LUBLIN STUDIES IN MODERN LANGUAGES AND LITERATURE Vol 48, No 3 (2024)
Maria Curie-Sklodowska University Press

E-ISSN 2450-4580

Tawei Wang, National Taipei University, Taiwan

DOI:10.17951/lsmll.2024.48.3.41-51

What Can We Learn from the 21st Century Translator Pen? A Case Analysis on Comparing English-to-Chinese Rendition Differences between Human and Machine Translation

ABSTRACT
This study investigated and compared differences in the rendering from the same English source text into two versions of the target text in Chinese produced by a human and a machine translator represented by a translator pen, respectively. Using news reports as case analysis, this study found that improper segmentation of punctuation marks appeared most frequently in machine translation, followed by lexical vacancy, and inconsistency of terms. This study also identified rendition differences between human and machine translation in the handling of terms as well as in the treatment of punctuation marks. Overall, the human translator showed more flexibility in the selection of words to match the target text expression than the machine translator.

KEYWORDS
machine translation; translator pen; source text; target text

1. Introduction

Technology has become an integral part of our daily lives entering the 21st century. It has fueled the development of human civilization in many aspects such as education, communication, entertainment, socializing, and work. Among them, the one that can be said to be influenced greatly by technology is the translation community where the role of translators is being refined or even partially replaced as we speak.

Over the past seventy years, particularly since the outbreak of the Second World War, research on machine translation (MT) has yielded fruitful results, being seen as a research discipline highly relevant to artificial intelligence (AI) and natural language processing (NLP). And, after a long evolution, translation with the aid of computer has been bettered gradually and become a comparatively developed field so far that is still being debated and explored by many scholars in the field of translation (Bahar, 2001; Celik, 2003; Furstenberg et al., 2001).

Tawei Wang, Department of Foreign Languages and Applied Linguistics, National Taipei University, No. 151, Daxue Rd., Sanxia Dist., New Taipei City 237, tawwang@mail.ntpu.edu.tw, https://orcid.org/0009-0003-2136-1550

The journal is financed by the Institute of Modern Languages and Literatures of Maria Curie-Sklodowska University.

However, the scope of existing studies on MT may still be limited when it comes to providing a systematic comparison and description of how human translation can be different from MT, on the condition that rendition is perceived as the only end product of the translation process.

This study therefore aims at addressing this limitation by adopting a text-analysis approach to identifying the rendition differences between human translation and MT represented by the translator pen. With the translator pen, the translation process is time-saving and entirely automatic, i.e., without human intervention such as post-editing. The only thing the user needs to do with the translator pen is simply scan through the source text (ST) verbatim and wait for the target text (TT) to be outputted either in the screen of the pen or in a personal computer document with the internet or Bluetooth connection on.

2. A brief review of terminology and the development of MT

MT in general refers to "computerized systems responsible for the production of translations with or without human assistance" (Hutchins, 1995, p. 431). The present study holds that MT is entirely automatic and responsible for the translations it produces without any human assistance and therefore "excludes computer-based translation tools which support translators by providing access to on-line dictionaries, remote terminology databanks, transmission and reception of texts, etc." (p. 431). Considering the notion, other terms related to MT such as machine-aided human translation (MAHT) and human-aided machine translation (HAMT) are not applicable in the present study as the core of MT should be "the automation of the full translation process" (p. 431). Although in common practice, the output of MT is usually post-edited either by human translators (e.g. the first and the second translator) or proofreaders, the ultimate and ideal goal of MT is to generate up-to-standard end product by rendering quality translation like a certified human translator.

MT has witnessed a long-winded development process, which can be divided into four periods: the sprouting period (1949–1960), the setback period (1960–1967), the recovery period (1967–1990), and the new period (1990-present) (Gao & Zhao, 2020, pp. 97–98) where the focal point of the present study lies. Although translation quality of MT has increased over the years, it would still be stretching to say that its quality is already comparable to that of quality human translation entering the 21st century (Qin & Xiang, 2022, p. 44). Some problems in its original output are unavoidable, thus making translation quality unsatisfactory.

