

Machine Translation in English Writing Education: A Comparative Study of Post-Editing by University Students with Different L2 Proficiency and Genre Knowledge

Miho YAMASHITA*

Abstract:

In recent years, generative AI tools for English writing, particularly Machine Translation (MT), have been increasingly integrated into foreign language teaching, especially in writing instruction. While MT has demonstrated potential benefits in improving linguistic accuracy and fluency, it remains imperfect, necessitating effective pre-editing and post-editing processes to achieve accurate translations (Lee 2020). Post-editing, in particular, is often challenging for learners with lower proficiency in the target language (L2) (Yamada 2021). However, little is known about how students with varying levels of L2 proficiency and genre knowledge engage with MT in educational contexts, particularly in Japan. This lack of investigation makes it challenging to develop instructional strategies that are both effective and tailored to diverse learner needs. To address this gap, the present study investigates the post-editing behaviors of Japanese undergraduate and graduate students at Ritsumeikan University. It examines how learners' English proficiency and familiarity with genre conventions affect their revision strategies when translating Japanese texts into English. The findings reveal that undergraduate students, especially those with lower proficiency, made minimal revisions and often accepted MT outputs uncritically. In contrast, graduate students, especially those with higher L2 proficiency and deeper understanding of genre conventions, were able to engage in more accurate and contextually appropriate revisions. These students actively reconstructed sentences and refined technical terminology, suggesting that genre familiarity also plays a critical role in effective post-editing alongside language proficiency. This study underscores the importance of developing instructional strategies that not only consider learners' proficiency and genre awareness but also empower students to critically engage with MT outputs and enhance their academic writing skills.

Keywords: *Machine translation (MT), post-editing, L2 proficiency levels, knowledge of genre*

* Professor, College of Life Sciences, Ritsumeikan University.

Email: mihoron@fc.ritsumei.ac.jp

Received on 2025/1/11, accepted after peer reviews on 2025/7/10.

1. Introduction

In recent years, a variety of generative AI tools have emerged, with Machine Translation (hereafter MT) being increasingly incorporated into foreign language education, especially for teaching L2 writing skills. Since 2016, the performance of neural machine translation has significantly improved, allowing students to use MT as a “good model” from which to learn (Klimova et al. 2022). Given this background, previous research at the university level has examined the effectiveness of MT use. Although the results have been mixed, a thirty-year review of past research showed that using MT has a positive effect in EFL settings (Jolley and Maimone 2022). MT can produce better texts in various linguistic subcategories such as lexical accuracy, syntactic accuracy, orthography, and overall writing quality (Lee 2020). For instance, a study by Chon et al. (2021) found that MT generates texts with higher lexical sophistication and syntactic complexity, thereby reducing the disparity between skilled and less skilled writers. This makes MT-assisted writing particularly beneficial for those with lower language proficiency, as it reduces the cognitive load and anxiety associated with writing in L2 (Garcia and Pena 2011).

Despite these potential benefits, perceptions towards MT use differ between students and educators. While students generally view MT as a useful tool for L2 writing, educators remain more skeptical, mainly due to concerns about historical inaccuracies¹ and potential overdependence on MT. MT is still not perfect; therefore, to achieve accurate translations when using MT, it is essential to pre-edit the original manuscript written in L1 and post-edit the translated L2 text to ensure they are translated as initially intended (Lee 2020). Through the process of post-editing, students can identify and correct errors, thereby enhancing their linguistic awareness (Lee 2020; Niño 2008). However, the effectiveness of post-editing can vary based on learner variables, particularly L2 proficiency levels. As MT can produce texts that far exceed the writing abilities of students with lower L2 proficiency, they may not be able to effectively post-edit and may simply incorporate MT outputs without modification. Other learner variables, such as knowledge of the genre or the contextual information of the texts, also affect their revising strategies and writing outcomes (Lee 2009; Qi and Lapkin 2001; Yamada 2021). Post-editing quality is higher when learners are familiar with the genre or topic, as it enables them to better understand the content and make more accurate corrections (Lee 2022; Niño 2009).

To address the gap in studies examining the effectiveness of MT use in writing, this study compares post-editing practices on texts written by undergraduate and graduate students in the College of Life Sciences at Ritsumeikan University. Students are encouraged to improve their English skills from the undergraduate level to become researchers active on the global stage, and writing papers becomes inevitable as they enter graduate school. The effective use of MT can undoubtedly help them in future academic paper writing. The author hopes this paper will provide insights into the potential benefits and challenges of MT use in higher education, contributing to a more nuanced understanding of its role in language learning and writing development.

1 *Historical inaccuracies* denote textual or grammatical inaccuracies that were typical of pre-neural MT systems. These limitations led many educators to approach the use of MT in language education with caution or doubt.

1. Literature Review

(1) Past Studies on the Use of Machine Translation in L2 learning

The use of Machine Translation (MT) in foreign language learning, particularly in L2 writing, has garnered significant attention in recent years. A number of studies have examined its effectiveness, revealing both advantages and limitations. MT tools, such as neural machine translation systems (NMT) developed since 2016, have demonstrated the ability to enhance L2 writing by improving various linguistic elements, including lexical accuracy, syntactic accuracy, orthography, and overall writing quality (Lee 2020). For instance, research by Chon et al. (2021) found that MT can reduce the disparity between skilled and less-skilled writers by enhancing lexical sophistication and syntactic complexity. This suggests that MT-assisted writing can provide a valuable support mechanism, particularly for learners with lower L2 proficiency, as it reduces cognitive load and anxiety (Garcia and Pena 2011).

Research has demonstrated that post-editing can significantly enhance learners' linguistic skills by helping them notice and correct errors, which fosters metalinguistic awareness and a deeper understanding of language use (Lee 2020; Niño 2008). The process encourages reflection, problem-solving, and critical thinking, as students evaluate both their own writing and the MT-generated outputs (Clifford et al. 2013; Niño 2009). Post-editing thus supports autonomous, self-directed learning by allowing students to actively engage with the language and take ownership of their learning process.

Despite these potential benefits, there are notable challenges and drawbacks to using MT in L2 writing. Several studies have reported mixed results regarding its effectiveness. For example, Chung and Ahn (2021) observed that MT's effectiveness varies significantly depending on proficiency levels and text genres, with notable improvements in accuracy but unclear benefits in syntactic and lexical complexity. Moreover, while MT can assist lower-proficiency learners by reducing their cognitive load, it may not promote as much learning as writing directly in L2, as noted by Garcia and Pena (2011). Over-reliance on MT is another concern, as it can lead to superficial learning, particularly when students neglect deeper linguistic understanding and critical thinking.

