

Vortrag 3: Kristina Smirnov „ Comparison of transfer Methods for low-resource morphology “. Betreuer: Katharina Kann

Die Motivation der Arbeit ist, dass man in der Computerlinguistik oft das Problem hat passende annotierte Daten zu finden. In ihre Arbeit versucht die Kristina zu schauen, in wie weit können die nicht annotierte Daten und die annotierte Daten anderer verwandten Sprachen dabei behilflich sein.

Die Aufgaben, für die die Korporas genutzt werden sind: gegeben lemma und target finde target; gegeben Form, source tag und target tag, finde target und gegeben Form, source tag finde target.

In der Arbeit werden folgende Kombinationen getestet und evaluiert: die annotierte Russische Beispiele mit annotierten Ukrainischen Beispielen, annotierten und nichtannotierten Russischen Beispielen und Kombination von alle drei Kategorien.

Die nicht annotierte Daten, die benutzt werden kommen aus Wikipedia und annotierte aus dem 2016 Shared Task on Morphological Reflection.

Es werden verschieden Kombinationen von LR und HR benutzt, wobei LR = 50, 200 und HR: 50,200,400,800,1600, 3200, 6400 und 12800.

Am Ende wird sowohl die statistische Evaluierung, als auch linguistische gemacht, um zu sehen, wo genau die Fehler auftreten