



Wissenschaftlich Schreiben (über NLP-Experimente)

HS Experimentelles Arbeiten in der Sprachverarbeitung

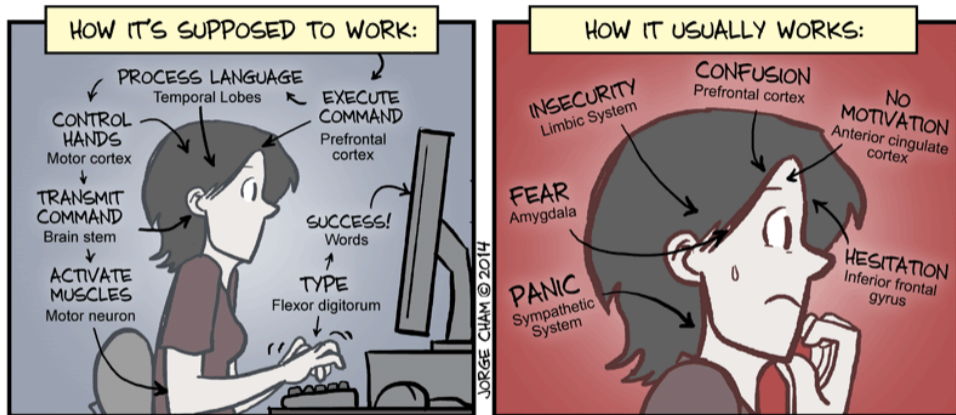
Nils Reiter

`nils.reiter@uni-koeln.de`

December 21, 2023

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

THE NEUROBIOLOGY OF WRITING



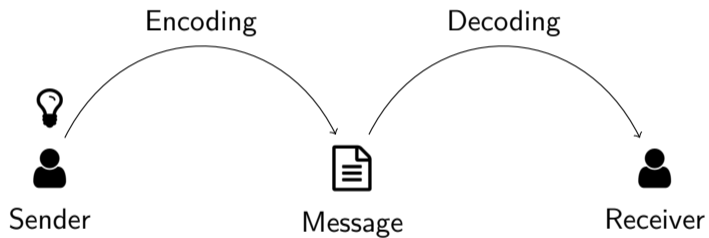
WWW.PHDCOMICS.COM

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



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

Tools und Hilfsmittel

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

Kleinigkeiten

- ▶ Adjektive klein, auch wenn sie Teil einer festen Wendung sind („überwachtes Machine Learning“)
- ▶ Screenshots nur wenn es gar nicht anders geht (Abbildungen oder Programme)
 - ▶ Nicht bei Tabellen

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 aus Blogs zitiert werden, sollte aber Ausnahme sein
 - ▶ Wissenschaftliche Themen nie anhand von 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, 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