

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

Morphological Processing and Word Reordering for Statistical MT of Highly Inflected Languages

Marcello Federico, FBK Trento, Italy

It is well known that phrase-based statistical MT performs best when source and target languages have similar sentence structure and morphology. My talk will overview recent work done at FBK on statistical MT from German, Arabic and Turkish to English. These languages are characterized by different morphological phenomena which require specific pre-processing methods to reduce data sparseness, to improve automatic word alignment at training time, and to induce better word reordering at decoding time. In particular, I will survey morphological decomposition and reduction techniques, as well as verb reordering methods that showed to positively impact on MT accuracy. Experimental results will be presented for the Arabic-English NIST 2009 task, the German-English WMT 2010 task, and the Turkish-English IWSLT 2010 task.