Computer-Assisted Language Learning Electronic Journal (CALL-EJ), 24(3), 1-19. 2023

L2 Learners' Use of Raw Machine-Translated Output in Reading Comprehension

Eun Seon Chung (echung@hknu.ac.kr)
Hankyong National University
School of Applied Humanities & Public Service, South Korea

Abstract

Many studies in the field of second language (L2) education have examined students' use of online machine translation (MT) in writing, but not much is known about the effect of MT use on reading comprehension. The present study thus investigated how L2 learners employ MT output in L2 reading and how it affects comprehension. Thirty advanced Korean learners of English participated in an online reading comprehension task in which unedited MT texts in Korean were presented before or after reading literary texts in English. Text reading times and responses to comprehension questions were collected and analyzed. The results revealed that the presentation order of MT output and source text has a significant effect on reading times as reading the MT output after the source text than before led to a closer reading. In the post-task survey, the majority of the learners strongly supported the use of MT as a convenient tool that can provide a quick overview of the text and easy access to vocabulary. As the first study to examine L2 learners' use of MT output in reading comprehension, this study presents important implications for MT as a "gisting" tool that increases learners' overall understanding of L2 texts. (199 words)

Keywords: online machine translation; L2 reading comprehension; gisting; reading times; text order

Introduction

Online machine translation (MT) has made significant improvements in quality and accuracy in recent years, and many second language (L2) learners report to frequently using MT as a resource that assists them in L2 learning (Briggs, 2018; Chang & Hsu, 2011; Chung & Ahn, 2022; Lee, 2021; Niño, 2009, 2022). MT applications can be divided into two main categories: assimilation and dissemination. Assimilation refers to the use of raw MT output for informational purposes to quickly gather the gist of foreign texts (i.e., gisting), whereas dissemination refers to using MT output as a rough draft translation that is then post-edited into a more publishable form. Much recent work on MT in the field of language education has investigated how learners use MT for dissemination such as L2 writing and post-editing of MT output (Chung, 2020; Chung & Ahn, 2022; Lee, 2020; Tsai, 2019), but not much is known about the use of MT in L2 reading comprehension despite the pervasive use of MT for gisting purposes. Assimilation is the most dominant use of MT in the number of words translated (Niño, 2020), but surprisingly the effect of MT on reading comprehension is still very much unexplored.

Studies in L2 reading have repeatedly shown that L1 knowledge is inextricably connected to L2 knowledge in reading comprehension (Cook, 2007; Kern, 1994). For example, mental translation, the act of using L1 forms to process the L2, is a widely used cognitive strategy for reading comprehension that can be successfully employed at any level of proficiency (Kern, 1994; Nourinezhad & Kashefian-Naeeini, 2020). Given such heavy involvement of L1 forms in L2 reading, it leads one to question how using MT and its output would affect the way L2 text is read and processed. Reading processes of MT texts are expected to be different from those of well-formed texts, as the readers need to guess the intended meaning for parts that are less coherent or accurate. The present study thus investigates this neglected area of research by observing L2 learners' use of MT output for gisting when reading challenging L2 texts and examining its effect on comprehension.

Literature Review

Studies in MT evaluation that have analyzed and compared raw (i.e., unedited) MT output with human-translated output have found raw MT texts to be a reliable resource for gisting purposes and even for answering reading comprehension questions (Macken & Ghyselen, 2018; Scarton & Specia, 2016). Scarton and Specia (2016) evaluated the comprehensibility of MT texts using reading comprehension tests. Documents from a German corpus were translated into English using several MT systems, and human translations were used as control. Fluent English speakers who participated in the study only read the raw MT output or the human-translated text and not the original documents, and reading performance was measured using open comprehension questions that inquire about explicit literal information as well as inferential information in the texts. The authors found that the reading scores of those who read the MT text was not significantly different from those who read the human-translated text. These findings indicate that the MT systems are highly sophisticated and can successfully deliver the message of the original text. Similarly, Macken and Ghyselen (2018) compared the results of 99 participants' performance in reading comprehension tests after reading either human-translated or raw MT texts and found no significant group differences in the reading scores. Although human-translated texts were rated higher than MT texts and had the best scores for clarity, human-translated texts did not confer a special advantage over MT texts when answering reading comprehension questions.

While evaluations of MT systems have been conducted as such, the topic of MT use in L2 reading comprehension is still poorly understood, and as far as could be ascertained to date, Karnal and Pereira (2015) is the only study that has specifically examined this topic. This preliminary study compared L2 reading strategies when using or not using Google Translate (GT) as a supplementary tool for reading. Ten intermediate learners of English with different L1s were asked to read academic texts in English, and their reading strategies were analyzed using think aloud protocols. Overall, GT use was found to be of great benefit to reading comprehension in various ways. Using GT enabled the learners to engage in a greater number of cognitive and metacognitive strategies that involved an interactive processing of top-down as well as bottom-up information, which led to a higher level of comprehension. Without the aid of GT, participants often skipped unfamiliar words and exhibited a shallow understanding of the text, but using GT reduced the learners' cognitive load and increased local understanding of the texts.

Such effectiveness of MT use is clearly reflected in students' perceptions of MT as a useful resource for reading comprehension (Alhaisoni & Alhaysony, 2017; Briggs, 2018; Clifford et al., 2013; Lee, 2020; Niño, 2020). Students in Clifford et al. (2013) reported that their main use of MT for L2-to-L1 translations was when reading a text (60%), and they found MT to be a fast and convenient tool that could facilitate reading. The majority of the L2 learners in Alhaisoni and Alhaysony (2017) also responded that MT helped them to understand unfamiliar words, phrases, and sentences in the L2 text and that MT was especially useful for vocabulary learning. Fairly proficient learners in Dennis (2019) pointed out that MT facilitated L2 reading comprehension by reducing approximately 25% of the usual reading time and providing rapid access to unfamiliar words. Niño (2020) also reported that the majority (83%) of the students found MT to be useful for reading comprehension. The learners acknowledged the limitations of MT but mostly agreed that the benefits of MT use outweigh the drawbacks for L2 reading comprehension similar to what has been found for MT use in L2 writing (Chung & Ahn, 2022; Garcia & Pena, 2011; Kol et al., 2018; Lee, 2020; Tsai, 2019). However, most of these surveys addressed the general use of MT, and it remains to be seen how the learners perceive the efficacy of MT use for gisting purposes in reading comprehension.

