



This work is licenced under a Creative Commons Attribution 4.0 International License  
Ovaj rad dostupan je za upotrebu pod licencom Creative Commons Imenovanje 4.0. međunarodna

---

IZVORNI  
ZNANSTVENI RAD

UDK: 81'25:81'322:378.091.3  
811.111:811.163.42  
DOI: 10.22210/strjez/50-2/2

Primljen: 12. 4. 2021.  
Prihvaćen 9. 5. 2021.

---

# Towards acquiring post-editing abilities through research-informed practical tasks

Nataša Pavlović  
npavlovi@ffzg.hr  
*University of Zagreb*

Goranka Antunović  
gantunovic@ffzg.hr  
*University of Zagreb*

The integration of machine translation (MT) in professional workflows, facilitated by recent gains in quality achieved by neural MT systems, means that translators are increasingly required to do post-editing (PE) in addition to their more traditional tasks. PE requires skills that are similar to, but also different from those expected of translators working without MT and revisors of human translation.

This paper takes a minimalist view of the PE skill set, describing it in terms of three interrelated abilities – to identify errors, to distinguish between necessary and unnecessary edits, and to implement the edits appropriately – and examining it in the context of translator training. An experimental study is conducted with translation students untrained in PE who are asked to post-edit a text in conditions involving two variables: time pressure and source text access. The observed gaps in their skill set are taken as a starting point to develop ideas for the creation of practical PE tasks to be used in training. The paper may be of interest to translator educators looking to diversify their PE classes, whether in dedicated PE courses/modules or as part of other courses.

Keywords: *post-editing, machine translation, error identification, translator training, post-editing skill set.*

## 1. BACKGROUND AND AIM

Translators, especially those of non-creative texts, are increasingly required to incorporate machine translation (MT) in their workflows or they do so of their own accord to achieve greater productivity. The integration of MT in translation processes, spurred by the development of neural machine translation (NMT) and ensuing gains in MT quality, has had a considerable impact on the translators' tasks. They now post-edit at least some of the time, along with modifying human translation (HT) segments retrieved from translation memories (TM) or translating segments without the help of either MT or TM. Post-editing (PE) refers to improving MT-generated text in line with the requirements of a specific translation assignment. Identification of errors and their correction play a crucial role in PE, much like in self- and other-revision in translation processes not involving MT. Like translators and revisors, post-editors use internal and external resources to detect and solve problems related to the assignment at hand, but research suggests they may do so in slightly different ways (Nitzke, 2019). When compared to translation, PE has been found to be a "more passive activity", with longer pauses and less time devoted to text production (Green, Heer & Manning, 2013). Differences have been observed between PE and translation in keystroke patterns (Carl et al., 2011), indicating a different attention distribution in the two activities. In PE tasks, the source text (ST) tends to be consulted more frequently but also more briefly than in translation, with PE tasks understandably requiring more focus on the target text (TT) (Carl et al., 2011).

Unlike revision of HT, one type of post-editing, so-called light PE, settles for lower quality levels, its purpose being to ensure that the text is "comprehensible and accurate but need not be stylistically adequate" (ISO, 2017: 10). Revisors of HT are not normally briefed to disregard style (although in practice tight deadlines might push them to adopt such a stance). Secondly, the errors encountered in machine-produced texts tend to be of different kinds than those found in HT. The latter difference was stressed already in early studies of PE (Löffler-Laurian, 1985; McElhaney & Vasconcellos, 1986; both as cited in O'Brien, 2002: 101). With the recent switch to NMT systems, it has been suggested that the errors they make are less predictable than those produced by statistical MT and might be more difficult to spot, even when they seriously distort the meaning (Burchardt, 2017; Castilho et al., 2017). Predictable or otherwise, errors produced by MT systems, including neural ones, tend to be different from those encountered by revisors of fully human translations.

It has been observed based on the above distinctions that the PE task requires a skill set similar to, but also different from that expected in translation and revision of texts produced without MT. In attempts to define PE skills, complex, multicomponent models have been proposed (e.g. Rico & Torrejón, 2012), inspired by those constructed for translation competence and/or revision competence. Thus Nitzke, Hansen-Schirra & Canfora (2019) recently proposed a model which references a translation competence model (PACTE, 2003) and a revision competence model (Robert, Remael & Ureel, 2017), and sees PE as consisting of core and subsidiary competences. The former group comprises risk assessment, strategic, consulting and service competences, while the latter includes bilingual, extralinguistic, instrumental, research, revision, translation, MT and PE competence (the latter presumably in a narrower sense). As can be seen, some of these fully or partially overlap with translation and/or revision (sub) competences, while others can be regarded as specific to post-editing.<sup>1</sup> The relationship between translation competence and PE competence was considered already by Krings (2001: 174), who posited that “the more distant translation competence and post-editing competence are from each other, the likelier the expectation that good translators are not automatically good post-editors and vice versa”. Research has suggested that translation experience may benefit PE performance, up to a point, but also act as hindrance (de Almeida & O’Brien, 2010).

The broad view of PE taken in multicomponent competence models is certainly desirable, as it highlights the complexity of PE and the various types of skills and knowledge required of post-editors today. On the other hand, it might be equally useful to adopt a minimalist approach, reminiscent of Anthony Pym’s (2003: 489) definition of translation competence as “the ability to generate a series of more than one viable target text [...] for a pertinent source text [and] to select only one viable TT from this series, quickly and with justified confidence”. Such a minimalist view of PE, focusing on the task itself rather than on other processes surrounding it, would enable us to take a closer look at the three interrelated abilities (drawing on de Almeida & O’Brien, 2010) that form the essence of PE, namely the ability to:

1. identify errors in the MT output under different conditions;
2. distinguish between edits that are necessary and those that are not, in a particular assignment; and

---

<sup>1</sup> On the blurring of lines between translation, revision, editing and post-editing, see Jakobsen (2019).

