The effect of L2 proficiency on post-editing machine translated texts

Pei-Jung, Wu（吳佩蓉）[1]

ABSTRACT

Research on machine translation (MT) has primarily focused on translation errors, mistranslations, learners' perceptions and attitudes towards MT, and the educational applications of MT in language learning. To address a gap in the existing literature, this study examines the impact of L2 proficiency on the post-editing skills of EFL learners when dealing with machine-translated texts. A total of 34 eleventh-grade EFL learners, whose L1 is Chinese and L2 is English, participated in the study. Their L2 proficiency levels were determined through the administration of the General English Proficiency Test (GEPT). Based on the test results, the participants were categorized into three proficiency groups: high, intermediate, and low. Subsequently, they were instructed to post-edit a machine-translated text, and their edited texts were recorded and analyzed at four linguistic levels: words, phrases, clauses, and sentences. The findings from a one-way ANOVA revealed that L2 proficiency significantly influenced the corrections made to the machine-translated texts. However, no noteworthy differences were observed among the proficiency groups concerning the percentage of errors corrected at the four linguistic levels.

Key word: MT, Machine translation post-editing, L2 proficiency, EFL learners

---

[1] 彰化女中英文教師

Introduction

Machine translation (MT) such as google translate has been readily available and widely used and as a tool for foreign language learning in the recent years. While its use as an aid for L2 writing and translation is common, research into MT remains limited to this day. Previous studies on MT examined translation errors and mistranslations of MT (Dhakar, Shinha & Pandey, 2013; Ghasemi & Hashemian, 2016), analyzed how learners revised their drafts with MT texts (Lee, 2019; Groves & Mundt, 2015), investigated learners perceptions and attitudes on the use of MT(Im, 2017; Larson-Guenette, 2013; White & Heidrich, 2013) and the pedagogical implications and ideas of MT (Correa, 2014; Enkin & Mejias-Bikandi, 2016; Garcia & Pena, 2011). While much discussion has been made on the educational uses of MT in writing, it is not clear how L2 learners actually integrate MT in their writing process and how their proficiency in L2 may affect the use of the online tool. The present study aims to fill in this gap in literature by exploring the relationship between EFL learners' English proficiency and their ability to post-edit MT text and the difficulties low-proficient learners may encounter in post-editing the MT texts.

Literature review

Limitations and benefits of MT

Challenges remain for MT to produce error-free translations. Previous research has reported MT errors in different aspects of language, including grammar (Enkin & Mejías-Bikandi, 2016; Groves &Mundt, 2015; Niño, 2009; Williams, 2006), register and cultural references (Correa, 2014; Niño, 2009; Somers et al., 2006), pragmatic usage(Ducar &Schocket, 2018)and proverbs and idioms (Somers et al., 2006; Correa, 2014; Luton, 2003; Kim, 2018). Though there has been significant improvement in MT texts in the recent decade, grammatical errors, register, cultural expectations as well as pragmatic breakdown continue to be the problems surrounding MT texts (Ducar &Schocket, 2018). Despite the limitations, the educational value of MT in foreign language learning cannot be overlooked. Previous studies have reported that the use of MT is beneficial in enhancing fluency(Garcia &Pena, 2011), accuracy (Tsai, 2019), paraphrasing skills (Niño,

2009) and word choice (Chen et al., 2015). Tsai (2019), for example, compared and analyzed college students' (N=124; L1=Chinese; L2=English) self-written (SW version) drafts and the versions made based on the Google translated texts (GT version) and found that the GT version had fewer grammatical errors, more accurate use of vocabulary words and expressions and greater lexical density compared to the SW version.

