SPECIAL FEATURE: AI AND LITERARY TRANSLATION

What's new on the "unlikely frontier"?

Recent research on AI for literary translation

Waltraud Kolb

Academic research on literary MT[1] has picked up pace since the mid-2010s on both sides of what Miguel Jiménez-Crespo[2] has recently called the "unlikely frontier", the meeting point of technology/computational linguistics and literary translation studies. Indeed, a dialogue between the two disciplines is getting underway, with researchers starting to take note of each other's work and interdisciplinary projects being initiated.

How to evaluate machine translations?
One central concern for both fields is MT quality and how to evaluate it. Automatic metrics, which compare MT output with human reference translations, are widely used, as they are much faster and less costly than manual evaluation. However, they do not allow for an in-depth assessment of literary translations, and recent studies tend to rely also (or exclusively) on manual evaluation, sometimes supplemented with corpus-linguistic analysis. Manual evaluation is a painstaking process of identifying and categorising MT errors and shortcomings; for literary texts, a number of taxonomies have been developed which take into account not just accuracy and fluency, as commonly used in non-literary contexts, but also features such as text-level coherence, cohesion, cultural references, style, or register.

For some language pairs, researchers have trained MT systems with huge amounts of literary texts (for example, Antonio Toral and Andy Way did a study published in 2018 where they used over 100 million words of literary text to train their English-to-Catalan engine), and such literary-adapted systems have indeed been found to outperform general-domain engines such as Google Translate. Another avenue that has been

---

[1] The term 'literary MT' is certainly controversial, as it might not only be read as MT of literature but also as MT output that has literary qualities of its own.

[2] All references mentioned in this article and a list of suggestions for further reading can be found here.

CEATL
European Council of Literary
Translators' Associations

explored more recently is 'personalising' MT by training generic systems not so much with a large corpus of different literary texts but rather with texts by one particular author and translator.

However, results regarding the performance of specific MT systems are difficult to compare as they depend on the study design and many variables, not least who the evaluators are: native speakers without a translation background, translation students, or professional (literary) translators? 'Success rates' indicating how much of the MT output is considered acceptable therefore vary quite a bit, with numbers frequently ranging between 30% and 40%. At the same time, studies have also revealed that literary post-editors hardly leave a sentence generated by MT untouched and tend to go well beyond achieving accuracy and fluency. Areas that are known to still pose a particular challenge to MT systems include context awareness, cohesion, reference, especially beyond sentence boundaries, ambiguity and polysemy, style, register, rare or unknown words, orthographically similar words, literalism, and omissions. Of course, research to improve the quality of MT in all these areas is ongoing.

How efficient is post-editing?

Raw MT output invariably needs to be post-edited, therefore the impact of different post-editing modalities on post-editing quality and user-friendliness are being explored. Apart from traditional post-editing or using a CAT-tool environment for post-editing, various forms of interactive post-editing are being investigated. An example is DeepL, which displays

alternative options for a word or phrase and automatically updates the rest of the text if such an alternative is clicked on; other interactive systems react to what the translator types in and then come up with translation completion proposals in real-time, which can be accepted or rejected. So far, no modality has emerged as clearly superior to others.

Regarding productivity gains of post-editing over human translation, findings are not clear-cut either. While many studies find that post-editing is faster than human translation, we need to account for the fact that they usually take place in research settings, which might skew results. Also, there is great inter-subject variation when it comes to working speed, and issues such as MT quality or working contexts also have a bearing.

The few real-life studies that we have so far on publication projects involving post-editing show that scenarios can differ considerably.
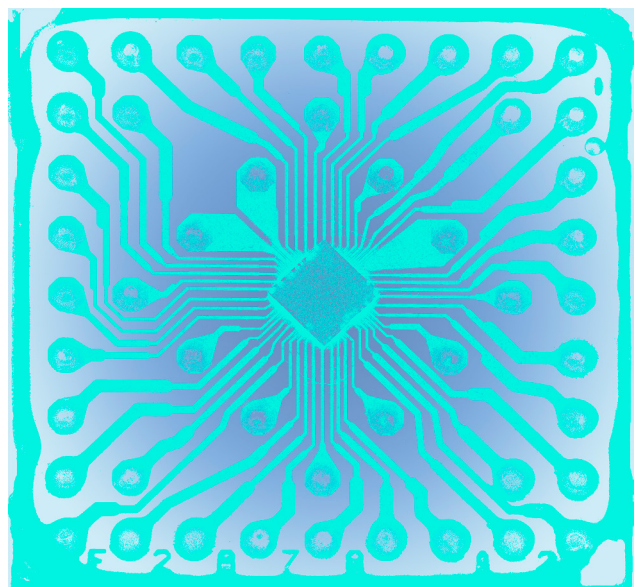
By Kaisa Ranta, based on a photo by Mister_fr *CC-SA*

In a 2022 study by Lieve Macken and colleagues of the workflow used by a book-translation company, a Dutch MT of an English literary source text was first post-edited by a professional literary translator and, as a second step, revised (mainly monolingually) by a different translator. In this two-pronged approach, the post-editor seemed to have chiefly been responsible for correcting MT errors, while the reviser's task then was to make the text more readable and acceptable for the target audience, with more edits actually made by the reviser than by the post-editor.

