Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

2024, Volume 28, Issue 1
pp. 1–23

ARTICLE

Second language processing of errors in Korean-to-English machine-translated output

Eun Seon Chung, Hankyong National University

Abstract

While previous investigations on online machine translation (MT) in language learning have analyzed how second language (L2) learners use and post-edit MT output, no study as of yet has investigated how the learners process MT errors and what factors affect this process using response and reading times. The present study thus investigates L2 processing of MT errors that are caused by syntactic, morphological, and semantic differences between the source and target language and also examines how L2 proficiency and visual display affect this process. Forty-seven Korean learners of English participated in an acceptability judgment task in which they read a Korean sentence and then its translated counterpart in English and had to judge the accuracy of the translated sentence on a four-point scale. The response latencies for the accuracy judgment as well as the total reading times of source and target sentences were measured. The results revealed that (a) learners generally find it harder to reject mistranslations than to accept correct translations, (b) high and low proficiency learners focus on different aspects of language when processing translated output, and (c) constant visual access to the source text does not facilitate but rather interferes with processing MT errors.

Keywords: *Machine Translation, Error Processing, L2 Proficiency, Visual Display*

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

APA Citation: Chung, E. S. (2024). Second language processing of errors in Korean-to-English machine-translated output. *Language Learning & Technology*, *28*(1), 1–23. https://hdl.handle.net/10125/73550

Introduction

With the widespread availability and rapid improvements in the quality of online machine translation (MT) such as Google Translate, many second language (L2) learners report using MT as a supplementary resource in language learning (Briggs, 2018; Chung & Ahn, 2022), and numerous studies have examined the educational value of using MT especially in L2 writing (Chung & Ahn, 2022; Garcia & Pena, 2011; Kol et al., 2018; S.-M. Lee, 2020a; O'Neill, 2016; Tsai, 2019). Recent works compare students' self-written L2 texts to written products that are aided by MT and generally report advantages of MT use; the use of MT not only resulted in a higher level of accuracy (Chung & Ahn, 2022; S.-M. Lee, 2020a; Lee & Briggs, 2021; O'Neill, 2016; Tsai, 2019), fluency (Garcia & Pena, 2011; Kol et al., 2018), and vocabulary (Fredholm, 2019; Kol et al., 2018) but also increased learners' metalinguistic awareness (Correa, 2014) and promoted self-directed learning (Garcia & Pena, 2011). Most of these works, however, examined the final product of MT use and did not investigate how MT output is analyzed or processed by the learners. It is important to examine how learners detect and process errors in the MT output, but this topic has received limited attention in the field. In fact, no study has systematically investigated how learners process certain types of errors in the MT output and the factors that affect this process, and therefore the current study aims to delve into such questions.

Despite the marked gains in the accuracy of Neural Machine Translation (NMT) in recent years, errors can still be found—especially for input that contains pragmatic expressions and discourse subtleties (Ducar & Schocket, 2018), cultural references and idioms (K. Kim, 2018), and complex embedded sentences (Kim

& Lee, 2017). Moreover, translation quality relies on the linguistic differences between the source language (SL) and the target language (TL), and MT output often needs to be post-edited or revised to be of acceptable quality (Chon et al., 2021). In fact, using MT output necessarily requires the ability to analyze and identify errors in the output and to edit/revise them accordingly, but it has been pointed out that some learners, especially those with low L2 proficiency, may heavily rely on MT and accept its output including its errors without critical analysis (Chung, 2020; S.-M. Lee, 2020a). Such concerns are not only confined to low proficiency learners, as undergraduate students majoring in translation were also found to mainly focus on micro-level errors without taking the context into account (S.-B. Lee, 2018). Teachers who are hesitant to allow this technology in L2 classrooms worry that learners' excessive reliance on and acceptance of MT output would hinder rather than facilitate language learning by depriving them of higher-order thinking processes (Crossley, 2018).

In light of these concerns, the present study examines how learners process MT errors that are caused by semantic, syntactic, or morphological differences in SL and TL and whether certain errors are more difficult to process than others. Research in L2 sentence processing reports that L2 learners often exhibit difficulty integrating morphosyntactic information and tend to rely on lexical cues (Clahsen & Felser, 2006; Roberts & Felser, 2011), but it is not yet clear whether this characteristic of L2 processing also applies to the processing of L1-to-L2 translated sentences. That is, would the learners show greater difficulty detecting and processing errors that originate from crosslinguistic differences in morphology and syntax than in semantics? In addition to examining L2 processing of MT errors of different linguistic categories, the study investigates the effect of L2 proficiency, which has been found to play an important role in MT use (Chung & Ahn, 2022), error detection (Kol et al., 2018; Lee & Briggs, 2021) and post-editing of MT output (Chung, 2020; Garcia & Pena, 2011). Given such wide-reaching effects of L2 proficiency, we can expect it to be an important factor that determines how the learners process MT errors. Furthermore, the study explores the effect of visual display on processing translated output, which has not been examined in previous work. Studies in translation processing have revealed that SL and TL interact in a parallel manner in translation (Chmiel & Lijewska, 2019; Macizo & Bajo, 2006; Seeber & Kerzel, 2011), but it is unclear whether constant visual access to the source text would facilitate or interfere with processing the MT output. In sum, the present study examines the processes involved in L2 learners' analyses of MT errors by investigating how Korean learners of English process different types of errors in Korean-to-English MT translations and how L2 proficiency and visual display affect this process. The results of this study are expected to increase our understanding of how L2 learners detect and identify errors in MT output in real-time, which has implications for the learners' ability to post-edit and integrate MT output in L2 production.

Literature Review

Machine Translation Errors

To examine MT errors of different linguistic categories, it is important to observe how investigations in the field of translation and post-editing research have analyzed MT errors. The quality of MT output has often been measured by estimating the amount of post-editing (PE) effort. Studies have assessed PE effort by asking bilinguals to rate the quality of MT sentences based on how much of the sentence needs to be edited using Likert scales (Specia et al., 2010) or using automatic metrics like BLEU (Papineni et al., 2002) that can evaluate a MT system's performance. Specific types of errors were linked to increased PE effort: edits involving word order, incorrect complex/compound sentences, incomplete sentences, or mistranslated idioms that were more cognitively demanding took longer to edit than word-level errors such as changing the word form or substituting lexical items (Koponen et al., 2012; Temnikova, 2010). To examine MT errors, previous studies have used various error taxonomies (Daems et al., 2017; Lee & Briggs, 2021; Vilar et al., 2006; Yamada, 2019). The linguistic category classification is based on general linguistic categories such as grammar, lexis, and morphology and sub-categories within (e.g., auxiliaries, prepositions, articles), whereas the surface structure taxonomy indicates how the translation has been revised (e.g., omission, addition, mis-ordering). Many studies have combined the two taxonomies into a single bidimensional taxonomy. For example, Yamada (2019) first classified the errors into linguistic categories which were then subcategorized using the surface structure taxonomy. Lee and Briggs (2021) identified error types that repetitively appeared in students' L2 texts and used categories in both linguistic and surface structure taxonomies to classify errors such as verb tense, articles, prepositions, insertions/deletions, and substitutions. Alternatively, Daems et al. (2017) analyzed translation quality from two perspectives—*acceptability* that measures adherence to target language norms (grammar, syntax, lexicon, spelling etc.) and *adequacy* that evaluates adherence to source text norms (meaning shifts and word sense). As such, no universal classification for MT errors has been proposed, and the error taxonomies in previous research are often specific to the languages observed. In fact, frequent error categories are largely attributed to the idiosyncrasies of the language pair and their differences. In order to accurately examine how learners process certain type of errors, it is thus important to classify the errors based on distinctive linguistic differences between SL and TL.

