CASMACAT: Cognitive Analysis and Statistical Methods for Advanced Computer Aided Translation

Philipp Koehn, University of Edinburgh, pkoehn@inf.ed.ac.uk
Michael Carl, Copenhagen Business School, mc.isv@cbs.dk
Francisco Casacuberta, Polytechnic University of Valencia, fcn@dsic.upv.es
Eva Marcos, Celer Solutions, eva.marcos@celersol.com
http://www.casmacat.eu/

Description

Productivity of human translators can be increased with computer aided translation (CAT) tools: translation memories are standard in the translation industry, but post-editing machine translation output is only slowly becoming an increasingly used practice. The current integration of machine translation technology into human translators' work processes is often done overly simplistic, breaks their work practices, and it is widely resisted. The CASMACAT project will build the next generation translator's workbench to improve productivity, quality, and work practices in the translation industry. Based on insights gained in the cognitive studies, novel types of assistance will be developed.

The CASMACAT project will carry out cognitive studies of actual unaltered translator behaviour based on key logging and eye tracking. The acquired data will aid understanding how interfaces with enriched information are used, help to determine translator types and styles, and to build a cognitive model of the translation process. Based on insights gained in the cognitive studies, the project will develop novel types of assistance to human translators and integrate them into a new workbench, consisting of an editor, a server, and analysis and visualization tools. The workbench will be designed in a modular fashion and can be combined with existing computer aided translation tools.

The project will develop new types of assistance along the following lines:
- *Interactive translation prediction*, where the CASMACAT workbench makes suggestions to the human translator how to complete the translation. The project will adapt the existing interactive machine translation paradigm by adding input modalities, especially electronic pens and basing the suggestions on better exploitation of novel statistical machine translation models, such as ones based on syntactic structure.
- *Interactive editing*, where the CASMACAT workbench provides additional information about the confidence of its assistance, integrates translation memories, and assists authoring and reviewing.
- *Adaptive translation models*, where the CASMACAT workbench learns from the interaction with the human translator by updating and adapting its models instantly based on the translation choices of the user.

The CASMACAT project, in close collaboration with the MATECAT project, will develop a web-based workbench for translators targeted at the European localisation industry. The project will demonstrate the workbench's effectiveness in extensive field tests of real-life practice of a translation agency. In addition, the project will also reach out to the wider language service industry and online volunteer translation platforms. The outcome of the CASMACAT project will be made available as open source software to industry, academia, and to individual end users.