

From a Discreet Role to a Co-Star: The Post-Editor Profile Becomes Key in the PEMT Workflow for an Optimal Outcome

Lucía Guerrero Romeo
Machine Translation Specialist
CPSL, Barcelona, Spain
lguerrero@cpsl.com

Abstract

Traditional MT (machine translation) workflows usually involve the post-editor only at the end. A lot of the time, post-editors are forced to rework unacceptable machine translation output. Even when the output is acceptable, they lack information about how the MT system works and what they can do to improve it, which causes frustration and unwillingness to post-edit again. Low-quality raw MT eventually leads to post-editors developing a negative attitude towards the MT system, which can result in poor quality post-edited texts which, in turn, contribute very little to the system training cycle, thus resulting in a static, never-improving process, and a tedious task for post-editors. This paper and the associated presentation aim to encourage post-editors to embrace the often-controversial MT processes, and companies to integrate post-editors into their MT workflows from the outset, even reshaping such processes to become linguist-focused rather than machine-centred, with the post-editor playing a central role. This change will enable higher quality and productivity and, therefore, a successful post-editing experience for everyone.

1 Towards a translator-centred process

Powered by neural networks and deep learning, AI (artificial intelligence) has nowadays become commonplace and proved useful in those businesses where vast amounts of data are available. The translation industry was one of the scenarios where neural networks were implemented first and, in many cases, these systems are offering amazingly good quality, thus contributing to the recurrent idea of machines taking over and replacing humans soon which usually appears with every new technological breakthrough.

Machine translation and post-editing are little by little finding their place as independent subjects in translation graduate and post-graduate programmes. There has also been a boom in academic writing dealing with several aspects related to the PEMT (post-editing of machine translation) process, but one that might go unnoticed amongst all the current hype about BLEU scores and productivity improvements is the suggestion to abandon the current machine-centred paradigm and work together towards a more translator-centred process, in which the post-editor is not only involved at the end to correct the errors stubbornly produced by the system but has a main role in each and every step.

When describing localization processes back in 2014, Anthony Pym already envisaged that translators were being forced to adapt to the interfaces and segmentation rules of modern CAT tools: “Translation memory and machine translation programs break texts into phrase and sentence units, inviting translators not to alter the renditions already in the database. The mind of the translator is consequently moved from the level of text and communication to that of phrase and formal equivalence” (Pym 2014).

In a similar way nowadays, immersed in the current research on NMT (neural machine translation), many developers assure that we now “don’t need to know much about a

language”¹: we just have to train the system with huge amounts of data and it will find out its own errors and will learn from them automatically. At the same time, some of the largest MT providers claim that they have reached ‘human parity’ with certain language combinations². While the methods they used to evaluate the system have been questioned (mainly the fact that evaluators were not professional translators)³, it is true that, since NMT appeared, machine translation is now in the hands of the developers. The intricacies of the systems are like a crystal ball which the translators can’t read; actually, not even the developers can explain some mechanisms from deep learning.

Such claims can inflate the expectations from MT buyers and the associated risks are MT buyers and LSPs sending raw MT output to post-editors regardless of the quality. Ignoring all the linguistic knowledge that translators can bring could mean lots of improvement possibilities lost. Finally, translators may end up considering post-editing as a tedious and alienating task, and not be willing to accept any more PEMT jobs.

So where does all this NMT hype leave translators? What kind of roles, as well as skills and abilities, are left for them in the translation pipeline? Celia Rico (2017) addresses these and other questions in some of her recent works. She depicts the translation process as an iterative cycle similar to the agile and scrum software localization management processes, with the translator at the very centre, administrating the computer and deciding how to best combine the materials they have to hand at each part of the process, either glossaries, translation memories or machine translation engines. When the iterative model is adapted to machine translation, the post-editor not only becomes responsible for delivering a final product but also performs a key role in several steps of the process, contributing to a successful outcome.

This perspective over-turns the idea of post-editors simply required to post-edit a file, without even knowing the kind of MT system they have to deal with or the reference material used to set it up, and considers machine translation as a tool that interacts with the user, instead of an object performing tasks on its own.

Works from Sharon O’Brien (2018) are in line with Rico’s and support the adoption of a human-centred process. She goes beyond workflows and proposes a deep customization of translation tools, either CAT tools or MT systems, according to variables such as a particular domain, specific circumstances (e.g. an urgent job in which the level of acceptance threshold of MT is lower than usual) or even the style of each translator. This is a promising topic that has been addressed by SDL Trados Studio AdaptiveMT and Lilt and can bring new scenarios to the current translation processes.

The translators’ opinions reinforce this research line as well. In a survey conducted as part of the EAMT 2018 21st Annual Conference (Pérez-Macías, Rico and Forcada, 2018), 52% of the translators were willing to accept post-editing jobs and 79% considered that translators contribute to MT development:

¹ Heard at the ‘Advances in Machine Translation – What is Exciting and Shows Promise Ahead’ GALA webinar offered by Ominiscient Technologies (<https://www.gala-global.org/ondemand/advances-machine-translation-what-exciting-and-shows-promise-ahead>)

² See Hassan *et al.* (2018).