Notwithstanding the significant improvements in MT accuracy since 2016 with the advances in NMT, it has been observed that certain crosslinguistic differences between English and Korean can still lead to notable mistranslations, especially when translating single sentences with no supporting contexts (K. Kim, 2018; S.-M. Lee, 2020b; S.-M. Lee, 2021; Park, 2017). Because this study examines MT errors in Korean-to-English translations, it is important to identify the errors that frequently occur in translations of this specific language pair, and the errors in the present study were divided into morphological, syntactic, and semantic errors that arise from crosslinguistic differences at these levels. Notable differences between English and Korean in the use of case markers, tense/aspect, pro-drop, adnominal clauses, and scope of negation were used as morphological and syntactic categories, and semantic ambiguities arising from homonyms and proverbial or idiomatic expressions were used in the semantic category to examine how Korean learners of English analyze MT errors (see Appendix A for detailed description and examples of these errors).

L2 Proficiency

Many studies have examined the effect of L2 proficiency on MT use and found advanced learners to be capable of detecting and correcting a higher number of errors than lower proficiency learners (Chung, 2020; Chung & Ahn, 2022; Kol et al., 2018; Lee & Briggs, 2021). In a preliminary study, Kol et al. (2018) conducted an awareness task to assess students' awareness of MT errors in Hebrew-to-English translated sentences and a correction task in which the identified errors were revised. They found that lower proficiency learners could identify only about half of the mistakes whereas the higher proficiency learners identified 73% and corrected 87% of the mistakes that were identified. In a more comprehensive study, Chung (2020) examined the effect of proficiency in learners' post-editing patterns of MT text and found a significantly higher number of corrections by learners with high proficiency scores. Advanced learners were better able to correct errors above the word-level, reconstruct whole sentences, and detect subtle semantic differences in the two languages, whereas beginning learners often simply adopted the MT output without making significant changes. The study also found that the higher the proficiency, the more willing the learners were to write their own original translations of the source text on the blank space instead of making direct revisions on the MT text or between the lines of the text. Such moderating effect of L2 proficiency on learners' analysis and correction of MT output is echoed by Lee and Briggs (2021) who found higher language proficiency groups to correct a higher percentage of errors and to be more effective in their detection and treatment of errors when they were asked to compare their L2 compositions to that of MT output.

Numerous studies have revealed that proficiency not only affects how learners treat the MT errors but also how they utilize MT outputs in their L2 texts (Chung & Ahn, 2022; Fredholm, 2019; S.-M. Lee, 2020a; Tsai, 2019). For example, Chung and Ahn (2022) investigated the effect of MT use on linguistic features in L2 writing and found that high and low proficiency groups both showed improvements in accuracy but were affected by MT use in different ways with regards to syntactic and lexical complexity. High proficiency learners showed greater gains in lexical complexity after MT use, while low proficiency learners gained higher scores in syntactic complexity. The moderating effect of L2 proficiency on MT use is uncontroversial, but previous findings are mixed regarding the relative advantages by learners of diverse L2 ability levels; some suggest that using MT is more beneficial for advanced students than for beginners (Stapleton & Kin, 2019; Tsai, 2019), while others argue that MT offers greater advantages for beginning than advanced learners (Briggs, 2018; Chung & Ahn, 2022; Garcia & Pena, 2011). Although the picture regarding the benefits of MT use for different proficiency levels is as yet rather scattered, what is clear in the literature is that advanced learners are more efficient in analyzing MT output and correcting errors than lower proficiency learners.

Visual Presentation

No study has explored the effects of visual presentation of SL and TL on processing translated output, and it remains unclear whether constant visual access to the source text would facilitate or interfere with processing the MT output. The presence of a source text could help reduce cognitive load by making the information available but could also lead to heavy reliance on the verbal form and linguistic features of SL when processing the output. Addressing this question is important as it can provide a window into the manner in which learners employ linguistic knowledge in the two languages when processing translated output.

Previous studies in translation processing provide ample evidence that translation involves a process that establishes linguistic matches between the two languages in a parallel manner and that greater processing resources are needed when translation involves greater transformations (Chmiel & Lijewska, 2019; Ruiz et al., 2008; Seeber & Kerzel, 2011; Viezzi, 1989). In experimental studies that explore the processes of sentence-level translation, there are two opposing views regarding how the three cognitive processes in translation—SL comprehension, reformulation, and TL production—operate. The vertical perspective of translation (Seleskovitch, 1976) states that translation is a serial and modular process with no direct interactions between the two languages; comprehension of SL produces a representation that lacks a specific linguistic form, and this deverbalized message is then restructured and recoded in TL in a sequential order. According to this view, there are no overlaps between the three processes in translation and different linguistic levels. Alternatively, the horizontal perspective (Gerver, 1976) states that translation involves direct and interactive processes between SL and TL in real-time even before SL comprehension has been completed. Lexical and morphosyntactic properties of both SL and TL are activated upon reading the SL, as comprehension and reformulation take place simultaneously. Previous research finds overwhelming support for the horizontal perspective (Chmiel & Lijewska, 2019; Macizo & Bajo, 2006; Ruiz et al., 2008; Seeber & Kerzel, 2011), and a substantial body of evidence points to co-activation of both languages during bilingual comprehension and production (Kroll et al., 2006; Marian et al., 2003). In accordance with such evidence, it can be predicted that the mechanisms of the two working languages are simultaneously active and interact in a parallel manner when processing the source text and its MT output. The learners are expected to establish linguistic matches between the two languages early on in SL comprehension before encountering the MT output. What is unclear, however, is whether constant visual access to the source text has an effect on how the learners process the translated output, which is a question that will be examined in the present paper.

In this context, the present study aims to gain a greater understanding of the processes involved in L2 learners' analysis of MT errors by addressing the following research questions:

1. How do Korean learners of English perceive and process errors in Korean-to-English MT output that arise from morphological, syntactic, and semantic differences between the two languages?
2. Does L2 proficiency have a moderating effect on how L2 learners process these errors?
3. Does having constant visual access to the source text facilitate or interfere with processing MT output?

Method

Participants

Forty-seven Korean L2 learners of English (*M* age = 23.5; range = 20–29; 33 female and 14 male) who are undergraduate or graduate students at a university in Seoul participated in the study. They learned about the study through the university's recruiting website and volunteered to be a part of the study. All reported that Korean is their first language, and their length of stay in a foreign country before the age of 15 did not exceed six months. They were each given a gift certificate that is worth KRW 5,000 for their participation.

A cloze test, which has been proven to be a reliable measure of learners' global language proficiency (Fotos, 1991), was administered to measure participants' proficiency. The text of the cloze test was taken from a passage in *American Kernel Lessons: Advanced Students' Book* (O'Neill et al., 1981) that has been adapted in previous L2 studies (Ionin et al., 2013). The test consisted of 40 multiple-choice questions and was generally completed in 15–20 minutes. The participants were divided into two proficiency groups based on the cloze test scores (out of 40 points; range = 22–39, mean = 30.74, median = 32); 24 participants with scores equal to or greater than the median were placed in the high proficiency group ($M = 33.83$, $SD = 1.88$), and 23 participants with scores below the median were placed in the low proficiency group ($M = 27.52$, $SD = 2.5$). The results of an independent samples *t*-test showed that the difference in the scores of the two proficiency groups was statistically significant ($t(45) = 9.67$, $p < .001$).

