Vortrag von Alexander Vordermaier, Betreuerin Katharina Kann
Thema: Comparison of Transfer Methods for low Resource Morphology

Zunächst gibt der Vortragende einen Überblick über sein Thema. Kern der Arbeit ist die Paradigmen Komplettierung von Sprachen, d.h. die Zuordnung eines Lemmas zu seinen flektierten Formen. Für manche Sprachen gibt es jedoch wenige Ressourcen. Die Frage ist ob man mit einer ähnlichen Sprache das „low-resource“ Problem umgehen kann. Beispiel-Sprachen sind hier: Bulgarisch „high-resource“ und Mazedonisch „low-resource“ . Folgende drei Methoden werden für die Paradigmen-Komplettierung eingesetzt:

1. Sprachübergreifende Paradigmen-Komplettierung
2. Auto Encoding
3. Kombination aus 1. und 2.

Punkt 1 läuft folgendermaßen ab: Man mischt annotierte Daten der „low-resource“ Sprache mit annotierten Daten der „high-resource“ Sprache. Diese Daten werden dann von einem Modell trainiert. Punkt 2 ist ein einfaches Verfahren bei dem die Eingabe in der einen Sprache die Ausgabe in der anderen Sprache ist d.h. viele gleiche Wortflexionen. Meist ist das aber nicht der Fall. Für Punkt 3 nimmt man die „Datenmischung“ aus Punkt 1 und mischt diese mit unannotierten Daten der „low-resource“ Sprache. Die Daten sind aufgeteilt in Source und Target. Source enthält die Daten, die dem Modell übergeben werden. Target enthält die Daten die ausgegeben werden. Die Source Daten sind annotiert mit den Tags Sprache, Art des Wortes (z.B Lemma), Wortart. Je größer ein Datenpaket ist umso bessere Ergebnisse erzielt man. Weitere Aufgaben sind u.a. Fehlerquellen zu identifizieren.