³ See Toral *et al.* (2018) and Läubli *et al.* (2018)

B) Use and perceptions on MT

DEGREE OF CONFORMITY WITH THE FOLLOWING STATEMENTS					
	Strongly agree	Agree	Indifferent	Disagree	Strongly disagree
<i>I mistrust MT</i>	15 %	29 %	16 %	29 %	12 %
<i>I'm willing to accept PE Jobs</i>	16 %	36 %	11 %	18 %	19 %
<i>Translators contribute to MT development</i>	40 %	39 %	13 %	6 %	2 %
<i>MT helps to improve productivity</i>	26 %	31 %	17 %	16 %	10 %

Figure 1: Degree of conformity with several statements about MT

However, it is particularly worrying that, when answering the question ‘How often the translators’ needs about MT are heard’, a total of 40% answered ‘never’ or ‘almost never’. As part of this question the translators added their suggestions on ways to contribute to the MT process:

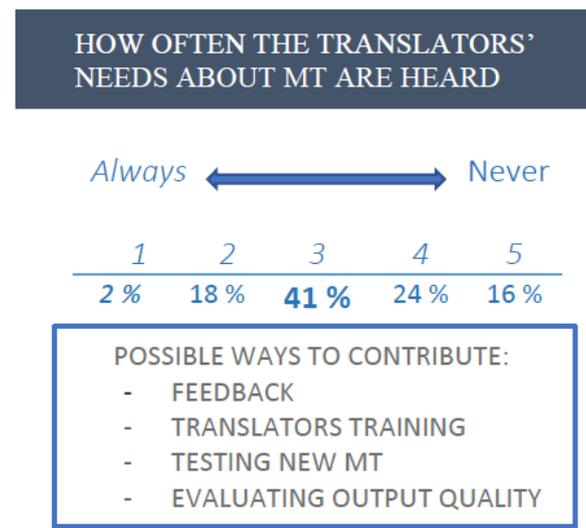


Figure 2: How often the translators’ needs about MT are heard and possible ways to contribute

If all parties involved in machine translation processes acknowledge that mutual collaboration is not only possible but also desirable, then the challenge for LSPs and machine translation buyers is to take up the torch from academic research and establish new relationships with post-editors, moving towards a more translator-centred process.

2 The role of the post-editors in the MT pipeline

At CPSL we started using machine translation in 2011. Since then, the number of language combinations, domains and systems used hasn’t stopped growing. We have also contributed to industry-leading round tables, blogs and webinars suggesting best practices, and we have learnt from our own experience the importance of engaging post-editors during the whole PEMT process, as well as giving a ‘human touch’ to evaluations and automatic scores. We are

now glad to share our best practices with the TC40 Asling conference audience, so we will try to describe how, when and what post-editors can bring in to the outcome.

2.1 Step 1: MT system setup

Although it is true that NMT more than ever requires professionals from the data science field, there are some tasks related to resources and evaluation in which translators' contribution can be decisive.

- Data selection

Training an engine for a specific domain usually implies selecting quality reference material, either monolingual or bilingual, such as translation memories and customer-approved glossaries. This task is typically carried out by LSP project managers but difficulties may arise, or opportunities can be lost if they are not proficient in the language pairs concerned or reference materials are not properly classified or up-to-date. A professional translator with experience in that particular subject can help decide which materials are relevant on a case-by-case basis and add further terminological resources created ad hoc.

- Evaluation

To avoid frustration and ensure the best results, it is strongly recommended to test any new engine before sending machine-translated output to post-editors, especially if the aim is to produce a good quality translation with a full post-editing. The best decisions will come from combining automatic metrics (quantitative) and human evaluation (qualitative).

While automatic metrics such as BLEU (based on similarity to a translation used as reference) or TER (based on the number of modifications done) can certainly save costs and time when compared to human evaluation, the post-editor's contribution to this step is crucial. A closer look at the changes, as opposed to simply sticking to a percentage, can reveal important details such as completely unacceptable mistakes.

Qualitative feedback can be requested in many ways, for example asking post-editors to analyse a comparison report of the test set (result of comparing the sample human translation with the raw MT) and classify errors according to severity and/or type. This valuable feedback can help understand the potential improvement areas, where translators can also be of help creating additional glossaries, lists of non-translatable terms or regular expressions to improve the edit distance.

Human evaluation of a new system becomes even more relevant with NMT: the results are more fluent and the semantic errors such as mistranslations can only be detected by professionals proficient in both source and target languages as well as in the subject area. Samuel Läubli *et al.* (2018) address this topic in a research in which they concluded that “as machine translation quality improves, translations will become harder to discriminate in terms of quality, and it may be time to shift towards document-level evaluation, which gives raters more context to understand the original text and its translation (...)”. To sum up, evaluation needs to take place as much as possible at a document-level and done by professional translators.