Educators also express skepticism regarding MT's use in language classrooms, primarily due to concerns over historical inaccuracies and the risk of over-dependence on technology (Lee 2022). This skepticism is supported by findings that less proficient students tend to incorporate MT into their writing more readily, while higher proficiency learners are often more cautious, avoiding unfamiliar expressions suggested by MT (Yamada 2023).

(2) Past Studies on Post-editing in MT

Post-editing plays a critical role in maximizing the benefits of MT in L2 writing. Research has highlighted the importance of post-editing from various perspectives, including its merits (Alsalem 2019; Escartín et al. 2017; Fredholm 2019), the post-editing process itself (Chung 2020; Jia et al. 2019; Shin and Chon 2023), and user experiences or familiarity with the genre (Harto et al. 2022). The post-editing process can significantly enhance learners' linguistic skills by highlighting errors, encouraging reflection, and fostering metalinguistic awareness (Lee 2020; Niño 2008). Furthermore, post-editing helps learners develop a more nuanced understanding of language use and promotes critical thinking by requiring them to evaluate both their writing and MT-generated outputs (Clifford et al. 2013; Niño 2009).

The effectiveness of post-editing, however, is influenced by several learner variables, particularly L2 proficiency levels, which affect the strategies used and the quality of the revised output (Cheng et al. 2015; Lee 2009; Qi and Lapkin 2001; Yamada 2023). Higher proficiency learners tend to make more extensive and accurate corrections during post-editing, while lower proficiency learners often face challenges in evaluating and correcting errors (Chung 2020; Lee 2022). For example, Shin and Chon (2023) found that advanced learners employ a wider range of post-editing strategies and focus more on broader content issues, whereas less proficient learners concentrate on word-level corrections. Thus, they argued that post-editing activities should be tailored to the learners' proficiency levels, providing structured guidance for lower proficiency learners and more challenging tasks for advanced learners.

Additionally, familiarity with the genre or context of the texts being translated affects post-editing success. While a lack of knowledge in specific areas can make the process more difficult, exposure to genre-aware post-editing tasks enhances students' linguistic sensitivity and structural awareness, enabling them to tailor revisions based on text types, whether academic or narrative (Harto et al. 2022). A Ukrainian study highlighted the common pitfalls in translating journalistic texts, where machines often misinterpret metaphors and expressions, reinforcing the need for human expertise in deciphering figurative and culturally nuanced language (Gusieva 2022). In Arabic post-editing training, students struggled primarily with genre-specific accuracy and terminology, emphasizing the importance of linguistic and genre familiarity in achieving acceptable outputs (Mohammed and Al-Rubai'i 2023). Collectively, these studies confirm that post-editing is not merely a linguistic activity, but one requiring a deep genre knowledge to bridge the gap between raw MT output and contextually appropriate human-like translations.

To effectively integrate MT in L2 writing, it is crucial to provide adequate pedagogical support and training. Studies emphasize the importance of teacher interventions and training sessions to enhance learners' ability to use MT effectively and to improve their post-editing skills (Samman 2022; Yoon and Chon 2022). Training can help students become more proficient in post-editing, thereby improving the quality of their writing over time. Without support, learners, especially those with lower proficiency, may struggle to benefit fully from MT, potentially widening the gap between different proficiency levels (Stapleton and Kin 2019; Tsai 2019). Udina (2019) also reinforces teaching through her analysis of LSP (Language for Specific Purposes), suggesting that pedagogical focus on genre and post-editing can foster students' ability to navigate such complexities (Udina 2019).

As discussed previously, many language teachers still remain hesitant to use MT due to historical inaccuracies and negative perceptions of translation (Lee 2022). In Japan, for example, many teachers are concerned about students simply copying and pasting MT outputs into their assignments (Oda 2019). Therefore, as discussed in past studies, effective use of MT and teaching appropriate post-editing techniques is crucial for facilitating L2 learning. To contribute to this goal, the present study focuses on two learner variables that have been considered as key determinants of post-editing behavior: genre knowledge and L2 (English) proficiency, including their potential interaction. However, empirical research examining these factors among Japanese college students remains limited. To address this gap, this study investigates the post-editing behaviors of undergraduate and graduate students at Ritsumeikan University's College of Life Sciences. These two groups are expected to differ in their overall English proficiency and, notably, in their genre knowledge, as graduate students are generally more familiar with the conventions of scientific writing than undergraduates. In addition to these group-level differences, this study also examines how varying

levels of L2 proficiency within each group affect post-editing behavior. As students advance in their academic careers, they are expected to strengthen these competencies to participate more actively in the global research community. Effective use of MT can play a critical role in supporting this development. Thus, the following research questions are formulated: (1) Do the post-editing behaviors of undergraduate students and graduate students differ according to their familiarity with a text genre? (2) In what way do different levels of L2 (English) proficiency affect the post-editing processes of students within the two groups?

2. Methodology

(1) Participants

The experiments in this study were conducted in two classes at the university where the author is employed. This university has three campuses, with the experiments being conducted at the campus where the science departments are located. The first class was a general English class with 13 undergraduate students (male = 6, female = 7) enrolled in 2023. In this class, students were introduced to the fundamental aspects of scientific writing, such as the structure of academic papers and the specific expressions used in such documents. As a class assignment, students selected a scientific paper of their interest, read it, and summarized it in English.

The second class was a science English course offered at the Graduate School of Life Sciences. This elective course, held during the spring semester of 2024, was attended by 8 graduate students (male = 4, female = 4). As in the first class, students were taught the basic structure of scientific papers, as well as grammar and expressions commonly used in such texts. However, in contrast to the undergraduate course, students in this class summarized their own research studies. The significant difference lies in the fact that most of the graduate students had already read some papers related to their studies in the laboratory. Thus, they must have had a basic knowledge of science papers well before coming to this class. To assess their proficiency levels, all students took an English assessment test developed by Speechace (<https://speak.speechace.co/placement/>). The participants' CEFR (Common European Framework of Reference for Languages Assessment) levels were as follows: Undergraduate (Level = number of students) A2 = 7, B1 = 6; Graduate A2 = 4, B1 = 4).

(2) Pre-Questionnaire Survey and Instruction

The author conducted a questionnaire survey regarding students' use of MT prior to the study. It was initially conducted as a part of a large-scale survey targeting Ritsumeikan University students. There were 5 major sections and 20 subsequent questions, both multiple-choice questions and open-ended questions. In this section, however, this author describes only the salient results related to the current study. As to the question: "Do you use MT when writing in English?", the results revealed that all undergraduate students and all but two graduate students were already using MT when writing in English. The reasons cited were consistent across both groups: the majority highlighted "time efficiency compared to writing in English manually" and "reduction in effort required for translation." Additionally, some students noted that they used MT because they "lacked confidence in their English proficiency" (5 undergraduate students and 3 graduate students).

