

Lernmodul *

Linguistisches Propädeutikum

Institut für Computerlinguistik
Universität Zürich

<http://www.cl.uzh.ch>

Simon Clematide
siclemat@cl.uzh.ch

Version von 28. März 2007

*PDF-Version: <http://www.cl.uzh.ch/siclemat/lehre/ling-prop/script/script.pdf>
HTML-Version:
<http://www.cl.unizh.ch/siclemat/lehre/ling-prop/script/html/script.html>
Folien-Version: <http://www.cl.unizh.ch/siclemat/lehre/ling-prop/slides-4up.pdf>

Inhaltsverzeichnis

1	Linguistisches Propädeutikum I	4
1.1	Wort	4
1.1.1	Definition	4
1.1.2	Token	5
1.1.3	Wortform	5
1.1.4	synt. Wort	5
1.1.5	Lexem	6
1.2	Wortarten	6
1.2.1	5 Hauptwortarten nach Glinz	6
1.3	Morphologische Merkmale	7
1.3.1	Genus	8
1.3.2	Zahl	8
1.3.3	Kasus	8
1.3.4	Modus	9
1.3.5	Zeit	9
1.3.6	Person	9
1.3.7	Grad	9
1.3.8	Flexion	10
1.4	STTS	10
1.5	Vertiefung	11
2	Linguistisches Propädeutikum II	12
2.1	Proben	12
2.1.1	Ersetzen	12
2.1.2	Einsetzen	13
2.1.3	Weglassen	13
2.1.4	Verschieben	14
2.1.5	Umformen	14
2.2	Satz	15
2.3	Syntaxanalyse	15
2.3.1	Konstituenz	15
2.3.2	Köpfe	15
2.3.3	Dependenz	17
2.3.4	Satzglieder	17
2.4	Baumbanken	18
2.4.1	NEGRA	18
2.4.2	TIGERSearch	19
2.5	Vertiefung	20
3	Literaturverzeichnis	21

Abbildungsverzeichnis

1.1	5 Hauptwortarten nach Glinz	7
2.1	Moderne Satzgliedlehre	18

1 Linguistisches Propädeutikum I

Lernziele

- Kenntnis der Begriffe Wort, Token, Lexem, Lemma, Lexemverband
- Kenntnis der Wortartenlehre für Deutsch
- Kenntnis der morphologischen Kategorien für Deutsch und Englisch
- Kenntnis und Anwendung des Stuttgart-Tübingen-Tagsets (STTS) mit Hilfe der Referenzkarte
- Erfahrungen mit computerlinguistisch unterstützter, halbautomatischer linguistischer Annotation von Morphologie und Wortarten

1.1 Wort

1.1.1 Definition

Wort

Definition 1.1.1 (nach [BUSSMANN 1990]). *Wort*. Intuitiv vorgegebener und umgangssprachlich verwendeter Begriff für sprachliche Grundeinheiten, dessen zahlreiche sprachwissenschaftliche Definitionsversuche uneinheitlich und kontrovers sind.

Beispiel 1.1.2 (Was ist ein Wort?).

1. Sie wollte vor allem am 1. Spiel teilnehmen.
2. Sie nahm z.B. an dem 2. Spiel teil.
3. Das gibts doch nicht!
4. Blick Online verlost zum Film-Start das «Ich bin Borat»-Kit [...]

Präzisierungsversuche des Wort-Begriffs

Sprachliche Ebenen zur Präzisierung des Wort-Begriffs

- *phonetisch-phonologisch*: Kleinstes durch Wortakzent und Grenzsignale wie Pause oder Knacklaut theoretisch isolierbares Lautsegment
- *orthographisch-graphemisch*: durch Leerstellen im Schriftbild isolierte Einheit
- *morphologisch*: Grundeinheit, welche flektierbar ist
- *lexikalisch-semantisch*: kleinster Bedeutungsträger, welcher im Lexikon kodifiziert ist
- *syntaktisch*: kleinste verschieb- und ersetzbare Einheit des Satzes

Gängige Wortauffassungen [LINKE 2001]

Wieviele verschiedene Wörter hat dieser Satz?

Wenn hinter Fliegen eine Fliege fliegt, fliegt eine Fliege Fliegen nach.

Antworten

- 11
- 7
- 6
- 5
- __

1.1.2 Token

Antwort 11: Wort als Vorkommen einer Wortform

Wieviele verschiedene Wörter hat dieser Satz?

Wenn¹ hinter² Fliegen³ eine⁴ Fliege⁵ fliegt⁶, fliegt⁷ eine⁸ Fliege⁹ Fliegen¹⁰ nach¹¹.

Definition 1.1.3 (Token). Ein *Token* ist ein einzelnes Vorkommen einer Wortform in einem Text.

Bemerkung

In der CL werden Interpunktionszeichen ebenfalls als textuelle Token betrachtet. Die Antwort wäre dann ...

