

Lost in literary machine translation? Explorations with more training data, literariness and going beyond isolated sentences

Research on machine translation (MT) for literary texts has become increasingly relevant in the last lustrum. In a popular strand in this area, in-domain data (i.e. literature) are used to tailor MT systems to the literary domain (Toral and Way 2015; Toral and Way 2018; Kuzman et al. 2019; Matusov 2019). A recent study shows that such an in-domain system delivers considerably better translations than a generic system for a language pair with scarce resources (English-to-Catalan) (Toral, Oliver and Ribas 2020).

In this project, we aim to 1) determine whether it is worthwhile to build a literary-adapted MT system for a better-resourced language pair (English-to-Dutch), for which generic systems (e.g. DeepL) are known to be competitive; and 2) to investigate the quality of the translations by such a system compared to those by a professional literary translator.

To achieve this, our literary-adapted MT system goes a step further in the amount of data used. The largest amount of such data used to date is 1.1 million bilingual sentence pairs (Toral and Way 2018) and 10 million monolingual sentences (Matusov 2019). We have, however, trained an English-to-Dutch MT system on over 400 English novels and their Dutch translations (5 million sentence pairs), and 5000 Dutch novels (22 million sentences). In addition, we go beyond translating isolated sentences by taking into account longer blocks of text.

To assess translation quality, we also go a step beyond what has been done before (e.g. relative ranking and postediting). Our automatic evaluation considers not only translation metrics (BLEU and COMET) but also the literariness (van Cranenburgh et al. 2019) of the resulting translations. Finally, our human evaluation includes a survey among readers to motivate their quality assessment, and a qualitative analysis of sentence- versus document-based MT systems.