

11th International Conference on Theoretical and Methodological Issues in Machine Translation

Invited Talks

Rule-based and Statistical Machine Translation with a Focus on Swedish

Anna Sgvall Hein

Professor of Computational Linguistics,
Department of Linguistics and Philology
Uppsala University
P.O. Box 635
SE-751 26 Uppsala
Sweden

anna@lingfil.uu.se

In the talk, I will discuss the pros and cons of rule-based (RBMT) versus statistical machine translation (SMT). In particular, I will present data from running one system of each kind, Convertus and Pharaoh, on a corpus of automotive service literature from Scania CV AB. Translation goes from Swedish to English. The focus will be on the different kinds of errors that are typical of the two approaches, and how the errors may be identified and corrected. Some of them are language-independent whereas others are typical of the language pair in focus.

In addition, data from running Pharaoh on Europarl involving Swedish in relation to Danish, English, German, French, Spanish, Dutch, Portuguese, Italian, Greek and Finnish will be brought into the discussion, and from running Convertus on a corpus of university syllabi from Uppsala University. In the Europarl case, an RBMT system trained for the domain was not available, and in the syllabus case, a parallel corpus for the domain was not available. This seems to be the typical case in a situation where MT is needed. The choice between an RBMT and an SMT approach is constrained by what is available in terms of corpora for an SMT system, and language resources for an RBMT system. I will conclude by discussing how the two approaches may be combined.