

Post-Editing Oriented Human Quality Evaluation of Neural Machine Translation in Translator Training: A Study on Perceived Difficulties and Benefits

Işın ÖNER* and Senem ÖNER BULUT**

The aim of this article is to investigate translation trainees' perceived difficulties and benefits of a post-editing oriented neural machine translation (NMT) error annotation and quality evaluation task which was carried out for the language pair English-Turkish and for two separate text types and domains, i.e., environmental blogs and movie/TV reviews, within the scope of an MA course on translation quality standards. The data to be analyzed were collected through a semi-structured questionnaire which was given to the trainees attending the course after the completion of the task. The questionnaire was prepared with the aim of understanding perceived difficulties and benefits of the task. Analysis of the answers revealed that most of the trainees were in the opinion that the task was difficult. Majority of the trainees also believed that the task was beneficial and enabled them to feel empowered to make decisions on human translation quality evaluation of machine translation (MT) and to carry out error annotation and post-editing activities in the future. According to a significant number of trainees, error annotation facilitated their post-editing process and reduced the effort in post-editing. Enhanced understanding of MT error annotation and enhanced ability to perform post-editing were the significant benefits stated by the trainees. The difficulties were associated with being introduced and assigned to perform tasks they were not familiar with. Yet, as displayed in the answers to the questionnaire, a considerable majority of the trainees were positive about the learning experience. The results have also shown that integration of MT-related activities into translator training with a focus on the empowerment of the human translator has its difficulties and benefits also for translation trainers. While the difficulties for the trainers concern the decisions on the design, implementation, and planning of the task and the responsibility to carry out the task in constant interaction and collaboration with the trainees, the benefits are the sense of fulfillment and enrichment brought by positive feedback from the trainees and the discovery of the fact that the so-called 'teaching experience' becomes a 'learning experience' for trainers as well as trainees.

Keywords: perceived difficulties and benefits; human translator; neural machine translation; human quality evaluation; error annotation; post-editing; translation quality standards

* Professor at Istanbul 29 Mayıs University.

E-mail: iboner@29mayis.edu.tr; ORCID ID: <https://orcid.org/0000-0001-7273-7229>.

** Associate professor at Istanbul Arel University.

E-mail: senemoner@arel.edu.tr; ORCID ID: <https://orcid.org/0000-0002-6186-4924>.

(Received 9 January 2021; accepted 31 May 2021)

1. Introduction

In the face of the widespread automatization of translation, the need to re-evaluate the existing and future professional roles of the human translator is more compelling than it has ever been. This is even more so considering that automatization leads to dehumanization and devaluation (O'Brien 2012) and anxiety (Vieira 2020) on the part of the human translator. It can also be argued that such anxiety is also felt on the part of the translation trainees and translator trainers.

In a working environment which is heavily shaped by machine translation (MT) technologies and practices, the interaction between the human translator and the machine is suggested to be a “dance of agency” (cf. Cadwell, O'Brien, and Teixeira 2018; Ruokonen and Koskinen 2017; Olohan 2011), which has unprecedented implications for translator training as well as for the translation profession.

The emerging professional roles of the human translator include, first and foremost, post-editing (e.g., Vieira 2019; Koponen 2016; O'Brien et al. 2014; Guerberof Arenas 2013; Garcia 2011). Albeit having received less attention than post-editing, pre-editing (e.g., Hiraoka and Yamada 2019; Mercader-Alarcón and Sánchez-Martínez 2016; Shei 2002; Somers 1997) has also emerged as a significant task for the human translator, which is closely related with the use of controlled language and technical writing (e.g., Alimen and Öner Bulut 2020; Öner 2018, 2019; Öner and Öner 2011; Yuste 2005; O'Brien 2002, 2006).

Along with post-editing and pre-editing, manual/human quality evaluation of MT is another topic which receives growing interest from researchers (e.g., Freitag et al. 2021; Castilho 2020; Seljan, Dunder, and Pavlovski 2020; Castilho et al. 2018; Klubička, Toral, and Sánchez-Cartagena 2018, 2017) and which constitutes the focus of the present paper together with post-editing.

Taking into consideration that manual/human quality evaluation of MT, which is closely related to the tasks of post-editing and pre-editing, is among the emerging professional roles of the human translator, the present paper aims to investigate the translation trainees' perceived difficulties and benefits of a post-editing oriented neural machine translation (NMT) error annotation and quality evaluation task. The task was carried out for the language pair English-Turkish and for two separate text types and domains, i.e., environmental blogs and movie/TV reviews, within the scope of an MA course on translation quality standards.

The data to be analyzed were collected through a semi-structured questionnaire which was given to the students after the completion of the task. The questionnaire was prepared with the aim of understanding perceived difficulties and benefits of the task.

The paper is structured as follows: Section 2 sketches out the requirements of translator training in the MT age with a focus on the empowerment of translation trainees. Section 3 presents the design and implementation of the post-editing oriented NMT error annotation and quality evaluation task. Section 4 presents the analysis of the students' answers to the questionnaire given at the end of the task. Section 5 is devoted to the concluding remarks.

2. Empowering the Translator Trainee in the Machine Translation Age

Critically reflecting on the existing practices in the global language and translation services sector and updating translator training curricula have always been regular tasks for researchers and trainers. Especially the advances in the field of MT in the last decade challenge translator training. The impact of MT is so profound that even the need for a re-definition of the concepts of the translator and translation (e.g., Sakamoto 2019; Gaspari, Almaghout, and Doherty 2015) and especially the need to answer “the new question of who a translator and post-editor is” (Sakamoto 2019, 65) are voiced in the related literature, which suggests the further need of research into the ways of including post-editing into translator training (cf. Öner Bulut 2019a).

Yet, post-editing is not the only role to be played by the human translator. As rightly claimed by Dorothy Kenny and Stephen Doherty (2014) in their seminal article titled “Statistical Machine Translation in the Translation Curriculum: Overcoming Obstacles and Empowering Translators,” an approach that limits the new expected roles of the human translator just to post-editing “marginalizes” (cf. 285) the human translator and thus has the potential to overlook other tasks such as “creation of parallel corpora, profiling of training data, intervening in the SMT process, **evaluation of outputs**, etc.” (cf. 286; our emphasis) which can be carried out by the human translator if s/he possesses “the necessary data, know-how and technology” (cf. 285).

