Název práce: Rozsáhlé diskriminativní modely pro trénování strojového překladu do morfologicky bohatých jazyků
Autor: Miloš Stanojević
Katedra: Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic
Vedoucí diplomové práce: RNDr. Ondřej Bojar Ph.D.

Abstrakt: Diplomová práce se zabývá diskriminativními modely ve strojovém překladu do jazyků s bohatou morfologií. Shrnujeme současné přístupy a vypichujeme problém výběru slovních tvarů v cílovém jazyce a problém automatického odhadu kvality překladu jednotlivých vět. V našich pokusech s překladem z angličtiny do češtiny a srbštiny pak používáme morfologické i syntaktické rysy. Pro tento účel řešíme technické překážky, jak potřebné informace doručit k diskriminativnímu modelu: používáme jednoduchý tagging v průběhu překladu a promítáme zdrojové závislostní stromy na cílovou stranu.

Klíčová slova: diskriminativní modely, MIRA, řídké rysy, strojový překlad, vyhodnocování strojového překladu, ROUGE-S, projekce závislostní stromů

Title: Large-Scale Discriminative Training for Machine Translation into Morphologically-Rich Languages
Author: Miloš Stanojević
Department: Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic
Supervisors: RNDr. Ondřej Bojar Ph.D. and Mr Mike Rosner

Abstract: The thesis deals with large-scale discriminative models for machine translation into morphologically rich languages. We survey the approaches and highlight the problems of selecting word forms in the target language and evaluating translation quality at the sentence level. In our experiments with English-to-Czech and English-to-Serbian, we make use of morphological as well as syntactic features, solving necessary technical issues when bringing this information to the discriminative learner by tagging on the fly and projecting source dependency trees.

Keywords: discriminative training, MIRA, sparse features, machine translation, evaluation metrics, ROUGE-S, dependency tree mapping