3. implement the edits appropriately, by finding a more suitable solution and by not introducing new errors in the process.

In the multicomponent model mentioned above (Nitzke et al., 2019: 248-250), ability 1 is subsumed under ‘post-editing competence’, ‘machine translation competence’ and, up to a point, ‘bilingual competence’; ability 2 appears under ‘revision competence’ and ‘strategic competence’, while ability 3 might be implied in ‘bilingual competence’, ‘extralinguistic competence’, ‘research competence’, ‘revision competence’ and ‘translation competence’.

Ability 2, related to finding a compromise between the required quality and speed, is particularly relevant for work under time constraints, which in professional practice means a vast majority of the assignments. Post-editors who do not possess it will in the best-case scenario waste a lot of time, defying a key purpose of using MT, and in the worst-case scenario deliver suboptimal quality. Experience in translation does not necessarily help; it has been observed that experienced translators may introduce more unnecessary edits than inexperienced ones (de Almeida & O’Brien, 2010). The number of unnecessary edits, also reported by other authors (e.g. Aranberri, 2017), has been found to decrease only as experience in *post-editing* increases (de Almeida, 2013). The discernment of necessary changes is an ability PE shares with HT revision, where it is stressed as one of the key principles (Robert et al., 2017: 114, 115). Even for ‘full PE’, aiming to achieve human translation quality, the recommendation is to use “as much of the MT output as possible” (ISO, 2017: 8).

All three of the above abilities are especially challenged in monolingual PE, i.e. post-editing without access to the ST. In such practices, which seem unfortunate but nevertheless exist, post-editors must be able to identify errors solely based on the (lack of) cohesion and coherence in the MT output, and to correct them appropriately, unaided by the ST. Koponen & Salmi (2015) studied monolingual PE in the training context to see if meaning could be derived despite errors in the MT output. Their participants, 48 students, were able to arrive at correct meaning in about half of the cases. Other studies suggest that monolingual PE is more susceptible than bilingual PE to oversight of semantic errors (Čulo et al., 2014; Mitchell, Roturier & O’Brien, 2013; Nitzke 2016). Nitzke (2019) posits that problem-solving translation strategies may be transferred to bilingual PE, but not to monolingual, as there is no ST to consult and simply translate the problematic part anew.

Given all of the above, most translator educators today would probably agree that students of (specialized) translation should be trained to post-ed-

it, whether in dedicated courses/modules or as part of courses/modules dedicated to translation and/or HT revision. Since O'Brien's (2002) initial proposal, many universities have introduced such content in their translation programmes (see e.g. Doherty & Kenny, 2014; Flanagan & Christensen, 2014; Guerberof Arenas & Moorkens, 2019; Koponen 2015). Recommendations on PE training show a broad consensus on which topics should be covered: knowledge of MT systems (types, how they work, what kinds of errors can be expected and why), MT assessment (human and automatic methods; error classifications), PE guidelines (light vs. full PE, levels of quality), pre-editing, as well as basic programming and perhaps terminology management (Guerberof Arenas & Moorkens, 2019; O'Brien, 2002).

While acknowledging that practical experience in PE is essential for the acquisition of PE skills, existing literature does not provide sufficient detail on how PE training tasks might be designed and conducted. This paper attempts to contribute to filling that gap by reporting on an empirical study carried out with translation student participants untrained in PE, and by proposing ways in which the results of that study can inform task design. In the study, we focused on the three abilities mentioned above. We wanted to examine to what extent students might already possess them – by virtue of being trained in translation – before receiving any training in PE, and to measure their lack of abilities, i.e. their 'negative skills' in Pym's (2013: 497) sense of the term. We expected that learning about their performance in an experiment, particularly about the shortcomings and challenges they face while post-editing, might help inform the creation of practical PE tasks to be used in training.

## 2. RESEARCH QUESTIONS

In this study, we posed three groups of research questions related to the three abilities mentioned above (cf. de Almeida & O'Brien, 2010), which we consider crucial for PE:

1. To what extent are translation students untrained in PE able to identify errors in MT output? How does time pressure affect their error identification? What about access to the ST?
2. To what extent are translation students untrained in PE able to distinguish between necessary and unnecessary edits in a specific PE task? Do they tend to overedit or underedit?
3. To what extent are translation students untrained in PE able to implement edits appropriately? Are their solutions more suitable to the

task at hand than the original MT solutions? Do they introduce new errors while attempting to correct existing ones?

### 3. METHODOLOGY

#### 3.1. Research design and participants

To answer the research questions listed above, we conducted a post-editing experiment with translation students untrained in PE. The participants were 49 translation track students at the University of Zagreb, with English as their L2 and Croatian as L1, who were about to complete their first or third semester of the MA programme. Data obtained from five participants were discarded as outliers, using the interquartile range rule, which brought the total number of participants to 44.

The PE experiment was conducted in two groups, in a computer room where the participants normally had classes, using Microsoft Word. In each group, the experiment was divided into three subsequent phases, with data collected after each phase (Table 1). Both groups worked on the same text for 10 minutes in Phase 1, an additional 12 minutes in Phase 2, and 12 more minutes in Phase 3. Group A (n=19) had access to the ST from the beginning, working bilingually over all three phases, while Group B (n=25) set out post-editing monolingually and got access to the ST only in Phase 3. This allowed us to investigate the impact of two variables – time pressure and access to the ST – on error identification and editing.