MT post-editing and L2 proficiency

MT post-editing, which is referred to as the process of correcting and editing machine-translated texts, is a practice often implemented by practitioners and language learners. Previous studies have recognized the value of MT post-editing for intermediate and advanced language learners (S. M.Lee, 2020; Niño, 2008). Niño (2008) compared advanced Spanish learners' (N=32) use of MT output to assistant their translation (L1 to L2) as opposed to those who translated texts without the help of MT and found that the experiemnt group produced significantly smaller nmebr of lexical, grammatical and spelling errors than the control group. In a similar vein, Lee (2020) have intermediate and high intermediate learners of English translate a text from L1 to L2 by themselves, and then compared their own translations with the MT versions to make revisions. The results showed that MT texts helped improve learners' writings at the lexoco-grammmatical levels and posiitively improve their writing strategies. Descipte the apparent benefits of MT for intermediate and sdvanced learners, it remains unclear whether MT is also beneficial to low-proficient learners of L2. Most research so far has argued that MT can be much more effectively used by advanced learners than beginners (Larson-Guenette, 2013; Niño, 2009; Kol, Schcolinik, and Spector-Cohen, 2018; Kaye, 2009)

Method

Participants
The participants in the present study are 34 eleventh-grade students (L1=Chinese; L2=English) from a senior high school in Taiwan. Their English proficiency was measured based on the reading and listening sections of the General

English Proficiency Test (GEPT). GEPT is held by *The Language, Training and Testing Center* (LTTC), a foundation established in 1951 Taiwan and formally recognized by the government of Taiwan to hold proficiency tests of different languages. It has been widely used by educational institutions at high school level for college application and university level as a threshold of graduation. The intermediate level of the GEPT is generally considered to be the level of English proficiency for senior high school graduates. It corresponds to B1 (reading and listening skills) and B1+ (writing and speaking skills) of the CEFR. Since the participants in the present were all in their second year of senior high school, it was deemed appropriate to use GEPT to measure their English proficiency.

Material

An excerpt from the fairy tale '*The Cowherd and the Weaver Girl*' written in Chinese was machine-translated into English using Google translate. The reason why a narrative text was chosen as the material for the current study was that high school learners in Taiwan practice this text type most for the requirement of college entrance exam. The ability to write a narrative essay in English is thus considered essential at the high-school level. The machine translated text (henceforth referred to as MT text) consisted of 290 words in English , which serves as the material for error correction for the participants. Samples of the source text and MT texts are presented below:

a.Sample source text:

傳說天上有個織女星，還有一個牽牛星。織女和牽牛情投意合，心心相印。可是，天條律令是不允許男歡女愛、私自相戀的。織女是王母的孫女，王母便將牽牛貶下凡塵了，令織女不停地織雲錦以作懲罰。織女的工作，便是用了一種神奇的絲在織布機上織出層層疊疊的美麗的雲彩，隨著時間和季節的不同而變幻它們的顏色，這是"天衣"。

b.Sample MT text:

Legend has it that there is a Vega and an Altair in the sky. The Weaver Girl and the Morning Bull are in agreement, and they are in harmony. However, the law of heaven does not allow men and women to love or fall in private. As the Queen Mother's granddaughter, the Queen Mother demoted the petunias to the mortal world, causing the Weaver Girl to weave the brocade for punishment. The work of

the Weaver Girl is to use a kind of magical silk to weave layered beautiful clouds on a loom, and change their colors with time and seasons. This is "Heavenly Clothes".

Procedure

The participants are instructed to read and compare the source text (Chinese) and the MT text (English). They are given 30 minutes to post-edit the combined MT text by detecting and correcting errors using paper and pencil. In terms of the format and method of error correction, they were instructed to underlie the part of sentences they deemed problematic, number it and make corrections in the revision column (See Appendix 1 for examples). Efforts were made to ensure the participants understood the procedures of error correction.

Data analysis

The post-edited texts will be analyzed by the researcher in the following steps. First, the number of corrected errors will be counted twice. Second, the unit of post-editing will be recorded and categorized into four levels: words, phrases, clauses and sentences. A corrected error was included in the clause level if the subject and its predicate in a part of a sentence was changed, while it was included in the sentence level if the whole sentence or the sentence structure is changed. When determining the level of corrected errors, the level of the corrected error, rather than the detected error, was counted. That is, if a participant changed a phrase into a word, then the error was counted in the word level. The detected error was deemed 'corrected' if the grammaticality and naturalness of the sentence was improved, but was not included in corrected errors if the revision did not improve in grammaticality and naturality. (e.g., replacing a wrong word with another ungrammatical/ inappropriate word), made the error even worse or more awkward, or the detected error was left uncorrected. Table 1 presents samples of corrected errors at the four levels.