Regarding post-editing practices (Question: "What do you do with MT translations?"), approximately half of the students in each group indicated that they "compared the MT output with the original Japanese text" (5 undergraduate students and 6 graduate students). Notably, however, some

students stated that they “did not review the MT output at all” (3 undergraduate students and 1 graduate student). These responses suggest that many students had already been using MT prior to taking this course and were at least somewhat familiar with post-editing methods.

When asked whether they trusted the accuracy of MT-generated English translations (Question: “Do you trust MT translations?”), most students responded that they “trusted” or “somewhat trusted” MT. However, a few students expressed skepticism, with 2 undergraduate students and 3 graduate students indicating that they “did not trust” MT. Their reasons included: “MT does not always produce accurate translations,” “occasionally strange translations occur,” and “while MT is reliable for everyday English, it may not accurately translate specialized content” (one graduate student). Based on these results, the author ensured that students understood the appropriate use of MT when writing in English. Her instruction emphasized that students should not simply copy and paste MT outputs. Instead, they were encouraged to carefully evaluate the MT-generated texts and incorporate only the portions they deemed suitable. Furthermore, students were advised to consult a dictionary whenever they encountered unfamiliar words in the MT output.

(3) Research Procedure

In both classes, the tasks involved several stages of writing. First, students wrote summaries of either their chosen papers or their own research in Japanese (L1). Next, they translated these summaries into English (L2) without assistance. Then, they used DeepL (<https://www.deepl.com>)² to produce machine-translated (MT) versions of their summaries. Students compared their self-translated texts (L2) with the machine translations and subsequently revised their L2 versions. This process was introduced in class, with students completing several revisions as homework until a final version was produced. Although the number of words for each summary was not strictly specified, students were required to write at least 150 words. Consequently, each student produced four texts: (1) the original L1 summary, (2) their own L2 translation, (3) the MT translation, and (4) a revised version of their L2 translation. Students were permitted to use dictionaries and other resources throughout the writing and revision process.

The use of MT in this task was justified for several reasons. As mentioned above, the pre-task questionnaire revealed that almost all students had already used MT for academic purposes prior to the class. However, their typical approach involved writing the entire text in Japanese and then translating it to English using MT, followed by post-editing. This method, though widely practiced, facilitates cognitive processes by encouraging students to compare L1 and L2 texts. However, this practice has limitations, as concerns persist that students may overly rely on MT translations, incorporating them with minimal error correction. Therefore, in this study, students were tasked with comparing their original L2 texts with MT translations to identify and correct potential errors. Additionally, this approach enabled the author to compare the L1 summaries, students’ L2 translations, MT translations, and final revised versions, with particular emphasis on tracking the revision process and identifying the sources of grammatical errors, as described by Lee (2022).

2 In this study, the classic (free) version of DeepL was used. The classic version is based on a conventional neural network architecture. As of April 2025, however, DeepL has introduced its next-generation language model—a large language model (LLM) specifically designed for language-related AI and translation tasks.

(4) Data Analysis

The students' revised L2 texts were analyzed by two researchers: the author and a Japanese teacher with over 20 years of experience in teaching English. To address research questions 1 and 2, the researchers counted the number of words from the MT translation incorporated into the students' revised texts. The post-editing units were categorized into three levels: word, phrase, clause, and sentence. Sentence-level revision involved changes to the entire sentence structure, while clause-level revisions were limited to modifying part of a sentence (e.g., the subject or predicate), following Chung's (2020) guidelines.³ Word-level revisions were counted when changes were made to individual words, phrase-level revisions were for phrases. Since students only identified errors that they perceived as such, the number of errors in their original L2 texts, as well as those left uncorrected in their final revisions, were also recorded.

The research questions aimed to investigate the features of the post-editing behaviors of the students in terms of their knowledge of a text genre and their L2 proficiency. Based on the previous literature, it was hypothesized that graduate students who had a better knowledge of academic papers in a particular subject area would detect and correct errors, such as adding the specific terminologies in their practices (Harto et al. 2022 and others). Similarly, it was expected that the students with a higher L2 proficiency would correct grammatical errors and perform more extensive revisions than those with lower proficiency (Cheng et al. 2015 and others). Therefore, the data were divided into two groups of affiliation and two proficiency levels according to the students' CEFR levels in both classes, and statistical comparisons were made. The revision processes of the students were also analyzed qualitatively, with particular attention paid to the revision strategies employed by both undergraduate and graduate students at different proficiency levels.

3. Results

A comparison between graduate and undergraduate students revealed both similar and different features. Firstly, there was a difference in the total number of words in both their original and revised texts (see Tables 1 and 2). In both cases, graduate students tended to produce a higher average word count than undergraduate students (original: $t(9) = 2.2891$, $p < .05$; revised: $t(9) = 2.9247$, $p < .05$). The mean word count in original texts was 201.25 (S.D = 91.04) for graduate students, compared to 116.25 (S.D = 48.64) for undergraduates. In revised texts, the mean was 228.25 (S.D = 90.59) for graduate students and 119.67 (S.D = 47.6) for undergraduates. Graduate students appeared to write more about their own research; however, the standard deviation suggests variability in word count across different L2 proficiency levels within each group.

With error correction, the original texts of undergraduate students had a higher average error rate (10.08%) compared to those of graduate students (4.91%). However, both groups significantly reduced their error rates in the revised texts (undergraduates: $t(11) = 6.166$, $p < .01$; graduates: $t(7) = 4.884$, $p < .01$), dropping to 3.41% for undergraduates and 2.41% for graduates. Considering the MT acceptance rate in their revised texts, where both groups incorporated approximately 30% of the MT output into

3 Sentence-level revisions refer to rewriting an entire sentence, as seen in Example 3 in the main text. In contrast, clause-level revisions involve modifying only part of a sentence. For instance, changing "The results indicate *that* how temperature affects enzyme activity" to "The results indicate *how temperature affects enzyme activity*." would be considered a clause-level revision.

their revised texts, this suggests that referring to MT during the post-editing process and integrating about 30% (undergraduates: 33.28%, graduates: 26.13%) of it helped both undergraduate and graduate students achieve a substantial reduction in errors.

Regarding the nature of revisions, both undergraduates and graduates had a higher frequency of sentence-level revisions, but undergraduates primarily focused on word-level revisions compared to graduates (undergraduates: 8.57%; graduates: 3.07%), a difference that was statistically significant ($t(9) = 2.924$, $p = .05$). This trend of word-level focus among undergraduates aligns with previous studies (Chung 2020 and others), suggesting a tendency to revise words individually while preserving sentence structure among the undergraduate students. An analysis of phrase-, clause-, and sentence-level revisions revealed no significant differences between the groups. Although undergraduates made slightly more clause-level revisions than graduates on average, this difference was not statistically significant.

