Undergraduate Translation Students' Performance and Attitude vis-à-vis Machine Translation and Post-editing: Does Training Play a Role?

Maria Stasimioti

Department of Foreign Languages, Translation and Interpreting
Ionian University

Corfu, Greece

stasimioti@ionio.gr

Vilelmini Sosoni

Department of Foreign Languages, Translation and Interpreting
Ionian University

Corfu, Greece

sosoni@ionio.gr

Abstract

In an effort to meet the demands in speed, productivity and low-cost, the translation industry has turned to Machine Translation (MT) and Post-editing (PE). Nowadays, MT output is used as raw translation to be further post-edited by a translator (Lommel and DePalma, 2016). Yet, translators still approach PE with caution and scepticism and question its real benefits (Koponen 2012; Gaspari et al 2014; Moorkens 2018). In addition, attitudes to MT and PE seem to affect PE effort and performance (Witczak, 2016; Çetiner and İşisağ, 2019). Under that light, this study aims to investigate the attitudes and perceptions of undergraduate translation students towards MT and PE and their performance before and after they receive training in MT and PE. Questionnaires are used to capture their attitudes and perceptions, a calculation of the technical effort and the temporal effort expended by the students while post-editing is also used, while a human evaluation of the post-edited output is carried out to assess their performance and the quality of the post-edited texts. The analysis reveals a change in the students' attitudes and perceptions; they report a more positive attitude toward MT and PE, they are more confident and faster, while they avoid over-editing.

1   Introduction

In recent years, the production of content in multiple languages has become one of the most significant aspects of communicating information, a fact that requires high levels of speed and productivity in translation services, i.e. high levels of efficiency (Bowker, 2005; Guerberof, 2009; Federico, Cattelan and Trombetti, 2012). In an effort to meet this demand, while keeping the cost low, the translation industry has turned to new translation technologies, including Machine Translation (MT) and Post-editing (PE), which were boosted by rapid technological advances – most notably the recent advent of neural machine translation (NMT) models. In fact, nowadays, it is common practice to include MT in the translation workflow by using machine translated text as raw translation to be further post-edited by a translator (Gaspari *et al.*, 2015; Lommel and DePalma, 2016). As a result, PE has become a generally accepted, separate part of the translational landscape. Yet, translators still approach PE with caution and scepticism and question its real benefits (Koponen 2012; Gaspari et al 2014; Moorkens 2018). Their scepticism is directly related to the nature of PE which has been found to differ from human translation as well as revision of human translation not only in terms of the practical goals set and the processes employed but also in terms of the cognitive processes employed (Krings, 2001; O'Brien, 2002). This is mainly explained by the fact that PE involves "working by correction rather than creation" (Wagner 1985: 2). In addition, translators have the perception that PE is slower than translating from scratch and view MT as a threat to their profession (Moorkens 2018: 58).

Under that light, several studies were carried out aiming at identifying the extent to which attitudes to MT and PE affect PE effort and performance (Ling *et al.*, 2016; Witczak, 2016; Çetiner and İşisağ, 2019). In particular, a positive attitude to MT has been found to be a factor in PE performance (de Almeida 2013; Mitchell 2015). In addition, experienced translators have been found to exhibit rather negative attitudes to PE as opposed to novice translators (Moorkens and O'Brien 2015) and to be rather reluctant to take on PE jobs, while novice translators appeared to be more positive towards MT and PE and more suited for PE jobs (Garcia 2010, Yamada 2015). Within that framework, the role of training is particularly important, as it may affect MT and PE perception and translators' performance. Several studies, in fact, have shown that the attitudes of students changed positively after training in MT and PE (Alotaibi, 2014; Koponen, 2015; Çetiner and İşisağ, 2019; Guerberof and Moorkens, 2019).

The present study adopts a mixed-methods approach and triangulates findings from different methods in order to investigate the attitudes and perceptions of undergraduate translation students towards MT and PE as well as their PE performance before they receive training (Stage 1) and after they receive training in MT and PE (Stage 2). In particular, students are asked to fill in questionnaires regarding their attitudes and perceptions before and after the receive training; they are also asked to carry out PE of NMT output generated by the NMT system developed by Google (Google Translate NMT system), before and after the training while measuring their keystrokes, i.e. technical effort, and time spent, i.e. temporal effort. Finally, a human evaluation of their post-edited MT output is used to assess their performance and the quality of the final post-edited texts both before and after they receive training.

