Identifying MT Errors for Higher-Quality Target Language Writing

Kayo Tsuji, Osaka Metropolitan University, Japan*

ABSTRACT

Second language education has arrived at a phase of proposing effective uses of neural machine translation (NMT). Previous research has explored various aspects of post-editing and suggested that it is crucial to manually edit NMT output to produce better target language (TL) texts. The purpose of this study was to identify NMT errors in output text, so that Japanese TL (English) learners can recognize what to be aware of. The study targeted the NMT output from Japanese-written academic reports, pre-edited by 73 Japanese students with intermediate TL proficiency. The data was analysed and primarily lexical and grammatical issues were detected and systematically classified. Results showed that the use of inappropriate TL vocabulary was the most frequent error, followed by misuse or lack of determiners. Some could be avoided in a pre-editing phase by carefully choosing precise source-language (SL) vocabulary or reducing SL ambiguity, while others required a deeper understanding of TL syntactic rules or the nuance of TL vocabulary. TL Learners need to raise their awareness of these NMT errors for effective post-editing.

KEYWORDS

Developing neural machine translation (NMT) has had a significant impact on second language education and research (e.g., Chung, 2020; Klimova et al., 2023; Lee, 2023), and its effective use has been the focus of much attention. Previous studies have reported on the value of pre-editing (Farhana et al., 2023; Kokanova et al., 2022; Liang & Han, 2022; Marzouk & Schirra, 2019), the effectiveness of pre-editing rules (Hiraoka & Yamada, 2019; Tsuji & Okamoto, 2022; Zheng et al., 2022), the necessity of post-editing (Lee, 2020; Niño, 2009; Udina, 2019), and the efficacy of post-editing strategies (Shih, 2021; Shin & Chon, 2023).

The most effective way to use MT includes both pre- and post-editing (Cheng et al., 2021; Zheng et al., 2022). Although pre-editing contributes to improving the quality of target-texts (TTs), it is indispensable to manually edit MT output to further enhance TT quality (Kuraya, 2019; Yamada, 2021). Considering that post-editing is a challenging task for learners with intermediate or lower target-language (TL) ability (Chung, 2020; Klimova et al., 2022; Lee, 2022), typical MT errors in the Japanese-English MT output should be identified for effective post-editing. To that end, the present study discusses what errors MTs tend to make and which elements TL learners should be aware of.

PREVIOUS RESEARCH

Post-editing[1] has received attention from a variety of perspectives. Researchers have identified the merits of post-editing (Alsalem, 2019; Escartín et al., 2017; Fredholm, 2019), explored the post-editing process (Chung, 2020; Jia et al., 2019; Shin & Chon, 2023), along with user experiences of post-editing (Harto et al., 2022). They have also analysed how TL proficiency affects the post-editing process (Chung, 2020; Shin & Chon, 2023) and how the process of post-editing is different between native and non-native TL speakers (Báez, 2018; Sánchez-Gijón & Torres-Hostench, 2014). In relation to education, research has explored whether MT can be used as a pedagogical support tool (Lee & Briggs, 2020; Tsuji & Okamoto, 2022), how post-edit training influences TT quality (Báez, 2018; Zhang & Torres-Hostench, 2022), as well as what MT errors are output in order to develop post-editing strategies (Shih, 2021; Shin & Chon, 2023).

Influential Elements on Post-editing Quality

The quality of post-editing can be influenced by learner's level of TL proficiency and the degree of the users' genre/topic familiarity. Previous studies have investigated the relationship between learners' TL proficiency and the post-editing quality. A certain TL proficiency is required for post-editing MT output (Escartín et al., 2017; Ishikawa, 2020; Kuraya, 2019). Some research, involving advanced-level TL students (e.g., Lee, 2022; Niño, 2009), reported that post-editing improved the quality of TTs, while other research, with learners of lower TL proficiency (e.g., Chung, 2020; Klimova et al., 2022), identified learner difficulties in detecting and correcting MT errors. For instance, Chung (2020) investigated how TL ability affects the post-editing process and concluded that students with intermediate or lower TL proficiency tended to focus only on the word-level correction, while advanced learners reinterpreted the content of TTs and then edited parts inconsistent with source texts (STs). Also, Shin and Chon's 2023 study revealed that TL proficiency affects learners' success with post-editing strategies: Learners with higher L2 proficiency employed a greater variety of post-editing strategies. Both skilled and less skilled learners tended to employ word- and phrase-level deletion/ paraphrase, however skilled learners used these strategies more frequently. Such post-editing requires a greater cognitive load on the part of students with relatively lower TL ability, with the result being that the content of TTs cannot be reconsidered with a wider perspective.[2] Similarly, Lee and Briggs (2020) also reported that learners' TL proficiency affects the number and type of error corrections made by students. Learners with advanced TL proficiency can post-edit TTs which are generally close in quality to those of native TL speakers (Sánchez-Gijón & Torres-Hostench, 2014). Not only the learner's TL ability but also the genre of STs factors into the effectiveness of post-editing. When TL learners do not have sufficient knowledge on the specific genre/topic, understanding the ST content is difficult (Harto et al., 2022). Choosing accurate TL vocabulary along with the particular context of an unfamiliar genre/topic is not easy. However, based upon the results of Harto's 2022 study, there are some advantages of MT use: Post-editing practices contribute to enhancing learners' TL vocabulary in the unfamiliar genre.

MT Use in Educational Settings

MT use can facilitate TL learners to produce higher-quality TL writing. In Lee's 2020 study, Korean learners with intermediate TL (English) proficiency (70-95 on the TOEFL iBT) were asked to compare their own translations with MT outputs in order to revise the former. Through the activity, the participants learned new TL vocabulary and expressions relevant to the ST context. Statistical analysis showed a decrease in lexical and grammatical errors in their revisions, thus MT use facilitated learning of new perspectives on the TL. In a similar vein, Lee and Briggs' 2020 study, involving 58 Korean university students, investigated whether MT can be an indicator for error analysis when translating Korean (STs) into English (TTs). The learners corrected errors in their own translations by comparing them with the TL vocabulary and expressions used in the MT output. Results showed that

the number of errors was significantly reduced in the revised TTs, thus MT served a feedback function to indicate errors in learners' own translations. Furthermore, MT use may lessen the disadvantages faced by non-native TL writers (Niño, 2009). For users who can correct grammatical errors and mistranslations through post-editing, the study of Escartín et al. (2017) concludes that MT can be effectively used as a writing assistant to produce TL-written research papers. Even for learners with relatively lower TL proficiency, Kliffer's 2005 study claimed positive perceptions of MT use as it reduces the translation workload on the part of TL learners.

