

Effects of Strategy Instruction on Machine Translation Use for Reading

Caleb Prichard

Okayama University

Reference Data:

Prichard, C. (2026). Effects of strategy instruction on machine translation use for reading. In B. Lacy, M. Swanson, & H. Yoder (Eds.), *Language S: Learning, Teaching, Assessing – JALT 50 Years – Challenges and Perspectives*. JALT. <https://doi.org/10.37546/JALTPCP2025-01>

Research has suggested that machine translation (MT) is considered extremely useful for L2 readers, but it can be abused. Scholars have therefore called for strategy training on MT usage. However, whether instruction on using MT strategically leads to more selective usage and increased content recall has not been examined. A pre-, posttest control group design was utilized to examine the efficacy of strategy instruction. Data collection involved screen recordings of participants' page navigation to determine frequency of MT access and to compare MT use of difficult and easy passages. Students were given an immediate recall test to measure comprehension and short-term retention of the content. The results suggested that strategy training led to reduced MT use, but with no significant results between the two groups. The experimental group did not improve recall scores overall, but those who reduced their MT use for easy paragraphs tended to improve their recall scores.

機械翻訳 (MT) は第二言語読者にとって非常に有用であるが、乱用される可能性があることが研究によって示唆されている。そのため、学者はMT使用に関するストラテジー指導の必要性を訴えてきた。しかし、こうした指導が選択的な使用や内容想起の向上につながるかについては、十分に検討されてこなかった。本研究では、ストラテジー指導の効果を検証するために、事前・事後テストを用いた統制群デザインを採用した。データ収集には、参加者の閲覧行動を画面録画し、MTアクセスの頻度や、難易度の異なる文章でのMT使用を比較した。学生には、内容の理解度と短期保持を測定するために、即時想起テストを実施した。結果として、ストラテジートレーニングによってMT使用の減少が促されたが、統制群と実験群の間に有意差は認められなかった。実験群全体の想起スコアは改善しなかったが、易しい段落でMT使用を減らした参加者はスコアが向上する傾向を示した。

DeepL Translator, Google Translate, and other machine translation (MT) tools can enable comprehension of L2 texts (Wei, 2021); at the same time, overuse of MT may prevent language acquisition and proficiency development (Murtisari et al., 2024; Musk, 2014). Frequently relying on translations rather than carefully reading comprehensible texts lessens exposure to the target language, hindering vocabulary acquisition (Wang & Pellicer-Sánchez, 2022). MT use may also hamper the retention of comprehended content because reading translations may lead to shallower processing compared to working through L2 texts (Prichard & Atkins, 2024, 2025). These studies also suggested that using MT for *difficult* passages can aid in comprehension, but hastily accessing MT instead of closely reading *comprehensible* passages can lead to less content and vocabulary form recall.

Since translation overreliance may affect reading outcomes, it has frequently been argued that learners may benefit from strategy instruction on using MT more selectively (Lee, 2023; Merschel & Munné, 2022; Murtisari et al., 2024). To address this gap in the literature and to inform L2 reading instruction, this study used a pre-, posttest control group design to examine the effect of an intervention on the strategic use of MT, focusing on less translation use for comprehensible passages.

Literature Review

MT Use of L2 Readers

Thanks to its effectiveness and convenience, MT use has become an extremely common tool among L2 readers to comprehend texts, confirm understanding, and to check vocabulary (Klimova, 2025). Surveys have suggested that looking up individual words was the most common use, but a majority sometimes translate larger chunks of text while reading (e.g., Murtisari et al., 2019).

One limitation of most prior studies on MT use is their reliance on surveys, as the accuracy of self-report measures may be compromised because participants are not

always conscious of their strategy use or they may give idealized responses (e.g., Bax & Weir, 2012). There have been a handful of studies using screen recording or eye tracking to more reliably capture L2 readers' actual behavior. One small-scale study in a Swedish English classroom utilized screen recording and found that half translated the whole page to avoid reading the English text (Musk, 2014). Musk argued that such language avoidance by students threatened the integrity of the reading activities, potentially hindering proficiency development.