Table 1. Group A and Group B sequences, with changes in condition bolded

	<i>Phase 1</i>	<i>Phase 2</i>	<i>Phase 3</i>
Group A	with ST time pressure	with ST <b>no time pressure</b>	with ST no time pressure
Group B	without ST time pressure	without ST <b>no time pressure</b>	<b>with ST</b> no time pressure

After a short warm-up assignment, the participants were told that they would be post-editing a Croatian machine translation of an English text (see 3.2.) to publishable quality. Since the participants had done *translation* assignments involving texts similar to the ST used in this study and for the same purpose, they were familiar with the quality level required. They were told they would have 10 minutes to turn on the Track Changes option and correct all the errors they noticed, but only those which they were certain

were errors. Ten minutes was assessed to constitute time pressure for these participants, while allowing them to reach the end of the text, which they were explicitly asked to do, and which all of them accomplished. When the time elapsed, the participants were asked to save the document and upload it to the e-learning platform they regularly used. At the beginning of Phase 2 they were instructed to reopen the document and resume the same PE task for an additional 12 minutes. As they had already gone through the whole text once in Phase 1, 12 more minutes was considered enough additional time for the total of 22 minutes to be treated as a no-pressure condition. When the time expired, the participants uploaded the post-edited translation to a different folder. In Phase 3 they repeated the procedure for 12 more minutes in order to complete the same task, with Group B now gaining access to the ST. We could observe that in that phase even the students in the latter group had ample time to finalize the task.

### **3.2. Material**

The English ST used in the experiment, 288 words long, was a ‘plain language summary’ of a medical review (Farooq et al., 2017), published by Cochrane, an organisation providing systematic reviews of research evidence to assist informed decision making about health. We used a 253 word long Croatian translation produced by Google’s NMT engine in January 2018, shortly before the experiment took place. All participants had translated such texts as part of their coursework, and this one was chosen because it could be considered not to contain medical terminology unknown to students. That consideration was important because we did not allow the participants to use external resources so that we would be able to control the time pressure and ST access variables.

### **3.3. Analysis**

The analysis focused on the three abilities discussed in Section 1 above: 1) to identify errors (with and without time pressure; with and without the ST), 2) to distinguish between necessary and unnecessary edits, and 3) to implement edits that improve the MT.

To investigate ability 1, we first had to ascertain the errors ourselves. This meant deciding which solutions in the MT output we would consider genuinely erroneous, that is, unacceptable if the translation were to be published. Our point of departure were the edits introduced by the participants in any phase of the experiment. We assessed all the MT solutions that



they edited and then selected 41, which in both our opinions undoubtedly required editing and which we would label as ‘indisputable errors’. An indisputable error was considered to have been *identified* if a participant had tried to edit it in any way, whether successfully or not. We counted all such attempts, in all three phases of the experiment, based on Track Changes data. We termed instances of overlooked indisputable errors ‘missed necessary edits’<sup>2</sup>.

We then compared how many errors all the participants identified when working under time pressure (Phase 1 of the experiment, 10 minutes) and without it (Phases 1 and 2 combined, 22 minutes). We tested the difference between the two conditions for statistical significance using the paired samples t-test. Phase 3 was not included in this stage of the analysis since it also involved a change in the ST access variable. Subsequently we compared the performance of the two groups with respect to error identification using the independent samples t-test. We compared the two groups’ scores at the end of Phase 2 (bilingual vs. monolingual condition), and at the end of Phase 3 (bilingual PE throughout the whole process vs. monolingual PE followed by bilingual PE).

In addition to the number of errors identified, we also looked at error type, labelling the errors as either semantic or non-semantic. In the literature on MT (e.g. Klubička et al., 2017), these two error categories are usually termed accuracy-related and fluency-related respectively, the latter group encompassing errors that affect such aspects of a translation as spelling, grammar or register. Of the 41 indisputable errors in the MT output we used in the experiment, 17 were semantic and 24 were non-semantic. We wanted to see if either type had proved more difficult to identify in each of the experimental conditions.

Ability 2 was tested by considering not only the edits related to indisputable errors (‘necessary edits’) but also all the other (‘unnecessary’) edits. The proportion of the latter type in the total number of edits made by the participants in the whole process was calculated.

Finally, in order to test ability 3, we assessed all the edits to see whether or not they improved the initial MT solution. Necessary edits could thus provide an acceptable solution or could be erroneous themselves; they could also be partial or even introduce a new error. Among the unnecessary edits we distinguished between four different types: those that introduced

---

<sup>2</sup> De Almeida (2013) calls them ‘essential changes not implemented’ and Rigouts Terryn et al. (2017) ‘missed necessary changes’.



a solution that was more appropriate than the initial MT one, those introducing a less appropriate solution, those that could be considered neutral i.e. the new solution was neither obviously better nor worse than the MT solution, and those resulting in a new error.

In addition to the described quantitative analyses, we also analysed the participants' performance qualitatively, focusing in particular on the errors that eluded identification and those that, once spotted, nevertheless proved difficult to edit.

## 4. RESULTS AND DISCUSSION

This section is divided into three subsections, each presenting the results related to one of the three PE abilities investigated in this study. In each subsection, the presentation of the results is followed by a discussion of the implications for training, ending with concrete suggestions for the creation of practical PE tasks.

### 4.1. The ability to identify errors under different conditions<sup>3</sup>

At the end of Phase 3, that is, with both groups having gained access to the ST and all participants working with no time pressure whatsoever (all had finished the assignment and stopped working), the overall percentage of identified errors was 70 per cent. On average, 29 errors of the 41 were identified per participant ( $SD=4$ ; range 21–37). This result shows that translation students are able to identify a majority of the errors in MT output even without previous training in PE, simply based on their language and translation competences. However, 30 per cent of the errors remain undetected: on average, 12 necessary edits (range: 4–20) are overlooked in a 253 word long text. This would clearly be unacceptable in a professional context and suggests that error identification ability is a skill to be trained rather than assumed to be resulting from translation competence or intuition.