Considering the aforesaid findings, post-editing can be of benefit in generating higher quality TTs for learners with intermediate or higher-level TL proficiency. Users need to raise their awareness of MT errors (Funakoshi et al., 2004), however, it is not always easy for TL learners to detect and correct them (Escartín et al., 2017; Shin & Chon, 2023; Vidal, 2021), even for students with higher TL proficiency. Harto et al. (2022), targeting Indonesian-English MT output, reported that the most significant problems students encountered in their MT post-editing (MTPE) practices are related to grammar and vocabulary. This difficulty is even greater for learners of lower TL ability (Klimova et al., 2022; Lee, 2022). In view of these difficulties, training is required for effective MTPE (Yamada, 2019; Zhang & Torres-Hostench, 2022), which may lead to further improvements in TT quality (Escartín et al., 2017). Recent research supports the importance of training sessions in educational settings: Some revealed the positive effects of post-edit training on learners' ability to correct MT errors (Samman, 2022; Yoon & Chon, 2022), others reported its positive influence on the language skills required for post-editing (Emara, 2023; Zhang & Torres-Hostench, 2022).

MT Error Analysis and Strategies for Effective MT Use

MT error analysis is vital to suggesting what elements TL learners should be aware of for effective MTPE. Research offers valuable insights into such analysis. Koo et al. (2022) identify the error types in a data platform including Korean-English MT output, and systematically classify them into four broad categories: spacing, punctuation, numerical, and spelling and grammatical errors. This study developed the evaluating criteria for Korean grammatical error correction. Similarly, Daem et al. (2017) labeled MT errors identified in English-Dutch MT output into six items: adequacy, grammar and syntax, coherence, lexicon, spelling, and style. Furthermore, Báez's 2018 study, targeting Spanish-English MT output after post-editing, evaluated the degree of TL learners' error correction in terms of four major problems: grammar/syntax, punctuation, spelling, and mistranslation. These studies have yet to identify MTPE strategies, but recent research goes beyond MT error analysis (e.g., Shih, 2021; Shin & Chon, 2023) and provides several useful strategies for successful MTPE.

The representative post-editing strategies are suggested in MTPE guidelines known as TAUS (2016). This provides a guide to, firstly, achieve comprehensible text-quality and, secondly, human-translation quality. While TAUS (2016) may apply to MT users in general, it has no specific delineation/examples whereby, for instance, Japanese learners of English can raise their awareness for MT error correction. Each language pair must be examined to identify MT errors and develop tailored MTPE strategies. Shih's 2021 study re-examined MT errors and MTPE strategies between Chinese and English and extracted three major errors: linguistic, pragmatic, and affective errors. This study presented MTPE strategies for each MT error-type. Linguistic errors needed to be modulated by the following five MTPE strategies: "use of words with correct meanings based on the context; adaptation of metaphorical expressions; adaptation of cultural references; use of correct grammar; and use of correct word order" (p. 140). Pragmatic errors should be amended by the strategies such as "elimination of redundant, mechanical, and repetitive words; use of consistent words; supplementation of additional words or information; rewriting or paraphrasing of the entire clause or sentence; and use of well-established, register-specific terms of the target language" (p. 144). Another significant study is Shin and Chon's 2023 examination of how TL (English) learners post-edit MT output translated from Korean to English. In this study, they classified MT errors into four categories: "mistranslation, missing words, ungrammaticality, and extra words" (p. 1). The following seven strategies required

for effective MTPE were identified: "word deletion, phase deletion, sentence/clause deletion, word paraphrase, sentence/clause paraphrase, and grammar correction" (p. 7). In their specific context, these findings will effectively raise TL learners' awareness for improved post-editing.

PURPOSE OF THE STUDY

Studies exploring the elements essential for efficiently post-editing MT output have provided profound pedagogical implications for TL learners to produce higher-quality TTs. Research dealing with the Japanese-English pair is, however, few and far between. The most relevant research is Tsutada's 2019 analysis focusing on a language activity combining pre- and post-editing. Based on the imperfections of pre-editing, this study supported the importance of post-editing activities. Some semantically and grammatically incomprehensible sentences caused by NMT errors were provided as examples of imperfect MT output. However, identifying all the errors in the MT output from pre-edited texts and then classifying them was beyond the scope of this study.

MT error analysis used as a reference for MTPE methods/strategies is still a new area of research in the field, although valuable findings related to MTPE have recently been recognised (Çetiner & İşisağ, 2019). Research dealing with a variety of language pairs needs to be more widely conducted (Harto et al., 2022) and, since each language has its own unique characteristics, it is important for Japanese learners of English to understand what they should be aware of while post-editing. To date, these elements have not been identified. The present study aims to identify the different types of error in NMT output to raise TL learners' awareness for more effective post-edit error correction. It attempted to examine the following research questions (RQ):

1. What categories of NMT error are identified in NMT output of pre-edited Japanese-written texts?
2. What specific NMT errors are identified in said NMT output?

The findings of this study could give Japanese learners of English a greater chance of improving the quality of translated TTs.

METHOD OF THE STUDY

Participants and Data Collection

This study involved 73 Japanese university students with intermediate TL (English) proficiency, who enrolled in a required English course during the spring semester of 2022. The task was to generate NMT output of pre-edited academic reports (STs) written by the students in Japanese. The participating students were allocated to classes according to their level of TL proficiency. All were around B1-B2 of the Common European Framework of Reference for Languages. The participants were asked to prepare a roughly 800-character extract of a Japanese-written academic report on their specialised subject written by themselves. Each student selected a report to be entered into NMT under the condition that their degree of familiarity with the topic should be relatively high, ensuring they had sufficient context-specific knowledge and vocabulary (Harto et al., 2022). Among the free online NMTs, this study focused on DeepL[3] since it has been shown to perform well on the BLEU (Bilingual Evaluation Understudy) score, that is, the most mainstream MT evaluation measure (Fujii et al., 2021). In general, the higher the BLEU score, the closer the translation is to a human translation.

Each student pre-edited the extract through the following process. The participants first entered the STs into NMT. Secondly, they compared TTs with STs and identified parts of the former which did not accurately reflect the latter. The corresponding parts in STs were rewritten (pre-edited), and then the pre-edited text was re-entered into NMT. This procedure[4] was repeated until they judged that

the meaning described in the TTs was consistent with that of the STs. The final TTs, consisting of about 20 sentences in each TT, and the revised STs were collected as raw data. Each student selected several (typically two-three) TL sentences (TSs) for which they could not determine the syntactic and semantic accuracy of, making a total of 183 TSs collected as data to be analysed.