However, subsequent studies of university L2 readers have found relatively little language avoidance. In a study of proficient Korean English learners, participants spent longer reading the English text than translations, and some skipped translations altogether (Chung, 2023). In an eye tracking study of Japanese English learners (Prichard & Atkins, 2024), readers tended to try to work out meaning on their own, as evidenced by re-reading sentences with pseudowords. Participants accessed paragraph-level MT before reading the English paragraph just 6% of the time. In a follow-up study utilizing screen recording (Prichard & Atkins, 2025), half did not use MT at all even though fifty percent of the text was above their proficiency level. A survey revealed most of these non-users of MT had low cognitive reading anxiety, suggesting that they had resilience to cope with difficult texts on their own.

Nevertheless, these studies have shown that some learners are over-reliant on MT. A majority of participants checked translations even in cases where pseudowords were accompanied by clear context cues (Prichard & Atkins, 2024). Anxiety has also been found to be related to MT use. While readers with low cognitive anxiety avoided MT, as noted above, those who reported significantly higher levels of anxiety used MT six times more often, translating big chunks of text (i.e., whole paragraphs) 36 times more often (Prichard & Atkins, 2025).

MT Use on Reading Outcomes

L2 readers often think that MT use can lead to vocabulary learning and comprehension (e.g., Murtisari et al., 2019; Oh, 2022); studies comparing reading outcomes with and without translation allocation (not necessarily MT) have indeed shown that L1 translations can increase vocabulary learning (e.g., Grace, 1998) and comprehension (e.g., Dirix et al., 2020). However, while translation assists comprehension, it does not necessarily mean better retention of content since reading translated passages may lead to more shallow processing, which in turn may lead

to weaker memory traces (Craik, 2002). Oh's study (2022) found that MT allocation affected comprehension scores only marginally (17.59 versus 18.96), and two other studies (Prichard & Atkins, 2024, 2025) showed that frequently accessing MT was not significantly correlated with greater content recall. Moreover, translation access can compete with the L2 for attention, hindering vocabulary form recall (Wang & Pellicer-Sánchez, 2022).

Because of concerns that MT reliance and language avoidance can have detrimental effects, scholars argued that learners need to develop competence for utilizing MT more selectively (Lee, 2023; Merschel & Munné, 2022). One key factor distinguishing strategic MT use from overuse may be the comprehensibility of the text. While L2 readers can make inferences when they have high lexical coverage of the text, they have great difficulty doing so when the coverage is low (e.g., Kremmel et al., 2023). Therefore, MT use may be unnecessary if a participant has enough vocabulary and reading proficiency to deal with the text on their own, but MT use may be helpful if the text is too difficult (Oh, 2023).

Prichard & Atkins (2024, 2025) found some evidence for this. MT use was more common for passages with difficult vocabulary, and frequently accessing translations for difficult passages led to somewhat better comprehension. In addition to helping understand novel lexica, the studies also suggested that it may have aided comprehension and recall by reducing cognitive load and lowering cognitive reading anxiety. In contrast, using MT for comprehensible passages was not associated with greater content recall, and tending to access MT before reading these easy passages correlated with lower recall.

MT and Reading Strategy Instruction

Many researchers have called for strategy training to prevent the overreliance on MT. Strategy instruction aims to enhance comprehension by teaching cognitive skills, such as inferring meaning and summarizing, and metacognitive strategies, like considering one's goal and monitoring comprehension (Grabe, 2009). Strategy instruction often involves explicit instruction, teacher modeling, independent practice, collaborative work, and reflection. Although several factors, such as the type of instruction, the strategy taught, and the length of instruction, influence the efficacy of the training, meta-analyses have suggested that L2 reading strategy instruction tends to be effective (Taylor et al., 2006; Yapp et al., 2021).

