Post-Editing Metaphorical Expressions: Productivity, Quality, and Strategies[1]

⊙ Yanfang Jia, Si Lai
Hunan Normal University, Shaanxi Institute of International Trade and Commerce

Abstract: This study aims to explore the impact of neural machine translation (NMT) post-editing on metaphorical expressions from English to Chinese in terms of productivity, translation quality, and the strategies employed. To this end, a comparative study was carried out with 30 student translators who post-edited or translated a text rich in metaphors. By triangulating data from keystroke logging, retrospective protocols, questionnaires, and translation quality evaluation, it was found that: (1) processing metaphorical expressions using NMT post-editing has significantly increased the translators' productivity compared to translating them from scratch; (2) NMT was perceived to be useful in processing metaphorical expressions and post-editing produced fewer errors in the final output than translation from scratch; (3) different strategies were used to process metaphorical expressions in post-editing and from-scratch translation due to the inherent differences in the two tasks, with "direct transfer" used most frequently in post-editing as translators usually rely on the NMT output to produce the final translation but more balanced strategies adopted in from-scratch translation as they need to seek for different solutions to rendering the metaphorical expressions; the quality of NMT output played a major role in what strategies were adopted to process the metaphorical expressions and their final product quality in post-editing, rather than the conventionality of the metaphorical expressions in the source text. Practical and research implications are discussed.

Keywords: neural machine translation post-editing, metaphorical expressions, productivity, translation quality, translation strategies

CLC: H059 **Document Code:** A **Article ID:** 2096-4374(2022)02-0028-16

Doi: 10.53397/hunnu.jflc.202202004

1. Introduction

In recent years, the newly developed neural machine translation (NMT) has become an increasingly hot topic both in translation research and industry. Compared to statistical machine translation (SMT), NMT has greatly improved translation quality in terms of adequacy and fluency (Popović, "Language-Related Issues" 237-253; Yamada 87-106; Jia et al., "Post-Editing" 9-29). Although NMT is not yet perfect, correcting its errors by human post-editing (PE) is expected to

be faster than translation from scratch (HT). PE is "the task of editing, modifying and/or correcting pre-translated text that has been processed by an MT system from a source language into a target language" (Allen 297). PE has now been adopted as a promising approach worldwide in the language service industry in order to meet the globally increasing translation demands. In China, according to a report of the Translation Association of China in 2020, among the 8,928 documented companies that specialized in language and translation services at the end of 2019, 36.8% reported that PE was frequently used in their daily practice.

Increasing research on PE has shown that PE is different from HT in several aspects, including productivity, translation quality, and how source text and target text are processed. However, the results are not always consistent. Post-editing MT output for domain-specific texts has often been found to be faster than HT without compromising the translation quality (Plitt and Masselot 7-16; Garcia 7-21; O'Curran 113-118; Wang et al., "Investigating" 100-123). For general texts, however, PE is not always superior. Daems et al., for instance, found that when translating English into Dutch, PE with SMT output was significantly faster than HT for news texts without decreasing its translation quality (245-270). However, Jia et al. found no significant productivity increase using PE with NMT for general type texts from English to Chinese ("How" 60-86). These inconsistent results may be caused by the different MT systems used in these comparison studies. It could also be due to the different source texts involved, as recent studies show that differences in task difficulty between PE and HT can be affected by not only MT quality but also source text complexity (Jia and Zheng 36-55). General texts tend to have more negative translatability indicators, such as metaphorical expressions (MEs) (O'Brien, "An Empirical Investigation" 83-136). News texts, for example, are characterized by a higher percentage of metaphors (Steen et al. 43-60), and these features have been reported to pose various challenges for HT (Zheng and Xiang, "Processing" 160-183; Sjørup 69-71). For PE, however, previous studies using general type source texts have scarcely investigated in detail how these specific features are processed, and the differences in translation quality and strategy from HT.

The present study, therefore, attempts to explore the possible impact of NMTPE on processing MEs from English into Chinese. It seeks to answer the following three questions:

Compared with HT,

（1）Does NMTPE boost productivity when processing source text rich in MEs?

（2）Does NMTPE improve the final translation quality of MEs?

（3）Does NMTPE affect translation strategies used to process MEs?

2. Related Work

2.1 Post-Editing versus Human Translation

MTPE research has witnessed great developments over the past decade. Extensive studies have been conducted to investigate the differences between MTPE and HT from various perspectives, such as process, product, and translators' perceptions.

As increasing productivity is the main concern for the industry to adopt PE, processing speed has

been a key factor investigated in earlier research. Research focusing on the productivity of PE indicates partly contradicting results. Some studies have reported productivity gains when comparing PE to HT (O'Brien, "Eye-Tracking" 185-205; Guerberof, "Productivity" web; Plitt and Masselot 7-16; Daems et al. 245-270; Wang et al., "Investigating" 100-123), while others have found no significant increase in speed using PE (García 7-21; Carl et al., "The Process" 131-142; Lee and Liao 105-149). Jia et al. ("How" 60-86) found NMTPE to be significantly faster than HT for domain-specific texts, but for general texts, there was no significant difference. In general, processing speed shows large variation in different studies and appears to depend, to a certain extent, on specific language pairs, text types, machine translation systems, post-editors' experience, etc.

Studies have also addressed the quality of post-edited texts, as the usability of PE will be rejected if its productivity increase is at the cost of final quality. Many studies have reported that PE generates translations of higher accuracy when compared to those translated from scratch. Fiederer and O'Brien (52-74) showed that the post-edited texts got higher scores in accuracy and clarity, but lower scores in style than manually translated versions from English to German. When comparing NMTPE to HT for English to Chinese translation, Wang et al. ("Investigating" 100-123) found NMTPE output to be superior both in adequacy and fluency. Some studies indicate that PE delivers comparable translation quality in comparison to human from-scratch translation (Carl et al., "The Process" 131-142; Jia et al., "How" 60-86). In addition, Plitt and Masselot (7-16), García (7-21), and Guerberof ("The Role" 51-76) consistently found PE reduced errors in translation output. The occurrence percentage of error types has also been found to be different in the translated and post-edited output (Daems et al. 245-270). The above studies indicate that PE can produce translations with quality similar to or even better than manually translated texts, especially in terms of accuracy.