This study focused on the problematic elements which were not detected by TL learners. Specifically, the analysts first detected syntactic and semantic issues, missing parts of the meaning in TTs or parts that were not translated in line with the STs. Next, the problematic elements which were detected, i.e., key areas that should be noted during post-editing were extracted and classified as either lexical, grammatical, semantic and formatting problems. Data analysis was conducted by three English educators (an English native and two Japanese native speakers). Manual analysis was carried out with the analysts working together, first independently, and then by tabulating and classifying the results upon mutual agreement. The cause of each error was carefully examined to distinguish errors caused by issues within the STs themselves from those caused by NMT error.

Reasons for Targeting NMT Output of Pre-edited Texts

There are two types of usage to produce comprehensive TTs: MT use with post-editing, and MT use with both pre- and post-editing (Cheng et al., 2021). Pym et al. (2022) compared the effects of pre-editing with those of post-editing concluding that pre-editing results in a greater reduction of errors than post-editing. Other research has revealed that pre-editing helped to improve the quality of MT output (e.g., Cheng et al., 2021; Farhana et al., 2023; Kokanova et al., 2022; Simonova & Patiniotaki, 2022; Tuzcu, 2021). While pre-editing is effective, MT output of pre-edited text still has some problematic issues (Mercader-Alarcón & Sánchez-Martínez, 2016; Tsuji & Okamoto, 2022). Hiraoka and Yamada (2019) present some cases in which the quality of MT output of pre-edited texts does not improve or even deteriorates, possibly as a result of the editors' technical skills. For that reason, post-editing is required to enhance the quality of MT output after pre-editing (Cheng et al., 2021). Accordingly, some research (e.g., Cheng et al., 2021; Zheng et al., 2022) claims that a combination of pre-editing and post-editing is more effective in improving the overall quality of MT output. Pre-editing reduces MT errors (Farhana et al., 2023; Kokanova et al., 2022; Li & Lu, 2021) and the human effort required for post-editing (ElBeheri, 2019; Güner & Güner, 2019; Kokanova et al., 2022). While post-editing improves the form and accordance with a particular domain or style and cultural appropriateness, pre-editing significantly enhances the semantic accuracy and appropriateness of MT output (Liang & Han, 2022). In light of these remarks, this study focused on TSs translated by NMT from pre-edited STs.

RESULTS

NMT errors in TTs were categorized in Table 1. Out of 183, the number of TSs with issues was 72. No issues were found by the analysts in more than half of the TSs selected by students. In the event that a sentence had multiple problematic elements, all were counted.

NMT Errors Identified in Pre-edited NMT Output Sentences

Errors Related to Lexical Issues

MT errors related to lexical issues were inappropriate TL vocabulary, and redundant or repeated TL expressions. Specific examples of the former are presented first, followed by those of the latter.

Inappropriate TL vocabulary. The most frequent problematic element was inappropriate TL vocabulary. The examples presented below have no grammatical and/or structural errors, but an incorrect use of TL vocabulary in the specific context.

Table 1. Classification Table of Problematic Elements in MT Output of Pre-edited Texts

Issue No.	Type of Issue	Problematic Elements	Number of MT Errors	Number of Errors due to ST
1	Lexical Issues	Inappropriate TL vocabulary	19[a]	5
		Redundant or repeated TL expressions	7	15
2	Grammatical Issues	Misuse or lack of determiners	8[b]	0
		Misuse or lack of prepositions	5	0
		Inappropriate singular and plural forms	3	0
		Inappropriate verb collocations	3	0
		Inappropriate verb use	3	1
		Inappropriate participle forms	2	0
		Inappropriate part-of-speech	1	0
		Lack of subject	0	2
3	Formatting Errors	Misuse of punctuation	5	0
		Misuse of upper- and lower-cases characters	2	0
		Missing quotation marks	2	0
		Others	2	0
4	Semantic Issues	Missing information	0	8
		Incomprehensible sentences	0	2

Note. This table demonstrates the problematic elements in MT output. Most of the issues were recognised as NMT errors, but some issues were caused by errors in the STs.

[a] The causes for this type of MT error are related to TL vocabularies translated incorrectly given the specific context (7), Japanese-specific expressions (6), and technical terminology (6). Numbers in parentheses indicate the number of occurrences for each cause.

[b] This type of MT error is caused by: articles (6) and quantifiers (2).

Ex. 1: Next, "information seeking" is the process by which consumers gather pertinent information in an attempt to make a reasonable decision.

The underlined vocabulary, "reasonable," is used inappropriately. In this case, "an informed decision" would be more in line with the meaning of the source-language sentence (SS) since a decision would be made after gathering detailed information. The difference between the two terms can be explained in the following example: We believe he made a reasonable decision because it was an informed decision. The SL vocabulary[5] corresponding to "reasonable" is aligned with the meaning in the SL, but not the TL. Rectifying this type of NMT error requires a higher level of TL proficiency on the part of TL learners/MT users in order to recognise the nuance in meaning that distinguishes the two terms.

The following sentence has the same issue as that of Ex. 1.

Ex. 2: It heals spontaneously, but the scar remains.

"Spontaneously" used with "heal" is unnatural, and suggests suddenness or unexpectedness. Instead of "spontaneously," it is better to use "naturally" or "by itself" to convey the original meaning. It is not easy to recognise this issue while pre-editing STs since the SL vocabulary used in the SS[6] is not particularly problematic. As Ex. 1 suggests, this type of error-correction requires a deeper knowledge of the subtle nuances of TL linguistic meanings and collocations.

As shown in the above two examples,[7] the nuances of TL vocabulary should be carefully checked. Especially since, even when the SL vocabulary is aligned in the ST context, it sometimes has issues in the TT. Other causes for this error type can be identified as follows: mistranslated SL technical terminology, mistranslated expressions specific to Japanese, and inappropriate selection of SL vocabulary. A student used a technical term, 真性半導体 in the ST, which NMT translated into "a true semiconductor," rather than "an intrinsic semiconductor." Japanese-specific expressions were also mistranslated, such as 若草色の紙, which resulted in "young grass-colored paper," while the correct translation may be yellowish-green colored paper. These two cases could be prevented by checking what TL technical terminologies are used in a specific TL context or how Japanese-specific expressions are translated in MT output. For a detailed discussion of cases related to the inappropriate selection of SL vocabulary, see *Inappropriate TL Vocabulary Caused by STs*, below.

