



Second language learners' post-editing strategies for machine translation errors

Dongkawang Shin, Gwangju National University Of Education

Yuah V. Chon, Hanyang University

Abstract

Considering noticeable improvements in the accuracy of Google Translate recently, the aim of this study was to examine second language (L2) learners' ability to use post-editing (PE) strategies when applying AI tools such as the neural machine translator (MT) to solve their lexical and grammatical problems during L2 writing. This study examined 57 students' MT output and post-edited (PEd) texts to analyze MT errors and the PE strategies that L2 learners employed to express target meaning. The MT errors occurred from mistranslation, missing words, ungrammaticality, and extra words. To modify the MT sentences, the learners employed PE strategies such as deletion, paraphrase, and grammar correction. Successfulness of PE was gauged by comparing sentence adequacy scores of the MT output and PEd texts. The results of the study highlight that L2 proficiency influences the learners' ability to deploy appropriate PE strategies. The taxonomy of MT errors and PE strategies provides a model for understanding the competence required as part of the new writing ability in the AI era. Implications are discussed as to how L2 learners are required to be trained in using MT by detecting MT errors and deploying appropriate PE strategies.

Keywords: *Machine translation, Post-editing, Errors, Sentence adequacy*

Language(s) Learned in This Study: *English*

APA Citation: 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>

Introduction

Over the past few decades, digital technologies such as online dictionaries, spell check, grammar check, and concordancers have helped second language (L2) learners solve lexical and grammatical problems encountered during the writing process. The paradigm shift from statistical to neural machine translation (NMT) with *Google Translate* (GT) has also made a more significant impact on the way L2 learners write since 2016 as language learners' dependence on the use of MT is rapidly increasing (Crossley, 2018; Ducar & Schocket, 2018). While NMT is a recent emerging approach, it attempts to build and train a single large neural network that reads a sentence and outputs a correct translation. Before 2016, computer translations were performed in a phrase-based manner; however, NMT produces translations that are more accurate than statistical machine translation (SMT) due to its superior ability to translate complete sentences at a time (Le & Schuster, 2016).

Within the field of translation, the act of editing or revising MT output has been referred to as “post-editing” (PE); however, limited research has been conducted on this around L2 writing. Indeed, there is the concern that lower level L2 learners' excessive dependence on MT may lead to the neglect of efforts to improve their own writing skills (Kol et al., 2018). However, the accessibility of MT is likely to make it a widely referenced tool for future L2 writers. That is, the use of MT comes with the advantage of being able to help remove L2 learners' cognitive overload that is often associated with human translation (Jia et al., 2019). With MT, learners can also devote more of their working memory to conceptualizing ideas and thinking of rhetorical devices rather than focusing on the literal translation of lexical items (Chon et al., 2021, Kol et

al., 2018; Lee, 2020). Nonetheless, competence in using MT subsumes the ability to revise raw MT output by identifying the errors in them and making appropriate choices in terms of PE strategies (Kliffner, 2008; Sun, 2017). PE competence is a relatively new area of research and there is limited research on the ways in which even translation students can acquire it (Blagodarna, 2020), not to mention its limited research in L2 writing.

With the aim of researching how L2 learners utilize PE strategies in response to the errors that they detect in an MT output, we asked learners to compose a text in Korean (L1) and submit it to GT. The researchers assessed the adequacy of the MT output by using sentences as the unit of analysis and identified MT errors. PE strategies were examined focusing on the types of MT errors. The effects of PE strategies were gauged by comparing the quality of MT output with the post-edited (PEd) text through a comparison of sentence adequacy scores.

Background

Quality of Machine-Translated Output

In the days of statistical MT, Vilar et al. (2006) aimed to present a framework for human error analysis of MT output. Temnikova (2010) enriched Vilar et al.'s (2006) four-category error classification with an interpretation that considered the cognitive effort required to correct different MT errors. A slightly modified version of the error difficulty scale was also used in a study by Koponen et al. (2012) to investigate error types found in sentences with long or short PE times but with a similar number of mistakes. More recently, further studies were conducted regarding MT errors by adopting the quality assessment of MT output (Daems et al., 2015; Koponen & Salmi, 2015; Lacruz, 2017). Moorkens (2018) examined undergraduate and PhD students of Translation Studies who had to employ an error annotation using a typology of errors consisting of word order errors, mistranslations, omissions, and additions. The students were asked to carry out evaluations of statistical and neural MT output using TQA metrics: adequacy, PE productivity, and an error taxonomy. In this study, "adequacy" gauged the extent to which the meaning expressed in the source appeared in the translation fragment.

In contrast, L2 writing researchers have only recently taken interest in analyzing MT output, where MT has been reported as a resource of assistance for L2 writing. Groves and Mundt (2015) found GT to be able to translate a number of stretches of Malay or Chinese into grammatically correct English; however, the GT output was far from perfect and the majority of errors occurred in sentence structure and word choice. According to Wallwork (2016), in the translations from Italian to English, GT was the cause of errors in word order, word forms (plural -s on acronyms, un/countable nouns), and misuse of tenses. Kol et al. (2018) conducted an awareness task to assess student awareness of GT mistakes. The intermediate level students identified 54% of the mistakes, while advanced students identified 73% and corrected 87% of the identified mistakes. Tsai (2019) asked English as a foreign language (EFL) learners to write first in Chinese, later draft the corresponding text in English, translate the Chinese into English using GT, and finally compare their self-written English texts with their machine-translated English texts. When both English drafts were analyzed, the machine-translated English texts had more words, fewer mistakes in spelling and grammar, and fewer errors per word. Chon et al. (2021) asked EFL university learners to produce compositions for direct writing, self-translated writing, and machine-translated writing. The results indicated that MT had narrowed the difference of writing ability between the skilled and less skilled learners. However, MT-assisted texts were found to contain a higher number of mistranslations and poor word choices compared to when the L2 learners had not received any aid in writing. Lee (2021) found that MT output and EFL university student-generated texts were equally comprehensible, but MT outperformed the students in spelling, vocabulary, and grammatical accuracy. Cacino and Panes (2021) conducted a study with EFL high school learners who were randomly assigned to one of three groups: GT without instruction, GT with instruction, and a group with no access to GT. Results suggested that syntactic complexity and accuracy scores were higher in the groups that had access to MT in comparison to the group that did not have the tool. Chung and Ahn (2021) examined how L2 learners' use of MT affects syntactic complexity, accuracy,

lexical complexity, and fluency in L2 writing. Text analysis of students' writing revealed major improvements in accuracy but unclear benefits in syntactic and lexical complexity. Furthermore, using MT helped learners increase lexical variation but it led to lower scores in lexical sophistication, suggesting that MT was recommending a wide range of common and frequently-used vocabulary.

PE of Machine Translation Output

Recently, PE has begun to attract considerable attention due to the quality of raw MT output that may not completely represent the meaning reflected in the source text or the inaccuracy of the MT output based on the linguistic rules of the target language (e.g., Chung, 2020; Garcia, 2011; Koponen, 2016; Sun, 2017). "PE typically refers to a situation where a language professional takes the raw machine translation output and corrects any errors in order to bring it up to an acceptable quality" (Bowker & Ciro, 2019, p. 25).

Most research on PE have been conducted within the area of translation studies (Allen, 2003; Lacruz, 2017). The limited number of studies on PE in L2 writing clearly calls for more research. One of the early studies that have implications for PE in L2 writing was conducted by Niño (2008). When EFL students were asked to post-edit a raw MT output by consulting different online resources, they were found to make use of PE strategies such as rewriting, paraphrasing, self-correction, guessing, inferencing, reflecting, and use of synonyms. Kliffer (2008) conducted experimental tasks with undergraduate students to compare performance between "from-scratch" translations and PEd MT output for French-to-English translations. Comparison between student translations and PEd versions indicated significant differences between *literal translations* and *word choice*, illustrating that MT errors had been rectified through PE. However, *word choice* remained the most frequent error category among the PE results. Even after PE, average frequency of total errors between PEd output (30.65) and student translations (45.15) was not significantly different.