In previous studies, most L2 learners reported to using MT to translate difficult words or phrases (Bower, 2010; Clifford et al., 2013; Dennis, 2019), but learners are showing a greater tendency to use MT to translate longer segments or long stretches of text (Niño, 2022). Up to 69% of L2 learners in Chen (2020) reported using MT to translate entire paragraphs or texts in L2 writing, and it is not uncommon for learners to use MT for large sections of L2 texts in reading comprehension as well. MT can be expected to have been more extensively used by language learners in recent years due to an unprecedented growth in online learning in the post-COVID era and the wide availability of advanced features that enable instant translation of whole texts via text recognition and embedded MTs in web browsers and social media platforms. As more learners are increasingly using MT to translate long stretches of text, there needs to be an examination of whether and how learners use MT for gisting purposes in reading comprehension. Gisting entails reading the MT output as a whole without constantly comparing it to the source text (ST) for certain parts of the text. A recent study (Chung, in press) has found that displaying ST and MT output together on the same screen resulted in slower reading times than showing ST and MT output consecutively, which suggests that constant visual access to the ST could possibly interfere with detecting and processing MT errors. Although this study was conducted with short sentences, it can be predicted that the simultaneous presence of lengthy ST and MT texts would also hinder rather than facilitate information processing. Therefore, the question of how learners use MT output *as a whole* when reading L2 texts must be further examined.

Against this backdrop of fast-evolving omnipresent MT systems and L2 learners' increasing reliance on online tools, more L2 studies must systematically investigate the effect of MT output in reading comprehension, but there is a dire lack of empirical studies on this topic. As a result, many questions remain. To what extent do the learners refer to raw MT output for gisting, and how does it affect their comprehension? If the learners do refer to the MT text to inform them about the L2 text, do they prefer to read the MT text before or after the L2 text and how does this order affect comprehension? The present study attempts to answer such questions by conducting a timed reading comprehension task in which Korean L2 learners of English read challenging literary texts in English and are provided with raw MT output of whole passages before or after reading the L2 texts. Literary texts were used because they encourage authentic language use and are deemed particularly suitable for language learning with their richness and cultural content (Bavendiek, 2022; Jones & Oakey, 2019). Also, the study focused only on advanced learners to maximally understand the extent to which MT output is deemed effective for L2 reading comprehension; proficient learners have been found to be generally less receptive to MT use and more keenly aware of the shortcomings of MT than low proficiency learners who tend to readily accept and use MT output (Chung & Ahn, 2022; Lee, 2020; Niño, 2020). The study also examined the learners' perception and attitude towards MT use for reading by conducting a post-task survey. By addressing the following research questions, the present study seeks to gain a greater understanding of the effect of MT output in L2 reading comprehension, an important but under-investigated topic which has frequently been overlooked in the research of language learning and teaching:

- How much do advanced Korean learners of English refer to raw MT output for gisting and how does it affect comprehension?
- How does reading raw MT output before or after reading the source text (i.e., text order) affect comprehension?
- What are the advanced learners' general perceptions and attitude towards using raw MT output for L2 reading?

Methods

Participants

Thirty Korean L2 learners of English (mean age 26.1, range 20-34) who are undergraduate or graduate students at two different universities in Seoul participated in the study. Having learned about the study through university recruiting websites, the students voluntarily participated in the study and were each given KRW 9,000 for their participation. The learners' length of stay in a foreign country before the age of 15 did not exceed six months, and they were all advanced learners whose scores on the Test of English for International Communication (TOEIC)[1] were above 785 points (mean score 865.5, range 785-990). In addition to the TOEIC scores, the learners' L2 proficiency was measured using LexTALE (Lexical Test for Advanced Learners of English; Lemhöfer & Broersma, 2012), a quick lexical test for advanced learners that has been found to be a reliable measure of vocabulary knowledge and possibly general proficiency (Nakata et al., 2020). The LexTALE test was included because numerous studies have found vocabulary knowledge to be a critical predictor of reading competence (Brevik et al., 2016; Jeon & Yamashita, 2014; Li & Clarina, 2019). Adapted from Lemhöfer and Broersma (2012), it consisted of 60 words as stimuli (40 words and 20 non-words), and the participants had to indicate for each word whether it is an existing word in English or not by selecting "Yes" or "No". This simple test could be completed in five minutes. Simple bivariate correlation analysis via Pearson coefficient showed that the participants' LexTALE scores (mean: 40.4; range 33-56) were moderately correlated ($r = .37$) with their TOEIC scores at a statistically significant level ($p < .001$).

---

[1] The Test of English for International Communication (TOEIC) offered by Educational Testing Service (ETS) is a widely recognized proficiency test in Korea which mainly tests comprehension and receptive language skills (listening and reading). As a test that is out of 990 points, scores that are above 785 (equivalent to CEFR level B2 and up) are described as "upper intermediate" and "advanced" (http://ets.org/toeic/html).

Materials and Procedure

The experiment was conducted using an online survey platform called *Alchemer* (https://app.alchemer.com/), which enables advanced questions logics and measurement of response times in milliseconds. All participants had to provide informed consent prior to partaking in the experiment. After giving their online consent, the participants completed three tasks in the following order: the LexTALE test, a reading comprehension task in English, and a survey in Korean.

