

# Einführung in die Computerlinguistik

## Wortarten

Alex Fraser / Robert Zangenfeind

Center for Information and Language Processing

2019-11-04

Die Grundfassung dieses Foliensatzes wurde von Dr. Benjamin Roth unter Zuhilfenahme von Materialien aus Vorlesungen von Prof. Dr. Tania Avgustinova erstellt. Fehler und Mängel sind ausschließlich meine Verantwortung.

- 1 Intro
- 2 Taxonomie der Wortarten
- 3 Details zu den Wortarten
- 4 Wortartenzuweisung
- 5 Anwendung

# Zum Begriff “Wortart”

- Wortarten bzw.
- Lexikalische Kategorien
- Syntaktische (Wort-)Kategorien
- Part-of-Speech (POS)

- Viele syntaktische Eigenschaften sind identisch für (große) Klassen von Wörtern.
- Regeln gelten nur für bestimmte Kategorien von Lexemen.
- Kategorisierung der Lexeme nötig, damit Regeln richtig anwendbar sind.
- Ambiguität: Wortart einer Form muss im Kontext bestimmt werden, damit sie richtig verarbeitet werden kann, vgl.: *Time flies like an arrow.*

- POS-tagging (Part-of-speech-tagging):  
Automatische Wortartbestimmung
- Lemmatisierung: Grundform eines Wortes kann gefunden werden, wenn Wortart bekannt, vgl.:  
*runde* (Verb) → “runden”; *runde* (Adjektiv) → “rund”
- Maschinelle Übersetzung: Richtige Übersetzung hängt von Wortart des fraglichen Wortes ab.
- Zusammen mit Tokenisierung einer der am häufigsten verwendeten Vorverarbeitungsschritte in der Computerlinguistik.

- Lexeme bilden z.T. offene Listen  $\Rightarrow$  nicht aufzählbar (vs. grammatische Morpheme – bilden geschlossene Listen)
- Linguistische Kriterien sind nötig zur Klassifizierung
- Eine gängige Art der Klassifizierung richtet sich nach **morphologisch-syntaktischen Kriterien**.

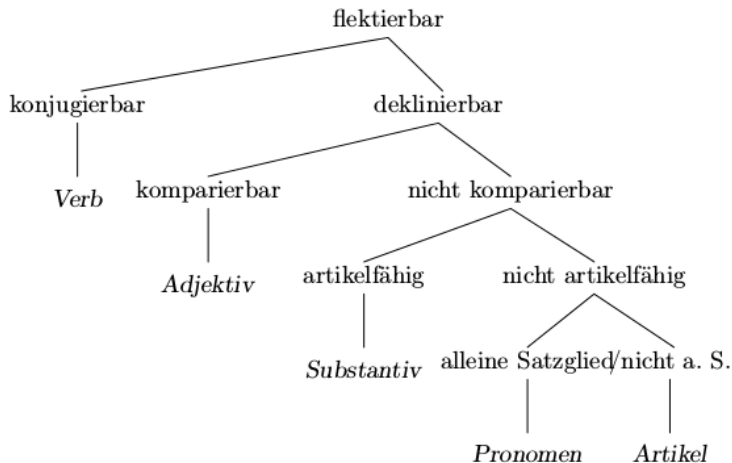
- Morphologisch:
  - **flektierbar**: Substantiv, Adjektiv, Pronomen, Numerale, Verb, Artikel
  - **nicht flektierbar**: Präposition, Konjunktion, Partikel

Problem:

  - Manche Adverbien sind flektierbar: Komparation
  - *öfter schlafen*, aber auch: *noch und nöcher*
- Syntaktisch:
  - die Fähigkeit **als Satzglied zu fungieren**
  - die Fähigkeit **einen Artikel zu binden**
  - die Fähigkeit **einen bestimmten Kasus zu fordern**
- Die Hauptunterscheidung wird zwischen **flektierbaren** und **nicht-flektierbaren** Lexemen getroffen, die Wortarten werden davon ausgehend weiter eingeteilt.