More recently, Lee (2020) asked EFL university students to translate their L1 writing into L2 without the help of MT and then correct their L2 writing using the MT translation for comparison. In the process of editing with MT, the students were found using writing strategies such as double-checking, using previous knowledge, inferencing, paraphrasing, and rewriting. Chung (2020) investigated how L2 proficiency affects the degree to which EFL university students can discern the accuracy of the MT output and whether proficiency affects the PE process. With increasing proficiency, the number of corrections increased especially above the word level, and significant group differences could be found in PE patterns of the MT text. Lee and Briggs (2021) examined error corrections made by Korean university students by comparing their original L2 texts to that of MT output. The results proved that the MT had mostly helped students to make corrections concerning errors in articles, prepositions, noun plurals, and substitutions.

In sum, the current state of research on MT most noticeably indicates that L2 writing studies which have taken interest in the PE process or strategies are scarce due to the evolving nature of literature on the use of MT for L2 writing. There is also lack of research that systematically analyzes PE strategies that are employed to improve the quality of raw MT output. The literature on MT indicates that there is no consensus among researchers regarding the classification of MT errors, for instance, due to different operations involved depending on the type of MT and different language pairs. More noticeably, PE has been discussed mainly in translation studies by analyzing the frequency of errors and their impact on PE (Daems et al., 2015), by assessing temporal (Koponen et al., 2012) and cognitive efforts (Lacruz et al., 2012; Lacruz, 2017) demanded by the MT output, and by reporting on quality assessment of MT output with interest in evaluating the usability value of MT in the translation industry. Although MT has demonstrated improved productivity, there is urgent need to examine how it is likely that MT can change the way L2 writers draft and post-edit MT output.

Research Questions

The use of neural MT for L2 writing can expedite the writing process. However, successful use of MT will require learners to notice the errors and employ appropriate PE strategies. We examined the MT output by identifying errors and evaluating the adequacy of the meaning which were expressed in L2, and performed

an analysis on how the MT output was improved by learners' PE. We also examined if L2 proficiency was a factor when the learners attempted to post-edit the MT output. To this end, the following research questions (RQs) guided our study.

1. How adequate was the machine-translated (MT) output in expressing the writer's intended meaning when examined through 1) MT errors and 2) sentence adequacy scores?
2. How according to L2 learners' proficiency levels, were PE strategies employed to rectify the errors in the MT output?
3. How according to L2 learners' proficiency levels, did the PEd text differ from the MT output in expressing the L2 learners' intended meaning when sentence adequacy scores were compared?

Method

Participants and Context of Study

Fifty-seven L2 university learners of English from South Korea (hereafter, Korea or Korean) participated in the study. There were 16 male and 41 female students. The participants, who were native Korean speakers, were from two universities: Gwangju ($N = 27$) and Seoul ($N = 30$). At the time of the study, the students, ranging from 22-25 years of age, were enrolled in writing courses as juniors. The students from Seoul were English Language Teaching majors, who were aiming to become secondary school teachers, and those from Gwangju were Education majors, who were being trained to be elementary school teachers. The diagnostic writing task conducted at the beginning of the semester indicated that according to the CEFR (Common European Framework of Reference) scale, the students from Seoul were advanced English learners (C1, hereafter "skilled learners") and the students from Gwangju were high-intermediate English learners (B2, hereafter "less skilled learners"). Although the students had started to learn English from their third year of elementary school (10 years old), most of the participants lacked experience in the productive skills of speaking and writing. This is due to how the learners' instruction of English was focused primarily on the grammar-translation method and standardized tests.

Instruments and Procedure

Main Writing Task

Before learners performed the main writing tasks, they received training on the history of MT and performed an awareness task based on some MT sentences to familiarize them with some of the language-specific translation errors that can occur between Korean-English, when Korean is used as the source language.

The students performed the main writing tasks in three stages. In stage one, the learners were required to write an argumentative essay in Korean (L1). The topic required the learners to take a position by providing reasons for either agreeing or disagreeing with the use of the Internet (see [Appendix A](#)). While this was a familiar topic to the students, two supporting ideas were provided with key expressions to reduce students' cognitive burden. The students had to add one supporting idea of their own. In stage two, the students were asked to submit the source text (L1) to GT to obtain a translation (L2). In the last stage, the learners were asked to post-edit any parts of the MT output that did not reflect the meaning stated in the source text, or any language problems that they felt required correcting. The students were asked to write approximately 300 words in English. During the 60 minutes provided to complete the whole writing task, learners first wrote in L1 for 30 minutes. Then the students translated their writing with MT, and post-edited the MT output for the next 30 minutes. The writing tasks yielded a source text (L1), MT output, and a PEd text.

Rating of Sentence Adequacy

To evaluate how well the MT output and PEd text conveyed the meaning stated in the source text, sentences were rated for adequacy on a 5-point scale (1 = None of it, 2 = Little of it, 3 = Some of it, 4 = Most of it,

and 5 = All of it). The source text (L1) was also analyzed for adequacy on a 5-point scale (1 = Very poor, 2 = Poor, 3 = Fair, 4 = Good, and 5 = Excellent) by evaluating how well the source text was semantically, syntactically, and grammatically correct in expressing meaning (Moorkens, 2018). While rating of sentence adequacy was a holistic method of evaluating how much of the meaning expressed in the source text (L1) appears in the machine-translated or post-edited sentences, the raters were asked to assign points according to the severity of word choice, grammar, or sentence structure errors. Descriptors for the scale were provided during the evaluation process (see [Appendix B](#)). Judgment of sentence adequacy was conducted on 2,793 sentences (3 Stages x 931 sentences) independently by each of the two raters. For the rating tasks, experts who knew both source and target languages on a high if not native level were required. The present study's researchers fulfilled this purpose. Before rating the sentences, the raters participated in a workshop where they first rated sample sentences before reaching a consensus regarding the sentence adequacy scheme and scale.

For the verification of sentence adequacy of MT output and PEd text, a native English speaker working at one of the researchers' universities and experienced in rating English compositions was also recruited. The rater examined approximately 10% of the English sentences (200 sentences) from the MT output and PEd text to check the adequacy and the raters' evaluation of the English sentences, and any disagreements were negotiated. The native speaker rater's role was to validate the two bilingual raters' judgment of sentence adequacy rather than to conduct an exhaustive evaluation of the sentences. In addition, the use of random sampling of the sentences, which yielded a sample that was the representative of the MT output and the PEd text that was studied, established external validity (Moore & McCabe, 2003). The bilingual and native-speaker raters were able to reach an agreement for more than 95% of the sentences. Any disagreement was resolved through discussion. Between the two bilingual raters, Pearson's correlation indicated high inter-rater reliability. Reliability coefficients were .804, .945, and .884 ($p < .01$) for the source text, MT output, and PEd text, respectively. The mean score derived from the two raters was used as the final sentence adequacy score.

Coding of Machine Translation Errors

For quality assessment of the MT output, coding of errors was also necessary for the individual sentences. The error scheme was first established by referring to previous studies on error analysis (e.g., Costa et al., 2015; Ferris, 2011; Groves & Mundt, 2015; Moorkens, 2018; Lee & Briggs, 2021; Wallwork, 2016). Through a reiterative process of analyzing the errors that appeared in the MT output, the researchers were able to finalize the error scheme as largely consisting of four types –Missing Words, Mistranslations, Ungrammaticality, and Extra Words (see [Table 1](#)). Mistranslations are instances when the *meaning* in the source text (L1) has not been clearly indicated in the machine-translated text (L2). This could occur at the level of *sentences/clauses*, *phrases* and *words*. For instance, the MT output was coded as a mistranslation at the clause level when the MT output produced “people who are out of shopping,” which should have been “people who waste their money because of shopping.” Ungrammaticality occurred when there were errors related to *verb tense*, *article*, *sentence fragment*, *missing preposition*, *wrong verb form*, *wrong word form*, *misplaced adverb*, and *word order*. Errors with Missing Words and Extra Words occurred for single word items. Within a sentence, more than one error could be coded.

