

# Einführung in die Computerlinguistik

## Wortarten

Hinrich Schütze

Center for Information and Language Processing

2018-10-29

Die Grundfassung dieses Foliensatzes wurde von Dr. Benjamin Roth erstellt. Fehler und Mängel sind ausschließlich meine Verantwortung.

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

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

# Zum Begriff “Wortart”

- Wortarten
- 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. *Time flies like an arrow.*

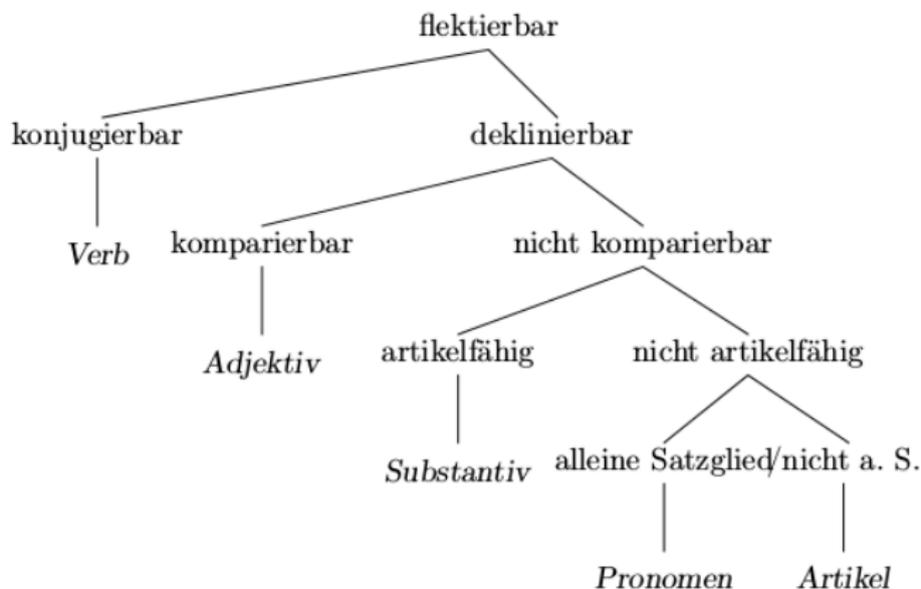
- POS-tagging (Part-of-speech-tagging):  
Automatische Wortartbestimmung
- Lemmatisierung: Grundform eines Wortes kann gefunden werden, wenn Wortart bekannt.  
“runde” (verb) → “runden”; “runde” (adjective) → “rund”
- Maschinelle Übersetzung: Richtige Übersetzung hängt von verwendeter Wortart ab.
- Zusammen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

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

- 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**.

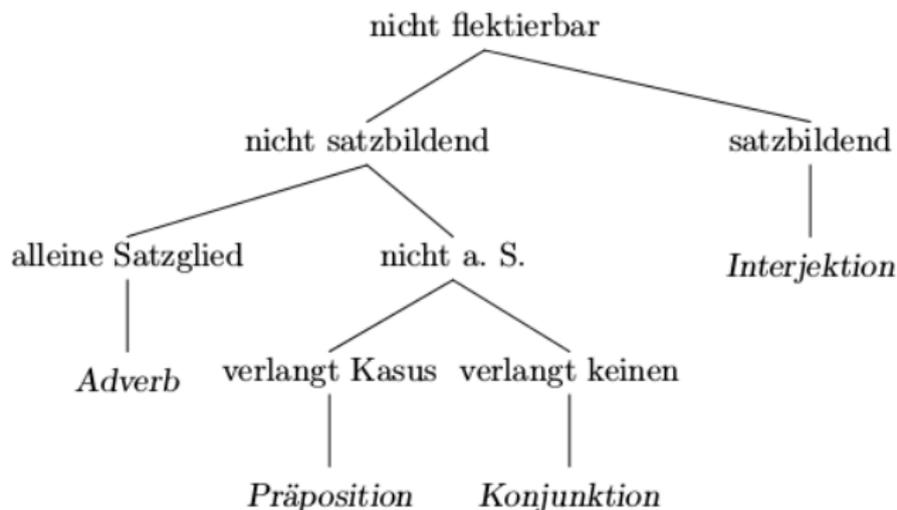
- Morphologisch:
  - **flektierbar**: Substantiv, Adjektiv, Pronomen, Numerale, Verb, Artikel
  - **nicht flektierbar**: Präposition, Konjunktion, Partikel
  - Viele Adverbien sind flektierbar: Komparativ.
  - “schneller laufen”, aber: “noch und nöcher”, “er hat mir sehrer geholfen”
- Syntaktisch:
  - 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), Hilfspartikel (zu)
- Pronomen, Präposition, Artikel und Partikel lassen sich schlecht 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 und somit in die Grammatik integrierbar:  
Präposition, Artikel, Konjunktion

Kann das Lexem den Kasus eines Satzgliedes bestimmen?