Materials

The main experimental task was an Acceptability Judgment Task (AJT) in which participants read a Korean sentence and then its translated counterpart in English and had to judge the accuracy of the translated sentence on a four-point scale (*Very accurate* '4,' *Accurate* '3,' *Inaccurate* '2,' *Very inaccurate* '1'). The task consisted of 36 Korean sentences and their machine translated English counterparts, half of which were accurate and the other half inaccurate. The accurate translations were translated and proofread by four advanced bilinguals who were native speakers of Korean with near-native proficiency in English, and the inaccurate translations were machine-translated sentences with errors in the three linguistic categories of interest (morphological, syntactic, semantic). Half of the inaccurate sentences were translated using Google Translate (https://translate.google.com) and the other half using Naver Papago (https://papago.naver.com); two MTs were used because they are both widely used by Korean learners of English and using both could encompass various types of errors that learners encounter when using MT. Of the 36 experimental items (Korean-English sentence pairs), there were 12 items for each linguistic category and three error types per category (i.e., four items for each error type). The morphological error types were based on crosslinguistic differences in the use of the nominative case, adverbial case, and tense/aspect. Syntactic error types were based on crosslinguistic differences in the scope of negation in coordinated constructions, pro-drop, and modification of adnominal clauses. Semantic error types pertained to homonyms, idioms, and proverbial expressions. All of these MT errors based on linguistic differences between the SL and TL resulted in meaning shifts but did not violate target language norms in grammar, sentence structure, spelling, or punctuation (see Appendix A for description and examples). The Korean sentences were controlled for length (30–35 characters) and were all embedded complex sentences that contained an independent clause and a dependent clause. To additionally investigate the effect of visual display, the sentences were either shown together on the same screen (synchronous display) or shown separately on consecutive screens (asynchronous display). The accuracy of the translated English sentence (accurate vs. inaccurate) and the display type (synchronous vs. asynchronous) were counterbalanced across participants and items resulting in four counterbalanced presentations lists with 11–12 participants randomly assigned to each list (see Appendix B).

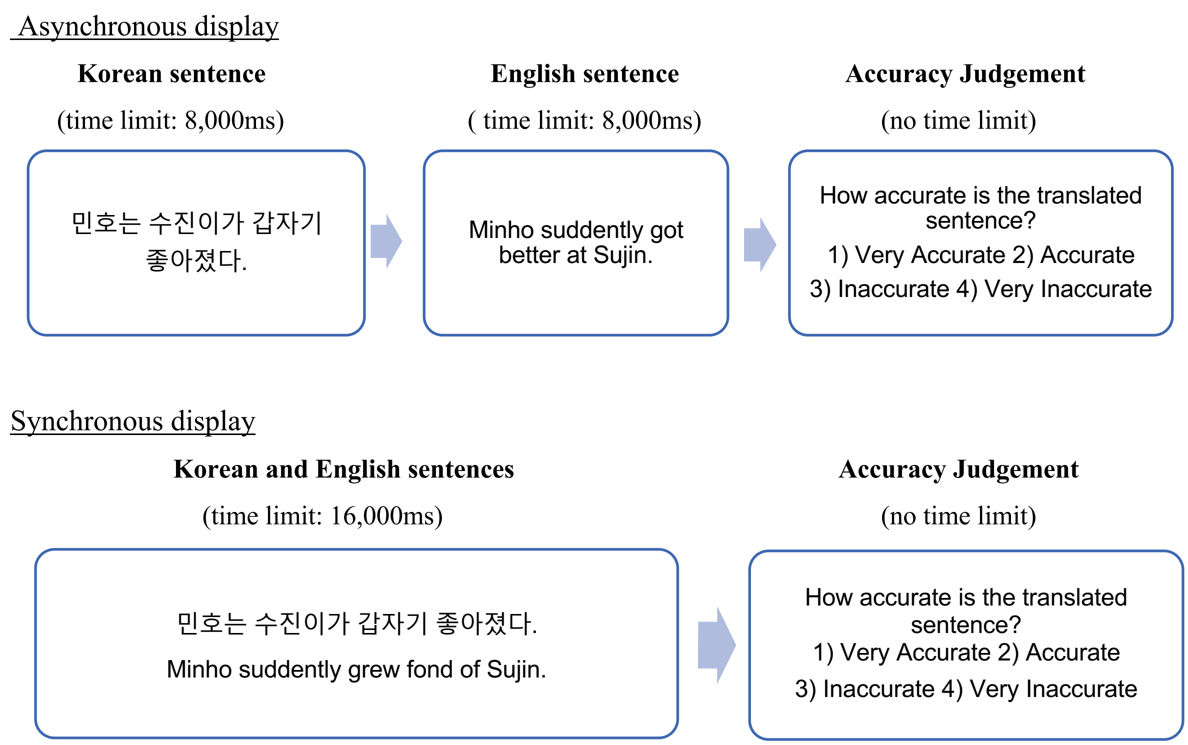
Procedure

The experiment was conducted using an online survey platform called Alchemer (http://www.alchemer.com/), which enables advanced question logics and measurement of response times.

The participants gave their online consent, took the cloze test, and were given two practice questions before partaking in the AJT. They were told that they would see a series of Korean sentences and their machine translated counterparts and would be asked to judge the accuracy of each translated sentence. In the asynchronous display condition, participants read the Korean sentence first and then pressed the 'NEXT' button to read its translated English counterpart on the subsequent screen. They had a time limit of eight seconds to read each sentence and could not go back to the previous page. In the synchronous display condition where the two sentences were shown together on the same screen, participants were given a maximum of 16 seconds to read the two sentences. After reading the two sentences, participants pressed the 'NEXT' button to judge the accuracy of the translation. There was no time limit for accuracy judgment. The use of dictionaries, MT, or other resources were strictly forbidden, and the entire procedure took about 30 minutes for each participant. The experimental procedure for the two display conditions is illustrated in Figure 1.

Figure 1

Experimental Procedure for the Two Display Conditions



Data Analysis

In psycholinguistics, delays in response or reading times are indicative of processing difficulty, while faster times indicate relative ease of processing. Such time measurements have been employed in numerous psycholinguistic studies (Clahsen & Felser, 2006; Roberts & Felser, 2011, among many others) as they offer reliable data that can account for speakers' linguistic and cognitive abilities. The present study measured accuracy judgment (*Very accurate* '4,' *Accurate* '3,' *Inaccurate* '2,' *Very inaccurate* '1'), response latencies, and total reading times of source and target sentences. These dependent variables were analyzed with logit mixed-effect models (Baayen et al., 2008) using the lme4 package (Bates et al., 2015) in the R environment (R Core Team, 2019). The fixed effects were Accuracy (accurate vs. inaccurate; effect-coded as -.5 and .5), Group (high vs. low; effect-coded as -.5 and .5), Display (asynchronous vs. synchronous; effect-coded as -.5 and .5), and Category (morphology vs. syntax vs. semantics; effect-coded as .5, -1, .5 and .25, 0, -.25) and their interactions (Jaeger, 2008). For random effects, the models started out with the maximal random effects structure with by-participant and by-item random intercepts and slopes (Barr et al., 2013), but the models with random slopes failed to converge and thus the models reported below only had by-participant and by-item random intercepts as random effects.

