

# Methoden zur digitalen Analyse sprachlicher Komplexität im Bildungskontext

---

Detmar Meurers und Zarah Weiss

Workshop bei der Jahrestagung des Mercator-Instituts, 18.02.2020

## 1 AGENDA

- 11:00 – 12:15 Einführung in die Analyse sprachlicher Komplexität
- 12:15 – 13:00 Einführung in die sprachliche Textanalyse mit CTAP
- 13:00 – 14:15 *Mittagspause*
- 14:15 – 15:15 Gruppen- oder Eigenarbeit mit CTAP
- 15:15 – 16:00 Diskussion der Ergebnisse und allgemeines Fazit

## 2 KOMPLEXITÄSANALYSEN MIT CTAP

**Anmerkung:** Sie finden alle Ressourcen für die hier beschriebenen Analysen unter

<http://www.sfs.uni-tuebingen.de/~zweiss/workshop.zip>

Bitte laden Sie die Datei herunter und entpacken Sie diese auf Ihrem Computer. Sollten Sie zu irgendeinem Zeitpunkt Fragen haben oder Hilfe benötigen, zögern Sie nicht sich an uns zu wenden.

### 2.1 Nutzung von CTAP

In diesem Schritt sammeln Sie eigene Erfahrungen mit der Nutzung von CTAP für Ihre Textanalysen. Sie werden ein Korpus in CTAP hochladen, ein eigenes Merkmal-Set erstellen und eine Analyse mit CTAP durchführen.

1. Loggen Sie sich mit Ihrem Account in CTAP ein (<http://ctapweb.com>). Sollten Sie noch keinen Account besitzen, registrieren Sie sich bei CTAP.

2. Navigieren Sie zum **Corpus Manager** und erstellen Sie das Corpus „Merlin (korrigiert)“. Öffnen Sie das Korpus und laden Sie die Dateien aus dem Ordner **data/merlin50-korrigiert/**, der sich in dem Material-Ordner befindet, den Sie heruntergeladen haben.
3. Navigieren Sie zum **Feature Selector** und erstellen Sie ein Merkmal-Set mit dem Namen „Workshop Merkmale“. Öffnen Sie das Merkmal-Set und fügen Sie die folgenden Merkmale hinzu, welche Sie auch im *workshop*-Ordner in der Textdatei „merkmale.txt“ finden.
  - Number of Tokens
  - Mean Sentence Length in Tokens
  - Mean Token Length in Syllables
  - Lexical Richness: Type Token Ratio (TTR)
  - Lexical Richness: Type Token Ratio (Uber)
  - Lexical Variation Feature: Lexical
  - Syntactic Complexity Feature: Verb Cluster per Clause
  - Syntactic Complexity Feature: Complex Nominals per Sentence
  - Syntactic Complexity Feature: Prenominal Modifier per Complex Noun Phrase
  - Syntactic Complexity Feature: Dependent clause ratio
  - Lexical Sophistication Feature: SUBTLEX Word Frequency (AW Token)
  - Lexical Sophistication Feature: Mean Age of Active Use in KCT (AW Token)
  - Cohesive Complexity Feature: Breindl All Connectives per Token
  - Cohesive Complexity Feature: Breindl Adversative and Concessive Connectives per Token
4. Navigieren Sie zum **Analysis Generator** und erstellen Sie eine Analyse, welche die im Merkmal-Set „Workshop Merkmale“ definierten Merkmale aus dem Korpus „Merlin (korrigiert)“ extrahiert. Achten Sie darauf, dass Sie Deutsch als Analysesprache auswählen. Starten Sie die Analyse und laden Sie die Ergebnisse im *Long Table*-Format herunter.

## 2.2 Visualisierung der Analyseergebnisse in R

In diesem Schritt werfen Sie einen Blick auf die Analyseergebnisse, die Sie mittels CTAP generiert haben. Im Rahmen dieses Tutoriums nutzen wir hierfür R und RStudio.

1. Öffnen Sie das R-Skript „ctap-analysen.R“ in RStudio, das sich in dem Material-Ordner befindet, den Sie heruntergeladen haben. Führen Sie die ersten Zeilen aus, um die Daten in das richtige Format zu bringen.
2. Erstellen Sie den vorbereiteten Plot, der Boxplots zur Textlänge über Kompetenzlevel zeigt und sehen Sie ihn sich genauer an. Was für ein Muster können Sie über Kompetenzstufen hinweg beobachten? Welche Interpretationsmöglichkeiten fallen Ihnen für diese Beobachtung ein?
3. Erstellen Sie Plots für die anderen Merkmale, die Sie mittels CTAP extrahiert haben und betrachten Sie diese genauer. Welche Muster können Sie erkennen? Welche scheinen Ihnen besonders bemerkenswert? Welche Erklärungen scheinen Ihnen für diese plausibel?

## 2.3 Fortgeschrittene Übungen

Sollten Sie Aufgabenpunkte 2.1 und 2.2 abgeschlossen haben, dienen die folgenden Analyse-vorschläge dazu, Ihr Verständnis der automatischen Analyse sprachlicher Komplexität im Bil-dungskontext zu vertiefen.

- Wiederholen Sie die vorigen Analysen. Nutzen Sie diesmal den gesamten Merlin-Datensatz und analysieren Sie die korrigierten sowie die nicht-korrigierten Daten. Vergleichen Sie die Ergebnisse für ausgewählte Merkmale (entweder die oben genannten oder ande-re Merkmale, die Sie interessieren). Welche Unterschiede können Sie beobachten und wie würden Sie diese interpretieren? Schauen Sie sich hierfür gerne auch die eigentli-chen Textdateien im Vergleich an. Zur Hilfe können Sie das R-Script „ctap-erweiterte-analysen.R“ nutzen.
- Splitten Sie die Boxplots nach Aufgaben. Die hierfür benötigte Funktion heißt **erstel-le.boxplot.mit.aufgabensplit** und kann äquivalent zur Funktion **erstelle.boxplot** genutzt werden. Zur Hilfe können Sie das R-Script „ctap-erweiterte-analysen.R“ nutzen. Sehen Sie Unterschiede über Aufgaben hinweg? Wie verhalten sich diese zu den zu beobach-tenden Kompetenzunterschieden? Wie würden Sie diese interpretieren? Welche Konse- quenzen würden Sie hieraus ziehen, wenn Sie Daten zur Analyse sprachlicher Komplexi- tät erheben müssten?
- Wofür steht „AW “bei den Frequenzmaßen? Welche anderen Optionen gibt es für Fre- quenzmaße? Nutzen Sie die Info-Box im Feature Selector, um nähere Informationen zu den einzelnen Merkmalen zu erhalten.
- Laden Sie Ihre eigenen Daten hoch und untersuchen Sie diese!