Methodology

As research has suggested a need for strategy instruction focusing on lessening MT overuse, this research examined the effect of such an intervention. It aimed to answer the following research questions. First, does strategy training reduce the frequency of MT use, particularly for easy passages? Second, does strategy training affect content recall of the passages? The study replicated previous MT research (Prichard & Atkins, 2025), utilizing the same materials, procedures, and analysis of both MT usage and content recall. However, this follow-up study involved two tests and an intervention for a pre-, posttest control group design to answer the research questions.

Participants

Forty-three students from the author's two second-year reading and writing classes at a national university in Japan participated in the study. One class was set as the experimental group ($n = 20$), which received the intervention described below, and the other class was set as the control group ($n = 23$). The two classes had similar proficiency scores before the term started. The experimental group's mean Linguaskill (4-skills) score was 153.44 ($SD = 4.36$) and the control group's was 154.00 ($SD = 7.85$). Informed consent was collected following institution rules for ethics.

Materials

Two 600-word texts on an American high school custom were used. One text advocated for school dances, and the other advocated for seasonal sports clubs. Each described four advantages of the custom in eight paragraphs, two paragraphs describing each advantage. Participants were randomly assigned one text in the pretest, and they read the other in the posttest.

To examine participants' use of MT according to passage difficulty, paragraphs were edited so that they were well below or well above the participants' proficiency level. While the participants were estimated to be at grade levels 8 or 9 based on their Linguaskills scores and Johnston's (2024) analysis, the easy paragraphs were edited to be at the grade level 4 (Kincaid et al., 1975) so that MT use would not be necessary. The difficult paragraphs were at level 15, meaning MT use could theoretically be used to enhance comprehension. The easier paragraphs all had at least 92% of words in the top 2,000 on the Classic General Service List (Cobb, 2020), while the harder texts each had less than 78% of words in this range. The Coh-Metrix L2 Reading Index (RDL2;

McNamara et al., 2014) confirmed that easy ($RDL2 = 8.08$) and difficult paragraphs ($RDL2 = 24.99$) differed greatly in readability.

Immediate Recall Test

An English immediate recall test was used to evaluate comprehension and retention of content as in Prichard and Atkins (2025). Immediate recall tests have long been considered valid and reliable as they measure both comprehension and short-term retention while avoiding the guessing and the ceiling effects of multiple-choice quizzes (Alderson, 2000; Bernhardt, 1983). Students were asked to write down all of the content they remembered from the text, including details and examples. For scoring, idea units were counted. Idea units are commonly defined as minimal units of meaning, often operationalized as subject-predicate structures or phrases representing a single idea. Each text contained 84 idea units. Each idea unit recalled was scored as one point regardless of grammar or spelling errors.

Procedures

The pre- and posttest readings took place in a room with desktop computers. Since learners often assume MT use is considered cheating (Merschel & Munné, 2022), participants were instructed to create an anonymous code. Participants were informed that their screens would be recorded but that the recordings were anonymous.

Participants were then told that they would read one text advocating for an American school custom and that after closing the text they would later write all of the content that they remembered. It was explained that MT use while reading was acceptable, and links to DeepL and Google Translate were provided. After screen recording began, participants answered pre-reading questions to activate their background knowledge. Then they opened the article, which they had ten minutes to read. They then closed the text, and they had ten minutes for the written recall task. Screen recordings were saved to the desktop and transferred to the researcher's device.

Intervention

The intervention included three hours of instruction by the author on strategic MT usage (e.g., avoiding MT use for comprehensible passages) for the experimental group, divided over a period of seven weeks. During this time, the control group instead focused on other skills, such as paraphrasing and discussing texts.

One week after the pretest, the first part of the intervention took place, and it lasted a full lesson (100min.). The experimental group discussed whether they used MT for reading, and, if so, when and how. The instructor then gave explicit guidelines based on prior research. For example, it was explained in simple terms that L2 readers may be able to infer meaning without MT use when they have high lexical coverage. In contrast, it was pointed out that MT may be beneficial if lexical coverage is lacking since inferring meaning is often impossible in such cases. Students were shown sentences with “XXXX” representing unknown words, and they discussed in groups which strategy they would use: trying to infer meaning, dictionary use, MT use, or reading on. The instructor then gave suggestions. For example, for “One of the biggest issues is XXXXXX since it is very XXXXX,” it was pointed out that it may be best to translate the sentence due to multiple unknown words. In contrast, for “The snake XXXXX through the grass,” it was suggested that students could likely infer meaning, making MT use unnecessary.