#### 4.1.1. Error identification with and without time pressure

To gauge the effect of time pressure on error identification we compared the data collected after Phase 1 (10 minutes) and after Phase 2 (22 minutes altogether) for all participants. In Phase 1, i.e. with time pressure, the par-

---

<sup>3</sup> The results from this section are reported in more detail in Antunović & Pavlović (2019). Here we summarize them and discuss possible training implications and ideas. Results in 4.2. and 4.3. are presented for the first time.

ticipants identified 39 per cent of all indisputable errors, while in the two phases combined, i.e. without substantial time pressure, the figure rose to 60 per cent. A paired samples t-test confirmed that the difference, evident at first glance, was indeed statistically significant ( $p < .001$ ).

This clearly shows that time pressure significantly reduces the ability to identify errors among translation students untrained in PE. Even the most glaring semantic error – the word *trial*, machine-translated to mean legal trial (*sudjenje*) rather than medical trial (*ispitivanje, pokus, studija*) – which was identified by almost everyone by the end of the PE process, was missed under time pressure by 40 per cent of the participants, and access to the ST did not make any difference in this respect.

However, even with sufficient time the students did not get close to 100 per cent error identification. We believe that, while students should be trained to cope with time pressure, it might be advisable for such exercises to be more prevalent in later stages of training, when the trainees have considerably improved their error identification ability without time constraints.

#### 4.1.2. Error identification in bilingual vs. monolingual PE

As we explained in 3.1., Group A had access to the ST from the beginning of the experiment, while the ST was made available to Group B only in Phase 3. Over the first two phases, with differing ST access variable, Group B identified slightly *more* errors. The difference was slim – 24.6 errors identified on average (60 per cent), as compared to 23.8 in Group A (58 per cent) – and proved not to be statistically significant ( $p = 0.761$ ). Pending further research, we might speculate that Group B's focus on the MT output, the only text they worked with, was perhaps more acute due to undivided attention. Nevertheless, the results obtained in this study do not yet warrant any firm conclusion as to the influence of ST access on the number of errors identified.

In Phase 3, when Group B got access to the ST, they identified slightly fewer errors than the group working with the ST throughout. A possible explanation is that their focus was now more directed to reading the ST, which divided their attention and weakened their concentration. The values are, however, too small to draw any firm conclusions. Over the whole process, Group B fared a bit better in their error identification, when we look at the *number* of identified errors alone.

In the next stage of the analysis, we examined the *types* of errors – semantic or non-semantic errors (the latter type encompassing, for example,

orthographic, morphosyntactic or phraseological errors) – identified in the different conditions. Time pressure did not prove to affect the type of error identified, but ST access did prove to make a difference (Table 2). In the first two phases, Group B, working without the ST, identified a higher proportion of non-semantic errors (67 per cent of all such errors) than semantic ones (50 per cent) and more non-semantic errors than were identified by Group A (55 per cent). Conversely, Group A identified a higher proportion of semantic errors (62 per cent) than non-semantic ones and more semantic errors than Group B did. The small total number of errors of each type (24 non-semantic, 17 semantic) was not propitious for statistical analysis but the percentages are different enough to suggest that this may not be accidental. It stands to reason that the group working without the ST would have more difficulty identifying meaning-related errors than the other group, as was also shown in previous research (Čulo et al., 2014; Mitchell, Roturier & O'Brien, 2013; Nitzke, 2016). It is in fact remarkable that the participants in the monolingual condition identified half of semantic errors, which is a result similar to the one in Koponen and Salmi (2015).

Table 2. Error types identified by the two groups in various phases of the experiment

	Ph. 1+2	Ph. 1+2	Ph. 3	Ph. 3	Ph. 1+2+3	Ph. 1+2+3
	non-semantic	semantic	non-semantic	semantic	non-semantic	semantic
<b>Group A</b>	55%	62%	9%	10%	64%	72%
<b>Group B</b>	67%	50%	4%	19%	71%	69%

In Phase 3, when Group B was given access to the ST, they identified almost double the number of semantic errors as Group A did in that phase. They still did not quite catch up with Group A but they did come very close to the other group's result over the whole process (phases 1, 2 and 3 taken together). Further, the number of semantic errors Group B identified in that phase was several times higher than the number of non-semantic errors that they noticed. Over the whole process (two rightmost columns in Table 2), Group B's result was nevertheless slightly lower than Group A's when it comes to semantic errors (69 and 72 per cent respectively) and higher with regard to non-semantic errors (71 and 64 per cent respectively). The group in the bilingual condition presumably divided their attention between the ST and the MT, identifying more semantic errors than the other group, but paying the price in terms of poorer identification of non-semantic ones.

As we mentioned earlier, the fluency of the NMT output can disguise semantic errors, making them difficult to spot at first glance. The machine

translation we used in our experiment had one such erroneous but “convincing” solution. The ST was about (a lack of) guidelines for health professionals on communicating a diagnosis of mental illness to patients, and the pertinent sentence said: “At the moment, there is no information or evidence regarding how best to disclose a diagnosis of schizophrenia.” In the MT output, the word *disclose* was mistranslated as *otkriti* to mean “discover a diagnosis”, but the target sentence was otherwise error-free. Outside of this text, such a sentence might make sense. Under time pressure, more than a half (57 per cent) of Group B, working without the ST, overlooked this error, and almost a half (48 per cent) of Group A did so despite having access to the ST. With more time in Phase 2, the gap widened: in the group working bilingually the error remained unidentified by 19 per cent of the participants, and in that working monolingually by 46 per cent. When they got access to the ST in Phase 3, the latter group still failed to notice this error in 36 per cent of the cases, and in Group A, with ST access throughout the process, the error remained undetected until the end by 14 per cent of the participants.

