Using MT post-editing for translator training

Mehmet Şahin
Izmir University of Economics
mehmet.sahin@ieu.edu.tr, rbsmsahin@gmail.com

Post-editing machine translation (MT) has emerged as a new requirement for translators due to the changing face of the translation industry. This need is augmented by the growing volume of texts to be translated and the constant progress in MT in the last decades. The current study explores the preliminary results of an experimental course where translation students post-edited texts from different fields translated by Google Translate MT system from English into Turkish. Logs kept by the students for each translation task, students’ answers to survey questions, and translation products were used to analyze the overall effect of this method on novice translators and their reaction to using MT in classroom, and to compare students’ performance in post-editing MT output and translating from scratch for similar texts. The results suggest that although MT post-editing seems to be a difficult and irritating task for the English-Turkish language pair, novice translators are positive about the place of this new component in the curriculum. The analyses of data also show that students’ performance does not seem to vary across two methods (MT post-editing and translating from scratch), which suggests that MT is not likely to contribute to students’ performance as translators. It can also be argued that using MT as a bad model has a potential to be a productive activity in translation courses.

lien video : http://webcast.in2p3.fr/videos-using_mt_postditing_for_translator_training
Introduction

“A decent MT machine is just around the corner. The computer computes chess, while Kasparov plays chess. A computer will never understand, but it can translate, at least to some extent. And, since translation without understanding is meaningless, the future of the human translator is proof-sensing what a machine has pre-translated.” Champollion (2001)

Post editing machine translation (MT) has emerged as a new requirement for translators (O’Brien, 2002). The main reasons are that the volume of texts to be translated is increasing in proportion to texts produced in the electronic environment and that we are witnessing continuous progress in MT in the last decades. Post-editing of machine translated texts is required especially for dissemination purposes, where publishable quality is sought (Hutchins and Somers, 1992). With current developments in statistical and corpus-based machine translation, the nature of the translation profession is also subject to changes where, sometimes, translators take the role of “revisers” rather than target text producers (Austermühl, 2011). Post-editing of machine translated texts has been used for language learning purposes in the recent years (Nino, 2009). MT post-editing seems to have a great potential for translator training as well (Somers, 2003), especially for less-commonly taught languages such as Turkish, where MT output usually needs substantive post-editing before it can be considered as quality translations. There is also a need to identify whether post-editing MT should be included in the translation curriculum as a skill and a means of improving translation skills, and to test students’ reaction to MT and post editing.

1. Literature review

Appendix 14 of the ALPAC report (1966) was titled Translation Versus Postediting of Machine Translation. In this appendix, results of an experiment comparing the process of post-editing machine-translation output with the process of ordinary translation were reported. The researchers measured the ease of post-editing, translation speed, impact of post-editing on output rates, and willingness to post-edit machine translation. This experiment had the following conclusions (pp. 97-8):

1. An adept translator’s skills will probably be wasted on postediting.
2. The slower the translator, the greater the likelihood that his output can be increased by having him postedit machine translation.
3. Machine translation is not yet of such quality as to allow postediting to be done without a copy of the original in the hands of the translator.
4. Translators are apt to be rather mediocre typists.
5. Either translators do not consider their time and effort to be overly dear, or our respondents were exaggerating the time necessary to perform postediting, since half indicated their willingness to do the same work for less pay.

Almost half a century has passed since the publication of those results and conclusions; and in the recent decade post-editing has been subject to many studies. The European Society for Translation Studies Research Committee State-of-the-Art Research Report titled “Machine translation and postediting” by Temizöz (2012) gives a comprehensive and elaborate insight into this topic by summarizing research studies and papers on machine translation and post-editing.