The impact of PE on translation strategies has scarcely been investigated. According to the few studies addressing translation strategies in PE and HT, post-edited texts were found to be closer to the source text in lexical and grammatical structure, compared to HT (Depraetere 1-7; Čulo 200-218; Carl et al. "Post-Editing" 145-175). Other studies have investigated the differences in translation strategies between PE and HT from the perspective of creativity. Moorkens et al. (240-262), Kenny and Winters (123-149), and Guerberof and Toral (255-282) reported that PE decreased creativity in translations, as the translators mainly relied on the MT output to generate the final PE output. Jia and Zheng (36-55) found that source text complexity and MT quality can interact to affect the speed, quality, and strategies used to process source text and MT in PE.

2.2 Post-Editing Metaphorical Expressions

With the publication of the groundbreaking work *Metaphors We Live By*, conceptual metaphors prompted new research interests in cognitive linguistics. Metaphor, henceforward, exists at two levels, namely, the conceptual level and the linguistic level. A conceptual metaphor is a cognitive concept deeply embedded in human culture through which an abstract concept can be understood, while a linguistic metaphor, or a ME, is the surface realization of a conceptual metaphor in a given text or speech (Lakoff 202-203). MEs, therefore, are easier to be definitively identified and observed. Linguistic metaphors (also referred to as ME hereafter) were investigated in this study.

Metaphors can be further divided according to the mapping relations between the source and target domains from different perspectives. Some classification schemes are finely granulated but overlap to different degrees (Newmark 84-97; Goatly 217-232; Deignan ch.16). For example, from conceptual and cognitive perspectives, Lakoff and Johnson have classified metaphors into structural, ontological and orientational metaphors, but the boundaries between these groups are not always clear.

Other studies perceived factors, such as conventionality, familiarity, and frequency, rather than the types of conceptual mappings as major contributors affecting the processing effort needed for metaphors. The career of metaphor theory proposed by Gentner and Bowdle groups metaphors into conventional and novel metaphors (223-247). Conventional metaphors become lexicalized as a secondary meaning of the topic term, so when a conventional metaphor is processed, two possible interpretations may be activated and then only the relevant one is kept for use. For novel metaphors, no interpretation is readily available and two conceptual concepts must be connected via a cross-domain mapping (229). Similarly, the Gradient Salience model brought forward by Giora suggests that it is the "saliency" of the linguistic expressions characterized by conventionality, prototypicality, familiarity, and frequency that determines whether metaphorical expressions can be understood rapidly. According to these classifications, conventional metaphors are more frequently used and well-established in language use than novel metaphors which may be more difficult to process. Novel metaphors have been found to cost more effort to process than conventional ones (Gentner and Wolff 331-355; Lai et al. 145-155).

Historically, there has been scant empirical research on MEs from a translation perspective. A few studies have reported that the inclusion of MEs is problem-prone and it, compared to literal expressions, slows down the processing speed of translators (Zheng and Xiang, "Processing" 160-183; Sjørup 203-209). Sjorup found that the conventionality of MEs affected the comprehension and production of these metaphors in translation. Whether the conventionality of a ME has an impact on MTPE needs to be further investigated. It is reasonable to assume that linguistic metaphors may pose new challenges to PE. A few recent studies have started to show interest in whether PE will outperform HT in processing MEs. The most related research was conducted by Koglin (126-141), which compared SMT post-edited metaphors to manually translated metaphors from English to Brazilian Portuguese in terms of PE effort. She found that for source text rich in MEs, the cognitive effort required to MTPE was lower than HT in terms of overall task time, pauses, and insertions. However, when checking whether processing MEs was cognitively less effortful than HT using eye-tracking data, the results were mixed. This study chose the whole source text area as the area of interest rather than the specific MEs in the source text when checking the cognitive effort and didn't check either the source text complexity or the translation quality of the SMT output for the MEs. Therefore, it is difficult to know how these MEs were processed in detail. For English to Chinese translation, Wang et al. found that PE reduced the cognitive effort needed to process MEs indicated by eye-tracking data and the overall quality of the post-edited MEs was better than the translated ones ("A Comparative Study" 115-126). However, the MT quality for the MEs was not reported in this study. Therefore, how the MT output impacted the product quality of the post-edited MEs was unknown.

The aforementioned studies are helpful references to this study, but NMTPE with logographic languages, such as Chinese for linguistic metaphors need to be further explored, especially in terms of how MEs are processed strategically, and the quality of the NMT output, detailed error analysis of the post-edited product quality, and productivity.

3. Research Design

3.1 Participant Profile

Thirty Master in Translation and Interpreting (MTI) students at a Chinese university, were recruited to participate in the present experiment. There were 25 females and 5 males with an average age of 24 (SD=3.58). All the participants had the same language background (Chinese as L1, English as L2) and similar English language proficiency as they all had learned English as their second language for more than ten years and had passed the Test for English Majors-Band 8.[2] They were at a similar level in terms of professional translation and PE experience. They all took part in a training session in PE before the experiment. The participants were divided into the HT group (Participants P01 to P15) and PE group (Participants P16 to P30), based on their scores in the latest translation exam, ensuring that the two group's translation competence were roughly identical. All the participants were guaranteed anonymity and confidentiality and received a gift worth ¥30 for their time of participating.