Students were reminded that frequently accessing MT instead of reading English texts could affect reading fluency development and lead to less attention to novel L2 vocabulary. The author also explained that aiming to read and re-read somewhat challenging texts leads to deeper processing, which could in turn lead to greater content recall and higher chances words would be acquired.

Throughout the four lessons in the following weeks, the experimental group had shorter follow-up instruction (20min. per lesson) on MT. They reviewed the above guidelines and did activities trying to infer meaning from context without relying on MT or dictionaries. To enhance metacognitive skills, the experimental group class was allowed MT use throughout the term but were told that they may be asked why they were using it. The last MT-related instruction for the experimental group was one week before the posttest.

Analysis

Based on the partially replicated study (Prichard & Atkins, 2025), screen recordings were analyzed to determine participants’ MT use, counting the number of times used for easy or difficult passages. Strategic MT use was defined as avoiding translating content from easy passages.

In order to find out the effect of strategy training on frequency of MT use, the following data analysis was conducted. Due to unexpected baseline differences in MT use between groups (control: $M = 1.85$; experimental: $M = 1.05$), ANCOVA was used to analyze posttest MT use (DV, scale) with the group as the independent variable (categorical: experimental vs. control) and pretest MT use as a covariate (scale),

controlling for initial differences. In order to determine whether strategy training affects content recall, an independent sample t -test was used to compare recall score gains (posttest minus pretest) between the two groups.

Results

Effect of Strategy Training on MT Use

Roughly half of the participants used MT in the pretest, including 11 of the 20 in the control group (55%) and nine of 19 in the experimental group (47%). In the pretest, most of the MT uses were for difficult texts, as is shown in Table 1. The pretest results showed unexpected baseline group differences, with the control group using MT more.

Table 1

The Mean Number of MT Uses per Test and Text Type for Both Groups

	Pre		Post	
	Experimental	Control	Experimental	Control
Text				
Difficult	.89	1.45	.21	.85
Easy	.16	.40	.00	.30
Total	1.05	1.85	.21	1.15

In the posttest, the experimental group had an 80% reduction in MT uses, and the control group had a 37% reduction. The decline for the experimental group was statistically significant, $t(19) = 2.73$, $p = .01$, $d = 0.61$, but it was not significant for the control group, $t(22) = .86$, $p = .40$, $d = 0.19$. The experimental group completely avoided translating chunks of text (i.e., whole paragraphs) in the posttest after doing so eight times in total in the pretest. However, overall, the experimental group’s reduction in the number of MT uses was not significant compared to that of the control group, $F(1, 36) = 3.52$, $p = .069$, $\omega_p^2 = .061$.

In the posttest, the experimental group had no uses of MT for easy texts. As no variance existed in the experimental group’s posttest easy text scores, inferential analysis was not possible. On difficult passages, after controlling for pre-test MT use, there was no significant effect of group on post-test MT use, $F(1, 36) = 3.20$, $p = .082$, $\omega_p^2 = .053$.

Effect of Strategy Training on Content Recall

Content recall scores of the experimental group did not vary from the pre- ($M = 5.87$, $SD = 1.39$) to posttest ($M = 5.79$, $SD = 1.23$). Scores of the control group likewise did not significantly increase (Pretest: $M = 4.95$, $SD = 2.61$; Posttest: $M = 5.40$, $SD = 2.28$). Overall, there was no significant difference in gains between the two groups, $t(37) = .96$, $p = .34$.