Reading Comprehension Task

In the reading comprehension task, excerpts from three short stories in English that are 480-500 words in length and have advanced Lexile scores (1210-1400) were used as source texts (ST).[2] The raw MT texts (RMT, henceforth) were translated using Naver Papago, which Korean learners of English strongly prefer over other MT systems (Ahn & Chung, 2020). The RMT was unedited and presented as is; each text contained approximately 8-10 errors that were mostly semantic in nature (i.e., literal translations or awkward expressions that do not reflect culturally specific knowledge or are taken out of context) but contained no grammatical errors. Five reading comprehension questions were formulated for each text in the format of True or False questions with "TRUE", "FALSE", or "NOT GIVEN" as possible answer choices. The five questions were of three different types: major detail (1 question), inference or application (1 question), and minor details (3 questions).[3] Minor detail questions specifically asked about the details in the passage which the RMT erred in. The order of these questions was randomized across participants. Three different text conditions appeared in the following order: (1) Control condition where no RMT was provided (2) Post-ST condition where RMT was made available after reading ST (3) Pre-ST condition where RMT was made available before reading ST. Each of the three stories was assigned to a different condition, but the presentation order of the short stories was randomized across participants so that a particular story was not always assigned the same text condition. That is, all participants were shown the short stories in three different text conditions with five questions in randomized order. ST, RMT, and comprehension questions were all presented on separate screens, and the participants could not go back to the previous screen. Participants were given a maximum of five minutes to read each ST or RMT, but there was no time limit for comprehension questions. In the instruction for the reading comprehension task, the participants were not told about the presence of RMT so as not to reveal the purpose of the study. It was only when they reached the screen with RMT that they were explicitly informed that it was an unedited machine-translated output by Papago and that they had the option of skipping it if they wished. Figure 1 presents an example of the reading comprehension task.

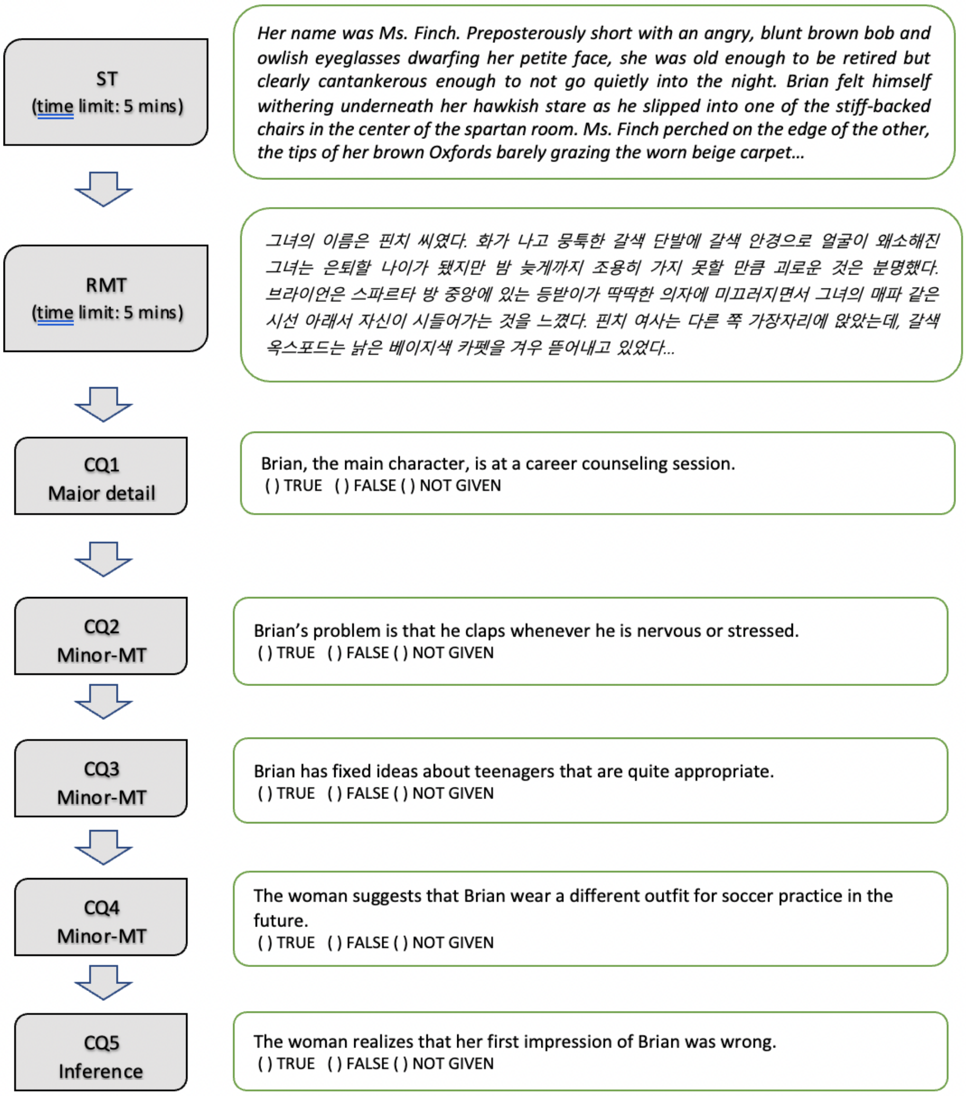
---

[2] The designated passages were taken from 1) *The Red-headed League* by Sir Arthur Conan Doyle, 2) *Discretion is the Better Part of Marriage* by Rebecca Christian, and 3) *The Interview* (https://www.readingvine.com/passages/the-interview)

[3] Major detail questions asked about the general setting or the main idea of the passage (e.g., *The main character is at a career counseling session*), whereas minor detail questions asked about details that pertained to a specific part of the passage (e.g., *The main character's fixed ideas about teenagers are quite appropriate).*

Figure 1

An example of the reading comprehension task procedure (Post-ST condition)



Her name was Ms. Finch. Preposterously short with an angry, blunt brown bob and owlish eyeglasses dwarfing her petite face, she was old enough to be retired but clearly cantankerous enough to not go quietly into the night. Brian felt himself withering underneath her hawkish stare as he slipped into one of the stiff-backed chairs in the center of the spartan room. Ms. Finch perched on the edge of the other, the tips of her brown Oxfords barely grazing the worn beige carpet…

그녀의 이름은 핀치 씨였다. 화가 나고 뭉툭한 갈색 단발에 갈색 안경으로 얼굴이 왜소해진 그녀는 은퇴할 나이가 됐지만 밤 늦게까지 조용히 가지 못할 만큼 괴로운 것은 분명했다. 브라이언은 스파르타 방 중앙에 있는 등받이가 딱딱한 의자에 미끄러지면서 그녀의 매파 같은 시선 아래서 자신이 시들어가는 것을 느꼈다. 핀치 여사는 다른 쪽 가장자리에 앉았는데, 갈색 옥스포드는 낡은 베이지색 카펫을 겨우 뜯어내고 있었다…

