

Title: Rich Features in Phrase-Based Machine Translation

Author: Kamil Kos

Department: Institute of Formal and Applied Linguistics

Supervisor: RNDr. Ondřej Bojar, Ph.D.

Supervisor's e-mail address: bojar@ufal.mff.cuni.cz

Keywords: machine translation, quality evaluation, source-context model, suffix array

Abstract:

In this thesis we investigate several methods how to improve the quality of statistical machine translation (MT) by using linguistically rich information. First, we describe SemPOS, a metric that uses shallow semantic representation of sentences to evaluate the translation quality. We show that even though this metric has high correlation with human assessment of translation quality it is not directly suitable for system parameter optimization. Second, we extend the log-linear model used in statistical MT by additional source-context model that helps to better distinguish among possible translation options and select the most promising translation for a given context.