Christopher D. Mellinger (2017) has also suggested the holistic inclusion of the tasks “controlled authoring, terminology management, engine tuning” along with “post-editing” into translator training at the curriculum level (cf. 281). Thus, in the specific context of the integration of MT-related tasks into translator training, training models to be developed with the aim of empowering the human translator are crucial.

Designing translator training models and activities that would help the translation trainee feel empowered and confident about the varying roles s/he would play in the MT age is closely related to the pedagogical approach to be adopted while designing and implementing such models and activities. One such approach, also adopted by the authors of the present paper, is Don Kiraly’s social constructivist approach to translator education in which the focus is on “learner empowerment” (2000, 17). What is of special concern for the purposes of the present paper is the construct of “self-efficacy beliefs” (Haro-Soler and Kiraly 2019) and its relevance for translator training in the MT age (cf. Doherty and Kenny 2014).

In a recent study on the investigation of “constructs related to the translator’s psychological ‘self’” (Haro-Soler and Kiraly 2019, 256), Maria del Mar Haro-Soler and Don Kiraly define “self-efficacy” as “the abilities that a person actually possesses to perform a specific task” and “self-efficacy beliefs” as “the individual’s perception of these abilities” and, in the context of translation, as “a translator’s or a translation student’s confidence in their ability to translate the text at hand under the specific conditions and situational constraints stated in and implied by the brief or assignment” (261).

Considering the aforementioned anxiety (see Introduction) caused by the growing prevalence of and especially the uncertainties associated with MT, we consider self-efficacy beliefs as a construct requiring utmost attention in designing learning activities for the training of the human translators and helping them develop “human translator competence” (Öner Bulut 2019b). The reason for this requirement is that future human translators will have to survive in future working environments where they will be constantly challenged by the increasing quality of MT. They will as well be demanded to “obtain appropriate social rewards for automation-resistant skills” (cf. Pym 2019). Thus, in order for the human translators to carry out value-added tasks that will keep them in the profession, translator training is required to help trainees develop a perception of their human abilities that would support them in their dance of agency with the machine (cf. Cadwell, O’Brien, and Teixeira

2018; Ruokonen and Koskinen 2017; Olohan 2011). And this is precisely the reasoning that underlies the post-editing oriented NMT error annotation and quality evaluation task designed and carried out within the scope of an MA course on translation quality standards. The details of the design and implementation of the task are presented in the following section.

3. Design and Implementation of the Post-Editing Oriented Neural Machine Translation Error Annotation and Quality Evaluation Task

The post-editing oriented NMT error annotation and quality evaluation task was implemented in the Translation Quality Standards course offered at the Translation Studies MA Program at Istanbul 29 Mayıs University, Turkey during the fall semester of the academic year 2020-2021. The students enrolled in the MA course had undergraduate training in translation or language departments and/or had work experience as translators. Although some of the students had limited training and/or experience in post-editing as part of their undergraduate study, none of the students had a special training or experience in MT error annotation and quality evaluation. As part of the course work, the students were divided into two groups to work on two text varieties, i.e., environmental blogs and movie/TV reviews, and the decision to work on either of the varieties was left to the students. Each student worked on a different source text and its raw MT output obtained using Google Translate.

For the error annotation sub-task, which consisted of identifying and annotating the errors in the raw MT output, the students were asked to use the following error categorization (table 1) which is the customized version of Multidimensional Quality Metrics¹ (MQM). Customization was performed through the collaboration of the course instructors and students.

¹ Available at <http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>.

Table 1. Customized version of MQM to be used in the error annotation sub-task

<p>1. Accuracy</p> <ul style="list-style-type: none"> • Addition • Mistranslation <ul style="list-style-type: none"> ○ Ambiguous translation ○ Date/time ○ Entity (such as name or place) ○ False friend ○ Number ○ Overly literal ○ Should not have been translated ○ Unit conversion • Omission <ul style="list-style-type: none"> ○ Omitted variable • Over-translation • Under-translation • Untranslated 	<p>2. Fluency</p> <ul style="list-style-type: none"> • Ambiguity <ul style="list-style-type: none"> ○ Unclear reference • Character encoding • Coherence • Cohesion • Duplication • Grammar <ul style="list-style-type: none"> ○ Function words ○ Word form <ul style="list-style-type: none"> ▪ Agreement ▪ Part of speech ▪ Tense/mood/aspect ○ Word order • Grammatical register • Inconsistency <ul style="list-style-type: none"> ○ Inconsistent abbreviations ○ Inconsistent with external reference • Offensive • Spelling <ul style="list-style-type: none"> ○ Capitalization ○ Diacritics • Typography <ul style="list-style-type: none"> ○ Punctuation ○ Unpaired quote marks or brackets ○ Whitespace • Unintelligible
<p>3. Style</p> <ul style="list-style-type: none"> • Register <ul style="list-style-type: none"> ○ Variants/slang • Awkward • Inconsistent style • Unidiomatic 	<p>4. Terminology</p> <ul style="list-style-type: none"> • Inconsistent with domain • Inconsistent use of terminology <ul style="list-style-type: none"> ○ Multiple terms for concept in source ○ Multiple translations of same term
<p>5. Verity</p> <ul style="list-style-type: none"> • Culture-specific reference • End-user suitability • Locale-specific content 	<p>6. Other</p>

Adopting a functionalist approach to translation, detailed briefs were given to the students at the beginning of the task for each text variety, and the students were asked to carry out first the error annotation sub-task. Then, on the basis of their annotation, they were asked to perform the sub-tasks of human translation, full-post editing, and light post-editing in line with the ISO 18587:2017 requirements of full and light post-editing, respectively. The below

table shows the briefs designed for the sub-tasks of human translation, full-post editing, and light post-editing.