2.1 Common problems facing MT

These common problems generally refer to 1) *inconsistency of terms*, 2) *improper segmentation of punctuation marks*, 3) *redundancy*, and 4) *lexical vacancy* (Qin & Xiang, 2022, p. 45). In the first case, it means "one term of the source language has different expressions, but the multiple expressions in the source text for the same

thing are translated into different versions in the target text by machine" (p. 45). Such difficulty for machine to analyze Chinese language accurately lies in that "the same part of speech in Chinese serves as grammatical components without morphological changes" (Guo & Wang, 2017, p. 78). Compared to its human counterpart, MT is specifically vulnerable to term inconsistency when it needs to process large texts where different colocations of the same term could appear frequently.

In the second case, a type of problem rooted in the MT punctuation system, "the punctuation marks used in Chinese are formulated based on the English punctuation system" (Qin & Xiang, 2022, p. 45). This contributes to analytical problem on the part of MT to convert the punctuation marks accurately between two languages. In others words, MT will copy them into the TT, giving rise to some translation problems.

In the third case, "redundancy refers to the functional repetition, overlapping or redundant expressions in the translation" (Cui & Li, 2015, p. 21). Since redundancy is a typical feature of the Chinese language expression, a common example of it would be "synonym with different words in the form of four-character words" (Qin & Xiang, 2022, p. 45). This feature is, nevertheless, opposite to that of the English language expression where repetition is usually avoided and replaced with pronouns and prepositions to substitute the repeated speech part.

In the last case, it "refers to the difficulty in achieving complete equivalence between the source language and the target language, resulting in lexical vacancy in translation" (Qin & Xiang, 2022, p. 45). This problem is caused primarily by cultural differences between the two languages and can be commonly observed in translating "culturally-loaded words" (p. 46). At present, as far as MT is concerned, it is not able to detect and interpret entirely accurately the precise meaning of terms rich in cultural connotation. Therefore, if lexical vacancy cannot be addressed by MT, the translation quality will for sure be compromised and will not be improved in a short period of time.

Even though all the above-mentioned problems can be solved with post-editing in the form of human intervention by using context-specific translation strategies such as replacement, omission, addition, or shift, it is not possible with the sole use of MT. Therefore, the four common problems facing MT will also serve as critical parameters for probing rendition differences between human and MT.

To identify and compare differences in rendering from the same ST produced between a human translator and the translator pen, the present study therefore proposed the following research questions for investigation: 1) Among the four common problems facing MT, how are they placed in terms of occurrence in the TT by translator pen? 2) Using the four common problems facing MT as parameters, what are the rendition differences observed in the present study? 3) Do these differences include more than the four common problems facing MT?

3. Research methodology

To answer the three research questions proposed, this study adopts a qualitative approach in the form of text analysis to compare and analyze the two renditions produced by a human translator and the translator pen separately based on the same selected ST.

3.1 Research design

The present study selected an English news report on *How the Coronavirus Steals the Sense of Smell* excerpted from New York Times as the ST to be translated by a news translator and the translator pen for text analysis. In the analysis, the four common problems facing MT will serve as parameters for identifying rendition differences between human and MT.

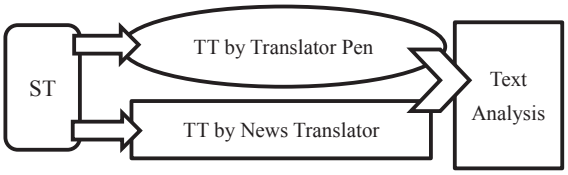

Figure 1: Research design

3.2 Instruments and data collection

In the present study, a ST, two versions of TT, and a translator pen were employed as the instruments. For the ST, it was an English news report of 362 English words excerpted from New York Times. The ST was used for outputting two versions of TT. For the two versions of TT, one was produced by a news translator, Li, a full-time UDN[1] journalist and translator who has translated more than 4,800 articles in the business to contain 668 Chinese characters. The other one was produced by the Muigic[2] translator pen, which is able to perform text scanning translation in English-Chinese language combination to contain 640 Chinese characters. Only the two versions of TT were collected for text analysis.