1.1.3 Wortform

Antwort 7: Wort als Wortform

Wenn¹ hinter² Fliegen³ eine⁴ Fliege⁵ fliegt⁶, fliegt⁶ eine⁴ Fliege⁵ Fliegen³ nach⁷.

Definition 1.1.4 (Wortform). Eine *Wortform* ist eine rein graphematische Einheit eines Textes.

Bemerkung

Dieselbe Wortform kann Wörter unterschiedlicher Bedeutung repräsentieren

- Die Fliege war tot.
- Er trug eine samtene Fliege.
- Fliege nicht so schnell!

1.1.4 Syntaktisches Wort

Antwort 7: Wort als syntaktisches Wort

Wenn¹ hinter² Fliegen³ eine⁴ Fliege⁵ fliegt⁶, fliegt⁶ eine⁴ Fliege⁵ Fliegen³ nach⁷.

Definition 1.1.5 (Syntaktisches Wort). Ein *syntaktisches Wort* ergibt sich aus einer Wortform kombiniert mit ihren morphosyntaktischen Merkmalen.

Rein orthographische Varianten einer Wortform werden für ein syntaktisches Wort normalerweise zusammengefasst.

Beispiel 1.1.6 (Wortform vs. syntaktisches Wort).

Die Wortform "Fliegen" kann mindestens 4 syntaktische Wörter repräsentieren: "Fliege" in Nominativ, Akkusativ, Dativ oder Genitiv Plural.

1.1.5 Lexem

Antwort 6: Wort als Lexem

Wenn¹ hinter² Fliegen³ eine⁴ Fliege³ fliegt⁵, fliegt⁵ eine⁴ Fliege³ Fliegen³ nach⁶.

Definition 1.1.7 (Lexem im engen Sinn). Ein *Lexem* ist eine Menge (Paradigma) von syntaktischen Wörtern, welche sich nur in bestimmten morphosyntaktischen Merkmalen (Kasus, Numerus, Tempus usw.) unterscheiden.

Definition 1.1.8 (Nennform, Zitierform, Grundform, Lemma). Ein *Lemma* ist eine lexikographische Standard-Notation für ein Lexem.

Beispiel 1.1.9 (Lexikographische Notationskonventionen im Deutschen).

Substantiv-Lexeme werden durch das syntaktische Wort im Nominativ Singular repräsentiert. Verb-Lexeme durch den Infinitiv. Wie werden Verb-Lexeme im Latein notiert?

Antwort 5: Wort als Lexemverband

Wenn¹ hinter² Fliegen³ eine⁴ Fliege³ fliegt³, fliegt³ eine⁴ Fliege³ Fliegen³ nach⁵.

Definition 1.1.10 (Lexemverband, auch Lexem im weiten Sinn). Ein *Lexemverband* ist eine Menge von Lexemen, welche den gleichen Wortstamm haben.

Bemerkung

Lexemverbände umfassen typischerweise Lexeme mit unterschiedlichen Wortarten.

Beispiel 1.1.11 (Satz als Menge von Lexemverbänden).

Wie lässt sich die Menge der Wörter des Beispielsatzes in Mengennotation als Lexemverband darstellen?

1.2 Lehre von den Wortarten

Wortarten nach [BUSSMANN 1990]

Definition 1.2.1 (Redeteile, engl. *parts of speech* (PoS)). *Wortarten* sind das Ergebnis der Klassifizierung der Wörter einer Sprache nach morphologischen, syntaktischen und/oder semantischen Kriterien.

Historisches [GALLMANN und SITTA 2001]

Die 8-Wortarten-Lehre von Dionysius Thrax (ca. 100 vor Chr.) wirkte stark bis ins 19. Jahrhundert. Für Deutsch wurde in den Schulen lange die traditionelle 10-Wortarten-Lehre (Numerale als eigene Kategorie) gemäss Adelung (1781) vermittelt. Mit der 5-Wortarten-Lehre nach Hans Glinz setzte sich ab den 50er-Jahren eine flachere Einteilung durch, welche explizite (operationalisierbare) und durchgängige Klassifikationskriterien anwendet.

1.2.1 5 Hauptwortarten nach Glinz

Klassifikationskriterien

Welcher Fachbegriff wäre präziser anstelle von "Wörter" in der Abbildung?