Table 2. Briefs for human translation, full post-editing, and light post-editing

Brief for Environmental Blogs	Brief for Movie/TV Reviews
General information: The purpose in the blogs is both to inform the readers and to change their attitudes towards environmental issues.	General information: The purpose in the reviews is to entertain, persuade, and inform the readers.
Human translation: The translation will be published on the website of Greenpeace Turkey for the general public. It is important to use persuasive language in order to convey the message in the translation in line with the purpose of the blogs. Special attention should be paid to idiomatic use of language, correct use of syntax and grammar, correct use of terms, and correct spelling and punctuation.	Human translation: The translation will be published on the related websites for the readers interested in movie/TV reviews. It is important to use expressive and persuasive language in order to convey the message in the translation in line with the purpose of the reviews. Special attention should be paid to style, idiomatic use of language, correct use of syntax and grammar, correct use of terms, and correct spelling and punctuation.
Full post-editing: The translation will be published on the website of Greenpeace Turkey for the general public. It is important to use persuasive language in order to convey the message in the translation in line with the purpose of the blogs. Special attention should be paid to idiomatic use of language, correct use of syntax and grammar, correct use of terms, and correct spelling and punctuation.	Full post-editing: The translation will be published on the related websites for the readers interested in movie/TV reviews. It is important to use expressive and persuasive language in order to convey the message in the translation in line with the purpose of the reviews. Special attention should be paid to style, idiomatic use of language, correct use of syntax and grammar, correct use of terms, and correct spelling and punctuation.
Light post-editing: The translation will not be published and will be used for gisting. It is NOT necessary to use persuasive language in order to convey the message in the translation in line with the purpose of the blogs. Informative content is important.	Light post-editing: The translation will not be published and will be used for gisting. It is NOT necessary to use expressive and persuasive language in order to convey the message in the translation in line with the purpose of the reviews. Informative content is important.

The implementation of the sub-tasks was carried out through weekly assignments and the workshops held a week after the assignments were given. Difficulties encountered by the students during the completion of assignments and their solutions for the difficulties were discussed in the workshops where the instructors continuously gave feedback to the students

on their assignments. In the workshops, the instructors also explained and presented their justification for the design of the learning experience as well as of each sub-task and activity to the students asking whether they had any comments or revision suggestions. After the completion of the task, the students were asked to anonymously fill out a semi-structured questionnaire (see appendices) which was prepared in Google Docs and consisted of close-ended and open-ended questions. The questions in the questionnaire were formulated with the aim of understanding the students' assessment and self-assessment of their learning experience with a focus on the perceived benefits and difficulties of the task. The questionnaire was completed by 11 out of 15 students. The answers obtained will be analyzed in the following section.

4. Analysis of the Students' Answers to the Questionnaire

Analysis of the answers to the close-ended questions related to the students' perceptions of the benefit of the task has shown that an overwhelming majority of the students found the task useful as 90.9% of the students either agreed or strongly agreed with the statement "I benefited considerably from error annotation" while 9.1% were neutral (part 1, question 1). Similarly, 90.9% of the students either agreed or strongly agreed with the statement "Annotating machine translation errors helped me considerably in the post-editing process" while 9.1% were neutral (part 1, question 6). The statements "Having completed the tasks of error annotation and post-editing, I feel more empowered and safe to make decisions on human translation quality evaluation as well as machine translation" (part 1, question 13) and "I feel experienced in anticipating machine translation errors" (part 1, question 10) were agreed or strongly agreed with by 81.9% of the students.

As for the students' perception of the benefit of their error annotation and post-editing experience for their future roles as post-editors, 90.9% of the students either agreed or strongly agreed with the statement "I think that the error annotation and post-editing experience I have gained during the study will help me considerably in my future endeavors as a post-editor" (part 1, question 12). However, the percentage of the students who agreed with the statement "I think that the error annotation and post-editing experience I have gained during the study will help me considerably in translating from scratch" (part 1, question 11)

was slightly lower (strongly agree = 45.5%, agree = 27.3%) while 9.1% disagreed with and 18.2% were neutral about the statement.

Analysis of the answers to the questions related to the perceived difficulty of the sub-tasks has shown that 72.7% of the students either agreed or strongly agreed with the statement “Overall error annotation was considerably difficult” (part 1, question 2) while 27.3% were neutral. The answers to the two questions related to the difficulty concerning the selection of the error type and the relevance of translation direction have shown that 63.6% of the students either agreed or strongly agreed with the statement “I had difficulty selecting the error type during error annotation of the raw machine translation output in the direction from English to Turkish” (part 1, question 3) while 9.1% disagreed with and 27.3% were neutral about the statement. On the other hand, 72.8% of the students either agreed or strongly agreed with the statement “I had difficulty selecting the error type during error annotation of the raw machine translation output in the direction from Turkish to English” (part 1, question 4) while 27.3% were neutral, which shows that the students found error annotation of the raw MT output slightly more difficult for the direction from Turkish to English while they found error annotation in Turkish easier.

In terms of the difficulty associated with the interpretation and use of ISO 18587:2017 requirements of full and light post-editing in post-editing sub-tasks, it has been understood that the majority of the students found the requirements easy to apply as 72.8% either agreed or strongly agreed with the statements “I did not have any difficulty interpreting and using the ISO 18587:2017 requirements of full and light post-editing” (part 1, question 7) and “If I had not had previous translation training and/or experience, I would not have easily interpreted and used ISO 18587:2017 requirements of post-editing” (part 1, question 8) while 9.1% disagreed with and 18.2% were neutral about the statements. We believe this finding is particularly significant as it points to the students’ perceived relationship between using the standards for post-editing and having a previous translation training and/or experience.

The said perception is also evident in the answers of the students to the questions on the impact of their previous training and/or experience on their error annotation and post-editing performance as 72.8% either agreed or strongly agreed with the statement “My previous translation training and/or experience helped me considerably both in error annotation and in post-editing” (part 1, question 14) while 27.3% were neutral about it. Yet,

interestingly, 54.6% either agreed or strongly agreed with the statement “If one does not know how to translate from scratch, s/he cannot make machine translation error annotation and post-editing” (part 1, question 15) while 9.1% disagreed with and 36.4% were neutral about it, which shows that nearly half of the students are not sure about or disagree with the relevance of translation training and/or experience for error annotation and post-editing.