Redundant or Repeated TL Expressions. Redundancy in TS was another notable error. Alternative descriptions or synonyms should be explored if the same vocabulary is used repetitively or unnecessarily.

Ex. 3: Therefore, newly $_{(1)}$ cleared land was $_{(2)}$ cleared, a canal was built to allow water to flow through, and the boats were moved on carts powered by hydroelectric power plants.[8]

"Cleared" is used twice in the first part. It could be improved as follows: "Therefore, land was newly cleared and…."

Ex. 4: In semiconductors, the things that play the role of charge $_{(1)}$ carriers are called $_{(2)}$ carriers.

In Ex. 4, the SS[9] has no issues, but the TS displays redundancy due to the duplication of "carriers." To be aligned with the SS meaning, the author should post-edit the sentence to "the things that play the role of carrying/transferring charge are called carriers."

As seen in these examples, TL learners/MT users need to explore alternative descriptions in the event that the same vocabulary is used repetitively.

Errors Related to Grammatical Issues

The most frequent MT errors related to grammatical issues were the misuse or lack of determiners, and prepositions. Issues with determiners are presented first, followed by those relating to prepositions.

Misuse or Lack of Determiners. The most frequent errors within grammatical issues (see Table 1) were the misuse or lack of determiners. Determiners, especially articles, are one of the most difficult grammatical elements for Japanese learners of English to acquire (Aoki, 2000) as the Japanese language has no equivalent.

Ex. 5: $_{(1)}$ Fatigue fracture is a fracture in which cracks caused by repeated stresses spread and lead to $_{(2)}$ final fracture.

"Fatigue fracture [underlined (1)]" and "final fracture [underlined (2)]" need "a" as they are not proper nouns, and the sentence is identifying these two things as different types of fracture.

Ex. 6: However, $_{(1)}$ society has become so overflowing with information that few people have the time to think for themselves, and it seems as if people are becoming more indifferent to $_{(2)}$ the society and the world behind the information.

Since the first mention of "society [underlined (1)]" is used generally (i.e., to mean "society at large") the subsequent use of "society [underlined (2)]," referring to the same thing, should not use the determiner "the." In doing so, the TS is either grammatically inaccurate or is using the term in a different way (e.g., to mean "an organisation known as a society"). Although the latter is plausible, it is confusing and does not accurately reflect the ST in any event. Another case caused by misuse of quantifiers is as follows:

Ex. 7: As a student of architecture, I strongly felt that I must acquire as much knowledge and (1) skills as I could at university and fully utilize (2) the knowledge and skills I had learned once I started working.

Since "skills [underlined (1)]" is a countable noun, "many" should be used to quantify it, rather than "much," making the correct phrase "…and as many skills…." which reconciles this issue. The issue related to the underlined (2) will be discussed later in this paper.

Misuse or Lack of Prepositions. Incorrect use of prepositions was the second most frequent grammatical error. One case relating to this is as follows:

Ex. 8: Formerly called (1) "adult diseases," the Ministry of Health and Welfare proposed in 1966 that the term (2) "lifestyle-related diseases" be changed (3) to "adult diseases" because they can be prevented by improving lifestyle habits even in adults, and because they can occur even in non-adults.

The relationship between statements in the SS[10] is reversed since the term "adult diseases" [underlined (1)] was to be changed to "lifestyle-related diseases [underlined (2)]." It can be solved if the preposition "to [underlined (3)]" is changed to "from." This type of error is not easy to solve during the pre-editing phase and careful proofreading is essential to notice such mistakes.

Other Notable Errors. Although relatively rare, the misuse of singular and plural forms and verb tense were other significant errors in the NMT output.

Ex. 9: There are two types of (1) impurity (2) semiconductors.[11]

"Types" is pluralised, therefore "semiconductors [underlined (2)]" should just be singular. As with the article, since Japanese has no equivalent, plural/singular forms are again one of the more difficult grammatical elements for Japanese learners of English to acquire. Furthermore, the noun "impurity [underlined (1)]" was classified as an inappropriate part-of-speech (see Table 1), as the correct word would be the adjective "impure."

The following sentence displays the issue related to inappropriate verb use.

Ex. 10: Nevertheless, the Japanese pair (1) had won the gold medal with an (2) outstanding combination to beat the Danes in an (3) outstanding performance.

In this case, use of past perfect tense [underlined (1)] suggests that although the "Japanese pair" previously held the gold medal, they no longer do. This is incongruent with the meaning of the SS,[12] wherein the pair still hold the medal. The selection of past perfect, rather than simple past or present perfect, alters the meaning of the TS when compared with the SS. Moreover, recognition of this discrepancy may require a relatively high TL proficiency in order to understand the nuance of using each verb tense. Separately, the underlined (2) and (3) can be regarded as repeated vocabulary, thus,

either could be replaced with synonyms, such as "fantastic" or "incredible." The corresponding SL words used in the SS is not problematic.

Errors in MT Output Caused by Source Texts

Redundancy Caused by STs

Next, the analysis will look towards errors in the MT output which were caused by problems in STs rather than any issue relating to the NMT itself. The repetitive usage of the same words or phrases in STs directly resulted in unnatural repetition in TTs.

Ex. 11: …because not only people living in Japan but also people all over the world can be affected by (1) <u>lifestyle-related diseases</u>. (2) <u>Lifestyle-related diseases</u> can lead to more dangerous (3) <u>diseases</u>.

"Disease" is used many times here. The redundancy of vocabulary can be avoided by changing "Lifestyle-related diseases [underlined (2)]" to "These," or by using the relative pronoun, "which," to smoothly connect the two sentences. Also, "diseases [underlined (3)]" could be replaced with a synonym such as "conditions" for a higher-quality TS. (Note: Unlike the former [underlined (2)], however, the SL vocabulary corresponding to "diseases [underlined (3)]" used in the SS[13] used synonyms, thus, this particular instance was recorded as an MT error, discussed in the previous section.) Similar to the underlined (1) and (2), "the knowledge and skills [underlined (2)]" in Ex. 7 was caused by redundancy in the SS[14] and should be replaced with "what" or "everything" to improve the quality of the TT. Alternative expressions or vocabularies should potentially be explored by MT users in the event that the same vocabulary item is used more than once, particularly within the same sentence.

Words having the same (or nearly the same) meaning used in a sentence can be redundant. One case relating to this is as follows:

Ex. 12: Bears have a habit of bringing back the salmon they catch to their own habitat, so during the salmon run, there are large numbers of <u>dead salmon carcasses</u> on the ground.

