

Interactive Models for Post-Editing

Marie Escribe, Ruslan Mitkov

Research Group in Computational Linguistics
University of Wolverhampton
{m.c.escribe, r.mitkov}@wlv.ac.uk

Abstract. Despite the increasingly good quality of Machine Translation (MT) systems, MT outputs require corrections. Automatic Post-Editing (APE) models have been introduced to perform these corrections without human intervention. However, no system has been able to fully automate the Post-Editing (PE) process. Moreover, while numerous translation tools, such as Translation Memories (TMs), largely benefit from translators' input, Human-Computer Interaction (HCI) remains limited when it comes to PE. This research-in-progress paper discusses APE models and suggests that they could be improved in more interactive scenarios, as previously done in MT with the creation of Interactive MT (IMT) systems. Based on the hypothesis that PE would benefit from HCI, two methodologies are proposed. Both suggest that traditional batch learning settings are not optimal for PE. Instead, online techniques are recommended to train and update PE models on the fly, via either real or simulated interactions with the translator.

Keywords: Automatic Post-Editing, Machine Translation, Human-Computer Interaction.

1 Introduction

The emergence of automatic translation dates back to 1949, when Warren Weaver, a researcher at the Rockefeller Foundation, presented a set of proposals for MT solutions which were based on information theory and successes in code breaking during the Second World War. However, MT research faced several challenges over the years, which led to what was almost a standstill in the field for many years to come. In 1966, the Automatic Language Processing Advisory Committee (ALPAC) report concluded that MT outputs were too disappointing to continue investigating such systems, especially since there were enough human translators to complete translation projects [1]. This conclusion no longer stands. Translation plays a crucial role in communication in today's globalised world, and the number of texts requiring translation continues to rise. With such high volumes and tight deadlines, translators now heavily rely on technological assistance. Most projects are indeed undertaken in a Computer-Assisted Translation (CAT) environment, in which professionals use TMs and/or post-edit MT outputs. As the quality of MT outputs continues to increase, PE has become a common step in translation workflows nowadays. Nevertheless, since transla-

tions generated by MT systems need editing, methods have been proposed to automate this process using APE models.

2 Automatic Post-Editing

APE was first suggested by Knight and Chander [2] as a module complementing MT systems. Allen and Hogan [3] also observed that recurring errors are often found in MT outputs and introduced an APE module based on a controlled language to address this issue. This module was trained using triplets (source, MT output, post-edited version) to extract PE rules and apply these to unseen texts. This is still a typical approach for most APE models today, but more sophisticated methods have been suggested to improve the training phase. In fact, the evolution of APE has tended to follow that of MT, as they both benefitted from very similar technological improvements. Just like MT systems, APE was first based on rules and then adopted statistical methods before utilising machine learning and neural networks. Since 2006, when Llitjós and Carbonell [4] raised the issue of a lack of fully automatic solutions to PE, APE has regained popularity with the first APE shared task at the WMT conference series [5]. These have been running regularly ever since, providing datasets and a forum for discussing the latest advances in this field. The first round of the APE shared task was not very successful, as no system could beat the baseline [6]. However, significant improvements were later achieved, when APE systems began to use neural approaches [7]. Despite this, it is common for APE systems to produce over-corrections and to fail to detect certain errors [5].

Moreover, APE became even more challenging, as Neural MT (NMT) provides translations of a higher quality compared to previous MT systems, thus making automatic correction a more complex task [5,8]. In fact, the usefulness of APE models resides in the improvement margin observed in MT outputs [5].

Nevertheless, APE models still have clear advantages. While it could be argued that retraining an MT system would yield similar results in terms of corrections applied to the target text, this might not always be feasible, in particular because it would require having access to the MT system parameters [6].

3 Human-Computer Interaction in MT & PE

Although APE might contribute to improving MT outputs, the target texts generated in this way require proofreading by human experts. It could even be argued that working with an APE output necessitates more effort than PE alone, as extra attention should be paid to overcorrections and unspotted errors. This creates a significant discrepancy between APE systems and the real needs of MT users. Moreover, professional post-editors benefit from translation technologies (e.g. CAT), but these are not optimal for PE. TMs are valuable in assisting professionals and benefit from interaction with the translator: they are populated with approved translations which can easily be reused, via either automatic propagation of fuzzy matches or a manual concordance search. TMs are therefore constantly adapted to the style of the post-editor and to

the register and terminology of the document being translated. When it comes to PE, however, HCI is very limited, as corrections made by post-editors are not exploited.