2   Methodology

The investigation of the perception vis-à-vis MT and PE as well as the PE performance draws on data gathered at the Department of Foreign Languages, Translation and Interpreting of the Ionian University during the 2018-2019 Spring Semester. In particular, in the context of the compulsory module "Translation Tools" which is taught in the 4th year of the BA in Translation Studies, PE training was offered consisting of five two-hour lectures, four independent assignments and several in-class exercises. It is important to note that the particular module is the only module in Translation and Technology included in the curriculum. This is in line with the rationale and the objectives of the programme which seeks to help students develop a comprehensive understanding of the theories and practice of translation, to acquire high foreign language skills, interpersonal communication skills and a deep knowledge about a variety of cultures, to cultivate advanced critical thinking skills and be able to translate for diverse audiences in a variety of contexts and genres and make optimal use of the latest translation technology. Translation technology is thus important, but more important is an all-round translation education which requires students to first learn how to approach a text holistically. It is in that spirit that translation technology is introduced during the 4th year of their studies. Participation in the study was optional for the students. Of the 168 students registered, 90 chose to participate. The latter signed a detailed consent form prior to the execution of the PE task, while all stored data were fully anonymized in accordance with Greek Law 2472/97 (as amended by Laws 3783/2009, 3917/2011 and 4070/2012). In the sub-sections that follow, information is provided about the participants and the PE training they received as well as the methods used to investigate i) their attitudes towards MT and PE and ii) their performance.

2.1 Participants

As can be seen in Table 1, almost all of the participants, i.e. 98%, belonged to the 18-24 age group and were studying for their first degree (99%), while the majority, i.e. 92%, were female.

Gender	Female	92%
	Male	6%
	Other	2%
Age group distribution	18-24	98%
	25-34	2%
Education level	High School leaving certificate	99%
	Undergraduate degree in a field other than Translation	1%

Table 1. Participants' demographics

2.2 PE training

The training offered in the context of the compulsory module "Translation Tools" aimed to introduce students to MT and PE as well as to the recent developments in the respective fields. Upon completion of the training, students were expected, among others, to be able to (i) use MT during the pre-translation process, (ii) evaluate MT output using both automatic and human evaluation metrics and (iii) post-edit MT output according to the expected level of quality (full/light PE).

To that end, the topics covered included, among others, the theory and history of MT and PE, the basic principles of MT technology, analysis of the dominant systems in the market (PBMT, SMT, Hybrid, NMT), the importance of controlled language and pre-editing for MT, quality metrics and evaluation of MT output, PE levels of quality, PE effort and productivity (temporal, technical and cognitive effort), MT output error identification, MT engine implementation in the translation workflow and post-editor profile and associated skills (O'Brien, 2002; Depraetere, 2010; Doherty et al, 2012; Pym, 2013; Doherty and Kenny, 2014; Kenny and Doherty, 2014; Koponen, 2015; Guerberof and Moorkens, 2019). As already mentioned, the training included four independent assignments: (i) Comparison of different MT systems (SMT, Google NMT, Neural NMT), (ii) MT quality evaluation using human metrics, i.e. adequacy and fluency evaluation (Snover *et al.*, 2009) and automatic metrics (e.g. BLEU score), (iii) Light PE and measuring PE time and (iv) Full PE and measuring PE time.

2.3 Questionnaires

As pointed out, for the aims of this study, the students had to fill in two questionnaires: one questionnaire prior to the training (pre-training questionnaire) and one questionnaire following the training (post-training questionnaire). The pre-training questionnaire consisted of 21 questions: 18 closed-ended questions (multiple choice) and 3 open-ended questions, all of which aimed at defining the participants' profile and capturing their perception and attitudes vis-à-vis MT and PE as well as their opinion about the importance of PE training prior to the training. The post-training questionnaire consisted of 15 questions: 10 closed-ended questions (multiple choice) and 5 open-ended questions, all of which aimed at receiving feedback on the PE tasks performed, at capturing their perception vis-à-vis MT and PE as well as their opinion about the importance of PE training after the training.

2.4   Process based approach: keystrokes and task duration

The participants were asked to carry out PE tasks while the temporal effort (total task time) and the technical effort (keystrokes: insertions and deletions) expended were registered using the Translog-II software (Carl 2012). The participants were asked to carry out the tasks at the speed at which they would normally work in their everyday work as translators; therefore, no time constraint was imposed [1].