When examining differences between proficiency levels within each group, the variations in text characteristics became more pronounced. The total word counts in both original and revised texts were significantly higher for high-proficiency students than for low-proficiency students in both the undergraduate group (original: high-level 136.40 words, low-level 101.86 words; revised: high-level 141.20 words, low-level 104.29 words) and graduate group (original: high-level 215.0 words, low-level 187.50 words; revised: high-level 267.75 words, low-level 188.75 words). Both high- and low-proficiency groups increased their word count in revised texts by using MT, but especially the high-proficiency groups.

A notable distinction in error rates emerged between the two proficiency levels within each group. Among undergraduates, the low-proficiency group made significantly more errors in their original texts than the high-proficiency group (high-level: 8.64%; low-level: 10.34%), suggesting that students with lower English proficiency produced more errors initially. However, post-editing substantially reduced error rates, with high-level students reducing their errors to 3.93% and low-level students to 3.04%, resulting in almost equivalent revision outcomes. Graduate students, however, displayed a different pattern. High-proficiency graduates had a higher error rate in their original texts (6.28%) compared to low-proficiency graduates (4.96%). However, after post-editing, error rates in both groups reduced to approximately 3% (high-level: 2.52%; low-level: 2.97%).

A closer examination of revision types revealed notable patterns. Among undergraduates, low-proficiency students made more revisions at the word and sentence levels compared to high-proficiency students (Table 1). Specifically, high-proficiency students revised 4.01% of their text at the word level, whereas low-proficiency students revised 11.83%. At the sentence level, high-proficiency students revised 6.09% of their text, while low-proficiency students revised 26.64%, indicating a significantly higher rate of sentence-level revisions for the latter. Regarding MT acceptance, low-proficiency students incorporated 47.10% of MT content, while high-proficiency students incorporated only 22.60%, suggesting that low-proficiency students relied more heavily on MT, particularly at the sentence level, to reduce their error rate. Notably, MT acceptance among lower-proficiency students was higher than average, reaching rates of 78.22%, 61.96% or 59.41% for some students. Given that most of their revisions occurred at the sentence level, there is a high likelihood that they simply adopted the MT translations without significant changes to their original texts. As previously mentioned, high-proficiency undergraduates reduced their error rate from 8.64% in the original to 3.93% in the revision, while low-proficiency undergraduates reduced theirs from 10.34% to 3.04%. This may suggest that high-proficiency students attempted to correct errors independently, resulting in

Table 1. Frequency of revisions by categories (Undergraduate students)

Students	total words (original)	total words (revision)	Frequency of word-level revisions	word-level revisions %	Frequency of phrase-level revisions	phrase-level revisions (number of words)	phrase-level revisions (number of words) %	Frequency of clause-level revisions	clause-level revisions (number of words)	clause-level revisions (number of words) %	Frequency of sentence-level revisions	sentence-level revisions (number of words)	sentence-level revisions (number of words) %	Number of T-unit	MT acceptance total (number of words)	MT acceptance rate	errors in original texts (in frequency)	error rate in original texts	errors in revision (in frequency)	error rate in revision
High: A	156	165	8	4.85	0	0	0	0.00	1	9	5.45	1	9	5.45	26	15.76	9	5.77	5	3.03
B	86	88	5	5.68	1	3	3.41	4	4	27	30.68	0	0	0.00	35	39.77	10	11.63	3	3.41
C	190	189	1	0.53	0	0	0.00	0	0	0	0.00	0	0	0.00	1	0.53	10	5.26	10	5.29
D	73	80	5	6.25	1	3	3.75	2	8	10.00	1	20	25.00	3	36	45.00	8	10.96	2	2.50
E	177	184	5	2.72	0	0	0.00	1	17	9.24	0	0	0	0.00	22	11.96	17	9.60	10	5.43
Low F	136.4	141.2	4.8	4.01	0.4	1.2	1.43	1.6	12.2	11.08	0.4	5.8	6.09	1.2	24	22.60	10.8	8.64	6	3.93
G	177	163	10	6.13	1	3	1.84	3	27	16.56	1	14	8.59	2	54	33.13	12	6.78	3	1.84
H	75	70	15	21.43	1	3	4.29	1	12	17.14	1	8	11.43	1	38	54.29	11	14.67	2	2.86
I	99	101	5	4.95	0	0	0.00	0	0	0.00	3	74	73.27	11	79	78.22	7	7.07	3	2.97
J	72	92	6	6.52	1	4	4.35	0	0	0.00	4	47	51.09	4	57	61.96	9	12.50	1	1.09
K	63	61	4	6.56	0	0	0.00	0	0	0.00	2	2	3.28	0	6	9.84	8	12.70	3	4.92
L	66	73	22	30.14	1	2	2.74	0	0	0.00	0	0	0.00	0	24	32.88	7	10.61	3	4.11
	161	170	12	7.06	1	4	2.35	2	19	11.18	4	66	38.82	7	101	59.41	13	8.07	6	3.53
Total average	101.86	104.29	10.57	11.83	0.71	2.29	2.19	0.86	8.29	6.41	2.14	30.14	26.64	3.57	51.29	47.10	9.57	10.34	3.00	3.04
	116.25	119.67	8.17	8.57	0.58	1.83	1.89	1.17	9.92	8.35	1.42	20.00	18.08	2.58	34.25	33.28	10.08	9.63	4.25	3.41

Table 2. Frequency of revisions by categories (Graduate students)

Students	total words (original)	total words (revision)	Frequency of word-level revisions	word-level revisions %	Frequency of phrase-level revisions	phrase-level revisions (number of words)	phrase-level revisions (number of words) %	Frequency of clause-level revisions	clause-level revisions (number of words)	clause-level revisions (number of words) %	Frequency of sentence-level revisions	sentence-level revisions (number of words)	sentence-level revisions (number of words) %	Number of T-unit	MT acceptance total (number of words)	MT acceptance rate	errors in original texts (in frequency)	error rate in original texts	errors in revision (in frequency)	error rate in revision
High M	190	297	5	1.68	1	3	1.01	0	0	0.00	4	110	37.04	11	118	39.73	15	7.89	12	4.04
N	271	297	2	0.67	0	0	0.00	0	0	0.00	4	85	28.62	11	87	29.29	12	4.43	6	2.02
O	189	249	1	0.40	1	3	1.20	3	25	10.04	6	69	27.71	11	98	39.36	17	8.99	10	4.02
P	210	228	3	1.32	0	0	0.00	2	11	4.82	1	4	1.75	2	18	7.89	8	3.81	0	0.00
Low Q	215	267.75	2.75	1.02	0.5	1.5	0.55	1.25	9	3.72	3.75	67	23.78	8.75	80.25	29.07	13	6.28	7	2.52
R	158	158	6	3.80	2	7	4.43	0	0	0.00	2	36	22.78	3	49	31.01	13	8.23	8	5.06
S	133	141	1	0.71	1	1	0.71	1	9	6.38	0	0	0.00	0	11	7.80	9	6.77	4	2.84
T	69	81	11	13.58	1	3	3.70	1	4	4.94	0	0	0.00	0	18	22.22	3	4.35	3	3.70
	390	375	27	7.20	3	9	2.40	3	20	5.33	3	63	16.80	8	119	31.73	2	0.51	1	0.27
Total average	187.5	188.75	11.25	6.32	1.75	5	2.81	1.25	8.25	4.16	1.25	24.75	9.90	2.75	49.25	23.19	6.75	4.96	4	2.97
	201.25	228.25	7	3.07	1.125	3.25	1.68	1.25	8.625	3.78	2.5	45.875	16.84	5.75	64.75	26.13	9.875	4.91	5.5	2.41

some remaining errors in their revised texts, whereas low-proficiency students incorporated larger portions of the MT text, effectively reducing errors.

