

Machine Translation and Morphologically-rich Languages: Research Workshop
of the Israel Science Foundation, University of Haifa, Israel, 23-27
January, 2011

Ido Dagan and Shachar Mirkin:
Textual entailment inference in machine translation

Translation is about generating semantically equivalent texts in a different language. Textual entailment is a framework for modeling semantic equivalences or implications between texts, typically within the same language. Taking a unified view, one could think of a process where a source sentence is first transformed to semantically equivalent or implied variants in the source language, then gets translated to a sentence in the target language which may be further converted to better-phrased semantic equivalents in the target-language. Such process may be uniformly viewed as sequentially generating semantically equivalent or implied forms, where some of the steps are intra-lingual (entailment) and some are cross-lingual (MT).

In this talk we first review the potential roles of textual entailment technology along such process. These include generating paraphrases or semantic generalizations (with some loss of information) of the source or target sentences, as well as comparing automatic and reference translations in MT evaluation. Then, we describe in detail a concrete work in which paraphrases and directional entailments (generalizations) were applied to overcome the problem of translating unknown words.

This is a joint work with Wilker Aziz, Nicola Cancedda, Mark Dymetman, Lucia Specia and Idan Szpektor.