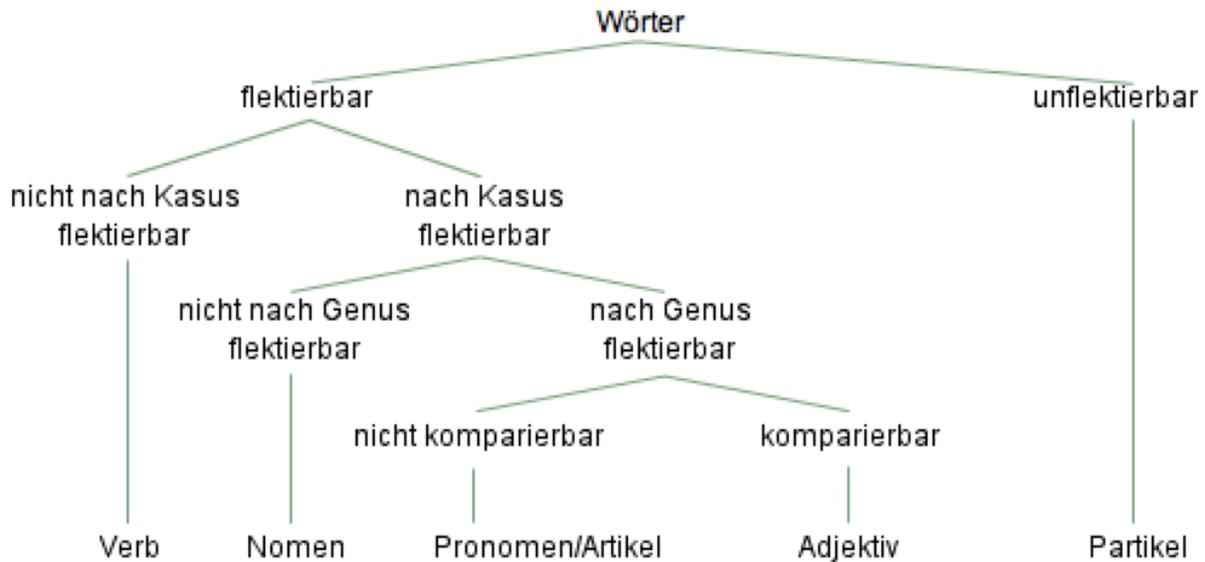


Abbildung 1.1: Die 5 Hauptwortarten nach Glinz [STOCKER et al. 2004]

1.3 Morphologische Merkmale

Morphologische Kategorien und ihre Merkmale

Definition 1.3.1 (auch morphosyntaktisches oder grammatisches Merkmal). Die *morphologischen Merkmale* sind Ausprägungen von morphologischen Kategorien wie Genus, Kasus, Numerus, Person, Tempus, Modus und Komparation, welche durch die Flexion realisiert werden.

Flexion: Bildung von den unterschiedlichen syntaktischen Wortformen eines Lexems

- Konjugation von Verben
- Deklination von Nomen, Adjektiven, Artikeln und Pronomen
- Steigerung von Adjektiven

Diskussion: Sind steigerbare Adverbien ein Widerspruch im System?

Nach Glinz zählen die Adverbien zu den Partikeln. Partikeln sind gemäss dieser Einteilung nicht flektierbar. Zur Flexion wird normalerweise auch die Komparation gezählt. Es gibt einige Adverbien, welche komparierbar sind (“sehr, mehr, am meisten”, “gern, lieber, am liebsten” ...). Ist dies ein Widerspruch? Ein bisschen schon. Aber: Steigerung wird oftmals nur für Adjektive als Flexion betrachtet.

Ist dies ein definitiver Zirkel? Ein bisschen schon. Aber: Was ein Adjektiv ausmacht, ist die Gesamtheit der Attribute im Klassifikationsbaum, d.h. flektierbar nach Kasus, Genus und komparierbar. Somit kann man Adjektiv-Lexeme von Adverb-Lexemen unterscheiden.

Aber es gibt doch auch Adjektiv-Lexeme, welche nicht flektierbar nach Kasus und Genus sind oder nicht attributiv (d.h. in flektierter Position) verwendet werden können? Hmmh, dann muss es wohl noch andere Gründe geben, ein Adjektiv-Lexem zu sein.

Welche denn?

1.3.1 Genus

Das Genus (engl. *gender*): Grammatisches Geschlecht

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
Maskulinum	männlich	masculine	Masc	Elephant
Femininum	weiblich	feminine	Fem	Gazelle
Neutrum	sächlich	neuter	Neut	Huhn
Unterspezifiziert	n/a	n/a	*	Ferien

Tabelle 1.1: Übersicht: Genera

Bemerkung

Der Plural von Genus lautet Genera.

Beispiel 1.3.2 (Bei welchen Wortarten ist die Kategorie Genus ausgeprägt?).

1.3.2 Numerus

Der Numerus (engl. *number*): Grammatische Zahl

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
Singular	Einzahl	singular	Sg	Huhn
Plural	Mehrzahl	plural	P1	Hühner

Tabelle 1.2: Übersicht: Numeri

Bemerkung

Der Plural von Numerus lautet Numeri.

Beispiel 1.3.3 (Bei welchen Wortarten ist die Kategorie Numerus ausgeprägt?).

1.3.3 Kasus

Der Kasus (engl. *case*): Fall

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
Nominativ	Werfall	nominative	Nom	der Baum
Genitiv	Wesfall	genitive	Gen	des Baumes
Dativ	Wemfall	dative	Dat	dem Baum
Akkusativ	Wenfall	accusative	Akk	den Baum

Tabelle 1.3: Übersicht: Kasus

Bemerkung

Der Plural von Kasus lautet Kasus.

Beispiel 1.3.4 (Bei welchen Wortarten ist die Kategorie Kasus ausgeprägt?).

