

Source or target first? Comparison of two post-editing strategies with translation students

Lise Volkart, Sabrina Girletti, Johanna Gerlach, Jonathan David Mutal, Pierrette Bouillon

▶ **To cite this version:**

Lise Volkart, Sabrina Girletti, Johanna Gerlach, Jonathan David Mutal, Pierrette Bouillon. Source or target first? Comparison of two post-editing strategies with translation students. Journal of Data Mining and Digital Humanities, 2022, Towards robotic translation?, 10.46298/jdmdh.9067 . hal-03546151v3

HAL Id: hal-03546151
https://hal.science/hal-03546151v3

Submitted on 12 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal of Data Mining and Digital Humanities

EPI sciences overlay journals

Source or Target First? Comparison of Two Post-Editing Strategies with Translation Students

Lise Volkart[1*], Sabrina Girletti[1], Johanna Gerlach[1], Jonathan David Mutal[1], Pierrette Bouillon[1]

[1]Universite de Genève, Suisse

*Corresponding author: Lise Volkart, lise.volkart@unige.ch

Abstract

We conducted an experiment with translation students to assess the influence of two different post-editing (PE) strategies (reading the source segment or the target segment first) on three aspects: PE time, ratio of corrected errors and number of optional modifications per word. Our results showed that the strategy that is adopted has no influence on the PE time or ratio of corrected errors. However, it does have an influence on the number of optional modifications per word. Two other thought-provoking observations emerged from this study: first, the ratio of corrected errors showed that, on average, students correct only half of the Machine Translation (MT) errors, which underlines the need for PE practice. Second, the time logs of the experiment showed that when students are not forced to read the source segment first, they tend to neglect the source segment and almost do monolingual PE. This experiment provides new insight relevant to PE teaching as well as the designing of PE environments.

Keywords

post-editing; post-editing strategies; post-editing teaching; machine translation; post-editing environment; reading strategy

INTRODUCTION

The importance of post-editing (PE) in the translation industry no longer needs to be demonstrated, as a quick survey of the growing number of PE training options for professional translators is enough to understand the value of this skill. In academia, machine translation (MT) and PE competences are also largely integrated in curricula under the European Master's in Translation (EMT) Competence Framework [EMT Expert Group, 2017]. The content and methods of some PE training sessions and classes are described in academic papers (see for example Koponen [2015] and Doherty and Kenny [2014]), where standards such as TAUS guidelines [TAUS and CNGL, 2010] or the ISO norm on PE[1] are often mentioned. However, very few authors give advice on the procedure to be followed when post-editing, i.e. should post-editors read the source or the target first, and does it make any difference? In a survey conducted by Ginovart Cid [2021] among PE teachers, half of the respondents confirmed that they do not touch on this issue or give clear-cut advice to their students. A third of the participants recommend reading the source text first, while a fifth recommend starting with the target text. As far as we know, there are no studies that compare the benefits of one strategy or the other. The study presented here aims to provide some preliminary insight into the question.[2]

[1] ISO 18587. Translation services – post-editing of machine translation output – requirements, 2017.

[2] We would like to warmly thank Professor Annarita Felici and Professor Mathilde Fontanet for their participation, as well as the Master students of the faculty who took part in this experiment.

The paper is structured as follows: we start by laying out our motivations and the goals of our study before introducing our experiment and detailing our methodology and results. Finally, we present our conclusions and some prospects for further research.