Table 1*Machine Translation Error Scheme*

Errors in MT Sentences	MT Error Scheme	Sample Excerpt
Missing Word	MW	MT As a result, you will waste time <i>scrolling</i> the screen.
		TS As a result, you will waste time <i>meaninglessly scrolling</i> the screen.
Mistranslation	MT	MT Because it is a stimulus that cannot be easily accessed by other everyday activities, people are immersed in the Internet without <i>their knowledge</i> .
		TS Because it is a stimulus that cannot be easily accessed by other everyday activities, people are immersed in the Internet without <i>realizing</i> .
Ungrammaticality	UG	MT The pleasures of not showing up in the real world but revealing one's own appearance in the Internet community <i>addicted</i> people and prevented them from living normal lives.
		TS The pleasures of not showing up in the real world but revealing one's own appearance in the Internet community <i>has addicted</i> people and prevented them from living normal lives.
Extra Word	EW	MT As a result, there are many sources of <i>unreliable and unreliable</i> information on the Internet.
		TS As a result, there are many sources of <i>unreliable</i> information on the Internet.

Note. MT = Machine Translation, TS = Target Sentence.

Coding of PE Strategies

The last stage of coding involved identifying and classifying the PE strategies adopted by the learners while revising the MT output (see Table 2). In line with the rating of adequacy and MT errors, coding for PE strategies was also based on individual sentences. The framework for PE strategies was initially conceptualized from previous work (Barreiro, 2008; Jia et al., 2019; Lee, 2018; Moorkens, 2018), and through a reiterative process of analyzing the PE text and strategies, a taxonomy of PE strategies could be established. Particularly, paraphrase was conceptualized as a PE strategy by referring to Barreiro's (2008) work, which explains that the "translation of a source language into a target language does not operate necessarily only at the sentence level in all circumstances" (p. 29) and that it is important to understand how paraphrase operates at different levels. For instance, Barreiro (2008) states that paraphrase is more often associated with synonymy and usually operates at the lexical or phrasal level (words and multiword expressions). In the end, analyzing the strategies yielded a PE strategy scheme that can be broadly classified into deletion, paraphrase, and grammar correction (see Table 2). Deletions occurred for words (*Word Deletion*), phrases (*Phrase Deletion*), and sentences (*Sentence Deletion*). Paraphrase strategies were employed when the learners needed to replace segments of the MT output with alternative expressions, such as with a word (*Word Paraphrase*), a phrase (*Phrase Paraphrase*), or a clause/sentence (*Sentence/Clause Paraphrase*). Grammar correction occurred when the learners needed to post-edit for *relative pronouns, verb tense, word form, singular/plural, articles/determiner, prepositions, word order, and punctuation*. Similar to errors, more than one strategy could be coded within a sentence.

Table 2*Post-editing Strategy Scheme*

Post-editing Strategies (PES)	PES Scheme	Sample Excerpt
Word Deletion	WD	MT In addition, with the advent of <i>simple</i> payment services such as SAMSUNG PAY, PAYCO, KAKAO PAY, we can live without a wallet <i>offline</i> . [Skilled: #58]
		PS In addition, with the advent of <i>simple</i> payment services such as SAMSUNG PAY, PAYCO, KAKAO PAY, we can live without a wallet <i>offline</i> .
Phrase Deletion	PD	MT Therefore, information on the Internet <i>is information that</i> cannot be trusted completely. [Less Skilled: # 783]
		PS Therefore, information on the Internet <i>is information that</i> cannot be trusted completely.
Sentence/Clause Deletion	SD	MT The cognitive fatigue is low due to the unlimited access to the field of interest through the Internet, <i>so it is not easy to feel the burden of using it for a long time</i> . [Skilled: #63]
		PS The cognitive fatigue is low due to the unlimited access to the field of interest through the Internet, <i>so it is not easy to feel the burden of using it for a long time</i> .
Word Paraphrase	WP	MT It can satisfy the people's right to know and also function to gather <i>the will</i> of the people. [Skilled: #16]
		PS It can satisfy the people's right to know and also function to gather <i>opinions</i> of the people.
Phrase Paraphrase	PP	MT Some people also <i>fall in love with</i> shopping on the Internet. [Skilled: #167]
		PS Some people also <i>addicted with</i> shopping on the Internet. (*Note: are also addicted)
Sentence/Clause Paraphrase	SP	MT <i>The Internet is a science and technology, so if you think about the neutrality of science, it depends on who uses it, but I think it's a very good source of information</i> . [Skilled: #238]
		PS <i>Although it depends on the way it is utilized considering the neutrality of science, I think it's a very good source of information</i> .
Grammar Correction	GC	MT The Internet can be used not only <i>for</i> computers, but also <i>for</i> various means such as laptops, tablet PCs, and mobile phones. [Less skilled: #451]
		PS The Internet can be used not only <i>by</i> computers, but also <i>by</i> various means such as laptops, tablet PCs, and mobile phones.

Note. MT = Machine Translation, PS = Post-edited Sentences.

^a [] indicates proficiency and sentence number.

Data Analysis

Statistical analyses were conducted with Statistical Package for Social Sciences (SPSS). For RQ 1, a paired t-test was conducted for sentence adequacy scores between the source text (L1) and the MT output (L2) to make a relative assessment of how well the MT output reflected the original message. To examine errors, the frequencies of MT errors *per* sentence were computed as mean values to conduct tests for descriptive (i.e., mean, standard deviation) and inferential statistics. Repeated-measures one-way ANOVA was used to examine any significant differences within the MT errors. For RQ 2, the frequencies of PE strategies *per* sentence were computed for the mean values and entered for repeated-measures one-way ANOVA. Independent measures t-test was used to examine whether there were any differences in the use of PE strategies by the two proficiency groups. For RQ 3, paired t-test was conducted with sentence adequacy scores to investigate any differences between the MT output and the PEd text. Linear multiple regression was also conducted to analyze the contribution of PE strategies to final sentence adequacy scores. For the analysis, calculating the frequencies of MT errors and PE strategies *per* sentence was considered a valid way to examine the MT output and the PE strategies since this allowed the researchers to examine how common they were within a sentence.

Results

Concerning RQ 1, the principal measure for gauging the quality of MT output was sentence adequacy scores. The scores were compared to those of the source text. With paired t-test, there was a significant difference between the two adequacy values ($t = 17.734, p < .001$). This indicated that the MT output ($M = 4.49$) had not been able to completely convey the meaning expressed in the source text ($M = 4.93$). The adequacy of the MT output was also operationalized by identifying the MT errors. To calculate the mean values of the MT errors, “translation acceptable” sentences (600 sentences) were excluded from the 931 sentences that had been produced from the MT output. The remaining sentences (331 sentences) were problematic due to errors of *mistranslation* ($M = .713$), *missing words* ($M = .184$), *ungrammaticality* ($M = .175$), and *extra words* ($M = .015$) (Table 3). The mean values indicated that there had been less than a single error for every sentence according to each type of error. As seen in Table 3, there was a total of 360 errors for the 331 sentences, meaning that there was an average of 1.09 errors per sentence. According to repeated measures one-way ANOVA, the errors occurred mostly as *mistranslation*, followed by *missing words/ungrammaticality* and *extra words* ($p < .001$).

Table 3

Types of Errors in Machine-translated Sentences

(Ns = 331)	Sum	M	SD	F	Post-hoc
Mistranslation	236	.713	.479	175.086***	1 > 2*** 2 > 4***
Missing Word	61	.184	.396		1 > 3*** 3 > 4***
Ungrammaticality	58	.175	.381		1 > 4***
Extra Word	5	.015	.122		2 = 3
TOTAL	360	.272	.073		

Note. Ns = Number of sentences, M = Mean frequency of error per sentence.