The underlined "dead salmon carcasses" is redundant as a carcass is always dead. However, this error was caused by redundancy in the SS[15] rather than NMT. This type of error must be noticed during the pre-editing phase.

Inappropriate TL Vocabulary Caused by STs

The examples presented below have an inappropriate TL vocabulary, caused by choosing SL vocabulary which is incompatible with the SS context.

Ex. 13: I had thought that the only effect of (1) <u>sleep deprivation</u> was physical exhaustion, but was surprised to learn that it is (2) <u>an action</u> that increases the risk of lifestyle-related diseases.

"Sleep deprivation [underlined (1)]" cannot typically be interpreted as "an action [underlined (2)]" performed on oneself. The nuance of this term is not semantically fitting even in the SS.[16] Such SL words were directly translated with an inappropriate nuance in this particular context, resulting in semantic inconsistency/unnaturalness in the TS. Instead of "an action," the use of "something" eliminates this issue. Alternatively, the TS can be revised as "…was surprised to learn that it increases the risk of lifestyle-related diseases."

Similarly, below is another problematic TS example caused by selecting incongruous SL vocabulary.

Ex. 14: Here, I believe, lies ₍₁₎the factor ₍₂₎behind what my ballet teacher once told me: that it is harder to soften the body as it grows.[17]

The term "factor [underlined (1)]" does not by itself collocate with the subsequent part [underlined (2)]. Rather, it should state, for instance, "…lies the meaning/truth/cause behind…." Using SL vocabulary equivalent to meaning/truth/cause will eliminate this unnaturalness in the translation, while retaining the ST's meaning. This type of error was caused by STs, thus, they can be avoided as part of the pre-editing phase.

To briefly summarize, the problematic elements to be corrected in relation to lexical issues were inappropriate TL vocabulary, and redundant or repeated TL expressions. The former (see Ex. 1, 2) required a deeper understanding of the subtle nuances of TL vocabulary, compared to the latter (see Ex. 3, 4, 10), which can easily be resolved by searching for synonyms or replacing them with alternative TL descriptions. TL learners need to take care with repetitive words or phrases used redundantly when writing STs. In terms of grammatical issues, the most frequent errors were the misuse or lack of determiners (see Ex. 5, 6, 7) and those related to the misuse or lack of prepositions (see Ex. 8). Advanced knowledge on TL grammar and syntactical rules is required for Japanese learners of English to correct misused determiners and prepositions, due to their absence in their L1. Other notable errors were the misuse of singular and plural forms (see Ex. 9) and verb tense (see Ex. 10). Apart from these NMT errors, there were errors caused by STs. Of these, the errors caused by redundancy or repetition in SSs (see Ex. 7, 11, 12) are relatively easy to correct. Even errors caused by inappropriate selection of SL vocabulary (see Ex. 13, 14) can generally be prevented by carefully choosing precise SL vocabulary and reducing ambiguity in STs when writing them.

DISCUSSION AND CONCLUSION

Most NMT errors did not significantly interfere with meaning since the target of this study was the NMT output of pre-edited STs.[18] The analysts did not find critical errors affecting communication such as missing sections of the message or omitted sentences, which are regarded as typical MT errors (Nakazawa, 2017; Tabata, 2021).

Out of the 183 pre-edited TSs, NMT errors which learners could not identify by themselves were found within 72 TSs. Amongst all errors in the output, issues caused by errors in the STs were excluded as they were not NMT errors but a matter of the students' Japanese-writing skills. When appropriate SSs were entered into NMT, major errors affecting intelligibility and meaning such as lack of sentence elements (Castilho et al., 2017; Wu et al., 2016) were relatively rare, thus, correction was mostly simple. The pre-editing phase can help users avoid such critical errors, so the quality of MT output translated from the pre-edited texts was largely close in meaning to the STs. However, considering the fact that the final NMT outputs still had some lexical and grammatical issues to be corrected, Japanese learners of English need to be aware of the following three elements in eliminating MT errors during a post-editing phase: lexical, grammatical, and formatting errors output by NMT. The formatting errors are relatively minor and can be solved during the reviewing process. The abovesaid three elements also provide the response to RQ1. To solve these NMT errors, training sessions for post-editing would be necessary for TL learners to effectively produce better-quality TTs. The errors identified in this study could be used as a reference for developing MTPE strategies for Japanese-English translations. That said, some issues were caused by errors remaining in the pre-edited STs, such as redundant or repeated SL expressions, missing information, inappropriate selection of SL vocabulary, the lack of subject, and incomprehensible sentences. TL learners need to proofread their pre-edited STs carefully and confirm whether the STs are comprehensible in terms of the aforesaid aspects.

In response to RQ2, the findings in this study are partially identical to NMT errors provided by TAUS (2016), Koo et al. (2022), and Daem et al. (2017). Firstly, the lexical issues mentioned in the present study, which are essential for obtaining satisfactory translation quality, are somewhat similar to the MTPE problems provided by TAUS (2016). However, most of the MTPE problems related to grammatical issues, shown in Table 1, were absent in TAUS (2016). More specifically, TAUS (2016) does not contain the following items: inappropriate determiner usage, prepositions, verb tense, and verb collocations. Likewise, the lexical issues found by Koo et al. (2022), focusing on Korean-English MT output, largely reflect those identified in the present study; in particular inappropriate TL vocabulary, and redundant or repeated TL expressions. While inappropriate TL words were mostly caused by Korean-specific expressions in Koo et al. (2022), in the present study, this error type was the result of not only Japanese-specific expressions but also technical terminology or inappropriate collocations. MT errors at the syntax and grammatical level, however, have no similarity between the two studies. From the view of grammatical issues, while the present study has identified inappropriate determiner use, preposition use, and singular and plural forms as major MT errors, Koo et al. did not regard them as critical. Additionally, Koo et al. (2022) identified the misuse of suffixes and auxiliary verbs, numeric, sequential, and semantic issues as MT errors, whereas these same errors were not found in this study. Finally, while errors related to inappropriate singular and plural forms and redundant or repeated TL expressions were found in the present study, these errors were absent in Daem et al. (2017), who analyzed English-Dutch NMT output. Daem at al. (2017) identified errors in structure, coherence, and spelling, whereas this study did not find any such issue. Otherwise, the errors identified in this study are similar to those of Daem et al. (2017). Misuse of determiners, prepositions and singular and plural forms could be considered as distinctive NMT errors particular to the Japanese-English language pair. Japanese has no linguistic elements equivalent to determiners, prepositions, and singular and plural forms. Likewise, Japanese can often be understood without clearly articulating the subject of SS in the STs, which, in some instances, may confuse the NMT and cause errors in its output. Japanese users should not delete sentence elements during the pre-editing process,[19] even in cases where the resulting sentence remains grammatically accurate and semantically comprehensible in Japanese-texts.