"*More edits were actually made by the reviser than by the post-editor*"

In one of my own studies earlier this year, I also investigated a real-life post-editing project, in this case a translation via post-editing from Brazilian-Portuguese into German. Here, the workflow was very different, with the post-editor being responsible for all aspects of the final product, much like a translator in a traditional scenario. It therefore did not come as a surprise that the post-editing process in this case was far from a fluent and straightforward affair, comprising at least six full-text (plus a number of partial) editing rounds. To compare post-editing with human translation, effort is also measured in terms of keystrokes and pauses. Post-editing can be expected to require fewer keystrokes to type in content, but more use of navigation

and erase keys. Similarly, process studies indicate that there are fewer pauses made in post-editing than in human translation. As pauses are commonly associated with cognitive effort, it would follow that post-editing is less cognitively demanding.

The cognitive effort involved in post-editing is certainly something that needs to be explored in more depth and in real-life contexts, as some literary translators experience post-editing as more demanding and tiring than human translation (cf. feedback provided by participants in a recent German study initiated from within the community of literary translators).

Pause patterns can also be analysed with a view to creativity, as pauses are often linked to a period in a cognitive process in which a creative idea is being incubated. In a 2022 study on creativity in post-editing and human translation, Ana Guerberof-Arenas and Antonio Toral confirmed such a correlation between the number of pauses and the number of creative solutions in the target text – both were higher in human translation than in post-editing.

Do machine-generated translations speak post-editese?
An interesting question is whether post-edited texts exhibit certain linguistic features that set them apart from human translations (so-called post-editese). A number of such features have indeed been uncovered, e.g., more standard vocabulary, less lexical diversity and density, more standard and simpler syntax, more interference from the original – in other words, traces of unedited MT output or echoes of what Gys-Walt van Egdom

and Joke Daems in an article from 2021 call MT's "mechanical voice".

Such traces are also a result of MT priming effects. In a study, in which I compared how five literary translators translated a short story by Hemingway into German and five others post-edited a DeepL version, priming in post-editing occurred not only on the level of semantics and syntax, where we might expect it, but also in the interpretation of whole scenes in the narrative.

While in human translation, translators construct the meaning of words, phrases and scenes from scratch as they read the original, the MT already presents them with a ready-made interpretation, which post-editors tend not to question unless faced with an obvious error or inconsistency. As one result, the post-edited versions of Hemingway's story turned out more similar to each other than the translations made by human translators.

Priming effects thus have a substantial impact on the extent to which a translator's or post-editor's personal style or voice is present in a target text, and studies have shown that it is less manifest in post-edited texts than in human translations. One such study by Dorothy Kenny and Marion Winters, published in 2020, was presented in some detail in *Counterpoint* No. 4 by Hans-Christian Oeser, the translator who participated in it [insert link 3]. In a follow-up study of a real-life translation via post-editing of a full novel by the same translator the focus was not so much on the loss of personal style in post-editing but rather on ways in which a translator can assert

his voice to a certain degree through his edits (Winters and Kenny 2023).

In any case, Hans-Christian Oeser's remark that he intends to use MT in the future "only to spot-check and not over a wide area of text" is very much in line with feedback from participants in other studies, who in most cases say they prefer translating from scratch, feel primed by the MT draft and constrained in their creativity, and find post-editing cognitively and emotionally draining. At the same time, though, they also acknowledge that the MT output can occasionally serve as a source of inspiration.

When it comes to the use of translation technology by literary translators, surveys indicate that CAT-tools are used to some extent, especially by translators who work with them routinely in non-literary contexts. Not so MT. In a survey by Paola Ruffo, carried out in 2018 and published in 2022, only ten out of 150 respondents from 35 countries stated that they use MT or had used it at least once for a literary translation job (compared to 38 mentions of CAT tools), but this number will probably have risen over the last five years. Apart from MT, researchers have also looked into potential benefits of other CALT tools (CALT standing for 'computer-assisted literary translation', a term coined by a team of researchers at the University in Swansea), such as corpus tools for text analysis, text visualisation software, or tools assisting in the translation of puns and wordplay.

What do readers think?
How are texts produced in different modalities read by the target audience?

Waltraud Kolb is Assistant Professor of Literary Translation at the University of Vienna. She studied translation (English, French, Portuguese/German) and holds a PhD in comparative literature. One focus of her research is literary translation processes, machine translation and post-editing. She has been working as a freelance translator since 1985 and is a member of the executive board of the Austrian Association of Literary Translators.

Waltraud Kolb
Photo: Private archive

So far, data shows that it takes readers longer to read an MT version of a whole novel than a human translation, probably due to MT errors – an eye-tracking study tracing the impact of certain types of errors on the reading process is still ongoing (Colman et al. 2022). MT has also been found to rank lower than post-editing, human translation and originals when it comes to narrative engagement, enjoyment and translation reception. However, results seem to vary with languages: in a 2023 article, Guerberof-Arenas and Toral report that in their reception study Catalan readers clearly preferred human translation to the other modalities, while Dutch readers seemed to prefer post-editing over human translation, the (English) original scoring highest for engagement and enjoyment.

The impact of AI on literary translation has grown into a very active field of research, and only a few topics of interest have been mentioned here; others that come to mind are the use of ChatGPT, the use of machine-translated bilingual e-books by second-language learners, MT of neologisms or metaphors – or, importantly, ethical issues of AI use, such as authors' and translators' rights, transparency and accountability, environmental implications, or long-term impacts on language and translatorial skills. It remains to be seen what new encounters on this "unlikely frontier" will yield.

Two new interdisciplinary research projects will address a number of aspects of AI in literary translation, including current technological needs of literary translators.

The project *Narrative Text, Translator and Machine: In Search of User-Friendly Translation Technology for Literary Texts* will be led by Kristiina Taivalkoski-Shilov and funded by the Academy of Finland. The EU-funded project *Uncovering the Creative Process: From Inception to Reception of Translated Content Using Machine Translation* will be led by Ana Guerberof-Arenas.

CEATL
European Council of Literary
Translators' Associations