For graduate students, as shown in Table 2, similar to undergraduates, low-proficiency students made more word-level revisions (low: 6.32%, high: 1.02%), whereas high-proficiency students made more sentence-level revisions (low: 9.90%, high: 23.78%). Additionally, high-proficiency students showed a higher MT acceptance rate (29.07%) than low-proficiency students (23.19%). This suggests that high-proficiency students tended to accept MT content at the sentence level, whereas low-proficiency students focused on word-level MT integration during the post-editing process.

To examine the impact of L2 proficiency on post-editing comprehensively, a chi-square test was conducted by dividing both undergraduate and graduate students into high- and low-proficiency groups. The results showed that low-proficiency students, regardless of their academic affiliation, made more word-level and sentence-level revisions, with a statistically significant difference observed in word-level revisions ($t(11) = 2.6182, p < .05$). This suggests that students with lower L2 proficiency made word-level revisions more frequently, irrespective of whether they were undergraduates or graduates. However, the statistical analysis did not identify which group contributed more strongly to this effect. Given that the average proportion of word-level revisions among low-proficiency undergraduates was 11.83%, compared to 6.32% among low-proficiency graduates, it is likely that the word-level revisions made by low-proficiency undergraduates contributed to the statistical significance.

4. Discussion

The present study aimed to investigate how differences in genre knowledge and L2 proficiency among undergraduate and graduate students affect their post-editing behaviors. To examine the effects of these factors, the author compared errors in their original texts and revisions, analyzed error corrections at different linguistic levels (word, phrase, clause, sentence), and assessed the extent to which they incorporated machine-translated outputs in their revisions. The findings indicate that the number of errors can be significantly reduced through the use of MT, regardless of students' affiliation or proficiency. These findings are aligned with Lee (2020) and Niño (2008), who showed that MT can improve overall writing quality and that students benefited from its use.

As to Research Question 1: (1) Do the post-editing behaviors of undergraduate students and graduate students differ according to their familiarity with a text genre? The most notable difference between these two groups was the nature of their revisions. Undergraduate students made word-level changes more frequently than graduate students, tending to revise smaller portions of sentences while leaving the sentence structures largely unchanged. Most errors include basic grammatical ones, such as singular/plural forms, verb forms, or third-person singular "s," or technical terms used in research papers. Of particular note here is that technical terms were mostly replaced by the vocabulary produced by MT, which indicates that undergraduate students were not able to discern the accuracy of words in their original texts in comparison with the words in the MT. Here is one example from the undergraduate students. As this example shows, most word-level revisions here were technical words. She changed *eczema* to *dermatitis*, *therapeutic products* to *remedies*, and *effective product* to *active medication*:

Example 1:**A part of the original text:**

Many kinds of conditions of the skin such as eczema and psoriasis are treated by tropical therapeutic products. Instead of applying effective product directly on the skin, they are used with vehicle and sink in the deep area of the skin called epidermis and dermis.

Machine-translated text:

Many skin conditions, such as dermatitis and psoriasis, are treated with tropical remedies. Instead of applying the active medication directly to the skin, it is combined with excipients to reach beyond the stratum corneum and into deeper areas of the skin, namely the epidermis and dermis.

Revised text after referring to the MT text (w=word-level, p=phrase-level revision):

(p) Many skin conditions, such as (w) dermatitis and psoriasis, are treated (w) with tropical (w) remedies. Instead of applying the (w) active (w) medication directly to the skin, (p) it is combined with vehicle and sink into deeper areas of the skin, (w) namely the epidermis and dermis.

While both undergraduate and graduate students made sentence-level revisions more frequently than phrase- or clause-level changes, graduate students were more likely to reconstruct entire sentences by comparing their original texts with MT outputs. This tendency is evident in the increased number of minimal terminable units (T-units⁴) in the revised texts of graduate students (5.75) compared to undergraduates (2.58), indicating that the revised texts of graduate students contained more complex structures. This trend—characterized by a deliberate examination of both their own texts and MT outputs to produce the best possible revisions—persisted throughout the semester. There was also a tendency to attempt to select more accurate technical terms, as discussed later in Example 3.

The differences observed between undergraduates and graduates may be attributed to the genre of the texts that students were tasked with summarizing. Graduate students summarized their own research, enabling them to critically evaluate the appropriateness of MT outputs due to their familiarity with the content and academic conventions of their field. In contrast, undergraduates summarized papers that were not necessarily related to their own studies. Lacking the expertise to assess vocabulary, especially technical terms and phrases suitable for academic papers in a particular subject area, undergraduates were more likely to incorporate MT-produced words without sufficient scrutiny. Indeed, while the difference was not statistically significant, the MT acceptance rate among undergraduates (33.28%) was higher than that of graduate students (26.13%). This finding aligns with Mohammed and Al-Rubai'i (2023) and Harto et al. (2022), which suggests that a lack of domain-specific knowledge can make the post-editing process more challenging.

As for Research Question 2: (2) In what way do different levels of L2 (English) proficiency affect the post-editing processes of students within the two groups? Within each group, proficiency levels also influenced post-editing behaviors. Among undergraduates, lower-proficiency students made more word- and sentence-level revisions than higher-proficiency students. For instance, two lower-level students made word-level revisions in 21.43% and 30.14% of their total revisions, the highest proportion among all revision types. These students typically revised individual words while leaving

4 “T-unit” in linguistics refers to a main clause plus any subordinate clauses that may be attached to it.

sentence structures unchanged. It was found in the previous studies (Shin and Chon 2023 and others) that learners with lower proficiency are less capable of making extensive revisions and instead focus on word-level changes. At the sentence level, particularly those with lower proficiency levels tended to rely heavily on MT-generated sentences. Three students revised sentences in 73.27%, 51.09%, and 38.82% of their total revisions, with MT acceptance rates of 78.22%, 61.96%, and 59.41%, respectively. This indicates that they incorporated the majority of the MT outputs into their texts with only minimal changes to vocabulary. Although this author's instruction to the students was to revise their original texts, referring to the MT translations, one of the students with lower proficiency discarded his original text and used all of the MT outputs with some words changed for revision. Another notable example shown below involved a student who revised only a few words in the first and last sentences of her text while replacing most of the middle section with unaltered MT output.