Even though NMTs have been making remarkable progress, they still output TTs with errors. Therefore, TL learners/MT users need to be aware of the aforementioned elements to effectively use such imperfect NMTs. It is crucial for TL learners to learn the correct use of determiners, prepositions, and singular and plural forms in the TS context. Therefore, they need to pursue further learning to acquire a higher level of TL proficiency required to notice the nuances of TL linguistic meanings and their agreement or disagreement with that of the ST.

LIMITATIONS

While the present study reported important insights into identifying NMT errors for effective MTPE, there are some limitations in its design. The first limitation is that this study focused on MT output of pre-edited STs. The results would be presumably different when analysing NMT output without pre-editing. The second is that the study involved a limited number of participants at a particular level of TL proficiency. Furthermore, it focused on translating from Japanese (SL) to English (TL) within the basic academic writing genre. Accordingly, the findings cannot necessarily be extrapolated to other languages, writing genres or NMT users in different contexts. Therefore, further research comparing elements users must be aware of when post-editing TTs (i.e., MT output from pre-edited texts) to those for raw TTs (without pre-editing) is required moving forward, along with a multi-genre

text design. This may contribute to determining how learners can effectively use NMT to produce higher quality TTs in a range of contexts and genres.

ACKNOWLEDGMENT

AUTHOR CORRESPONDENCE

Correspondence should be addressed to Prof. Kayo Tsuji; tsujikayo@omu.ac.jp

REFERENCES

Alsalem, R. (2019). The effects of the use of Google Translate on translation students' learning outcomes. *AWEJ for Translation & Literary Studies, 3*(4), 46-60. http://doi.org/10.24093/awejtls/vol3no4.5

Aoki, N. (2000). A study of Japanese university students' judgments on English article use. *Hiroshima Journal of International Studies, 6*, 117-130. https://hiroshima-cu.repo.nii.ac.jp/record/294/files/HJIS6-117.pdf

Báez, M. C. T. (2018). Machine translation and post-editing: Impact of training and directionality on quality and productivity. *Tradumatica, 16*, 24-34. https://doi.org/10.5565/rev/tradumatica.215

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics, 108*(1), 109-120. http://doi.org/10.1515/pralin-2017-0013

Çetiner, C., & İşisağ, K. U. (2019). Undergraduate level translation students' attitudes towards machine translation post-editing training. *International Journal of Languages' Education and Teaching, 7*(1), 110-120. http://doi.org/10.182 98/ijlet.3242

Cheng, Y., Yue, S., Li, J., Deng, L., & Quan, Q. (2021). Errors of machine translation of terminology in the patent text from English into Chinese. *ASP Transactions on Computers, 1*(1), 12-17. https://www.sciencegate.app/document/10.52810/tc.2021.100022

Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. *Journal of Asia TEFL, 17*(1), 182-193. http://doi.org/10.18823/asiatefl.2020.17.1.11.182

Daem, J., Vandepitte, S., Hartsuiker, R., & Machken, L. (2017). Identifying the machine translation error types with the greatest impact on post-editing effort. *Frontiers in Psychology, 8*, 1-15. https://doi.org/0.3389/fpsyg.2017.01282

ElBeheri, N. (2019). Collaboration between machine translation and human translation for higher quality and more production in translation. *Journal of English and Comparative Studies, 1*, 446-473. http://doi.org/10.21608/ttaip.2019.123829

Emara, N. Y. (2023). Using machine translation error identification to improve translation students' post-editing skills. *Transcultural Journal of Humanities and Social Science, 4*(1), 345-357. https://doi.org/10.21608/tjhss.2023.289357

Escartín, C. P., O'Brien, S., Goulet, M. J., & Simard, M. (2017). Machine translation as an academic writing aid for medical practitioners. *Proceedings of MT Summit XVI: Research Track*, 254-267. https://core.ac.uk/download/pdf/200762284.pdf

Farhana, B. C. D., Baharuddin, W. A. L., & Farmasari, S. (2023). Academic text quality improvement by English department students of University of Mataram: A study on pre-editing of Google neural machine translation. *Jurnal Ilmiah Profesi Pendidikan, 8*(1), 247-254. http://doi.org/10.29303/jipp.v8i1.1186

Fredholm, K. (2019). Effects of Google translate on lexical diversity: Vocabulary development among learners of Spanish as a foreign language. *Revista Nebrija, 13*(26), 98-117. http://doi.org/10.26378/rnlael1326300

Fujii, R., Mita, M., Abe, K., Hanawa, K., Morishita, M., Suzuki, J., & Inui, K. (2021). Phenomenon-wise evaluation dataset towards analyzing robustness of machine translation models. *Natural Language Processing, 28*(2), 450-478. https://doi.org/10.5715/jnlp.28.450

Funakoshi, K., Fujishiro, Y., Nomura, S., & Ishida, T. (2004). Analysis of interactivity in intercultural collaboration environments-lessons learned of ICE2002. *Information Processing Society of Japan, 45*(1), 112-120. https://jglobal.jst.go.jp/detail?JGLOBAL_ID=200902272659978417

Güner, P., & Güner, E. S. (2019). Assessing post-editing effort through semantic similarity. *Proceedings of International Linguistics and Language Studies Conference LILA '19*, 17-28. https://www.researchgate.net/publication/353411340

Harto, S., Hamied, F. A., Musthafa, B., & Setyarini, S. (2022). Exploring undergraduate students' experiences in dealing with post-editing of machine translation. *Indonesian Journal of Applied Linguistics, 11*(3), 696-707. http://doi.org/10.17509/ijal.v11i3.42825

Hiraoka, Y., & Yamada, M. (2019). Pre-editing plus neural machine translation for subtitling: Effective pre-editing rules for subtitling of TED talks. *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, 64-72. https://www.aclweb.org/anthology/W19-6710.pdf

Ishikawa, Y. (2020). Consideration for effective use of Japanese-English machine translation: Application of an English writing strategy. *LET Journal of Central Japan, 31*, 23-37. https://doi.org/10.20656/letcj.31.0_23

Jia, Y., Carl, M., & Wang, X. (2019). Post-editing neural machine translation versus phrase-based machine translation for English-Chinese. *Machine Translation, 33*, 9-29. https://link.springer.com/article/10.1007/s10590-019-09229-6

Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv:2301.08745v4*, 1-9. https://arxiv.org/abs/2301.08745

