
Literary Machine Translation to Produce Translations that Reflect Translators' Style

Mehmet Şahin^{*1}, Tunga Güngör^{*1}, Ena Hodzik^{*1}, Sabri Gürses^{*1}, Zeynep Yirmibeşođlu^{*1}, Harun Dallı^{*1}, and Olgun Dursun^{*1}

¹Bogazici University – Turquie

Résumé

In our presentation, we will be reporting on a research study* that involves building an English-Turkish machine translation (MT) model to translate literary texts whose style has been defined using corpus tools. We aim to determine the extent to which an MT system can reproduce a translator's style if it is trained on a corpus of previous translations by the same translator. Our presentation will focus on the following questions:

Does corpus stylistics allow us to better study the style of the translator?

Can we teach machines to replicate a translator's style?

To answer the first question, we adapted Leech and Short's (2007) methodology of literary style analysis to Turkish and used AntConc (Anthony, 2022) and WordSmith (Scott, 2021) as corpus visualization tools to analyze the style of a particular English-Turkish translator both qualitatively and quantitatively (Youdale, 2019). Our preliminary findings suggest consistent use of particular lexical and morphological forms as discourse markers. Our corpus analysis findings will be presented against a reference corpus representing Turkish literary translations between 1946 and 2015 in order to determine whether the stylistic features we identified are exclusive to our translator or general linguistic phenomena.

We have also applied text classification methods used in authorship attribution to identify different translators. One key question here is how these classifications are affected when two variables, namely author and translator, are involved instead of only the author.

To replicate a translator's style in machine translation, we built several machine translation models based on both LSTMs and transformers. We are using our own datasets consisting of chosen translators alongside publicly available bilingual corpora. We will conclude our presentation with the initial findings and implications for our future work.

*The study is part of a scientific research project funded by TÜBİTAK (the Scientific and Technological Research Council of Turkey). The title of the project is "LiteraryMachine Translation to Produce Translations that Reflect Translators' Style andGenerate Retranslations" and its grant number is 121K221.

*Intervenant

Mots-Clés: literary machine translation, translator style, corpus, based translation studies, stylistics, authorship