Although non-semantic errors are undoubtedly important in PE, even in its light variety, semantic errors are critical in that they can jeopardize the key message of the text. If future research were to take up this issue, it would be useful to see which of the following scenarios might lead to best identification of errors: single-stage bilingual PE; two-stage PE, with monolingual followed by bilingual; or two-stage PE, with bilingual followed by monolingual.

If confirmed by future research, what might be the implications for PE training of our results related to the impact of ST access? Several ideas come to mind. First, monolingual PE tasks could be used to improve trainees’ focus on non-semantic error identification. Such tasks should be relatively easy as they can mobilize the students’ target language competence as well as experience with self- and other HT revision and proofreading. Secondly, and perhaps more interestingly, monolingual exercises could also aim to improve trainees’ ability to detect *semantic* errors solely on the basis of the MT, by focusing on the sections that lack coherence. Honing this ability could help future post-editors to identify semantic errors more easily even when working *with* the ST. In this respect exercises can also raise awareness of typical NMT errors, esp. “sneaky” ones such as that mentioned above.

Finally, tasks that combine mono- and bilingual PE of the same text in different order could be done to allow students to experiment, e.g. work monolingually first, focusing on non-semantic errors, followed by bilingual

PE, during which the ST is consulted to check for sense; or the other way round. As with translation, there might be 'PE styles' best suited to individual post-editors, and this type of exercise would allow students to find their own preferred style.

Among the 'different conditions' from the heading of this section, another constraint that is increasingly present in translation workflows involving MT deserves to be mentioned, namely the fact that segments requiring PE are often interspersed with segments requiring revision of translation memory results or even with translation from scratch (if there is no match in the memory and the quality of a machine-translated segment is judged by the translator to be too low to be helpful). We did not focus on this variable in our research so we can only share our belief that including such mixed scenarios in practical PE training seems highly desirable.

#### **4.2. The ability to distinguish between necessary and unnecessary edits**

We now shift our attention from the edits of indisputable errors to *all* edits made during the PE experiment. We expected our participants, untrained post-editors, to make many unnecessary edits, mostly with no impact on the quality of the translation. This proved to be true: almost 40 per cent of all edits introduced by the participants were in our opinion unnecessary (841 out of the total of 2123), even for publication purposes. This would suggest that translation students without PE training are prone to over-editing, even as they underedit, as reported in 4.1. Since we noticed these two parallel trends, we wanted to see if they correlated, that is, if the focus on unnecessary edits might have distracted the participants and prevented them from identifying indisputable errors. The result of Pearson's correlation test did not, however, confirm our expectation; in fact, it indicated a weak negative correlation ( $r = -0.318$ ;  $p = 0.035$ ) between the number of unnecessary edits and the number of missed necessary edits. This means that at least some of those participants who made more unnecessary edits also made more necessary ones. That is similar to a previous finding about experienced translators post-editing more accurately but also making more unnecessary changes (de Almeida & O'Brien, 2010). This result would suggest that the ability to identify errors and the ability to distinguish between necessary and unnecessary edits, although connected, are separate abilities and both deserve to be addressed in training.

The ability to distinguish what needs to be edited in MT output from what is an acceptable, although perhaps not the best, solution is apparently not something that translation trainees (or experienced translators) intuitively possess to high degrees. Previous or concurrent training in HT revision would likely be beneficial, as this ability is something PE shares with other types of revision. Our students had had some experience revising their peers' and their own work, where they had faced similar insecurities as to what constitutes a genuine error and what might be an acceptable solution even if not to their taste. This problem is inherent in editing and revision in general: How much to change? Learning to work with what you have is an important principle that trainees should acquire. In this respect, exercises in PE and HT revision, although requiring a slightly different focus, could be devised to cross-fertilize each other.

For PE, exercises should specifically address the fact that not all PE is done to publishable quality, and that the point of using MT in a translation process is to increase productivity. To this end awareness of PE guidelines might be helpful, but may not be enough, as they tend to be too general, do not consider the specificities of target contexts, and may not always be clear (cf. Flanagan and Christensen, 2014). Practical tasks involving PE to different levels of quality would be a useful way for students to learn the skill of recognizing different levels of acceptability and setting apart the unacceptable from the acceptable-for-the-purpose. For example, a hypothetical Task 1 could be accompanied by a brief such as the following: "The client, a retailer, has requested an informative translation of a washing machine user manual, which will help them decide whether to include the product in their assortment. Use the MT output the client has sent to produce a translation for information purposes." This could be followed by Task 2, with a brief such as this: "Use the MT output to produce a publishable translation of a washing machine user manual that will accompany the product documentation. Bear in mind that poorly translated user manuals negatively affect sales and may have legal repercussions for your client." Different time limits can be provided for Task 1 and Task 2. The tasks could be followed by discussion, with the decision-making framed in terms of best speed-to-quality ratio.

Exercises such as these require appropriate feedback to help students advance in their ability to distinguish between the necessary and unnecessary edits.<sup>4</sup> To make feedback provision easier for the teachers it might be useful

---

<sup>4</sup> That this may be challenging for trainers, particularly if they themselves lack training and/or experience in PE, is not an irrelevant circumstance, but it falls outside the scope of this article.

to compile and share learning corpora consisting of ST – raw MT – lightly post-edited MT – fully post-edited MT, with comments and explanations where appropriate. Trainees could use them for self-study, perhaps with increasing overall time pressure, to supplement coursework, since it is unrealistic to expect that any course could dedicate enough time to practical tasks for trainees to gain the necessary confidence (as stressed already in O'Brien, 2002).