The experiment consisted of one session per participant per stage, i.e. Stage 1 and Stage 2. Before the sessions, a group meeting was organised during which the participants were informed about the nature of the experiments, the task requirements and the general as well as task-specific guidelines they had to follow. In particular, during Stage 1 and given that the participants had not received any training in MT and PE, they were asked to correct the NMT output with a view to creating a final text of publishable quality. During Stage 2 and after they had received training in MT and PE, they were asked to carry out full PE of the NMT output according to the respective guidelines, i.e retain as much raw MT translation/output as possible, transfer the message accurately, fix any omissions and/or additions (at the level of sentence, phrase or word), correct mistranslations, correct morphological errors, correct misspellings and typos, fix incorrect punctuation if it interferes with the intended message, correct erroneous terminology, fix inconsistent use of terms and do not introduce stylistic changes. The session during Stage 1 began with a warm-up PE task in order to familiarise each participant with the procedure. No warm-up PE task was required during Stage 2, as the participants were already familiar with the procedure. The actual experimental task involved the full PE of the NMT output of two semi-specialised texts during Stage 1 and the full PE of the NMT output of two different but comparable semi-specialised texts during Stage 2 (see section 2.4.1). During the experiment, the ST was displayed in the Translog-II software at the top half of the screen and the MT output at the bottom half, as suggested by previous studies (Hvelplund 2011, Carl et al 2011, Mesa-Lao 2014, Carl et al 2015). The participants worked directly on the MT output.

2.4.1   The texts

The source texts (STs) used in this study were short (~140-word) educational semi-specialised texts from the domain of Psychology selected from a public digital library of open educational resources. They all had comparable Lexile® scores (between 1200L and 1300L), i.e. they were suitable for 11th/12th graders (see Table 2). Two texts (Texts 1 and 2) were used for PE during Stage 1, two texts (Texts 3 and 4) were used for PE during Stage 2 and one text was used exclusively for the warm-up session during Stage 1 and is not included in the ensuing analysis and discussion.

---

[1] The study also involved the collection of eye-tracking data – the analysis of which is not presented here. For that reason, the participants did not have Internet access and were not allowed to use online or offline translation aids as these could lead to a reduction in the amount of recorded eye-tracking data.

	Text 1	Text 2	Text 3	Text 4
Lexile® Measure	1200L-1300L	1200L-1300L	1200L-1300L	1200L-1300L
Number of sentences	7	7	7	7
Mean sentence length	20.86	21.00	20.86	21.14
Word count	146	147	146	148

Table 2. The source texts

The NMT-core engine used to produce the raw MT output was Google Translate (output obtained March, 2019). The NMT output was evaluated with the BLEU metric using the Tilde Custom Machine Translation toolkit[2]. All texts had a very good score, ranging from 51.23 for Text 2 to 62.21 for Text 4, which means that PE could be used to achieve publishable translation quality in all cases.

2.5 Product based approach: fine grained human error analysis for quality evaluation

For our product-based analysis and similarly to previous studies (Carl and Buch-Kromman 2010, Moorkens and O'Brien 2015, Koponen and Salmi 2017, Koponen Salmi & Nikulin 2019), we used an automatic metric, i.e. Human-targeted Translation Edit Rate (HTER) (Snover *et al.*, 2006), which calculates the edits performed to measure the distance between the MT output and its post-edited version. However, the edits do not always reflect actual errors (Koponen Salmi & Nikulin 2019). Previous studies (de Almeida 2013, Koponen & Salmi 2017, Koponen Salmi & Nikulin 2019) have shown that post-editors either over-edit the MT output making preferential choices or they under-edit it leaving errors uncorrected, while sometimes they also introduce new errors. For that reason, following the calculation of HTER, each edit operation was annotated manually by one annotator –a professional translator with 10 years of translation experience– with one of the following categories suggested by Koponen & Salmi (2017) and Koponen Salmi & Nikulin (2019):

- unedited: no change;
- form changed: different morphological form;
- word changed: different lemma;
- deleted: word removed;
- inserted: word added;
- order: position of a word changed.

It should be noted that similarly to Koponen & Salmi (2017) and Koponen Salmi & Nikulin (2019) in those cases where a word has been affected by more than one edit type, it was annotated with both categories. For example, some words had both their morphological form and their position changed (form + order) or a different lemma was used and its position was also changed (word + order).

Each word-level edit was then assessed for correctness of meaning (accuracy) and language (fluency) as well as for necessity, i.e. for establishing whether the edits were necessary to correct the meaning and/or the language or whether they were preferential edits in terms of style or word

---

[2] https://www.letsmt.eu/Bleu.aspx

choice. According to Koponen and Salmi (2017) and Koponen et al (2019), each edit could be either correct or incorrect and necessary or unnecessary. However, in some cases no edit may be required in the MT output, while in some other cases post-editors may leave errors uncorrected or make preferential changes. Therefore, we decided in our study to cater for such cases by adding additional options. In particular, as far as correctness is concerned, we added the "correct no edit" option, which means that the post-editor was right to leave a segment unedited given that there was no error in the MT output; we also added the "edit missing" option for cases when an error in the MT was not corrected by the post-editor; finally, we added the "redundant edit" option, which indicates a preferential change made by the post-editor. As far as necessity is concerned, we added the "necessary no edit" option, which means there was no error in the MT output and therefore no edit was required, and the "edit required" option, which means that there was an error in the MT which was not corrected by the post-editor. Summarizing the above, the options for correctness and necessity are the following:

Correctness

➔ correct no edit
➔ correct edit
➔ incorrect edit
➔ edit missing
➔ redundant edit

Necessity

➔ necessary no edit
➔ necessary edit
➔ unnecessary edit
➔ edit required

3 Findings and discussion

3.1 Perception analysis: Perception of MT before and after the training

During Stage 1, apart from a specific question on their perception of MT, the participants had to answer questions regarding the use of CAT tools and Google Translate, in order to establish to what extent they are familiar with translation technology. With respect to the use of CAT tools, half of the participants (51%) said that they do not use any CAT tools, while the rest noted that they use Matecat (43%), SmartCAT (11%), Memsource (6%) or WordFast (1%). Although the majority (74%) use Google Translate in their translation work, most of them said that they either rarely (42%) or sometimes (22%) use it. Only 10% said that they use it often (9%) or always (1%). This can be easily explained if we take into account their response as regards the quality of Google Translate. Only 6% find the quality of Google Translate good, while none of them finds it excellent. 7% find it very poor, 44% poor and 43% fair.

As regards their perception of MT before receiving training, the majority of the participants (74%) said that MT is a useful/helpful tool. 27% consider it to be necessary, while 9% observed that it has made significant progress in the past years. However, 26% approach it with caution and mistrust due to its shortcomings at the level of fluency and at the level of accuracy. After receiving training, a significant change in the participants' perception of MT was observed. In particular, 92% noted that their perception of MT has been positively affected. 3% noted that their perception of MT has been negatively affected, while 5% were not sure. As it emerges from Figure 1, slightly

more students (77% as opposed to 74% in Stage 1) thought that MT is a useful tool. More importantly, 39% noted that MT saves them time and increases their productivity as opposed to 2% in Stage 1. In addition, 8% as opposed to 0% in Stage 1 said that it is accurate, more fluent and more trustworthy than what they initially thought and acknowledged the improvement of MT quality at the level of fluency and at the level of accuracy. It is important to mention that after receiving training only 4% approach MT with caution and mistrust, as opposed to 26% in Stage 1. Another interesting finding is that although in Stage 1, only 1% of the students thought that MT is far from achieving human parity, in Stage 2 this percentage climbed to 9%.
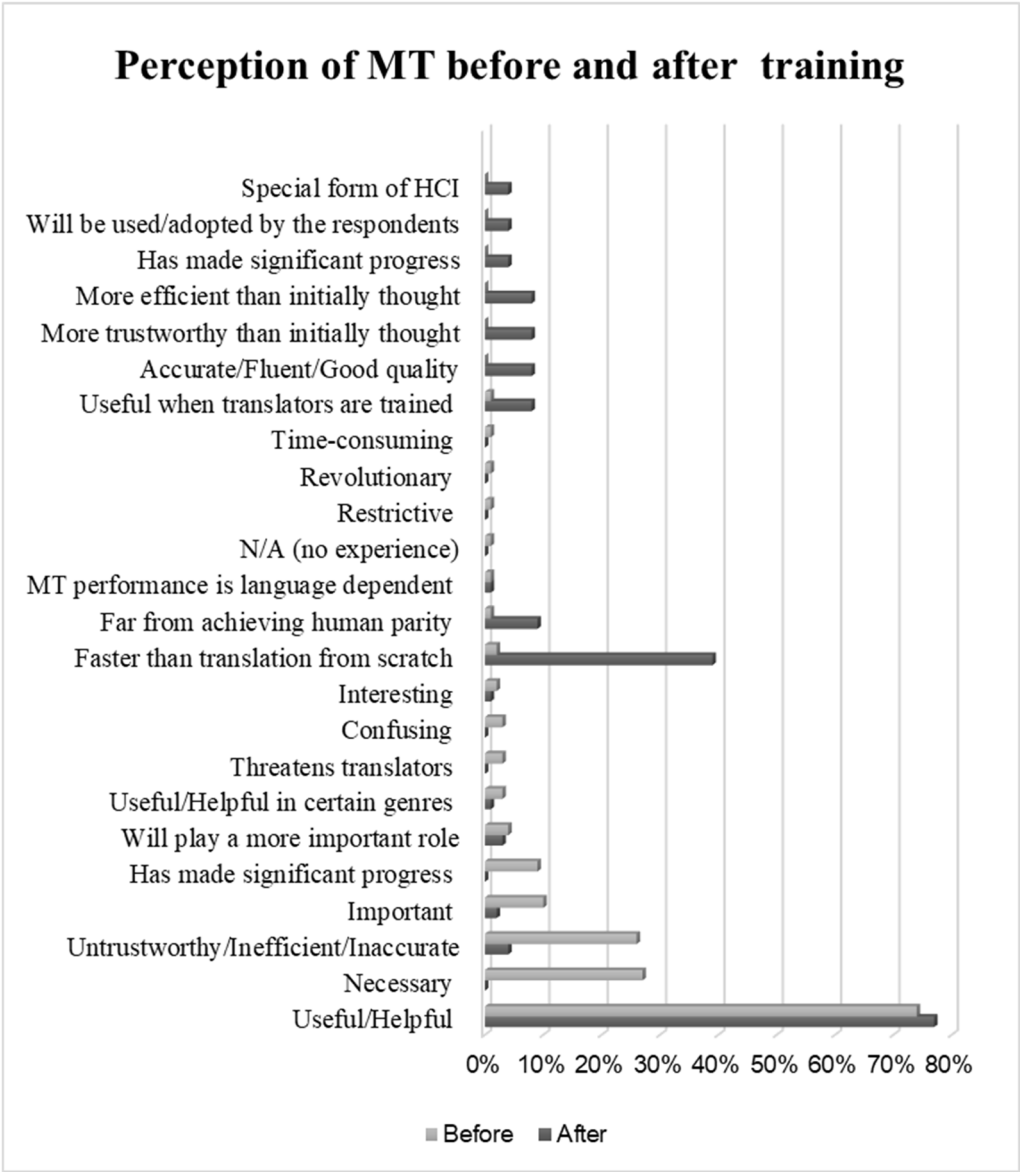

