Complexity and Errors. System, 96.
[2] Dahal,S. B., Auon, M.(2023). Exploring the Role of Machine Translation in Improving Health Information Access for Linguistically Diverse Populations, Journal of Intelligent Information Systems, 8(2):4-6.
[3] Dreisbach, J. L. and Mendoza-Dreisbach, S. (2021). Unity in Adversity: Multilingual Crisis Translation and Emergency Linguistics in the COVID-19 Pandemic, The Open Public Health Journal, 14, 1, 94–97.

ISSN 2616-5783 Vol.6, Issue 24: 98-103, DOI: 10.25236/AJHSS.2023.062417

[4] Koehn, P. & Knowles, R. (2017). Six Challenges for Neural Machine Translation[J]. Proceedings of the First Workshop on Neural Machine Translation, 28-39.
[5] Koehn, P. (2010). Statistical Machine Translation[M]. New York, the United States of America by Cambridge University Press, 217-235.
[6] Nugraha, D. S. and Dewanti,R. (2022). English-Indonesian crisis translation: accuracy and adequacy of Covid-19 terms translated by three MT tools, Theoretical and Applied Linguistics, 8 (1), 122-134.
[7] Siu, S.C.(2023). COVID-19 MT Evaluator:A Platform for the Evaluation of Machine Translation of Public Health Information Related to COVID-19, published in Translation and Interpreting in the Age of COVID-19, edited by Liu, K., Cheung, A.K.F., Springer.
[8] Vieira, L. N., O'Hagan, M. and O'Sullivan, C. (2020). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases, Information, Communication and Society, 24, 11, 1515-1532.