### **4.3. The ability to implement edits appropriately**

In the final stage of our analysis, we assessed all the edits introduced by participants to see if they had a positive, negative or neutral effect on the quality of the translation's final version.

With regard to necessary edits, some 12 per cent failed to improve the erroneous solution, improved it only partially or even introduced a new error while fixing the existing one. This might not seem a bad percentage if it did not combine with the unsatisfactory score of error identification reported in 4.1. Regarding unnecessary edits, a majority (68 per cent) resulted in a solution that was neither evidently better nor worse than the previous one, but merely compromised the productivity of the process. Some 19 per cent of the edits, albeit unnecessary, did improve the existing MT solution, while 10 per cent were in fact less appropriate than the original solution. In three per cent of the cases, an unnecessary edit introduced a new error. When both necessary and unnecessary edits are considered together, some 61 per cent improved the quality of the translated text; around 27 per cent neither improved nor made it worse, and the remaining 12 per cent resulted in deterioration.

This leads us to conclusion that untrained post-editors not only underedit and overedit but also, in the words of one of our students, "wrongedit". The reasons for this are various. Some of the added errors are typos that result from the students' carelessness during the editing process. Further, many of the unnecessary edits that made the translation neither better nor worse were stylistic changes resulting from the participants' insecurity regarding the target language (their L1). They would cross out a perfectly acceptable word and write an equally acceptable synonym, forgetting at the same time to ensure grammatical congruence in all the relevant places. These types of added errors are relatively easy to address in training, as students can be shown the different ways of viewing the tracked changes and invited to experiment with the view that suits them best. Trainees can also



be guided to incorporate proofreading and spellchecking into their routine to see for themselves if these additional steps improve the quality of their final product.

In some cases, the participants edited wrongly because they did not trust the MT system for terminology and did not have access to external resources to check it. This was a limitation of our research design; as already explained, we chose the particular text for the experiment because we expected the students to be familiar with all the terminology it contained, having dealt with similar texts in their translation classes. This proved not to be the case with the term *randomised controlled trials*, which was correctly translated by the MT system as *randomizirani kontrolirani pokusi*, yet some of the participants thought the translation was erroneous. A healthy dose of mistrust towards MT is highly advisable, but trainees also need to learn when to trust it (cf. Nitzke et al., 2019: 249; Pym, 2013: 495ff), or rather which system to trust for what type of text and domain. Experimenting with various freely available systems, and in different language combinations and directions, should help the students learn what they can expect from each system and to what extent they can trust it (cf. Guerberof Arenas & Moorkens, 2019: 224). Even more importantly, such experiments should turn familiarization with different MT engines and their typical strengths and weaknesses into an important habit.

Another reason for the wrong edits was undoubtedly the lack of ST in Group B in the first two phases of the experiment. Some participants noticed a semantic error but were unable to correct it appropriately without the ST, and later, when the ST became available, failed to return to the problematic part. For example, the ST mentions a *study-based register*, which was machine-translated as *studija bazirana na studiji* (“study-based study”). This type of error, whereby the MT system repeats a word when unsure what else to do, is rather typical of NMT. In the group post-editing without the ST, understandably, none of the participants were able to fix this error in Phase 1 and only one was in Phase 2. However, even at the end of Phase 3, during which these participants had access to the ST and could see what it said, the error was left uncorrected by as many as 21 per cent of the participants (as compared to only 5 per cent of the participants in the group which had the ST from the start).

Monolingual tasks of the type mentioned in 4.1.2., in addition to improving error identification, can be used to help trainees enhance their sense-making abilities by filling in gaps and guessing what the ST might have said, by relying on the MT output and their knowledge of the world.

Students quickly learn by experience that backtranslation is a useful strategy (cf. Nitzke, 2019) in this respect, and generally find this type of task both challenging and entertaining (variations include getting groups of students to compete in correcting the errors accurately without having access to the ST). Such exercises in sense-making could have the added bonus of helping the trainees learn how to improve the coherence of a text even when post-editing bilingually (or translating).

Occasionally, the participants made erroneous edits because they did not understand the text correctly even if they had access to the ST. An example is the expression *long-term outlook*, which was used in the text to mean “prognosis”. The MT wording (*dugoročni pogled*) misled some of the participants into thinking that the phrase referred to the patient’s outlook on life. The error in the MT may have contributed to their confusion, but we cannot know whether they would have made the same kind of mistake translating without the help of MT. This brings us to a crucial observation, namely that our participants are still developing their translation competence, and in some cases also their SL and TL competences. This is a point at which PE competence and translation competence overlap to a high degree. Choosing the most appropriate from among multiple solutions in PE is similar to choosing the most appropriate solution in translation, with one important difference: in PE, one solution is already offered, that in the MT output. Having something to work with can be helpful, as it can generate other ideas, but it can also block creativity and lead one astray. Other than that, the same considerations – purpose of the translation, text type, target readers, client requirements – are involved. The trainees’ underlying translation and language competences need to be improved along with, and as a basis for, the trainees’ PE skills as such.