*** $p < .001$.

The *mistranslation* errors were most common for sentences ($Ns = 75, 31.8\%$), followed by phrases ($Ns = 61, 25.8\%$), and words ($Ns = 50, 21.2\%$). Mistranslations also occurred due to learners' poorly written L1

sentences ($N_s = 50$, 21.2%). For *ungrammaticality*, problems of *verb tense* were most common ($N_s = 23$, 39.7%), followed by cases of *sentence fragment* ($N_s = 13$, 22.41%), *missing articles* ($N_s = 6$, 10.34%), *missing prepositions* ($N_s = 3$), *wrong verb form* ($N_s = 3$), *wrong word form* ($N_s = 3$), and *wrong word order* ($N_s = 7$, 12.07%). When the MT output was *missing words* ($N_s = 61$), they ranged from sentences ($N_s = 8$) and phrases ($N_s = 19$) to different forms of words (adverb: $N_s = 16$; noun: $N_s = 9$; adjective: $N_s = 6$; verb: $N_s = 1$; conjunctions: $N_s = 2$). *Extra words* occurred for words (adjectives, adverb, noun, verb) when there were unintended language items in the MT output.

Research question 2 was related to examining seven types of PE strategies employed for 331 sentences. The calculation was conducted by excluding 185 “translation acceptable” sentences for which PE strategies were adopted by the learners (see later for results). The sum of strategy use was 566, indicating that an average of 1.71 strategies had been used for each sentence. Mean values, which were calculated for frequency of strategy per sentence, indicated that PE strategies had been employed mainly via *paraphrase*, *deletion*, and *grammar correction*. Both *paraphrase* and *deletion* occurred at the single word, phrase, and sentence level. As seen in Table 4, *paraphrase* was the primary strategy employed by the learners to modify MT output. Repeated measures one-way ANOVA indicated no significant difference in mean values between the employment of *word paraphrase* ($M = .254$), *phrase paraphrase* ($M = .266$), and *sentence paraphrase* ($M = .257$) strategies. *Grammar correction* proved that this was less frequently employed compared to all paraphrase strategies ($p < .001$). Details of *grammar correction* indicated that the PE strategies had been employed to correct verb tense ($N_s = 13$, 43.3%), word form ($N_s = 7$, 23.3%), prepositions ($N_s = 3$, 10%), singular/plural ($N_s = 2$, 6.67%), relative pronouns ($N_s = 1$, 3.33%), article/determiners ($N_s = 2$), word order ($N_s = 1$), and punctuation ($N_s = 1$). There were also cases when words, phrases or sentences were deleted from the MT output to express the target meaning. There were no significant differences between *word deletion* and *phrase deletion* ($p = 1.00$) and between *phrase deletion* and *sentence deletion* ($p = .814$). However, the mean of *word deletion* was statistically higher than that of *sentence deletion* ($p < .01$), indicating that deletion at the single word level was more salient as a PE strategy.

Table 4

Mean of Post-editing Strategies with Multiple Comparisons

(N _s = 331)		Sum	M	SD	F	Post-hoc		
Deletion	Word Deletion	23	.069	.288	29.080***	1 = 2	2 < 4***	3 < 7***
	Phrase Deletion	13	.039	.210		1 > 3**	2 < 5***	4 = 5
	Sentence Deletion	4	.012	.109		1 < 4***	2 < 6***	4 = 6
Paraphrase	Word Paraphrase	84	.254	.507		1 < 5***	2 = 7	4 > 7***
	Phrase Paraphrase	88	.266	.506		1 < 6***	3 < 4***	5 = 6
	Sentence Paraphrase	85	.257	.458		1 = 7	3 < 5***	5 > 7***
Grammar Correction		30	.091	.336		2 = 3	3 < 6***	6 > 7***
TOTAL		327	.141	.123				

Note. N_s = Number of sentences, M = Mean frequency of strategy per sentence.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Since MT errors had triggered the use of strategies, it was necessary to analyze PE strategies with this regard. When the number of PE strategies used for each type of MT error was calculated, for mistranslations, 0.99 strategies were used for each error (233 strategies/236 errors). For missing words, 0.92 strategies were used (56/61); for ungrammaticality, 1 strategy was used (58/58), and for extra words, 1.2 strategies were used (6/5). As indicated in Table 5, Table 6, and Table 7, for errors of mistranslation, missing words, and ungrammaticality, paraphrase was most frequently employed by the learners. However, there was no significant difference between the subcategories of paraphrase strategies. There was also no significant difference between the subcategories of deletion strategies, which is not surprising considering the small number of the strategies. This suggested that, in general, the MT errors had prompted the use of paraphrase strategies; however, this did not influence the level of the paraphrase adopted (word, phrase, clause/sentence). For mistranslation errors in particular, there were significant differences between the subcategories of deletion and paraphrase strategies ($p < .001$). In a similar vein, for missing word errors, phrase and sentence deletion were respectively different from the use of phrase paraphrase strategies ($p < .05$). Paraphrase being a dominant strategy even for *ungrammaticality* errors, for which we expected grammar correction to be the dominant strategy, suggests that the type of error did not influence the learners' choice of PE strategies. In other words, this may also indicate that the learners were not able to employ the relevant strategies for grammar problems or opted for an alternative PE strategy to solve the MT error.

Table 5

Post-editing Strategies for Mistranslations

Mistranslation (N_s = 232)							
	Sum	M	SD	F	Post-hoc		
Word Deletion	13	.056	.249	23.738***	1 = 2	2 > 4***	3 < 7*
Phrase Deletion	10	.043	.224		1 = 3	2 > 5***	4 = 5
Sentence Deletion	2	.009	.093		1 < 4***	2 > 6***	4 = 6
Word Paraphrase	60	.259	.503		1 < 5***	2 = 7	4 > 7***
Phrase Paraphrase	64	.276	.511		1 < 6***	3 < 4***	5 = 6
Sentence Paraphrase	65	.280	.478		1 = 7	3 < 5***	5 > 7***
Grammar Correction	19	.082	.345		2 = 3	3 < 6***	6 > 7***
TOTAL	233	.144	.126				

Table 6*Post-editing Strategies for Missing Word*

Missing Word (N_s = 60)							
	Sum	M	SD	F	Post-hoc		
Word Deletion	6	.100	.354	5.251**	1 = 2	2 = 4	3 = 7
Phrase Deletion	2	.033	.181		1 = 3	2 < 5*	4 = 5
Sentence Deletion	1	.017	.129		1 = 4	2 = 6	4 = 6
Word Paraphrase	13	.217	.490		1 = 5	2 = 7	4 = 7
Phrase Paraphrase	19	.317	.596		1 = 6	3 = 4	5 = 6
Sentence Paraphrase	12	.200	.403		1 = 7	3 < 5*	5 = 7
Grammar Correction	3	.050	.220		2 = 3	3 < 6*	6 = 7
TOTAL	56	.133	.123				

When scrutinized for comparison, the sum and the total mean of PE strategies in Tables 5, 6, and 7 did not add up to the results in Table 4 for some PE strategies. That is, when analyzed by error type, there were extra numbers of strategies used for word deletion (3 strategies), word paraphrase (6 strategies), phrase paraphrase (6 strategies), sentence paraphrase (6 strategies), and grammar correction (8 strategies). Rather than being cases of miscalculation, this indicates that sometimes, there was more than one PE strategy used for a single MT error. Other cases were when a single PE strategy had been used to solve multiple MT errors. The pattern of results demonstrates the configuration of learners' PE strategy use, and the statistical results were able to provide a parsimonious explanation for the relationship between MT errors and PE strategies.

In principle, "translation acceptable" (TA) sentences ($N_s = 600$) would not have required PE. However, the learners in practice used PE strategies for 30.8% ($N_s = 185$) of the TA sentences. As indicated in 8, paraphrase strategies at the word, phrase, and sentence levels were the most employed particularly compared to *grammar correction* ($p < .01$) and *deletion* ($p < .001$). Even though grammar correction was not necessary for these TA sentences, it occurred when the learners PEd the sentences to make changes in the use of prepositions, tense, number, modifiers, and articles. Some of the TA sentences are presented in 9. However, since the sentences were already acceptable for communicating the target meaning, PE did not necessarily contribute toward improving the quality of the text. We interpreted them as cases of "overcorrection" or alternatives to how the learners wanted to express TA sentences (to be discussed further in the Discussion).