The answers of the students to the questions about the revision of their error annotation has displayed a certain degree of confidence in their error annotation practice as only 45.5% either agreed or strongly agreed with the statement “Most of the error types I had selected were changed by the instructors” (part 1, question 5) while 36.4% disagreed with and 18.2% were neutral about it. Similarly, only 36.4% agreed with the statement “I had to correct errors which I had not annotated previously in the full post-editing process” (part 1, question 9) while 27.3% either disagreed or strongly disagreed with and 18.2% were neutral about the statement.

The overall results of the answers to the close-ended questions indicate that majority of the students benefited from the task, felt empowered and confident to carry out MT quality evaluation tasks in the future despite experiencing significant difficulty annotating errors. Majority of the students believed that their previous translation training and/or experience had a positive effect on their error annotation and post-editing performance. However, the answers of the students diverged in response to the question concerning the relevance of the ability to translate from scratch for post-editing, which points out the need for further research into the relationship between translating from scratch and post-editing.

Analysis of the students’ answers to the open-ended question “If you think that you have benefited (considerably) from error annotation, in what way or ways have you benefited? Please first state the domain of the texts you worked on.” (part 2, question A) has shown that the students thought that they benefited from error annotation in several ways. In the answers to this question, the focus was on the benefit of error annotation for MT error anticipation and post-editing. For instance, according to one student, the task gave her/him “a chance to see what kinds of errors could occur during MT and how to spot them,” and the text analysis s/he performed made her/him “more comfortable working on this ‘new’ type of text (environmental blog) and method” (S6). Another student stated that error annotation helped her/him “analyze the machine translation output” and provided “a roadmap” to be followed in

post-editing (S11). A similar emphasis was evident also in the answer of another student according to whom “error annotation and seeing where most of the errors occurred” enabled her/him to “understand the capabilities and shortcomings of MT systems” and thus for her/him “made it much easier to post-edit the MT” (S8).

Benefit of practice and gaining experience in a “real” project was emphasized by the student who stated, “I believe that working on the errors helped me understand the error categories in a way that simply reading about them could not have” (S2). Another student also emphasized that the feedback received from the instructors in the workshops enabled her/him to “experience this process in a real post-editing project,” that s/he feels “more decisive at making decisions,” and that this experience will help her/him “in future projects as a post-editor” (S8). Finally, one student stated that “despite the fear among some translators that human translators will lose their jobs because of machine translation,” s/he realized that “there will always be need for professional translators or translation scholars because machines also need human guidance,” which made her/him “encouraged to take part in post-editing in the future” (S1).

As for the perceived difficulty of error annotation, the answers to the open-ended question “If you think that overall error annotation was considerably difficult, could you explain these difficulties?” (part 2, question B) have displayed that most of the students found error annotation and error type selection difficult. According to one student, error annotation was difficult for two specific reasons. First, it was her/his first time annotating errors of MT, and second, error identification of NMT output was more difficult than that of statistical machine translation (SMT) outputs since “[I]n SMT, you can spot errors so easily, but when it comes to NMT, you just don’t know if it’s an error or not” (S11). Another student explained her/his perception of difficulty with a focus on the “subjectivity” inherent to error identification and on the fact that the metrics used were based on the English grammar as follows:

S1:

First of all, I found identifying an error incredibly subjective despite all the metrics which are supposed to be described clearly. I couldn’t be sure if an error was actually an error. Second problem occurs when an error is identified and needs to be categorized. Sometimes I couldn’t specify the error even though I was sure it was an error. I’d like to add one thing on error categorization here. Sub-categories under “Grammar” according to MQM metrics may be insufficient to determine the

grammatical errors in Turkish since they are based on English grammar. To my view, metrics should include grammatical issues which are innate to agglutinative languages.²

The answers to the open-ended question “If you think that annotating machine translation errors helped you (considerably) in the post-editing process, can you explain how?” (part 2, question C) have illustrated that majority of the students thought that error annotation facilitated their post-editing process and reduced the effort in post-editing. One student stated that error annotation facilitated the process because s/he was “sure about the errors at that point,” adding that “[T]hanks to the categories I annotated, I could come up with a solution more easily” (S1). According to another student, error annotation “enhances the post-editor’s ability to analyze the text thoroughly,” and identification of “characteristic error categories can improve the post-editor’s ability to predict the potential errors in further machine translation projects” (S3). In other students’ views, error annotation “saves time when you’re post-editing since you get to comprehend the source text beforehand and know where the errors are. Even though post-editing is not just ‘finding errors,’ it could be said that error annotation reduces the effort in the post-editing process” (S11) and “makes you think about what’s wrong with the MT sentence; helps you focus on the part that needs editing and so makes editing easier” (S8). According to another student, the task “was a great experience for the future” (S7).

The answers to the open-ended questions “If you had difficulty interpreting the ISO 18587:2017 requirements of full and light post-editing, what were those difficulties and how did you overcome them?” (part 2, question D) and “If you think that on condition that you had not had previous translation training and/or experience, you would not have easily interpreted and used ISO 18587:2017 requirements of post-editing, could you explain how your previous translation training and/or experience facilitated your interpretation and usage of the requirements?” (part 2, question E) which concern the ease of use of ISO 18587:2017 light and full post-editing requirements have shown that the majority of the students did not experience much difficulty interpreting and using them. However, two students stated that it was difficult to decide “where light post-editing ends and full post-editing begins” (S2) and that “[T]he rule about ‘editing inappropriate content’ was a little vague. Some examples

² The students answered the questions in the questionnaire in English. The statements were not altered or revised except the correction of minor grammar and spelling mistakes.

would make it easier for me to understand what it meant, I think” (S8). As for the perceived relationship between previous translation training and/or experience and interpretation/use of requirements, the answers have shown that although some students were not sure about the said relationship, others believed that their previous training and/or experience helped them interpret and use the requirements. One student stated that “the biggest plus of” her/his previous experience “was self-confidence because I’d edited and translated numerous texts before at work and also as part of my training” (S8). Another student also underlined the significance of her/his previous translation training as follows:

S9:

I don’t think it would be this easy to use ISO 18587:2017 if I had not had previous translation training or experience. When I was an undergraduate Translation Studies student, I took courses related to MT-PE and International Standards and we have learned post-editing process. We also learned how to work in accordance with standards so in this project, I knew how to conduct a post-editing process and I can easily implement the requirements and stages in ISO 18587:2017. If I had not had any previous experience, following the requirements in standards would be a lot harder for me.

