

Jakob Sharab  
CIS, LMU München  
Matr.Nr.: 11186993

## **Repetitorium zu Computerlinguistisches Arbeiten -**

### **Protokoll zur Sitzung vom 29.05.2017**

#### **Repetitorium: Testen von Signifikanz**

Zu Beginn erklärte Herr Roth, weshalb es Sinn macht seine Daten in Training-, Development-, und Testset zu unterteilen, woraufhin wir auch die Funktionen einzelnen der Sets klärten.

Anschließend ging es um die Bewertung verschiedener Systeme. Dazu meinte Herr Roth, dass wenn man zwei Systeme vergleicht, man wissen will, ob das eine tatsächlich besser ist als das andere, oder ob dies nur auf zufällige Faktoren zurückzuführen ist. Und wenn man nun die unterschiedlichen Ergebnisse hat, man auch wissen will, ob diese signifikant sind. Dafür, erklärte Herr Roth, wie man eine Nullhypothese  $H_0$  aufstellt und unter welchen Bedingungen man  $H_0$  verwirft oder beibehält. Danach gingen wir auf verschiedene Metriken ein, um unsere Systeme zu bewerten, wobei wir uns aus zeitlichen Gründen hauptsächlich auf die Accuracy konzentrierten, wobei wir auch kurz den Paired t-Test und den Vorzeichentest sprachen.