TABLE 1

Samples of Corrected Errors at the Four Levels

	Level	Detected Errors	Corrected Errors
Word	word > word phrase > word	*He* *were born*	*They* *died*
Phrase	Ø > phrase word > phrase phrase > phrase	with time and seasons *today* *agreed to them*	with *the changes of* time and seasons *that day* *gave them the permission*
Clause	clause > clause	*when his parents were born*	*when his parents left this world behind*
Sentence	sentence > sentence	*After the petunias were demoted, they were born in a farmer's house*	*After being demoted, Niulang was born in a farmer's house.*

Results

A one-way between-group ANOVA was conducted to explore the impact of English proficiency on how learners post-edited MT texts. Participants were divided into three proficiency groups according to their GEPT scores: high-proficient (the top 25%), low-proficient (the bottom 25%), and intermediate (the remaining 50%). There was a statistically significant difference at the $p < .05$ level in learners' post-editing patterns for the three proficiency groups: $F(2, 31) = 4.78$, $p = .015$, indicating that L2 proficiency does have a significant effect on learners' post-edited MT output. A LSD post-hoc test was performed to examine the between-group differences in the number of corrected errors. The results showed that the high-proficient learners corrected significantly more errors than the intermediate ($p < .03$) and the low-proficient learners ($p < .005$), while the intermediate learners did not significantly outperform the low-proficient learners ($p = .23$) in error correction.

Regarding the levels of corrected errors for the three proficiency groups, The high- and intermediate proficient groups made the highest number of corrections at the word level, whereas the low- proficient group made the most corrections at the phrase levels. For all proficient learners, around half of the corrections (44-50%) made were words, and the other half (44%-48%)were phrases. The corrections made at the clause (0-4%) and sentence (0-8%) level were minimum. The percentage,

means and standard deviations of the four error levels across proficiency groups were presented in Table 2.

Table 2

Descriptive Statistics of Corrected Errors at the Four Levels across Three Proficiency Groups

	Error Level	Percentage(%)	Mean	SD
High	Word	49.30	3.89	2.21
	Phrase	47.9	3.78	2.19
	Clause	2.80	0.22	0.67
	Sentence	0.00	0.25	0.00
Int	Word	50.0	2.41	3.14
	Phrase	43.9	2.12	1.45
	Clause	3.70	0.18	0.39
	Sentence	2.40	0.12	0.33
Low	Word	44.0	1.38	1.85
	Phrase	48.0	1.50	1.20
	Clause	0.00	0.00	0.00
	Sentence	8.00	0.12	0.71

The Kolmogorov-Smirnov test was performed to ensure the normality of the number of corrected errors (p =.84). An one-way ANOVA followed by an LSD post-hoc test was conducted to examine the group differences at four error levels. The results showed that L2 proficiency had a significant effect only for phrases ($F(2,31)$= 4.87, p=.014), but not for words, clauses and sentences. Specifically, at the phrase level, the mean number of corrections for the high-proficient group was significantly higher than that for the intermediate (p = .017) and low-proficient groups (p = .006), while there was no difference between intermediate and low-proficient groups in the correction of phrases. As for the corrections of words, clauses and sentences, there was no significant difference among the three proficiency groups. The results are presented in Table 3.

Table 3

Main Effect of Proficiency Group in Four Error Levels

	F	*p*	Post-hoc (LSD)
Word	1.930	.162	
Phrase	4.869	.014*	Low < High**, Int < High*
Clause	.614	.548	
Sentence	.779	.467	

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

When the percentage of corrections (the number of corrected errors compared with the number of detected errors) was compared across proficiency groups, it was found that high-proficient group had the highest percentage of error corrections (83%). However, the percentage of corrections of intermediate and low-proficient groups was not different (66%).

