

Computerlinguistik

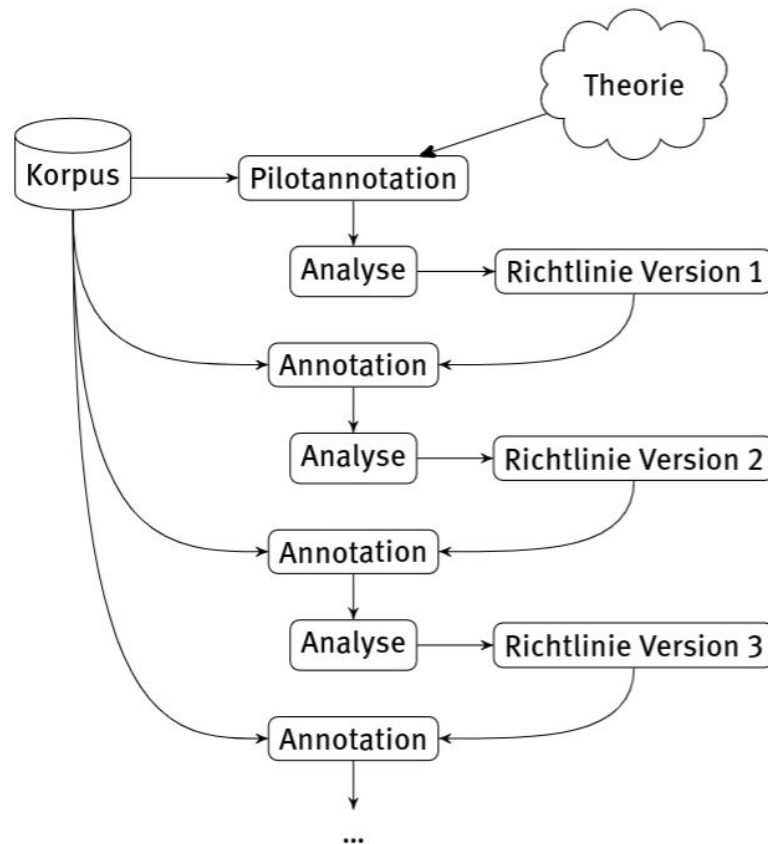
E03: Linguistische Annotation

Linguistische Annotation

- **Auszeichnung** von (Text)Einheiten mit Labels / Verknüpfung von Sprachdaten mit deskriptiven oder analytischen Notationen. Beispiel: INCEPTION
- **Zweck (I):** Überprüfung linguistischer Theorien
- **Zweck (II):** Training von sprachverarbeitenden Anwendungen
- **Annotationsschema:** Definiert, welche Labels gültig sind und mit welchen Einheiten sie assoziiert werden sollen. Labels sollten dabei sprechend gewählt werden, so dass sie wiedergefunden werden können. Beispiel PoS-Tags.
- **Annotationsrichtlinien:** Legen fest, wie bei der Zuweisung von Labels zu Einheiten vorgegangen werden soll.

Annotationsrichtlinien

- Sollten sowohl **generisch**, als auch **präzise** sein.
- Sicherstellung von zuverlässiger, intersubjektiver Annotation.
- Identifikation von Lücken und Mehrdeutigkeiten für Zuweisung und Lokalisation.
- Iterativer Prozess: Schrittweise Verbesserung der Richtlinien



(aus Reiter 2020:194)

Theorie: Zeichentypen

Piktogramme (Bildzeichen): Bildlich erkennbare, festgelegte Inhaltseite, aber keiner Konvention für die Ausdrucksseite. Sprachübergreifend verständlich.

Ideogramme (Begriffszeichen): Wie Piktogramme, aber nicht bildlich erkennbar

Logogramme (Wortzeichen): Festgelegte Inhalts- und Ausdrucksseite

Phonogramme (Lautzeichen): Festgelegte Ausdrucksseite, inhaltsunabhängig

Aufgabe: Annotation der Zeichen

1. Annotieren Sie im Beispieltext (Seite 1) alle Logogramme, Ideogramme und Piktogramme.
2. Setzen Sie sich in Dreiergruppen zusammen und ermitteln Sie, wo Sie unterschiedlich annotiert haben. Dokumentieren Sie diese Fälle.
3. Erstellen Sie eine erste Version für Annotationsrichtlinien, gerne auch mit Beispielen versehen.
4. Annotieren Sie Seite 2 des Beispieltextes in der Gruppe. Dokumentieren Sie, ob die Annotationsrichtlinien geändert werden müssen was der Grund dafür war.

Literatur / Hausaufgabe

- Reichen Sie die Annotationsrichtlinien Ihrer Gruppe bis zum 18.11. in ILIAS ein. Bitte spezifizieren Sie zu Anfang der Datei, wer bei den Richtlinien mitgewirkt hat (Namen und Matrikelnummer), da ich für diejenigen einen INCEPTION-Account anlegen werde.
- **Zur Nachbereitung:**
 - Ide (2017): Kapitel 1 (S. 1-18)
 - Reiter (2020): Anleitung (S. 193-201)
- Die Texte finden Sie im Ilias-Seminarordner.