The answers to the open-ended question “If you think that the error annotation and post-editing experience you have gained during the study will help you (considerably) in translating from scratch experience, in what way or ways will this experience help you?” (part 2, question F) have shown that some students were not sure about the potential contribution of error annotation and post-editing experience to translation from scratch. Instead, they emphasized that the experience would rather help them in “post-editing” (S5), (S9) and “editing” (S6), (S7) processes while two students clearly stated that they do not think that the experience will help them translate from scratch (S8), (S11). In contrast, the answers to the question “If you think that the error annotation and post-editing experience you have gained during the study will help you (considerably) in your future endeavors as a post-editor, in what way or ways will this experience help you?” (part 2, question G) have shown that an overwhelming majority of the students believed that the experience they gained would help them in their future roles as post-editors. One student stated that having completed the task, s/he would know “how to handle the source text and MT output” and “in what kind of a point of view” s/he “should analyze them” (S11). According to other students, the experience will help them “in another post-editing project” (S5) and “work more carefully” (S7) and enable

them to “feel more comfortable at a job meeting” (S6). In another student’s view, “[E]rror annotation makes post-editing easier. Everything goes smoother and faster” (S8). Two students who explained the perceived benefit of the experience also underlined the need for further experience as follows:

S1:

I used to have reservations about MTPE and I thought it is no match for human translators. Yet I am now more optimistic about it provided that human translators and scholars contribute to the development of MT. Additionally, I now feel much more confident about error annotation since I got familiar with the metrics and applied them. However, I know that I need more experience to do it on professional level.

S9:

I actually have no experience as a professional post-editor, since I work as a translator for three years. I only worked as a post-editor in our group projects, as an undergraduate student. This project has taught me the more detailed part of the process, the error annotation part and helped me to have an insight about ISO standard. I have post-edited two texts within this project and all the work I have done, including creating an error annotation table and analyzing the International Standard for this process will help me as a post-editor in the future, if I attend any projects as a post-editor.

The answers to the open-ended question “If, having completed the tasks of error annotation and post-editing, you feel more empowered and safe to make decisions on human translation quality evaluation as well as machine translation, in what way or ways will this experience help you?” (part 2, question H) have displayed varied perceptions concerning empowerment. One student stated that he did not think he was empowered and safe to make decisions on human translation quality evaluation on the grounds that the implemented task “showed” her/him “how hard it was to identify errors in a translation,” yet s/he still added that s/he will “evaluate” her/his “own translation more consciously from now on” (S1). Another student made a distinction between empowerment and competence as s/he stated that s/he feels “a bit empowered, but not competent enough” (S5). Two students emphasized that the experience will help them “make corrections or decisions confidently” (S6) and “make easier decisions” (S7) in the editing and/or post-editing processes. Another student stated that carrying out the task “step by step” made her/him “feel” that “a quality job” was done “at the end” (S10). Emphasis was put on the feedback received from the instructors by one student who stated, “What helped me through this project was my advisers’ feedback on my error

annotation. After I saw what I did wrong in that part, I feel more empowered to make decisions” (S9). One student, on the other hand, made a distinction between feeling empowered related to MT quality evaluation and human translation quality evaluation as follows:

S11:

I am neutral about this. I feel empowered to decide on machine translation quality and these experiences will help me as a future professional post-editor, but I feel like I can't say the same thing about human translation quality since you don't look for the same kind of errors or the same aspects in general. It feels like evaluating human translation is another dimension.

The answers to the open-ended question “If you think that your previous translation training and/or experience helped you (considerably) both in error annotation and in post-editing, in what way or ways have you benefited from your previous training and/or experience?” (part 2, question I) have also shown varied responses. While some students (S1) expressed that they were not sure whether previous translation training and/or experience helped them in the implemented tasks or that they did not have previous experience in post-editing (S4), (S5), others were in the belief that their previous training and/or experience helped them (S2), (S3), (S6), (S7). Three students explained how they felt confident thanks to their previous experience and/or training as follows:

S8:

I'd never had post-editing training. I'd never used error annotation either. Thanks to my training and experience, I knew how to approach a text and I knew there are multiple parties involved in the translation process like the client or clients, other translators, revisers, editors etc. The only direct benefit of my training and experience was my attitude toward error annotation and post-editing. I was confident in myself.

S9:

I feel neutral on this one because it's the first time I prepared an error annotation table and used it as a reference in the post-editing process but I am definitely sure that this experience will help me in my future projects. But with the help of my training as a translator, I can detect problems easily and I make the right decisions in light of my theoretical knowledge.

S11:

I think I have benefited from the classes related to linguistics, since you are deeply involved in how the language works, also the classes related to translation technologies since I had the chance to experience nearly every step of a translation

project and those classes raised awareness of machine translation post-editing, and lastly the classes related to text analysis made me define the source texts.

The answers to the open-ended question “If you think that on condition that one does not know how to translate from scratch, s/he cannot make machine translation error annotation and post-editing, could you explain why?” (part 2, question J) have displayed two main and conflicting perceptions among the students. Some students stated that no experience as a translator is necessary for carrying out the tasks in question (S2), (S4) or that they were unsure or neutral about this relationship since “a bilingual person might be able to detect and may even annotate errors in a translation with the help of metrics” (S1) or “it depends on the machine translation quality and what kind of post-editing (light or full) required” (S11). Other students were in the opinion that “in order to complete a task like this one should have the basic knowledge on translation” (S7), that “[I]f someone doesn’t know how to translate a text, it would be almost impossible to catch mistakes specific to text type for example” (S6), that “[S]omeone who does not know how to translate from scratch cannot know and see where there was a mistake and what to dictate while editing” (S10), and that “proper training” is needed “to understand metrics in error annotation” (S5). Two students, who firmly believed that knowing to translate from scratch was a prerequisite for MT error annotation and post-editing, explained their opinions as follows:

S8:

Some post-editing requirements involve restructuring sentences and producing content that is appropriate for the audience and ensuring cohesion etc. If you don’t have the capability to translate a text from scratch, you wouldn’t catch all the errors in the MT output. The post-edited translation would still have errors. If those errors are minor ones, it might not be a big issue. However, if they are critical, that means all you’re doing is waste time and resources.

