Integration of Machine Translation and Translation Memory: Post-editing efforts

Rocío Caro Quintana

University of Wolverhampton, UK
R.Caro@wlv.ac.uk

Abstract. The development of Translation Technologies, like Translation Memory and Machine Translation, has completely changed the translation industry and translator's workflow in the last decades. Nevertheless, TM and MT have been developed separately until very recently. This ongoing project will study the external integration of TM and MT, examining if the productivity and post-editing efforts of translators are higher or lower than using only TM. To this end, we will conduct an experiment where translation students and professional translators will be asked to translate three short texts; then we will check the post-editing efforts (temporal, technical and cognitive efforts) and the quality of the translated texts.

Keywords: Translation Memory, Machine Translation, Integration, Post-editing efforts, Eye-tracking

1   Introduction

The way translators work has changed considerably in the last decades: with a world that is more globalised each day, the creation of texts has increased, and this has also affected and transformed the translation industry. The development of Translation Memory (TM) tools and Machine Translation (MT) systems is the result of these changes.

Translation Memory and Machine Translation as applications have been used for many years by academics and professional translators, but they have been developed and studied in isolation [1]. Nonetheless, recent research suggests there is an increase in the interest of integrating these two applications.

The integration of TM and MT creates new questions for all the parties involved in the translation process (translators, language service providers, clients) and academia. Will the integration save time? Therefore, will it also save money? Will the quality of the texts be as good as a translation from scratch?

Even if the answer to these questions turns out to be positive (e.g., the time will be reduced and the quality maintained), another question remains: Will the cognitive efforts of translators decrease? With the integration of TM and MT, more options (or segments) are presented to the translator and the translator's cognitive efforts may vary as the translation process would have one more step (explained in five simple steps):

1) Read source segment
2) Read TM output
3) Read MT output
4) Compare and choose which option is better
5) Translate or post-edit

The aim of this work-in-progress paper is to compare the different types of post-editing efforts (temporal, technical and cognitive) of translating with only TM or MT and translating with the combination of TM and MT. That is, calculating if the time it takes to finish a translation task is higher or lower using both technologies, checking if the technical efforts (studying the keystrokes and edits) are the same, lower, or higher; and examining if the cognitive efforts are also higher or lower using TM and MT. Our initial hypothesis is that the three efforts will decrease with the combination of TM and MT, however, while there are studies about the differences between translating with TM and MT [2, 3], empirical research about translating with a Computer-Assisted Translation (CAT) tool that integrates TM and MT is still limited.

Therefore, this ongoing project will try to answer the following research questions:
RQ1: Will the translator save time using a system that integrates TM and MT?
RQ2: Will the translator invest less cognitive efforts when using TM and MT?
RQ3: Will the quality of the translation not be deteriorated when using TM systems with integrated MT?

2   Related work

The integration of TM and MT was first envisaged by Peter Arthern [4] and nowadays most CAT tools have an option to integrate both technologies.

According to Zaretskaya et al. [5] the integration of TM and MT can be separated in two main categories: internal and external. The internal integration aims at improving the quality of the TM systems by using MT techniques, or vice versa, thus providing only one option to the translator. The external integration provides another suggestion, the MT output, in addition to the matches from the TM.

One of the first studies on the internal processing integration of TM and MT is from authors Lange and Bennet [6], in which they integrated MT into the translation process by treating the MT segments as fuzzy matches. Several papers working on internal integration were published the following years [7, 8, 9].

Concerning the external integration, Kanavos and Kartsaklis [10] is one of the first studies about this topic, however, the integration is not straightforward as the MT is not suggested with every segment, but only with the segments with a low fuzzy match. Other relevant studies are The MateCat project [11], and Eriguchi et al. [12] where TM segments were combined with Neural Machine Translation.

Herbig et al. [13] briefly discuss that "machines can generate a variety of probable translations from MT and TM instead of a single one" and that this could bias or confuse the human translator. The translators could, for example, see an output from the TM or the MT and consider it accurate since the other aspects of the sentence (grammar,

162

punctuation, coherence, etc.) may be correct, whereas the translation is not. They claimed, therefore, that future research should focus on estimating the cognitive efforts of translators in these types of scenarios.

3 Methodology

3.1 Experiments

The experiments we will carry out are, according to Zaretskaya et al. [5], within the external integration category. We will check if the translators' post-editing efforts increase or decrease in a setting where both segments from TM and MT are suggested.

As the main objective of the project is to examine the differences between translating with the help of only TM or MT and translating with the combination of both technologies, professional translators and Translation students will be asked to translate three short texts (around 300 words) with a CAT tool from English into Spanish.

The first text will be translated using TM, the second text will be translated with MT, and the third text will be translated using TM and MT. As the texts are different the results may not be conclusive, therefore the same texts should be translated again changing the order.

Table 1. Distribution of the texts

1st text	2nd text	3rd text
TM	MT	TM + MT

The study will be divided in two stages: the pilot project and the main project. For the pilot project and, due to the COVID-19 pandemic, participants can carry out the task remotely (in their homes, university, library) as long as they have an internet connection; in the main project, the translation task will be carried out in the presence of the researcher using an eye-tracking device that records the eye movements and the size of the pupil of the translators.

3.2 Post-editing efforts

The post-editing efforts, first described by Krings [14], will be studied as follows:

Temporal effort: The total time it takes for each translator to complete the task will be recorded, as well as the seconds per segments and the seconds per word.