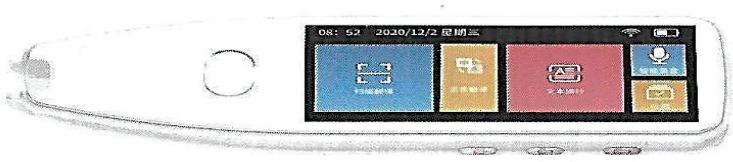

Figure 2: Translator pen

[1] It is a Taiwan-based online open access news media established in 1999.
[2] It is a brand that features intelligent appliances.

3.3 Data analysis

The analytical data process focused primarily on comparing the two versions of TT based on the four parameters mentioned earlier: inconsistency of terms, improper segmentation of punctuation marks, redundancy, and lexical vacancy to probe rendition differences. These differences were then presented based on categories (i.e. which parameter) and occurrences in the two renditions to address the research questions.

4. Findings

Through text analysis, the present study found that out of the four common problems facing MT, three of them appeared in the TT by translator pen. They were inconsistency of terms, improper segmentation of punctuation marks, and lexical vacancy. Redundancy was not observed.

Precisely, in the TT by translator pen, improper segmentation of punctuation marks registered more than nine occurrences, followed by lexical vacancy to register five occurrences, and by inconsistency of terms to register one occurrence as indicated in Figure 3.

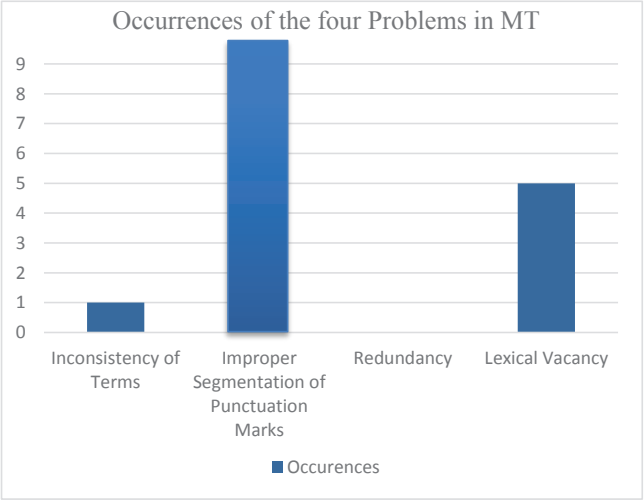

Figure 3: Occurrences of the four problems in MT

Using the four common problems facing MT as parameters for investigation, rendition differences between human translator and MT were also present in the three parameters mentioned earlier, namely, the inconsistency of terms, lexical vacancy, and the improper segmentation of punctuation marks. As shown in Table 1 where referenced Chinese translations (hence RCT) were provided for individual ST terms, the term COVID appeared four times in total and was consistently

translated in the TT by news translator either as 新冠肺炎 or as 新冠 (i.e. a shorter form for 新冠肺炎) to refer to the disease. In the case of TT by translator pen, out of the four appearances of COVID, three of them were lexically vacant (i.e. not translated) and only one of them was translated as 多科疾病, which did not suggest any propositional meaning to relate to the disease in Chinese. This would therefore be counted as one inconsistent handling of the term COVID on the part of translator pen. Another difference was spotted in the handling of the word *indirectly* in the ST – it was translated by news translator as 間接 in Chinese but was lexically vacant in the TT by translator pen. A difference was also observed in the handling of the verb *line* in the ST – it was translated by news translator as 內側 to refer to the inner side of the nasal cavity in Chinese but was lexically vacant in the TT by translator pen.