Klimova, B., Pikhart, M., Benites, A. D., Lehr, C., & Sanchez-Stockhammer, C. (2023). Neural machine translation in foreign language teaching and learning: A systematic review. *Education and Information Technologies, 28*(1), 663-682. http://doi.org/10.1007/s10639-022-11194-2

Kliffer, M. (2005). An experiment in MT post-editing by a class of intermediate/advanced French majors. *Proceedings of European Association for Machine Translation (EAMT) 10th Annual Conference*, 160-165. https://aclanthology.org/2005.eamt-1.22.pdf

Kokanova, E. S., Berendyaev, M. V., & Kulikov, N. Y. (2022). Pre-editing English news texts for machine translation into Russian. *Language Studies and Modern Humanities, 4*(1), 25-30. https://doi.org/10.33910/2686-830X-2022-4-1-25-30

Koo, S., Park, C., Seo, J., Lee, S., Moon, H., Lee, J., & Lim, H. (2022). K-NCT: Korean neural grammatical error correction gold-standard test set using novel error type classification criteria. *IEEE Access, 10*, 118167-118175. https://doi.org/10.1109/ACCESS.2022.3219448

Kuraya, N. (2019). The potential use of machine translation services as a learning tool in English writing. *International Information Research, 16*(1), 24-35. https://doi.org/10.11424/gscs.16.1_24

Lee, S-M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning, 33*(3), 157-175. http://doi.org/10.1080/09588221.2018.1553186

Lee, S-M. (2022). Different effects of machine translation on L2 revisions across students' L2 writing proficiency levels. *Language Learning & Technology, 26*(1), 1-21. https://hdl.handle.net/10125/73490

Lee, S-M. (2023). The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis, *Computer Assisted Language Learning, 36*(1-2), 103-125. https://doi.org/10.1080/09588221.2021.1901745

Lee, S-M., & Briggs, N. (2020). Effects of using machine translation to mediate the revision process of Korean university students' academic writing. *ReCALL, 33*(1), 18-33. http://doi.org/10.1017/S0958344020000191

Li, Y., & Lu, X. (2021). Study on post-editing for machine translation of railway engineering texts. *SHS Web of Conferences, 96*, 1-11. https://doi.org/10.1051/shsconf/20219605001

Liang, Y., & Han, W. (2022). Source text pre-editing versus target text post-editing in using Google Translate to provide health services to culturally and linguistically diverse clients. *Science, Engineering and Health Studies, 16*, 1-5. https://doi.org/10.14456/sehs.2022.25

Marzouk, S., & Hansen-Schirra, S. (2019). Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. *Machine Translation, 33*, 179-203. https://link.springer.com/article/10.1007/s10590-019-09233-w

Mercader-Alarcón, J., & Sánchez-Martínez, F. (2016). Analysis of translation errors and evaluation of pre-editing rules for the translation of English news texts into Spanish with Lucy LT. *Traduccio i dispositius mobils Revista Tradumatica: tecnologies de la traduccio, 14*, 172-186. http://doi.org/10.5565/rev/tradumatica.164

Nakazawa, T. (2017). New paradigm for machine translation: How the neural machine translation works. *Journal of Information Processing and Management, 60*(5), 299-306. https://doi.org/10.1241/johokanri.60.299

Nishimura, S., Nevgi, A., & Tella, S. (2008). Communication style and cultural features in high/low-context communication cultures: A case study of Finland, Japan, and India. *Proceedings of a Subject-Didactic Symposium*, 783-796. https://api.semanticscholar.org/CorpusID:202955957

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL, 21*(2), 241-258. http://doi.org/10.1017/S0958344009000172

Pym, A., Ayvazyan, N., & Prioleau, J. (2022). Should raw machine translation be used for public-health information? Suggestions for a multilingual communication policy in Catalonia. *Journal of Language Rights & Minorities, Revista de Drets Linguistics i Minories, 1*(1-2), 71-99. https://doi.org/10.7203/Just.1.24880

Sánchez-Gijón, P., & Torres-Hostench, O. (2014). MT post-editing into the mother tongue or into a foreign language? Spanish-to-English MT translation output post-edited by translation trainees. *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*, 5-19. https://aclanthology.org/2014.amta-wptp.1.pdf

Samman, H. M. (2022). *Evaluating machine translation post-editing training in undergraduate translation programs: An exploratory study in Saudi Arabia*. [Doctoral dissertation, University of Southampton]. University of Southampton Institutional Repository. http://eprints.soton.ac.uk/id/eprint/469163

Shih, C. L. (2021). Re-looking into machine translation errors and post-editing strategies in a changing high-tech context. *Compilation and Translation Review, 14*(2), 125-166. https://ctr.naer.edu.tw/v14.2/ctr140204.pdf

Shin, D., & Chon, Y. V. (2023). Second language learners' post-editing strategies for machine translation errors. *Language Learning & Technology, 27*(1), 1-25. https://hdl.handle.net/10125/73523

Simonova, V., & Patiniotaki, E. (2022). Pre-editing for the translation of life-science texts from Russian into English via Google Translate. *Proceedings of New Trends in Translation and Technology 2022*, 259-265. https://www.researchgate.net/profile/Abdelalah-Alsolami/publication/371681915

Tabata, F. (2021). Verification of Chinese patent translation: Verification of the Chinese Patent English abstracts provided by China Patent Office and machine translation provided by Japan Patent Office. *Proceedings of 18th Information Professional Symposium*, 7-12. https://www.jstage.jst.go.jp/article/infopro/2021/0/2021_7/_pdf/-char/ja

TAUS. (2016). *Post-editing guidelines*. https://www.taus.net/resources/reports/mt-post-editing

Tsuji, K. (2020). *Bogo shiyō o toriireta gaikokugo raitingu ni kansuru kenkyū* [*Research on L1 formulation during L2 writing process: Rethinking L2 writing process for novice writers*]. [Unpublished doctoral dissertation]. Kyoto University.