Results

Statistical analyses of the three dependent variables (accuracy judgment, response latencies, and total reading times) revealed that MT error categories, learners' L2 proficiency, and visual presentation all interact to determine how MT output is processed. First, analysis of response latencies—the time taken for the learners to make an accuracy judgment after reading the source and target sentences—indicated that the ways in which learners process MT errors of different categories (morphological, syntactic, and semantic) largely depend on their L2 proficiency level. In the statistical analysis of response latencies, extreme response times above 10,000 ms or 2.5 standard deviations above or below the mean were excluded, which affected around 8% of all responses. No main effect was found, but a significant interaction between Category and Group ($\beta = .07$, $SE = .02$, $t = 2.99$, $p = .003$) showed that the two proficiency groups significantly differed in their response times for the semantic category. Advanced learners took longer to respond to items in the semantic category than the other two categories, whereas the less proficient learners spent the least amount of time to respond to items in the semantic category and longest for those in the syntactic category. When separate analyses were conducted for accurate vs. inaccurate items within each category, high proficiency learners took significantly more time to respond to accurate semantic items than low proficiency learners ($\beta = .08$, $SE = .03$, $t = 2.46$, $p = .014$). The advanced learners also took significantly more time to respond to inaccurate morphological and semantic items than the low proficiency learners ($\beta = -.25$, $SE = .11$, $t = -2.23$, $p = .026$). High proficiency learners' response times were longest in the semantic condition, while low proficiency learners' response times were longest in the syntactic and shortest in the semantic condition. Table 1 and Figure 2 illustrate these group differences in response times of the three linguistic categories.

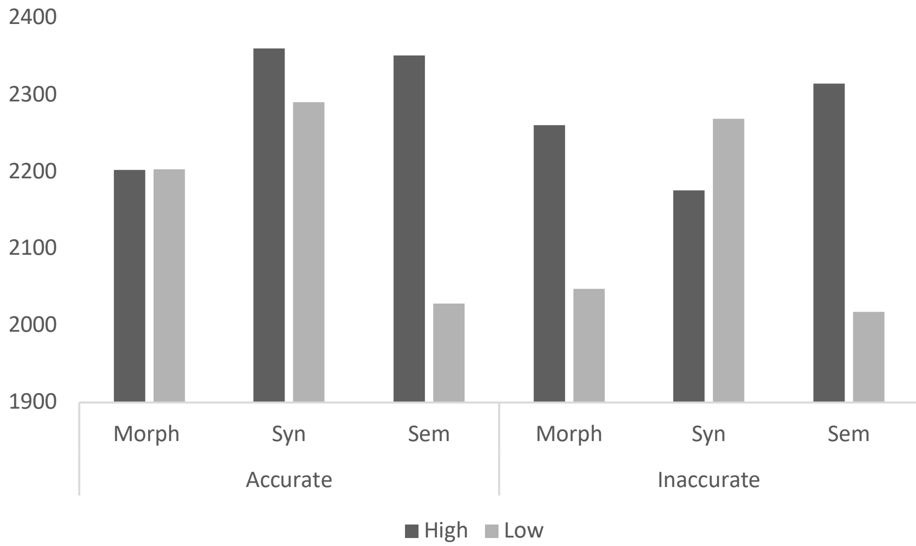
Table 1

Mean Values of Response Times in Milliseconds (SD in Parentheses)

	Morphological		**Syntactic**		**Semantic**	
Proficiency	Accurate	Inaccurate	Accurate	Inaccurate	Accurate	Inaccurate
High	2202.33 (869.84)	2260.52 (836.31)	2360.58 (932.17)	2176.38 (767.67)	2351.71 (824.81)	2315.14 (951.43)
Low	2203.54 (1049.48)	2047.94 (943.79)	2290.41 (1036.19)	2268.91 (1030.85)	2029.08 (901.36)	2017.51 (803.02)

Figure 2

Mean Values of Response Times



The effect of proficiency level was also evident in accuracy judgments. The responses revealed a main effect of Accuracy (β = 1.11, *SE* = .04, *t* = 28.61, *p* < .001) and a significant interaction between Accuracy and Group (β = -.40, *SE* = .08, *t* = -5.23, *p* < .001). No other factor or interaction was significant. Accurate items received significantly higher scores than inaccurate items, and high proficiency learners were better at differentiating accuracy than low proficiency learners, as expected. The two proficiency groups were significantly different in their judgments of inaccurate items (β = .27, *SE* = .09, *t* = 2.75, *p* = .009), but not for accurate items (β = -.13, *SE* = .08, *t* = -1.74, *p* = .089), which shows that the advanced learners were better at rejecting inaccurate items than lower proficiency learners. Table 2 and Figure 3 display the mean values of accuracy responses by high and low proficiency groups in accurate and inaccurate conditions.

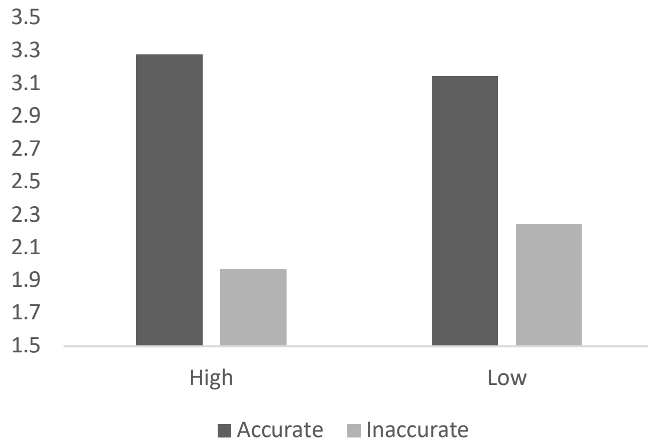
Table 2

Mean Values of Accuracy Responses (SD in Parentheses)

Proficiency	Accurate	Inaccurate
High	3.28 (.75)	1.97 (.92)
Low	3.15 (.80)	2.24 (1.03)

Figure 3

Mean Values of Accuracy Responses by Proficiency Group



Additional analyses of accuracy responses were conducted by coding the responses as either correct ('1') or incorrect ('0') with 'Very accurate' and 'Accurate' as correct responses for accurate items and 'Inaccurate' and 'Very inaccurate' as correct responses for inaccurate items. The number of correct responses was significantly higher for accurate than inaccurate items (β = .17, *SE* = .02, *t* = 8.684, *p* < .001), especially in the semantics category (β = -.06, *SE* = .03, *t* = -2.09, *p* =.03). This indicates that the learners were better at accepting correct translations than rejecting mistranslated sentences. Moreover, advanced learners provided a significantly higher number of correct responses than low proficiency learners (β = -.09, SE = .02, *t* = -4.22, *p* < .001).

Lastly, analyses of total reading times—the time taken for learners to read and process the source and target sentences—revealed a significant effect of Display. There was a time limit of 8,000 ms to read each sentence, so the maximum reading time was 16,000 ms. To correct for outliers, response times that were 2.5 standard deviations above or below the mean were excluded, but the reading time data largely deviated from a normal distribution and could not be corrected by log-transformation and other standardization procedures because it was highly skewed towards the maximum value, 16,000 ms. Thus, total reading times that were equal to the maximum value were excluded, which affected 23% of the data. In the models that were used to analyze reading times, four-way interactions of fixed effects were removed due to convergence issues. There was a main effect of Display, with longer reading times found for synchronous (*M* = 8458.05, *SD* = 2786.08) than asynchronous (*M* = 8315.28, *SD* = 2731.73) display. No other main effect was found, but there were significant interactions between Category and Display, Category and Accuracy, Display and Accuracy, Display and Group, as well as Display, Accuracy, and Group, as shown in Table 3.

