Human-Adapted MT for Literary Texts: Reality or Fantasy?

Damien Hansen
LIG-GETALP, Université Grenoble Alpes, 38000 Grenoble, France
CIRTI, Université de Liège, 4020 Liège, Belgium
damien.hansen@univ-grenoble-alpes.fr

Emmanuelle Esperança-Rodier
LIG-GETALP, Université Grenoble Alpes, 38000 Grenoble, France
Emmanuelle.Esperanca-Rodier@univ-grenoble-alpes.fr

Type of submission: Oral presentation
Word count: 717

Keywords: machine translation, human-adapted MT, MT evaluation, literary translation, fantasy

As the interest for the use of new technologies in the literary field continues to grow, in particular with the subject of literary machine translation (Toral and Way, 2018; Matusov, 2019; Kuzman et al., 2019), our recent work has focused on building an MT system specifically tailored to the work of a particular author and translator in the fantasy genre. This personal MT engine leverages the advances of the latest Transformer architecture (Vaswani, 2017) and was developed using the default parameters of the model (6 layers, model dimension of 512, feed-forward layers of size 2048, 8 attention heads, dropout of 0.1) for 200,000 steps. The actual training was carried out with the publicly available OpenNMT toolkit (Klein et al., 2017), meaning that it could be reproduced by anyone, as were the training sets we used, namely *Books*, *Europarl-v8*, *GlobalVoices*, *News-Commentary-v16*, and *TED2020* (Tiedemann, 2012). The resulting model was then fine-tuned on a much smaller subset of custom training data made up of 45,000 segments, all taken from the same literary saga (Sage, 2005–2013; Serval, 2005–2013).

Actualizing a research avenue that had not been explored for this combination since Besacier (2014), we tested our MT engine on a particularly challenging text for the English–French pair. This work is indeed characterized by its relatively free translation, a great variation in register among characters, invented words, language inspired by old or regional French and other typical challenges of fiction novels. Our results nevertheless showed that it was possible to adapt such a system not only on literary texts, but also on a specific genre, series and, more importantly, on the writing and style of human translators. In light of these findings, the aim of the present paper is to delve more deeply into the actual productions of our human-adapted MT system through various automatic and human evaluation methods. With these more fine-grained analyses, we intend to highlight the improved features and remaining drawbacks of such a tool and, in doing so, to broach questions of creativity (Guerberof-Arenas and Toral, 2020) and translator's voice (Kenny and Winters, 2020).

These confirm, on the one hand, that an engine tuned on literary data leads to improved scores in terms of lexical richness (Covington and McFall, 2010) or syntactic freedom (Vanroy et al., 2021), for instance. What is more, a review and error classification task performed on the output revealed that it reproduced not only typical patterns and word combinations, but also meaningful choices from the translator. The custom classification scheme specifically developed for this literary translation was inspired by those described in Vilar et al. (2006), Tezcan et al. (2019), and Schumacher (2020), with two main goals in mind: to reduce as much as possible the number of categories so as to simplify the annotation process, all the while highlighting the specific issues that an MT system could be facing in the field of literature. It further indicated that there remained, unsurprisingly, many errors that show the necessity of human intervention and the added value of human translation. These mistakes, however, are already well documented for MT and not so much related to the literary nature of the translation, since they primarily involve lexical choices for common words, use of determiners and literal translations.

New Trends in Translation and Technology, pages 78–80
4–6 July, 2022. ©NeTTT 2022

From these examinations, we make two observations, beginning with the dependence of NMT systems on training data. This does not only apply to specific domains, for which more and more custom engines are created and put on the market today, but also, as we think, to translator style. Indeed, we were able to trace back every surprising outcome in the target text, whether positive or negative, to the training data, contradicting the metaphor of the black box that was often used with the first apparition of NMT engines. Clever paraphrases and particularly free translation of certain phrases, for example, were systematically used by the translator in previous novels, whereas hallucinations of the NMT system could similarly be traced to specific linguistic phenomena in the training set. Secondly and consequently, our experience suggests, in our opinion, a new perspective for the use of MT. One of a personalized and interactive tool much better suited to the translation of creative texts, which would truly empower rather than constrain translators and leave them in control of the entire process.

References:

Besacier, Laurent. (2014). Traduction automatisée d'une œuvre littéraire : une étude pilote. In Proceedings of TALN 2014 (pp. 389–394).

Covington, Michael A., and Joe D. McFall. (2010). Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). Journal of Quantitative Linguistics 17 (pp. 94–100).

Matusov, Evgeny. (2019). The Challenges of Using Neural Machine Translation for Literature. In Proceedings of the Qualities of Literary Machine Translation (pp. 10–19).

Guerberof-Arenas, Ana, and Antonio Toral. (2020). The impact of post-editing and machine translation on creativity and reading experience. Translation Spaces 9-2 (pp. 255–282).

Kenny, Dorothy, and Marion Winters. (2020). Machine translation, ethics and the literary translator's voice. Translation Spaces 9-1 (pp. 123–149).

Klein, Guillaume, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. (2017). OpenNMT: Open-Source Toolkit for Neural Machine Translation. In Proceedings of ACL 2017, System Demonstrations (pp. 67–72).

Kuzman, Taja, Špela Vintar, and Mihael Arčan. (2019). Neural Machine Translation of Literary Texts from English to Slovene. In Proceedings of the Qualities of Literary Machine Translation (pp. 1–9).

Schumacher, Perrine. (2020). Post-édition et traduction humaine en contexte académique : une étude empirique. Transletters 3 (pp. 239–274).

Tiedemann, Jörg. (2012). Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (pp. 2214–2218).

Toral, Antonio, and Andy Way. (2018). What Level of Quality can Neural Machine Translation Attain on Literary Text? In Joss Moorkens, Sheila Castilho, Federico Gaspari and Stephen Doherty (eds). Translation Quality Assessment: From Principles to Practice. Cham: Springer (pp. 263–287).

Sage, Angie. (2005–2013). Septimus Heap. 7 vols. London: Bloomsburry Publishing.

Serval, Nathalie. Translator. (2005–2013). Magyk. By Angie Sage. 6 vols. Paris: Albin Michel.

Vanroy, Bram, Orphée De Clercq, Arda Tezcan, Joke Daems, and Lieve Macken. Metrics of syntactic equivalence to assess translation difficulty. In Michael Carl (ed). Explorations in empirical translation process research. Vol. 3. Cham: Springer (pp. 259–294).

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. (2017). Attention is All you Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 6000–6010).

Vilar, David, Jia Xu, Luis Fernando D'Haro, and Hermann Ney. (2006). Error Analysis of Statistical Machine Translation Output. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (pp. 697–702).