Discussion and Conclusion

The present study investigated the effect of L2 proficiency on EFL learners' post-editing skills of machined-translated texts. In particular, the study examined how high-, intermediate and low-proficient learners corrected errors at the four language levels: word, phrase, clause and sentence. The results revealed that L2 proficiency did have a significant effect on the corrections of MT texts. In terms of the frequency of corrections, word- and phrase-level corrections were found to be the most common across the three proficiency groups. No notable differences were found regarding the percentage of errors corrected at the four levels across proficiency groups. The findings are not in consistent with previous research (Chung, 2020; Lee, 2019) on MT post-editing, where word-level corrections were twice as frequent as phrase or clause/sentence level ones. In terms of the main effect of L2 proficiency, among the four error levels, only the phrase-level errors were found to show significant between-group differences. That is, high-proficient learners made significant more phrase-level corrections than the other two proficiency groups. The finding did indicate that high-proficient learners seemed to be more willing to reconstruct longer sentences as compared with intermediate and low-proficient learners. The fact that learners from all proficiency groups produced

very few clause- and sentence-level corrections can possibly be explained by the composition and the learning environment of the participants. First, the participants in the present study are students from a gifted (in math and science) class of a senior high school in Taiwan. In fact, this senior high school is considered to be the first choice for local junior high graduates who aim for academic success. Therefore, although the GEPT results were shown to be significantly different among the three proficient groups, it is believed that the three groups are still highly homogeneous in their English ability. Another possible reason for the low number of corrections made beyond the phrase level is that Taiwanese high school students generally did not have the chance to translate or write English essays until they enter the twelfth grade, when they would prepare for the translation and writing sections of the college entrance exam. As the participants in the study are eleventh graders, their school training in English was mainly on the input (reading and listening), and not the output (writing and speaking). It could be due to this reason that learners were less willing to consider making revisions at a more holistic level and the predominant corrections were made at a local level. In fact, among the 34 participants who took the GEPT (intermediate level) in the study, 85% (n=29) of them did pass the test, with their total scores exceeding 160 points, which is the threshold of passing the GEPT. Thirdly, the fact that the participants were instructed to first underline the problematic part, number it and then make corrections in the revision column may also have played a role in the overall insignificant differences at the four error levels among the three proficiency groups. They could have been misled into changing only the lexical- and phrasal-level parts of the original sentences due to the instruction.

Despite the limitations, the present study did show that EFL learners' MT-post editing skills were significantly influenced by their L2 proficiency, which supports the findings of previous studies on the effect of L2 proficiency indicating that advanced learners produced longer, more complex and more accurate sentences with vocabulary diversity (Crossley & McNamara, 2012; Grant & Ginther, 2000; Hwang, 2012; Jarvis et al., 2003; Jung et al., 2019; Shin & Kim, 2014). While the results of the present study did not reveal significant difference in the length and complexity of revised words and phrases among the three proficiency groups, the high-proficient learners did show higher accuracy in their revisions (83%) and

produced significantly higher number of phrase-level corrections than the other two groups, which suggested longer revisions.

The results of the present study yielded several pedagogical implications. First, post-editing MT texts seems to be an appropriate task for teachers to measure the underlying linguistic knowledge of the students and to help enhance their awareness of the similarities and differences of L1 and L2. Second, at high- school level, attention should be paid not only at the grammaticality and the meaning of words and phrases. For advanced learners, teachers can help students focus more on mistranslations at the discursive levels (e.g. cohesion and coherence) as well as the appropriate use of sentence structures in particular contexts and for different registers. For low-proficient learners, while they should continue to work on the grammaticality of sentences, word choice and collocational use should not be overlooked. Furthermore, in post-editing MT texts and MT use, low-proficient learners may need specific guidelines and scaffoldings as to when and how to use it since they tend to accept the MT output without critical evaluation of the structure and contextual use of lexicons. In designing MT-related activities, for instance, teachers can differentiate instructions by giving high-proficient learners more challenging tasks such as reconstructing sentences to make more precise translations while providing low-proficient learners more guidelines in detecting grammatical and lexical errors such as introducing tools for checking corpus-based collocational uses. The current study, due to its preliminary nature, has several limitations. Apart from the lack of heterogeneity (mostly high-and intermediate proficient) and the relative small sample size (N=34), the genre of the text (narrative), and the instructions and format of post-editing (offline) as well as the translation direction (L1 to L2) can all be further extended in the future research.