Additional analyses of the significant interactions revealed that the effect of Display was significant for the syntactic category (β = 650.20, *SE* = 209.24, *t* = 3.11, *p* = .002) and for inaccurate items (β = 596.42, *SE* = 178.50, *t* = 3.34, *p* < .001) with reading times of the synchronous display being significantly longer than those of the asynchronous display. High proficiency learners were significantly slower to read sentences in the synchronous than asynchronous condition (β = 468.00, *SE* = 163.37, *t* = 2.87, *p* = .004) and accurate than inaccurate items (β = 333.61, *SE* = 158.21, *t* = 2.11, *p* = .04). Low proficiency learners' reading times displayed a significant interaction between Accuracy and Display (β = -1039.05, *SE* = 385.80, *t* = -2.693, *p* = .007); significantly longer reading times were found for accurate than inaccurate items in the asynchronous display condition but longer reading times were found for inaccurate than accurate items in the synchronous display condition. Additional analyses of reading times in the asynchronous display conditions revealed that most participants (86%) took longer to read the second sentence (TL) than the first

(SL), and the difference in the time spent on each sentence was significantly greater for accurate than inaccurate items (β = 563.63, *SE* = 163.63, $t = 3.445$, $p < .001$). Also, the difference in reading times between the two sentences was significantly greater for low proficiency than high proficiency learners (β = 858.55, *SE* = 420.90, $t = 2.04$, $p = .05$). The mean values of total reading times in each experimental condition for the two groups are presented in Table 4 and Figure 5.

Table 3

Fixed Effects Estimate for Total Reading Times

Fixed effects	**β**	***SE***	***t***	***p***
Intercept	8744.45	278.58	31.39	<.001 ***
Category1: Morph/Syn vs. Sem	294.43	211.29	1.39	.17
Category2: Morph vs. Syn	-594.87	733.82	-.81	.42
Display	290.75	120.54	2.41	.02 *
Accuracy	108.63	119.15	.91	.36
Group	708.90	484.74	1.46	.15
Category1 x Display	342.12	167.39	2.04	.04 *
Category1 x Accuracy	-350.04	167.76	-2.09	.04 *
Display x Accuracy	-571.95	252.02	-2.27	.02 *
Display x Group	-510.39	244.24	-2.09	.04 *
Display x Accuracy x Group	-1058.10	500.84	-2.11	.03 *

Note. Only significant interactions are reported for space reasons. Results on other interactions can be viewed upon request to the author; *SE* = standard error.
Significance codes: '***' 0.001 '**' 0.01 '*' 0.05

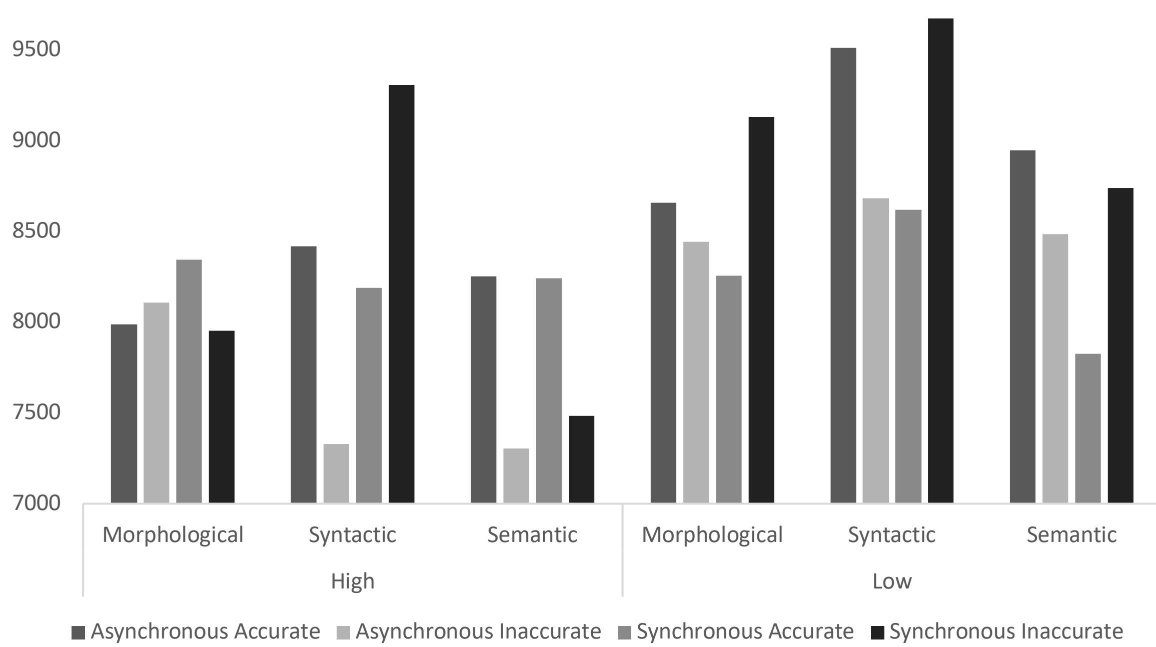
Table 4

Mean Values of Total Reading Times in Milliseconds (SD in Parentheses)

		Asynchronous		Synchronous	
		Accurate	**Inaccurate**	**Accurate**	**Inaccurate**
High	Morphological	7987.00 (2675.06)	8108.98 (2899.36)	8342.55 (2717.66)	7951.88 (2857.67)
	Syntactic	8416.44 (2521.67)	7328.60 (2286.24)	8186.92 (2486.24)	9306.90 (2850.62)
	Semantic	8251.76 (2727.50)	7302.93 (2413.42)	8240.02 (2851.88)	7483.60 (2427.79)
Low	Morphological	8656.23 (2840.35)	8442.90 (3061.79)	8256.07 (2554.27)	9130.65 (3124.26)
	Syntactic	9511.90 (2762.80)	8682.00 (2614.09)	8619.29 (2396.30)	9672.13 (2974.03)
	Semantic	8946.93 (2606.30)	8483.43 (2789.20)	7825.46 (2617.90)	8737.28 (2977.49)

Figure 5

Mean Values of Total Reading Times



Discussion

The present study examined how Korean learners of English perceive and process errors in Korean-to-English MT output that arise from morphological, syntactic, and semantic differences in the two languages and investigated the effect of L2 proficiency level and visual display on this process using accuracy judgments and response/reading times. Overall, distinctive differences between the two proficiency groups were found in how they process sentences with inaccurate vs. accurate translations, errors in the three categories (morphological, syntactic, semantic), and asynchronous vs. synchronous display format. The results also revealed that learners generally take longer to process and accept accurate translations and are better at accepting correct translations than rejecting mistranslated sentences. This suggests that learners do not indiscriminately accept the translated output but scrutinize the sentences for errors even when the translated sentences are acceptable. In addition, constant visual access to the source text resulted in longer reading times and did not play a facilitative role when processing MT errors. The present findings are discussed in further detail below.

The L2 learners as a whole did not perform significantly better in one linguistic category than the other; there was no main effect of linguistic category in the three dependent variables of interest (accuracy responses, response latencies, or reading times). However, when the learners were divided into two proficiency levels, group differences were found in the response latencies of different linguistic categories. Learners of different proficiency levels focused on different aspects of language when processing translated output. High proficiency learners' response times for the semantic category was significantly slower than those of the other categories, and this special attention to semantic differences in the two languages was apparent not only when rejecting mistranslations but also when accepting correct translations. Low proficiency learners, on the other hand, spent the least amount of time to respond to items in the semantic category and longest for those in the syntactic category. When compared to the high proficiency learners, they were significantly slower to respond to inaccurate items in the syntactic category but significantly faster to respond to items in the semantic category. These results suggest that advanced learners tend to focus on resolving differences in meaning and low proficiency learners on resolving structural differences. When processing L1-to-L2 translated output, less skilled learners pay greater attention to structural aspects of the language, but as proficiency develops, they focus on refining the meaning by noticing semantic differences. In other words, morphosyntactic differences in the two languages pose the greatest processing cost for low proficiency learners whose priority in translation is to form acceptable sentences that structurally conform to the rules of the target language. In contrast, advanced learners who are fairly confident in their knowledge of L2 syntactic properties need not spend as much processing resources on structural differences but instead use them to discern subtle meaning differences. These results resonate with previous findings on different language goals in MT use by high and low proficiency learners. In their study of learners' MT use in L2 writing, Chung and Ahn (2022) found high proficiency learners to focus on producing lexically diverse sentences and low proficiency learners to prioritize forming grammatically acceptable sentences. Ahn and Chung (2020) also reported that the main purpose for using MT for high proficiency learners was to find the right word or expression, whereas that of the low proficiency learners was to check the grammar. Putting these findings together, L2 proficiency can be seen as an integral factor in determining how errors of different linguistic categories are processed in L1-to-L2 translated output.