Given that the experimental group did not access MT for easy passages without corresponding gains in recall, post hoc analyses were conducted to explore the relationship overall between MT use and recall performance. Across all participants and testing occasions, the number of MT uses for easy passages was weakly negatively correlated with recall scores, $r(76) = -.26$, $p = .01$. Within the experimental group, reductions in MT usage for easy passages from pre- to posttest were associated with small gains in recall, $r(19) = -.29$, $p = .04$. In contrast, the number of MT uses for difficult passages had no significant correlation with recall scores across all participants and tests, $r(76) = -.06$, $p = .30$.

Discussion

Overall, although the experimental group showed significant declines in their MT usage, there was no significant difference compared to the control group. Many participants completely avoided MT in the first place, so much of the intervention was not relevant to much of the class. Nevertheless, the experimental group participants who did access translations in the pretest greatly reduced their usage (just 15% of their pretest total), and they avoided all paragraph-level translations. This suggests that the intervention may be useful for the student populations who are more dependent on MT.

There were no significant gains in content recall scores for either the experimental or control group. Once again, one explanation was that the experimental group did not significantly reduce their MT use since many did not access translations in the pretest. Nevertheless, reductions in MT use were weakly associated with gains in content recall.

However, another possible reason for the lack of significant change in content recall scores may simply be that MT use may not affect content recall as much as expected. Overall, the negative correlation between content recall and the number of MT uses was relatively weak. Moreover, MT use for difficult texts was not positively associated with content recall, despite expectations.

Limitations

As noted above, the main limitation of the study was that many participants rarely used MT even before the intervention. As a result, there was limited room for change, reducing the likelihood of detecting intervention effects. In addition, the time constraints imposed on the recall task may have limited participants' ability to fully demonstrate comprehension and recall differences. Some participants were still completing their responses when time expired. Future studies should employ untimed recall protocols to more fully capture potential intervention effects on comprehension. Follow-up research should also examine the effect of MT use on other outcomes, such as vocabulary recall, reading fluency, and cognitive reading anxiety.

Conclusion

Rather than relying on surveys, this study utilized screen recording to more reliably examine how frequently participants used MT for reading. As scholars have suggested that L2 readers need to develop competence in MT use, an intervention was carried out aiming to decrease overreliance on translations, especially reducing MT use for comprehensible passages.

In the posttest, the experimental group had no uses of MT for easy texts. However, the instruction did not lead to significantly less MT use for the experimental group than the control group overall since relatively few participants used MT in the pretest. The participants in this study were more proficient than the average student at the institution, and they may have been more self-reliant and confident. Prior research suggested that readers with high cognitive reading anxiety are much more reliant on MT (Prichard & Atkins, 2025). Instruction aiming to reduce overreliance on MT may be more appropriate for other student groups.

While the experimental group participants showed some signs of more selective use of MT (i.e., not using MT for comprehensible texts), this was not enough to impact their ability to recall content overall. Selective MT use may not affect content recall use as much as expected. There was a significant correlation between using MT for easy passages and recall scores, but it was a weak correlation.

However, considering the possible negative effects of translation overreliance on reading fluency and vocabulary form recall, recalling the same amount of content while using less MT could be considered a positive sign. Experimental group participants were more self-reliant, spending more time on the L2 instead of translations, and their content recall was not impacted. From this perspective, the intervention may have had pedagogical value.

Bio Data

Caleb Prichard is an Associate Professor from Okayama University. In addition to researching reading strategies, he researches L2 humor competence and language program administration.