1.3.4 Modus

Der Modus (engl. *mode, mood*): Aussageweise

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
Indikativ	Wirklichkeitsform	indicative	Ind	er geht
Konjunktiv	Möglichkeitsform	subjunctive mood	Konj	er gehe

Tabelle 1.4: Übersicht: Modi

Bemerkungen

Der Plural von Modus lautet Modi.

Die Befehlsform (Imperativ) wird normalerweise auch als Modus aufgefasst. Im STTS ist dies durch das V.IMP ausgedrückt.

1.3.5 Tempus

Das Tempus (engl. *tense*): grammatische Zeit, Zeitform

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
Präsens	Gegenwart	present tense	Pres	er geht
Präteritum	Vergangenheit	past tense	Past	er ging

Tabelle 1.5: Übersicht: Tempora

Bemerkungen

Der Plural von Tempus lautet Tempora.

1.3.6 Person

Die Person (engl. *person*)

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
1. Person	Sprecher	first person	1	ich gehe
2. Person	Angesprochene	second person	2	du gehst
3. Person	Person, Sachverhalt	third person	3	er geht

Tabelle 1.6: Übersicht: Personen

Bemerkungen

Der Plural von Person lautet Personen; die Verwendung ist allerdings ungebräuchlich.

Beispiel 1.3.5 (Bei welchen Wortarten ist die Kategorie Person ausgeprägt?).

1.3.7 Grad

Der Grad (engl. *degree*): Steigerung, Komparation

Beispiel 1.3.6 (Bei welchen Wortarten ist die Kategorie Grad ausgeprägt?).

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
Positiv	Normalform	positive	Pos	schlau
Komparativ	Vergleichsform	comparative	Comp	schlauer
Superlativ	Höchststufe	superlative	Sup	schlauste

Tabelle 1.7: Übersicht: Komparation

1.3.8 Adjektiv-Flexion

Die Adjektiv-Flexion (engl. *adjective inflection*)

Fachbegriff	Deutsch	Englisch	STTS	Beispiel
Schwache Flexion	mit best. Artikel	weak	Schw	der schlaue Fuchs
Gemischte Flexion	mit ein, kein,...	mixed	Mix	keine schlaunen Füchse
Starke Flexion	ohne Artikel	strong	St	schlaue Füchse

Tabelle 1.8: Übersicht: Adjektiv-Flexion

Beispiel 1.3.7 (Bei welchen Wortarten ist die Kategorie Adjektiv-Flexion ausgeprägt?).

1.4 STTS

Stuttgart/Tübingen-Tagset (STTS)

- <http://www.cl.unizh.ch/clab/hilfe/stts/>
- Das Standard-Tagset des Deutschen [SCHILLER et al. 1999] (vergleichbar mit dem Penn-Treebank-Tagset für Englisch)
- Wichtige linguistische Korpora wie NEGRA-Korpus oder TIGER-Korpus sind mit STTS annotiert.
- Frei verfügbare Werkzeuge zum automatischen Bestimmen von Wortarten (sogenannte "Tagger") liefern mit STTS-Tags versehenen Output.
- Eine Notationsvariante von STTS ist als europäischer Standard für Deutsch (EAGLES ELM-DE) [EAGLES 1996] spezifiziert worden.
- Alternativen: Münsteraner Tagset [STEINER 2003]

Besonderheiten von STTS

- 'Wortarten' für satzinterne und -finale Interpunktion
- Kategorie für fremdsprachliches Material (FM)
- Feine Aufgliederung bei Partikeln
- Durchgängige Unterscheidung von attribuierend vs. substituierend bei Pronomen
- Eigennamen sind eine semantisch definierte Klasse.
- Die Ordinalzahlen (ORD) sind im neuesten Standard zu Adjektiven geworden...

1.5 Vertiefung

- Referenzkarte zu STTS und dem NEGRA-Sprachmodell
- Manuals für das Programm **annotate** [PLAEHN 2000, PLAEHN 1998]
- <http://www.canoo.net> aus Basel hat eine sorgfältige linguistische Terminologie im Web und viel (!) Anschauungsmaterial aus ihrem Morphologiesystem.

Pfichtlektüre

Nachschlagen von mindestens 30 Tags im Annotationshandbuch [SCHILLER et al. 1999] bei Unsicherheit im Annotieren von Wortart und morphologischer Kategorie

2 Linguistisches Propädeutikum II

Lernziele

- Kenntnis der klassischen syntaktischen Proben
- Kenntnis der Begriffe Satz, Konstituenz, Dependenz, Konstituente, Kern, Kopf, Satzglied, Phrase
- Kenntnis der wichtigsten syntaktischen Funktionen
- Kenntnis und Anwendung des NEGRA-Annotationsschemas mit Hilfe der Referenzkarte
- Erfahrungen mit computerlinguistisch unterstützter, halbautomatischer linguistischer Annotation von syntaktischen Strukturen
- Erfahrungen mit der einfachen Suche von syntaktischen Strukturen in einer Baumbank

2.1 Linguistische Proben

Linguistische Testverfahren

Definition 2.1.1 (Linguistische Proben). Eine *linguistische Probe* ist ein experimentelles Analyseverfahren der strukturellen Linguistik, das gezielt die sprachliche Kompetenz benutzt, um grammatische Aussagen verifizieren (bestätigen) oder falsifizieren (verwerfen) zu können.

