

Morphologie, Lemmatisierung und Wortartenklassifikationen

Morphologieanalyse und Lexikonbau (1. Vorlesung)

Übersicht

- [Was ist Morphologie?](#)
 - [Warum braucht man Morphologieanalyse in der CL?](#)
 - [Was ist ein Wort?](#)
 - [Was ist ein Lemma?](#)
 - [Wortarten-Klassifikation](#)
 - [Historische Entwicklung der Wortarteneinteilung](#)
 - [Wortarten nach `Thomas von Erfurt'](#)
 - ['Klassische' Einteilung fürs Englisch](#)
 - [Wortarten in der Duden-Grammatik](#)
 - [Wortarteneinteilung in Morphix](#)
 - [Wortarteneinteilung im STTS](#)
 - [Zusammenfassung](#)
-

Was ist Morphologie?

Morphologie (in der Sprachwissenschaft) ist die Lehre von

- Flexion (Deklination, Konjugation, Komparation) und
- Wortbildung (Derivation, Komposition)

Ziele der morphologischen Analyse (nach [\[Bußmann 83\]](#))

- Gewinnung von Kriterien zur Bestimmung der Wortarten
- Beschreibung der Regularitäten der Flexion
- die Untersuchung von grammatischen Kategorien wie Tempus, Modus und ihre sprachlichen Entsprechungen
- bei der Wortbildung die Untersuchung der Basiselemente, Kombinationsprinzipien und semantischen Funktionen von Wortneubildung

Warum braucht man Morphologieanalyse in der CL?

[[Sproat 92](#)]: In der CL Praxis sind Anwendungen ohne Morphologieanalyse möglich.
(Ein syntaktischer Parser z.B. kann jede Wortform in einem Vollformenlexikon nachschlagen)

Aber

- Vollformenlexika sind unübersichtlich und schwerer verwaltbar.
- Sämtliche morphologischen Regularitäten werden missachtet.
- Das Erkennen neuer Wortformen wird wesentlich schwerer.
- Während englische Nomen tatsächlich nur 2 Wortformen haben, sind es bei englischen Verben schon 4, bei deutschen Verben mindestens ein Dutzend.
- In stark flektierenden Sprachen wie Türkisch oder Finnisch gibt es oft hunderte bis tausende von Wortformen für eine Grundform.
- Wortbildungsphänomene sind so kaum handzuhaben, z. B. deutsche Nominalkomposita kennen fast grenzenlose Kombinationsmöglichkeiten und sprengen jeden Lexikonrahmen.

Was ist ein Wort?

Beispielsatz

Er sah sich die 5 Ein- und Ausgaben des MS-Word-Programms an.

Wieviele Wörter kommen in diesem Satz vor? (SPIEGEL, 13.3.95):

Jahrelang wurde in den IDG-internen Info-Briefen der Konkurrent unter dem Kürzel ZD argwöhnisch verfolgt, der von William Ziff sen. schon 1927 mit Special-Interest-Titeln für Flieger, Ski- und Autofahrer gegründet worden war.

Wort: "Intuitiv vorgegebener und umgangssprachlich verwendeter Begriff für sprachliche Grundeinheiten." (nach Bußmann)

- auf phonetisch-phonologischer Ebene: Wort ist kleinste durch Wortakzent und Grenzsignale wie Pause, Knacklaut u.a. theoretisch isolierbares Lautsegment.
- auf orthographisch-graphemischer Ebene: Wort ist durch Leerstellen im Schriftbild isoliert
- auf morphologischer Ebene: Wort ist die Grundeinheit eines grammatischen Paradigmas wie Flexion (i. Ggsatz zu Wortform)
- auf lexikalisch-semantischer Ebene: Wort ist kleinster selbständiger Träger von Bedeutung
- auf syntaktischer Ebene: kleinste verschiebbare und ersetzbare Einheit eines Satzes

Was ist ein Lemma?

nach Bußmann: Eintrag bzw. einzelnes Stichwort in einem Lexikon

nach Metzler ('Lexikon Sprache'): Repräsentant eines Lexems, zu dem in einem Wörterbuch Sprach- und Sachinformationen geboten werden.

allgemein in der CL: normierte Grundform für alle zu einem Wort gehörenden Wortformen

Was ist Lemmatisierung?