Table 7*Post-editing Strategies for Ungrammaticality and Extra Word*

Ungrammaticality (Ns = 58)							
	Sum	M	SD	F	Post-hoc		
Word Deletion	3	.052	.223	5.015**	1 = 2	2 < 4*	3 < 7*
Phrase Deletion	1	.017	.131		1 = 3	2 = 5	4 = 5
Sentence Deletion	1	.017	.131		1 = 4	2 < 6*	4 = 6
Word Paraphrase	16	.276	.586		1 = 5	2 < 7*	4 = 7
Phrase Paraphrase	11	.190	.396		1 = 6	3 < 4*	5 = 6
Sentence Paraphrase	13	.224	.421		1 = 7	3 = 5	5 = 7
Grammar Correction	16	.276	.586		2 = 3	3 < 6*	6 = 7
TOTAL	61	.150	.140				
Extra Word (Ns = 4)							
	Sum	M	SD	F	Post-hoc		
Word Deletion	4	.800	.837	2.571	N/A		
Phrase Deletion	0	.000	0				
Sentence Deletion	0	.000	0				
Word Paraphrase	1	.200	.447				
Phrase Paraphrase	0	.000	0				
Sentence Paraphrase	1	.200	.447				
Grammar Correction	0	.000	0				
TOTAL	6	.171	.120				

Note. N_s = Number of sentences, M = Mean frequency of strategy per sentence.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 8*Post-editing Strategies for Translation Acceptable Sentences*

N_s = 600	Sum	M	SD	F	Post-hoc		
Word Deletion	9	.015	.122	25.163***	1 = 2	2 < 4***	3 < 7***
Phrase Deletion	11	.018	.146		1 = 3	2 < 5***	4 = 5
Sentence Deletion	1	.002	.041		1 < 4***	2 < 6***	4 = 6
Word Paraphrase	77	.128	.390		1 < 5***	2 = 7	4 > 7***
Phrase Paraphrase	68	.113	.366		1 < 6***	3 < 4***	5 = 6
Sentence Paraphrase	51	.085	.285		1 = 7	3 < 5***	5 > 7***
Grammar Correction	22	.037	.197		2 = 3	3 < 6***	6 > 7*
TOTAL	239	.057	.098				

Note. N_s = Number of sentences, M = Mean frequency of strategy per sentence.

* $p < .05$, ** $p < .01$, *** $p < .001$.

In contrast, there were instances of *ignoring*, which refer to cases when MT errors are either ignored or not realized by the learner so that no further action is taken toward them. Out of the 331 sentences that required PE, 98 (29.6%) were ignored by the learners for 106 errors (mistranslations: 68, missing words: 17, ungrammaticality: 19, extra words: 2). There was a significant difference between the two proficiency groups for *ignoring* (skilled: $M = .228$, $SD = .421$; less skilled: $M = .358$, $SD = .481$; $t = -2.633$, $p < .01$), indicating that the less skilled learners found PE to be more difficult than the skilled learners.

Since PE strategies had been employed for the MT output as a whole (including the TA sentences), difference in proficiency was examined for the complete set of sentences ($N_s = 931$). When independent groups t-tests were conducted, there were significant differences between the skilled and less skilled learners in terms of *sentence deletion*, *word paraphrase*, and *phrase paraphrase* strategies where the skilled learners were consistently using more strategies than the less skilled learners (see Table 10). This indicates, particularly for *sentence deletion*, that the less skilled learners preferred to work within the sentence structures provided by the MT output rather than be involved in the process of constructing new sentences to express the target meaning. The way the skilled learners made more effort to paraphrase words or phrases indicates that they were attempting to retrieve single word or multi-word items that would most effectively express the target meaning (see Table 2 for examples).

Table 9*Examples of Post-editing for Translation Acceptable Sentences*

Machine Translation Output and Post-edited Text	
Word Paraphrase	[U]sing the Internet can waste <u>our</u> time and expose <u>us</u> to unreliable information and the risk of addiction. → [U]sing the Internet can waste <u>your</u> time and expose <u>you</u> to unreliable information and the risk of addiction.
Phrase Paraphrase	The first advantage is <u>easy information access</u> . → The first advantage is <u>an easy access to information</u>
Sentence Paraphrase	Increasingly, creativity has become more important, and more people are looking for their own differentiated talents to reveal to the world. → With the focus shifted from instilled knowledge to creativity, the number of people increased who want to explore and express their own differentiated talents to the world.
Word Deletion	Second, the Internet has <u>brought about</u> effective communication. → Second, the Internet has <u>brought</u> effective communication.
Phrase Deletion	<u>Before the Internet was used</u> , we had to go to a library to borrow books or to meet people. → <u>Before the Internet</u> , we had to go to a library to borrow books or to meet people.
Grammar Correction	...[P]eople can communicate with distant people and even get in touch with people <u>from</u> other countries. → ...[P]eople can communicate with distant people and even get in touch with people <u>in</u> other countries.

Table 10*Post-editing Strategies for Total Machine-translated Sentences*

(Ns = 931)	Skilled (Ns = 445)		Less Skilled (Ns = 486)		t	df	Sig.
	M	SD	M	SD			
Strategy NOT used†	.413	.493	.477	.500	-1.962	923.915	.050
Word Deletion	.047	.233	.023	.162	1.853	784.908	.064
Phrase Deletion	.034	.204	.019	.135	1.327	759.273	.185
Sentence Deletion	.011	.106	.000	.000	2.246	444.000	.025*
Word Paraphrase	.211	.494	.138	.379	2.527	831.442	.012*
Phrase Paraphrase	.211	.475	.128	.375	2.966	843.498	.003**
Sentence Paraphrase	.169	.381	.126	.350	1.790	902.187	.074
Grammar Correction	.058	.253	.053	.259	.293	929	.770

Note. N_S = Number of sentences, M = Mean frequency of strategy per sentence.

^a Unit used is 'sentence.'

* $p < .05$, ** $p < .01$, *** $p < .001$.

Research question 3 aimed to determine how the MT output had changed after learners' employment of PE strategies. Any progress made using learners' PE strategies could be explained by comparing sentence adequacy scores between the MT output and the PEd text. Paired t-test indicated a significant difference between the two texts (MT: $M = 4.49$, PE: $M = 4.57$; $t = 3.692$, $p < .001$). However, as indicated in 11, analysis revealed a significant increase in adequacy scores only with the skilled learners ($p < .001$) but not with the less skilled learners ($p = .479$). The results suggest that L2 language proficiency is a factor for explaining the ability to detect errors in the MT sentences and use appropriate PE strategies.

Table 11

Sentence Adequacy for Machine-translated Output and Post-edited Text

	M	SD	t	df	Sig.
<u>Skilled Learners</u> ($N_S = 444$)					
Machine-Translated Text	4.46	0.82	-4.291	443	.000***
Post-Edited Text	4.60	0.66			
<u>Less Skilled Learners</u> ($N_S = 486$)					
Machine-Translated Text	4.51	0.73	-.708	485	.479
Post-Edited Text	4.53	0.70			

Note. N_S = Number of sentences, M = Mean of sentence adequacy score.

*** $p < .001$.

To validate our claim, linear regression was conducted for the less skilled and skilled learners' texts with sentence adequacy score as the dependent variable and the frequency of PE strategies as the independent variables (predictors). The outcome of the analyses were significant regression models (Skilled: $F_{(9, 434)} = 28.715$, $p < .001$, $R^2 = .373$; Less Skilled: $F_{(8, 477)} = 58.475$, $p < .001$, $R^2 = .495$). Beta values (Skilled: .323, Less Skilled: .251) indicated an increase in sentence adequacy scores with "Strategies not used" ($p < .01$) (see Tables 12, 13). The results suggest that the learners were in fact correct in deciding not to use a PE strategy for the cases of TA sentences. On the other hand, for both groups, "Ignoring" led to a plunge in sentence adequacy scores (Skilled: Beta = $-.391$; Less Skilled: Beta = $-.439$), demonstrating that failing to notice an MT error or not taking any action will lead to detrimental effects to the quality of the text. Most noticeably, the employment of PE strategies with the less skilled learners resulted in a drop of sentence adequacy scores for word deletion (Beta = $-.164$), phrase paraphrase ($-.169$), sentence paraphrase ($-.193$), and grammar correction ($-.212$). The results demonstrate that the less skilled learners' deployment of strategies rather had negative effects on the PEd texts. However, these patterns of results did not occur with the skilled learners when they used PE strategies.