Several interactive models have been introduced for MT, which use an interface to collect and reuse corrections made by post-editors. Well-known Interactive MT (IMT) models include Transtype [9], Transtype 2 [10], MIPRCV (Multimodal Interaction in Pattern Recognition and Computer Vision) [11] and CASMACAT (Cognitive Analysis and Statistical Methods for Advanced Computer-Aided Translation) [12]. These models offer a form of autocompletion in which suffixes are suggested based on translation prefixes validated by the post-editor, who might select to accept or modify the predictions. More recently, several studies have suggested implementing IMT models based on NMT systems [13, 14, 15, 16]. This typically involves adapting the search algorithm to make predictions which are constrained to a given prefix. IMT is therefore a human-centred approach [17], as IMT models evolve based on interactions with the translator. Despite this, certain studies comparing the PE effort in a traditional setting versus IMT yielded mixed results. For instance, in the paper by Underwood et al. [18], certain participants did not find the interactive setting helpful, while others reported positive experiences. Alves et al. [19] found that IMT did not yield improvements in efficiency (time and number of keystrokes) but could contribute to reduce the cognitive effort (shorter fixation durations were observed using eye-tracking equipment in the case of IMT). Several explanations to these results can be put forward, such as the lack of familiarity of the participants with such interfaces and the time and effort required to engage with predictions which are constantly being updated.

While HCI has been explored using glass-box approaches in IMT settings, several studies have also introduced ‘interactive’ APE models. APE is a black-box approach, as it does not require access to the MT system parameters but only to an MT output. Augmenting such systems by adding an interactive component enables APE models to learn from corrections during the PE process and to make informed correction predictions. Consequently, such models do not fall under the category of APE, as this process would not be fully automatic. Instead, they would be Interactive PE (IPE) systems. The study by Knight and Chander [2] was the first to briefly formulate this concept, which they called an ‘adaptive post-editor’. This suggestion is in line with the findings of the user survey conducted by Lagoudaki [20], which revealed that translators found the concept of an adaptive PE system highly relevant.

With the objective to implement this theory, Simard and Foster [21] introduced PEPr (Post-Edit Propagation), a model inspired by TMs which takes a phrase-based MT output and uses an APE module based on an online method to learn from human corrections on the fly. The assumption is that propagating corrections automatically can be beneficial when the number of repetitions in a text is high. Building on this model, Lagarda et al. [22] proposed an online APE system specifically designed to improve domain adaptation.

More recently, Chatterjee et al. [23] emphasised the need for APE systems capable of handling continuous streams of data to adapt to evolving settings and to the variety of domains present in real-world translation workflows, and presented a statistical online APE model designed to address this challenge. Building on [23], Negri et al. [24]

developed an online APE system which can learn from simulated interaction with the human post-editor. However, these interactive APE models all simulate interaction with the translator, which has several implications. The most salient limitation is that corrections are constrained to the pre-existing references, which allows for one possible correction only. This is a rather unrealistic scenario, as translation is an open-ended problem. Consequently, retrieving human input in an interactive environment would provide a unique opportunity to improve the PE process.

4 Suggested Methodologies

Two methodologies are proposed to address this issue. In both cases, the underlying assumption is that batch training is not optimal for PE tasks. Most machine learning algorithms behind MT and APE are trained offline, which impedes any form of interaction, as adding new attributes involves retraining the model from scratch. Online learning, on the other hand, allows for handling continuous streams of data and updating the model parameters on the fly. This approach thus appears more appropriate for implementing or simulating interactive PE settings. Furthermore, it should be noted that the objective of the approaches presented here is not to improve APE but rather to enhance the PE experience.

4.1 Fully Interactive PE Model

The first option is to design an IPE environment. As it seems very unlikely to produce fully automatic translations, an IPE setting would enable a PE system to learn from human corrections in real time. To the best of our knowledge, such a setting has not yet been proposed in research, as HCI is typically simulated in APE (e.g. [21, 24]). This approach would require implementing an interface to collect human corrections. Such a system would be similar to an IMT model (e.g. CASMACAT), but it would not require continuously updating the MT system parameters to make translation predictions. Instead, it would incrementally learn real-time PE actions to suggest corrections.

Investigating the benefits of such an interactive scenario would be insightful to design more effective IPE environments. Several alternatives are conceivable for IPE, as the translator can be shown either autocompletion suggestions which are updated when new prefixes are entered (this would be similar to IMT) or an entire MT correction suggestion which can be accepted or edited.