Rückführung einer Wortform auf ihr Lemma

(oft nicht eindeutig möglich: Bsp. *rasten*)

Was ist ein Morphem?

kleinste bedeutungstragende Einheit (oder: kleinste lexikalisch definierte Einheiten, aus denen ein Wort zusammengesetzt ist); Bsp.:

- Kernmorphem (*Berg, seh, gross*)
- Partikelmorphem (*bei, auf, sehr*)
- Derivationsmorphem (*-heit, -tum, ur-, un-*)
- Flexionsmorphem (*e, -test, -ste*)

(vgl. [Vorlesung 2](#))

Wortarten-Klassifikation

siehe [\[Trask 99\]](#) für eine Einführung

Historische Entwicklung der Wortarteneinteilung

Plato 400 v.Chr.	Aristoteles 350 v. Chr.	Stoiker 250 v. Chr.	Dion. Thrax 100 v. Chr.	heute
onoma	onoma	onoma	onoma	Nomen
		prosegoria		Eigenname
		mesotes	epirrhema	Adverb
rhema	rhema	rhema	rhema	Verb
			metoche	Partizip
	syndesmoi	arthra	arthron	Artikel
			antonymia	Pronomen
		syndesmos	prothesis	Präposition
			syndesmos	Konjunktion

Wortarten nach `Thomas von Erfurt'

(spätes Mittelalter; nach [\[Robins 90\]](#) S.89)

nomen: ein Wort, das etwas Beständiges oder Permanentes bezeichnet, oder etwas mit typischen Eigenschaften (distinctive characteristics)

verbum: ein Wort, das den zeitlichen Vorgang bezeichnet, losgelöst von einer Substanz (die es jedoch beschreibt)

participium: ein Wort, das den zeitlichen Vorgang bezeichnet, jedoch nicht losgelöst von einer Substanz (die es beschreibt)

pronomen: ein Wort, das etwas Beständiges oder Permanentes bezeichnet, ohne typische Eigenschaften (distinctive characteristics)

adverbium: ein Wort, das zusammen steht mit einem Wort, das den zeitlichen Vorgang bezeichnet und diesen weiter bestimmt

conjunctio: ein Wort, das den Zusammenschluss zweier anderer Terme bezeichnet

praepositio: ein Wort, das eine syntaktische Konstruktion aufbaut mit einem Kasus-flektierten Wort, und dieses mit einer Handlung verbindet

interiectio: ein Wort, das ein Verb oder Partizip näher bestimmt und Gefühle ausdrückt

Lindley Murray's 'klassische' englische Wortarten:

Die klassische englische Einteilung stammt vom amerikanischen Linguisten Murray aus dem Jahr 1795. Sie basiert auf den Gegebenheiten der englischen Sprache und nicht mehr hauptsächlich auf Latein, was zwei Änderungen bewirkt:

- Adjektive werden als eigene Wortart anerkannt, nicht als Untergruppe der Nomen
- Partizipien werden unter die Gruppe der Verben eingeordnet

Ihre 8-fache Grundeinteilung wurde und wird von sämtlichen englischen Schulgrammatiken befolgt:

- Nomen
 - Verb
 - Adjektiv
 - Adverb
 - Pronomen
 - Präposition
 - Konjunktion
 - Interjektion
-

Wortarten in der Duden-Grammatik:

(4. Auflage 1984 S. 88; vgl. dort auch Tabelle auf S. 91 und in [\[Duden 95\]](#) S. 88)

Def.: Auf Grund der unterschiedlichen Funktionen im Satz und der damit eng verknüpften Formmerkmale, Anordnung und Beziehungen zueinander können verschiedene Gruppen oder Klassen von Wörtern unterschieden werden, die sich auch semantisch voneinander abgrenzen lassen und die man Wortarten nennt.

Verben: Mit Verben wird das Prädikat des Satzes gebildet. Mit Verben bezeichnet der Sprecher, was geschieht oder was ist: Zustände (*wohnen*), Vorgänge (*verblühen*), Tätigkeiten und Handlungen (*bauen*.)

- Vollverben
- Hilfsverben
- Modalverben

Substantive:

- können mit Artikel verbunden werden
- werden als Subjekt, Objekt, adv. Best. oder Attribut gebraucht
- sind deklinierbar (in Kasus, Numerus)
- stehen für Lebewesen, Sachen, Begriffe, Abstrakta

Adjektive:

- sind deklinierbar (in Kasus, Numerus, Genus)
- meist steigerbar
- bezeichnen Eigenschaften und Merkmale
- Beispiele: *gut, schnell, tragbar, provinziell*

Begleiter und Stellvertreter des Substantivs:**Artikel:**

- sind deklinierbar (in Kasus, Numerus, Genus)
- werden nur in Verbindung mit einem Substantiv gebraucht
- Beispiele: *der, den, eine, einen*

Pronomen:

- sind deklinierbar (in Kasus, Numerus, Genus, teilw. Person)
- werden nur in Verbindung mit einem Substantiv oder als Stellvertreter eines Substantivs gebraucht