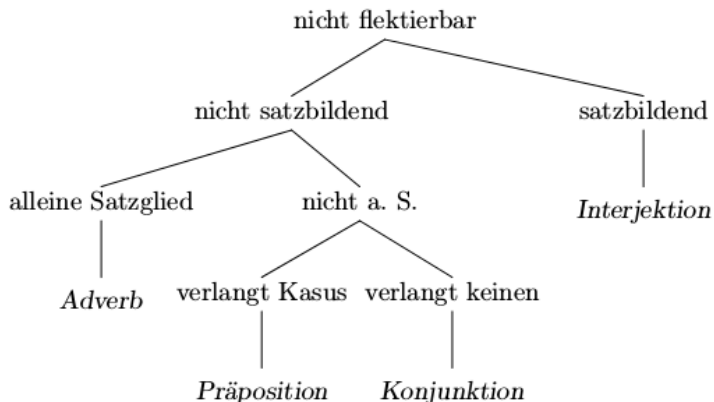


# Wortarten (flektierbare Lexeme)



Klassifizierung von Heringer, H.-J.: Morphologie. Paderborn 2009.

# Wortarten (nicht-flektierbare Lexeme)



Klassifizierung von Heringer, H.-J.: Morphologie. Paderborn 2009.

- Autosemantika: Substantiv, Adjektiv, Adverb, (Voll-)Verb
- Synsemantika:  
Hilfsverb (*sein, haben, werden*), Partikel (*zu*)

Problem:

- Pronomen, Präposition, Artikel und Partikel lassen sich schwer in dieses Schema einordnen

- Offene Klassen sind Bestandteile des Lexikons und können durch Wortbildungsregeln jederzeit erweitert werden:  
Verb, Nomen, Adjektiv, Adverb
- Geschlossene Klassen sind im Prinzip aufzählbar:  
Präposition, Artikel, Konjunktion

Kann das Lexem den Kasus eines Satzgliedes bestimmen?

- z.B. Verben mit Nominativ für Subjekt sowie Akkusativ, Dativ bzw. Genitiv für Objekte
- mehr zu Satzgliedern bei Syntax

	NOM	AKK	DAT	GEN
<i>Verb</i>	X	X	X	X
<i>Pröp.</i>		X	X	X
<i>Adj.</i>		(X)	X	X
<i>Nomen</i>				X

# Wortarten: Übersicht und Beispiele

*tree, dog, freedom; Baum, Ofen*  
*run, kick, work; sprechen, müssen*  
*big, red, beautiful; braun, ehrlich*  
*a(n), some, any, the, this; der, die*  
*du, sie*  
*sieben, anderthalb*  
*today, there, well, strangely; heute, sehr*  
*in, on, below, against; für, auf*  
*that, because, although; wenn, weil*  
*ouch, oops; oh, psst; !*

**Substantiv / Nomen** (Hauptwort)  
**Verb** (Zeitwort)  
**Adjektiv** (Eigenschaftswort)  
**Artikel / Determinator** (Geschlechtswort)  
**Pronomen** (Fürwort)  
**Numerale** (Zahlwort)  
**Adverb** (Umstandswort)  
**Präposition** (Verhältniswort)  
**Konjunktion** (Bindewort)  
**Interjektion** (Empfindungs-/Ausrufewort)

- Konjugierbar: morphologische Kennzeichnung nach Person, Numerus, Tempus, ...
- Kongruenz in Person, Numerus und/oder Genus mit einem oder mehreren Argumenten (z.B. dt. mit Satzsubjekt)
- Einteilung nach Stelligkeit: Valenzklassen
  - Verben ohne Ergänzung:  
*Es **schneit**.*
  - Intransitive Verben (nur Subjekt):  
*Martin **schnarcht**.*
  - Transitive Verben (Subjekt und Akkusativobjekt):  
*Die Professorin **lobt** ihre Studenten.*
  - Ditransitive Verben (Subjekt, Akkusativ- und Dativobjekt):  
*Hans **verkauft** sein Auto einem Freund.*
  - Verben mit Genitiv- oder Dativobjekt (ohne Akkusativobjekt):  
*Wir **gedenken** der Toten. / Die Spieler **danken** dem Trainer.*
  - Verben mit Präpositionalobjekt:  
*Sie **zieht nach** Hamburg.*

# Drei besondere Verbklassen

- **Kopulaverben** (*sein, werden, bleiben*) spezifizieren lediglich das Tempus, während der semantische Gehalt vom Nomen oder Adjektiv beigetragen wird:  
*Die Vorwürfe **sind** schwerwiegend.*
- Bei sog. **Stützverbkonstruktionen** (engl. Lightverb Constructions) ergibt sich die Hauptbedeutung durch ein Satzglied, mit dem das Verb eine lexikalisierte Verbindung eingegangen ist:  
*Ich **ziehe** alle Optionen **in Erwägung**.*  
*Er **erhebt** schwere **Vorwürfe**.*
- **Modalverben** (*können, müssen, sollen, ...*) spezifizieren die Möglichkeit oder Notwendigkeit etc. einer Aussage:  
*Ich **kann** morgen nicht zum Training kommen.*