Austermühl (2011) underlines the importance of post-editing as a form of revision in the translation process and explains differences between the definitions of “revision” by Mossop (“the process of checking a draft translation for errors and making appropriate amendments” (Mossop 2001: 169)) and Haar (“full scale revising may include major new sections of text or even a substantially new try at a document, while editing involves spelling, grammar, mechanics, word-usage, and other local concerns” (Mossop 2001: 16)). Austermühl (2011) suggests “Future research on the post-editing of MT outputs will need to indicate how much editing and revising (in Haar’s definition) is actually involved.” (p. 9).
Another study, comparing machine translation and human translation by Fiederer and O’Brien (2009) investigated the contribution of post-editing to the quality of translation. Based on ratings of 30 source sentences by qualified raters in terms of clarity, accuracy and style; the researchers found that “the machine translated, post-edited output [from English into German] was judged to be of higher clarity and accuracy, while the translations were judged to be of better style. When asked to pick their “favourite” sentence, the majority of the evaluators chose translated (as opposed to machine translated) sentences.” (p. 52)

A study by Bowker and Ehgoetz (2007) focused on user acceptance of machine translation output in terms of quality, time and cost for the French-English language pair. The researchers compared raw MT output, (rapid-) post-edited MT output and human translation of administrative memos. Three major findings of this study can be summarized as follows:

- raw MT output is completely unacceptable,
- although post-edited MT output can be deemed acceptable it still lacks stylistic richness of human translation, and
- post-edited MT can be a complementary to the translation workflow as it saves time and energy of the translators for long-lived texts.

In her master’s thesis, Guerra (2003) examined the time gain obtained by post-editing a machine-translated text fully rather than translating the same text from scratch without using MT for the English-Spanish language pair. The texts were marketing brochures and Prompt XT - a rule-based MT system was used. The researcher found that using MT as a basis for translation and then post-editing it took half of the time spent on human translation of the same texts, with no mention of the comparison of quality between two modes of translation.

As one of the most recent studies on MT post-editing, Péraldi (2012) tried to find a solution to accelerate the post-editing phase for translation students by developing a tool with an interface that facilitates certain post-editing acts such as marking non-translated words, suggestions of synonyms, possibility of relocating words, dialog boxes, list of choices accessible with one click, etc.). In this study, where the focus was on English-French-Spanish-German language pairs, it was argued that the post-editing phase can be optimized by developing tools through building vast corpora, enriching dictionaries, creating new syntactic rules and an accelerated revision of the MT errors, which in turn is likely to reduce the cost of MT post-editing in the translation process.

While a growing body of research on post-editing MT is becoming available in the translation field, it is paradoxical to see that this aspect of translation is rarely included in translation curriculum. About a decade ago, O’Brien (2002) proposed course content for post-editing and outlined the skill-set for post-editors. O’Brien (p. 100) argues that post-editing should be included in the translation curriculum because

- it would help meet the increasing demand for translation and for faster production times;
- post-editing skills are different from translation skills and we cannot assume that a qualified translator will be a successful post-editor;
- it would produce graduates who are already “comfortable” with postediting and who are more ready to be productive in a machine translation environment upon graduation;
- and it could improve the uptake of machine translation technology by improving translators’ perceptions of MT and its capabilities.
The current study focuses on the initial results out of an experimental course designed for novice translators of English-Turkish language pair. The primary aim of the course was to increase students’ translation competence through post-editing machine translation output for texts translated from English into Turkish using Google Translate. The lack of large-scale parallel text resources in English and Turkish was indicated as a problem of practical significance for English-Turkish statistical MT by Oflazer and Durgar El-Kahlout (2007). This is also a problem for Google Translate MT output as the amount of parallel texts for the English-Turkish language pair is relatively limited. The study also aimed at exploring students’ attitudes toward machine translation and the practice of post-editing. Comparison of students’ performance on post-editing translation tasks with their performance on actual translation tasks is another aspect investigated in the study.

2. Method

2.1 Participants

Participants were 15 fourth-year translation and interpretation students in a foundation-funded university in Turkey. All participants went through a traditional translation curriculum where students translate texts from English into Turkish at home and get feedback on their work in class. Students worked on texts from a variety of fields such as technical, legal, literary, financial, and media texts. All students took a course on translation technologies where they explored new tools and applications involved in translation practice. All students are native speakers of Turkish and they have at least intermediate level of proficiency in a language other than English. Before this course, participants did not have much post-editing experience and the main machine translation service that they knew of was Google Translate.