I MOTIVATIONS AND GOALS

While the source segment (eventually alongside a translation memory suggestion) constitutes the primary source of information for translators, post-editors, in contrast, have the choice to primarily orient their attention toward the source or the MT suggestion (or target text, TT), as underlined by Krings [2001]. Nitzke [2019] showed that post-editors do not all adopt the same strategy, however several studies indicate that the majority of post-editors tend to look at the target first [Carl *et al.*, 2011; Mesa-Lao, 2014; Belam, 2003]. Furthermore, numerous studies on cognitive processes during translation and/or PE indicate that less attention is paid to the source text (fewer fixations, shorter gaze time) in PE compared to human translation (HT) [Carl *et al.*, 2011; Bangalore *et al.*, 2015; Nitzke, 2019; Mesa-Lao, 2014; Daems *et al.*, 2017]. Carl *et al.* [2011], Čulo *et al.* [2014] and Carl and Schaeffer [2017] have all formulated the hypothesis that this predominance of the MT suggestion/TT might have an influence on the final product through priming or directing effects. Indeed, post-edited texts tend to be more literal, include a high number of typical ST constructions and formal equivalences, and also tend to be closer to the source text than HT [Depraetere, 2010; Čulo *et al.*, 2014; Martikainen and Kübler, 2016]. Furthermore, studies have shown that post-edited texts include more unidiomatic or ungrammatical constructions, especially when post-editors are students [Daems *et al.*, 2017; Schumacher, 2019]. Student translators tend to be more tolerant towards MT output and are often liable to accept sub-optimal translations [Schumacher, 2019; Depraetere, 2010; Carl and Schaeffer, 2017; Casas, 2020). Finally, a study conducted on bilingual revision (an activity comparable to PE in the sense that the reviser, like the post-editor, has the choice of primarily orienting his/her attention toward the source or target, but attention also seems to be mainly focused on the target text during revision) by Ciobanu *et al.* [2019] showed that revisers produce better quality revisions (especially in terms of correcting accuracy errors) when they listen to the source segment (via speech synthesis) while revising.

Considering the results of the above-mentioned studies, the question of the role of reading order in PE and the potential consequences of a mainly target-oriented PE strategy appears to be worth investigating. We therefore decided to compare PE by students in two different conditions: when they start by reading the source segment (source condition) and when they begin with the target segment (target condition). In doing so, we aimed to observe whether or not adopting a strategy closer to the HT process would influence the amount of corrections made to the MT suggestion and the total PE time. Additionally, we surveyed students after the experiment to gauge their satisfaction with regard to one strategy or the other.

II EXPERIMENT

2.1 Methodology

2.1.1 Variables

The experiment we designed aimed to measure the influence of the PE strategy (reading the source or target segment first) on three variables:

- Total PE time per source word;

Journal of Data Mining and Digital Humanities
eISSN 2416-5999, an open-access journal
http://jdmdh.episciences.org
https://doi.org/10.46298/jdmdh.9067

- Ratio of corrected errors (i.e., number of errors post-edited by participants out of the total number of MT errors in the raw output);
- Number of so-called 'optional modifications' per word (i.e., the number of PE actions performed by students on elements that were not indicated as MT errors).

Our hypotheses are that the ratio of corrected errors and the number of optional modifications per word will be higher in source condition as the students will possibly be less primed by the MT output.

2.1.2 Text

We performed our experiment on a news article in English on a general topic from *Times Magazine*, which consisted of around 540 source words. The text was translated into French and Italian using DeepL in February 2020 and raw MT output for each language was annotated by a translation professor from the Faculty. Professors annotated MT errors (i.e., errors that students should correct in a PE task) using their usual correction system for HT. 52 errors were identified in the French translation and 42 in the Italian.

2.1.3 Participants

We recruited 20 Master translation students from the University of Geneva to take part in our experiment. 12 of them were native French speakers and the other 8 were Italian native speakers, and all had English in their language combination. Before the experiment, they completed a questionnaire on their experience in PE. It revealed that they all had little to no practical experience, however they had some theoretical knowledge, as they all had taken at least one course in the Faculty that included content on MT and PE. Students were not paid for their participation but they received a voucher as compensation.

2.1.4 Design

The experiment was conducted on the tailor-made PE platform COPECO[3] [Mutal *et al.*, 2020]. The platform allows two kinds of PE tasks to be created: a 'source condition' task, where the source segment is displayed by default and the user has to click on a button to display the MT suggestion, and a 'target condition' task, where the MT suggestion is displayed by default and the source segment has to be manually displayed. The platform records the time the user spends on the segment displayed by default before clicking on the button to display its counterpart (so-called 'default reading time'). Figures 1 and 2 provide a view of the platform in source and target condition, respectively.