Brian, the main character, is at a career counseling session.
( ) TRUE ( ) FALSE ( ) NOT GIVEN

Brian’s problem is that he claps whenever he is nervous or stressed.
( ) TRUE ( ) FALSE ( ) NOT GIVEN

Brian has fixed ideas about teenagers that are quite appropriate.
( ) TRUE ( ) FALSE ( ) NOT GIVEN

The woman suggests that Brian wear a different outfit for soccer practice in the future.
( ) TRUE ( ) FALSE ( ) NOT GIVEN

The woman realizes that her first impression of Brian was wrong.
( ) TRUE ( ) FALSE ( ) NOT GIVEN

Post-task survey

After completing the reading comprehension task, the learners were asked to respond to a short survey that inquired about whether and how they used RMT in the present task, previous experience of using MT for L2 reading, thoughts about the usefulness of MT, and possible future use of this tool for L2 reading. The questions were adapted from surveys conducted in Clifford et al., (2013) and Chung & Ahn (2022). Twelve out of 16 questions were close-ended questions (multiple choice, Yes/No, Likert scale) and the rest were open-ended questions that elicited more detailed responses.

The use of resources such as dictionaries, MT, or internet references were strictly banned, and the whole procedure took approximately 30 minutes.

Data analysis

Time measurements (reading/response times) have been employed in numerous studies in the field of psycholinguistics as reliable data for cognitive processes as they can provide a window into the extent to which participants refer to the texts in different conditions and the ease or difficulty of processing particular texts or questions (Clahsen & Felser, 2006; Roberts & Felser, 2011 among many others). Statistical analyses were conducted with participants' reading times of ST and RMT, response accuracy to comprehension questions, and the time taken to respond to the questions as dependent variables. Independent variables were text type (ST and RMT), text condition (Control, Post-ST, and Pre-ST), and question type (inference, major detail, and minor details related to MT errors).

First, the effect of text type (ST and RMT) and text condition (Control, Post-ST, and Pre-ST) on participants' reading times was examined by using a linear mixed-effect model in the R environment (R Core Team, 2021). Text type (effect-coded as −.5 and .5), text condition (effect-coded as −1, .5, .5 and 0, .25, −.25) and their interactions were set as fixed effects (Jaeger, 2008). The model started out with a maximal random effects structure with by-participant and by-item random intercepts and slopes (Barr et al., 2013), but the model only had by-participant random intercepts as random effect because models with random slopes and by-item random intercepts did not converge.

Second, participants' responses on comprehension questions were coded for accuracy (1 = accurate, 0 = inaccurate), and the effect of text condition and question type (inference, major detail, and minor details related to MT errors) on response accuracy was analyzed using a mixed logit model. The model had text condition and question type (effect-coded as .5, .5, −1 and .25, −.25, 0) as fixed effects and only had by-participant random intercepts as random effect because the models with random slopes failed to converge.

Lastly, the effect of text condition and question type on response times to comprehension questions were analyzed using a linear mixed-effect model with text condition and question type as fixed effects and by-participant and by-item random intercepts as random effects. The relationship between these response times and text reading times was also examined using simple bivariate correlation analyses via Pearson coefficient.

As for the post-task survey, responses to multiple-choice and Likert scale questions were counted and calculated in percentages. Mean and standard deviation values of responses for Likert scale questions were additionally analyzed. The responses for open-ended questions were analyzed using a thematic analysis that centers on key recurring themes (Creswell, 2003).

Results

Online Reading Comprehension Task

The first two research questions were examined by conducting statistical analyses of participants' text reading times (i.e., the time taken for the learners to read ST or RMT), response accuracy to comprehension questions, and the time taken to respond to the questions.

Text Reading Times

Statistical analyses of text reading times revealed that the learners generally spend a longer time reading ST than RMT, but these reading times are dependent on whether RMT is presented before or after ST. Outliers were removed using interquartile ranges, which affected less than 1 percent of all observations, and the remaining reading times were centered and standardized. There was a main effect of text type ($\beta = .95$, SE = .10, $t = 9.16$, $p < .001$) with participants taking significantly longer to read ST than RMT in the Post-ST condition ($\beta = 1.57$, SE = .14, $t = 11.47$, $p < .001$) but this difference was only marginally significant in the Pre-ST condition ($\beta = .33$, SE = .17, $t = 1.89$, $p = 0.07$). There was also a main effect of text condition with texts in the Post-ST condition taking significantly longer to read than those in the Pre-ST condition ($\beta = .88$, SE = .21, $t = 4.27$, $p < .001$). Most importantly, there was a significant interaction between text type and text condition ($\beta = 2.48$, SE = .41, $t = 5.99$, $p < .001$): the time taken to read ST was significantly longer in the Post-ST than the Pre-ST condition ($\beta = 2.12$, SE = .33, $t = 6.47$, $p < .001$), whereas the time taken to read RMT was significantly longer in the Pre-ST than the Post-ST condition ($\beta = .18$, SE = .08, $t = 2.24$, $p = .03$). When ST reading times of the experimental conditions were compared to those of the Control condition, ST reading times in the Control condition were significantly longer than those in the Pre-ST condition ($\beta = .41$, SE = .16, $t = 3.57$, $p < .001$), but no significant difference was found between the Control and Post-ST conditions ($\beta = .32$, SE = .33, $t = .97$, $p = .33$). That is, the participants' tendency to depend on RMT was significantly stronger when they read the RMT before the ST, but when they read the RMT after ST, they predominantly relied on ST for comprehension as they did in the Control condition.

To sum up, the learners generally spend a longer time reading ST than RMT, but the text condition had a significant effect on the reading times: 1) overall reading times were significantly longer for Post-ST than Pre-ST condition, 2) ST reading times in Post-ST and Control conditions were significantly longer than those in the Pre-ST condition, and 3) RMT reading times were significantly longer in the Pre-ST than Post-ST condition. These text reading times are displayed in Table 1.