## 5. FINAL REMARKS

While the literature, acknowledging today’s professional practice, leaves little doubt as to the need to include the acquisition of PE competence in translator training, the trainers may feel insecure and lack guidance on how to go about the task. Facing the challenge ourselves, we thought it would be useful to first get a better understanding of what our students may or may not be able to do before receiving any training in PE, owing to their previous training and experience in translation. On that basis we would then try to determine which elements of PE competence we needed to target and what practical tasks could help students develop them. We therefore

decided to conduct an experiment that would indicate how skilful the students already were with regard to the three interrelated abilities that may be seen as the core components of a minimalist model of PE competence, namely the ability to identify errors in MT output, the ability to distinguish between edits that really need to be made in a particular assignment from those that are not necessary, and the ability to implement such edits that will genuinely correct the unacceptable MT solutions, without introducing any new errors along the way.

The minimalist view of PE as comprising three key abilities, which we have embraced here, should not be taken to exclude all the other knowledge and skills covered in broader multicomponent PE competence models mentioned in Section 1. Likewise, the acquisition of the three key abilities need not be attempted in isolation from those related skills. Collaborative authentic translation and/or localization projects in which the whole competence package is practiced present highly desirable learning opportunities and should be incorporated in training programmes whenever possible. However, such projects may not be feasible in all contexts and, despite their many strengths, they also have their weaknesses (Pavlović, 2016). That is why we believe there is also a place in translator/post-editor training for more focused exercises, such as the ones we have suggested in this paper, targeting specifically one, two or all three essential PE abilities. In fact, we believe that the best scenario, where possible, would be for the two types of learning activities – larger projects and targeted exercises – to go hand in hand and complement each other.

## REFERENCES

- Antunović, G. & Pavlović, N. (2019) Redaktura strojnih prijevoda – sve važniji prevoditeljski zadatak. In Matešić, M. & Vlastelić, A. (Eds.) *Jezik i um*. Zagreb: Srednja Europa & HDPL, 147-167.
- Aranberri, N. (2017) What do Professional Translators do when Post-Editing for the First Time? First Insight into the Spanish-Basque Language Pair, *Hermes – Journal of Language and Communication in Business*, 56, 89-110.
- Burchardt, A. (2017) Comparing Errors: Neural MT vs. Traditional Phrase-Based and Rule-Based MT. <https://www.gala-global.org/knowledge-center/professional-development/articles/comparing-errors-neural-mt-vs-traditional-phrase> (5 April 2021)
- Carl, M., Dragsted, B., Elming, J., Hardt, D. & Jakobsen, A. L. (2011) The Process of Post-Editing: a Pilot Study, *Copenhagen Studies in Language*, 41, 131-142.
- Castilho, S., Moorkens, J., Gaspari, F., Iacer, C., Tinsley, J. & Way, A. (2017) Is Neural Machine Translation the New State of the Art?. *The Prague Bulletin of Mathematical Linguistics*, 108, 109-120.

- Čulo, O., Gutermuth, S., Hansen-Schirra, S. & Nitzke, J. (2014) The Influence of Post-editing on Translation Strategies. In O'Brien, Sh., Winther Balling, L., Carl, M., Simard, M. & Specia, L. (Eds.) *Post-editing of Machine Translation: Processes and Applications*. Newcastle Upon Tyne: Cambridge Scholars Publishing, 200–218.
- de Almeida, G. (2013) *Translating the Post-editor: An Investigation of Post-editing Changes and Correlations with Professional Experience across two Romance Languages*. (Doctoral dissertation). Dublin City University. <http://doras.dcu.ie/17732/> (18 March 2021)
- de Almeida, G. & O'Brien, S. (2010) Analysing Post-Editing Performance: Correlations with Years of Translation Experience. In Hansen, V. & Yvon, F. (Eds.) *Proceedings of the 14th annual conference of the European Association for Machine Translation*. <https://www.aclweb.org/anthology/2010.eamt-1.19.pdf> (5 April 2021)
- Doherty, S. & Kenny, D. (2014) The Design and Evaluation of a Statistical Machine Translation Syllabus for Translation Students. *The Interpreter and Translator Trainer*, 8 (2), 295-315.
- Farooq, S., Johal, R., Ziff, C. & Farooq, N. (2017) Different communication strategies for disclosing a diagnosis of schizophrenia and related disorders, Plain language summary title: Sharing information about the diagnosis and outcomes of schizophrenia: evidence from well-designed clinical trials. In *Cochrane Database of Systematic Reviews* 10: CD011707. <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011707.pub2/full#CD011707-abs-0003> (16 March 2021)
- Flanagan, M. & Paulsen Christensen, T. (2014) Testing Postediting Guidelines: How Translation Trainees Interpret them and how to Tailor them for Translator Training Purposes. *The Interpreter and Translator Trainer*, 8 (2), 257-275.
- Green, S., Heer, J. & Manning, C. (2013) The efficacy of human post-editing for language translation. In *Proceedings of ACM Human Factors in Computing Systems (CHI)*, 439–448, <http://idl.cs.washington.edu/files/2013-PostEditing-CHI.pdf> (5 April 2021)
- Guerberof Arenas, A. & Moorkens, J. (2019) Machine translation and post-editing training as part of a master's programme. *The Journal of Specialized Translation*, 31, 217-38.
- ISO [International Standards Organization] (2017) *International Standard 18587, Translation services – Post-editing of machine translation output – Requirements*. Geneva: ISO.
- Jakobsen, A. L. (2019) Moving translation, revision and post-editing boundaries. In Dam, H. V., Nisbeth Brøgger, M. & Korning, K. (Eds.) *Moving Boundaries in Translation Studies*. London & New York: Routledge, 64-80.
- Klubička, F., Toral, A. & Sanchez-Cartagena, V. (2017) Fine-Grained Human Evaluation of Neural Versus Phrase-Based Machine Translation. *The Prague Bulletin of Mathematical Linguistics*, 108, 121-132.
- Koponen, M. & Salmi, L. (2015) On the Correctness of Machine Translation: A Machine Translation Post-editing Task. *Journal of Specialised Translation*, 23, 118-136.
- Koponen, M. (2015) How to Teach Machine Translation Post-editing? Experiences from a Post-editing Course. In O'Brien, S. & Simard, M. (Eds.) *Proceedings of 4th Workshop on Post-Editing Technology and Practice*. 2-15, [https://amtaweb.org/wp-content/uploads/2015/10/MTSummitXV\\_WPTP4Proceedings.pdf](https://amtaweb.org/wp-content/uploads/2015/10/MTSummitXV_WPTP4Proceedings.pdf) (18 March 2021)
- Krings, H. P. (2001) *Repairing Texts. Empirical Investigations of Machine Translation Post-Editing Processes*, Koby, G.S. et al. (Ed. and Trans.). Kent, Ohio and London. The Kent State University Press.