S9:

I agreed to this statement, “but” I think, if a person does not know how to translate from scratch, s/he actually may perform a post-editing from an MT output, from “A” language into her/his native language, just because she may understand the main idea of the topic, but the result could only be a “good enough quality” text in the target language and the truth is disputed. S/he cannot make an error annotation because of the lack of knowledge on error types and how to decide on. The post-editor has to be fully competent in the source language and the target language to perform this task and has to have an education or experience in the field. This is the reason why I think s/he cannot make machine translation error annotation and post editing.

As for the overall assessment of the learning experience, the answers to the open-ended question “How do you evaluate the overall learning experience and are there any other comments you would like to make?” (part 2, question K) have shown that overwhelming majority of the students had a positive perception of the experience and stated that it was “useful” (S6), “beneficial” (S5), “fun and challenging” (S9), “challenging yet unexpectedly thought-provoking and enjoyable” (S1), “inspiring” (S4), “a great experience” that would help students “in the future” (S8), and “an important opportunity to grow as a post-editor and translator” (S7), the only criticism being that the design could have been more “clear and structured” (S2). While assessing the learning experience, it was also observed that the students re-emphasized the difficulty they experienced, especially concerning error annotation, as seen in the following remarks:

S1:

It (the learning experience) made me see that nothing is easy as it seems. Error annotation was way more difficult than I imagined. On the other hand, I found out that I should have more confidence in the advances in machine translation. Some of the MT output was surprisingly successful.

S9:

It was fun and challenging. Sometimes I felt lost, especially while I tried to understand all the error types and tried to find them in the MT. After this class, I have a better understanding of quality standards and why we need them. I also learned a lot about post-editing. Since I didn't have any previous training on it, even basic information was new to me. Overall I had a positive experience.

S10:

It was a nice experience since I didn't attend any MT error annotation and post-editing project as a professional post-editor and my experience on this topic was on undergraduate level. I learned to prepare a detailed error annotation table. With the help of error annotation part, my knowledge on detecting errors and selecting error types have improved, thanks to my Professors' feedbacks. I feel more empowered and decisive now. I have benefited so much from the feedbacks on my Error Annotation table. I also learned MQM-DQF error typology template and had a chance to learn about the details in ISO 18587:2017 stages. Thank you for giving us the chance to attend such a fruitful project.

S3:

I greatly enjoyed the experience. The classes were engaging and filled with information that I had never seen or heard of or ones that I had but not studied about.

Even in cases where the reading or work felt tiring or boring, I still never felt that it was unnecessary.

S4:

It was an inspiring learning experience which led me to think on machine learning more and how to redefine translation as a profession in today's world.

The overall results of the answers to the close-ended questions show that the task of error annotation was beneficial for MT error anticipation and post-editing, and it reduced the effort in post-editing, helping students feel empowered. Yet, majority of the students also found error annotation and error type selection difficult. Accordingly, a number of students underlined the difference between empowerment and competence, expressing awareness of the further need of practice and experience. The considerably positive perception of the overall assessment of the task reveals that the difficulties were moderated by perceived benefits.

5. Conclusions

The post-editing oriented human quality evaluation of NMT task, which was carried out in an MA course on translation quality standards and the results of which have been presented above, has shown that integration of MT-related activities into translator training with a focus on the empowerment of the human translator has its difficulties and benefits both for translation trainees and translation trainers.

One difficulty for the trainers was to decide how to design and implement a particular task and carefully plan the steps involved. Another difficulty was to carry out the task in constant interaction and collaboration with the trainees, helping the trainees feel themselves as active participants of the whole learning experience, trying to explain and convince them of the necessity of the sub-tasks to be performed, and asking for their comments and suggestions for the design and implementation of the task on a regular basis. The benefits, on the other hand, are the sense of fulfillment and enrichment that comes from receiving positive feedback that the learning experience actually worked for the trainees and from discovering that the so-called 'teaching experience' becomes a 'learning experience' also for trainers as well as trainees.

The difficulties for the trainees, as inferred through the analysis of their answers to the semi-structured questionnaire prepared with the aim of understanding their perceived difficulties and benefits, seem to have stemmed from being introduced and assigned to perform tasks they were not familiar with as part of their previous training and/or experience. Yet, the benefits seem to be many. As the answers to the questionnaire have displayed, a considerable majority of the students were positive about the learning experience. The results have shown that having completed the tasks, most of the students either expressed that they became aware of the challenges and requirements of error annotation and post-editing tasks or that they felt confident about their ability to carry out these tasks in the future, which is in harmony with the very definition of self-efficacy beliefs in the context of translation (Haro-Soler and Kiraly 2019, 261).

Further empirical studies are needed to design and implement innovative learning activities and investigate their results with a focus on the empowerment of translation trainees. One such study would be to examine the relationship between post-editing and translating from scratch for specific language pairs and domains. We believe such studies would shed light on various questions about the future roles of translator trainers and human translators in the MT age.