Figure 1. View of the 'source condition'.

[3] https://copeco.unige.ch

Journal of Data Mining and Digital Humanities
eISSN 2416-5999, an open-access journal
http://jdmdh.episciences.org
https://doi.org/10.46298/jdmdh.9067

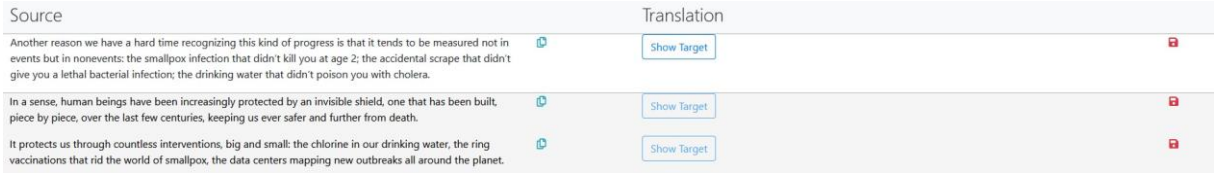


Figure 2. View of the 'target condition'.

Before the experiment, participants were assigned a test task to familiarize themselves with the platform. For each language pair, the text was split into two parts (A and B) and the participants were divided into two groups (1 and 2). The experiment was set up as a cross-over design in which group 1 does the PE of part A in source condition and part B in target condition, while group 2 post-edits part A in target condition and part B in source condition. During the experiment, participants had access to all online resources of their choice, with the exception of MT engines. We recorded the screens of participants during all the tasks in order to be able to observe participant behaviour during the experiment.

2.1.5 Instructions to the participants

Participants were given instructions to perform the two PE tasks with the aim of producing a high-quality translation (or one of publishable quality). They were not given precise information on the goal of the experiment and did not receive any instructions on the reading strategy to adopt (reading source or target first). With this design, our aim was to prompt students to adopt one strategy or the other, rather than providing them with a specific guideline that might influence their PE behaviour. Participants were also asked to fill out a questionnaire after the task in order for us to collect their impressions and comments.

2.2 Analysis

Because of the small number of participants, results for both languages were analysed together. We collected the different kinds of data recorded by the platform, as well as the screen recordings. The latter were used to verify whether or not participants responded as expected to the incentive of reading the source or target segment first. Only three participants out of 20 did not behave as expected in source condition and one in target condition. They clicked directly on the display button without spending any time reading the segment displayed by default. We performed our statistical analysis to both include and exclude these participants, but as it did not change the overall tendency of our results, and as we only have a small number of participants, we ultimately decided to include them. We manually counted the number of annotated errors that were post-edited and the number of optional modifications for each text part and each participant. Annotated errors and optional modifications were counted following the principle of single logical edits as described in Blain *et al.* [2011]. This means that one correction/modification can involve several mechanical actions (insertion, deletion, substitution, etc.) that are interdependent. Figure 3 shows an example of a single logical edit.

Aujourd'hui, les universités sont confrontées à l'un des chapitres les plus difficiles de leur **longue histoire.**	Aujourd'hui, les universités sont confrontées à l'un des chapitres les plus difficiles de leur **histoire, pourtant longue.**

Figure 3. Example of an optional modification. Here, three mechanical actions (a word shift and two additions) were counted as one logical edit, as adding the word *longue* implies adding a comma and the adverb *pourtant* in order to build a correct sentence.

Journal of Data Mining and Digital Humanities
eISSN 2416-5999, an open-access journal
http://jdmdh.episciences.org
https://doi.org/10.46298/jdmdh.9067

2.3 Results

2.3.1 Main variables

Table 1 presents the results obtained for each variable. The association between the total PE time per word and reading condition was investigated using the U-Mann-Whitney test (as normal distribution is not achieved). The total PE time per word was lower in the target condition, but this difference is not statistically significant (difference = -0.173s; Z = -1.722; p = 0.087). The associations between the ratio of corrected errors and the reading condition, and the number of optional modifications per words and the condition, were investigated using a linear regression model with mixed-effects. A random effect was set on the intercept to account for participant variability. The ratio of corrected errors was not significantly higher in the source condition (difference = +1.2%; 95% CI; -4.7-7.3; p = 0.675), but the number of optional modifications was significantly higher in the source condition (difference = +0.009; 95% CI; 0.0013-0.016; p = 0.021). It is interesting to note that the ratio of corrected errors is relatively low, as the participants corrected on average 50% of the MT errors annotated by the professors. This is in line with the results of previous studies, which show that students tend to be tolerant toward MT output.

