

Vortrag von Christina Smirnov, BA-Betreuerin Katharina Kann
Thema: Comparison of Transfer Methods for Low-Resource Morphology

Zunächst gibt die Vortragende einen Überblick über das Thema. Im Kern geht es um die Entwicklung eines Modells, basierend auf einem Verschlüssler und Entschlüssler, das in Text-Daten Lemmas und flektierte Wortpaare findet. Ein ähnliches Modell wurde auch für das LMU System beim SIGMORPHON 2016 Projekt verwendet. Die Verschlüssler und Entschlüssler Methode ermöglicht es auch bei geringfügig vorhandenen Trainingsdaten Experimente auf Korpora durchzuführen und dabei brauchbare Ergebnisse zu erzielen. Dem System werden Parameter übergeben. Diese können z.B. Flexionen (Art- und Formmerkmale) sein. Als Rückgabewert erhält das System ein sog. „target“. Ein „target“ ist z.B ein Lemma in einer bestimmten Sprache mit vorgegebenen Art- und/oder Formmerkmalen. Auf Nachfrage was denn der Sinn sei ein „target“ zu erhalten, wenn die Eigenschaften schon vorgegeben sind, wurde erläutert, dass es sich um Sprachpaare bzw. Paare aus annotierten und unannotierten Trainingsdaten handelt. Hier:

- (i) annotierte russische Daten und annotierte ukrainische Daten
- (ii) annotierte russische Daten und unannotierte russische Daten
- (iii) gemischte Beispielskombinationen aus (i) und (ii)

Es wird also überprüft ob das definierte „target“ in einem Datenset auch so in einem anderen Testset vorkommt. Zur Evaluierung des Systems werden Accuracy, Precision, Recall und F-Score verwendet. Des Weiteren wird noch eine manuelle Fehleranalyse durchgeführt. Die Evaluation wurde bisher noch nicht ausgeführt, da noch keine Ergebnisse aus (i)-(iii) vorliegen.