Problems with the second generation architecture and new trends in MT

Harold Somers
Centre for Computational Linguistics
UMIST, Manchester, UK

ABSTRACT

In this tutorial presentation, we will first consider the characteristics of the typical "second generation" rule-based architecture for MT, and suggest some of the drawbacks of this approach, including development effort, robustness, scalability, and translation quality. There will be some discussion of the second-generation architecture and comparison with the architecture it was supposed to replace, with particular reference to currently available *commercial* systems.

The main focus of the tutorial will however be on a recently emerged new paradigm in MT, namely *example-based MT* (also known as analogy-based or memory-based MT).

The basic ideas behind example-based MT (EBMT) will be explained and exemplified. Various approaches will be discussed, ranging from the use of a corpus of translation examples within a more traditional architecture and in commercial systems ("Translation memory"), via so-called hybrid systems, through to recent attempts at wholly corpus-based MT system design.

Background reading:

Arnold, D., L. Balkan, R. Lee Humphreys, S. Meijer & L. Sadler, *Machine Translation: an Introductory Guide,* NCC Blackwell, Oxford, 1994, Ch. 10.

Isabelle, P. & S. Warwick-Armstrong, 'Les corpus bilingues: une nouvelle ressource pour le traducteur', in P. Bouillon & A. Clas (eds) *La Traductique: études et recherches de traduction par ordinateur,* Les Presses de l'Université de Montréal, 1993, pp. 288-306.

Somers, H.L. 'La traduction automatique basée sur l'exemple ou sur les corpus', in P. Bouillon & A. Clas (eds) *La Traductique: études et recherches de traduction par ordinateur,* Les Presses de l'Université de Montréal, 1993, pp. 149-166.