- Mitchell, L., Roturier, J. & O'Brien, S. (2013) Community-based Post-editing of Machine-translated Content: Monolingual vs. Bilingual. In O'Brien, Sh., Simard, M. & Specia, L. (Eds.) *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice*. Nice: EAMT, 35-43.
- Nitzke, J., Hansen-Schirra, S. & Canfora, C. (2019) Risk management and post-editing competence. *The Journal of Specialized Translation*, 31, 239-259.
- Nitzke, J. (2016) Monolingual post-editing: An exploratory study on research behaviour and target text quality. In Hansen-Schirra, S. & Gruzca, S. (Eds.) *Eyetracking and Applied Linguistics*. Berlin: Language Science Press, 83-108.
- Nitzke, J. (2019) *Problem solving activities in post-editing and translation from scratch. A multi-method study (Translation and Multilingual Natural Language Processing 12)*. Berlin: Language Science Press.
- O'Brien, S. (2002) Teaching Post-editing: A Proposal for Course Content. In *Proceedings of 6th EAMT Workshop Teaching Machine Translation*. Manchester: EAMT, 99-106.
- PACTE (2003) Building a Translation Competence Model. In Alves, F. (Ed.) *Triangulating Translation: Perspectives in Process Oriented Research*. Amsterdam: Benjamins, 43-66.
- Pavlović, N. (2016) Stocktaking and prospects: Five years of the translation track in the University of Zagreb's Department of English. In Zovko Dinković, I. & Mihaljević Djigunović, J. (Eds.) *English Studies from Archives to Prospects 2 – Linguistics and Applied Linguistics*. Newcastle upon Tyne: Cambridge Scholars Publishing, 165-181.
- Pym, A. (2003) Redefining Translation Competence in an Electronic Age. In Defence of a Minimalist Approach. *Meta* 48 (4), 481-497.
- Pym, A. (2013) Translation Skill-Sets in a Machine-Translation Age. *Meta*, 58 (3), 487-503.
- Rico, C. & Torrejón, E. (2012) Skills and Profile of the New Role of the Translator as MT Post-editor. *Revista Tradumàtica*, 10, 166-178.
- Rigouts Terry, A., Robert, I. S., Ureel, J. J. J., Remael, A. & Hanouille, S. (2017) Conceptualizing translation revision competence: a pilot study on the acquisition of the knowledge about revision and strategic subcompetences. *Across Languages and Cultures*, 18 (1), 1-27.
- Robert, I. S., Remael, A. & Ureel, J. J. J. (2017) Towards a model of translation revision competence. *The Interpreter and Translator Trainer*, 11(1), 1-19.

## **Poučavanje vještine redakture strojnih prijevoda utemeljeno na rezultatima istraživanja**

Nataša Pavlović  
npavlovi@ffzg.hr  
*Sveučilište u Zagrebu*

Goranka Antunović  
gantunovic@ffzg.hr  
*Sveučilište u Zagrebu*

Poboljšanje kvalitete strojnih prijevoda (SP) omogućeno uvođenjem neuronskih sustava potaklo je uključivanje SP-a u profesionalne prijevodne procese, uslijed čega prevoditelji sve češće, uz svoje uobičajene zadatke, obavljaju i redakturu SP-a. Taj zadatak iziskuje skup vještina donekle različit od onoga koji se inače očekuje od prevoditelja, kao i od onoga koji moraju posjedovati redaktori prijevoda koje je napravio čovjek bez pomoći SP-a.

U ovom se radu zauzima minimalistički pristup vještini redakture SP-a te se ona, u kontekstu obrazovanja prevoditelja, promatra u smislu triju međusobno povezanih sposobnosti – sposobnosti uočavanja pogrešaka, sposobnosti razlikovanja između nužnih i preferencijalnih preinaka te sposobnosti provođenja preinaka na primjeren način. Provedeno je eksperimentalno istraživanje u kojem su sudjelovali studenti prevoditeljskoga smjera diplomskog studija prije početka obuke iz redakture SP-a. Od njih se tražilo da redigiraju SP u dvjema eksperimentalnim situacijama, u kojima su nezavisne varijable bile vremenski pritisak i mogućnost pristupa izvornome tekstu.

Rezultati istraživanja, osobito uočeni nedostaci u vještinama sudionika, poslužili su kao polazišna točka u razmišljanju o praktičnim zadacima iz redakture SP-a koji se mogu primijeniti u obrazovanju prevoditelja. Rad bi mogao biti zanimljiv nastavnicima prevođenja koji žele obogatiti poučavanje redakture SP-a, bilo u kolegijima/modulima posvećenima upravo toj temi, bilo u sklopu drugih kolegija.

*Ključne riječi: obrazovanje prevoditelja, redaktura strojnih prijevoda, strojno prevođenje, uočavanje pogrešaka*