Table 12*Post-editing Strategies as Predictors of final Adequacy Scores of Skilled Learners*

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Skilled Learners (N = 30)					
(Constant)	4.564	.087		52.601	.000***
Strategy NOT used	.434	.095	.323	4.556	.000***
Word Deletion	-.045	.117	-.016	-.388	.698
Phrase Deletion	-.042	.127	-.013	-.332	.740
Sentence Deletion	-.456	.274	-.065	-1.666	.096
Word Paraphrase	.024	.065	.018	.374	.708
Phrase Paraphrase	-.128	.069	-.092	-1.838	.067
Sentence Paraphrase	-.129	.098	-.074	-1.325	.186
Grammar Correction	-.179	.107	-.069	-1.679	.094
Ignoring	-.947	.122	-.391	-7.793	.000***

Note. N_S = Number of sentences.

** $p < .01$, *** $p < .001$.

Table 13*Post-editing Strategies as Predictors of final Adequacy Scores of Less Skilled Learners*

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Less Skilled Learners (N = 27)					
(Constant)	4.638	.101		46.029	.000***
Strategy NOT used	.354	.106	.251	3.334	.001**
Word Deletion	-.710	.155	-.164	-4.589	.000***
Phrase Deletion	-.176	.184	-.034	-.958	.339
Word Paraphrase	-.148	.088	-.080	-1.674	.095
Phrase Paraphrase	-.318	.089	-.169	-3.549	.000***
Sentence Paraphrase	-.388	.105	-.193	-3.706	.000***
Grammar Correction	-.574	.109	-.212	-5.286	.000***
Ignoring	-.926	.118	-.439	-7.814	.000***

Note. N_S = Number of sentences.

** $p < .01$, *** $p < .001$.

Discussion

The recent neural MT appears to be a viable tool that can support the efforts of L2 writers to write in English. Nevertheless, the use of MT compels users to detect errors and post-edit MT output. For successful PE, the learners' ability to distinguish the different types of errors in the MT output, which is related to learners' L2 proficiency, is a prerequisite for the L2 learners in deploying appropriate PE strategies (Stapleton & Kin, 2019). With the aim of examining the pattern of PE strategies and their effects on the raw MT output, the current study on MT use for L2 writing was conducted in three stages by 1) assessing the quality of the MT output via sentence adequacy scores and classification of MT errors, 2) identifying the types and frequencies of PE strategies, and 3) evaluating the effectiveness of PE strategies by comparing the quality of MT output and PE text via sentence adequacy scores. The use of PE strategies and the outcome of the PE text were also analyzed considering the learners' L2 proficiency.

The types of errors identified by the researchers clarified that the texts needed PE, especially for 35.55% ($N_S = 331$) of the sentences. *Mistranslation* errors were significantly more frequent than those of *missing words*, *ungrammaticality*, and *extra words*. Regardless of type of MT error, the main types of strategies deployed by the learners were paraphrase strategies—at the word, phrase, and sentence levels. Examination of PE strategies by type of error indicated that it was not necessarily the characteristic of the error that prompted the use of specific PE strategies. For instance, contrary to our expectation considering the characteristic of the problem, deletion strategies were used not only for extra words but also for

mistranslation errors and paraphrase strategies were used not only for mistranslation errors but also for those related to ungrammaticality. Likewise, when the MT output provided “According to the National Statistical Office, about 85.1 percent of Koreans use* the Internet in 2015, with more than 90 percent of them spending more than half of the day using the Internet,” the learner decided to rewrite the whole sentence (i.e., sentence paraphrasing), and produced “According to the National Statistical Office, about 85.1 percent of Koreans use* the Internet in 2015. Among them, more than 90 percent spend* more than half of the day using the Internet.” However, the learner failed to notice that there was an error in the tense of “use” and even added to the existing error by writing “spend” rather than “spent”. A particular selection of PE strategies by the learners demonstrates that the L2 learners’ level of metacognitive awareness toward the MT errors was so low that they failed to choose a PE strategy that would most efficiently solve the MT error. Another explanation for some mismatch between the errors in the MT output and the PE strategies may be due to how the L2 writers’ initial attempt to post-edit raw MT output was focused on revising for meaning through improving word choice or expressing cultural concepts. With more time and training provided for PE, some learners may have noticed the severity of grammar errors and the requirement to employ appropriate PE strategies.

The use of paraphrase strategies indicated that the learners most often conveyed their intended message using lexical items and syntactic structures that were familiar to them. This sometimes happened even when the MT provided TA sentences. The pattern of results demonstrates that a necessary skill set for PE includes learning to evaluate the MT suggestions and the usability of segments, learning to make only necessary changes, and discarding suggestions that require too many changes (Pym, 2013). In line with how the skilled learners’ use of PE strategies was significantly more effective than those of the less skilled learners, the extent to which the learners can successfully detect errors seems to be related to L2 proficiency (Chung, 2020; Kol et al., 2018; Stapleton & Kin, 2019). That is, although MT may have the potential to provide students with viable alternatives to their L2 texts, only students who have reached the threshold level of proficiency (e.g., high) may be able to use its output to improve their L2 writing (Chon et al., 2021; Lee, 2020; Tsai, 2019). That is, the skilled learners demonstrated adeptness at matching PE strategies to the MT errors that they detected, while the less skilled learners seemed to lack the metacognitive knowledge on task requirements needed to select appropriate strategies.

Most noticeably, the use of MT requires learners to understand that there may be errors in the MT output, their source being mistranslation, missing words, ungrammaticality, extra words, or a combination of these. For instance, the sentence “As you can see by subway or bus, the scenery that most people look into smartphones is no stranger to us” was marked as being problematic due to mistranslation and ungrammaticality errors when the learners’ target was “As you can see by just taking the subway or bus, the scenery where most people are looking at their smartphones is no longer unfamiliar to us.”

Errors may not only arise from MT but also from cross-linguistic differences, which occurred for 7.73% ($N_s = 72$) of the sentences. The first case may be when the learners omit pronouns in the Korean sentences since the Korean language is a pro-drop language (Kim, 2000). Other language-specific problems arise when four-character Chinese idioms are used as the source since they are not directly translatable. The Chinese idiom “이역만리 (異域萬里)” referring to “A very remote place in another country” was non-translatable using GT and was treated in a way that only provided the phonetic transcription as in “Yeokman-ri.” To tackle this error, the learner paraphrases by using the phrase “so far away” when writing. Language-specific acronyms can also be a source of error. At the time the learners were using GT, the Korean acronym, coined partially from English, “재테크” [jaetekeu] (financial technology, “財 + tech”) became a source of MT error when the learner was only able to obtain “jettech.” The learner writes “[I]t is also easy to access various fields such as invest, health, clothes, etc. with just one search.” The sentence is PED; however, the form of the word is not quite correct as it should have been “investment.”

Even though the L2 learners were asked to improve the MT output by PE, the analysis of learners’ PE strategies revealed that they were unconscious of or deliberately *ignored* the MT errors (perhaps because of their judgment that the MT output was better than what they could do). This may indicate that the learners’

level of linguistic knowledge essential for identifying MT errors and strategic competence for deploying appropriate PE strategies were not always present. Knowing this, it seems that learners chose to rely on the MT translations. While communication strategies (CS) are “potentially conscious plans for solving what to an individual presents itself as a problem in reaching a particular communicative goal” (Færch & Kasper, 1983, p. 36), the CS framework may provide the theoretical explanation for how *ignoring* can be considered a type of reduction strategy in L2 writing (Chon, 2008). On the other hand, paraphrase is an achievement behavior through which the learners can use alternative linguistic means to express the target message. In fact, the significant difference in using ignoring between the skilled and less skilled learners demonstrates that it is the lower level learners who need explicit attention in taking a more proactive role in the use of PE strategies.

