

# Vorträge leiten, Modulprüfungen machen und über NLP Schreiben

HS Sprachtechnologie für eine bessere Welt (Winter term 2022/23)

Nils Reiter,  
`nils.reiter@uni-koeln.de`

10. Januar 2023

# Vorträge – Reflexion

- ▶ Wie war es, einen interaktiven Teil vorbereiten zu müssen?
- ▶ Wie zufrieden seid Ihr damit wie es gelaufen ist?
- ▶ Welchen Ansatz findet Ihr aus Publikumperspektive am besten? Warum?
- ▶ Haben Euch Inhalte gefehlt? Welche?

## Themen

- ▶ Authorship obfuscation / attribution
- ▶ Hate Speech Detection
- ▶ Mental Health
- ▶ Fake News Detection

## Section 2

### Modulprüfungen

# Modul: Verarbeitung von Textdaten

## MA Informationsverarbeitung

Veranstaltung	Kontaktzeit	Selbststudium
Hauptseminar	30	60
Übung	30	60
Kolloquium	30	60
Modulprüfung	–	270

Tabelle: Lehrveranstaltungen im Modul

- ▶ 18 Leistungspunkte
- ▶ 30% der Fachnote

# Modul: Profilmodul Computerlinguistik (1C)

MA Linguistik

Veranstaltung	Kontaktzeit	Selbststudium
Hauptseminar	30	60
Projektseminar	30	60
Projektseminar	30	60
Modulprüfung	–	180

Tabelle: Lehrveranstaltungen im Modul

- ▶ 15 Leistungspunkte
- ▶ 40% der Fachnote

# Modulprüfungen

- ▶ Thema
  - ▶ Findung und Wahl: Ihre Aufgabe
  - ▶ Kann, muss aber nicht, etwas mit dem Seminar zu tun haben
  - ▶ Mit mir absprechen

# Modulprüfungen

- ▶ Thema
  - ▶ Findung und Wahl: Ihre Aufgabe
  - ▶ Kann, muss aber nicht, etwas mit dem Seminar zu tun haben
  - ▶ Mit mir absprechen
- ▶ Praktischer Anteil: Grundsätzlich offen. Beispiele:
  - ▶ Experiment zur automatischen Identifikation eines Textphänomens
  - ▶ Annotationsexperiment
  - ▶ Quantitativer Vergleich verschiedener Korpora
  - ▶ ...
- ▶ Am Ende: Hausarbeit von max. 8 S Länge
- ▶ Übung vor der Master-Arbeit

## Brainstorm

- ▶ Wer hat schon eine Idee?
- ▶ Alles ist erlaubt!
- ▶ Lasst Euch von Euren Kommiliton:inn:en inspirieren!

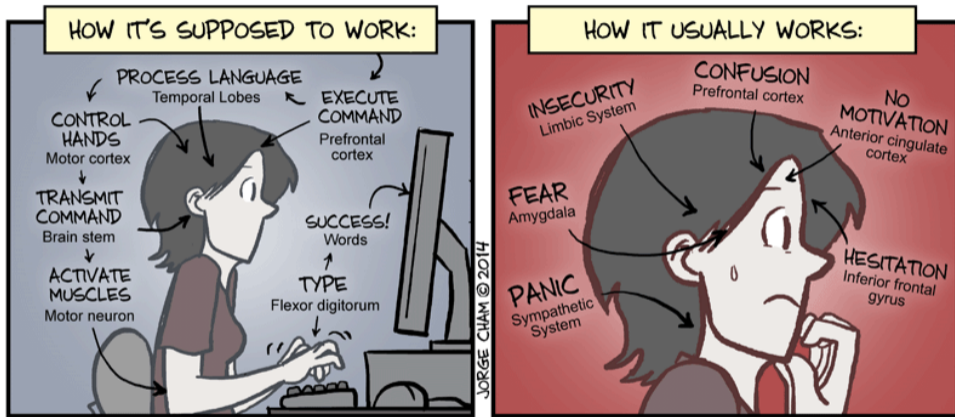


## Section 3

Schreiben über NLP-Experimente

Was zeichnet einen guten wissenschaftlichen Text aus?

# THE NEUROBIOLOGY OF WRITING



WWW.PHDCOMICS.COM

Abbildung: PhDComics, <https://phdcomics.com/comics/archive.php?comicid=1734>

## Schreiben über NLP-Experimente

- ▶ Schreiben ist Arbeit und braucht Zeit
- ▶ Iterativer Prozess aus Schreiben, lesen, überarbeiten, lesen, überarbeiten, ...
- ▶ (Wiss.) Schreiben ist eine Fähigkeit, die man lernen kann und muss
- ▶ Schreiben ist individuell: Mit der Zeit weiß man wie man funktioniert
- ▶ Kompetenzzentrum Schreiben: <https://schreibzentrum.phil-fak.uni-koeln.de>

## Schreiben über NLP-Experimente

- ▶ Schreiben ist Arbeit und braucht Zeit
- ▶ Iterativer Prozess aus Schreiben, lesen, überarbeiten, lesen, überarbeiten, ...
- ▶ (Wiss.) Schreiben ist eine Fähigkeit, die man lernen kann und muss
- ▶ Schreiben ist individuell: Mit der Zeit weiß man wie man funktioniert
- ▶ Kompetenzzentrum Schreiben: <https://schreibzentrum.phil-fak.uni-koeln.de>

### Verschiedene Aspekte

- ▶ Formalia (Layout, Zitate, Rechtschreibung)
- ▶ Stil
- ▶ Inhalt

# Formalia

- ▶ Grammatik und Rechtschreibung
- ▶ Layout, Schriftgrößen etc.
- ▶ Zitatwerk: Auszeichnungen im Text und Bibliographie

# Formalia

## Tools und Hilfsmittel

- ▶ Rechtschreib- und Grammatikprüfung
- ▶ Standardlayouts, Formatvorlagen: ‚Nichts Wildes‘
- ▶ Literaturverwaltung
  - ▶ Zotero, Endnote, BibTeX, ...

