

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

# Flexionsparadigmen (Nomen)

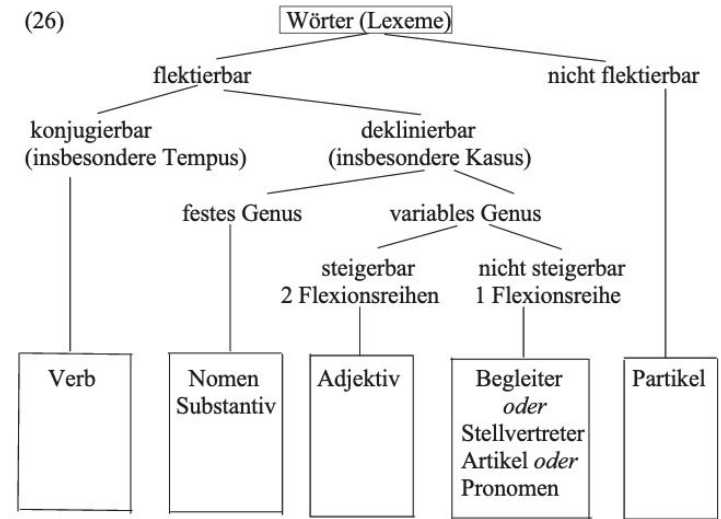
	KIND:		WOLF:	
<b>Nominativ</b>	das Kind	die Kind-er	der Wolf	die Wölf-e
<b>Genitiv</b>	des Kind-es	der Kind-er	des Wolf-es	der Wölf-e
<b>Dativ</b>	dem Kind	den Kind-er-n	dem Wolf	den Wölf-e-n
<b>Akkusativ</b>	das Kind	die Kind-er	den Wolf	die Wölf-e
	<b>Singular</b>	<b>Plural</b>	<b>Singular</b>	<b>Plural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

# 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).

# 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:**

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

## ➤ **Zur Vorbereitung:**

- Recherchieren Sie, was sich hinter dem Begriff “Interrater-Reliabilität” verbirgt.
- Die Texte finden Sie im Ilias-Seminarordner.