Example 2:

A part of the original text:

We introduce Jukebox, a model that generates music with singing in the raw audio domain. We tackle the long context of raw audio using a multi-scale VQ-VAE to compress it to discrete codes, and modeling those using autoregressive Trans-formers. We show that the combined model at scale can generate high-fidelity and diverse songs with...(cut)..We are releasing large amounts of non-selected samples with model and code.

Machine-translated text:

We present Jukebox, a model for generating sung music in the raw audio domain. We address this by compressing long contexts of raw audio into discrete codes using multi-scale VQ-VAE and modeling them using an autoregressive transformer. As a result, we show that it is possible to generate high-fidelity and diverse musical compositions with coherence that spans several minutes...(cut)... Thousands of samples are available, along with sample weightings and chords.

Revised text after referring MT text (w=word-level, s=sentence-level revision):

We introduce Jukebox, a model that generates music with singing in the raw audio domain. (s) We address this by compressing long contexts of raw audio into discrete codes using multi-scale VQ-VAE and modeling them using an autoregressive transformer. As a result, we show that it is possible to generate high-fidelity and diverse musical compositions with coherence that spans several minutes...(cut)... We are releasing large amounts of non-selected samples with model and code.

This student answered that she was not confident in writing in English in the pre-survey, suggesting that the lower-proficiency students struggled to evaluate and correct errors by comparing their original and MT texts, likely due to their limited knowledge of English (Chung 2020; Lee 2022; Yamada 2023); thus, they tended to overly incorporate MT. As a result, they missed opportunities to self-edit even basic errors, such as singular/plural forms, verb forms, or third-person singular “s,” all of which should have been easily correctable even for lower-proficiency learners.

On the other hand, higher-proficiency students were willing to and attempted to choose the better words appropriate to their L1 (Japanese) texts and reconstruct complete sentences in revision. However, here is one notable finding about the errors among the students with higher proficiency. Two

students stuck to their original texts with minimal changes in revision. Their MT acceptance rate was significantly low at 0.53% and 11.96% and the error rate remained higher than the average at 5.29% and 5.43% even after referring to MT outputs. Considering that they were senior students who had started to study in a laboratory, they tried to write and revise their texts by themselves, not relying on MT. In fact, they responded to the survey that they could manage to write in English by themselves and continuously showed a rather negative stance to using MT. They might even have refused to accept the MT translations. Thus, even after they tried to revise their texts, when they were compared to MT translations, most of the errors, particularly the technical errors, remained. Even with a high proficiency of English, they were not yet sufficiently knowledgeable about the technical terms to discern such mistakes. These findings highlight that effective post-editing not only demands L2 proficiency but also a solid understanding of genre conventions, which significantly impacts the quality and appropriateness of the final text. High proficiency in the target language alone is insufficient without subject-specific knowledge to appropriately revise MT outputs. In this case, students were provided with a room to revise their final texts with the help of their supervisors who could confer with them about the technical terms.

As for the graduate students, high-proficiency students tried to produce better revisions with both word-level and sentence-level revisions. They compared their original version with the MT outputs, and if subtle word-level revisions would not be effective enough to convey their intended meaning, they chose to replace the whole sentence with the MT translation. In fact, their effort to scrutinize their texts with MT repeatedly continued until they submitted their final revisions. This seemed to be the reason why the MT acceptance rate is higher among students with a high proficiency than those with a lower proficiency. Example 3 shows that this student replaced the words or sentences with the MT outputs, trying to retain his original text. Although he changed two sentences into MT ones, he tried to stick to his original text as much as possible. As to the technical terms, he changed from *flocking behavior* (of organisms such as insects) to *swarming behavior* underlined (p) generated by MT. He chose the latter one because he thought it was better in his text since he was able to distinguish the difference between the two terms.

Example 3:

Original text:

In this study, we simulated Boid Model using genetic algorithm for clarifying the evolutionary mechanism by analyzing flocking behavior dynamics in nature. The swarming behavior of organisms are believed to be evolved by several strategies related to their survival such as capturing preys, avoiding predators, breeding behavior. This research purpose is to deeply understand flocking behaviors through mathematical model. Boid model is mathematical model based on three rules (separation, cohesion, alignment) for simulating swarming behaviors. In this research, we added a force heading to the preys, that can simulate the interactions between preys and predators. In the simulation, we analyzed 4 types of flocking behaviors (Swarm, Torus, Dynamic Parallel, Highly Parallel) to clarify how it works for prey acquisition.

Machine-translated text:

The purpose of this study is to analyze the dynamics of swarming behavior of organisms in nature and to clarify its evolutionary mechanism. It is believed that swarming behavior of organisms has evolved as a variety of strategies directly related to survival, such as food capture, escape from

predators, and reproductive behavior. In this study, we aimed to deepen our biological understanding of the dynamics of swarming behavior by understanding it through mathematical models. The Boids model is a model that simulates swarming behavior based on the basic behavioral rules of association, separation, and alignment. In this study, we added suction to the model to mimic the dynamic interaction between predators and prey and capture the additional complexity of swarming behavior. In the simulations, we considered four major swarming morphologies, swarm, torus, dynamic parallel, and highly parallel, and compared and analyzed how each shape affects efficiency in prey acquisition.

Revised text after referring to the MT text (w=word-level, P=phrase-level, s=sentence-level revision):

In this study, we simulated (p) *the dynamics of swarming behavior* by using Boid Model and used genetic algorithm for clarifying the evolutionary mechanism in nature. (s) *It is believed that swarming behavior of organisms has evolved as a variety of strategies directly related to survival, such as food capture, escape from predators, and reproductive behavior.* This research purpose is to deeply understand swarming behaviors through mathematical model. Boid model is (w) *a* mathematical model based on three (w) *behavioral* rules (w) *of* separation, cohesion, alignment for simulating swarming behaviors. In this research, we added a force heading to the preys, that can simulate the interactions between preys and predators. In the (w) *simulations*, (s) *we considered four major swarming morphologies. Swarm, Torus, Dynamic parallel, and highly parallel, and compared and analyzed how each shape affects efficiency* in prey acquisition.

