



UNIVERSITÄT  
ZU KÖLN

# Wissenschaftlich Schreiben (über NLP-Experimente)

## Anwendungen der Computerlinguistik

Nils Reiter

`nils.reiter@uni-koeln.de`

December 19, 2024

# Nächste Hausaufgabe

(bis 09.01.)

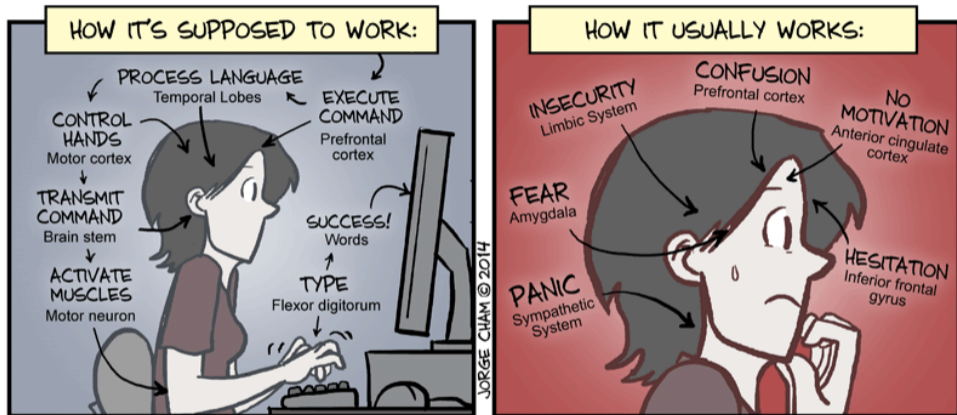
Schreiben Sie einen wissenschaftlichen Text über das von Ihnen geplante Experiment. Richten Sie den Text an jemanden, der allgemein mit CL-Verfahren vertraut ist, aber nicht mit ihrem konkreten Task.

Der reine Text sollte etwa eine A4-Seite lang sein (Schriftgröße 11, Times).

Literaturverzeichnis etc. kann dann länger sein. Stellen Sie sich vor, dass Sie die Arbeit ausgeführt hätten, und erfinden Sie Ergebnisse. Schreiben Sie dann einen Bericht über das Experiment, wie es hätte stattfinden können. Ja, dabei ist es erlaubt Dinge zu erfinden, sie müssen aber realistisch sein.

Wie unterscheiden sich wissenschaftliche Texte (die Sie kennen) von anderen Texten?  
Was zeichnet wissenschaftliche Texte aus?

# THE NEUROBIOLOGY OF WRITING



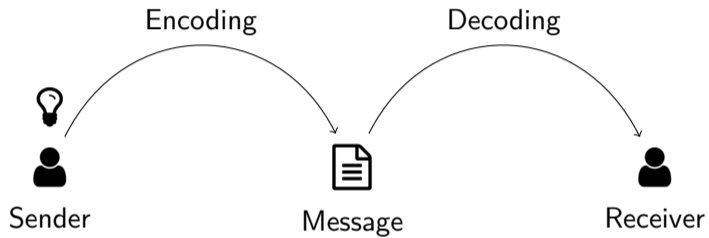
WWW.PHDCOMICS.COM

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

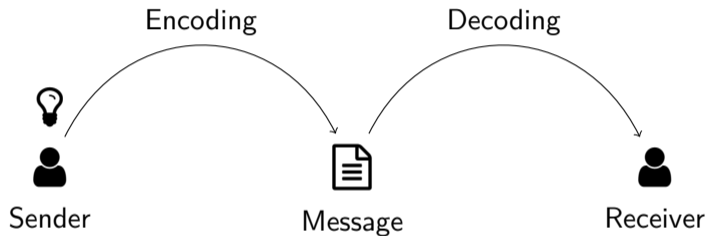
# Wissenschaftliches Schreiben

- ▶ 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
- ▶ Wissenschaftliche Texte sind Teil einer Kommunikation

# Kommunikation im Allgemeinen



# Kommunikation im Allgemeinen



- ▶ Schreiben: Kein direkter Feedbackkanal vom Receiver zum Sender
- ▶ Aufgabe der Schreibenden: Antizipieren, wie der Receiver liest

# Inhalt

- ▶ Verschiedene Aspekte
  - ▶ Formalia (Layout, Zitate, Rechtschreibung)
  - ▶ Stil
  - ▶ Inhalt
- ▶ Support: [Kompetenzzentrum Schreiben](#)



# 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
  - ▶ Quelle wird im Text genannt, fehlt aber im Literaturverzeichnis
  - ▶ Bei zusammenfassenden Aussagen ist oft nicht klar, ob die Zusammenfassung aus der Quelle kommt, oder die Quelle ein Beispiel ist
- “Lastly, numerous studies have integrated sentiment analysis with collocation analysis to examine how word associations influence emotional tone and help to improve the overall accuracy of the experiments (Hung and Cao 2020).”

# 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, hat aber im wiss. Text nichts zu suchen
  - ▶ 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”

- ▶ 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 ggf. überprüfbar sind
- ▶ Regelfall: Nicht alles was wir gemacht haben, landet im Artikel
- ▶ Artikel beschreibt einen etwas idealisierten Ablauf

Gliederung: Welche Teile brauchen wir?



# 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
- ▶ 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 – ...
- ▶ Mindset: Bis zur Deadline/Abgabe ist alles im Fluss
- ▶ Man muss nicht auf Anhieb perfekte Sätze hinschreiben
- ▶ Nicht: Vor ein leeres Dokument setzen und erwarten Text hinzuschreiben
- ▶ 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

# References I