

Computerlinguistik

E05: Wortebene

Morphologie

- **Begriffe: Wort - Morphem - Allomorph**
- **Flexion und Wortbildung**
- **Sprachtypologie anhand der Wortstruktur**
- **Morphologische Analyse**
- **Lemmatisierer**
- **POS-Tagger**

Begriffe: Wort - Morphem - Allomorph

- lexikalisches Wort (Lexem) vs. flexivisches Wort (Wortform)
- phonologisches Wort vs. syntaktisches Wort
- **Morphem**, funktionale Definition: Kleinste bedeutungstragende Einheit einer Sprache
 - freie vs. gebundene Morpheme
 - Wurzelmorpheme vs. Affixe
 - Derivationsaffixe vs. Flexive
- **Allomorphe**: Bedeutungs- und funktionsgleiche Varianten eines Morphems

Flexion und Wortbildung

- **Flexion:** Bildung flexivischer Wörter
 - Konjugation (Verben) - Deklination (Nomen und Adjektive) - Komparation (Adjektive)
- **Wortbildung:** Bildung lexikalischer Wörter
 - Wichtigste Formen:
 - Komposition (Wurzel + Wurzel)
 - Derivation (Wurzel + Affix)
 - Weitere Formen: Konversion - Abkürzung - Amalgamierung (- Entlehnung)

Flexionsparadigmen (Verben)

	WARTEN:		HALTEN:	
Sing, 1. Pers	Ich	wart-e wart-et-e	Ich	halt-e hielt-ø
Sing, 2. Pers	Du	wart-est wart-et-est	Du	hält-st hielt-est
Sing, 3. Pers	Es	wart-et wart-et-e	Es	hält-ø hielt-ø
Plur, 1. Pers	Wir	wart-en wart-et-en	Wir	halt-en hielt-en
Plur, 2. Pers	Ihr	wart-et wart-et-et	Ihr	halt-et hielt-et
Plur, 3. Pers	Sie	wart-en wart-et-en	Sie	halt-en hielt-en
Partizip	warten-d	ge-wart-et	halten-d	ge-halt-en
	Präsens	Präteritum	Präsens	Präteritum

Flexionsparadigmen (Nomen)

	KIND:		WOLF:	
Nominativ	das Kind	die Kind-er	der Wolf	die Wölf-e
Genitiv	des Kind-es	der Kind-er	des Wolf-es	der Wölf-e
Dativ	dem Kind	den Kind-er-n	dem Wolf	den Wölf-e-n
Akkusativ	das Kind	die Kind-er	den Wolf	die Wölf-e
	Singular	Plural	Singular	Plural

Sprachtypologie anhand der Wortstruktur

- **Isolierende Sprachen:** 1 Wort = 1 Bedeutung (keine Flexion)
- **Flektierende Sprachen:** Stämme + Flexive mit mehreren Bedeutungen
- **Agglutinierende Sprachen:** 1 Morphem = 1 Bedeutung (feste Flexive)
- **Polysynthetische Sprachen:** Verschmelzung von grammatischen Einheiten zu einer Worteinheit (inkorporierend).

Literatur / Hausaufgabe

➤ **Zur Nachbereitung:**

- Lesen Sie: Vater (2002): Kapitel 3 (S. 61-74)
- Bearbeiten Sie die Aufgabe in Inception (Anweisung in ILIAS)

➤ **Zur Vorbereitung:**

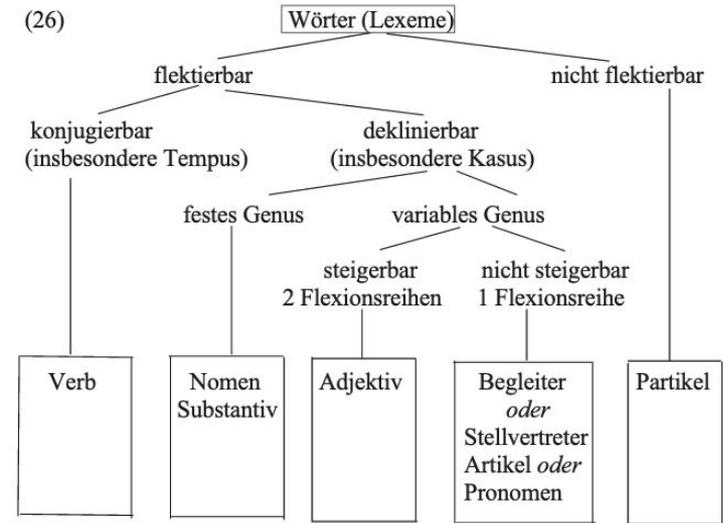
- Lesen Sie: Ramers (2000): Kapitel 1 (S. 11-34)
- Die Texte finden Sie im Ilias-Seminarordner.

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.

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)

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

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.