- z.B. Subjekt, Akkusativ-, Dativ-, Genitiv-Objekte bei Verben
- Später mehr zu Satzgliedern.

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

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

# 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, Tempus, Numerus, ...
- Kongruenz in Person, Numerus und/oder Genus mit einem oder mehreren Argumenten (z.B. mit Subjekt)
- Einteilung nach Stelligkeit: Valenzklassen
  - Verben ohne Ergänzung  
Es **schneit**.
  - Intransitive Verben (nur Subjekt)  
Martin **schnarcht**.
  - Transitive Verben (Subjekt und Akkusativobjekt)  
Der Professor **lobt** seine 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: “nach China ziehen”

- **Kopulaverben** (sein, werden) spezifizieren lediglich das Tempus, während der semantische Gehalt vom Nomen oder Adjektiv beigetragen wird.  
*Die Vorwürfe **sind** schwerwiegend.*
- Bei sog. **Lightverb Constructions** (keine eigene Wortart; dt. *Stützverben*) kommt 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 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 ...  
sich seiner Schuld bewusst ...  
bei uns beliebt ...  
in Köln wohnhaft ...  
seiner Überzeugung sicher ...  
der Idee dienlich ... sein

## Übung

Bestimme die Wortarten der Wörter, die sie im folgenden Satz haben:

“Das Ützlipütz prümft den pliemen Wenzipil krät.”

- Modifizieren Verb, Adjektiv oder Satz.  
Sie ist **schon** da.  
Ich werde **bald** gehen.  
Das hat mir **sehr** geholfen.  
Ein **äußerst** hilfreiches Buch.  
**Leider** konnte ich nicht teilnehmen.
- Nicht flektierbar.
- Manche steigerbar.
- Konvention: Adverbial gebrauchte Adjektive bleiben in ihrer Wortart-Kategorie Adjektiv.  
*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:  
*Er schläft, **aber** sie arbeitet noch.*
- **Subjunktionen** (satzeinbettende Konjunktionen):  
dass, weil, obwohl, . . .  
***Weil** er berühmt ist, lassen sie ihn durch.*

- Syntaktisch unverbundene, satzwertige Äußerungen.
- Drücken Empfindung, Bewertung oder Willen des Sprechers aus (z.B. aha, igitt, richtig, ja, nein, Danke)
- Übermitteln Aufforderung oder Signal der 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 wie Verben, denn
  - sie können Akkusativ 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 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  
\*Die Frau ist laufend.
- Konvention: Dem Partizip wird die Wortart *Verb* zugewiesen (Wortart vs. syntaktische Funktion).

## Übung

Bestimme die Wortarten der folgenden Wörter: Million, Mist, Stuhl, aber, abstract, abstrakt, automobile, between, damit, each, euch, eure, heuer, increase, interessant, interessanterweise, jeglichem, jenen, jenseits, million, somit, tun, wow

# 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)

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

## Wortartwechsel

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

## Zugehörigkeit zu mehreren Wortarten / Ambiguität

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

## 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*

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

# Part-of-Speech Tagging (POS Tagging)

- 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: Verb

- 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/](http://cistern.cis.lmu.de/marmot/)
  - CIS, LMU München: TreeTagger  
[www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/](http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/)
  - Stanford:  
<http://nlp.stanford.edu/software/tagger.shtml>

# Stemming (Stemmatisierung)

- Alternatives Verfahren zum Lemmatisieren
- Flexionsmorpheme von Wortform werden beseitigt  
⇒ Wort-” Stamm” (wird der Wortform zugeordnet)
- z.B. engl. Wortformen *process*, *processing*, *processed*  
⇒ ”Stamm” *process*
- Problem: sinnvolle Unterscheidungen können verloren gehen:
- z.B. *stocks* (Aktien etc.) und *stockings* (’Strümpfe’ etc.)  
⇒ ”Stamm” *stock*
- Programm in Netz z.B. Porter Stemmer:  
Porter-Stemmer (Demo und download):  
<http://snowball.tartarus.org>

- 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, . . .