2.2 Course design and procedure

The course is titled “Subject-Specific Translation” and the major focus of the course was machine translation and post-editing machine translation. The aim was to raise novice translators’ awareness about machine translation systems and their place in translation practice and introduce them to the practice of post-editing. The ultimate goal was to contribute to the development of translation skills through post-editing.

The class session was three hours a week and it was held in a computer laboratory. Each student had a Google Mail account for the course. Texts were sent to students one week prior to the class session. Students were asked to complete a background log on the texts which involved:

- terminology research
- finding similar texts in Turkish
- previous translation experience on the text topic and text genre
- views about the language use
- expectations of MT

Background logs were composed using Google Drive and shared with the instructor. Source texts and MT outputs were sent to students and students post-edited the MT output in class using Microsoft Word with Track Changes feature activated. Students were allowed to use any kind of resources in the translation process. Students were asked to produce final products following the post-editing guidelines by TAUS “for achieving quality similar or equal to human translation” which were as follows:
During and after the post-editing, students completed a translation process log where they expressed their views about the MT output and the overall post-editing activity. In this log, students answered questions based on the TAUS guidelines for post-editing and revision guidelines outlined by Mossop (2001).

Students worked on five different texts and were then tested on three texts different from the previous ones. The texts were all authentic and not pre-edited. News articles on technology start-ups, health screening, cross-modal perception; a European Union press release and an informative text on bridge types were the texts covered in class. The texts used in the test were another European Union text (legal text), a text on instructions about changing an AC filter (technical text), and a news article on neuroscience (media text).

Students also used Google Translate for their B and C languages (respectively English and other commonly taught languages such as Spanish, French, Italian, German, Russian, and Japanese) and explored the potential of machine translation from and into these languages. Finally, students completed a survey on this overall experience and expressed their views about MT in general and post-editing. After the sessions on post-editing ended, traditional translation activities resumed.

### 2.3 Evaluation and Analysis

Students translations were evaluated by the instructor using a grid adapted from Orlando (2011) for process-oriented translation evaluation. The main categories of the evaluation grid were as follows:

- Overall comprehension of the ST
- Overall translation accuracy
- Omissions / Insertions
- Terminology / Word Choices
- Grammatical Choices / Syntactical Choices
- Spelling Errors
- Punctuation Errors
- Formatting Errors
- Appropriateness for Target Audience
- Readability / Idiomatic correctness
- Function / Completeness
- Style / Presentation / Genre

The surveys were administered using Google Forms. Survey data were analyzed using descriptive and inferential statistics in SPSS. Participants’ comments are included as qualitative data. All translations were evaluated by the course instructor.
3. Results and Discussion

The main focus of this study was to explore the use of MT post-editing in translator training and investigate novice translators’ attitude towards machine translation and post-editing MT output. To this end, participants were asked to post-edit MT outputs of three different texts: a legal text, a technical text, and a media text, each consisting of about 1000 characters. In the analysis were included three post-edited translations and three similar texts translated totally by the participants three weeks after the post-editing experience.

3.1 Translation performance

Using an evaluation grid adapted from Orlando (2011), post-edited translations and human translations were evaluated by the course instructor. For the post-edited translations, the average score of all participants for the technical text was the highest (78.53). It was followed by the media text (72.67) and the lowest average score was for the legal text (51.60). This is in line with the participants’ self-report on the difficulty level of each text (see Chart 1). For the human translations, again, the average score for the legal text was the lowest (57.80); and the average score for the media text (77.73) was slightly higher than the technical text (76.67).

Although students had full access to the Internet resources and were supposed to have an advanced level of proficiency in English, it would be safe to argue that the major factor that led to lower scores for the legal text for most of the participants seemed to be “lack of comprehension of the source text” rather than the bad MT output. The relatively higher scores for the technical text and the media text suggest that participants are likely to create “better translations” once they get a grasp of the source text well.

