

# PROTOKOLLE ZU COMPUTERLINGUISTISCHES ARBEITEN

**Ines Röhrer**

Centre for Information and Speech Processing, LMU

`I.Roehrer@campus.lmu.de`

## 1 Referat

Das letzte Referat am 15.05. hält Kristina Smirnov. Ihre Bachelorarbeit wird von Katharina Kann betreut und trägt den Titel "Comparison of transfer methods for low resource morphology". Sie arbeitet mit einem Modell eines neuronalen En- und Decoder. Das Modell soll in einem Datenset Lemmas und flektierte Wortpaare finden. Hilfreich ist eine solche Methode dann, wenn nicht ausreichend Trainingsdaten einer Sprache verfügbar sind, da es möglich gemacht wird, Experimente auf Korpora anderer Sprachen zu testen. Außerdem wird es erleichtert autoannotierte Daten zu verwenden.

Übergibt man dem System verschiedene Informationen/Parameter, wie z.B. Flektionseigenschaften, wird ein "Target" zurückgegeben.

Das Modell hat verschiedene Funktionen, welche mehr oder weniger vollständig besprochen wurden. Eine dieser Funktionen besteht darin, dass dem Modell eine Sprache (z.B. Ukrainisch), ein Lemma, verschiedene "Target" Informationen, wie z.B. Nomen, Genitiv, Singular und das Lemma in der Zielsprache übergeben wird, woraufhin es dann das Vorkommen dieser Form im Text zurückgibt.

Für ihre Arbeit hat Kristina verschieden große Sets solcher Anfragen für verschiedene Korpora zu Verfügung. Ihre Aufgabe besteht aus drei Punkten, sie vergleicht die Ergebnisse, welche das Modell liefert für:

- Annotierte russische Daten mit annotierten ukrainischen Daten
- Annotierte russische Daten mit nicht annotierten russischen Daten
- Eine Kombination mit Beispielen aus allen 3 oben genannten Kategorien

Für ihre Evaluation verwendet sie sowohl statistische Methoden, indem sie Werte wie den Accuracy, Precision und F-Score berechnet und auswertet, als auch eine eigene Fehleranalyse. In dieser Fehleranalyse möchte sie, durch ihre Kenntnis von Ukrainisch und Russisch herausfinden, worin die Fehler bestehen die das System macht, sodass in Zukunft vielleicht sogar eine Verbesserung des Systems vorgenommen werden kann.

Zur Zeit wartet sie auf ihre Ergebnisse, damit sie in ihrem nächsten Schritt mit der Evaluation weitermachen kann.