Consequently, it appears relevant to study the number of edits necessary to make the system responsive enough as well as to examine translators' experiences. More specifically, comparing different settings, such as a single correction suggestion versus a list of n-best suggestions, and other adjustments (e.g. the length of predictions), as suggested in by Barrachina et al. [25], would be beneficial. Measuring the effort (e.g. time and keystroke logging as in [26]) in each case would also help to formulate IPE models which would suit user needs. Nevertheless, this would require building a rather complex system, which might be difficult due to time and cost constraints.

4.2 Online APE Model

The second option is to simulate human corrections by adopting an online learning method using pre-existing post-edited texts (such as the datasets made available for the WMT APE shared tasks). While this has been done in previous work, only Negri et al. [24] have implemented this method on NMT outputs. This therefore leaves room for further experiments. It should be noted that Negri et al. worked with eSCAPE (Synthetic Corpus for Automatic Post-Editing [27]), a synthetic corpus designed to train APE systems. While eSCAPE is a valuable resource (it is freely available and contains over 7 million triplets), the post-edited segments are artificial (they are pre-existing translations). This is understandable due to the scarcity of training data for APE. However, since the MT output and its corresponding ‘post-edited’ version are likely to be very different (or at least more distant than in a real PE scenario), a system created using this might be prone to overcorrections. Therefore, it seems relevant to examine the performance of a similar system using training data in which the PE side is comprised of corrections performed by human translators. As pointed out by Ortiz-Martínez [28], further research on online APE would also benefit from corpora containing documents with a high rate of repetitions, which could serve to examine the performance of online APE models in technical translation.

5 Conclusion

This paper discussed how translation technologies could benefit from the creation of more interactive PE environments. Two methods were proposed, both differing from traditional batch training approaches. The first method consists of creating an interactive environment for PE, and the second entails the use of already-available post-edited translations to simulate human interaction. In both cases, the systems would be based on incremental learning and would learn from corrections in a continuous feedback loop, thus suggesting corrections based on the PE actions observed.

References

1. Hutchins, J.: The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954 (2005).
2. Knight, K., Chander, I.: Automated postediting of documents. In: Proceedings of AAAI 1994, pp. 779-784. AAAI Press, Seattle, Washington, USA (1994).
3. Allen, J., Hogan, C.: Toward the development of a post editing module for raw machine translation output: a controlled language perspective. In: Kuhn, T., Fuch, N. E. (eds.) Proceedings of the third International Controlled Language Applications Workshop (CLAW), pp. 62-71. Springer, Zurich, Switzerland (2000).
4. Llitjós, A. F., Carbonell, J. G.: Automating post-editing to improve MT systems. Institute for Software Research (2006).