- Deklinierbar: morphologische Kennzeichnung von Kasus, Genus und Numerus
- Nomen: festes Genus, offene Klasse
- Pronomen:
  - geschlossene Klasse
  - verweisen auf etwas, haben als Zeichen alleine keine Referenz.
- Unterklassen:
  - 1 Personalpronomina: *ich, du er, sie, es, mich, dir*
  - 2 Reflexivpronomina: *sich*
  - 3 Possessivpronomina: *mein, dein, sein*
  - 4 Demonstrativpronomina: *diesen*
  - 5 Relativpronomina: *der, welcher*
  - 6 Interrogativpronomina: *welcher, wer, was*
  - 7 Indefinitpronomina: *jemand, etwas, alle, kein*

- **attributive Verwendung:** *das große Haus.*
- **prädikative Verwendung:** *Das Haus ist groß.*
- rein **attributive** Adjektive: *der ehemalige Präsident*  
vs. *\*der Präsident ist ehemalig*
- rein **prädikative** Adjektive: *die Regierung ist schuld*  
vs. *\*die schulde Regierung*
- **deklinierbar** (nur wenn attributive Verwendung möglich!)
- meist **komparierbar**
- bestimmte Adjektive verlangen **Ergänzungen:**  
*seinem Bruder ähnlich sein*  
*sich seiner Schuld bewusst sein*  
*in Köln wohnhaft sein*  
*seiner Überzeugung sicher sein*  
*der Idee dienlich sein*

- Modifizieren Verb, Adjektiv oder Satz, vgl.:  
*Sie ist schon da.*  
*Ich werde bald gehen.*  
*Das hat mir sehr geholfen.*  
*Leider konnte ich nicht teilnehmen.*
- nicht flektierbar
- manche steigerbar
- Konvention: adverbial gebrauchte Adjektive bleiben in ihrer Wortart-Kategorie Adjektiv, vgl.:  
*Er fährt schnell.*  
⇒ Wortart vs. syntaktische Funktion

# Wortart Artikel (Determinierer)

- Geschlossene Liste
- syntaktische Funktion: komplettieren eine Nominalphrase
- Definite Artikel: *der Hut, die Katze, das Haus*  
⇒ verweisen auf Entitäten, die bereits bekannt sind, schon in den Diskurs eingeführt wurden, oder deren Existenz aus anderen Informationen folgt.
- Indefinite Artikel: *ein Hut, eine Katze, ein Haus*  
⇒ führen z.B. neue Referenten in den Diskurs ein, auf die später referenziert werden kann.
- z.B.: *Hans kaufte ein Haus. Der Kredit war günstig.*
- weitere Artikel: Demonstrativart. (z.B. *diese, jene, dieselben, solche*); Quantifikatoren (z.B. *alle, jeder, viele, beide*); Negatoren (z.B. *kein, keine*); Possessivart. (z.B. *mein, ihr*); Interrogativart. (z.B. *welche*)
- Artikel (wie auch Adjektive) sind typischerweise kongruent zu einem Nomen in Numerus, Genus und Kasus.

# Wortart Präposition (Adposition)

- weisen Nomen Kasus zu
- **Präpositionen** stehen links (z.B. *in, auf, für*):  
*nach München, wegen der Kinder*
- Seltener Adpositionen:
  - manche rechts (z.B. *zufolge*):  
*seiner Frau zuliebe, den Freunden entgegen*
  - wenige: links und rechts möglich (z.B. *wegen*)
  - manche umschließen Nomen (z.B. *um ... willen*):  
*um der Liebe willen, von Gesetzes wegen*

- **Konjunktionen** verbinden syntaktische Einheiten der gleichen syntaktischen Kategorie (Sätze, Phrasen, Wörter, Wortteile) (z.B. *und, oder, aber, entweder ... oder*)
- geschlossene Liste
- **Koordinierende** (nebenordnende) Konjunktionen: (z.B. *und, oder, aber*):  
*Er schläft, **aber** sie arbeitet noch.*
- **Subjunktionen** (satzeinbettende Konjunktionen) (z.B. *dass, weil, obwohl*):  
***Weil** er berühmt ist, lassen sie ihn durch.*

