

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

Chris Dyer:  
Inflectional Morphology in Probabilistic Translation Models

In conventional translation models, words that differ from each other in any way are modeled independently of each other. From a modeling perspective, this is unsatisfying since closely related morphological forms of an underlying stem are likely to share many characteristics that are important for translation. And, more practically, this independence assumption means data sparsity is a significant issue in translation between morphologically complex languages.

I compare two new probabilistic translation models that relax this "lexical independence assumption" and share statistics across morphologically related word forms. The first model is generative, based on hierarchical Pitman-Yor processes, in which the translation distributions for different inflection variants of a stem share a common base distribution. The second model is based on Markov random fields and uses morphological features to share information across related forms.