The effect of L2 proficiency was also significant in the learners' judgments of inaccurate translations. Advanced learners could detect a higher number of MT errors and were better at judging the mistranslations as inaccurate than lower proficiency learners, which is consistent with previous findings (Chung, 2020; Kol et al., 2018; Lee & Briggs, 2021). The number of correct responses was significantly higher for accurate than inaccurate items, and the learners found it harder to reject mistranslations than to accept correct translations. Interestingly, however, the learners generally took more time to read accurate than inaccurate items, possibly because they took extra processing time to be certain of the accuracy of the target sentence. Unlike the common concern that students tend to use MT indiscriminately, the learners in the study seemed to accept the MT output only after checking for possible errors. Overall, the learners seem generally more eager to accept than reject errors in the MT output, but the ability to detect MT errors can be expected to improve with increasing proficiency.

As for the third research question regarding access to the source text, the visual display format did not affect learners' accuracy responses or response latencies, but it had a significant effect on total reading times.

Significant delays in processing speed in the synchronous display condition suggested that constant visual access to the source text does not facilitate but rather interferes with processing the translated output. Learners generally took longer to read the two sentences when they were shown together on the same screen than when the target sentence was presented on the next screen, and such tendency was most prominent when processing inaccurate translations in the syntactic category. It can be speculated that the constant presence of the source text led to a heavier reliance on SL linguistic features which could interfere with processing MT errors that especially involve structural differences in the two languages. That is, constant visual access to the source text could lead the learners to make unnecessary linguistic comparisons and syntactic matches that are not helpful in determining the accuracy of MT output. The effect of display, however, heavily depended on the accuracy of the items in the low proficiency group's reading times; the reading times were significantly slower for accurate than inaccurate items in the asynchronous display condition but significantly slower for inaccurate than accurate items in the synchronous display condition. Simply put, having visual access to both source and target sentences facilitated processing of accurate translations, whereas reading one sentence after another reduced processing cost for mistranslated output. It seems that constant visual access to the source text aided the low proficiency learners in establishing linguistic matches between the two languages but had an adverse effect when processing crosslinguistic differences. This could be because the learners relied on surface SL forms instead of the deverbalized message when judging MT accuracy. Advanced learners' reading times, on the other hand, were significantly slower in the synchronous than asynchronous display conditions irrespective of accuracy. Processing the source and target sentences on separate screens significantly reduced the high proficiency learners' reading times, possibly by avoiding unnecessary reliance on SL and helping them to focus on the translated output only. The overall results suggest that learners with lower proficiency tend to rely more heavily on the source text to judge the accuracy of the translated sentences, but that such reliance may not facilitate processing of TL in MT output.

The present findings show how L2 learners identify and process MT errors in real-time and have various pedagogical implications for incorporating MT as a computer-assisted language learning (CALL) tool for L2 production. Recent studies have reported MT to be a useful tool that can help learners produce better-quality outputs (Briggs, 2018; S.-M. Lee, 2021; Tsai, 2019), and the learners' ability to identify errors in the MT output is essential to this process. Explicit instruction and training sessions on MT use can increase the learners' MT instrumental competence (S.-M. Lee, 2021; O'Neill, 2016), and the current findings identify important factors that must be taken into consideration in MT instruction. First, MT instruction and activities must be tailored to differences between proficiency levels. L2 proficiency plays a critical role in how MT output is processed, and noticeable differences between proficiency levels are suggestive of different processing mechanisms at work. Less skilled learners show difficulty processing structural differences in the two languages and tend to accept inaccurate literal translations of the SL structure. Therefore, instructors must train these learners to be knowledgeable of frequent error types that arise from distinctive morphosyntactic differences between SL and TL using error correction activities that compare different properties of the two languages. Advanced learners, on the other hand, tend to focus on semantic differences and would benefit more from activities that compare proverbial/idiomatic expressions and discourse subtleties in the two languages. Explicit instructions on subtle nuances, connotative meanings, and context-appropriate expressions should be provided in MT instruction for advanced learners. In addition to providing different language focus by proficiency level, MT instruction should be more specific about what constitutes an unacceptable translation by providing numerous examples of MT output in the instructional material. Learners generally took longer to process and accept accurate than inaccurate items and were better at accepting correct translations than rejecting the mistranslated sentences in the present study. That the learners are scrutinizing the output for errors even for seemingly accurate translations suggests that learners do not blindly accept good-enough output nor excessively rely on this tool. However, the learners, especially those with low proficiency, were often uncertain about what makes the output unacceptable and failed to reject mistranslated sentences. Accuracy judgment questions like the ones used in the present study that contain various types of mistranslations can serve as practice questions for distinguishing erroneous output. Lastly, based on the present finding that learners' processing speed is significantly affected by how the sentences are presented, instructors must be aware of how visual access to both source and target sentences can affect the learners' ability to detect errors in the MT output. Viewing both sentences could help learners to compare the surface form of the two languages, but the constant presence of the source text could lead to unnecessary reliance and interference. Therefore, it is highly recommended that the source and target sentences be presented separately, and the learners be encouraged to refrain from viewing the source text when evaluating or post-editing MT output.

Conclusion

As one of the first studies that have examined L2 processing of MT errors using response and reading times, the present work delineates the manner in which learners employ their linguistic knowledge in L1-to-L2 translation processing and contributes to our understanding of the factors involved in L2 learners' real-time processing of MT errors. With MT being increasingly viewed as a CALL tool that can facilitate language learning, the present findings underline the importance of L2 proficiency, crosslinguistic differences between SL vs. TL, and the visual presentation of texts as significant factors that affect L2 processing of errors in the MT output.

Future works must further investigate how MT errors are processed using alternative error taxonomies or metrics of post-editing effort, which could be a good predictor of difficulty in processing MT sentences. Various other language pairs must also be explored. More importantly, using a self-paced reading task or eye-tracking would provide a more comprehensive picture of how learners process individual words/clauses and what causes processing delay. Also, as mentioned by one of the reviewers, the effect of display may depend on the length of the translated text, and thus longer sentences or paragraphs should be examined in future investigations. Focusing on these issues will help to further validate and examine the effectiveness of MT as a CALL tool.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea under Grant NRF-2019S1A5B5A02041012.