2.1.1 Ersatzprobe

Ersatzprobe

Definition 2.1.2. In der *Ersatzprobe* wird ein mehrdeutiger Ausdruck in einem Satz durch einen grammatisch eindeutig zu interpretierenden Ausdruck ersetzt. Die Interpretation des Satzrestes muss dabei unverändert bleiben.

Beispiel 2.1.3 (Bestimmung des Kasus).

Probe	Mehrdeutig	Eindeutig	Kasus
Baumprobe	Peter ist zornig	Der Baum ist zornig	Nom
Baumprobe	Peter ist heiss	Dem Baum ist heiss	Dat
Frageprobe	Peter ist zornig	Wer ist zornig	Nom
Frageprobe	Peter ist heiss	Wem ist heiss	Dat

Manchmal kann das Scheitern einer Reihe von Ersatzproben Eindeutigkeit schaffen.

Beispiel 2.1.4 (Bestimmung der Wortart von “das”).

- Das ist das Angebot, das uns überzeugt hat.

- Dieses ist das Angebot, welches uns überzeugt hat.
- *Welches ist das Angebot, dieses uns überzeugt hat.
- ?Das ist dieses Angebot, welches uns überzeugt hat.

Probleme der Interpretation

- Test bestanden vs. nicht bestanden ist manchmal abhängig von der ursprünglichen Interpretation des Satzes.
- Ob die Interpretation des Satzrestes sich ändert durch den Ersatz, ist nicht immer leicht zu beurteilen.

2.1.2 Einsetzprobe

Einsetzprobe

Definition 2.1.5. In der *Einsetzprobe* wird ein mehrdeutiger Ausdruck in einem Kontext eingesetzt, der ihn grammatisch eindeutig interpretiert.

Beispiel 2.1.6 (Bestimmung der Wortart).

Probe	Mehrdeutig	Eindeutig	Wortart
Attributivprobe	Das war billig	Der billige Baum	Adjektiv
Attributivprobe	Das war gratis	*Der gratis Baum	Adverb

Einsetzen in Flexionsparadigmen

Eine Variante der Einsetzprobe ist das Anwenden eines Flexikonparadigmas wie Konjugation für Verben oder Steigerung für Adjektive.

2.1.3 Weglassprobe

Weglassprobe

Definition 2.1.7. In der *Weglassprobe* wird von einem grösseren mehrdeutigen Ausdruck soviel Material wie möglich entfernt, um einen eindeutigen Ausdruck zu erhalten.

Beispiel 2.1.8 (Bestimmung eines Satzglieds).

1. Schon einen Tag nach der Abreise seiner Freundin fühlte er sich einsam.
2. Schon einen Tag nach der Abreise seiner Freundin fühlte er sich einsam.
3. Schon einen Tag nach der Abreise seiner Freundin fühlte er sich einsam.
4. Schon einen Tag nach der Abreise seiner Freundin fühlte er sich einsam.

Frage

Warum nicht Variante 4?

2.1.4 Verschiebeprobe

Verschiebeprobe

Definition 2.1.9. In der *Verschiebeprobe* werden Wörter und Wortgruppen im Satz umgestellt, so dass der Satz grammatisch bleibt und sich am Inhalt höchstens die Gewichtung ändert. Damit lassen sich die Anfang und Ende von Satzgliedern erkennen.

Beispiel 2.1.10 (Bestimmung von Satzgliedern).

1. Die Sitzung mit dem Chef bereite ich morgen vor.
2. Morgen bereite ich die Sitzung mit dem Chef vor.
3. Ich bereite die Sitzung mit dem Chef morgen vor.

Regeln: Tausche immer 2 Kandidaten aus, um nicht unnötig ungrammatische (Pseudo-)Sätze zu erzeugen. Ersetze in einfachen Sätzen immer das Satzglied vor dem flektierten Verb.

Verschiebeprobe

Beispiel 2.1.11 (Unzulässiges Verschieben).

1. Die Sitzung mit dem Chef bereite ich morgen vor.
2. *Morgen ich bereite die Sitzung mit dem Chef vor.
3. #Die Sitzung bereite ich morgen mit dem Chef vor.

Gründe

Pseudo-Satz 1 ist ungrammatisch. Satz 2 hat eine andere Bedeutung bekommen durch die Umstellung.

2.1.5 Umformungsproben

Umformungsprobe

Definition 2.1.12. In der *Umformungsprobe* werden Sätze umfassend umgebaut.

Beispiel 2.1.13 (Funktion von Nebensätzen).

1. Es würde mich freuen, wenn du mitkäme.
2. Dein Mitkommen würde mich freuen.

Der Nebensatz mit “wenn” erfüllt eine analoge Funktion wie “Es”.