Technical effort: An analysis of the keystrokes will be made with a key-logging tool, as well as the edit distance, which is the number of changes that were made from the source text to the translated text. and the post-edit distance. Post-edit distance is calculated by dividing the number of words or characters in the original text between the number of words or characters in the post-edited text.

Cognitive effort: To study the cognitive efforts, we will use an eye-tracking device. The measures that will be studied with the eye-tracker are the following: pauses or fixations (fixation duration, fixation count, first and second fixation duration), eye-movement, pupil dilation and regression behaviour. The eye-tracking software generates a video recording of the screen and the face/eyes of the participant, as well as an Excel file with the data collected. It also includes heat maps and gaze plots.

3.3 Qualitative data

In order to collect qualitative data, this project will also include two questionnaires: one before the translation task and one after the translation task. The first questionnaire will contain questions about the background of the participants (education, job experience on translation), their knowledge of CAT tools and their general opinion of working with translation technologies. The questionnaire after the translation task will include questions regarding their opinion of the experiment: if they think using TM and MT is beneficial, if they prefer working only with TM, etc. Previous studies have also conducted similar surveys [2, 15, 3].

The quantitative results (gathered from the translation task) may differ from the qualitative results. For instance, the eye-tracking results could show that the translator invests less cognitive efforts during the translation of the texts using TM and MT combined, but the experience of the participants could be the opposite, i.e., they find it more difficult to translate with TM and MT, rather than with only TM. Therefore, we also plan to analyze the times each translator chooses each option (translate from scratch, edit TM or post-edit MT), and if the option they choose is the most beneficial in terms of cognitive and temporal efforts.

3.4 Quality Evaluation Methods

The results will be evaluated manually and automatically. For the manual evaluation, the TAUS DQF/MQM Error Typology [16] will be used.

Regarding the automatic evaluation, all the translations made by the participants will be assessed with the automatic metrics BLEU [17], TER [18] and METEOR [19] comparing the results of the participants with a 'gold standard' to check if they correspond to the manual evaluation.

4 Conclusion

To sum up, we will carry out an experiment to check whether the external integration of TM and MT, where translators are able to see segments proposed by both technologies, is beneficial for translators. To do that, we will conduct a translation task, and we will assess if the temporal, technical and cognitive efforts increase or decrease. We will also check the quality of the translations, and we will collect qualitative data about the opinion of the participants with two questionnaires.

References

1. Koehn, Philipp, and Jean Senellart. "Convergence of translation memory and statistical machine translation." Proceedings of AMTA Workshop on MT Research and the Translation Industry. (2010)
2. Guerberof Arenas, Ana. "Productivity and quality in the post-editing of outputs from translation memories and machine translation." Localisation Focus The International Journal of Localisation 7.1 (2008): 11-21.
3. Sánchez-Gijón, Pilar, Joss Moorkens, and Andy Way. "Post-editing neural machine translation versus translation memory segments." Machine Translation 33.1 (2019): 31-59.
4. Arthern, Peter J. "Machine translation and computerized terminology systems: a translator's viewpoint." Translating and the Computer: Proceedings of a Seminar, London. Vol. 14. (1978)
5. Zaretskaya, Anna, Gloria Corpas Pastor, and Miriam Seghiri. "Integration of Machine Translation in CAT tools: State of the art, evaluation and user attitudes." Skase Journal of Translation and Interpretation 8.1 (2015): 76-89.
6. Lange, Carmen Andres, and Winfield Scott Bennett. "14. Combining Machine Translation with Translation Memory at Baan." Translating into success. John Benjamins, (2000) 203-218.
7. Marcu, Daniel. "Towards a unified approach to memory-and statistical-based machine translation." Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, (2001)
8. Langlais, Philippe, and Michel Simard. "Merging example-based and statistical machine translation: an experiment." Conference of the Association for Machine Translation in the Americas. Springer, Berlin, Heidelberg, (2002)
9. Groves, Declan, and Andy Way. "Hybrid data-driven models of machine translation." Machine Translation 19.3-4 (2005): 301-323.
10. Kanavos, Panagiotis, and Dimitri Kartsaklis. "Integrating machine translation with translation memory: A practical approach." (2010).
11. Federico, Marcello, Alessandro Cattelan, and Marco Trombetti. "Measuring user productivity in machine translation enhanced computer assisted translation." Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA). (2012)
12. Eriguchi, Akiko, Spencer Rarrick, and Hitokazu Matsushita. "Combining translation memory with neural machine translation." Proceedings of the 6th Workshop on Asian Translation. (2019)
13. Herbig, Nico, et al. "Integrating Artificial and Human Intelligence for Efficient Translation." arXiv preprint arXiv:1903.02978 (2019).
14. Krings, Hans P. Repairing texts: Empirical investigations of machine translation post-editing processes. Vol. 5. Kent State University Press, (2001)
15. Bundgaard, Kristine. "Translator Attitudes towards Translator-Computer Interaction– Findings from a Workplace Study." Hermes 56 (2017): 125-144.
16. Harmonized DQF-MQM Error Typology https://www.taus.net/qt21-project#harmonized-error-typology
17. Papineni, Kishore, et al. "Bleu: a method for automatic evaluation of machine translation." Proceedings of the 40th annual meeting of the Association for Computational Linguistics. (2002)

18. Snover, Matthew, et al. "A study of translation edit rate with targeted human annotation." Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers. (2006)

19. Banerjee, Satanjeev, and Alon Lavie. "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments." Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. (2005)