It is important to mention that statistical analyses revealed an influence of the text part (A or B) on the ratio of corrected errors and the number of optional modifications. Participants made on average more corrections (difference = +7.6%; 95% CI; 1.5-13.7%; p = 0.014) and more optional modifications (difference = +1.7%; 95% CI; 1.05-2.53%; p = 0.014) to the second part (B) of the text. Thanks to the cross-over design, this aspect had a limited effect on our results.

	Source condition	**Target condition**
Total PE time/word	6.596	6.423
Ratio of corrected errors	0.505	0.500
Optional modif./words	0.051	*0.043

Table 1. Average total time per word, ratio of corrected errors and additional modification per word for each condition. (* indicates that scores are significantly different at $p < 0.05$).

2.3.2 Other variables

Alongside our three variables, we made an interesting observation on the 'default reading time' (i.e., the time participants spend on the side that is displayed by default before clicking to display the other part of the segment). In the source condition (i.e., when the source is displayed by default), the average default reading time by segment is 4.2 s, which corresponds to an average reading speed of 240 to 300 words per minute (comparable to standard English reading speed according to Brysbaert [2019]), while in the target condition (i.e., when the target is displayed by default), the average default reading time is 57.8 s which is far more than what is needed to read the MT output. The screen recordings revealed that most participants start doing research or begin post-editing the MT output before displaying the source. Some participants even postedit the whole segment and only look at the source when they are done. This observation confirms what other studies (see introduction) have shown, which is that the source text tends to be neglected in PE and attention is mainly oriented toward the MT suggestion/target text.

2.3.3 Questionnaire

In the post-task questionnaire, participants were asked which task they preferred. The answers were almost perfectly equally distributed, with 10 students indicating a preference for source condition, 9 for target condition and 1 indicating 'none of them'. Finally, we asked them if they thought the condition had influenced their PE behaviour. 11 replied 'yes', 7 said 'no' and 2 'I don't know'. It was interesting to note that a majority of them was conscious that the condition could influence their PE approach.

CONCLUSION AND FUTURE PROSPECTS

Our experiment has shown that the ratio of corrected errors was not significantly influenced by the order in which the source and target are displayed, which contradicts our first hypothesis. However, our results tend to confirm our second hypothesis, as the number of optional modifications was slightly higher in the source condition. These results might give rise to the assumption that the more attention students give to the source text, the more likely they are to deviate from the raw MT output. As for the time spent on the PE tasks, no significant difference was found between the two conditions.

This experiment also produced other interesting results. First, the fact that students spotted just half of the MT mistakes is in line with the results of other studies and emphasises the importance of MT and PE teaching in the translation curriculum. Second, it showed that the display design has a great influence on the post-editor's behaviour: when presented with the MT suggestion first, participants tend to omit the source segment and consult it only after having done a great part of, if not all, the PE. This aspect, even if its implications still need to be further investigated, cannot be ignored when giving PE guidelines to students, preparing specific PE exercises, as well as choosing, setting up and/or designing PE environments. Our results further support the idea already formulated by Moorkens and O'Brien [2017] that post-editors, who generally work with classical HT interfaces (CAT-tools), would benefit from environments optimized for the specific nature of PE tasks. Those environments could improve PE efficiency by offering, for instance, text-to-speech synthesis of the source text (as investigated by Ciobanu [2021]), or different positioning options for the source text and MT suggestions.

Even if our study provided valuable insight into the PE processes of students, we acknowledge that our experiment has some limitations, including the small number of participants and the fact that MT errors were only annotated by one annotator per language.