Beispiel 2.1.14 (Infinitivumformung zur Subjekterkennung).

1. Die Lärche ist ein Nadelbaum.
2. ein Nadelbaum sein / die Lärche

Definition 2.3.4. Ein *Kopf* oder *Kern* einer Konstituente ist diejenige Unterkonstituente, welche die grammatischen Eigenschaften ihrer Mutterkonstituente bestimmt und von der die Geschwisterkonstituenten abhängig sind. Der Kern ist normalerweise nicht weglassbar (Weglassprobe).

Beispiel 2.3.5 (Welches ist der Kern der geklammerten Konstituenten?).

- Er [hält ihm den Ausweis unter die geschwollene Nase].
- Sie rennt [mit dem Mobile hantierend] zum Kiosk.
- Es wird [viel zu oft] telefoniert.
- [Die Frau, die zum Kiosk rannte,] war aufgeregt.
- Sie fühlte sich [wie in einem schlechten Film].
- Aber sie war auch [in einem ziemlich schlechten Film].

Konstituententypen

Typen von Konstituenten

1. *Verbalgruppe, -phrase (VP)*: Konstituente mit einem verbalen Kern. In NEGRA umfasst VP nur infinite, d.h. nicht-flektierte Kerne. Die Kategorie Satz (S) hat finite Verbalkerne.
2. *Nominalgruppe, -phrase (NP)*: Konstituente mit Nomen oder Pronomen als Kern.
3. *Adjektivgruppe, -phrase (AP)*: Konstituente mit Adjektiv oder adjektivisch verwendetem Partizip als Kern.
4. *Adverbgruppe, -phrase (AVP)*: Konstituente mit Adverb als Kern.
5. *Präpositionalgruppe, -phrase (PP)*: Konstituente mit Präposition oder Postposition als Kern.
6. *Konjunkionalgruppe, -phrase*: Konstituente mit der Konjunktion “als” oder “wie” als Kern. In NEGRA ebenfalls mit PP bezeichnet.

Koordination, Konstituenz und Köpfe

Definition 2.3.6. *Koordination* (Beiordnung) ist eine syntaktische Struktur, welche aus zwei oder mehr Konjunkten (Wörter, Wortgruppen oder Sätzen) besteht.

Beispiel 2.3.7 (Koordination von Konstituenten aus NEGRA-Korpus).

- Selbst die flotteren Passagen werden nie *ausgelassen und fröhlich*. [Satz 35]
- Oder saß es wieder einmal immer nur mit der Teeflasche *im Auto und in der Sportkarre?* [Satz 9328]
- Wenn es eine Organisation gibt, der vertraut wird und die etwas erreichen kann, ist das die Kirche. [Satz 11'978]

Problem

Was ist der Kopf?

Keine Entscheidung in NEGRA, denn es gibt die Konstituenten CS, CNP, CAP usw.

2.3.3 Dependenz bzw. syntaktische Funktion

Dependenz und syntaktische Funktion

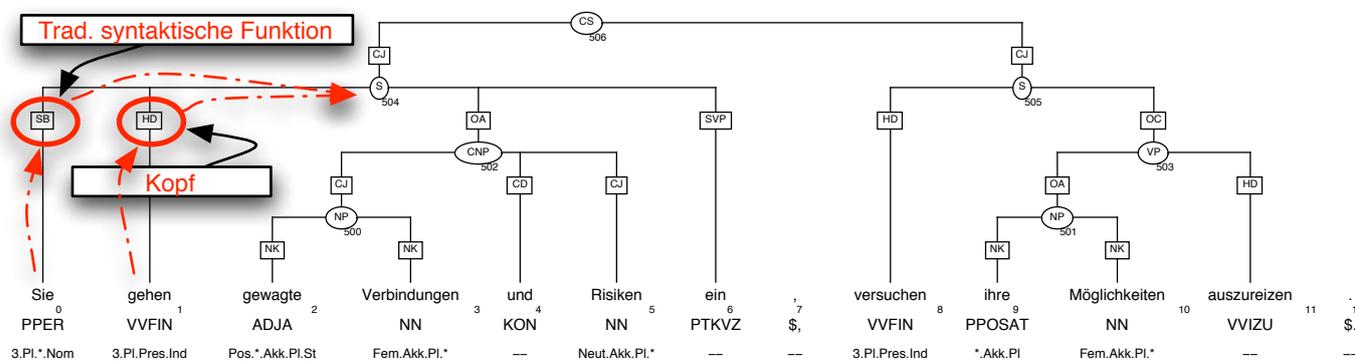
Definition 2.3.8. *Dependenz* ist die syntaktische Relation der Abhängigkeit zwischen Konstituenten.

Definition 2.3.9 (nach [BUSSMANN 1990]). *Syntaktische Funktion* ist ein Sammelbegriff für Beschreibungsgrößen wie “Subjekt”, “Objekt”, “Prädikat”, “Adverbial”, “Attribut” u.a., welche nach Sprachtyp oder Grammatiktheorie unterschiedlich verwendet werden, um die Beziehung zwischen abhängigen Konstituenten zu bestimmen.