3.2 Source Text and the Identification of MEs

An excerpt from Bill Clinton's farewell speech in 2001 was employed as the raw material in this study.[3] To improve the density of MEs, several sentences were extracted from the speech and combined into the source text (157 words). Two external well-educated English native speakers assessed the readability and comprehensibility of the source text to ensure it was self-contained and could be understood without additional context. On account of the basic information about the source text (listed in Table 1), it was of acceptable difficulty, i.e., neither too difficult nor too easy for the participants. And no Chinese translation of the text was found on the Internet.

Table 1. Quantitative profiling of the source text.

Indicators	The source text
Word Count	157
Mean Sentence length	31.60
Lexile Measure	1400L-1500L
Flesch Reading Ease Score	55.3
Grade Level	13
Reading Level	Fairly difficult to read
Reader's Age	18-19 yrs. Old (college-level entry)

Seven MEs[4] (designated as M1-M7) were identified in the source text based on the definition

of linguistic metaphor and relevant references (Zheng and Xiang, "The Impact" 5-24). These metaphors were further classified into conventional and novel metaphors by checking dictionaries and analyzing the linguistic context (Way ch.6). If the metaphorical sense of a lexical unit was listed in dictionaries, it was labeled as conventional; otherwise, it was identified as novel. *Macmillan English Dictionary for Advanced Learners* (Rundell), *Longman Dictionary of Contemporary English* (Longman) and *Oxford English Dictionary* (Simpson and Edmund) were used as main reference sources to identify and classify the MEs. The identification and classification of MEs were carried out by two English native speakers. In the end, M1 to M6 were ascertained as conventional MEs, while M7 was labeled as a novel one. The source text was pre-translated by the Google Neural Machine Translation system before PE.

3.3 Experimental Procedures

To triangulate the data, key-logger, retrospective protocols and questionnaires were used to collect quantitative and qualitative data from the participants. Before the formal experiment, a pilot study was conducted with two participants who were not part of the main experiment to further modify our research design. The main experiment included eight stages:

(1) The experimenter introduced the task and the experiment stages.
(2) The participants signed an informed consent form, which was approved by the Ethics Committee of the College of Foreign Languages at the University.
(3) The participants completed the pre-task questionnaire concerning their educational background and professional experience as translators and post-editors, as well as their perceptions of MT and PE.
(4) The participants were asked to read translation briefs containing the target audience and the expected final quality of their translations. The PE group, additionally, was provided with PE guidelines for publishable quality (TAUS web).
(5) The participants were assigned a warm-up task, which required translating a non-experimental English text (50 words) into Chinese, followed by cued retrospective protocols.
(6) The participants manually translated the source text or post-edited the Google NMT output. Their typing activities were recorded by Translog-II. There was no time limit and they had access to the Internet and other external resources. According to the translation briefs and the PE guidelines, the participants could press "stop logging" button when they thought their target texts had reached the publishable quality criteria.
(7) While Translog-II was replaying their translation process, the participants were asked to make cued retrospective protocols about their processing of the seven MEs, including translation problems they encountered, translation strategies they adopted, and whether they were satisfied with the translations. The protocols were recorded.
(8) The participants filled out the post-test questionnaire, to enquire about their attitudes towards HT and PE in terms of processing speed, translation quality, and their translation challenges and strategies when dealing with MEs.

3.4 Data Analysis

Four streams of data, namely, from-scratch translations and post-edited texts in Translog files, key-logging data in Translog files, audio recordings of the participants' retrospective protocols and questionnaires, were collected during the experiment for triangulation purposes.

The final Translog-II xml-files were manually aligned in the YAWAT tool (Germann 20-23) and then they were processed into a set of tables with user activity data containing over 200 features, which include the process and product data of the HT or the PE tasks (Carl et al., "The CRITT" 13-54). A preliminary analysis revealed that all 30 key-logging data were valid for future analysis, as all of the participants had not read the source text before (as shown in the pre-task questionnaire). Besides, their logged files were correctly saved and the instructions in the translation briefs and the PE guidelines were followed.

The data analysis was performed using SPSS version 19.0 statistical software. Independent samples t-test was conducted on the experimental data to test whether the differences between the HT group and the PE group were statistically significant. The cut-off point for the significance level was set at 0.05.

4. Results and Discussion

The following sections present the results for the differences in process and product between NMTPE and HT in English-Chinese metaphor translation. The impact of NMTPE was examined and compared with HT from three perspectives: (1) productivity, as recorded by keystroke logging; (2) translation quality, as established by error analysis; (3) translation strategies, as revealed by retrospective protocols and final output. Qualitative analyses based on cued retrospective protocols and questionnaires were incorporated in each section to explain the quantitative results.

4.1 Productivity

Productivity is measured by the time taken to process each source text token in seconds. Overall, the task speed for the PE group was significantly faster than the HT group (Table 2). PE has been constantly reported to be faster than HT for technical texts (O'Brien, "An Empirical Investigation" 83-136; Zhechev 2-24; Jia et al., "How" 60-86), our results show that it can also boost the student translators' productivity when processing source text rich in MEs, which is in line with Koglin (126-141). The time-saving effect in PE to a certain extent confirms the usefulness of the NMT output, as the participants didn't have to translate the source text from scratch. The post-task questionnaires also showed that all participants agreed that MT output was useful for their PE task with 33.33% choosing "totally agree" and 66.67% "agree". This was supported by the keylogging data which showed that a large proportion of the MT output remained unchanged in the post-edited texts for all participants in the PE group.