5. do Carmo, F., Shterionov, D., Moorkens, J., Wagner, J., Hossari, M., Paquin, E., Schmidtke, D., Groves, D., Way, A.: A review of the state-of-the-art in automatic post-editing. *Machine Translation*, pp. 1-43 (2020).
6. Junczys-Dowmunt, M.: Are we experiencing the golden age of automatic post-editing? In: *Proceedings of the AMTA 2018: Workshop on Translation Quality Estimation and Automatic Post-Editing*, pp. 144-206. Publisher, Boston, Massachusetts, USA (2018).
7. Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huang, S., Huck, M., Koehn, P., Liu, Q., Logacheva, V., Monz, C., Negri, M., Post, M., Rubino, R., Specia, L., Turchi, M.: Findings of the 2017 conference on machine translation (WMT17). In: *Proceedings of the Conference on Machine Translation (WMT), Volume 2: Shared Task Papers*, pp. 169–214. Association for Computational Linguistics, Copenhagen, Denmark (2010).
8. Chatterjee, R., Negri, M., Rubino, R., Turchi, M.: Findings of the WMT 2019 shared task on automatic post-editing. In: *Proceedings of the Third Conference on Machine Translation, Volume 3: Shared Task Papers*, pp. 710–725. Association for Computational Linguistics, Belgium, Brussel (2018).
9. Foster, G., Langlais, P., Lapalme, G.: User-friendly text prediction for translators. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 148-155. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA (2002).
10. Esteban, J., Lorenzo, J., Valderrábanos, A. S., Lapalme, G.: Transtype2 – an innovative computer-assisted translation system. In: *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pp. 94-97. Association for Computational Linguistics, Barcelona, Spain (2004).
11. Toselli, A. H., Vidal, E., Casacuberta, F.: *Multimodal interactive pattern recognition and applications*. Springer Science & Business Media (2011).
12. Alabau, V., Bonk, R., Buck, C., Carl, M., Casacuberta, F., García-Martínez, M., González, J., Koehn, P., Leiva, L., Mesa-Lao, B., Ortiz, D., Saint-Amand, H., Sanchis, G., Tsoukala, C.: CASMACAT: An open-source workbench for advanced computer aided translation. *The Prague Bulletin of Mathematical Linguistics*, 100, pp. 101-112 (2013).
13. Wuebker, J., Green, S., DeNero, J., Hasan, S., Luong, M. T.: Models and inference for prefix-constrained machine translation. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 66-75. Association for Computational Linguistics, Berlin, Germany (2016).
14. Knowles, R., Koehn, P.: Neural interactive translation prediction. In: Green, S., Schwartz, L. (eds.) *Proceedings of AMTA 2016*, vol. 1: Researchers’ track, pp. 107-120. Association for Machine Translation in the Americas, Austin, Texas, USA (2016).
15. Santy, S., Dandapat, S., Choudhury, M., Bali, K.: INMT: Interactive neural machine translation prediction. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pp. 103-108. Association for Computational Linguistics, Hong Kong, China (2019).
16. Peris, Á., Casacuberta, F.: Online learning for effort reduction in interactive neural machine translation. *Computer Speech & Language*, 58, pp. 98-126 (2019).
17. Casacuberta, F., Civera, J., Cubel, E., Lagarda, A. L., Lapalme, G., Macklovitch, E., Vidal, E. Human interaction for high-quality machine translation. *Communications of the ACM*, 52(10), pp. 135-138 (2009).
18. Underwood, N. L., Mesa-Lao, B., García-Martínez, M., Carl, M., Alabau, V., González-Rubio, J., Leiva, J. A., Sanchis-Trilles, G., Ortiz-Martínez, D., Casacuberta, F. Evaluating

- the effects of interactivity in a post-editing workbench. In: *Proceeding of LREC*, pp. 553-559. European Language Resources Association, Reykjavik, Iceland (2014).
19. Alves, F., Koglin, A., Mesa-Lao, B., García Martínez, M., de Lima Fonseca, N. B., de Melo Sá, A., Gonçalves, J. L., Sarto Szpak, K., Sekino, K., Aquino, M. Analysing the impact of interactive machine translation on post-editing effort. *New directions in empirical translation process research*, pp. 77-94. Springer, Cham (2016).
 20. Lagoudaki, E.: The value of machine translation for the professional translator. In: *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas*, pp. 262-269. Association for Machine Translation in the Americas, Waikiki, Hawaii, USA (2008).
 21. Simard, M., Foster, G.: Pepr: Post-edit propagation using phrase-based statistical machine translation. In: Sima'an, K., Forcada, M.L., Grasmick, D., Depraetere, H., Way, A. (eds.) *Proceedings of the XIV Machine Translation Summit*, pp. 191-198. International Association for Machine Translation and the European Association for Machine Translation, Nice, France (2013).
 22. Lagarda, A. L., Ortiz-Martínez, D., Alabau, V., Casacuberta, F.: Translating without in-domain corpus: Machine translation post-editing with online learning techniques. *Computer Speech & Language*, 32(1), pp. 109-134 (2015).
 23. Chatterjee, R., Gebremelak, G., Negri, M., Turchi, M. Online automatic post-editing for MT in a multi-domain translation environment. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 525-535. Association for Computational Linguistics, Valencia, Spain (2017).
 24. Negri, M., Turchi, M., Bertoldi, N., Federico, M.: Online neural automatic post-editing for neural machine translation. In: Cabrio, E., Mazzei, A., Tamburini, F. (eds.) *Proceedings of the 5th Italian Conference on Computational Linguistics*. CEUR-WS, Torino, Italy (2018).
 25. Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomás, J., Vidal, E., Vilar, J. M. Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1), pp. 3-28 (2009).
 26. Aziz, W., de Sousa, S. C. M., Specia, L.: PET: a tool for post-editing and assessing machine translation. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, pp. 3982-3987. European Language Resources Association, Istanbul, Turkey (2012).
 27. Negri, M., Turchi, M., Chatterjee, R., Bertoldi, N.: eSCAPE: a Large-scale Synthetic Corpus for Automatic Post-Editing. In: *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association, Miyazaki, Japan (2018).
 28. Ortiz-Martínez, D. Online learning for statistical machine translation. *Computational Linguistics*, 42(1), pp. 121-161 (2016).