Appendices

Questionnaire – Part 1

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. I benefited considerably from error annotation.					
2. Overall error annotation was considerably difficult.					
3. I had difficulty selecting the error type during error annotation of the raw machine translation output in the direction from English to Turkish.					
4. I had difficulty selecting the error type during error annotation of the raw machine translation output in the direction from Turkish to English.					
5. Most of the error types I had selected were changed by the instructors.					
6. Annotating machine translation errors helped me considerably in the post-editing process.					
7. I did not have any difficulty interpreting and using the ISO 18587:2017 requirements of full and light post-editing.					
8. If I had not had previous translation training and/or experience, I would not have easily interpreted and used ISO 18587:2017 requirements of post-editing.					
9. I had to correct errors which I had not annotated previously in the full post-editing process.					
10. I feel experienced in anticipating machine translation errors.					
11. I think that the error annotation and post-editing experience I have gained during the study will help me considerably in translating from scratch.					
12. I think that the error annotation and post-editing experience I have gained during the study will help me considerably in my future endeavors as a post-editor.					
13. Having completed the tasks of error annotation and post-editing, I feel more empowered and safe to make decisions on human translation quality evaluation as well as machine translation.					
14. My previous translation training and/or experience helped me considerably both in error annotation and in post-editing.					
15. If one does not know how to translate from scratch, s/he cannot make machine translation error annotation and post-editing.					

Questionnaire – Part 2

- A. If you think that you have benefited (considerably) from error annotation, in what way or ways have you benefited?
- B. If you think that overall error annotation was considerably difficult, could you explain these difficulties?
- C. If you think that annotating machine translation errors helped you (considerably) in the post-editing process, can you explain how?
- D. If you had difficulty interpreting the ISO 18587:2017 requirements of full and light post-editing, what were those difficulties and how did you overcome them?
- E. If you think that on condition that you had not had previous translation training and/or experience, you would not have easily interpreted and used ISO 18587:2017 requirements of post-editing, could you explain how your previous translation training and/or experience facilitated your interpretation and usage of the requirements?
- F. If you think that the error annotation and post-editing experience you have gained during the study will help you (considerably) in translating from scratch experience, in what way or ways will this experience help you?
- G. If you think that the error annotation and post-editing experience you have gained during the study will help you (considerably) in your future endeavors as a post-editor, in what way or ways will this experience help you?
- H. If, having completed the tasks of error annotation and post-editing, you feel more empowered and safe to make decisions on human translation quality evaluation as well as machine translation, in what way or ways will this experience help you?
- I. If you think that your previous translation training and/or experience helped you (considerably) both in error annotation and in post-editing, in what way or ways have you benefited from your previous training and/or experience?
- J. If you think that on condition that one does not know how to translate from scratch, s/he cannot make machine translation error annotation and post-editing, could you explain why?
- K. How do you evaluate the overall learning experience and are there any other comments you would like to make?

References

- Alimen, Nilüfer, and Senem Öner Bulut. 2020. “Çevirinin Teknolojikleşmesi Bağlamında İnsan Çevirmenin Rollerini Yeniden Düşünmek: Çevirmen Eğitiminde Teknik Metin Yazarlığı.” [Re-thinking the roles of the human translator in the context of the technologization of translation: Technical writing in translator training.] *Turkish Studies – Language and Literature* 15 (3): 1047–1062. doi:10.47845/TurkishStudies.45679.
- Cadwell, Patrick, Sharon O’Brien, and Carlos S. C. Teixeira. 2018. “Resistance and Accommodation: Factors for the (Non-) Adoption of Machine Translation Among Professional Translators.” *Perspectives* 26 (3): 301–321. doi:10.1080/0907676X.2017.1337210.
- Castilho, Sheila. 2020. “On the Same Page? Comparing Inter-Annotator Agreement in Sentence and Document Level Human Machine Translation Evaluation.” In *Proceedings of the 5th Conference on Machine Translation (WMT)*, 1150–1159. Association for Computational Linguistics (ACL). <https://www.aclweb.org/anthology/2020.wmt-1.137.pdf>.
- Castilho, Sheila, Joss Moorkens, Federico Gaspari, Rico Sennrich, Andy Way, and Panayota Georgakopoulou. 2018. “Evaluating MT for Massive Open Online Courses: A Multifaceted Comparison Between PBSMT and NMT Systems.” In “Human Evaluation of Statistical and Neural Machine Translation,” edited by Andy Way and Mikel L. Forcada. Special Issue, *Machine Translation* 32 (3): 255–278. doi:10.1007/s10590-018-9221-y.
- 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. doi:10.1080/1750399X.2014.937571.
- Freitag, Markus, George Foster, David Grangier, Viresh Ratnakar, Qijun Tan, and Wolfgang Macherey. 2021. “Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation.” arXiv preprint arXiv:2104.14478, April 29. <https://arxiv.org/pdf/2104.14478.pdf>.
- Garcia, Ignacio. 2011. “Translating by Post-Editing: Is It the Way Forward?” *Machine Translation* 25 (3): 217–237. doi:10.1007/s10590-011-9115-8.
- Gaspari, Federico, Hala Almaghout, and Stephen Doherty. 2015. “A Survey of Machine Translation Competences: Insights for Translation Technology Educators and Practitioners.” *Perspectives* 23 (3): 333–358. doi:10.1080/0907676X.2014.979842.
- Guerberof Arenas, Ana. 2013. “What Do Professional Translators Think About Post-Editing?” In “Machine Translation and the Working Methods of Translators,” edited by Louise Brunette. Special Issue, *JoSTrans The Journal of Specialised Translation*, no. 19, 75–95. https://www.jostrans.org/issue19/art_guerberof.pdf.