Figure 1. Perception of MT after receiving training

3.2 Perception analysis: Perception of PE before and after the training

In a similar vein and as regards their perception of PE before receiving training, the majority of the participants (62%) pointed out that PE is necessary because MT cannot be used without human editing, while 20% pointed out that it is useful and also faster than translation from scratch. As it emerges from Figure 2 and like in the case of MT, their perception of PE after receiving training has changed and has been positively affected for the majority of the participants (96%). For only 1% the perception of PE has been negatively affected, while 3% could not decide. The percentage

of the participants who found PE to be faster than translation from scratch and useful for their work increased dramatically, 44% in Stage 2 as opposed to 17% in Stage 1 and 32% in Stage 2 as opposed to 19% in Stage 1. Having seen first-hand the improvement of the MT system, less than 30% found PE to be mandatory, as opposed to 60% in Stage 1. After receiving training, 20% of the participants, as opposed to 2% in the pre-training stage, recognised the importance of PE training, while 12% stated that they were prepared to perform PE as opposed to 0% in Stage 1. Finally, 7% found PE to be easier than they initially thought.
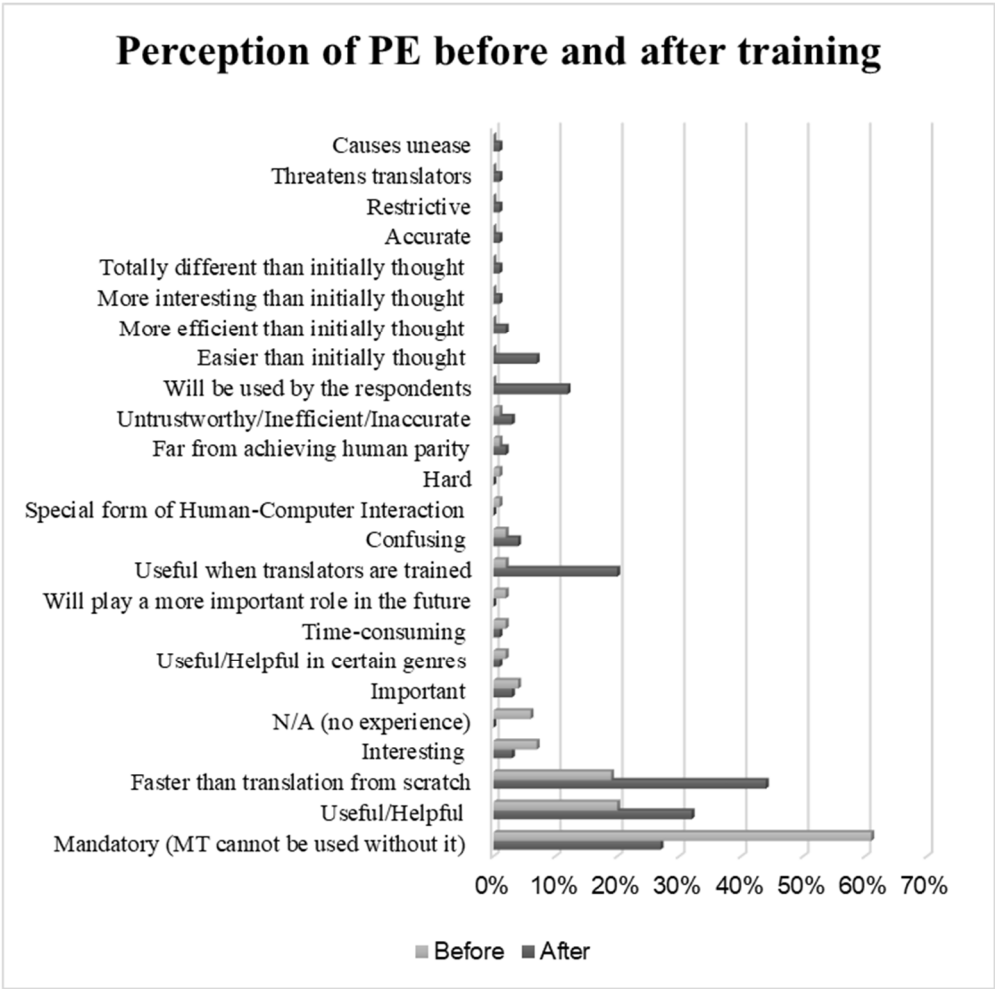