These results can be further explained using minimax theory, which takes translation as a process of decision-making (Levy 148-159). The translator, from a set of possible solutions, resolves for the one "promises a maximum of effect with a minimum of effort" (156). One of the goals of PE is to maximize productivity and meanwhile, achieve expected end quality by making

the best of the raw MT output, so when the MT output is regarded as good enough to fulfill the effects, the post-editors will stop investing more effort in seeking for alternative solutions. In HT, translators are more likely to generate various translated versions and then select the most desirable one that optimally realizes the effects. PE can, therefore, yield productivity gains by saving the time for decision-making and external resources consulting.

Table 2. Task speed in terms of the time taken to process each source text token in seconds for HT and PE.

	Mean	Std. Deviation
HT (P01-P15)	9.29	2.64
PE (P16-P30)	7.23	2.39
independent sample t-test	t=2.17; p<0.05; df=26; ES=0.85	

4.2 ME Translation Quality

The final translation quality of the MEs from the PE and HT groups was analyzed based on the customized error types developed by the Multidimensional Quality Metrics, in terms of accuracy errors (mistranslation, addition, omission, over-translation and under-translation) and fluency errors (grammar, wrong collocation, logical problem, typing and unintelligible) (Lommel 109-127). The total number of errors was calculated as an indicator of the overall ME translation quality. Two professional translators annotated all the ME translations for errors in accuracy and fluency. The operational definition of each error type and examples were demonstrated in Table 3.

Table 3. Samples of Error analysis for the ME translations based on MQM.

Dimensions	Error types	Operational Definition	Examples
Accuracy	Addition	The target text includes text not present in the source.	ST: be ignited by our indifference (M5) TT: 被我们的冷漠无情所点燃 Back translation: be ignited by our indifference and hard-heartedness
	Mistranslation	The target content does not accurately represent the source content.	ST: the cutting edge (M2) TT: 飞速发展的时代 Back translation: an era of rapid development
Fluency	Grammar	Issues related to the grammar or syntax of the text.	ST: weave the threads of our coat of many colors into the fabric of one America (M7) TT: 消除不同肤色的偏见,实现融合 Back translation: eliminate the prejudice against different skin colors and achieve integration
	Wrong collocation	Issues related to uncommon combinations of words	ST: the same as above (M7) TT: 和平地对待每个种族 Back translation: treat every race peacefully

The two raters annotated the translated and the post-edited MEs (105 MEs in total for each group) from all the participants in a randomized order, without being informed of whether they were translated from scratch or post-edited. A ME translation may present no error, errors in

accuracy, fluency, or both. All error classifications were discussed by both annotators, and only the ones they both agreed on were retained for further analysis. The results of the detailed error analysis were presented in Figure 1.

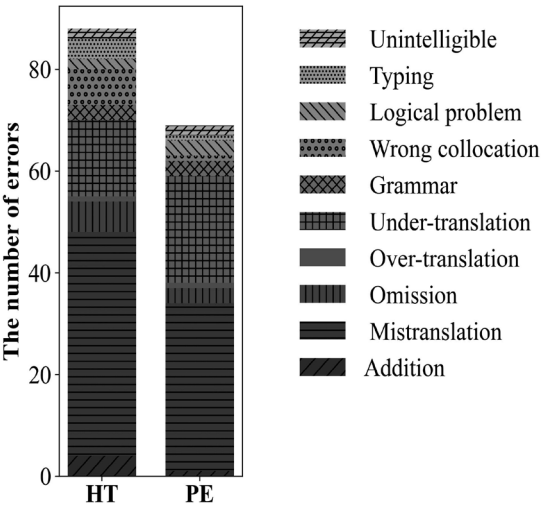

Figure 1 The total number of errors in the final ME translation from HT and PE.

Overall, the post-edited MEs were higher in translation quality with fewer total errors than those manually translated. There were 69 errors in the PE and 88 in the HT groups' output and the difference was statistically significant ($t=2.24$; $p<0.05$; $df=28$; $ES=0.83$). This result indicates that PE based on NMT output has produced MEs with better translation quality compared with HT. The detailed error types analysis showed that mistranslation was the most common error type both for HT (50%) and PE (48%) out of all error types. For the PE tasks, all participants perceived the NMT output to be helpful and they further reported in their retrospections that the lexical information provided by the NMT helped them achieve a better understanding of the MEs in the ST.

Table 4. Translation quality of the NMT output for the MEs based on MQM.

MEs	MT output	Error types
M1	消除差距 (closed the gap)	None
M2	前沿 (the frontier)	Under-translation
M3	在刀刃上 (the knife's edge)	Under-translation Logical problem
M4	火药桶 (a powder keg)	None
M5	…是我们漠不关心的… (… which we treat with indifference)	Omission
M6	在...中树立人性的面貌 (establish the characteristics of humanity on…)	Mistranslation
M7	将我们的多种颜色外衣的线织到一个美国的织物中 (weave the threads of our multicolored coats into the fabric of one America)	Mistranslation Unintelligible

* Back translations in brackets were provided by the authors.

Although NMT output has been found helpful to increase the product quality of the final

translation quality of the MEs using post-editing compared to human translation from scratch, the quality of the NMT output was not perfect. The error analysis of the NMT output based on the customized error types of MQM as introduced above showed that five out of the seven ME translations produced by the NMT involved different types of errors (see Table 4). For example, most mistranslations were found in the post-edited output for M6 (put a human face on the global economy) and M7 (weave the threads of our coat of many colors into the fabric of one America), which were mistranslated by the NMT system (see Table 4). These two English MEs are highly language-specific (Jakobsen et al. 217-249) and strongly culturally conditioned (Dobrzyńska 595-604). Although M6 is a conventional metaphor in English, there is no TT equivalent for it in Chinese just as the novel metaphor M7.

