Smart Computer Aided Translation Environment

IWT-SBO 130041

http://www.ccl.kuleuven.be/scate

List of consortium partners
University of Leuven (CCL - Centre for Computational Linguistics + ESAT/PSI - Centre for Processing of Speech and Images + LIIR - Language Intelligence & Information Retrieval + Thomas More Translation school)
University of Ghent (LT3 Language and Translation Technology Team)
University of Hasselt (Expertise Centre for Digital Media)

We aim at improving the translators' efficiency. Current commercial translation tools do not meet the productivity requirements imposed by the globalisation of business activities and the increasing information flow.

The usefulness of already translated data can be improved through using syntactic models for fuzzy matching, and retrieving syntactically similar constructions from the translation memory. We will investigate complex types of translation grammar induction and tree alignment that allow to *transduce* source syntax trees into target trees.. We will investigate how to seamlessly integrate MT into a translation memory, by automatically resolving the syntactic fuzziness of the match through MT techniques.

We will investigate translation evaluation in order to automatically judge whether MT output is worth post-editing, or whether the suggested translation can be applied to resolve the fuzzy match in the translation memory. We will build an annotated data set and a taxonomy of typical translation errors and combine this with loggings and analysis of human-machine interaction during post-editing, which should lead to improvements in automatic confidence estimations of machine translation output.

We will investigate automated terminology extraction from comparable corpora in order to speed-up the translation process and make translations more consistent. Therefore we will study translator's methods in acquiring domain terminology. We will also research methods to determine which texts in different languages contain comparable information, and we will improve current methods of terminology extraction from comparable corpora through techniques such as cross-lingual topic modelling.

We will investigate the interaction of speech recognition with machine translation by integrating the language model of the MT engine with the language model of the speech recogniser, in two directions. We will study the adaptation of the recogniser as input method for the post-editor, and we will study the translation of speech. Furthermore we will also study how to perform automatic domain-adaptation for speech recognition, in order to automatically adapt the language models of the recogniser to the domain.

We will investigate the work flows and personalised user interfaces in order to obtain a higher comfort and productivity for the translators, by analysing and modelling of current translation systems and translator's work flows and practices, investigating new visualisations of translation features, and developing and testing new interfaces for translation work.