Figure 2. Perception of PE after receiving training

The importance of PE training during Stage 2 is highlighted not only in the answer to the open-ended question about their perception of PE, but also in a closed-ended question, where they noted that PE training is extremely important (70%), very important (27%) or moderately important (3%). None of them found it slightly important or not at all important. It should be noted that the respective percentages before the training where 56% for extremely important, 36% for very important, 7% for moderately important and 2% for not at all important.

In addition, almost all the students (99%) said that the PE training helped them carry out the tasks more efficiently (63%), i.e. with greater accuracy, more easily (62%), as they knew what errors to anticipate, and more quickly (58%).

3.3 Process based approach: keystrokes and task duration

As already mentioned, apart from the investigation of the effect that PE training may have on undergraduate translation students' perception of MT and PE, the study also aims at the investigation of the effect that it may have on their performance when carrying out PE tasks as well as on the final post-edited text. For that reason, we measured the participants' technical effort (keystrokes) and temporal effort (task duration), during both Stages (Stage 1 and Stage 2), to assess their performance and we also evaluated their post-edited text to assess their quality both before and after they receive training.

3.3.1 Task duration (temporal effort)

As far as the temporal effort is concerned, we measured the average time (in minutes) the students needed to post-edit the NMT output during Stage 1, on the one hand, and during Stage 2, on the other hand. As it emerges from Table 4, the students needed less time to post-edit the NMT output during Stage 2 ($M = 8.09$, $SD = 2.61$) compared to Stage 1 ($M = 15.05$, $SD = 4.50$). According to a t-test, that difference in average task time between Stage 1 and Stage 2 is statistically significant $t(179) = 17.75$, $p < 0.05$.

Stage	Task duration (in mins)	
	Mean	Std. Deviation
Stage 1	15.05	4.50
Stage 2	8.09	2.61

Table 3. Temporal effort per Stage: Mean and standard deviation values of the task duration

3.3.2 Keystrokes (technical effort)

As far as the technical effort is concerned, we measured the average number of keystrokes (insertions and deletions) the students performed when post-editing the NMT output during each stage. As it emerges from Table 4, the students performed less keystrokes during Stage 2 ($M = 276$, $SD = 178$) compared to Stage 1 ($M = 749$, $SD = 271$). Similarly to temporal effort, a statistically significant difference $t(179) = 14.98$, $p < 0.05$ was reported.

Stage	Total number of keystrokes	
	Mean	Std. Deviation
Stage 1	749	271
Stage 2	276	179

Table 4. Technical effort per Stage: Mean and standard deviation values of the total number of keystrokes (insertions and deletions)

3.4 Product based approach: fine grained human error analysis for quality evaluation

As emerges from Table 5, PE training affects not only the perception and the actual performance of the undergraduate translation students, but also the quality of the final post-edited texts. Both during Stage 1 and Stage 2 the students performed equally well while carrying out the PE task

either by making correct and necessary edits or by leaving the MT output as is, in cases where no edit was necessary (correct and necessary no edit). The main difference we noticed was the lower percentage of redundant and unnecessary edits during Stage 2. This means that after receiving training the students were able to understand the requirements of the PE task, to recognise the errors in the MT output more easily and to follow the specific PE guidelines, while performing only the necessary edits without interfering with the quality of the final post-edited text.