Table 1. Rendition differences in the handling of terms

ST Terms	**News Translator**	**Translator Pen**
COVID RCT: 新冠肺炎	新冠肺炎/ 新冠	1. X 2.多科疾病
Indirectly RCT: 間接	間接	X
that **line** the nasal cavity RCT: (貼著)內側	內側	X

Apart from differences in the handling of terms and lexical vacancy mentioned above, differences in the improper segmentation of punctuation marks were spotted between the two versions of TT as shown in Table 2.

The first difference in this category was present in converting the quotation mark in the ST. It appeared four times in the ST in total and was converted into Chinese corner brackets consistently in the TT by news translator. In the case of TT by translator pen, it remained unchanged. It should be noted that in the TT, the Chinese text, English quotation mark does not exist in the writing system.

The second difference in this category was embodied in the conversion of a colon, which appeared twice in total in the ST. For its first appearance in the ST, both TT versions treated it the same way into a Chinese colon. Yet, for its second appearance, this mark was converted into a Chinese full stop by news translator but still a Chinese colon by translator pen.

The third difference in this category lay in the treatment of a scholarly title- *Dr. Sandeep Robert Datta* in the ST. In the TT by news translator, it was translated as 塔達博士 with no punctuation marks added in the form of a Chinese last name

followed by how it is called to address a doctoral degree holder, which is a common combination in the Chinese expression without having to address a person's first and middle name. In the case of TT by translator pen, however, two hyphenation points were added to distinguish the first and the middle name. This may therefore seem awkward in the Chinese text although such treatment is context-specific.

The fourth difference was observed in the handling of a semicolon in the ST. In the TT by news translator, it was converted into a Chinese full stop whereas in the TT by translator pen a Chinese semicolon. The last difference lay in the treating of four commas in the ST. In the TT by news translator, they were converted sequentially into a Chinese full stop, a Chinese comma, a Chinese colon, and a Chinese full stop whereas in the TT by translator pen a Chinese comma, a Chinese ideographic comma, a Chinese comma, and a Chinese comma.

Table 2. Rendition differences in improper segmentation of punctuation marks

ST	News Translator	Translator Pen
“ ”	「 」	“ ”
:	。	：
Dr. Sandeep Robert Datta	X	• •
;	。	；
,	。 , ： 。	， 、 ， ，

5. Discussion and conclusion

To answer the first research question, the present study discovered that improper segmentation of punctuation marks occurred most frequently in the TT by translator pen to record over nine occurrences, followed by lexical vacancy to record five occurrences and inconsistency of terms to record one occurrence. MT is still vulnerable to three of the four common problems.

To answer the second research question, the present study found that rendition differences between the two versions of TT were identified in terms of the handling of terms (lexical vacancy included) and in converting punctuation marks.

To answer the third research question, the present study found that two primary differences not listed in the four parameters were observed. The first one is the treating of word order. It was observed that compared to its human counterpart, translator pen in most cases followed the word order of the ST for conversion. This would easily lead to a situation in the TT where the agent, the doer of an

action, was not conceptually identical to the one in the ST and thus hinder TT's readability and affect the reader's understanding. The second primary difference not listed in the four parameters is word choice. It was observed that compared to the machine counterpart, human translator was more flexible in word selection to fit a specific contextual expression. Overall, the above-mentioned details were the additional differences identified between the two versions of TT.

6. Limitations of the study

The presentation of the current findings is limited by certain constraints regarding the selection and use of the research instrument. The study was mainly a qualitative and a case-based text analysis, which may not be able to contribute much to its generalizability and reliability. A possible direction for future studies could be the employment of a mixed methods approach encompassing the combination of quantitative and qualitative tools to encourage a cross-verification of the results, and thus giving potential researchers more room to look at the differences between human and MT.

Finally, it has to be pointed out that the scope of the current research is also limited due to the length of the ST selected and the number of words analyzed in the two versions of TT. The analyzed sample may not be representative enough given the many types of text that can be used for analysis. It would be advisable to investigate more texts from different genres for future studies, other than news report. Despite these limitations, the results of this study may open up possibilities and hopefully attract the attention and interests for future research.