Table 1
Mean values of reading times by text type and condition in milliseconds (SD in parentheses)

	Control	Post-ST	Pre-ST
ST	193,034.8 (79,950.33)	205,093.13 (76,651.92)	109,667.5 (87,415.06)

RMT	NA	64,186.87 (42,960.51)	80,220.1 (50,194.43)

Response Accuracy

In the statistical analyses of response accuracy to comprehension questions, there was no significant effect of text condition ($\beta = .18$, SE = .14, $t = 1.28$, $p = .20$), but a main effect of question type ($\beta = .46$, SE = .14, $t = 3.26$, $p = .001$) with responses to questions on minor details related to MT errors (Minor-MT, henceforth) being significantly less accurate than responses to Inference ($\beta = 1.20$, SE = .53, $t = 2.26$, $p = .02$) and Major-detail ($\beta = 1.54$, SE = .55, $t = 2.82$, $p = .005$) questions as shown in Table 2. Minor-MT questions were most difficult to answer in all conditions including the Control condition where no RMT was provided. This suggests that the difficulty had more to do with understanding or recalling minor details than being misled by MT errors. Although there was no significant difference between text conditions, mean values of response accuracy to all question types were lowest in the Control condition, which suggests that referring to RMT before or after ST does aid in reading comprehension.

Table 2

Mean values of response accuracy by question type and text condition (SD in parentheses)

	Inference	Major-detail	Minor-MT
Control	.67 (.48)	.63 (.49)	.54 (.50)
Post-ST	.73 (.45)	.73 (.45)	.57 (.50)
Pre-ST	.70 (.47)	.83 (.40)	.59 (.49)

Note. 1 = accurate, 0 = inaccurate

When the correct responses were tallied up as accuracy scores, no statistically significant relationships could be found between text reading times and these scores, most likely because of the small number of comprehension questions and the lack of statistical power. However, when the relationship between text reading times and the accuracy scores were descriptively analyzed for each passage (out of 5 points) in the experimental conditions, there was a general pattern in which longer RMT reading times led to lower accuracy scores and longer ST reading times to higher accuracy scores. This pattern was especially prominent in the accuracy scores for Minor-MT questions in each passage (out of 3 points). Those who relied heavily on RMT seemed to have also accepted its errors which led to lower accuracy scores, while those who paid greater attention to ST generally received higher accuracy scores as can be seen in Table 3 and Figure 2 below.

Table 3

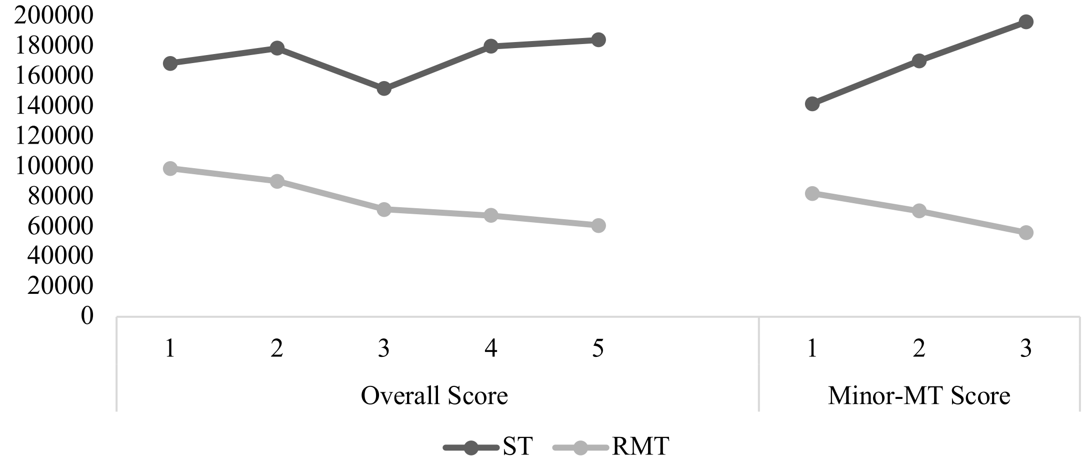
Mean values of reading times by accuracy score per passage (in milliseconds)

Overall Score	1	2	3	4	5
ST	168,676.3	178,717.5	151,924.0	180,025.3	184,286.5
RMT	98,794.5	90,269.8	71,382.9	67,391.9	60,868.6

Minor-MT score	1	2	3
ST	141,788.5	170,399.8	196,193.6
RMT	82,145.2	70,447.7	56,028.1

Figure 2

Mean values of reading times by accuracy score per passage (in milliseconds)



Response Times

For statistical analyses of response times to comprehension questions, outliers were removed using interquartile ranges as has been done for text reading times. This affected approximately 6 percent of all observations, and the remaining data was centered and standardized. The results revealed no significant effect of text condition but significant effects of question type: participants were significantly faster to respond to Inference questions than Major-detail ($\beta = -.96$, SE = .29, $t = -3.29$, $p = .008$) and Minor-MT ($\beta = -.91$, SE = .24, $t = -3.85$, $p = .003$) questions, but there was no significant difference between the response times for Major-detail and Minor-MT questions ($\beta = .05$, SE = .24, $t = .20$, $p = .84$) as displayed in Table 4. This is somewhat surprising because inference questions that do not have a direct answer in the text could be more challenging. It seems that the advanced learners in the present study did not experience much difficulty applying their cultural knowledge to the overall context and understanding of the stories but took longer to answer questions related to major and minor details.

Table 4

Mean response times by question type and text condition in milliseconds (SD in parentheses)

	Inference	Major-detail	Minor-MTs
Control	5858.21 (3142.70)	7861.52 (3885.34)	7492.44 (3545.52)
Post-ST	5550.04 (2531.89)	6843.19 (3013.48)	7395.24 (3182.59)
Pre-ST	6360.83 (2777.66)	7623.07 (2821.06)	7332.26 (3321.01)

The relationship between text reading times and response times to comprehension questions was additionally examined using simple bivariate correlation analyses via Pearson coefficient. Both RMT and ST reading times were positively correlated with question response times (RMT: r = .45, ST: r = .35) at a statistically significant level (*p* < .001), which indicates that the longer the participants read the texts, the longer they took to answer comprehension questions. That is, dwelling on the texts longer seems to have led the participants to ponder on the questions more and entertain different possible answers. The findings from the online reading comprehension task are summarized in Table 5.