Hinweis

Traditionelle Ansätze blenden gerne die syntaktische Funktion des “Kerns” sowie der koordinativen Verknüpfungen aus.

Syntaktische Funktionen in der Annotate-Darstellung



Jedes eckige Kästchen repräsentiert eine Dependenz zwischen der dominierten und der dominierenden Konstituente.

2.3.4 Satzglieder

Einige Satzglieder mit NEGRA-Funktion

- *Subjekt (SB)*: Infinitivprobe oder Frageprobe (Wer oder was?). Subjekt kann auch ein S oder eine VP sein!
- *Akkusativ-Objekt (OA)*: Frageprobe (Wen oder was?) oder Ersatzprobe (“ihn”)
- *Dativ-Objekt (DA)*: Frageprobe (Wem?) oder Ersatzprobe (“ihm”)
- *Genitiv-Objekt (OG)*: Frageprobe (Wessen?) oder Ersatzprobe (“seiner”)
- *Adverbiales und prädikatives Präpositionalglied, Adverbglied, Konjunkionalglied u.ä. (MO)* für Modifikator
- *Nicht-finite Verbalteile (OC)*: Abhängig von flektiertem oder nicht-flektierten Verb

Neue Satzgliedlehre

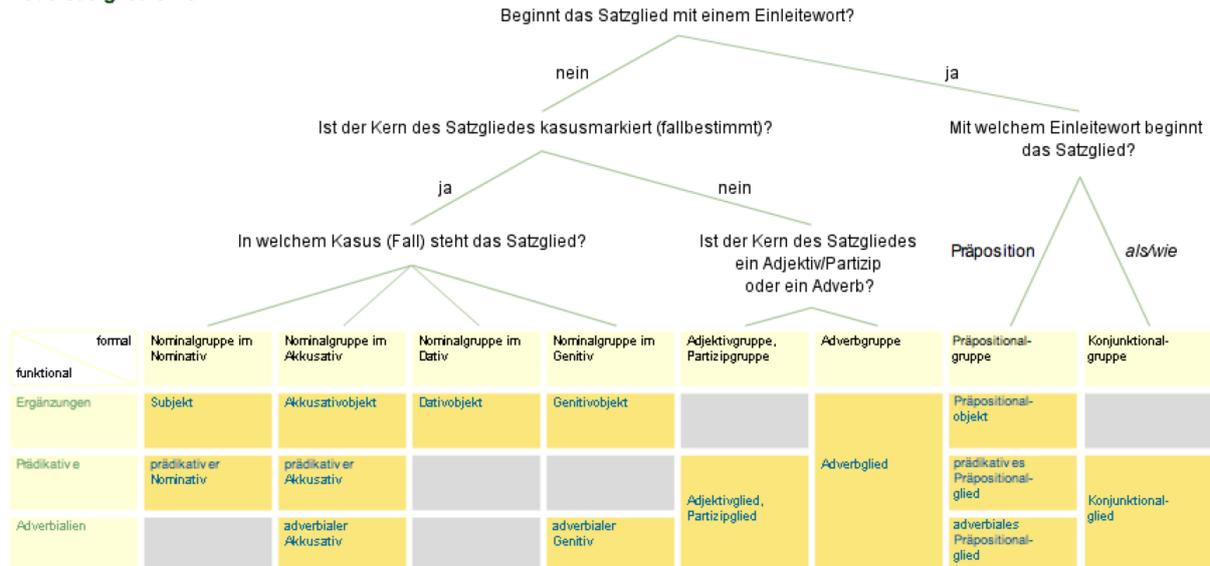


Abbildung 2.1: Moderne Satzgliedlehre nach [STOCKER et al. 2004]

Einige Gliedteile mit NEGRA-Funktion

- Artikel, attributive Adjektiv(phrasen), begleitende Pronomen (NK)
- Präpositionen und konjunktionales “als” oder “wie” (AC)
- Genitivattribute von Nominalphrasen (AG) oder feinere Unterscheidung: pränominaler (GL) oder postnominaler (GR) Genitiv
- Koordinierende Konjunktion (CD) und die koordinierten Konstituenten (CJ)
- Platzhalter (PH) und wiederholte Elemente (RE)

2.4 Baumbanken

2.4.1 NEGRA

NEGRA-Baumbank

- Das erste grosse (20'000 Sätze) Baumbank-Projekt für Deutsch (1997-2001): <http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html>
- PP und NP werden mit flacher Hierarchie verbaut
- Ketten von infiniten Verben (VP) sind zu verschachteln
- Syntaktische Phrasenknoten werden nur für Wortgruppen gebildet, nie für Einzelwörter
- Keine annotierten Köpfe in NP
- Keine explizite Unterscheidung der Satztypen

- Annotationshandbuch [BRANTS et al. 1999] mit detaillierten Regeln
- Die TIGER-Baumbank (TIGER-Korpus) (50'000 Sätze im 2006) ist "Nachfolgeprojekt"

2.4.2 TIGERSearch

TIGERSearch: Ein modernes Suchwerkzeug

Lernen aus Beispielen

Beim Annotieren ist es hilfreich, aus bereits (hoffentlich korrekt!) annotiertem Material zu lernen!