As to those with a lower proficiency, another possible explanation for their minimal text revisions is that they used MT to generate their original texts from the beginning. In fact, both the original texts and the MT translations submitted by the two students were almost the same, and their revisions were extremely scarce, showing that writing the summary of a scientific study must be a tedious task for graduate students with lower English proficiency, even if they had a basic knowledge of the content. This finding is reflected by the fact that high-proficiency graduates had a higher error rate in their original texts compared to low-proficiency graduates. In order to reduce errors in revised texts, MT proved to be a useful tool for students at both proficiency levels. However, similar to the findings for undergraduates, those who made a deliberate effort to revise their texts by comparing them with MT outputs had a better chance of learning vocabulary and expressions appropriate to their research.

Overall, the findings of the current study suggest that the way MT output is used and analyzed is largely influenced by learners' proficiency, as indicated in previous studies (Chung 2020; Chung and Ahn 2021; Lee 2022). Students with higher L2 proficiency are more critical and better able to detect and correct errors. At the same time, the task type and genre of the text are significant factors influencing the quality of the final products. In this study, graduate students who were knowledgeable about their field of study and familiar with technical terms and genre conventions made better choices when utilizing MT outputs. Therefore, it can be concluded that both L2 proficiency and genre knowledge play critical roles in successful post-editing. While L2 proficiency facilitates error detection and linguistic refinement, genre knowledge enables learners to make contextually appropriate lexical and structural choices, particularly in the case of technical content.

Based on the results, this study can provide several pedagogical implications for using MT in

English writing assignments. First, although it can be beneficial to use MT in writing, merely using technology cannot bring benefits to all students (Cheng et al. 2015). Lee (2022) argues that the use of MT might further increase the gap between higher- and lower-level students. As the current study showed, some students with lower proficiency levels might not have understood MT outputs, which tend to produce far better texts than they can create. They just incorporated MT translation without trying to analyze the reasons for errors. On the other hand, those with a higher proficiency level scrutinized the errors and could intake the correct or appropriate words or phrases, learning from MT, which is the expected way of using MT for the sake of its educational benefits. The use of MT and tasks involving MT with learners of lower proficiency should be carefully designed and supported with clear guidelines. For such students, the focus should be on identifying and correcting basic, manageable errors, such as simple grammar mistakes or word choice issues. In contrast, students with higher proficiency, who are more capable of critically analyzing the MT outputs, can be assigned more advanced tasks, such as reconstructing sentences with significant mistranslations. In other words, MT post-editing activities should be thoughtfully structured to provide appropriate support in the areas where learners need the most help.

In this study, summarizing research was a particularly challenging task for both undergraduate and graduate students. For students who are not familiar with the conventions and subject matter of a particular academic field, teachers or supervisors can provide valuable support by reviewing final drafts and offering feedback on technical terms and content. Since MT may generate inaccurate terminology, expert guidance in the field is essential. Therefore, students aiming to become researchers should be encouraged to read as many academic papers as possible to become accustomed to the terminology, structural conventions, and rhetorical norms characteristic of scholarly texts in their discipline. Strengthening both English proficiency and knowledge of fundamental academic writing practices is likely to improve learners' ability to produce higher-quality research papers with the support of MT.

First and foremost, it is imperative for educators to instruct students on the capabilities and limitations of MT tools. Given that MT outputs are not flawless, students should be guided to critically assess their accuracy rather than relying on them uncritically, fostering a sense of ownership and responsibility over their writing. To this end, it is crucial to first have students compose texts independently and cultivate their "revision skills" through iterative improvements based on feedback. When encountering unfamiliar words or phrases generated by MT, students should be encouraged to verify their accuracy using reliable resources such as dictionaries. This approach not only enhances their writing skills but also promotes effective learning through the integration of MT tools. Furthermore, educators themselves should engage with MT and AI tools to gain a comprehensive understanding of their practical applications, advantages, and limitations, thereby enabling them to provide informed guidance on their use.

5. Conclusion

This study explored the integration of Machine Translation (MT) into English writing instruction, with a particular focus on post-editing behaviors among Japanese undergraduate and graduate students, analyzed from the perspective of L2 proficiency and genre familiarity. While the study is limited by its small sample size and the lack of qualitative data on learners' revision processes (e.g., through interviews), it nevertheless offers important insights into how students interact with MT based

on their academic affiliation and language proficiency. The findings suggest that MT can be an effective tool for reducing linguistic errors in student writing. However, its educational value depends heavily on learners' ability to critically evaluate and revise MT-generated outputs. Students with higher L2 proficiency and greater familiarity with academic discourse engaged in more sophisticated revisions, including sentence reconstruction and lexical refinement. In contrast, lower-proficiency students tended to rely more heavily on MT outputs with limited post-editing, highlighting the need for differentiated instructional support. Furthermore, graduate students, likely due to their disciplinary knowledge, demonstrated more effective use of MT in academic writing tasks than undergraduates. These results underscore the importance of both linguistic competence and genre literacy in achieving successful post-editing. To ensure equitable learning outcomes, it is essential that educators provide training that aligns with learners' proficiency levels and genre awareness. Explicit instruction in revision strategies, combined with structured teacher feedback and the development of metalinguistic awareness, can empower students to use MT more critically and productively in their academic writing.

Future research could investigate the long-term effects of MT-assisted learning on students' independent writing abilities. Specifically, it would be valuable to explore whether consistent use of MT fosters genuine improvement in language proficiency or creates a reliance on the tool, potentially hindering linguistic growth. Additionally, as AI technologies continue to evolve, studies could examine how advancements in MT, such as context-sensitive corrections and adaptive learning features, may help bridge the proficiency gap among learners. These innovations might enable MT to provide more personalized feedback, helping lower-proficiency students better understand and revise their work while offering advanced learners nuanced suggestions for improvement. In parallel, the increasing use of generative AI tools such as ChatGPT in language education should be critically examined. These tools offer real-time, context-aware feedback and can support learners in drafting, revising, and refining texts. However, there is a risk that overreliance on such tools may hinder the development of a deeper understanding of language structure and academic conventions. From the author's perspective, in order for learners to effectively assess whether the outputs generated by MT or AI tools align with their intended meaning, a higher L2 proficiency and strong genre knowledge are still essential. Future research should consider how generative AI can be used not only as a productivity tool but also as a means to enhance metalinguistic awareness and critical language use. In any case, as English education increasingly integrates MT and AI technologies, we as teachers must stay informed of these developments and continue to propose effective ways to utilize them beneficially in education.