Table 5

Fixed effects estimate for text reading time, response accuracy, and response time

	Fixed effects	*β*	SE	*t*	*p*
Text reading time	Text type	.95	.10	9.16	<.001***
	Text condition	.88	.21	4.27	<.001***
	Post-ST (ST > RMT)	1.57	.14	11.47	<.001***
	Pre-ST (ST > RMT)	.33	.17	1.89	.07
	Text type x Text condition	2.48	.41	5.99	<.001***
Response Accuracy	Text condition	.18	.14	1.28	.20
	Question Type	.46	.14	3.26	.001**
Response Time	Text condition	-.05	.06	-.788	.44
	Question Type	-.95	.29	9.86	.008**

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

Post-task Survey

The results of the brief survey at the end of the reading task revealed the advanced learners' perceptions and attitudes towards using RMT for L2 reading. In the first part of the survey that asked about the present experimental task, most participants found the stories difficult and used RMT to aid them in comprehension. When the difficulty of the stories was evaluated on a 5-point scale (1= very easy, 5 = very difficult; Mean= 3.90, SD =.54), 80% of the participants (n=24) indicated that the three stories were "very difficult" or "difficult" and none thought the stories were easy. The majority of the participants read the RMT either all the time (43%, n=13) or partially (50%, n=15), and only two people (6%) skipped all RMTs in the task. Most learners who answered "partially" read parts of RMT in both Pre- and Post-ST conditions (73%, n=11), but a few learners read the RMT only in a certain text condition (Post-ST: 20%, n=3; Pre-ST: 6%, n=1). For those who referred to the RMT (either always or partially), additional questions were asked about its accuracy and usefulness in the present task. Although more learners perceived RMT to be inaccurate (50%, n=14) than accurate (18%, n=5), up to 80% of participants reported that RMT clarified the stories.

Twenty-two learners who found RMT to be helpful in reading comprehension were almost equally divided when asked about the effect of text condition; half of these learners (50%, n=11) indicated it was more helpful to read RMT before ST, while others (45%, n=10) said reading RMT after ST was more helpful. In response to the open-ended question that asked about how RMT was helpful, participants largely mentioned usefulness for gisting and vocabulary. The two most recurring responses were that RMT helped them to gain a general understanding of the story and its context (n=12) and to understand the meaning of unfamiliar words or expressions (n=7).

In the second part of the survey, 80% of the participants (n = 24) indicated that they had previously used MT for reading comprehension tasks in English primarily for reasons related to time efficiency/convenience (n = 13), to look up unfamiliar words/phrases (n = 10), or to check their overall understanding of the texts (n =1). Most of these learners used MT during ST reading (63%, n = 15), but some used MT only in a certain text condition with more people preferring the Pre-ST than Post-ST condition (Pre-ST: 29%, n = 7; Post-ST: 8%, n = 2). "Word/phrase" was selected as the most common language unit for MT input (42%, n = 10) followed by "multiple sentences" (29%, n = 7), "whole passage" (21%, n = 5), and "a single sentence" (8%, n = 2). More than half of these participants (54%, n=13) expressed satisfaction towards MT output, and only 8% (n=2) expressed dissatisfaction. Participants were then asked to elaborate on the reasons why they were satisfied or dissatisfied with MT output in an open-ended question, for which there was a total of 17 responses as shown in Table 6. For those who were satisfied, the reasons included time efficiency, clear translations of unfamiliar words or sentences, and fast delivery of the overall context of the stories, which is similar to the reasons stated for previous MT use in reading comprehension. Additionally, the learners mentioned that MT was helpful in parsing complex sentences as it helped them to chunk long sentences into comprehensible units. For those who expressed dissatisfaction, inaccuracy and awkwardness of translations were viewed as critical flaws of MT, which outweighed its convenience.

Table 6

Reasons for satisfaction or dissatisfaction

	"Why were you satisfied or dissatisfied"
Satisfied	● "time efficient" ● "provides key information and the general context" ● "can quickly find out the meaning of difficult words and phrases" ● "fairly accurate and helps to understand difficult passages" ● "can get a bigger picture of what is happening in the story" ● "helps chunk long sentences into comprehensible units" ● "helps parse complex sentences" ● "can deliver the overall context of the story" ● "can check the meaning of vague or unclear sentences"
Not satisfied	● "inaccurate" ● "awkward translations that are taken out of context" ● "time efficient but translation is not perfect" ● "can understand the gist of the story but cannot fully comprehend" ● "cannot clearly translate difficult sentences" ● "homonyms and idiomatic expressions are inaccurately translated" ● "translations of entire sentences are inaccurate and awkward"

The last part of the survey addressed the learners' future use of MT for L2 reading comprehension and asked them about their willingness to use MT in the future on a scale of 1 to 5. Sixty-three percent of the participants (n=19) said they would use MT in the future, whereas only 10% (n=3) were unwilling. An open-ended question asked the participants to state their reasons for using or not using MT in the future for English reading comprehension, and two key themes—advantages and disadvantages—were developed, some of which resonated with their responses for previous questions on satisfaction and usefulness of MT. The learners who were willing to use MT in the future found MT to be a time efficient tool that quickly delivers the meaning of unfamiliar vocabulary and provides an overall gist of the text fairly accurately. They also mentioned that reading RMT alongside ST increased comprehension. On the other hand, those who were unwilling or undecided about using MT in the future focused on disadvantages such as imprecision or awkwardness of translations that interfered with comprehending ST and mentioned issues related to academic integrity. Overall, it seems that these advanced learners are generally aware of the downsides of heavily relying on MT but view MT as a convenient gisting tool that allows a quick overview of the text. These responses are summarized in Table 7.

Table 7

Reasons for using or not using MT for English reading comprehension in the future

	"Why are you willing or not willing to use MT in the future"
Advantages	● "can better comprehend than when reading English text only" [x5] ● "to look up unfamiliar vocabulary" [x3] ● "to get a rough idea of what the text is about" [x3] ● "to understand the text more quickly (time efficiency)" [x6] ● "fairly accurate" [x4]
Disadvantages	● "inaccurate and awkward" [x2] ● "machine-translated sentences are even harder to interpret" [x3] ● "interferes with reading in English" [x1] ● "using MT for reading comprehension is not appropriate" [x1] ● "using MT for reading (English-to-Korean) is not as efficient as using it for writing (Korean-to-English)" [x2]

Discussion

The present study investigated the effect of using MT output in L2 reading comprehension and found that using RMT can aid comprehension as a supplementary resource, but heavy reliance on RMT can lead to an inaccurate understanding of ST. When the participants read RMT before ST, they often allocated significantly more attentional resources to RMT and significantly less attentional resources to ST than when they read RMT after ST. The advanced learners in the study were largely aware of the inaccuracies of RMT but still agreed that it is a convenient tool for gisting and vocabulary. These findings are discussed in further detail in the order of the three research questions raised in the study.