In these cases, a word-by-word direct transfer of the ME image in the ST is not ideal, as it is hardly apt in the target language. When the NMT system translated M6（在…中树立人性的面貌）and M7（将我们的多种颜色外衣的线织到一个美国的织物中）word by word, the literal interpretations inevitably led to incomprehensibility in the target language. As a result, for M6 and M7, the NMT output failed to serve as useful references for the post-editors, and they were misled by the MT output. This was particularly evident in the processing of M6. Except for P29, all translations for M6 in the PE group were mistranslations as they mostly kept the literal NMT expressions with only minor revisions. Therefore, the quality of MT output plays a major role in deciding the quality of the post-edited output for the MEs, rather than the conventionality of the MEs in the source language, especially when there are no direct TT equivalents for them in the target language.

In addition, NMT sometimes generates fluent translations with additions or omissions difficult to be detected without checking the source text carefully, which becomes problematic for PE (Castilho et al. 109-120; Popović, "Comparing" 209-220; Yamada 87-106; Jia and Zheng 36-55). For example, in the translation of M5 (be ignited by our difference), the NMT did not translate it completely. It combined M5 with M4 and rendered the two into "全球贫困是我们漠不关心的火药桶 (global poverty is a powder keg which we treat with indifference)", which led to the omission of the keyword "ignited" in M5 for five post-editors. They adopted the MT expression without noticing that "ignited" was omitted.

Regardless of task types, the total number of errors in fluency was substantially fewer than that in accuracy for both the HT (20.45% in fluency and 79.55% in accuracy) and PE (14.49% in fluency and 85.51% in accuracy). The post-test questionnaires and retrospections showed that when asked "at which step do the difficulties occur, understanding or reformulating the MEs?", in both groups, most participants reported that it was understanding the metaphors (67% for the HT group and 60% for the PE group). This indicates that comprehension was more challenging than reformulation while processing MEs in both tasks. The present study was conducted based on English-Chinese translation, i.e., L1 translation (into their native language). According to the Revised Hierarchical Model (Kroll and Steward 149-174), translation asymmetry caused by translation directionality results in different difficulties in L2 and L1 translation. In L1 translation, difficulties are more likely to occur in understanding the source text rather than reformulating a fluent target text, which explains why both groups offered translations with fewer errors in fluency. This also further verifies that the

conventionality of the MEs in the source language plays a less important role in their translation process. If there is a direct equivalent for the image of the MEs in the target language, it will be easy to be understood and reformulated during L1 translation regardless of the conventionality condition of the MEs in its source language.

4.3 ME Translation Strategies

According to Dobrzyńska, a translator can choose among three possible solutions for translating metaphors: (1) use an exact equivalent of the original metaphor, (2) seek another metaphorical phrase that would express a similar sense, (3) replace an untranslatable metaphor with its approximate literal paraphrase (595). Jenson's classification of metaphor translation strategies adds a complementary one: omit the ME (183-209). Based on these two studies and Zheng and Xiang ("The Impact" 5-24), the strategies for metaphor translation in this study could be classified into direct transfer (M-M), substitution (M1-M2), paraphrase (M-P), and omission (M-Ø). Based on the participants' retrospective reports, all 210 translations of MEs were classified by two transcribers and the results were shown in Figure 2.

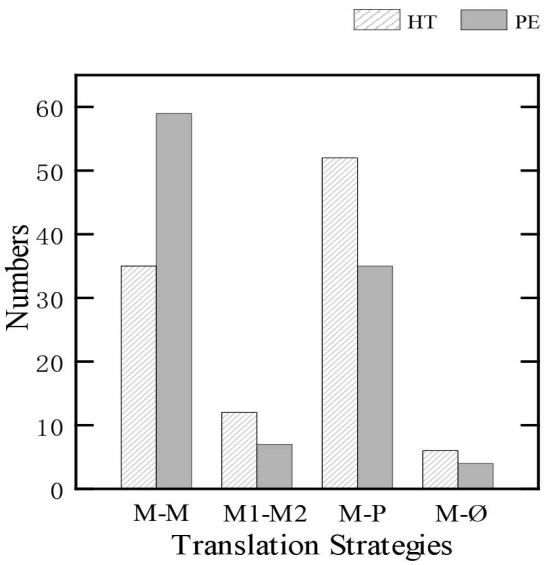

Figure 2 Numbers of four metaphor translation strategies used in HT and PE (105 MEs in total for each group).

The statistical analysis between the HT and PE groups in the ME translation strategies was further demonstrated in Table 5. The results showed that direct transfer (M-M, 56.19%) was the most favored solution adopted by the PE group, while, for the HT group, paraphrase (M-P, 49.52%) was used most frequently. When comparing the two groups, M-M was used more frequently by the PE group than the HT group, while M-P, M1-M2, and M-Ø were used more often by the HT group. Statistical significance between HT and PE was only found in using the M-M strategy ($p<0.01$) and M-P strategy ($p<0.05$).

Table 5. Comparison between HT and PE groups in ME translation strategies.

Task type	M-M (%)	M1-M2 (%)	M-P (%)	M-Ø (%)
HT	35(33.33)	12(11.43)	52(49.52)	6(5.72)
PE	59(56.19)	7(6.67)	35(33.33)	4(3.81)
Independent sample t-test	t=-3.15; p<0.01; df=28; ES=1.15	t=1.13; p>0.05; df=28; ES=0.43	t=2.50; p<0.05; df=28; ES=0.95	t=0.76; p>0.05; df=28; ES=0.29

The differences in translation strategies during HT and PE may be due to the inherent differences in these two tasks. During PE, with the raw MT available, post-editors, in most cases, only revise the MT output, and often "find it more difficult to think outside the box" (Guerberof and Toral 263). That may explain why direct transfer (M-M) was used more frequently in PE than HT. For example, in the translation of M1 (closed the gap), all the post-editors except for one (93%) made no edit on its MT output 消除差距 (closed the gap). However, NMT tends to translate the source text literally (Popović, "Language-related Issues" 237-253). This sometimes leads to translation errors especially when it comes to ME translation. As discussed in the section above (see Table 4), all seven MEs were translated literally by the NMT, and five of them were not acceptable due to the different types of errors involved.