References

Bahar, I. B. (2001). Linguistic relativity and the translation dilemma: Reading between the lines in Malay literatures in English. *Nottingham Linguistic Circular*, *16*, 19–29.

Celik, M. (2003). *Learning stress and intonation in English*. Ankara Gazi.

Cui, Q., & Li, W. (2015). The character of error types of post-editing: Perspective of machine translation based on scientific and technology materials. *Chinese Science & Technology Translators Journal, 28*(4), 19–22.

Furstenberg, G., Levet, S., English, K., & Maillet, K. (2001). Giving a virtual voice to the silent language of culture: The "Cultura" Project. *Language Learning & Technology*, *5*(1), 55–102.

Gao, L., & Zhao, W. (2020). An overall study on machine translation. *Foreign Languages in China*, *17*(6), 97–103.

Guo, G., & Wang, Z. (2017). Research on the pre-edit and post-edit of machine translation in science and technology text translation. *Journal of Zhejiang International Studies University*, *3*, 76–83.

Hutchins, W. (1995). Machine translation: A brief history. In E. Koerner, & R. Asher (Eds.), *Concise history of the language sciences*: *From the Sumerians to the cognitivists* (pp. 431–445). Pergamon. https://doi.org/10.1016/B978-0-08-042580-1.50001-5

Qin, F., & Xiang, X. (2022). On the translation strategies of human-computer interaction based on machine translation. *Education, Language and Sociology Research*, *3*(4), 43–50. https://doi.org/10.22158/elsr.v3n4p43

譯／李京倫
新冠病毒如何偷走嗅覺

Few of COVID-19's peculiarities have piqued as much interest as anosmia, the abrupt loss of smell that has become a well-known hallmark of the disease. COVID patients lose this sense even without a stuffy

Appendix A

Source Text
2022/04/08 第377期 New York Times
How the Coronavirus Steals the Sense of Smell 新冠病毒如何偷走嗅覺
文／Roni Caryn Rabin
nose; the loss can make food taste like cardboard and coffee smell noxious, occasionally persisting after other symptoms have resolved.

Scientists are now beginning to unravel the biological mechanisms, which have been something of a mystery: The neurons that detect odors lack the receptors that the coronavirus uses to enter cells, prompting a long debate about whether they can be infected at all.

Insights gleaned from new research could shed new light on how the coronavirus might affect other types of brain cells, leading to conditions like "brain fog," and possibly help explain the biological mechanisms behind long COVID — symptoms that linger for weeks or months after the initial infection.

The new work, along with earlier studies, settles the debate over whether the coronavirus infects the nerve cells that detect odors: It does not. But the virus does attack other supporting cells that line the nasal cavity, the researchers found.

The infected cells shed virus and die, while immune cells flood the region to fight the virus. The subsequent inflammation wreaks havoc on smell receptors, proteins on the surface of the nerve cells in the nose that detect and transmit information about odors.

The process alters the sophisticated organization of genes in those neurons, essentially short-circuiting them, the researchers reported.

Their paper significantly advances the understanding of how cells critical to the sense of smell are affected by the virus, despite the fact that they are not directly infected, said Dr. Sandeep Robert Datta, an associate professor of neurobiology at Harvard Medical School, who was not involved in the study.

"It's clear that indirectly, if you affect the support cells in the nose, lots of bad things happen," Datta said. "The inflammation in the adjacent cells triggers changes in the sensory neurons that prevent them from working properly."

Indeed, many complications of COVID appear to be caused by the immune system's friendly fire as it responds to infection by flooding the bloodstream with inflammatory proteins called cytokines.