RQ1: How much do advanced Korean learners of English refer to raw MT output for gisting and how does it affect comprehension?

All but two participants used RMT to inform them of ST either partially or all the time, but the reading times for ST were significantly longer than those of RMT, which indicates that the learners mainly referred to ST for comprehension while using RMT as an additional resource. Although no statistically significant relationships were found between text reading times and accuracy scores of comprehension questions possibly due to the small number of questions and lack of statistical power, mean values of response accuracy were higher in the experimental conditions where RMT was provided before or after ST than the Control condition where only ST was provided. This suggests that reading RMT in addition to ST could lead to higher levels of comprehension. Such results are consistent with previous studies that found raw MT texts to be a reliable resource for gisting and answering comprehension questions (Macken & Ghyselen, 2018; Scarton & Specia, 2016). However, there was a general pattern in which longer ST reading times led to higher accuracy scores, and longer RMT reading times led to lower accuracy scores. It seems that referring to RMT can be helpful only if used with moderation because heavily relying on RMT could also lead to acceptance of its errors. Another notable finding was that the longer the reading times for both ST and RMT, the longer the response times were for comprehension questions. This could mean that the participants who paid greater attention to the texts spent a longer time recalling the details of the stories instead of merely guessing when uncertain.

RQ2: How does text order affect comprehension?

There was a robust effect of text condition (Post-ST and Pre-ST) on text reading times. When participants read RMT after ST (Post-ST), they spent a significantly longer time reading ST than RMT and primarily relied on ST for comprehension as they did in the Control condition where no RMT was provided. When participants read RMT before ST (Pre-ST), however, ST reading times fell drastically, and RMT reading times were significantly longer than those of the Post-ST condition. These results show that the order in which ST and RMT are presented plays a critical role in how much attention the learners allocate to these texts. Although ST is given priority over RMT in all conditions, the participants' tendency to depend on RMT was significantly stronger when they read RMT before ST. The difference in response accuracy between text conditions was not statistically significant in the present study, but there was a general pattern in which longer ST reading times led to higher accuracy scores and longer RMT reading times to lower accuracy scores. That is, the Post-ST condition seemed more conducive to accurate understanding of the texts than the Pre-ST condition. Further studies with a greater number of comprehension questions must be conducted to confirm such speculations.

RQ3: What are the advanced learners' general perceptions and attitude towards using raw MT output for L2 reading?

The advanced learners in the present study were largely aware of the inaccuracies of RMT, but up to 80% of them reported that RMT aided them in the comprehension task by providing a "rough idea" of the story and easy access to the meaning of unfamiliar vocabulary or complex sentences. Some learners, however, expressed ambivalence or dissatisfaction because of imprecise awkward translations that interfered with rather than helped comprehension. Despite the downsides of MT, the majority of the participants were quite confident that MT can be effectively used for L2 reading as a convenient tool for gisting. Most participants reported that they read RMT in both Pre- and Post-ST conditions, but some individuals showed a stronger preference for a certain text condition with more people preferring to read the RMT before than after ST. Given that 40% of the learners selected "multiple sentences" or "whole passage" as the most common language unit for MT input, it seems reasonable to assume that even advanced learners are increasingly using MT for translating longer stretches of L2 text before reading ST. The learners found translations of whole passages mostly useful for gisting and focused on the overall context rather than precision of meaning. As proficient learners, the participants in the present study did not seem to accept the RMT indiscriminately but use it accordingly to complement their understanding of ST. These results resonate with the findings of previous studies where the learners overwhelmingly support the use of MT as a resource for reading comprehension despite its limitations (Clifford et al., 2013; Dennis, 2019; Niño, 2020).

Overall, these findings revealed that the advanced learners in the study were clearly aware of the limitations and inaccuracies of MT output and could use it appropriately to aid comprehension in the present reading task. Unlike the common concerns of language educators that learners' reliance on MT may hinder their use of language knowledge and analytical skills (Niño, 2020), the participants mainly used RMT to gain a general understanding of the story and its context without heavily resorting to the MT output for details. In fact, ST was given precedence over RMT even in the Pre-ST text condition where RMT was presented first. The order in which RMT and ST were presented did influence how long the participants dwelt on each text as the Post-ST condition led to a closer reading of the original text, but the participants used ST for main reference and seemed capable of guessing the meaning for parts that were less accurate in the RMT. The present findings on the learners' performance on the comprehension questions suggest that longer reading times for RMT may be negatively correlated with response accuracy, but this question must be further investigated with a greater number of comprehension questions in future research. Overall, the present results are in line with previous research that found MT to be useful in understanding L2 texts (Karnal & Pereira, 2015) and beneficial for advanced students (Niño, 2009; Tsai, 2019).

Conclusion

This study is the first to examine the effect of using raw MT output on L2 reading comprehension using an online reading comprehension task that measures reading and response times. With an increasing number of learners using MT for large sections of L2 text, it is important to examine the extent to which learners attend to lengthy MT output and how this affects their understanding of ST. The time measurements provide reliable data for how the learners process the texts in real-time, and the present findings have important pedagogical implications for using MT as a language learning tool for L2 reading. First, it is strongly recommended that learners read RMT after ST for checking purposes as an additional resource, and that ST be given a higher priority as the main text. This is because reading RMT before ST gives precedence to RMT and leads one to allocate greater attentional resources to RMT than ST. Longer reading times of RMT were found to result in difficulty in accurately recalling details and a higher acceptance of MT errors. It is highly possible that reading RMT before ST can interfere with cognitive processes in L2 reading as found in Resende and Way (2021) where MT output was found to prime and affect L2 sentence processing. Second, RMT can be helpful for gisting and understanding the overall context but not so much for understanding details, and the learners must read the ST closely for a deeper understanding of the L2 text. Although RMT can be helpful for vocabulary and parsing complex sentences as the learners mentioned in the present study, learners may still be misled by MT errors if there is no careful reading of ST. That is, the purpose and effectiveness of using RMT should be focused on providing a general overview rather than specific content. Lastly, in today's rapidly evolving digital era that highly values efficiency, learners' main reasons for using MT and its output in L2 reading were time-efficiency and convenience. As this could lead to disproportionate dependence on RMT, educators must provide specific guidelines and engage students in tasks that require careful reading and in-depth discussion of reading materials.