Tsuji, K. (2021). Developing and evaluating a scoring rubric for argumentative essays: A module-based approach. *Urban Scope, 12*, 1-13. https://urbanscope.lit.osaka-cu.ac.jp/

Tsuji, K., & Okamoto, K. (2022). Improving the process of L1 paraphrasing independent language learning on L1 paraphrasing: Using machine translation as a support tool. *Journal of Japan Association for College and University Education, 44*(2), 95-105. https://doi.org/10.60182/jacuejournal.44.2_95

Tsutada, K. (2019). Assessing awareness and acceptability of neural machine translation among Japanese university students in relation to its practical application. *Journal of Inquiry and Research, 110*, 35-52. https://doi.org/10.18956/00007874

Tuzcu, A. (2021). The impact of Google Translate on creativity in writing activities. *Language Education and Technology, 1*(1), 40-52. https://langedutech.com/letjournal/index.php/let/article/view/18

Udina, N. (2019). Using post-editing in translation and LSP courses. *Proceedings of 6th International Conference on Education and Social Sciences*, 1097-1101. https://www.researchgate.net/profile/Natalie-Udina/publication/330997355

Vidal, S. A. (2021). *Post-editing effort and linguistically motivated evaluation of machine translation*. [Doctoral dissertion, Universitat Pompeu Fabra]. Theses and Dissertations Online (TDX). https://www.tdx.cat/handle/10803/672723?locale-attribute=en#page=1

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M, Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv:1609.08144v2*, 1-23. https://arxiv.org/abs/1609.08144

Yamada, M. (2019). The impact of Google neural machine translation on post-editing by student translators. *The Journal of Specialised Translation, 31*(1), 87-106. http://www.jostrans.org/issue31/art_yamada.pdf

Yamada, M. (2021). Post-editing and a sustainable future for translators. *Journal of Foreign Language Studies, 24*, 83-105. https://doi.org/10.32286/00023085

Yoon, C. W., & Chon, Y. V. (2022). Machine translation errors and L2 learners' correction strategies by error type and English proficiency. *English Teaching, 77*(3), 153-175. http://doi.org/10.15858/engtea.77.3.202209.153

Zhang, H., & Torres-Hostench, O. (2022). Training in machine translation post-editing for foreign language students. *Language Learning & Technology, 26*(1), 1-17. http://hdl.handle.net/10125/73466

Zheng, Y., Peng, C., & Mu, Y. (2022). Designing controlled Chinese rules for MT pre-editing of product description text. *International Journal of Translation, Interpretation, and Applied Linguistics (IJTIAL), 4*(2), 1-13. http://doi.org/10.4018/IJTIAL.313919

ENDNOTES

[1]	There are two types of post-editing: 1) Post-editing TTs written in a first language (L1), and 2) post-editing TTs written in a second language (L2). The Previous Research section focuses mainly on the latter as the present study aims to identify NMT errors written in the L2 (English). The TL can be either an L1 or L2, as it refers to the language output by MT, however this paper mainly refers to TL as the L2.

[2]	This is drawn from results of Tsuji (2020). The researcher conducted a systematic review of previous studies on L2 writings and described the differences in the reviewing process between advanced and beginner L2 writers.

[3]	TTs output by DeepL are comparable to those from ChatGPT (Jiao et al., 2023).

[4]	The student's awareness prompted by pre-editing was categorized in two ways: 1) an awareness of syntactic development to improve syntactic consistency, and 2) an awareness of semantic development to improve content consistency. More details are described in Tsuji and Okamoto (2022).

[5]	The pre-edited SS is as follows: 次に、「情報探索」とは、消費者が妥当な決定を下そうと、適切な情報を集めていくプロセス。The underlined SL term corresponds to the underlined part in Ex. 1.

[6]	The pre-edited SS is as follows: それは自然回復するが、傷跡は残ってしまう。The underlined SL term corresponds to the underlined part in Ex. 2.

[7]	All Japanese words in Ex. 1 and 2 make sense in STs, but the English translation was inaccurate in the given context.

[8]	The pre-edited SS is as follows: そのため、新たに土地を(1)(2)切り開いて、通水させるために水路を設け、それに加え水力発電所の動力を使い台車に船をのせて移動させるという手法がとられた。The underlined SL term relates to the numbered terms in Ex. 3.

[9]	The pre-edited SS is as follows: 半導体では、電荷を(1)運ぶ役割を果たすもののことを(2)キャリアと呼びます。The underlined SL terms are those related to the numbers in Ex. 4.

[10]	The pre-edited SS is as follows: 以前は、(1)「成人病」と呼ばれていたが、成人であっても生活習慣の改善により予防でき、成人でなくても発症する可能性があることから、1966年に厚生省が(2)「生活習慣病」と改称することを提唱した。The underlined SL terms are those related to the numbers in Ex. 8. The preposition "to [underlined (3) in Ex. 8]" has no equivalent in the SS.

[11]	The pre-edited SS is as follows: 2種類の(1)不純物(2)半導体がある。The underlined SL terms are those related to the numbers in Ex. 9.

[12]	The pre-edited SS is as follows: それにもかかわらず、日本のペアは、(2)抜群のコンビネーションでデンマークを(3)見事に下し、金メダルを(1)獲得していた。The underlined SL terms are those related to the numbers in Ex. 10.

[13]	The pre-edited SS is as follows: なぜなら、日本に住んでいる人々だけでなく世界中の人々も(1)生活習慣病になってしまう可能性があるからだ。(2)生活習慣病はもっと危険な(3)疾患につながる。The underlined SL terms are those related to the numbers in Ex. 11.

14   The pre-edited SS is as follows: 建築学科の学生として、私は大学で学び得ることのできる知識や技術を₍₁₎できるだけ多く習得し、就職後は学び得た₍₂₎知識や技術を十分に活用しなければならないと強く感じた。 The underlined SL terms are those related to the numbers in Ex. 7.

15   The pre-edited SS is as follows: クマは捕まえたサケを自分の住処に持ち帰る習性があるので、サケの遡上の時期は死んだサケの死骸が大量に地面に落ちている。 The underlined SL term corresponds to the underlined part in Ex. 12.

16   The pre-edited SS is as follows: 私は、₍₁₎睡眠不足の影響は体が疲れるだけだと思っていたが、生活習慣病のリスクを高める₍₂₎行為であることに驚いた。 The underlined SL terms are those related to the numbers in Ex. 13.

17   The pre-edited SS is as follows: ここに、₍₂₎かつてバレエの先生に言われた、成長すると身体を柔らかくしにくくなるという₍₁₎言葉の要因があると考える。 The underlined SL terms are those related to the numbers in Ex. 14.

18   The activity of pre-editing contributes to syntactic and semantic development in STs: The former is improved by, for instance, clarifying necessary sentence parts, simplifying sentence structure, and using appropriate linking words, etc.; and the latter by avoiding long SSs, simplifying lengthy phrases, and reducing the ambiguity in each word, etc. More details are described in Tsuji and Okamoto (2022).

19   Japan's ethnic and cultural homogeneity influences students to heavily rely on their implicit understanding during communication (Nishimura et al., 2008) in the case written communication, too (Tsuji, 2021).