Evaluating Machine Translation in Use:
from Theory to Practice

FEMTI is a tool that helps evaluators of MT systems to define contextualized quality models, by relating the context of use of an MT system to the quality model used to evaluate it [1]. First defined by the Evaluation Working Group of the ISLE European/NSF (USA) project, FEMTI is in reality based on feedback obtained through several types of workshops[1] and has evolved in the past years to a web-based open tool for MT evaluation [2, 3], which has been used in workshops including hands-on exercises in MT evaluation.

This tutorial will focus on a number of typical contexts of use of MT systems and will explore the possibility of applying the FEMTI[2] framework to a real-world scenario. In addition, the tutorial will propose a hands-on exercise in which participants will use the FEMTI guidelines to describe a context of use and to propose a relevant quality model.

Although many MT developers and users were involved in the creation and improvement of FEMTI, an effort must still be made to extend its use towards corporate and individual users of MT, who often need guidelines and inspiration for establishing evaluation criteria, and would benefit from the use of a normalized reference such as FEMTI. To reduce the time needed to setup an evaluation, and to increase the completeness and applicability of FEMTI, it is now useful to define also a list of typical evaluation plans, i.e. typical scenarios of use accompanied by typical quality characteristics that should be evaluated in those cases, and possibly the most frequently used metrics. This tutorial, along with the exercise, will thus help to refine FEMTI to enable the definition of real environments of MT use, thus making this research initiative more applicable to practical situations and providing the community with a more valuable resource for evaluation, which is FEMTI enriched with typical use cases. The tutorial would help the community make progress on these issues in the typical bottom-up approach specific to the FEMTI normalization process.

The intended audience of the tutorial is people with experience in the use of MT systems, including professionals using MT systems as part of their daily work, researchers and developers of MT systems who would like to consider the actual uses of their systems in context.

The outline of the tutorial is as follows.
1. Introduction to FEMTI
2. Practical exercise
   - Description of the scenario of use of an MT system
   - Specification of a related quality model in FEMTI
3. Break
4. Synthesis of results by organizers and plenary discussion

---

[1] See a list at http://www.issco.unige.ch/projects/isle/ewg.html.
[2] http://www.issco.unige.ch/femti

References

[1] Hovy E. H., King M. and Popescu-Belis A. (2002). "Principles of Context-Based Machine Translation Evaluation". *Machine Translation*, vol. 17, n° 1, pp. 1-33.

[2] Estrella P., Popescu-Belis A. & Underwood N. "Finding the System that Suits you Best: Towards the Normalization of MT Evaluation". *Proceedings of the 27th International Conference on Translating and the Computer, ASLIB*, 24-25 November 2005, London.

[3] Popescu-Belis A., Estrella P., King M. & Underwood N. "A model for context-based evaluation of language processing systems and its application to machine translation evaluation". *Proceedings of LREC 2006 (Fourth International Conference on Language Resources and Evaluation)*, Genoa, Italy, p.691-696.

Organizing committee

Anthony Hartley
Centre for Translation Studies
School of Modern Languages and Cultures
University of Leeds
Web: http://www.leeds.ac.uk/cts/staff/tony_hartley.htm

Andrei Popescu-Belis
IDIAP Research Institute, Switzerland
Web: http://www.idiap.ch/~apbelis/

Maghi King
ETI/TIM/ISSCO,
University of Geneva, Switzerland
Web: http://www.issco.unige.ch/staff/king/

Paula Estrella
ETI/TIM/ISSCO
University of Geneva, Switzerland
Email: paula.estrella@issco.unige.ch
Web: http://www.issco.unige.ch/staff/estrella/