Smart Computer-Aided Translation Environment (SCATE): Highlights

Vincent Vandeghinste
Tom Vanallemeersch
Bram Bulté
Liesbeth Augustinus
Frank Van Eynde
Joris Pelemans
Lyan Verwimp
Patrick Wambacq
Geert Heyman
Marie-Francine Moens
Iulianna van der Lek-Ciudin
Frieda Steurs
KU Leuven
first.lastname@kuleuven.be

Ayla Rigouts Terryn
Els Lefever
Arda Tezcan
Lieve Macken
Ghent University
first.lastname@ugent.be

Sven Coppers
Jens Brulmans
Jan Van den Bergh
Kris Luyten
Karin Coninx
UHasselt – tUL – EDM
first.lastname@uhasselt.be

Abstract

We present the highlights of the now finished 4-year SCATE project. It was completed in February 2018 and funded by the Flemish Government IWT-SBO, project No. 130041.[1]

We present key results of SCATE (Smart Computer-Aided Translation Environment). The project investigated algorithms, user interfaces and methods that can contribute to the development of more efficient tools for translation work.

Improved fuzzy matching: Levenshtein distance is not the best predictor for post-editing effort. Linguistic metrics and different metrics (such as TER) combined show better results.

Integration of Translation Memory (TM) and Machine Translation (MT): Combining TM matches, fuzzy match repair and SMT shows improvements over a baseline SMT.

Informed Quality Estimation: Accuracy and fluency error detection systems form the basis of the sentence-level Quality Estimation system, which results in better correlations with temporal post-editing effort compared to the Quest++ baseline. Detected errors can additionally be highlighted in the MT output.

Identifying bilingual terms in comparable texts: We found improvements when combining word embeddings with character-based models using a neural classifier trained on a seed lexicon. This includes short multi-word term phrases.

Post-Editing via Automated Speech Recognition (ASR): ASR for post-editing can benefit from additional information sources, such as the source language, the MT translation model and the activation of domain-specific terminology, for which we boosted ASR language model probabilities. The ASR language model is also enriched with character-level information, making it possible to model out-of-vocabulary words, which are very common in new domains.

Intelligible Translator Interfaces: We iteratively developed a functional prototype that integrates several of the aforementioned translation aids. In contrast with other approaches, our system applies the design concept of intelligibility to support translators' decision-making process when they interact with their translation environment. The evaluation showed that the prototype allows translators to better evaluate translation suggestions from MT, TM and term base but it had no major impact on their performance in terms of speed and quality. Furthermore, a small-scale lab experiment revealed no significant difference in efficiency between translating with the prototype and with a commercial tool, which shows less suggestions by default.

Integration: We created an interactive demo so that translators can experience and evaluate our research results: http://scate.edm.uhasselt.be/.

[1] http://www.ccl.kuleuven.be/scate

Pérez-Ortiz, Sánchez-Martínez, Esplà-Gomis, Popović, Rico, Martins, Van den Bogaert, Forcada (eds.)
Proceedings of the 21st Annual Conference of the European Association for Machine Translation, p. 367
Alacant, Spain, May 2018.