# Formalia

## Zitatwerk

Was kann schiefgehen?

- ▶ Keine Quellenangaben
- ▶ Zitate ohne Quelle, Quellenangabe unvollständig oder kaputt, ...
- ▶ Einer Quelle wird etwas ‚untergeschoben‘, was sie gar nicht gesagt hat
- ▶ Nicht zitierfähige Literatur



# Formalia

## Zitatwerk

### Was kann schiefgehen?

- ▶ Keine Quellenangaben
- ▶ Zitate ohne Quelle, Quellenangabe unvollständig oder kaputt, ...
- ▶ Einer Quelle wird etwas ‚untergeschoben‘, was sie gar nicht gesagt hat
- ▶ Nicht zitierfähige Literatur
  - ▶ ML-Kontext: ‚Gefahr durch Blogs‘
  - ▶ Grundlagen sollten aus Lehrbüchern o.ä. zitiert werden
  - ▶ Spezialkonzepte können auch via Blog zitiert werden, sollte aber Ausnahme sein
  - ▶ Wissenschaftliche Themen nie anhand populärwissenschaftlicher Quellen zitieren (Spektrum, F.A.Z., ...)

# Stil

- ▶ Wissenschaftliche Arbeiten sollen sachlich und präzise geschrieben sein
- ▶ Also:
  - ▶ Keine Wertungen wie z.B. „das wenig überzeugende Experiment von Müller (2017) ...“
  - ▶ Keine persönlichen Erfahrungsberichte
    - ▶ Dass jemand viel gelernt hat ist schön, sollte aber nicht im Text stehen
  - ▶ Gleiche Dinge gleich benennen
    - ▶ Auch wenn das zu Wortwiederholungen führt
  - ▶ Kein Geschwurbel
  - ▶ Zeitlosigkeit: Text muss auch außerhalb des Seminarkontextes funktionieren
- ▶ „Wörter auf die Goldwaage legen“

# Stil

## Geschwurbel

- ▶ Lange Sätze mit vielen Nebensätzen
- ▶ Unklare Referenzen (Pronomen)
- ▶ Steigerungsformen
  - ▶ Es ist fast nie wichtig, eine Datenmenge als „sehr groß“ im Gegensatz zu „groß“ zu bezeichnen
- ▶ Nichtssagende Füllwörter oder -Sätze
  - ▶ „Der Forschungsstand der Computerlinguistik in ihrer ganzen Breite ist zum aktuellen Zeitpunkt bereits sehr weit fortgeschritten.“

Welche Probleme haben die ausgeteilten Beispiele?

# Inhalt

- ▶ Leitlinien: Reproduzierbarkeit und Transparenz
- ▶ Experimente sollen so dokumentiert sein, dass sie überprüfbar sind
- ▶ Regelfall: Nicht alles was wir gemacht haben, landet im Artikel

# Inhalt

## Forschungsstand

- ▶ NLP-Papiere berichten über Fortschritt für eine bestimmte Aufgabe
- ▶ Forschungsstand gibt wieder, was man vor dem vorliegenden Papier wusste
- ▶ Konkret genannt werden Arbeiten:
  - ▶ die sich mit exakt dem gleichen Problem beschäftigt haben
  - ▶ die sich mit einem strukturell ähnlichen Problem beschäftigt haben
  - ▶ die eine Methode präsentieren, die für unser Problem einsetzbar sein könnte
- ▶ Fokus auf Methoden, nicht Tools

# Inhalt

## Häufige Probleme

- ▶ Unvollständige/ungenauere Informationen
- ▶ Abweichungen von der Gliederung
  - ▶ Z.B.: Im Abschnitt ‚Forschungsstand‘ geht es nicht um die eigene Arbeit
- ▶ Falsch verwendete Fachbegriffe
- ▶ Fehlende Konzentration auf das, was relevant ist
  - ▶ Nebenbei werden Fässer aufgemacht, die gar nicht nötig sind
- ▶ Zu wenig Abstraktion: Implementierungsdetails haben in NLP-Texten nichts verloren
- ▶ Selten: Experimente die nichts zeigen können, weil der Aufbau nicht durchdacht wurde

## (Meine) Tipps

- ▶ Überarbeiten – lesen – überarbeiten – lesen – überarbeiten – lesen – ...
- ▶ Bis zur Deadline ist alles im Fluss
- ▶ Man muss nicht auf Anhieb perfekte Sätze hinschreiben
- ▶ Nicht: Vor ein leeres Dokument setzen und Text hinschreiben
- ▶ Erst Notizen (was will ich eigentlich sagen?) machen, dann ausformulieren
- ▶ Den Text mal eine Woche liegen lassen und dann wieder lesen
- ▶ Am Anfang Kapitel schreiben, bei denen man weiß was man schreiben muss (Experimente)
- ▶ Einleitung und Schluss als letztes schreiben
- ▶ Denglish: Ein ML/NLP-Text ist durchsetzt von englischen Begriffen





WWW.PHDCOMICS.COM

This message brought to you by that manuscript you're supposed to be writing.

Abbildung: PhD Comics <https://phdcomics.com/comics.php?f=1785>

Fragen?