Among the writers, there were instances of *overcorrection*, which occur when sentences are PEd even when the MT output offers the correct translation. When overcorrection did not deteriorate the quality of the text, it was referred to as “positive overcorrection.” When the overcorrection led to a deterioration in the quality of the MT output, we referred to this as “negative overcorrection.” As such, positive correction did not result in deduction of sentence adequacy scores whereas negative correction did. Among the skilled learners, positive overcorrection occurred for 11.5% ($N_s = 51$) of the MT sentences; for the less skill skilled learners, it occurred for 9.7% ($N_s = 47$) of the sentences. In comparison, negative overcorrection occurred for 6.1% ($N_s = 27$) of the MT sentences for the skilled learners, whereas this occurred for 8.4% ($N_s = 41$) of the sentences for the less skilled learners. Calculation of mean values indicated that positive overcorrection was more frequent among skilled learners, whereas negative overcorrection occurred more frequently with the less skilled learners, but these were not significantly different.

Most of all, the occurrence of negative overcorrection demonstrates that the current neural MT may be far more accurate than what L2 learners can handle. The results, especially with the less skilled (high-intermediate) learners, demonstrate that they were not always able to conduct PE successfully when correcting errors. In contrast, the skilled (advanced) learners in our study were able to improve their text through PE similar to how professional translators would use MT for assistance. For the less skilled learners, the MT output even provided better translations of what the learners would have wanted to say as observed in the cases of negative overcorrection. For instance, the MT produces “[T]he development of the Internet has brought us into contact with a variety of cultures” (Source: The development of the Internet has made it possible to encounter various cultures). The learner post-edits by rephrasing the whole sentence to produce “[T]he development of the Internet knows more a variety of cultures.” As a result of the correction, there was a deduction in the sentence adequacy score for this senseless sentence.

Although it is not always possible to extrapolate the exact cause of overcorrection, the researchers were able to find in the process of analyzing the PEd texts that the learners tended to prefer “lexical teddy bears” (Hasselgren, 1994), that is, words that students felt safe, and sentence structures that were familiar. For instance, when MT produced “Of course, books made of paper have long been considered a treasure trove of knowledge, but now the information contained in paper books can be found on the Internet,” the learner instead decides to write “books made of paper have long been considered a treasury of knowledge” when not assured about the word “trove.” Another example is when MT produced “The Internet is doing a good job of reducing *information asymmetry* among people.” For PE, the learner retrieves the familiar collocation “information gap” instead.

In sum, although MT has proven useful in assisting L2 writers, the results of the study explain how using MT involves a range of competencies in rectifying MT errors. There is limited focus on competencies necessary for PE in second language writing/acquisition, but scholars of translation studies have conceptualized categories of PE competences. Rico and Torrejón (2012) present three categories of PE competences: *linguistic skills*, *core competences*, and *instrumental competences*. They divide core competences into attitudinal or psycho-physiological competence and the strategic competence required to arrive at informed decisions when choosing among alternatives (p. 170). On the other hand, instrumental competences include knowledge and capabilities of MT systems, skills in assessing corpus quality, while

linguistic skills cover language and textual skills as well as cultural knowledge. The skills presented by Pym (2013) are grouped under three headings: “learning to learn,” “learning to trust and mistrust data,” and “learning to revise.” Learning to revise comprises the detection and correction of errors, substantial stylistic revising, and the ability to revise and review in teams (Pym 2013, p. 496). Kontinen et al. (2020) also note that metalinguistic knowledge falls under the category of strategic subcompetences. As observed in our study, being able to detect the correct types of MT errors and adopt appropriate PE strategies would be an important competence of metalinguistic knowledge when using MT to express the learners’ target message. Given sufficient time and training on PE strategies, learners may decide to post-edit their text for improved cohesion and clarity by PE beyond sentential level.

Conclusion

While the use of MT is recently on the rise within L2 writing, the present study contributes to this development by providing a taxonomy of MT errors and PE strategies by grounding its analyses on MT output and L2 learners’ PEd text. To judge the quality of the MT output and the PEd text, sentence adequacy scores were assessed to grasp a microscopic perspective of how L2 learners deal with the MT output at the sentential level. While the most common MT errors occurred from *mistranslations* of the source text, *paraphrasing* was the dominant strategy employed by the L2 learners for most MT errors with the exception of ungrammaticality errors. However, it was only with the skilled learners that the PE strategies functioned to improve the quality of the PEd text. In comparison to the less skilled learners, the skilled learners employed significantly more *sentence deletion*, *word paraphrase*, and *phrase paraphrase* PE strategies. A majority of the less skilled learners’ employment of PE strategies led to a decline in sentence adequacy scores. The results demonstrate that the development of AI and the inseparable use of online reference materials for L2 writing including MT, are likely to change our mental representation of how L2 learners attend to writing.

For effective use of MT for L2 writing, some pedagogical implications are suggested. Training should be conducted for learners to be able to deal with the PE process (Cancino & Panes, 2021). There is need to incorporate lessons based on authentic examples from the MT output for detecting MT errors. L2 learners can be informed on the types of PE strategies to improve their metacognitive awareness on different types of MT errors and form-focused activities can be provided to help L2 learners develop grammatical and lexical accuracy. When learners are trying to correct MT errors, not only the dictionary, but also the concordancer (e.g., *Corpus of Contemporary American English*) can allow learners to make more appropriate choice of words, find fixed multi-word expressions, collocates, and information on grammar (preposition, articles). Alternatively, learners can be instructed to become more sensitive toward the sources (e.g., pro-drop languages, polysemous words, acronyms) of MT errors with particular language pairs of interest.

This study has a few limitations as well as implications for future studies. While research on PE is in its early stage within L2 writing, further qualitative and quantitative studies should be conducted within this area. Studies such as with think-aloud are needed on L2 learners’ MT error detection and PE process. There is also a need to investigate how different means of instructional support during the PE process can aid learners. For instance, learners can have access to online resources, peers for collaboration, and teacher feedback on the MT errors before PE. Overall, there needs to be more research on the strategies that learners use during MT use which can be highlighted as an important element of L2 writing competence, especially for L2 learners who need to be prepared for the challenges of working with evolving technologies.

Acknowledgements

We sincerely thank the three anonymous reviewers who have contributed their time, expertise, and effort to the peer review process.

References

- Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 297–317). John Benjamins.
- Barreiro, A. M. (2008). *Make it simple with paraphrases: Automated paraphrasing for authoring aids and machine translation* [Unpublished doctoral thesis]. New York University.
- Blagodarna, O. (2020). Acquisition of post-editing competence: Training proposal and its outcomes. In *Fit-for-market translator and interpreter training in a digital age edition*, p.p. 85–101 .
<https://vernonpress.com/file/10766/4833e92b981ffcc497a9d5137a0b12aa/1576153406.pdf>
- Bowker, L., & Ciro, J. B. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Publishing.
- Cancino, M., & Panes, J. (2021). The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System*, 98, 1–11.
<https://doi.org/10.1016/j.system.2021.102464>
- Chon, Y. V. (2008). The electronic dictionary for writing: A solution or a problem? *International Journal of Lexicography*, 22(1), 23–54. <https://doi-org.eres.qnl.qa/10.1093/ijl/ecn034>
- Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity, and errors. *System*, 96, 1–12.
<https://doi.org/10.1016/j.system.2020.102408>
- Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. *Journal of Asia TEFL*, 17(1), 182–193. <http://dx.doi.org/10.18823/asiatefl.2020.17.1.11.182>
- Chung, E. S., & Ahn, S. (2021). The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer Assisted Language Learning*, 35(9), 2239–2264. <https://doi.org/10.1080/09588221.2020.1871029>
- Costa, Â., Ling, W., Luís, T., Correia, R., & Coheur, L. (2015). A linguistically motivated taxonomy for machine translation error analysis. *Machine Translation*, 29(2), 127–161.
<https://doi.org/10.1007/s10590-015-9169-0>
- Crossley, S. A. (2018). Technological disruption in foreign language teaching: The rise of simultaneous machine translation. *Language Teaching*, 51(4), 541–552.
<https://doi.org/10.1017/S0261444818000253>
- Daems, J., Vandepitte, S., Hartsuiker, R., & Macken, L. (2015). The impact of machine translation error types on post-editing effort indicators. In S. O'Brien & M. Simard (Eds.), *Proceedings of 4th Workshop on Post-Editing Technology and Practice* (pp. 31–45). Association for Machine Translation in the Americas.
- Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google Translate. *Foreign Language Annals*, 51(4), 779–795.
<https://doi.org/10.1111/flan.12366>
- Færch, C., & Kasper, G. (1983). On identifying communication strategies in interlanguage production. In C. Færch, & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 210–238). Longman Pub Group.
- Ferris, D. (2011). *Treatment of error in second language student writing*. University of Michigan Press.
- Garcia, I. (2011). Translating by post-editing: Is it the way forward? *Machine Translation*, 25(3), 217–237. <https://doi.org/10.1007/s10590-011-9115-8>

- Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, 37, 112–121. <http://dx.doi.org/10.1016/j.esp.2014.09.001>
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–258. <https://doi.org/10.1111/j.1473-4192.1994.tb00065.x>
- Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation?: A product and process study. *The Journal of Specialised Translation*, 31, 60–86.
- Kim, Y. J. (2000). Subject/object drop in the acquisition of Korean: A cross-linguistic comparison. *Journal of East Asian Linguistics*, 9(4), 325–351. <https://doi.org/10.1023/A:1008304903779>
- Kliffer, M. D. (2008). Post-editing machine translation as an FSL exercise. *Porta Linguarum*, 9, 53–67. <https://doi.org/10.30827/Digibug.31745>
- Kol, S., Schcolnik, M., & Spector-Cohen, E. (2018). Google Translate in academic writing courses? *The EuroCALL Review*, 26(2), 50–57. <https://doi.org/10.4995/eurocall.2018.10140>
- Kontinen, K., Salmi, L., & Koponen, M. (2020). Revision and post-editing competences in translator education. In M. Koponen, B. Mossop, I. S. Robert, & G. Scocchera (Eds.), *Translation revision and post-editing: Industry practices and cognitive processes* (pp. 187–202). Routledge.
- Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *Journal of Specialised Translation*, 25, 131–148.
- Koponen, M., Wilker, A., Luciana, R., & Specia, L. (2012). Post-editing time as a measure of cognitive effort. In S. O'Brien, M. Simard, & L. Specia (Eds.), *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice* (pp. 11–20). Association for Machine Translation in the Americas.
- Koponen, M., & Salmi, L. (2015). On the correctness of machine translation: A machine translation post-editing task. *Journal of Specialised Translation*, 23, 118–136.
- Lacruz, I. (2017). Cognitive effort in translation, editing, and post-editing. In J. Schwieter, & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 386–401). Wiley-Blackwell.
- Lacruz, I., Shreve, G. M., & Angelone, E. (2012). Average pause ratio as an indicator of cognitive effort in post-editing: A case study. In S. O'Brien, M. Simard, & L. Specia (Eds.), *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice* (pp. 21–30). Association for Machine Translation in the Americas.
- Le, Q. V., & Schuster, M. (2016). *A neural network for machine translation, at production scale*. <https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html>
- Lee, I. J. (2018). A quality comparison of English translations of Korean literature between human translation and post-editing. *International Journal of Advanced Culture Technology*, 6(4), 165–171. <https://doi.org/10.17703/IJACT2018.6.4.165>
- Lee, S. M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, 33(3), 157–175. <https://doi.org/10.1080/09588221.2018.1553186>
- Lee, S. M. (2022). An investigation of machine translation output quality and the influencing factors of source texts. *ReCALL*, 34(1), 81–94. <https://doi.org/10.1017/S0958344021000124>
- Lee, S. M., & Briggs, N. (2021). Effects of using machine translation to mediate the revision process of Korean university students' academic writing, *ReCALL*, 33(1), 18–33. <https://doi.org/10.1017/S0958344020000191>

- Moore, D. S., & McCabe, G. P. (2003). *Introduction to the practice of statistics* (4th ed.). W. H. Freeman and Company.
- Moorkens, J. (2018). What to expect from neural machine translation: A practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, 12(4), 375–387. <https://doi.org/10.1080/1750399X.2018.1501639>
- Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, 21(1), 29–49. <https://doi.org/10.1080/09588220701865482>
- Pym, A. (2013). Translation skill-sets in a machine-translation age. *Meta*, 58(3), 487–503. <https://doi.org/10.7202/1025047ar>
- Rico Pérez, C., & Torrejón, E. (2012). Skills and profile of the new role of the translator as MT post-editor. *Tradumàtica*, 10 166–178. <https://revistes.uab.cat/tradumatica/article/view/n10-torrejon-rico/pdf>
- Stapleton, P., & Kin, B. L. K. (2019). Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes*, 56, 18–34. <https://doi.org/10.1016/j.esp.2019.07.001>
- Sun, D. (2017). Application of post-editing in foreign language teaching: Problems and challenges. *Canadian Social Science*, 13(7), 1–5. <http://dx.doi.org/10.3968/9698>
- Temnikova, I. (2010). Cognitive evaluation approach for a controlled language post-editing experiment. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner & D. Tapias (Eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation* (pp. 3485–3490). European Language Resources Association.
- Tsai, S.-C. (2019). Using Google Translate in EFL drafts: A preliminary investigation. *Computer Assisted Language Learning*, 32(5-6), 510–526. <https://doi.org/10.1080/09588221.2018.1527361>
- Vilar, D., Xu, J., D'Haro, L. F., & Ney, H. (2006). Error analysis of machine translation output. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odiijk & D. Tapias (Eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation* (pp. 697–702). European Language Resources Association.
- Wallwork, A. (2016). Using Google Translate and analysing student-and GT-generated mistakes. In A. Wallwork (Ed.), *English for academic research: A guide for teachers* (pp. 55–68). Springer.

Appendix A. Writing Prompt

Please refer to the table below to choose your position on the pros and cons of using the Internet and write your opinion in around 300 words by adding one reason to the given two reasons. (Originally conducted in Korean)

Advantages	Disadvantages
Introduction ↓ 1. access to information 2. efficient communication 3. <u>your own idea</u> ↓ Conclusion	Introduction ↓ 1. waste of time 2. unreliable information 3. <u>your own idea</u> ↓ Conclusion

Appendix B. Rubric on Sentence Adequacy for Machine-translated Output and Post-edited Text

Sentence Adequacy Score	<i>After machine translation/post-editing, how much of the meaning expressed in the source text (L1) appears in the translated sentence?</i>	Description of sentence
1	None of it	The meaning of the source text (L1) is not conveyed due to serious problems in word choice, grammar and, word order.
2	Little of it	There is a limited display of source text (L1) when errors in word choice, grammar, or sentence structure convey misleading information.
3	Some of it	The meaning of the source text (L1) is conveyed, but flaws in word choice, grammar, or sentence structure result in lack of clarity and obscure meaning.
4	Most of it	The meaning in the source text (L1) is conveyed, though it may have minor lexical or grammatical errors.
5	All of it	Meaning in the source text (L1) is completely conveyed.

About the Authors

Dongkwang Shin received his PhD in applied linguistics from Victoria University of Wellington and is currently an associate professor at Gwangju National University of Education, South Korea. His research interests include corpus linguistics, CALL, and AI-based language learning.

E-mail: sdhera@gmail.com

Yuah V. Chon received her PhD in English Language Teaching from University of Essex and is currently a professor at Hanyang University, South Korea. Her research includes learner strategies, vocabulary research, and technology-enhanced language learning. All correspondence regarding this publication should be addressed to her.

E-mail: vylee52@hanyang.ac.kr