Although direct transfer by MT was useful for most conventional metaphors during PE, it caused a major error in translating the conventional ME M6. As there is no TT equivalent in Chinese for M6, it is "novel" in this sense to the translators when doing L1 translation. While for the novel metaphor M7, no direct TT equivalent in the target language can be found for it either. Therefore, a direct transfer by MT for them can hardly serve as a favorable solution. Instead, in both cases, the translators must revise the MT output and create a new TT metaphorical image, namely, adopting M1-M2 strategy (Dagut 77-83). As post-editors were prone to be constrained by the MT output, relying too much on the NMT output in these cases sometimes resulted in errors in the final ME translation output. For example, while processing M6, six post-editors used M-M strategy by accepting the NMT output without any revision, which all resulted in mistranslation.

In HT, by contrast, translators are likely to seek more solutions to render the source text, which often leads to more creative target language expressions. In HT, M1 was translated into several expressions: 消除差距 (closed the gap), 缩小差距 (narrowed the gap), 消除鸿沟 (closed the chasm), 填补差距 (fill the gap). This happens less frequently in a more "mechanical" PE task (Carl et al., "The Process" 131-142). Although the student translators have not found ideal translations for all MEs as shown in the quality analysis, they chose translation strategies more diversely, with higher frequent use of M-P, M1-M2, and M-Ø strategies than PE. Therefore, from-scratch translations are more likely to be diversified compared to those in PE, while the post-edited texts are much closer to the original structure of the source texts (Čulo et al. 200-218; Guerberof and Toral 255-282).

M-Ø (omission) was the least used strategy both in PE and HT. A close examination of the retrospective reports reveals that M-Ø was used not when the participants' had no clue of what the metaphor meant, but as a strategic choice after ensuring that the metaphor's function has been

fulfilled elsewhere in the target text and the omission would not result in unexpressive text. For example, P05 omitted M2 (the cutting edge) and M3 (the knife's edge) because he thought "the source sentence is rather long and redundant, so it would be better to omit the two metaphors as the semantic meaning won't be compromised".

These results in translation strategies support previous findings that the differences between PE and HT should be investigated not only in terms of speed but also in terms of the translation process (Daems et al. 245-270; Jia et al., "How" 60-86). In addition, a recent study (Jia and Zheng 36-55) reported that how source text and MT were processed in PE was affected by the interaction effect between the source text complexity and MT quality. Further study is needed to check whether the strategies used to post-edit MEs will change when the quality of the NMT output is different from the one used in this study.

In addition, the quality of the NMT output relies on the parallel corpora and algorithms used to train the system to increase its context awareness and produce more fluent and accurate output than previous MT methods (Fernandes et al. web). Despite these developments, metaphor translation still poses difficulties for the NMT, as metaphorical expressions can be found in varying linguistic and cultural contexts, with morphosyntactic variations or idiomatic and literal (non-idiomatic) usages (Salton et al. 36-41). As this study shows, for these MEs whose metaphorical meanings are not readily available, it is still impossible for the NMT system to re-create their various logical and inferential properties in the source text into the target text. As the meaning of these metaphors cannot be interpreted from the meanings of the individual words, relying on the literal translation of the NMT output leads to various translation errors. Therefore, PE after machine translation is necessary and requires human translators to identify, interpret and transfer these MEs to guarantee the success of their cross-culture communication. This is important not only for news texts but also for other text types as metaphors are pervasive in language use.

5. Conclusion

This study has explored the impact of NMTPE on translating English MEs into Chinese in terms of productivity, product quality, and translation strategies. It triangulated quantitative and qualitative data from key logging, retrospective protocols, questionnaires, and translation quality evaluation to address the research questions.

For student translators, NMTPE has significantly boosted productivity in comparison to HT for source text rich in linguistic metaphors. The student translators perceived NMT to be useful and PE task helped them produce MEs with fewer errors compared to HT, although the NMT output for MEs translated from English to Chinese contained different errors. The differences in the choice of translation strategies were caused by the inherent differences in the two tasks, NMTPE has significantly affected the translation strategies adopted. Compared to HT, M-M strategy was more frequently used in PE, while M-P was used more often in HT. In PE, the student translators often relied too much on the literal NMT output evidenced by adopting mainly M-M strategy among others, which led to errors in the final translation. In contrast, different strategies were more evenly used in HT to seek more solutions to render the MEs in the source text, which resulted in more

diversified translations. Our findings showed that ME translation posed challenges both for HT and PE. In addition, the MT quality for MEs plays a more important role in the strategies used to process MEs and their final translation quality in PE, rather than the conventionality of the MEs in the source language. Therefore, further training is necessary for student translators to make better use better of MT output in PE and to improve their translation competence for source text rich in MEs.

This study only involved student translators and focused on post-editing MEs from English to Chinese, which are limitations that future studies could address by drawing on professional translators and investigating the other directionality and other language pairs. Moreover, the quality assessment and the classification of metaphor translation strategies cannot be objective. Finally, this study has only investigated NMT from Google translate. In continuing this research, MT with different quality levels from other MT engines can be selected to check whether MT quality levels impact how MEs are processed in PE in terms of productivity, quality, and strategies used.

Notes

1. Acknowledgments: The authors would like to thank the editor, the anonymous reviewers, and Prof. Zhi Lu for their comments and feedback on the earlier drafts of this article. This work was supported by Scientific Research Fund of Hunan Provincial Education Department [Grant Number 20B370], Scientific Research Fund of the Research Institute of Languages and Cultures at Hunan Normal University [Grant Number 2020QNP01], and the Model Curricula Project for Computer Aided Translation at Hunan Normal University [Grant Number 202101012].