References

Ahn, S., & Chung, E. S. (2020). Students' perceptions of the use of online machine translation in L2 writing. *Multimedia-Assisted Language Learning, 23*(2), 10–35. https://doi.org/10.15702/mall.2020.23.2.10

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–8. http://CRAN.R-project.org/package=lme4

Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses. *The JALT CALL Journal*, *14*(1), 3–24. https://doi.org/10.29140/jaltcall.v14n1.221

Chmiel, A., & Lijewska, A. (2019). Syntactic processing in sight translation by professional and trainee interpreters. *Target: International Journal of Translation Studies, 31*(3), 378–397. https://doi.org/10.1075/target.18091.chm

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*, 102408. https://doi.org/10.1016/j.system.2020.102408

Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. *The 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. (2022). 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. http://doi.org/10.1080/09588221.2020.1871029

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Correa, M. (2014). Leaving the "peer" out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content and Language Integrated Learning, 7*(1), 1–20. https://doi.org/10.5294/laclil.2014.7.1.1

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. J., & Macken, L. (2017). Identifying the machine translation error types with the greatest impact on post-editing effort. *Frontiers in Psychology, 8*, 1282. http://doi.org/10.3389/fpsyg.2017.01282

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. http://doi.org/10.1111/flan.12366

Fotos, S. S. (1991). The cloze test as an integrative measure of EFL proficiency: A substitute for essays on college entrance examinations? *Language Learning, 41*(3), 313–336. https://doi.org/10.1111/j.1467-1770.1991.tb00609.x

Fredholm, K. (2019). Effects of Google translate on lexical diversity: Vocabulary development among learners of Spanish as a foreign language. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas, 13*(26), 98–117. https://doi.org/10.26378/rnlael1326300

Garcia, I., & Pena, M. I. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning, 24*(5), 471–487. http://doi.org/10.1080/09588221.2011.582687

Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. W. Brislin (Ed.), *Translation: Application and research* (pp. 165–207). Wiley.

Ionin, T., Montrul, S., & Crivos, M. (2013). A bidirectional study on the acquisition of plural noun phrase interpretation in English and Spanish. *Applied Psycholinguistics, 34*(3), 483–518. https://doi.org/10.1017/S0142716411000841

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kim, K. (2018). The influence of case markers on the machine-translation of Korean proverbs into English. *The Linguistic Association of Korea Journal, 26*(3), 139–157. https://doi.org/10.24303/lakdoi.2018.26.3.139

Kim, J.-B., & Choi, I. (2004). The Korean case system: A unified, constraint-based approach. *Language Research, 40*(4), 885–921. http://s-space.snu.ac.kr/bitstream/10371/90786/1/5.%20The%20Korean%20Case%20System.pdf

Kim, S. (2019). Playing with machine translation in language classroom: Affordances and constraints. *Multimedia-Assisted Language Learning, 22*(2), 9–28.

Kim, S.-W., & Lee, H. J. (2017). A study on machine translation outputs: Korean to English translation of embedded sentences. *The Mirae Journal of English Language and Literature, 22*(4), 123–147.

Kol, S., Schcolnik, M., & Spector-Cohen, E. (2018). Google Translate in academic writing courses?. *The EuroCALL Review, 26*(2), 50–57. http://doi.org/10.4995/eurocall.2018.10140

Koponen, M., Aziz, W., Ramos, L., & 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).

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition, 9*(2), 119–135. https://doi.org/10.1017/S1366728906002483

Lee, S.-B. (2018). Process research into post-editing: How do undergraduate students post-edit the output of Google Translate?. *The Journal of Translation Studies, 19*(3), 259–286. http://doi.org/10.15749/jts.2018.19.3.010

Lee, S.-M. (2020a). 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. (2020b). Error analysis and the quality evaluation of neural-network machine translation. *Studies in Foreign Language Education, 34*(4), 129–153. http://dx.doi.org/10.16933/sfle.2020.34.4.129

Lee, S.-M. (2021). An investigation of machine translation output quality and the influencing factors of source texts. *ReCALL, 34*(1), 1–14. http://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

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same process?. *Cognition, 99*(1), 1–34. https://doi.org/10.1016/j.cognition.2004.09.012

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86*(1), 70–82. https://doi.org/10.1016/S0093-934X(02)00535-7

O'Neill, R., Cornelius, E. T., & Washburn, G. N. (1981). *American kernel lessons: Advanced students' book*. Longman.

O'Neill, E. M. (2016). Measuring the impact of online translation on FL writing scores. *The IALLT Journal of Language Learning Technologies, 46*(2), 1–39. https://doi.org/10.17161/iallt.v46i2.8560

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In P. Isabelle, E. Charniak, & D. Lin (Eds.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318). https://aclanthology.org/P02-1040.pdf

Park, J. (2017). Analysis of the types of translation errors by the Google Translator. *Studies in English Language and Literature, 59*(4), 279–299. http://doi.org/10.18853/jjell.2017.59.4.014

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics, 32*(2), 299–331. https://doi.org/10.1017/S0142716410000421

Ruiz, C., Paredes, N., Macizo, P., & Bajo, M. T. (2008). Activation of lexical and syntactic target language properties in translation. *Acta Psychologica, 128*(3), 490–500. https://doi.org/10.1016/j.actpsy.2007.08.004

Seeber, K. G., & Kerzel, D. (2011). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism, 16*(2), 228–242. https://doi.org/10.1177%2F1367006911402982

Seleskovitch, D. (1976). Interpretation: A psychological approach to translating. In R.W. Brislin (Ed.), *Translation: Application and research* (pp. 92–116). Wiley.

Seol, O.-S. (2009). English-Korean translation strategies on tense and aspect. *The Journal of Translation Studies, 10*(2), 23–60. http://doi.org/10.15749/jts.2009.10.2.002

Specia, L., Raj, D., & Turchi, M. (2010). Machine translation evaluation versus quality estimation. *Machine Translation, 24*(1), 39–50. http://doi.org/10.1007/s10590-010-9077-2

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

Temnikova, I. (2010). Cognitive evaluation approach for a controlled language post-editing experiment, In A. Abad, L. Bentivogli, I. Dagan, D. Giampiccolo, S. Mirkin, E. Pianta, … & D. Tapias (Eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation* (pp. 3485–3490). http://www.lrec-conf.org/proceedings/lrec2010/pdf/437_Paper.pdf

Tsai, S.-C. (2019). Using google translate in EFL drafts: A preliminary investigation. *Computer Assisted Language Learning, 32*(5–6), 510–526. http://doi.org/10.1080/09588221.2018.1527361

Viezzi, M. (1989). Information retention as a parameter for the comparison of sight translation and simultaneous interpretation: An experimental study. *The Interpreters' Newsletter, 2*, 65–69. http://hdl.handle.net/10077/4806

Vilar, D., Xu, J., D'Haro, L. F., & Ney, H. (2006). Error analysis of statistical machine translation output. In C. Strapparava, A. Valitutti, & O. Stock (Eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (pp. 697–702). http://www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf.pdf

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

Appendix A. Error Types in the Three Linguistic Categories

Category	Type	Description and example
Morphological	Nominative case	Korean overtly marks grammatical functions using case markers whereas English assigns grammatical roles structurally based on surface position. The nominative case in Korean (*-i/ka*) generally marks the subject of a phrase but can also occur with non-subjects in auxiliary constructions (K. Kim, 2018; Kim & Choi, 2004). <u>Example</u> Hoysa tonglyo-tul-kwa kathi hayngsalul Company colleague-PL-with together event-ACC cinhayng-ha-myense ku-nun tonglyo-tul-**i** cohacye-ss-ta. run-do-ing he-TOP colleague-PL-**NOM** like-PST-DECL 'He grew fond of his colleagues as he ran an event with them.' 'Running the event with his colleagues at work, he got better with his colleagues.' (MT: Papago)
	Adverbial case (*-ey*)	Adverbial case markers in Korean such as '*-ey'* can denote multiple semantic relations such as static locative, dative, allative, causative, and succession, which can result in frequent MT errors (Park, 2017). <u>Example</u> Hoyuy naynay kunye-uy etwuwun phyoceng-**ey** Meeting throughout she-POSS dark/grim expression-**CAUS** pwunwiki-ka coh-ci anh-ass-ta. atmosphere-NOM good-NOM NEG-PST- DECL. 'The atmosphere was not good because of her grim expression throughout the meeting.' 'The mood wasn't good for her dark look throughout the meeting.' (MT: Google)
	Tense/aspect	Tense and aspect are expressed using verb forms in English, but Korean represents tense and aspect beyond the verb-level using time adverbials, attaching inflections on an adjective or noun, or just by inferring from the sentence context without using external tense/aspect cues. When two consecutive clauses have the same tense/aspect, Korean often expresses tense/aspect only in the second clause and allows the first clause to be expressed in the present tense, which MT incorrectly perceives as a shift in tense/aspect (Seol, 2009). <u>Example</u> Pisangsathay-ka palsayng-ha-**ko** himtul-ess-ten