References

Bangalore S., Behrens B., Carl M., Ghankot M., Heilmann A., Nitzke J., Schaeffer M. and Sturm A. The role of syntactic variation in translation and post-editing. *Translation Spaces*. 2015; 4: 119-143.

Belam J. "Buying up to falling down": A deductive approach to teaching post-editing. In *MT Summit IX. Third Workshop on Teaching Machine Translation*. New Orleans, 2003: 1-10.

Blain F., Senellart J., Schwenk H., Plitt M. and Roturier J. Qualitative analysis of post- editing for high quality machine translation. *Conference: Machine Translation Summit XIII*. Xiamen (China), 2011.

Brysbaert M. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*. 2019; 109(104047). https://doi.org/10.1016/j.jml.2019.104047

Carl M. and Schaeffer M. Why translation is difficult: A corpus-based study of non-literality in post-editing and from-scratch translation. *HERMES. Journal of Language and Communication in Business*. 2017; special issue: 43-57.

Carl M., Dragsted B., Elming J., Hardt D. and Jakobsen A. L. The process of post-editing: A pilot study. *Copenhagen Studies in Language*. 2011; 41: 131-142.

Casas S. Intégration de la post-édition dans un cours de traduction et revision : étude de cas d'étudiant·e·s à la faculté de traduction et d'interprétation. 2020. https://archive-ouverte.unige.ch/unige: 131300

Ciobanu D. Benefits of speech technologies for translators and revisers. *Colloque #AFFUMT : Former aux métiers de la traduction aujourd'hui et demain*. Online, 2021.

Ciobanu D., Ragni V. and Secară A. Speech synthesis in the translation revision process: Evidence from error analysis, questionnaire, and eye-tracking. *Informatics*; 2019; 6(4): 51.

Daems J., De Clercq O. and Macken L. Translationese and post-editese: How comparable is comparable quality? *Linguistica Antverpiensia*. 2017; 16: 89-103.

Depraetere I. What counts as useful advice in a university post-editing training context? Report on a case study. *Proceedings of the Annual conference of the European Association for Machine Translation*. Saint-Raphaël (France), 2010. https://halshs.archives-ouvertes.fr/ halshs-00849535

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

EMT Expert Group. *European master's in translation competence framework 2017*, 2017.

Ginovart Cid C. *The need for practice in the acquisition of the post-editing skill-set: lessons learned from the industry*. PhD thesis, Universitat Pompeu Fabra (Spain), 2021.

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

Krings H. P. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Kent State University Press (Kent), 2001.

Martikainen H. and Kübler N. Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*. 2016; 27. https://doi.org/10.4000/ilcea.3863

Mesa-Lao B. Gaze behaviour on source texts: An exploratory study comparing translation and post-editing. In O'Brien S., Winther Balling L. *et al.* (eds). *Post-editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing (Newcastle upon Tyne), 2014: 219-245.

Moorkens J. and O'Brien S. Assessing user interface needs of post-editors of machine translation. In Kenny D. (ed.). *Human issues in translation technology*. Routledge (London), 2017: 127-148.

Mutal J. D., Bouillon P., Schumacher P. and Gerlach J. COPECO: A collaborative post-editing corpus in pedagogical context. *North American component of the International Association for Machine Translation. 1st Workshop on Post-Editing in Modern-Day Translation.* Orlando, 2020.

Nitzke J. *Problem solving activities in post-editing and translation from scratch: A multi-method study*. Language Science Press (Berlin), 2019.

Schumacher P. Avantages et limites de la post-édition. *Traduire*. 2019; 241: 108-123. https://doi.org/10.4000/traduire.1887

TAUS and CNGL. Machine translation post-editing guidelines. TAUS, 2010. https://info.taus.net/mt-post-editing-guidelines

Čulo O., Gutermuth S., Hansen-Schirra S. and Nitzke J. The influence of post-editing on translation strategies. In O'Brien S., Winther Balling L. *et al.* (eds). *Post-editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing (Newcastle upon Tyne), 2014: 200-218.