

Protokoll 3 für die Sitzung vom 15.05.2017  
Ivana Daskalovska 11139620

**Thema:** Comparison of transfer methods for low-resource morphology

**Student:** Kristina Smirnov

**Betreuer:** Katharina Kann

Am Anfang der Präsentation hat die Frau Smirnov den Titel ihrer Bachelorarbeit und ihre Ideen für die Gliederung ihrer Bachelorarbeit vorgestellt.

Das Ziel ihrer Bachelorarbeit ist es anhand bestimmter Kriterien flektierte Wortformen im Russischen zu finden. Leider ist die Anzahl der annotierten Daten in diesem Gebiet für das Russische sehr gering. Aus diesem Grund musste sich Frau Smirnov Gedanken darüber machen, was sie als Hilfe benutzen kann.

Frau Smirnov hat entschieden das gesamte über das Ukrainische und über nicht annotierten Daten im Russischen zu tun.

Sie trainiert die Daten dabei in drei Kombinationen:

1. Russisch (annotierten Daten) mit Ukrainisch (annotierte Daten)
2. Russisch (annotierten Daten) mit Russisch (nicht annotierte Daten)
3. Kombination aus Russisch (annotierten Daten), Ukrainisch (annotierte Daten) und Russisch (nicht annotierte Daten)

Das Modell mit dem sie arbeitet wurde vom Herrn Hinrich Schütze und Frau Katharina Kann entwickelt.

Es handelt sich um einen Morphologischen Encoder-Decoder:

The LMU System for the SIGMORPHON 2016 Shared Task on Morphological  
Reinflection

Dieses Modell ist erfolgreich bei folgenden Aufgaben:

1. Lemma und Target gegeben  $\hat{=}$  Target finden
2. Sorceform, Sourcetag und Targettag gegeben  $\hat{=}$  Target finden
3. Sourceform und Sourcetag gegeben  $\hat{=}$  Target finden

Man gibt die Sprache und das Lemma ein und bestimmt in welcher grammatikalischen Form das Wort vorkommen soll. Das System soll anhand von diesen Informationen die flektierte Form des Lemma ausgeben.

Frau Smirnov erhöht stufenweise die Anzahl an Trainingsdaten.

Es werden dabei verschiedene Kombinationen aus

LR(Low Resource) = 50,200 und

HR(Hight Resource) =50, 200,400,800,1600,3200,6400,12800 Daten verwendet.

Resultate und Evaluation: Das System wurde bereits angewendet aber die Ergebnisse sind noch nicht da, sie folgen in den nächsten Tagen. Die Ergebnisse werden tabellarisch dargestellt. Als russischer Muttersprachler hat die Frau Smirnov vor, zusätzlich zu der statistischen Evaluation, die Ergebnisse manuell zu evaluieren. Ihr Ziel ist es eventuelle Regelmäßigkeiten und Ursachen für falsche Zuordnungen zu finden, die bei der Verbesserung des Systems möglicherweise behilflich sein können.