![Chart 1. Difficulty level of the legal text](image)

3.1.1 Comparison with human translation

While doing classroom research, it is often difficult to create a pure experimental setting, where variables are expected to be clear-cut and data fully reliable and valid. In this study, the data came from the midterm exam, where students were asked to post-edit machine translation outputs in a computer lab in a given time (about two hours) and from another session, where students translated similar texts from scratch in a similar amount of time.

The second session was held three weeks after the MT post-editing sessions. The legal text and the media was just another part of the document used for post-editing and the technical text was a very similar how-to article; all three were almost half as long as the ones used in
the post-editing task and a proportionate amount of time was allocated. The students were allowed to use paper dictionaries, as well as online resources. The translations were evaluated using the same grid based on Orlando (2011). Two of the students did not use any Internet resources. All of the students used bilingual (English-Turkish) dictionaries, whereas only three used monolingual (English) dictionaries. All participants were able to complete the translation task. The minimum amount of time used was 40 minutes and the maximum 100 minutes, with the average of 80 minutes. Time effect on the translation result was not measured in this experiment. When compared with MT post-editing results, human translations were not any more or less successful.

<table>
<thead>
<tr>
<th>SCORE AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT MEDIA</td>
</tr>
<tr>
<td>POST-EDIT MEDIA</td>
</tr>
<tr>
<td>HT TECHNICAL</td>
</tr>
<tr>
<td>POST-EDIT TECHNICAL</td>
</tr>
<tr>
<td>HT LEGAL</td>
</tr>
<tr>
<td>POST-EDIT LEGAL</td>
</tr>
</tbody>
</table>

**Chart 2. Translation scores for post-edited MT outputs (POST-EDIT) and human translations (HT)**

To determine whether the translation scores varied according to the medium of translation (post-editing or human translation) both Mann-Whitney U test and Student’s t-test were conducted with scores as the dependent variable and the medium of translation as the independent variable. The reason behind the decision to run both tests was that the sample size was small (N=15). An alpha level of .05 was used for all statistical tests.

As a post-hoc test, Wilcoxon signed rank test was conducted for each pair of tests for each text (i.e., comparing MT post-editing scores and human translation scores individually for legal, technical, and media texts). The results of those tests suggested that there was no statistically significant difference between the scores for all three texts based on the translation method. Using MT post-editing in the translation process did not elicit a statistically significant change in students’ performance for legal texts ($Z = -1.052, P = 0.293$), technical texts ($Z = -0.654, P = 0.513$), or media texts ($Z = -1.726, P = 0.084$).

All three tests showed that there was no statistically significant difference between the participants’ performance in post-editing MT output and translating from scratch for all three texts (i.e. legal, technical, and media texts). In addition, participants’ average scores for all three texts were compared across two methods: post-editing and human translation. Again, no statistically significant difference was observed.

### 3.1.2 Logs

Another important aspect of this experiment was to encourage students to keep logs for their translation activities to focus their attention on the process of translation as well as the product.
For each translation activity, participants answered a set of questions before, during and after the translation work.

### 3.1.2.1 Background Logs

The participants completed a background log for each text before they started translating the texts. The questions were as follows:

- How familiar are you with the topic of the text?
- What is the difficulty level of the text for you?
- What is the level of your experience with such text type(s) and text genre both as a reader and as a translator?
- What can you say about the language used in this text?
- Are there similar texts available in Turkish? If so, how can those texts help you in your translation?
- What is the level of quality that you expect from the machine translation for this text? Explain the reasons.
- What kind of resources have you used for understanding the text? List all of them.
- How much time did you spend for this background work?

The participant spent from 5 to 30 minutes completing a background log for each text and their answers to the questions listed above ranged from a couple of words to two paragraphs. An important contribution of integrating such a component to the translation process is the fact that novice translators get to know the text in hand better by reflecting on its topic, genre, and language. Although translation students are always advised to conduct background work before starting their translations in all classes, they tend to start translating texts at once. In their answers to the survey questions, twelve out of fifteen participants stated that doing background work before post-editing MT made the post-editing process quicker and easier and ten out of fifteen that finding internet resources about the topic of the text helped them understand and post-edit the MT output better. This finding is important especially considering the initial resistance from the participants against this new practice of keeping logs.