The present study is not without limitations. The time measurements are not enough to reveal the underlying comprehension processes, and a greater number of comprehension questions with systematic categories is necessary for a more detailed analysis of response accuracy. Moreover, the present experimental setup did not allow for an examination of the use of RMT *during* reading, and future works must employ other experimental methods such as eye-tracking to tease apart the times spent on ST vs. RMT during reading. It must also be pointed out that the present study focused on advanced learners and literary texts with advanced Lexile scores, and the results may look different depending on the proficiency level, text difficulty, and text genre. That is, the present implications may not be wholly applicable to lower proficiency learners or easier texts in a different text genre, and the effects of RMT on reading and the educational value of MT as a language learning tool needs to be further elucidated with different populations and reading materials.

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.

Alhaisoni, E., & Alhaysony, M. (2017). An investigation of Saudi EFL university students' attitudes towards the use of Google Translate. *International Journal of English Language Education, 5*(1), 72-82.

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.

Bavendiek, U. (2022). Using machine translation as a parallel text to access literature for modern language learning. In C. Hampton, & S. Salin (Eds.), *Innovative language teaching and learning at university: facilitating transition from and to higher education* (pp.68-78). Research-publishing.net. http://doi.org/10.14705/rpnet.2022.56.1373

Bower, J. (2010). Japanese university students' use of online machine translators for English writing tasks. *Studies in Linguistics and Language Teaching*, *21*, 1–19.

Brevik, L. M., Olsen, R.V.. & Hellekjær, G. O. (2016). The complexity of second language reading: Investigating the L1-L2 relationship. *Reading in a Foreign Language*, *28*(2), 161-182.

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.

Chen, W. (2020). Using Google Translate in an authentic translation task: The process, refinement efforts, and students' perceptions. *Current Trends in Translation Teaching and Learning E*, *7*, 213-238.

Chang, C-K., & Hsu, C-K. (2011). A mobile-assisted synchronously collaborative translation–annotation system for English as a foreign language (EFL) reading comprehension. *Computer Assisted Language Learning*, *24*(2), 155-180. http://doi.org/10.1080/09588221.2010.536952

Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. The Journal of Asia TEFL, *17*(1), 182–193.

Chung, E. S. (in press). Second language processing of errors in Korean-to-English machine-translated output. *Language Learning & Technology.* https://doi.org/10.35542/osf.io/y4q35

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*, 3–42. https://doi.org/10.1017/S0142716406060024

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *The Acquisition of Second Languages and Innovative Pedagogies*, *10,* 108–121. http://doi.org/10.7203/attic.10.2228

Cook, G. (2007). A thing of the future: Translation in language learning. *International Journal of Applied Linguistics*, *17*(3), 396–401. http://doi.org/10.1111/ijal.2007.17.issue-3

Creswell, J. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Dennis, M. O. (2019). Learner Perceptions of Using Machine Translation Tools in the EFL Classroom, *The SNU Journal of Education Research, 28*(2), 63-83.

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

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.

Jeon, E. H., & Yamashita, J. (2014). L2 Reading Comprehension and Its Correlates: A Meta-Analysis. *Language Learning*, *64*(1), 160-212.

Jones, C., & Oakey, D. J. (2019). Literary dialogues as models of conversation in English language teaching. *Journal of Second Language Teaching and Research, 7*(1), 107-135.

Karnal, A. R., & Pereira, V. V. (2015). Reading Strategies in a L2: A Study on Machine Translation. *The Reading Matrix: An International Online Journal, 15*(2), 69-79.

Kern, R. G. (1994). The role of mental translation in second language reading. *Studies in Second Language Acquisition*, *16*(4), 441–461. http://doi.org/10.1017/S0272263100013450

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

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

Lee, S.-M. (2021). The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis. *Computer Assisted Language Learning*, http://doi.org/10.1080/09588221.2021.1901745

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. http://doi.org/10.3758/s13428-011-0146-0

Li, P., & Clarina, R. B. (2019). Reading comprehension in L1 and L2: An integrative approach. *Journal of Neurolinguistics, 50,* 94-105. http://doi.org/10.1016/j.jneuroling.2018.03.005

Macken, L., & Ghyselen, I. (2018). Measuring Comprehension and User Perception of Neural Machine Translated Texts: A Pilot Study. *Proceedings of the 40th Conference Translating and the Computer*, *UK*, 120–126.

Nakata, T., Tamura, Y., & Aubrey, S. (2020). Examining the Validity of the LexTALE Test for Japanese College Students. *The Journal of AsiaTEFL*, *17*, 335-348. http://doi.org/10.18823/asiatefl.2020.17.2.2.335

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

Niño, A. (2020). Exploring the use of online machine translation for independent language learning. *Research in Learning Technology*, *28*. http://doi.org/10.25304/rlt.v28. 2402

Niño, A. (2022). Online Translators in Online Language Assessment. *CALL-EJ, 23*(3), 115-135.

Nourinezhad, S., & Kashefian-Naeeini, S. (2020). Iranian EFL university learners and lecturers' attitude towards translation as a tool in reading comprehension considering background variables of age, major and years of experience, *Cogent Education*, *7*(1), 1746104. http://doi.org/10.1080/2331186X.2020.1746104

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Resende, N., & Way, A. (2021) Can Google Translate Rewire Your L2 English Processing? *Digital, 1,* 66–85. http://doi.org/10.3390/digital1010006

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

Scarton, C., & Specia, L. (2016). A Reading Comprehension Corpus for Machine Translation Evaluation. *Proceedings of the Tenth International Conference on Language Resources and Evaluation, Slovenia (LREC'16)*, European Language Resources Association (ELRA), 3652–3658. https://aclanthology.org/L16-1579

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