

Computerlinguistik

E05: Wortebene (2)

Inter-Rater-Agreement

- Maß für die Übereinstimmung unterschiedlicher Annotierender
- **Cohens Kappa:** Maß für die nicht-zufällige Übereinstimmung zweier Rater. p_o ist der gemessene, p_e der erwartete Wert.
$$\kappa = \frac{p_o - p_e}{1 - p_e}$$
- **Fleiss' Kappa:** Wie Cohens Kappa, nur für mehr als 2 Rater. p_o und p_e werden dabei über die Summen der Übereinstimmungen berechnet.
- Beide Maße nur sinnvoll auf Nominalskalenniveau anwendbar.
- **Krippendorff's Alpha:** Auch auf höherskalierte Daten anwendbar.

$$\alpha = 1 - \frac{\text{tatsächliche Varianz der Differenzen}}{\text{erwartete Varianz der Differenzen}}$$

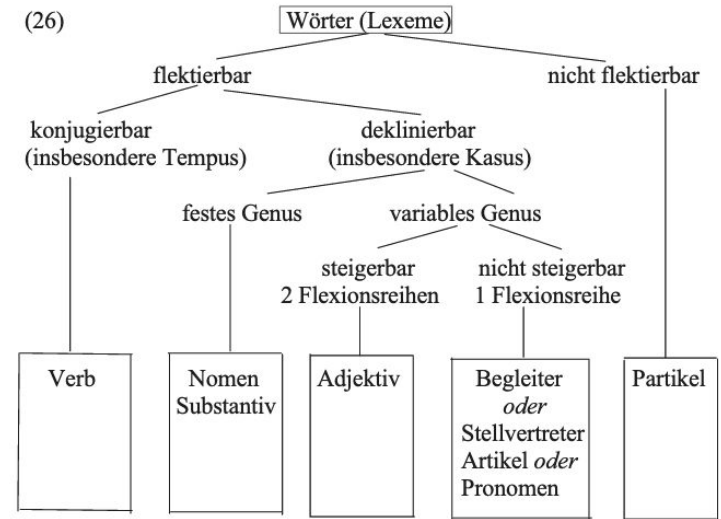
Morphologische Analyse

- **Segmentierung:** Zerlegung des Inputs in Morpheme
- **Klassifizierung:** Zuordnung der identifizierten Morpheme zu Klassen
- **Strukturierung:** Konstruktion des hierarchischen Aufbaus der Morpheme bzw. Morphemkomplexe

- **Verwendung:**
 - Lemmatisierung (Auch “Stemming”, Reduktion auf die Grundform)
 - Ermittlung morphosyntaktischer Merkmale zur Auszeichnung von Satzbestandteilen (z.B. mit Universal Features)
 - Webservice: CST online tools

Part-of-Speech-Tagging

- Zuordnung von Bestandteilen sprachlicher Äußerungen (meist Wörtern) zu Klassen (meist Wortklassen).
- Wortartenklassifikation:
 - Morphosyntaktische Eigenschaften (Merkmale)
 - Syntaktische Eigenschaften (Distribution)
- Tagsets:
 - Klassisch: 10 Wortarten
 - Computerlinguistisch: Universal POS-Tags -
 - Distributionell: Bis zu 150 Klassen
 - Übersicht zum state of the art
 - Webservice Weblicht



aus Ramers (2007:24)

Literatur / Hausaufgabe

- **Zur Nachbereitung:**
 - Bearbeiten Sie die Aufgabe in Inception (Anweisung in ILIAS)
- **Zur Vorbereitung:**
 - Lesen Sie: Ramers (2000): Kapitel 1 (S. 11-34)
- Die Texte finden Sie im ILIAS-Seminarordner.