# Wortart Interjektion (Satzwort)

- Syntaktisch unverbundene, satzwertige Äußerungen
- Drücken Empfindung, Bewertung oder Willen des Sprechers aus (z.B. *aha*, *igitt*, *richtig*, *ja*, *nein*, *Danke*)
- Übermitteln Aufforderung oder Signal zur Kontaktaufnahme (z.B. *Hallo*, *Prost*, *Hey*)

- Übernehmen lediglich syntaktische oder pragmatische Hilfsfunktionen.
- Bilden keine eigene Phrase.
- Beispiele:  
*Das kann man **aber** so nicht sagen.*  
*Das ist **halt** so.*  
***am** schönsten*  
***zu** schnell*
- Lassen sich oft schwer in ein Schema einordnen.



# Das Partizip: zwischen Verb und Adjektiv

- Partizipien verhalten sich einerseits wie Verben, denn
  - sie können Kasus zuweisen
  - sie “erben” die Argumentstruktur des Verbs, aus dem sie abgeleitet werden.

*die Tätigkeit befriedigt mich ⇒ eine mich befriedigende Tätigkeit;*

*der Schüler liest das Buch ⇒ der das Buch lesende Schüler*

- Partizipien verhalten sich andererseits wie Adjektive, denn sie flektieren wie Adjektive (können aber oft nicht prädikativ verwendet werden)

*die befriedigenden und nützlichen Tätigkeiten;*

*Freude an befriedigender und nützlicher Tätigkeit;*

*Eine singende Frau;*

*\*Die Frau ist singend.*

- Konvention: Dem Partizip wird die Wortart Verb zugewiesen (Wortart vs. syntaktische Funktion).

## Wortartwechsel

- *Leid* (vgl. z.B.: *Das tut mir leid*) (Nomen vs. Verbpartikel)
- *Klasse* (vgl. z.B.: *ein klasse Buch*) (Nomen vs. Adjektiv)
- *ja* (vgl. z.B.: *Das war ein klares Ja*) (Satzwort vs. Nomen)

## Zugehörigkeit zu mehreren Wortarten (Wortartenambiguität)

- *Er las, aber er war sehr unkonzentriert* (Konj.)
- *Das kann man aber so nicht sagen* (Partikel)

## Kontraktionen

- *im*
- *ins*
- *zum*
- ...

## Zahlwörter

- *eins/ein/eine* ...: deklinierbar (ähnlich zu Determinerern oder Pronomen?)
- *zwei*: auch deklinierbar: z.B. *der Bund zweier Kaiser*
- *tausend* (ebenso)
- *Million*: eher wie Nomen

## Sonderfall *viel*:

- teils wie Determinierer:  
*Vieles Erfreuliche stand in dem Brief*  
*Er trank viel Bier*
- teils wie Adjektiv:  
*viele Tiere*  
*die vielen Tiere*  
*das viele Laub*

# Part-of-Speech Tagging (POS Tagging) (1)

- Wörter eines Textes mit dazugehörigen Wortarten (engl. Part-of-Speech) kennzeichnen.
- eine Art der Annotierung des Textes/Korpus
- Wortart gibt viele Informationen über das Wort und seine benachbarten Wörter im Text
  - z.B. Possessivpronomen (z.B. *mein, dein, sein, unser*) ⇒ rechts davon: häufig Nomen
  - Personalpronomen (z.B. *ich, du, er, wir*) ⇒ rechts davon: (meist) Verb

# Part-of-Speech Tagging (2)

- Tagging manuell oder durch Algorithmen (regelbasierte oder statistische Methoden (z.B. Hidden-Markov-Modelle))
- Programme im Netz:
  - CIS, LMU München: MarMoT  
`cistern.cis.lmu.de/marmot/`
  - CIS, LMU München: TreeTagger  
`www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/`
  - Stanford:  
`http://nlp.stanford.edu/software/tagger.shtml`

- Begriff “Wortart”
- Part-of-speech tagging
- Warum brauchen wir Wortarten / POS tagging?
- flektierbar vs. nicht-flektierbar
- offen vs. geschlossene Klassen
- morphologische/syntaktische Klassifizierungskriterien
- Grenzen dieser Kriterien:  
Partizipien, Zahlwörter, “squishy” cases (“klasse”)
- Wortartenambiguität
- alle Wortarten: Verb, Nomen, Adjektive, ...