References

- Alsalem, R. 2019. The Effects of the Use of Google Translate on Translation Students' Learning Outcomes. *AWEJ for Translation and Literary Studies*, 3(4), 46–60. <<http://doi.org/10.24093/awejtls/vol3no4.5>>
- Cheng, K. H., J. C. Liang and C. C. Tsai. 2015. Examining the Role of Feedback Messages in Undergraduate Students' Writing Performance During an Online Peer Assessment Activity. *The Internet and Higher Education*, 25, 78–84. <<https://doi.org/10.1016/j.iheduc.2015.02.001>>
- Chon, Y. V., D. Shin and G. E. Kim. 2021. Comparing L2 Learners' Writing Against Parallel Machine-Translated Texts: Raters' Assessment, Linguistic Complexity and Errors. *System*, 96, 1–12. <<https://doi.org/10.1016/j.system.2020.102408>>
- Chung, E. S. 2020. The Effect of L2 Proficiency on Post-editing Machine Translated Texts. *The Journal of Asia TEFL*, 17(1), 182–193. <<http://dx.doi.org/10.18823/asiatefl.2020.17.1.11.182>>
- Chung, E. S. and S. Ahn. 2021. The Effect of Using Machine Translation on Linguistic Features in L2 Writing Across

- Proficiency Levels and Text Genres. *Computer Assisted Language Learning*, 35(9), 2239–2264. <<https://doi.org/10.1080/09588221.2020.1871029>>
- Clifford, J., L. Merschel and J. Munné. 2013. Surveying the Landscape: What is the Role of Machine Translation in Language Learning. *Research in Education and Learning Innovation Archives*, 10, 108–121.
- Escartín, C. P., S. O'Brien, M. J. Goulet and M. Simard. 2017. Machine Translation as an Academic Writing Aid for Medical Practitioners. *Proceedings of MT Summit XVI: Research Track*, 254–267. <<https://core.ac.uk/download/pdf/200762284.pdf>>
- Fredholm, K. 2019. Effects of Google Translate on Lexical Diversity: Vocabulary Development. Among Learners of Spanish as a Foreign Language. *Revista Nebrija*, 13(26), 98–117. <<http://doi.org/10.26378/rnlael1326300>>
- Garcia, I. and M. Pena. 2011. Machine Translation-Assisted Language Learning: Writing for. Beginners. *Computer Assisted Language Learning*, 24(5), 471–487. <<https://doi.org/10.1080/09588221.2011.582687>>
- Gusieva, O. 2022. Automated Translation and Its “Post-machine” Editing. *Bulletin of Mariupol State University*, 15(26–27), 261–267. <<https://doi.org/10.34079/2226-3055-2022-15-26-27-261-267>>
- Harto, S., F. A. Hamied, B. Musthafa and S. Setyarini. 2022. Exploring Undergraduate Students’ Experiences in Dealing with Post-Editing of Machine Translation. *Indonesian Journal of Applied Linguistics*, 11(3), 696–707. <<https://doi.org/10.17509/ijal.v11i3.42825>>
- Jia, Y., M. Carl and X. Wang. 2019. Post-Editing Neural Machine Translation Versus Phrase-Based Machine Translation for English-Chinese. *Machine Translation*, 33, 9–29. <<https://link.springer.com/article/10.1007/s10590-019-09229-6>>
- Jolley, J. R. and L. Maimore. 2022. Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature. *L2 Journal*, 14(1), 26–44.
- Klimova, B., M. Pikhart, A. D. Benites, C. Lehr and C. Sanchez-Stockhammer. 2022. Neural. Machine Translation in Foreign Language Teaching and Learning: A Systematic Review. *Education and Information Technologies*, 28(1), 663–682. <<http://doi.org/10.1007/s10639-022-11194-2>>
- Lee, S. M. 2009. Exploring the Writing and Editing Patterns of the Different Language Proficiency Groups on a Wiki. *Foreign Languages Education*, 16(3), 95–122.
- . 2020. The Impact of Using Machine Translation on EFL Students’ Writing. *Computer Assisted Language Learning*, 33(3), 157–175. <<https://doi.org/10.1080/09588221.2018.1553186>>
- . 2022. Different Effects of Machine Translation on L2 Revisions Across Students’ L2 Writing Proficiency Levels. *Language Learning and Technology*, 26(1), 1–21. <<https://hdl.handle.net/10125/73490>>
- Mohammed, S. S. and Al-Rubai’i, A. M. H. 2023. Post-editing of Machine Translation of An English-into-Arabic Text by Translation Students. *The Journal of the University of Duhok*, 26(1), 1227–1235. <<https://doi.org/10.26682/hjuod.2023.26.1.75>>
- Niño, A. 2008. Evaluating the Use of Machine Translation Post-Editing in the Foreign Language Class. *Computer Assisted Language Learning*, 21(1), 29–49. <<https://doi.org/10.1080/09588220701865482>>
- . 2009. Machine Translation in Foreign Language Learning: Language Learners’ and Tutors’ Perceptions of its Advantages and Disadvantages. *ReCALL*, 21(2), 241–258. <<https://doi.org/10.1017/S0958344009000172>>
- Oda, T. 2019. Foreign Language Learning Activities Coexisting with Machine Translation. *Journal of Humanities and Natural Sciences*, 145, 3–27.
- Qi, D. S. and S. Lapkin. 2001. Exploring the Role of Noticing in a Three-Stage Second Language. Writing Task. *Journal of Second Language Writing*, 10(4), 277–303. <[https://doi.org/10.1016/S1060-3743\(01\)00046](https://doi.org/10.1016/S1060-3743(01)00046)>
- Samman, H. M. 2022. *Evaluating Machine Translation Post-Editing Training in Undergraduate Translation Programs: An Exploratory Study in Saudi Arabia*. [Doctoral dissertation, University of Southampton]. University of Southampton Institutional Repository. <<http://eprints.soton.ac.uk/id/eprint/469163>>
- Shin, D. and Y. V. Chon. 2023. Second Language Learners’ Post-Editing Strategies for Machine Translation Errors. *Language Learning and Technology*, 27(1), 1–25. <<https://hdl.handle.net/10125/73523>>
- Stapleton, P. and B. Kin. 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>>
- Tsai, S.-C. 2019. Using Google Translate in EFL Drafts: A Preliminary Investigation. *Computer Assisted Language*

- Learning*, 32(5–6), 510–526. <<https://doi.org/10.1080/09588221.2018.1527361>>
- Udina, N. 2019. *Using Post-editing in Translation and LSP Courses*. <http://www.ocerints.org/intcess19_e-publication/abstracts/a265.html>
- Yamada, M. 2021. Post-Editing and a Sustainable Future for Translators. *Journal of Foreign Language Studies*, 24, 83–105. <<https://doi.org/10.32286/00023085>>
- . 2023. *Using Machine Translation for English Communication or Learning English Through MT Utilization*. Presented at the 14th Industrial Japanese Studies Symposium: Language Communication in the Age of Artificial Intelligence, Rikkyo University, (February 9, 2023).
- Yoon, C. W. and Y. V. Chon. 2022. Machine Translation Errors and L2 Learners' Correction Strategies by Error Type and English Proficiency. *English Teaching*, 77(3), 153–175. <<http://doi.org/10.15858/engtea.77.3.202209.153>>