- Haro-Soler, Maria del Mar, and Don Kiraly. 2019. “Exploring Self-Efficacy Beliefs in Symbiotic Collaboration with Students: An Action Research Project.” In “Training the Trainers,” edited by Gary Massey, Don Kiraly, and Maureen Ehrensberger-Dow. Special Issue, *The Interpreter and Translator Trainer* 13 (3): 255–270. doi:10.1080/1750399X.2019.1656405.
- Hiraoka, Yusuke, and Masaru Yamada. 2019. “Pre-Editing Plus Neural Machine Translation for Subtitling: Effective Pre-Editing Rules for Subtitling of TED Talks.” In *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, 64–72. <https://www.aclweb.org/anthology/W19-6710.pdf>.
- 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. doi:10.1080/1750399X.2014.936112.
- Kiraly, Don. 2000. *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. Manchester: St. Jerome.
- Klubička Filip, Antonio Toral, and Víctor M. Sánchez-Cartagena. 2017. “Fine-Grained Human Evaluation of Neural Versus Phrase-Based Machine Translation.” *The Prague Bulletin of Mathematical Linguistics*, no. 108, 121–132. doi:10.1515/pralin-2017-0014.
- . 2018. “Quantitative Fine-Grained Human Evaluation of Machine Translation Systems: A Case Study on English to Croatian.” In “Human Evaluation of Statistical and Neural Machine Translation,” edited by Andy Way and Mikel L. Forcada. Special Issue, *Machine Translation* 32 (3): 195–215. doi:10.1007/s10590-018-9214-x.
- Koponen, Maarit. 2016. “Is Machine Translation Post-Editing Worth the Effort? A Survey of Research into Post-Editing and Effort.” In “The Translation Profession: Centres and Peripheries,” edited by Helle V. Dam and Kaisa Koskinen. Special Issue, *JoSTrans The Journal of Specialised Translation*, no. 25, 131–148. https://www.jostrans.org/issue25/art_koponen.pdf.
- Mellinger, Christopher D. 2017. “Translators and Machine Translation: Knowledge and Skills Gaps in Translator Pedagogy.” *The Interpreter and Translator Trainer* 11 (4): 280–293. doi:10.1080/1750399X.2017.1359760.
- Mercader-Alarcón, Julia, and Felipe Sánchez-Martínez. 2016. “Analysis of Translation Errors and Evaluation of Pre-Editing Rules for the Translation of English News Texts into Spanish with Lucy LT.” *Traducció i dispositius mòbils Revista Tradumàtica: tecnologies de la traducció*, no. 14, 172–186. doi:10.5565/rev/tradumatica.164.
- O’Brien, Sharon. 2002. “Teaching Post-Editing: A Proposal for Course Content.” In *Proceedings of 6th EAMT Workshop: Teaching Machine Translation*, 99–106. <https://www.aclweb.org/anthology/2002.eamt-1.11.pdf>.

- . 2006. “Machine-Translatability and Post-Editing Effort: An Empirical Study Using Translog and Choice Network Analysis.” PhD diss., Dublin City University.
- . 2012. “Translation as Human–Computer Interaction.” *Translation Spaces* 1 (1): 101–122. doi:10.1075/ts.1.05obr.
- O’Brien, Sharon, Laura Winther Balling, Michael Carl, Michel Simard, and Lucia Specia, eds. 2014. *Post-Editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars.
- Olohan, Maeve. 2011. “Translators and Translation Technology: The Dance of Agency.” *Translation Studies* 4 (3): 342–357. doi:10.1080/14781700.2011.589656.
- Öner, Işın. 2018. “Why Technical Writing is One of the Basic Courses in Translation Studies.” *Intelligent Information Blog*. January 31. <https://intelligent-information.blog/en/why-technical-writing-is-one-of-the-basic-courses-in-translation-studies>.
- . 2019. “Teknik Metin Yazarlığı Nedir? Çeviri Eğitimindeki Yeri Nedir?” [What is technical writing? What is its position in translation training?] Speech delivered at the İstanbul Üniversitesi Edebiyat Fakültesi Çeviribilim Bölümü Çeviri Söyleşileri 16 [Istanbul University Faculty of Letters Department of Translation Studies Conversations on Translation 16], March 1.
- Öner, Işın, and Senem Öner. 2011. “Exploring the Writing and Translation of Technical Documentation for Its Integration into Translator Training Programs in Turkey: A Preliminary Model.” Paper presented at the Aspects in Production and Translation of Technical Documentation International Tekom Conference, Istanbul, December 12.
- Öner Bulut, Senem. 2019a. “Future Professional Profile and Agency of the Human Translator: A Survey on Human-Machine Tension in the Context of the Technologization of Translation.” In *Çeviribilimde Araştırmalar* [Research in Translation Studies], edited by Seda Taş, 93–122. Istanbul: Hiperyayın.
- . 2019b. “Integrating Machine Translation into Translator Training: Towards ‘Human Translator Competence’?” *transLogos Translation Studies Journal* 2 (2): 1–26. doi:10.29228/transLogos.11.
- Pym, Anthony. 2019. “How Automation Through Neural Machine Translation Might Change the Skill Sets of Translators.” Accessed September 10, 2019. https://www.academia.edu/40200406/How_automation_through_neural_machine_translation_might_change_the_skill_sets_of_translators.
- Ruokonen, Minna, and Kaisa Koskinen. 2017. “Dancing with Technology: Translators’ Narratives on the Dance of Human and Machinic Agency in Translation Work.” *The Translator* 23 (3): 310–323. doi:10.1080/13556509.2017.1301846.

- Sakamoto, Akiko. 2019. “Unintended Consequences of Translation Technologies: From Project Managers’ Perspectives.” *Perspectives* 27 (1): 58–73. doi:10.1080/0907676X.2018.1473452.
- Seljan, Sanja, Ivan Dunđer, and Marko Pavlovski. 2020. “Human Quality Evaluation of Machine-Translated Poetry.” In *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1040–1045. doi:10.23919/MIPRO48935.2020.9245436.
- Shei, Chi-Chiang. 2002. “Teaching MT Through Pre-Editing: Three Case Studies.” In *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*. <https://www.aclweb.org/anthology/2002.eamt-1.10.pdf>.
- Somers, Harold. 1997. “A Practical Approach to Using Machine Translation Software: ‘Post-Editing’ the Source Text.” *The Translator* 3 (2): 193–212. doi:10.1080/13556509.1997.10798998.
- Vieira, Lucas Nunes. 2019. “Post-Editing of Machine Translation.” In *The Routledge Handbook of Translation and Technology*, edited by Minako O’Hagan, 319–335. London: Routledge.
- . 2020. “Automation Anxiety and Translators.” *Translation Studies* 13 (1): 1–21. doi:10.1080/14781700.2018.1543613.
- Yuste, Elia. 2005. “Computer-Aided Technical Translation Workflows – Man-Machine in the Construction and Transfer of Corporate Knowledge.” *Linguistik Online* 23 (2): 67–75. doi:10.13092/lo.23.647.