### 3.1.2.2 Translation Process Logs

In translation process logs, students were asked to answer the questions presented in Table 1 for each text. The questions were based on Mossop (2001). In Table 1, the evaluation of each text by the course instructor is presented so that the overall performance of MT for the English-Turkish language pair can be reflected.

Although these questions do not reflect the whole of the translation process, they attempted to channel participants’ attention to the “success” of the MT so that they could identify which parts of the MT output they should focus on. The analysis of the raw MT output shows that the findings are similar to Bowker and Ehgoetz’s (2007) in that it is completely unacceptable. Considering the fact that the amount of parallel text resources in English and Turkish is much less than the English-French language pair, the raw MT outputs in the current study needed extensive post-editing to be of publishable quality, which may have caused negative attitude towards MT and the act of post-editing in participants’ view.

From the analysis of the translation process logs, completed by the participants during and after the post-editing session; it is clear that most of the participants were able to identify the overall quality of the MT output for each text. About half of the participants gave superficial answers to questions, whereas the rest supported their observations with examples from the texts, indicating specific shortcomings of the MT system. This process was highly likely to lead to better post-editing results. However, considering the difficulty levels of the texts reported by the participants, the quality level of the post-edited MT outputs was in parallel with the actual
MT output for most of the students. In other words, the more difficult and the more complex the source texts were, the worse the MT outputs and post-editing results were likely to be obtained. It should also be added that overall translation competence, language skills, level of familiarity with the text type and genre, and other psychological and physiological factors may all have an impact on the post-editing results and these factors are not within the scope of the current study.

In their answers to the survey questions, six of the participants reported that keeping a translation process log helped them reflect on their decisions for post-editing whereas six others were neutral and three did not find the log helpful. On the other hand, only two of the participants disagreed with the proposition that keeping translation process logs helped them become a more conscious translator, whereas four of them were neutral.
3.1.3 Participants’ Reaction to MT and Post-Editing

At the beginning of the course (experiment), students were required to do readings on machine translation and post-editing. These topics were also discussed in class and the emergence and importance of MT post-editing was supported and highlighted with references to these resources. However, in the first weeks of the experiment, participants expressed their reluctance toward and discomfort in post-editing MT outputs as a translation exercise.

According to the results of the survey administered at the end of the experiment, almost all participants reported that they found MT post-editing a difficult and an irritating task. In their answer to an open-ended question; participants reported the following as the most negative things about the MT output:

- it created confusion,
- it was “meaningless” most of the time,
- it generated wrong word order and wrong word choices,
- it took a lot of time to post-edit,
- it lacked subject-verb agreement

The most positive things about the MT output were as follows:

- it reflected the overall message of the text
- some of the word choices matched
- some segments were translated correctly
- it saved time for some texts

More than two thirds of the participants believed that they could improve their translation skills better in a traditional translation class where they translate texts themselves without the “help” of the MT system. Again, the same percentage of participants preferred translating texts from scratch rather than post-editing MT output (see Chart 3). Only two of the participants agreed that machine translation can ease the burden of translators (see Chart 4). The overall negative attitude toward MT did not seem to have resulted from the fact that the participants considered MT a threat to their profession, because only two of them reported having this opinion in the survey. The relatively “bad” output for the English-Turkish language pair might have contributed to this line of thought.
The participants were also asked about the potential place of machine translation and post-editing in the translation curriculum. Six of the participants agreed with the proposition that having an idea of the impact of MT systems on translation and of their actual possibilities and potential should be a part of the professional background of present day translators, whereas six others disagreed. One third of the participants disagreed with the proposition that translators should know how to use MT effectively for their professional development, whereas eight agreed. Nine of the participants believed that post-editing is a skill that should be included in translation education (Chart 5); whereas only six participants thought that post-editing is a skill that translators should have in order to survive in today’s translation market. One third of the participants were neutral about the proposition that post-editing machine translation output helped them improve their translation skills whereas seven of them disagreed.

