

Referat Kristina Smirnov (Montag, den 15.05.2017)

## Comparison Of Transfer Methods For Low-Resource Morphology

Zunächst wurden das Overview zu dem Referat gezeigt.

Genaue Unterteilungen wurden hier aufgezählt. Als erstes „Objective“, dann das „Modell“, des weiteren die „TASK“, dann die „DATA“ und zuletzt die „RESULTS, EVALUATION“.

Zu „OBJECTIVE“: Hier wurden einige Fragen gestellt, welche hier abgearbeitet werden.

Die erste Frage hier lautet: „What can we make use of if there is not enough language data?“

Als nächstes wird gefragt, „Would it help to add a data of a similar language?“

Außerdem, „Would it help to add a non-annotated data of the same language?“

Und zuletzt wird hier gefragt, „Is it sensible to combine these two?“

Zu Model werden hier 3 verschiedene TASKS erläutert. Die erste TASK gibt die „lemma & target“ wieder und man soll die dazu gesuchte target suchen.

TASK 2, die „source form, source tag & target tag“ sind gegeben, und man soll die target finden.

Und zu TASK 3, hier ist die „source form & source tag“ gegeben, und man soll die target wiederfinden.

Nachdem das Model erklärt wurde, wurden die einzelnen TASKS erläutert.

TASK 1, man kombiniert annotierte russische Beispiele mit annotierten ukrainische Beispiele.

Zur TASK2. Hier kombiniert man annotierte russische Beispiele mit nicht annotierten russischen Beispiele.

Und zum Ende hin, die TASK 3. Hier kombiniert man Beispiele aus allen drei Kategorien.

Zu Data wurde erklärt wie viele genaue Informationen man hier nutzt, um für so ein TASK durchlaufen zu können. Sprich „Low Ressource“ werden 50,200 verwendet. Bei „High Ressource“ werden 50, 200, 400, 800, 1600, 3200, 6400, und 12800 angewandt.

Für jede Kombination werden „training set“, „development set“ und „test set“ gefunden.

Bei den „Results“ werden die Resultate auf einer graphischen Ebene verdeutlicht.

Bei der „Evaluation“ werden hier Fragen gestellt und auch versucht auf beste Weise zu antworten.

Die erste Frage lautet hier, was für Fehler hier entstehen können? Gibt es eigentlich Fehlermuster und können wir das System dadurch verbessern?