

## Protokoll zur Sitzung am 15.05.2017

### Thema: Comparison of Transfer Methods for Low-Resource Morphology.

**Studentin: Kristina Smirnov**

**Betreuerin: Katharina Kann**

Die Studentin Kristina Smirnov schreibt ihre Bachelorarbeit im Rahmen des Projekts SIGMORPHON 2016-Shared Task-Encoder-Decoder. Die Dateien, die hier verwendet werden, enthalten das Lemma und die flektierte Form des Wortes (*target*).

Das Ziel dieses Projekts ist es, flektierte Wörter (hier als *target* bezeichnet) zu finden. Das Grundprinzip des Projekts ist es, dass wenn es nicht genügend Trainingsdaten zur Verfügung stehen, kann man auch Daten von anderen Sprachen oder der gleichen Sprache mit nicht annotierten Daten benutzen.

Der Zweck von dieser Bachelorarbeit ist es, diese beide Methode zu kombinieren und am Ende festzustellen, ob es funktioniert oder nicht.

Dafür präsentiert sie eine Liste von 3 generellen Tasks (**Blocktask 1**), bei denen das vorhandene Modell bereits erfolgreich ist:

Task 1: Lemma und *target* sind vorhanden → Finde *target*

Task 2: *Source* Form, *source* Tag und *target* Tag sind vorhanden → Finde *target*

Task 3: *Source* Form und *source* Tag sind vorhanden → Finde *target*

Die Vortragende geht von diesen 3 generellen Tasks aus, um drei neue spezifische Tasks zu erzeugen (**Blocktask 2**):

Task 1: annotierte Texte auf Russisch mit Ukrainisch kombinieren

Task 2: nicht annotierte Texte auf Russisch kombinieren

Task 3: Beide kombinieren

Sie kombiniert dann beide Blocktasks, indem sie für jeden Task des Blocktasks 1, alle drei Tasks des Blocktasks 2 anwendet.

#### **Data:**

Für jede Kombination findet man Trainingsset für *source* und *target*, Developmentset, Testset, Wortschatz für beide Trainingssets und die Länge des Wortschatzes.

LR(lower resource) = 50, 200

HR (higher resource) = 50, 200, 400, 800, 1600, 3200, 6400, 12800

### **Resultate und Evaluierung:**

Am Ende der Präsentation hat Kristina Smirnov erklärt, dass sie sich das Verfahren der Evaluierung selber ausgedacht hat. Sie hat es vor, folgende Aspekte zu überprüfen:

- Was für Fehler aufgetaucht sind?
- Kann man irgend ein Muster (engl.: pattern) finden, mit der man das Finden verbessern kann?

Smirnov hat ihr System bereits angewendet. Aber im Moment der Präsentation hatte sie die Resultate noch nicht bekommen. Sie werden in den nächsten Tagen erhalten.