It is important to note here that when traditional translation activities resumed after MT post-editing sessions ended, some of the participants, who had shown a certain degree of reluctance towards doing post-editing activities during the experiment, verbally reported that “they would rather post-edit” the new translation assignments. Their major argument was that they got used to working on MT output and it was taking more time to complete translations from scratch. This is likely to suggest that the more MT output students post-edit, the more comfortable they get with the process.
3.1.4 MT experience with B and C languages

As pointed out on Google Translate website, “[s]ince the translations are generated by machines, not all translation will be perfect. The more human-translated documents that Google Translate can analyze in a specific language, the better the translation quality will be. This is why translation accuracy will sometimes vary across languages.” Therefore, it would not be surprising to see better results between English and FIGS (French, Italian, German, and Spanish) languages.

In order to introduce the potential of machine translation for these languages, the participants were asked to experiment with translating texts between their B language (English) and C language (the language they are studying at the university throughout their undergraduate years). All participants agreed that the MT output was much better than the English-Turkish output and except for one participant (who was studying Russian), all participants reported that they would or were neutral about using MT for their translation work between their B and C languages (German, Italian, French, Spanish, and Japanese) (see Chart 6). This finding suggests that the novice translators might take a more positive stance toward MT in general and the practice of post-editing when they get better MT results.

![Chart 6. Willingness to use MT for B and C languages](image)

4. Conclusion

The current study investigated the potential of using MT post-editing activity in the translation curriculum. Fifteen fourth-year translation students took part in post-editing activities for five weeks and were then tested on their post-editing performance for three different texts: legal, technical and media texts. Students were also asked to complete background and process logs during the post-editing process. Their reaction to the overall experience and to the use of machine translation in general was explored through survey questions. Students also completed translation tasks where they translated three similar texts on their own without using any machine translation system.

One of the important findings of the study was that most of the students, who went through product-oriented translation activities during their previous courses before this experimental course, benefited from using background and translation process logs. This kind of activity seemed to help them do extensive research on the topic of the translation and focus on the language by analyzing MT output. However, the lack of comprehension of the actual source text seemed to impede students from producing quality translations through post-editing MT output.

Another aspect of the study was exploring novice translators’ reaction to the act of post-editing and to machine translation in general. The data showed that the post-editing experience
was difficult and irritating for most of the students because of the low quality of the raw MT output for most of the texts. Nevertheless, as the participants got used to the practice of post-editing, the level of negativity towards this activity decreased. The change in students’ reaction was obvious when they were given traditional translation assignments: some of them openly stated that they would rather use MT and post-edit in order to gain time by not dealing with terminology and by using correctly translated segments.

Finally, the comparison of students’ performance in post-editing and translation activities showed that there is no statistically significant difference between the two activities. In other words, students’ translations were not any better when they translated similar texts from scratch or compared with post-editing MT output. The most important factor that seemed to have determined the quality of the post-editing product or the translation product was the perceived difficulty level of the text for the students.

MT post-editing activities can and should be part of the translation curriculum so that novice translators get more familiar with this new type of work demand and they become more aware of the potential benefits and drawbacks of using machine translation in the translation process. Using process-oriented components such as keeping background logs and translation process logs is likely to increase novice translators’ consciousness level in their translation decisions. Also, based on the overall observations of the instructor; the ability to identify the problematic aspects of the raw MT output seems to lead to better translation performances. In line with this observation, it can be argued that using MT as a bad model has a potential to be a productive activity in translation courses.

4.1 Limitations of the current study and suggestions for further research

The current study has multiple limitations. First of all, the sample size was small (n=15) and the translation products were evaluated by only one person, the course instructor. Due to technical limitations; variables such as time allotted for each translation task and academic achievement of the participants were not included in the data analyses. Other aspects such as the amount of post-editing for each task, keyboard tracking, the amount of time spent using resources are also absent in the study due to technical limitations, as the current study was completed in the classroom setting and the instructor had to fulfill curriculum requirements at the same time. For future research, the abovementioned components can be included in order to present a clearer picture and further evidence in terms of the advantages and disadvantages of using post-editing in the translation process.

Bibliography