2. The Test for English Majors-Band 8 (TEM-8) is a national English test for English majors in China. It is taken by the end of the senior year. The test contains four testing components, namely, listening, reading, writing and translation, and it requires a candidate to master 13,000 words.
3. The STs and MT output can be found at pan.baidu.com/s/1RofA_leC9Crt0dJcY4hNZg (Password: 9zno).
4. The Seven MEs identified in the ST can be found at pan.baidu.com/s/1ryJArlBL3OQ8RPwxBEECyQ (Password:8f8g).

Works Cited

Allen, Jeff. "Post-editing." *Computers and Translation: A Translator's Guide*, edited by Harold Somers, John Benjamins, 2003, pp. 297-318.

Carl, Michael, et al. "The CRITT Translation Process Research Database." *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*, edited by Michael Carl, et al., Springer, 2016, pp. 13-54.

——. "Post-editing Machine Translation: Efficiency, Strategies and Revision Processes in Professional Translation Settings." *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting*, edited by Aline Ferreira and John W. Schwieter, John Benjamins, 2015, pp. 145-175.

——. "The Process of Post-editing: A Pilot Study." *Human-Machine Interaction in Translation: Proceedings of the 8th International NLPCS Workshop, Frederiksberg, Denmark, 20-21 Aug. 2011*, edited by Bernadette Sharp, et al., Samfundslitteratur, 2011, pp. 131-142.

Castilho, Sheila, et al. "Is Neural Machine Translation the New State of the Art?" *The Prague Bulletin of Mathematical Linguistics*, no. 108, Jun. 2017, pp. 109-120.

Čulo, Oliver, et al. "The Influence of Post-Editing on Translation Strategies." *Post-Editing Machine Translation: Processes and Applications*, edited by Sharon O'Brien, et al., Cambridge Scholars Publishing, 2014, pp. 200-218.

Daems, Joke, et al. "Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-Editing with Students and Professional Translators." *Meta*, vol. 62, no. 2, Aug. 2017, pp. 245-270.

Dagut, Menachem. "More about the Translatability of Metaphor." *Babel*, vol. 33, no. 2, Jan. 1987, pp. 77-83.

Deignan, Alice. *Metaphor and Corpus Linguistics*. John Benjamins, 2005.

Depraetere, Ilse. "What Counts as Useful Advice in a University Post-Editing Training Context? Report on a Case

Study." *Proceedings of the 14th Annual Conference of the European Association for Machine Translation, St Rapheal, France, 27-28 May 2010*, edited by Francois Yvon and Viggo Hansen, The European Association for Machine Translation (EAMT), 2010, pp. 1-7.

Dobrzyńska, Teresa. "Translating Metaphor: Problems of Meaning." *Journal of Pragmatics*, vol. 24, no. 6, Dec.1995, pp. 595-604.

Fernandes, Patrick, et al. "Measuring and Increasing Context Usage in Context-aware Machine Translation." *arXiv preprint* arXiv: 2105.03482 (2021).

Fiederer, Rebecca, and Sharon O'Brien. "Quality and Machine Translation: A Realistic Objective?" *The Journal of Specialised Translation*, no. 11, Jan. 2009, pp. 52-74.

Garcia, Ignacio. "Is Machine Translation Ready Yet?" *Target*, vol. 22, no. 1, Jan. 2010, pp. 7-21.

Gentner, Dedre, and Phillip Wolff. "Alignment in the Processing of Metaphor." *Journal of Memory and Language*, vol. 37, no. 3, Oct. 1997, pp. 331-355.

Gentner, Dedre, and Brian F. Bowdle. "Convention, Form, and Figurative Language Processing." *Metaphor and Symbol*, edited by Rachel Giora, Psychology, 2001, pp. 223-247.

Germann, Ulrich. "Yawat: Yet Another Word Alignment Tool." *Proceedings of 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Columbus, the US, 16 Jun. 2008*, edited by Jimmy Lin, Omnipress Inc., 2008, pp. 20-23.

Goatly, Andrew. *The Language of Metaphors*. Routledge, 1997.

Guerberof-Arenas, Ana, and Toral Antonio. "The Impact of Post-Editing and Machine Translation on Creativity and Reading Experience." *Translation Spaces*, vol. 9, no. 2, Dec. 2020, pp. 255-282.

Guerberof-Arenas, Ana. "The Role of Professional Experience in Post-Editing from a Quality and Productivity Perspective." *Post-Editing of Machine Translation: Processes and Applications*, edited by Sharon O'Brien., et al., Cambridge Scholars Publishing, 2014, pp. 51-76.

——. "Productivity and Quality in MT Post-editing." *MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT*, 26-30 Aug. 2009, aclanthology.org/2009.mtsummit-btm.

Jakobsen, Arnt Lykke, et al. "Comparing Modalities: Idioms as a Case in Point." *Interpreting Studies and Beyond,* edited by Fran Pöchhacker, et al., Samfundslitteratur, 2007, pp. 217-249.

Jensen, Astrid. "Coping with Metaphor: A Cognitive Approach to Translating Metaphor." *Hermes-Journal of Linguistics*, vol. 18, no. 35, Mar. 2005, pp. 183-209.

Jia, Yanfang, et al. "Post-editing Neural Machine Translation versus Phrase-Based Machine Translation for English-Chinese." *Machine Translation*, no. 33, Jun. 2019, pp. 9-29.

——. "How Does the Post-Editing of Neural Machine Translation Compare with From-Scratch Translation? A Product and Process Study." *The Journal of Specialised Translation*, no. 31, Jan. 2019, pp. 60-86.

