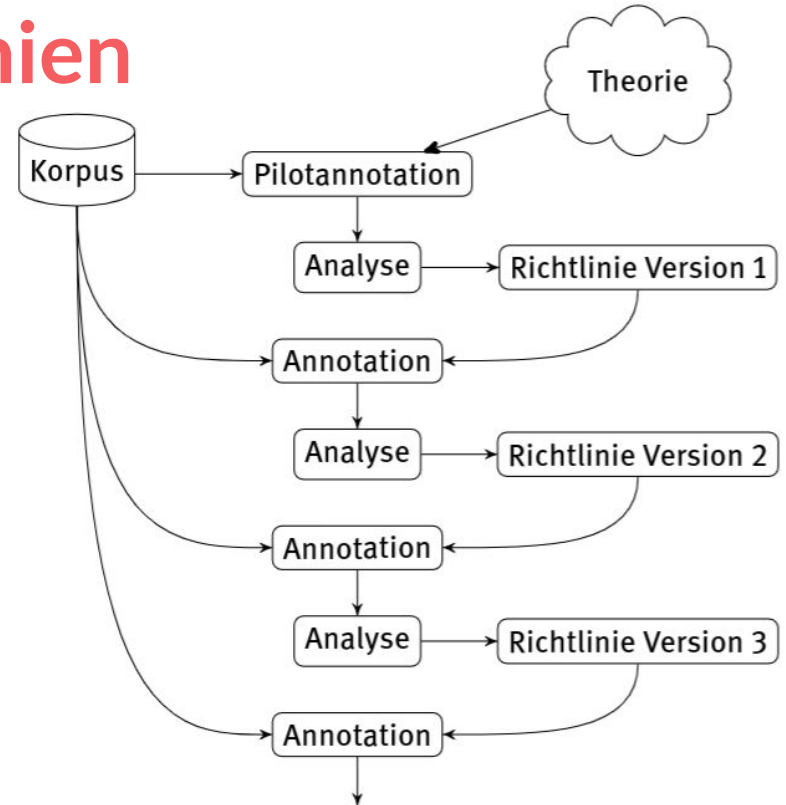


Computerlinguistik

E04: Zeichenebene (2)

(WH) 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)

(WH) Beispiel-Theorie: Zeichentypen

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

Bsp:   

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

Bsp:   

Logogramme (Wortzeichen): Festgelegte Inhalts- und Ausdrucksseite

Bsp: @ € 3

Phonogramme (Lautzeichen): Festgelegte Ausdrucksseite, inhaltsunabhängig

Bsp: a N ə ç

(WH) Graphemklassen (nach Gallmann 1985)

formal definierte Graphemklassen (<i>graphische Mittel</i>)						funktional definierte Graphemklassen
Grapheme		Supragrapheme				
		linear		flächig		
selbständig	unselbständig	konkret	abstrakt	konkret	abstrakt	
<ul style="list-style-type: none"> • Buchstaben • Hilfszeichen • Leerzeichen • Ziffern • Sonderzeichen 	<ul style="list-style-type: none"> • diakritische Zeichen 	<ul style="list-style-type: none"> • Unterstreichen 	<ul style="list-style-type: none"> • Anfangsgroßschreibung • allgemeine Großschreibung • Kapitälchen • Schriftart • Schriftauszeichnung • Schriftgröße • Sperren • Ligaturen • Hoch-/Tiefstellung 	<ul style="list-style-type: none"> • Umrandung • Ton-, Rasterfläche 	<ul style="list-style-type: none"> • Zeile, Textblock • Ausschluss • Einzug • Zeilenanfang, -ende 	<ul style="list-style-type: none"> • (Grund-)Grapheme • Ideogramme • Klassifikatoren • Grenzsignale • Satzintentionssignale • Auslassungssignale

Aufstellung aus der [Wikipedia](#) nach Gallmann (1985): Graphische Elemente der geschriebenen Sprache. Grundlagen für eine Reform der Orthographie.

(WH) Aufgabe: Annotation der Zeichen

1. Annotieren Sie im Beispieltext (zunächst ausschließlich auf 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 (bspw. in einer Tabelle).
3. Erstellen Sie eine erste Version für Annotationsrichtlinien (vgl. Reiter 2020 Abb. 2), 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. Reichen Sie Ihre Richtlinien bei ILIAS ein (siehe nächste Seite).

Annotationsrichtlinien

- Was soll annotiert werden?
- Worauf basieren die Richtlinien?
- Wer erstellt die Annotationen, an wen richten sie sich?
- Was sind die zu annotierenden Einheiten?
- Sind alle Einheiten zu annotieren oder nur bestimmte?
- Welche Kategorien werden den Einheiten zugewiesen?
- Wie bestimmt man die Kategorie?
- Welche Kategorien können leicht verwechselt werden?
- Was sind schwierige Fälle, wie werden diese entschieden?
- Wie wird das Annotationstool verwendet?

INCEpTION

- INCEpTION ist ein Tool zur manuellen und automatisierten Annotation und kann darum für eine Vielzahl von Fragestellungen, in denen Texte eine Rolle spielen, eingesetzt werden.
- Entwicklung an der TU Darmstadt, Webpräsenz inklusive Dokumentation, Downloads usw. unter <https://inception-project.github.io/>
- Das IDH hat eine eigene Instanz von INCEpTION, an der Sie sich auch anmelden sollen – <https://www.spinfo.uni-koeln.de/inception> (Anmeldedaten per Mail und im Unterricht).
- Sie sollten nach der Anmeldung bereits ein Projekt zur Auswahl haben (Computerlinguistische Grundlagen 2425). In diesem werden Sie die Hausaufgaben und das Abschlusstest bearbeiten können.

Literatur / Hausaufgabe

- Sie müssten jetzt einen Inception-Zugang haben und ein Annotationsprojekt haben. In diesem Annotationsprojekt müssten Sie den Text "Zeichentypen.txt" finden. Bitte zeichnen Sie dort *Piktogramme*, *Ideogramme* und *Logogramme* aus.
- Die Aufgabe ist nicht allzu schwer, weil im Text ja steht, um welchen Zeichentyp es sich handelt. Die Übung dient vor allem dazu, sich schon einmal mit Inception vertraut zu machen.
- Bitte klicken Sie nach der Bearbeitung auf das **Schloss-Symbol**, damit wir ihre Lösung in der nächsten Stunde kuratieren können!
- Falls Sie Probleme mit Login oder Handling haben, melden Sie sich bitte.