FAUST: Feedback Analysis for User Adaptive Statistical Translation

William Byrne
University of Cambridge
`bill.byrne@eng.cam.ac.uk`

Lluís Màrquez
Technical University of Catalonia
`lluism@lsi.upc.edu`

http://faust-fp7.eu

The FAUST project[1] focused on developing machine translation (MT) technology that responds rapidly and intelligently to user feedback. The FAUST project is unusual in its focus on interactive machine translation with non-professional users. This presents unique challenges and opportunities to observe and model user interaction. The project objectives were: 1) Enhance the high-volume Reverso.net translation website with an experimental and evaluation infrastructure that will enable the study of instantaneous user feedback in MT; 2) Deploy novel web-oriented, feedback collection mechanisms that reduce noise in feedback provided by users and increase the utility of the web contributions; 3) Automatically acquire novel data collections to study translation as informed by user feedback; 4) Develop mechanisms for instantaneously incorporating user feedback into the MT engines that are used in production environments, such as those that power the Reverso.net website; 5) Create novel automatic metrics of translation quality that reflect preferences learned from user feedback; and 6) Develop new translation models driven by user feedback data and integrate natural language generation directly into MT to improve translation fluency and reduce negative feedback from users.

FAUST was a productive international collaboration between academia and industry. Project partners were University of Cambridge (UK), Technical University of Catalonia (Spain), Charles University (Czech Republic), Language Weaver / SDL (USA and Romania), and Softissimo Inc. (France).

All of the project objectives were met within the project schedule. Highlights are: 1) Development of a web-based computing infrastructure for deploying university and commercial research systems at the http://labs.reverso.net website. All user interactions with the systems were logged and collected into a feedback database that was distributed within the project for modeling user interaction with statistical translation; 2) Detailed analyses of the types of feedback provided and the effectiveness of user interface designs to encourage users to interact with the system and to contribute useful feedback. A community of users of the project tools was established and nurtured by Softissimo; 3) Tools for robust syntactic annotation of the types of data likely to be submitted for translation were developed for use within the project and for public dissemination; 4) The Asiya open toolkit for automatic evaluation of MT systems (http://asiya.lsi.upc.edu), which provides an online graphical interface, an online search facility for translation and translation feedback, and a RESTful web service; 5) The development of novel approaches to incorporate deep and shallow natural language generation into SMT. Improvements in translation are found through better morphology and generation in the target language; and 6) Softissimo integrates technology developed in FAUST into its Reverso Localise product, and SDL is planning to use the feedback exploitation mechanisms developed in the context of FAUST primarily in professional post-editing environments, where data is produced in massive quantities and can be trusted.