Jia, Yanfang, and Binghan Zheng. "The Interaction Effect between Source Text Complexity and Machine Translation Quality on the Task Difficulty of NMT Post-Editing from English to Chinese: A Multi-Method Study." *Across Languages and Cultures*, vol. 23, no. 1, May 2022, pp. 36-55.

Kenny, Dorothy, and Marion Winters. "Machine Translation, Ethics and the Literary Translator's Voice." *Translation Spaces*, vol. 9, no. 1, Aug. 2020, pp. 123-149.

Koglin, Arlene. "An Empirical Investigation of Cognitive Effort Required to Post-Edit Machine Translated Metaphors Compared to the Translation of Metaphors." *Translation & Interpreting*, vol. 7, no. 1, Jan. 2015, pp. 126-141.

Kroll, Judith, and Erika Stewart. "Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations." *Journal of Memory and Language*, vol. 33, no. 2, Apr. 1994, pp. 149-174.

Lai, Vicky Tzuyin, et al. "Comprehending Conventional and Novel Metaphors: An ERP Study." *Brain Research*, vol. 1284, no. 11, Aug. 2009, pp. 145-155.

Lakoff, George. "The Contemporary Theory of Metaphor." *Metaphor and Thought (2nd edition)*, edited by Andrew Ortony, Cambridge UP, 1993, pp. 202-251.

Lee, Jason, and Posen Liao. "A Comparative Study of Human Translation and Machine Translation with Postediting." *Compilation and Translation Review*, vol. 4, no. 2, Sept. 2011, pp. 105-149.

Levy, Jiri. "Translation as a Decision-making Process." *The Translation Studied Reader*, edited by Lawrence Venuti, Routledge, 2000, pp. 148-159.

Lommel, Arle. "The Multidimensional Quality Metrics and Dynamic Quality Framework." *Translation Quality Assessment: From Principles to Practice*, edited by Joss Moorkens, et al., Springer, 2018, pp. 109-127.

Moorkens, Joss, et al. "Translators' Perceptions of Literary Post-Editing Using Statistical and Neural Machine Translation." *Translation Spaces*, vol. 7, no. 2, Nov. 2018, pp. 240-262.

Newmark, Peter. *Approaches to Translation*. Pergamon, 1981.

O'Brien, Sharon. "An Empirical Investigation of Temporal and Technical Post-Editing Effort." *Translation & Interpreting Studies*, vol. 2, no. 1, Jan. 2007, pp. 83-136.

——. "Eye-Tracking and Translation Memory Matches." *Perspectives,* vol. 14, no. 3, Jan. 2007, pp. 185-205.

O'Curran, Elaine. "Translation Quality in Post-Edited versus Human-Translated Segments: A Case Study." *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas, Vancouver, Canada, 22-26 Oct. 2014*, edited by Sharon O'Brien, Association for Machine Translation in the Americas (AMTA), 2014, pp. 113-118.

Plitt, Mirko, and François Masselot. "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localization Context." *The Prague Bulletin of Mathematical Linguistics*, no. 93, Jan. 2010, pp. 7-16.

Popović, Maja. "Language-Related Issues for NMT and PBMT for English-German and English-Serbian." *Machine Translation*, vol. 32, no. 3, Sept. 2018, pp. 237-253.

——. "Comparing Language Related Issues for NMT and PBMT between German and English." *The Prague Bulletin of Mathematical Linguistics*, no. 108, Jun. 2017, pp. 209-220.

Salton, Giancarlo, et al. "An Empirical Study of the Impact of Idioms on Phrase Based Statistical Machine Translation of English to Brazilian-Portuguese." *Proceedings of the 3rd Workshop on Hybrid Approaches to Translation (HyTra), Gothenburg, Sweden, 27 Apr. 2014*, edited by Rafael E. Banchs, et al., Association for Computational Linguistics, 2014, pp. 36-41.

Simpson, John and Edmund Weiner, editors. *The Oxford English Dictionary (2nd edition)*. Oxford UP, 1989.

Sjørup, Annette Camilla. *Cognitive Effort in Metaphor Translation: An Eye-tracking and Key-Logging Study*. Diss. Copenhagen Business School, 2013.

Steen, Gerard, et al. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins, 2010.

TAUS. "TAUS Post-Editing Guidelines." *TAUS the Language Data Network*, 2016. info.taus.net/mt-post-editing-guidelines.

Wang, Xiangling, et al. "Investigating Usability in Postediting Neural Machine Translation: Evidence from Translation Trainees' Self-Perception and Performance." *Across Languages and Cultures*, vol. 22, no. 1, May 2021, pp. 100-123.

——. "A Comparative Study of HT and NMT Post-Editing: Data from Eye-Tracking and Key-Logging on Metaphor Translation." *Foreign Language Learning Theory and Practice*, no. 4, Nov. 2021, pp. 115-126.

Way, Eileen Cornell. *Knowledge Representation and Metaphor*. Cromwell, 1994.

Yamada, Masaru. "The Impact of Google Neural Machine Translation on Post-Editing by Student Translators." *The Journal of Specialised Translation*, no. 31, Jan. 2019, pp. 87-106.

Zhechev, Ventsislav. "Analysing the Post-editing of Machine Translation at Autodesk." *Post-Editing of Machine Translation: Processes and Applications*, edited by Sharon O'Brien, et al., Cambridge Scholars Publishing, 2014, pp. 2-24.

Zheng, Binghan, and Xia Xiang. "The Impact of Cultural Background Knowledge in the Processing of Metaphorical Expressions: An Empirical Study of English-Chinese Sight Translation." *Translation & Interpreting Studies*, vol. 9, no. 1, Jan. 2014, pp. 5-24.

——. "Processing Metaphorical Expressions in Sight Translation: An Empirical-Experimental Research." *Babel*, vol. 59, no. 2, Jan. 2013, pp. 160-183.