References

Chen, M. H., Huang, S. T., Chang, J. S., &Liou, H. C. (2015). Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Computer Assisted Language Learning*, *28*(1). https://doi.org/10.1080/09588221.2013. 783873

Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. *Journal of Asia TEFL*, *17*(1). https://doi.org/10.18823/asiatefl. 2020. 17.1.11.182

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

Crossley, S. A., &McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2). https://doi.org/10.1111/j.1467-9817.2010.01449.x

Dhakar, B. S., Sinha, S. K., &Pandey, K. K. (2013). A Survey of Translation Quality of English to Hindi Online Translation Systems (Google and Bing). *International Journal of Scientific and Research Publications*, *3*(1), 2250–3153.

Ducar, C., &Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*, *51*(4). https://doi.org/10.1111/flan.12366

Enkin, E., &Mejías-Bikandi, E. (2016). Using Online Translators in the Second Language Classroom: Ideas for Advancedlevel Spanish. *Latin American Journal of Content & Language Integrated Learning*, *9*(1). https://doi.org/10. 5294/laclil.2016.9.1.6

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

Grant, L., &Ginther, A. (2000). Using Computer-Tagged Linguistic Features to Describe L2 Writing Differences. *Journal of Second Language Writing*, *9*(2). https://doi.org/10.1016/S1060-3743(00)00019-9

Groves, M., &Mundt, K. (2015). Friend or foe? Google translate in language for academic purposes. *English for Specific Purposes*, *37*. https://doi.org/10.1016/j.esp.2014.09.001

Hwang, E. (2012). Korean EFL Learners' Language Development across Proficiency Levels in Written Productions. *English Teaching*, *67*(3), 27–50. https://doi.org/10.15858/engtea.67.3.201209.27

Im, H. J. (2017). The university students' perceptions or attitudes on the use of the English automatic translation in a general English class: Based on English writing lessons. *Korean Journal of General Education*, *11*(6), 727751.

Jarvis, S., Grant, L., Bikowski, D., &Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, *12*(4). https://doi.org/10.1016/j.jslw.2003.09.001

Jung, Y. J., Crossley, S., &McNamara, D. (2019). Predicting second language writing proficiency in learner texts using computational tools. *Journal of Asia TEFL*, *16*(1). https://doi.org/10.18823/asiatefl.2019.16.1.3.37

Kim, K. (2018). The Influence of Case Markers on the Machine-Translation of Korean Proverbs into English. *The Linguistic Association of Korea Journal*, *26*(3). https://doi.org/10.24303/lakdoi.2018.26.3.139

Kol, S., Schcolnik, M., &Spector-Cohen, E. (2018). Google Translate in Academic Writing Courses? *The EuroCALL Review*, *26*(2). https://doi.org/10.4995/eurocall.2018.10140

Larson-Guenette, J. (2013). "It's just reflex now": German Language Learners' Use of Online Resources. *Die Unterrichtspraxis/Teaching German*, *46*(1). https://doi.org/10.1111/tger.10129

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

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

Luton, L. (2003). If the computer did my homework, how come i didn't get an "A"? *French Review*, *76*(4).

Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, *21*(1). https://doi.org/10.1080/09588220701865482

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

Shin, H. W., &Hyun Jung, K. (2014). Linguistic differences in the writing performance of adolescent EFL learners: The influence of independent and integrated tasks. *Journal of Asia TEFL*, *11*(1).

Somers, H., Gaspari, F., &Niño, A. (2006). Detecting inappropriate use of free online machine translation by language students - A special case of plagiarism detection. *EAMT-2006 - 11th Annual Conference of the European Association for Machine Translation*.

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

White, K. D., &Heidrich, E. (2013). Our Policies, Their Text: German Language Students' Strategies with and Beliefs about Web-Based Machine Translation. *Die Unterrichtspraxis/Teaching German*, *46*(2). https://doi.org/10.1111/tger.10143

Williams, L. (2006). Web-based machine translation as a tool for promoting electronic literacy and language awareness. *Foreign Language Annals*, *39*(4). https://doi.org/10.1111/j.1944-9720.2006.tb02276.x