Finally, when the translators get involved in the system's evaluation they obtain valuable information about the strengths and weaknesses of the engines they will have to work with. If they cannot be involved at least they must receive information about the decisions taken in this step and the kind of system they will be using (rule- or phrase-based, statistical or neural).

- Settings definition

These refer to the creation, when relevant, of post-editing guidelines and the definition of the level of post-editing required. Usually the project manager or the customer should define these, but post-editors can get involved as well especially in compiling examples of the most common errors for the post-editing rules, for example, in the form of a list.

2.2 Step 2: Post-editing

- Pre-editing

Errors in the source text may prevent the system from finding the best matches for each segment. To avoid this, sometimes it's just a matter of running a spelling and grammar check on the source text—either a post-editor if they are proficient enough in the source language, or a native speaker of the source language, for other problems a more complex pre-editing is required (e. g. to make a document comply with a controlled language).

- Post-editing

This is the actual review and modification of the machine-translated text. When working with CAT tools, the post-editor can receive the files in different ways: either all pre-translated (i. e. the raw MT directly downloaded in the target segments), or only pre-translated with the translation memory, when available, and the machine-translated segments compiled in a separate TMX.

The second option is more in line with the works by Rico and O'Brien, as the post-editor is not forced to accept raw MT output as is and they can decide where to download the segments from—either from the translation memory or the MT system, or when to translate them from scratch. Simply extracting all raw MT segments in a separate TM does not only improve productivity (the translator does not have to delete useless segments) but also makes the post-editing task less annoying. Asking the post-editing team which of these options is more comfortable for them can contribute to make the task more pleasant.

- Final quality checks

From time to time companies face urgent projects where even machine translation is not enough to meet the tight deadlines, so texts must be not only be machine-translated but also split between several post-editors. Using a CAT tool and a shared remote TM and glossary, as well as QA (quality assurance) tools, can certainly help, but urgencies can cause oversights, especially when using NMT—as explained above, the resulting translations are so fluent that it's easy to miss semantic errors. Rush projects may also require releasing MT output into production before it has been tested, risking some ridiculous or meaningless translations.

QA tools cannot detect semantic errors automatically, so the post-editor can contribute by creating a list of completely unacceptable mistranslations and search for them in the post-edited text to make sure all instances have been properly corrected.

2.3 Step 3: Evaluation and optimization

Engines must be tested not only after being set up, but also after entering the production pipeline to make sure the results still match the expectations and to allow for a continuous improvement. In this step the evaluation can also be automatic or human, and the combination of both will offer the most objective information, as the successes from one can compensate the failures from the other.

There are many ways to collect feedback from the post-editors about any given system. It can be gathered using exhaustive reporting systems and tools such as DQF⁴ or Qualityity⁵, which allow for severity and error categorization, but also in a plain Excel file, an email or even a call. What is important is that we allow the post-editors who have worked with the system to give their own opinion about the errors found and use this to improve the system.

Updating TMs and glossaries with corrections from reviewers, language leads or the customer is also a task that must be carried out by a professional post-editor. These can be used to optimize the system and avoid the same errors happening in future PEMT projects.

Finally, if post-editors are sent a comparison report they can spot repetitive error patterns and define regular expressions to correct them faster, allowing an increase in productivity.

3 Conclusions

Based on our experience, changing the mindset of all the individuals involved in PEMT projects towards a human-centred workflow offers several advantages for all stakeholders:

- Relevant resources and information are shared amongst everyone involved
- The post-editors, instead of tools, take centre stage
- The more information and tasks are shared with post-editors, the more engaged and willing to collaborate they become in the future
- The interpretation of the results is more precise and allows for an optimal system fine-tuning

References

- Hassan, Hani, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. 2018. Achieving Human Parity on Automatic Chinese to English News Translation. arXiv:1803.05567 [cs.CL].
- Läubli, Samuel, Rico Semrich, and Martin Volk. 2018. Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. arXiv:1808.07048 [cs.CL].
- Pérez-Macías, Lorena, Celia Rico, and Mikel Forcada. 2018. EAMT 2018 Translators' track: survey report. http://www.eamt.org/translators_documents/eamt_2018_survey_report.pdf [last accessed October 2, 2018].
- Pym, Anthony. 2014. Localization, Training, and Instrumentalization. In *Translation Research Projects 5*, pages 37-50.
- Rico, Celia. 2017. La formación de traductores en traducción automática. In *Revista Tradumàtica. Tecnologies de la traducció*, 15, pages 75-96.
- Toral, Antonio, Sheila Castillo, Ke Hu and Andy Way. 2018. Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. arXiv:1808.10432v1 [cs.CL].

⁴ <https://www.taus.net/quality-dashboard-lp>

⁵ Free Trados Studio plug-in available at the SDL AppStore.