

# Study assesses the quality of AI literary translations by comparing them with human translations

November 8 2022, by Ingrid Fadelli

---

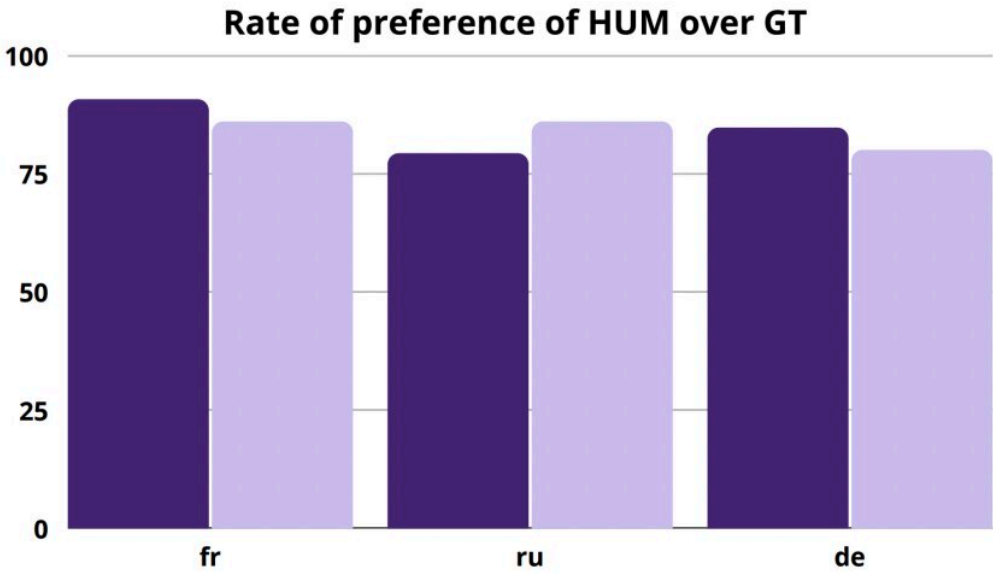


Figure 1: The percentage of cases in which raters preferred the human-written translation to the Google translation by source language. Note that the value for monolingual raters is the average of 3 percentages for 3 monolingual raters.

Credit: Thai et al.

Recent advancements in the field of machine learning (ML) have greatly improved the quality of automatic translation tools. At present, these tools are primarily used to translate basic sentences, as well as short texts or unofficial documents.

Literary texts, such as novels or short stories, are still fully translated by expert human translators, who are experienced in grasping abstract and complex meanings and translating them in another language. While a few studies have investigated the potential of computational models for translating literary texts, findings in this area are still limited.

Researchers at UMass Amherst have recently carried out a study exploring the quality of literary text translations produced by machines, by comparing them with same text-translations created by humans. Their findings, pre-published on arXiv, highlight some of the shortcomings of existing computational models to translate foreign texts into English.

"Machine [translation](#) (MT) holds potential to complement the work of human translators by improving both training procedures and their overall efficiency," Katherine Thai and her colleagues wrote in their paper. "Literary translation is less constrained than more traditional MT settings since translators must balance meaning equivalence, readability, and critical interpretability in the target language. This property, along with the complex discourse-level context present in literary texts, also makes literary MT more challenging to computationally model and evaluate."

The key objective of the recent work by Thai and her colleagues was to better understand the ways in which state-of-the-art MT tools still fail in

the translation of literary texts when compared to human translations. Their hope was that this would help to identify specific areas that developers should focus on to improve these models' performance.

"We collect a dataset (PAR3) of non-English language novels in the [public domain](#), each aligned at the paragraph level to both human and automatic English translations," Thai and her colleagues explained in their paper.

PAR3, the new dataset compiled by the researchers for the scope of their study, contains 121,000 paragraphs extracted from 118 novels originally written in different languages other than English. For each of these paragraphs, the dataset includes several different human translations, as well as a translation produced by Google translate.

The researchers compared the quality of human translations of these literary paragraphs with the ones produced by Google translate, using common metrics for evaluating MT tools. Concurrently, they asked expert human translators which translations they preferred, while also prompting them to identify issues with their least preferred translation.

"Using PAR3, we discover that expert literary translators prefer reference human translations over machine-translated paragraphs at a rate of 84%, while state-of-the-art automatic MT metrics do not correlate with those preferences," Thai and her colleagues wrote in their paper. "The experts note that MT outputs contain not only mistranslations, but also discourse-disrupting errors and stylistic inconsistencies."

Essentially, the findings gathered by Thai and her colleagues suggest that metrics to evaluate MT (e.g., BLEU, BLEURT, and BLONDE) might not be particularly effective, as human translators did not agree with their predictions. Notably, the feedback they gathered from human

translators also allowed the researchers to identify specific issues with translations created by Google translate.

Using the human experts' feedback as a guideline, the team ultimately created an automatic post-editing model based on GPT-3, a deep learning approach introduced by a research group at OpenAI. They found that expert human translators preferred the literary translations produced by this model at a rate of 69%.

In the future, the findings of this study could inform new studies exploring the use of MT tools to translate literary texts. In addition, the PAR3 dataset compiled by Thai and her colleagues, which is now [publicly available on GitHub](#), could be used by other teams to train or assess their language models.

"Overall, our work uncovers new challenges to progress in literary MT, and we hope that the public release of PAR3 will encourage researchers to tackle them," the researchers concluded in their paper.

**More information:** Katherine Thai et al, Exploring Document-Level Literary Machine Translation with Parallel Paragraphs from World Literature, *arXiv* (2022). [DOI: 10.48550/arxiv.2210.14250](https://doi.org/10.48550/arxiv.2210.14250)

© 2022 Science X Network

Citation: Study assesses the quality of AI literary translations by comparing them with human translations (2022, November 8) retrieved 24 April 2024 from <https://techxplore.com/news/2022-11-quality-ai-literary-human.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.