	Stage 1	Stage 2
correct and necessary edit	32.49%	14.27%
correct and necessary no edit	38.12%	70.35%
incorrect and necessary edit	8.93%	4.37%
incorrect and unnecessary edit	1.70%	2.20%
edit missing and required	5.17%	4.87%
redundant and unnecessary edit	**11.93%**	**3.34%**

Table 5. Error analysis per Stage

4 Conclusion and future work

In line with previous studies, the findings of the questionnaire analysis reveal a change in the students' attitudes and perceptions after their training in MT and PE. In particular, the students reported a more positive attitude towards MT and PE; they were also more confident and faster and they avoided over-editing. It thus appears that training is crucial if MT is to be accepted and adopted by translators and if PE is to produce target texts with human-like quality.

There are a number of limitations to this study. First, the students could not use any resources during the PE tasks, i.e. they could not use online or offline resources, such as dictionaries, termbases, parallel texts, etc. In addition, they were asked to carry out the tasks in an environment that differed from their usual work environment, while the error analysis was carried out by only one annotator and the sample size was small. These limitations will be addressed in future research.

References

Alotaibi, Hind. 2014. Teaching CAT Tools to Translation Students: An Examination of Their Expectations and Attitudes. *Arab World English Journal*, 3(1): 65–74.

Bowker, Lynne. 2005. Productivity vs quality: A pilot study on the impact of translation memory systems. *Localisation Focus*, 4(1):13–20.

Carl, Michael. 2012. Translog-II: A program for recording user activity data for empirical reading and writing research. In *LREC*, pages 4108–4112.

Carl, Michael, and Matthias Buch-Kromann. 2010. Correlating Translation Product and Translation Process Data of Professional and Student Translators. In François Yvon and Viggo Hansen (eds) (2010). EAMT 2010: Proceedings of the 14th annual conference of the European association for machine translation, 27-28 May 2010 Saint-Raphaël

Congrès, Saint-Raphaël, France, European Association for Machine Translation. http://www.mt-archive.info/EAMT-2010-Carl.pdf [last accessed September 22, 2019]

Carl, Michael, Barbara Dragsted, Jacob Elming, Daniel Hardt, and Arnt Lykke Jakobsen. 2011. The process of post-editing: A pilot study. *Copenhagen Studies in Language,* 41: 131–142.

Carl, Michael, Silke Gutermuth, and Silvia Hansen-Schirra. 2015. Post-editing machine translation: Efficiency, strategies, and revision processes in professional translation settings. In Aline Ferreira and John W. Schwieter (eds), *Psycholinguistic and cognitive inquiries in translation and interpretation studies*. Cambridge Scholars Publishing, Newcastle upon Tyne, pages 145–174.

Çetiner Caner, and Korkut Uluç İşisağ. 2019. Undergraduate Level Translation Students' Attitudes towards Machine Translation Post-Editing Training. *International Journal of Languages' Education and Teaching*, 7(1), 110–120. Doi: 10.18298/ijlet.3242.

de Almeida, Giselle. 2013. *Translating the post-editor: An investigation of post-editing changes and correlations with professional experience.* PhD Thesis, Dublin City University, Dublin.

Depraetere, Ilse. 2010. What counts as useful advice in a university post-editing training context? Report on a case study. In François Yvon and Viggo Hansen (eds) (2010). *EAMT 2010: Proceedings of the 14th annual conference of the European association for machine translation*, 27-28 May 2010 Saint-Raphaël Congrès, Saint-Raphaël, France, European Association for Machine Translation. http://www.mt-archive.info/EAMT-2010-Depraetere-2.pdf [last accessed September 22, 2019]

Doherty, Stephen, Kenny, Dorothy, and Andy Way. 2012. Taking statistical machine translation to the student translator. In *Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*, 28 Oct – 1 Nov 2012, San Diego, USA. http://doras.dcu.ie/17669/1/AMTA-2012-Doherty-1.pdf [last accessed September 22, 2019]

Doherty, Stephen, and Dorothy Kenny. 2014. The design and evaluation of a statistical machine translation syllabus for translation students. *The Interpreter and Translator Trainer*, 8(2): 295-315.

Federico, Marcello, Alessandro Cattelan, and Marco Trombetti. 2012. Measuring user productivity in machine translation enhanced computer assisted translation. In *Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*, 28 Oct – 1 Nov 2012, San Diego, USA. http://www.mt-archive.info/AMTA-2012-Federico.pdf [last accessed September 22, 2019]

Gaspari, Federico, Antonio Toral, Sudip Kumar Naskar, Declan Groves, and Andy Way. 2014. Perception vs reality: Measuring machine translation post-editing productivity. Paper presented at the 3rd workshop on post-editing technology and practice (WPTP-3), within the 11th biennial conference of the Association Human Factors in Computing Systems (CHI). http://vis.stanford.edu/papers/post-editing [last accessed September 22, 2019]

Gaspari, Federico, Hala, Almaghout and Stephen Doherty. 2015. A survey of machine translation competences: insights for translation technology educators and practitioners. *Perspectives: Studies in Translatology*, 23(3): 333–358.