References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University.
- Bax, S., & Weir, C. J. (2012). Investigating learners' cognitive processes during a computer-based CAE Reading test. *Cambridge ESOL Research Notes*, 47, 3–14.
- Bernhardt, E. B. (1983). Testing foreign language reading comprehension: The immediate recall protocol. *Die Unterrichtspraxis*, 16, 27–33.
- Chung, E. S. (2023). L2 Learners' use of raw machine-translated output in reading comprehension. *Computer-Assisted Language Learning Electronic Journal*, 24(3), 1–19.
- Cobb, T. (2020). Web vocabprofile v.5 [Computer software]. <http://www.lexutor.ca/vp/b>
- Craik, F. I. (2002). Levels of processing: Past, present... and future? *Memory*, 10(5–6), 305–318. <https://doi.org/10.1080/09658210244000135>
- Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., & Duyck, W. (2020). Reading text when studying in a second language: An eye-tracking study. *Reading Research Quarterly*, 55(3), 371–397. <https://doi.org/10.1002/rrq.277>
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Grace, C. A. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning level CALL software. *Modern Language Journal*, 82, 533–544. <https://doi.org/10.1111/j.1540-4781.1998.tb05541.x>
- Johnston, M. C. (2024). A comparison of text analysis tools: Levels of agreement and disagreement. *SHIKEN*, 28(1), 38–55.
- Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel. *Research Branch Report*, 8–75. Chief of Naval Technical Training: Naval Air Station Memphis.
- Klimova, B. (2025). Use of machine translation in foreign language education. *Cogent Arts & Humanities*, 12(1), 2491183. <https://doi.org/10.1080/23311983.2025.2491183>
- Kremmel, B., Indrarathne, B., Kormos, J., & Suzuki, S. (2023). Unknown vocabulary density and reading comprehension: Replicating Hu and Nation (2000). *Language Learning*, 73(4), 1127–1163. <https://doi.org/10.1111/lang.12622>
- Lee, S. M. (2023). The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis. *Computer Assisted Language Learning*, 36(1-2), 103–125. <https://doi.org/10.1080/09588221.2021.1901745>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- Merschel, L., & Munné, J. (2022). Perceptions and practices of machine translation among 6th-12th grade world language teachers. *L2 Journal*, 14(1), 60–76. <https://doi.org/10.5070/L214154165>
- Murtisari, E. T., Kristianto, A. K., & Bonar, G. (2024). Self-directed use of machine translation among language learners: Does it lead to disruptive L2 avoidance? *Foreign Language Annals*, 57(4), 1094–1114. <https://doi.org/10.1111/flan.12768>
- Murtisari, E. T., Widiningrum, R., Branata, J., & Susanto, R. D. (2019). Google Translate in language learning: Indonesian EFL students' attitudes. *Journal of Asia TEFL*, 16(3), 978. <http://dx.doi.org/10.18823/asiatefl.2019.16.3.14.978>
- Musk, N. (2014). Avoiding the target language with the help of Google: Managing language choices in gathering information for EFL project work. *TESOL Quarterly*, 48(1), 110–135. <https://doi.org/10.1002/tesq.102>
- Oh, E. (2022). Exploratory study on the use of machine translation for reading in college English classes. *Multimedia-Assisted Language Learning*, 25(4), 66–92. <https://doi.org/10.15702/mall.2022.25.4.66>
- Prichard, C., & Atkins, A. (2024). Examining readers' use of machine translation through eye tracking. *System*, 125, 103415. <https://doi.org/10.1016/j.system.2024.103415>
- Prichard, C., & Atkins, A. (2025). Variables affecting the use and outcomes of machine translation for reading: Text difficulty and cognitive reading anxiety. *System*, 134, 103822. <https://doi.org/10.1016/j.system.2025.103822>
- Taylor, A., Stevens, J., & Asher, J. W. (2006). The effects of explicit reading strategy training on L2 reading comprehension. In L. Ortega & J. M. Norris (Eds.), *Synthesizing research on language learning and teaching* (pp. 213–244). John Benjamins.
- Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning* 72(3), 765–805. <https://doi.org/10.1111/lang.12495>
- Wei, L. K. (2021). The use of Google Translate in English language learning: How students view it. *International Journal of Advanced Research in Education and Society*, 3(1), 47–53. <https://myjms.mohe.gov.my/index.php/ijares/article/view/12459/6181>
- Yapp, D. J., de Graaff, H. C. J., & van den Bergh, H. H. (2021). Improving second language reading comprehension through reading strategies: A meta-analysis of second language reading strategy interventions. *Journal of Second Language Studies*, 4(1), 154–192. <https://doi.org/10.1075/jsls.19013.yap>