		Emergency-NOM occur-do-CONV hard/tough-PST-MODF il-nyen-i keuy mamwuli toye-kan-ta. one-year-NOM almost finish become-PROG-DECL. 'A tough year has almost passed since the emergency crisis.' 'An emergency occurs and the hard year is almost over.' (MT: Google)
Syntactic	Coordination & Negation	MT sometimes assigns inappropriate scope of negation in reference to coordinated constructions in Korean, assigning negation to all conjuncts (words/phrases) in the coordination when only one should be negated (Park, 2017). Example Yocum-un kyelhon-ul **milwu-kena** **ha-ci** Nowadays-TOP marriage-ACC postpone-either do-NOM **anh-ulye-nun** itul-i nulko-iss-ta. not-intend-MODF people-NOM increasing-PROG-DECL 'Nowadays, an increasing number of people postpone marriage or do not get married at all.' 'Nowadays, more and more people do not want to postpone marriage.'(MT: Google)
	Pro-drop	Pro-drop in Korean can be problematic for MT because the number of subjects does not match with that of predicates (Kim & Lee, 2017). Although this error has significantly improved since the introduction of NMT and missing noun phrases can be correctly added when entire texts are translated (S.-M. Lee, 2021), MT may still fail to restore missing subjects in single sentence translations with no previous context (S. Kim, 2019). Example Ku-nun **(*pro*)** coychaykkam-ul nukkil philyo-ka He-TOP guilt-ACC feel need-NOM epstamye nay sakwa-lul kecel-hay-ss-ta. no-QT my apology-ACC reject-do-PST-DECL. 'He did not accept my apology saying that I didn't have to feel guilty.' 'He refused my apology, saying he didn't have to feel guilty.' (MT: Google)
	Adnominal Clause	Unlike English that uses relative pronouns like 'which' to mark the embedded clause, Korean does not alert the start or end of the embedded adnominal clause with separate markers, and MT often

		fails to apply the modification in the right place (Kim & Lee, 2017; Park, 2017). <u>Example</u> Halwycongil pappun kunye-ka **yeyu-lul** **pwuli-nun** All.day busy-MODF she-NOM relax-ACC indulge-MODF **kutul-ul** caychok-han-ta. them-ACC prod-PRES-DECL. ‘Being busy all day long, she prods those who are taking their time.’ ‘Busy all day long, she hastens them to relax.’ (MT: Papago)
Semantic	Homonym	Semantic ambiguities arising from homonyms, discourse subtleties, and proverbial or idiomatic expressions with cultural references were all found to result in frequent MT errors (K. Kim, 2018; Park, 2017). <u>Example</u> Yenghwa sok-ey-nun nulkum-kwa celmum-ul **taypi**-han Movie in-LOC-TOP old.age-and youth-ACC contrast-MODF cangmyen-i manhi nathana-n-ta. scene-NOM many appear-PRES-DECL. ‘There are many scenes that contrast old age and youth in the movie.’ ‘There are many scenes in the movie that prepare for old age and youth.’ (MT: Papago)
	Idiomatic expressions	<u>Example</u> Cikak-ul **pap** **mek-tus ha-nun** keyulun haksayng-tul-un Tardy-ACC meal eat-like do-MODF lazy student-PL-TOP enu hakkyo-eyna iss-ta. any school-LOC present-DECL. ‘There are lazy students who are always tardy in every school.’ ‘There are lazy students in every school who are late for eating.’ (MT: Google)
	Proverbial expressions	<u>Example</u> **Holangi-to** **cey** **malha-myen** **onta**-teni ku-ka machim Tiger.also self talk-COND come-as he-NOM just natha-nass-ta.

	appear-PST-DECL.
	‘Speaking of the devil, he just appeared.’ ‘The tiger came when I said it, and he finally appeared.’(MT: Google)

Note. Yale Romanization is used to transliterate the Korean examples. ACC: accusative, CAUS: causative, CONV: converb, DECL: declarative, MODF: modifier, NEG: negative, NOM: nominative/nominalizer, PL: plural, PRES: present, PST: past, POSS: possessive, PROF: progressive, QT: quotative, TOP: topic.

Appendix B. Counterbalanced Presentation Lists (Accuracy and Display)

		List 1		List 2		List 3		List 4	
Q#	Category	Acc.	Display	Acc.	Display	Acc.	Display	Acc.	Display
Q1	morph	1	async	0	sync	1	sync	0	async
Q2	syn	1	async	0	sync	1	sync	0	async
Q3	sem	1	async	0	sync	1	sync	0	async
Q4	morph	1	async	0	sync	1	sync	0	async
Q5	sem	0	async	1	sync	0	sync	1	async
Q6	syn	1	async	0	sync	1	sync	0	async
Q7	syn	1	async	0	sync	1	sync	0	async
Q8	morph	0	async	1	sync	0	sync	1	async
Q9	sem	0	async	1	sync	0	sync	1	async
Q10	syn	0	sync	1	async	0	async	1	sync
Q11	sem	1	sync	0	async	1	async	0	sync
Q12	morph	0	sync	1	async	0	async	1	sync
Q13	sem	0	sync	1	async	0	async	1	sync
Q14	morph	1	sync	0	async	1	async	0	sync
Q15	syn	1	sync	0	async	1	async	0	sync
Q16	sem	0	sync	1	async	0	async	1	sync
Q17	syn	0	sync	1	async	0	async	1	sync
Q18	morph	0	sync	1	async	0	async	1	sync
Q19	morph	0	async	1	sync	0	sync	1	async
Q20	syn	1	async	0	sync	1	sync	0	async
Q21	sem	0	async	1	sync	0	sync	1	async
Q22	morph	0	async	1	sync	0	sync	1	async
Q23	sem	1	async	0	sync	1	sync	0	async
Q24	syn	1	async	0	sync	1	sync	0	async
Q25	syn	1	async	0	sync	1	sync	0	async
Q26	morph	1	async	0	sync	1	sync	0	async
Q27	sem	1	async	0	sync	1	sync	0	async
Q28	syn	1	sync	0	async	1	async	0	sync
Q29	sem	0	sync	1	async	0	async	1	sync
Q30	morph	0	sync	1	async	0	async	1	sync
Q31	sem	0	sync	1	async	0	async	1	sync
Q32	morph	0	sync	1	async	0	async	1	sync
Q33	syn	0	sync	1	async	0	async	1	sync
Q34	sem	1	sync	0	async	1	async	0	sync
Q35	syn	0	sync	1	async	0	async	1	sync
Q36	morph	1	sync	0	async	1	async	0	sync

Note. 1 = Accurate, 0 = Inaccurate.

About the Author

Eun Seon Chung is an assistant professor at the Department of English Language and Culture at Hankyong National University in South Korea. She conducts research in the field of second language acquisition and is currently interested in the topic of online machine translation, second language writing, and language processing.

E-mail: prolingesc@gmail.com