García, Ignacio. 2010. Is Machine Translation Ready Yet? *Target,* 22(1):7–21.

Guerberof, Anna. 2009. Productivity and quality in the post-editing of outputs from translation memories and machine translation. Localisation Focus. *The International Journal of Localisation*, 7(1): 11–21.

Guerberof, Anna, and Joss, Moorkens. 2019. Machine translation and post-editing training as part of a master's programme. *Jostrans: The Journal of Specialised Translation*, 31: 217-238.

Hvelplund, Tangsgaard Kristian. 2011. *Allocation of cognitive resources in translation: An eye-tracking and key-logging study*. Copenhagen, Denmark: Copenhagen Business School - PhD Series.

Kenny, Dorothy and Stephen Doherty. 2014. Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer*, 8(2): 276-294.

Koponen, Maarit. 2012. Comparing human perceptions of post-editing effort with post-editing operations. In *Proceedings of the 7th workshop on statistical machine translation*. Montreal, Canada

Koponen, Maarit. 2015. How to teach machine translation post-editing? Experiences from a post-editing course. In Proceedings *of the 4th Workshop on Post-Editing Technology and Practice* (WPTP4), pages 2-15.

Koponen, Maarit, and Leena Salmi. 2017. Post-editing quality: Analysing the correctness and necessity of post-editor corrections. *Linguistica Antverpiensia* New Series, 16: 137–148.

Koponen, Maarit, Leena Salmi, and Markku Nikulin. 2019. A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation*, 33: 61-90

Krings, Hans. 2001. *Repairing texts: empirical investigations of machine translation post-editing processes*. Kent: Kent State University Press

Ling Teh Hooi, NG Yim San, and Thomas Chow Voon Foo. 2016. The efficacy of Machine Translation tools in the translatıon of technical and non-technical texts: perceptions of undergraduate student users. *Laglit*, 3(2): 1–14.

Lommel, Arle Richard, and Donald DePalma. 2016. *Europe's Leading Role in Machine Translation: How Europe Is Driving the Shift to MT. Technical report*. Boston: Common Sense Advisory.

Mesa-Lao, Bartolome. 2014. Gaze behaviour on source texts: An exploratory study comparing translation and post-editing. In Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia (eds), *Post-editing of machine translation*. Cambridge Scholars Publishing, Newcastle upon Tyne, pages 219–245.

Mitchell, Linda. 2015. *Community post-editing of machine-translated user-generated content*. PhD thesis. Dublin City University, Dublin.

Moorkens, Joss. 2018. Eye tracking as a measure of cognitive effort for post-editing of machine translation. In Calum Walker, Federico M Federici FM (eds) *Eye tracking and multidisciplinary studies on translation*. John Benjamins, Amsterdam, pages 55–69.

Moorkens, Joss, and Sharon O'Brien. 2015. Post-editing evaluations: Trade-offs between novice and professional participants". In *Proceedings of the European Association for Machine Translation (EAMT)*, Antalya, Turkey, European Association for Machine Translation, pages 75–81.

O'Brien, Sharon. 2002. Teaching Post-editing: A Proposal for Course Content. *Proceedings of 6th EAMT Workshop Teaching Machine Translation*, Manchester, UK, European Association for Machine Translation, pages 99–106. http://mt-archive.info/EAMT-2002-OBrien.pdf

Pym, Anthony. 2013. Translation skill-sets in a machine-translation age. *Meta* 58(3): 487–503.

Raptis, Spyros and Giagkou, Maria. (2016). From capturing to generating human behavior: closing the interaction loop at the HUBIC Lab. In: *Proceedings of the 20th Pan-Hellenic Conference on Informatics (PCI) with International Participation*, Patras, Greece, 10-12 November 2016. ACM Digital Library, International Conference Proceedings Series. https://doi.org/10.1145/3003733.3003814.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA)*, pages 223–231.

Snover, Matthew, Nitin Madnani, Bonnie Dorr, and Richard Schwartz. 2009. Fluency, adequacy, or HTER? Exploring different human judgments with a tunable MT metric. *In Proceedings of the fourth workshop on statistical machine translation. Association for Computational Linguistics*, Athens, Greece, 259–268.

Wagner, Elizabeth. 1985. Post-editing Systran – A Challenge for Commission Translators. *Terminologie et Traduction*, (3):1–7.

Witczak, Olga. 2016. Incorporating post-editing into a computer-assisted translation course. A study of student attitudes. *Journal of Translator Education and Translation Studies*, 1(1): 33–55.

Yamada, Masaru. 2015. Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing settings. *Machine Translation*, 29: 49–67.