Appendix B

Target Text by News Translator Li

少有新冠肺炎的特點像嗅覺喪失一樣激起那麼多關注。嗅覺喪失是突然失去嗅覺，已成為這種疾病眾所周知的特徵。新冠肺炎患者甚至沒有經歷鼻塞就失去嗅覺。失去嗅覺會讓食物嘗起來像硬紙板，咖啡氣味難聞，這種症狀偶爾會在其他症狀消退後持續。

科學家現在開始弄懂這個向來可說是個謎的生物機制：感知氣味的神經元並無受體供新冠病毒用來進入細胞，引發關於這些神經元究竟能否被感染的長期爭論。

從新近研究收集來的洞見或許能進一步闡明，新冠病毒如何侵襲其他種類的腦細胞，導致「腦霧」等症狀出現，而且或許能解釋新冠長期症狀的生物機轉。新冠長期症狀是在最初感染後持續數周或數月的症狀。

除了稍早的研究之外，新研究也解決了關於新冠病毒會不會侵擾察覺氣味的神經細胞爭論。答案是不會。不過，研究人員發現，新冠病毒攻擊的是位在鼻腔內側的其他支持細胞。

被感染的細胞擺脫病毒後死亡，同時免疫細胞蜂擁到這個區域對抗病毒。隨後的發炎嚴重破壞嗅覺受體，即鼻內神經細胞表面能察覺並傳達氣味資訊的蛋白質。

研究人員說，這個過程改變了這些神經元基因的複雜組織，實質上使神經元基因短路。

並未參與這分研究的美國哈佛大學醫學院神經生物學副教授達塔博士說，他們的論文大幅增進了這方面的了解：對嗅覺至關重要的細胞儘管不會被新冠病毒直接感染，卻會被新冠病毒侵襲。

達塔說：「顯然，如果你攻擊鼻子的支持細胞，有很多壞事會間接發生。鄰近細胞發炎會引起感覺神經元改變，使感覺神經元無法正常工作。」

的確，許多新冠肺炎併發症似乎由免疫系統對自己人的攻擊引起，因為免疫系統應付感染的方式是讓血液中充滿名為細胞激素的發炎蛋白質。

Appendix C

Target Text by Translator Pen

19的特異性與厭食的興趣很少，嗅覺的突然喪失已經成為這個疾病的一個顯著特徵，很少有人會這樣做。

患者即使沒有鼻塞也會失去這種感覺；這種損失會使食物的味道像紙板和咖啡的氣味一樣有毒，偶爾會堅持等其他癥狀后已解決。

科學家們現在開始解開生物機制，這是一個謎：檢測氣味的神經元缺少了冠狀病毒用於進入細胞的受體，這促使人們對它們是否能被完全感染進行了長期的爭論。

從新的研究收集的洞察力可以揭示，冠狀病毒如何可能影響其他類型的腦細胞，導致諸如 “腦霧”的條件，並可能有助於解釋的生物學機制背後長長的癥狀，持續幾個星期或幾個月後，最初感染。

這項新的工作，連同早期的研究，解決了關於冠狀病毒是否感染了檢測氣味的神經細胞的爭論：它沒有。研究人員發現，這種病毒確實攻擊了鼻腔的其他支持細胞。
被感染的細胞會使病毒和死亡，而免疫細胞在該地區氾濫，以對抗病毒。隨後的炎症對嗅覺受體、鼻子上的神經細胞表面的蛋白質進行了破壞，可以檢測和傳播有關氣味的信息。

研究人員報道，這一過程改變了這些神經元中複雜的基因組織，基本上是對它們的短路。

哈佛醫學院的神經生物學副教授桑迪普・羅伯特・達塔達說・他們的論文大大地了解了病毒對嗅覺影響的細胞是如何影響的，儘管事實上它們沒有直接感染。

“很明顯，如果你影響鼻子上的支撐細胞，就會產生很多不好的東西，”達塔說。相鄰細胞中的炎症會觸發感覺神經元的改變，防止它們正常工作。

事實上，多科疾病的併發症似乎是由免疫系統的友好火災，因為它回應感染的血液與炎症蛋白稱為細胞因子。