- Beispiele: *ich, mein, dieser, sich, niemand, keine*

Partikel:

- sind nicht veränderbar (Bsp. *zu* gehen, gibt ... *auf, am* schönsten)

Adverbien:

- Umstandsangabe oder Attribut (Bsp.: *bald, sehr, fast, abends, probeweise*)

Konjunktionen:

- verbinden Wörter oder Wortgruppen (Bsp.: *und, weil, dass*)

Präpositionen:

- bestimmen Kasus des zug. Substantivs
- bezeichnen Verhältnisse oder Beziehungen
- Beispiele: *auf, wegen, seit, jenseits*

Interjektionen:

- syntaktisch isoliert
- steuern und gliedern Dialoge
- Ausrufe, die Empfindungen ausdrücken
- Beispiele: *ach, hm, oh, ei, hoi*

Aufgabe: Wie sollten die folgenden Wörter eingeordnet werden?

Er kannte ihn schon **als** kleines Kind.

Die Wissenschaft **selbst** ist ein kompliziertes System.

Ich habe noch nie **solch** eine Geschichte gehört.

Sie ist **erkrankt**.

Auf der **einen** Seite ist es so, aber ...

Er stellt **zuvielen** Fragen.

Der Mann, von **dessen** Vater das Buch handelt, ist ...

Er kam plötzlich **und ohne** anzuklopfen herein.

Er wartete **bis um** 5 Uhr.

Sie ass die Suppe **auf**.

Das System wurde **flächendeckend** eingeführt.

Es gibt manchmal, je nach Argumentation, verschiedene Einteilungsmöglichkeiten. Wortarteneinteilung ist also auch eine Frage der Kriterien, der Konvention und der Objektsprache. Einteilungskriterien sind :

- Distribution (mögliche syntaktische Positionen)
- Flexion
- Wortbildung
- Bedeutung
- Pragmatik

Nicht alle Wortarten kommen in jeder Sprache vor. z.B. Postpositionen gibt es

- im Englischen den meisten Grammatiken zufolge keine
- Türkisch, Baskisch, Japanisch haben sehr viele Postpositionen
- Im Deutschen gibt es wenige Postpositionen (der Grammatik *zufolge*, der Sache *wegen*) und die in vielen Sprachen (z.B. Englisch) völlig unbekanntes Zirkumpositionen (*um* etwas *herum*, *von* morgen *an*).

Einzig die Wortarten *Nomen* und *Verb* werden von den meisten Linguisten als universell angesehen (siehe Plato oben, 400 v. Chr.).

Wortarteneinteilung in Morphix (nach [\[Finkler und Neumann 86\]](#))

- Vollformenlexikon für
 - *nicht-flektierende Wortarten*
 - Adverb
 - nebenordnende Konjunktion
 - unterordnende Konjunktion
 - Partikel
 - Präposition
 - Verbzusatz
 - Stammformenlexikon für
 - *endungsflektierende Wortarten*
 - Verben
 - Adjektive
 - Nomen
 - Ordinalzahlen
 - Possesivpronomen
 - Relativpronomen
 - Quantoren
 - *nicht-endungsflektierende Wortarten*
 - Personalpronomen
 - Reflexivpronomen
 - Interrogativpronomen
 - Determinativ
 - Hilfsverb
 - Modalverb

Wortarteneinteilung im STTS

Das folgende Tagset ist das “Stuttgart/Tübingen TagSet” (STTS), das von Anne Schiller, Christine Thielen, Simone Teufel und Christine Stöckert in den Jahren 94-95 entwickelt wurde.

Siehe auch: [Tagging-Guidelines etc.](#)

1. Adjektive

ADJA	attributives Adjektiv	[das] große [Haus]
ADJD	adverbiales oder prädikatives Adjektiv	[er fährt] schnell [er ist] schnell

2. Adverbien

ADV	Adverb	schon, bald, doch
-----	--------	-------------------

3. Adpositionen

APPR	Präposition; Zirkumposition links	in [der Stadt], ohne [mich]
APPRART	Präposition mit Artikel	im [Haus], zur [Sache]
APPO	Postposition	[ihm] zufolge, [der Sache]
	wegen	
APZR	Zirkumposition rechts	[von jetzt] an

4. Artikel

ART	bestimmter oder unbestimmter Artikel	der, die, das, ein, eine, ...
-----	---	----------------------------------

5. Zahlen

CARD	Kardinalzahl	zwei [Männer],[im Jahre] 1994
ORD	Ordinalzahl	[der] neunte [August]

6. Konjunktionen

KOUI	unterordnende Konjunktion mit ``zu'' und Infinitiv	um [zu leben], anstatt [zu fragen]
KOUS	unterordnende Konjunktion mit Satz	weil, daß, damit, wenn, ob
KON	nebenordnende Konjunktion	und, oder, aber
KOKOM	Vergleichskonjunktion	als, wie

7. Substantive

NN	normales Nomen	Tisch, Herr, [das] Reisen
NE	Eigennamen	Hans, Hamburg, HSV

8. Pronomen

PDS	substituierendes Demonstrativpronomen	dieser, jener
PDAT	attribuierendes Demonstrativpronomen	jener [Mensch]
PIS	substituierendes Indefinitpronomen	keiner, viele, man, niemand
PIAT	attribuierendes Indefinitpronomen ohne Determiner	kein [Mensch], irgendein [Glas]
PIDAT	attribuierendes Indefinitpronomen mit Determiner	[ein] wenig [Wasser], [die] beiden [Brüder]

PPER	irreflexives Personalpronomen	ich, er, ihm, mich, dir
PPOSS	substituierendes Possessivpronomen	meins, deiner
PPOSAT	attribuierendes Possessivpronomen	mein [Buch], deine [Mutter]
PRELS	substituierendes Relativpronomen	[der Hund ,] der
PRELAT	attribuierendes Relativpronomen	[der Mann ,] dessen [Hund]
PRF	reflexives Personalpronomen	sich, einander, dich, mir
PWS	substituierendes Interrogativpronomen	wer, was
PWAT	attribuierendes Interrogativpronomen	welche [Farbe], wessen [Hut]
PWAV	adverbiales Interrogativ- oder Relativpronomen	warum, wo, wann, worüber, wobei
PAV	Pronominaladverb	dafür, deswegen, trotzdem

9. Partikel

PTKZU	``zu'' vor Infinitiv	zu [gehen]
PTKNEG	Negationspartikel	nicht
PTKVZ	abgetrennter Verbzusatz	[er kommt] an, [er fährt] rad
PTKANT	Antwortpartikel	ja, nein, danke, bitte
PTKA	Partikel bei Adjektiv oder Adverb	am [schönsten], zu [schnell]

10. Verben

VVFIN	finites Verb, voll	[du] gehst, [wir] kommen [an]
VVIMP	Imperativ, voll	komm [!]
VVINFINF	Infinitiv, voll	gehen, ankommen
VVIZU	Infinitiv mit ``zu'', voll	anzukommen, loszulassen
VVPP	Partizip Perfekt, voll	gegangen, angekommen
VAFIN	finites Verb, aux	[du] bist, [wir] werden
VAIMP	Imperativ, aux	sei [ruhig !]
VAINFINF	Infinitiv, aux	werden, sein
VAPP	Partizip Perfekt, aux	gewesen
VMFIN	finites Verb, modal	dürfen
VMINFINF	Infinitiv, modal	wollen
VMPP	Partizip Perfekt, modal	gekonnt, [er hat gehen] können

11. Sonstiges

XY	Nichtwort, Sonderzeichen enthaltend	3:7, H2O, D2XW3
FM	Fremdsprachliches Material	[Er hat das mit] 'A big fish' [übersetzt]
SGML	SGML Markup	
SPELL	Buchstabierfolge	S-C-H-W-E-I-K-L
TRUNC	Kompositions-Erstglied	An- [und Abreise]
ITJ	Interjektion	mhm, ach, tja
\$,	Komma	,
\$.	Satzbeendende Interpunktion	. ? ! ; :
\$(sonstige Satzzeichen; satzintern	- [,] ()

Zusammenfassung

- Der Begriff 'Wort' ist ungenau und muss je nach Anwendung definiert werden.
 - Wichtige, gut definierte Begriffe sind 'Lemma' und 'Morphem'.
 - Die Wortarteneinteilung kann nach unterschiedlichen Kriterien geschehen. Dazu gehören morphologische, syntaktische, semantische und pragmatische Kriterien.
 - Probleme bei der Wortarteneinteilung entstehen z.B. durch:
 - Mehrwortausdrücke (New York, Windows 98, Prof. Hess, Bescheid sagen)
 - Fremdsprachige Ausdrücke (a la carte)
 - Abkürzungen (z.Zt., USA, Ein- und Ausfahrt)
 - Zahlausdrücke (14jährige, 1999, 1-Stunden-Takt)
 - Derivativa (die Illustrierte, das Schöne, die Beiden)
 - Kontrahierte Wortformen (am, ins, geht's, darüber)
 - Internet-Adressen (<http://www.ifi.unizh.ch>)
 - Anwendung verschiedener Einteilungskriterien
 - Andere Sprachen haben teilweise andere Wortarteneinteilungen. Es gibt weiterhin Streitfälle.
-