Die 3 Hauptfunktionen von TIGERSearch



Abfrage-Sprache (*investigation*)



Visualisierung der Suchresultate und Baumbank (*exploration*)



Einfache statistische Auswertung (*condensation*)

Die wichtigsten Suchrezepte

- Wie wird das Wort "der" verwendet?
[word="der"]
- Mit welchen Wortarten kommt das Wort "der" wie oft vor? ("#w:" braucht's für das Statistikmodul als Knoten-Identifikator)
#w: [word="der"]
- Welche Adverbien kommen vor?
#w: [pos="ADV"]
- Welche koordinierten Nominalphrasen kommen vor?
#p: [cat="CNP"]
- Welche Dativobjekte kommen vor?
[] >DA #da: []

NB: [] > [] = unmittelbare Dominanz und [] . [] = lineare Präzedenz ...

Weiteres

- Homepage von TIGERSearch: <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/>
- Sofortbenutzung via X11: `ssh -Y tigersearch@hal.ifi.unizh.ch` (ev. -Y durch -X ersetzen; Joe-Account; kein VPN erforderlich)
- Anleitung mit Bildern http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=weitere_tiger.html

2.5 Vertiefung

- Das Kapitel Baumbanken [CARSTENSEN et al. 2004, 414ff.]
- QUIZ Die Form der Satzglieder

3 Literaturverzeichnis

- [BRANTS et al. 1999] BRANTS, THORSTEN, R. HENDRIKS, S. KRAMP, B. KRENN, C. PREIS, W. SKUT und H. USZKOREIT (1999). *NEGRA Annotierschema*, unpublished, Arbeitsmaterial. 19
- [BUSSMANN 1990] BUSSMANN, HADUMOD (1990). *Lexikon der Sprachwissenschaft*. Kröner, Stuttgart, 2., völlig neu bearbeitete Aufl. 4, 6, 15, 17
- [CARSTENSEN et al. 2004] CARSTENSEN, KAI-UWE, C. EBERT, C. ENDRISS, S. JEKAT, R. KLABUNDE und H. LANGER, Hrsg. (2004). *Computerlinguistik und Sprachtechnologie : Eine Einführung*. Elsevier, München, <http://www.ifi.unizh.ch/CL/CLBuch/>. 20
- [DUDENREDAKTION 2005] DUDENREDAKTION, Hrsg. (2005). *Duden, die Grammatik: unentbehrlich für richtiges Deutsch*, Bd. 4 d. Reihe *Der Duden*. Dudenverlag, 7. Aufl. 15
- [EAGLES 1996] EAGLES (1996). *ELM-DE: EAGLES Specifications for German morphosyntax: Lexicon Specification and Classification Guidelines*, electronic, http://www.ilc.cnr.it/EAGLES96/pub/eagles/lexicons/elm_de.ps.gz, Arbeitsmaterial. 10
- [GALLMANN und SITTA 2001] GALLMANN, PETER und H. SITTA (2001). *Deutsche Grammatik*. Lehrmittelverlag, 3. Aufl., Konzis und verständlich. 6
- [LINKE 2001] LINKE, ANGELIKA ET AL., Hrsg. (2001). *Studienbuch Linguistik*. Niemeyer, Mit interaktivem Lernprogramm auf CD-ROM. 5
- [PLAEHN 1998] PLAEHN, OLIVER (1998). *ANNOTATE: Bedienungsanleitung*, Arbeitsmaterial. 11
- [PLAEHN 2000] PLAEHN, OLIVER (2000). *ANNOTATE v3.6 – Quick Reference*, Arbeitsmaterial. 11
- [SCHILLER et al. 1999] SCHILLER, ANNE, S. TEUFEL und C. STÖCKERT (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*, <http://www.ims.uni-stuttgart.de/projekte/corp-lex/TagSets/stts-1999.pdf>. 10, 11
- [STEINER 2003] STEINER, PETRA (2003). *Das revidierte Münsteraner Tagset / Deutsch (MT/D). Beschreibung, Anwendung, Beispiele und Problemfälle*, http://santana.uni-muenster.de/Publications/tagbeschr_final.ps. 10
- [STOCKER et al. 2004] STOCKER, CHRISTA, D. MACHER, R. STUDLER, N. BUBENHOFER, D. CREVLIN, R. LINIGER und M. VOLK (2004). *Studien-CD Linguistik: Multimediale Einführungen und interaktive Übungen zur germanistischen Sprachwissenschaft*, Max Niemeyer Verlag, <http://www.ds.unizh.ch/studien-cd>. 7, 18

Index

Adjektiv-Flexion, 10
Dependenz, 17
Einsetzprobe, 13
Ersatzprobe, 12
Genus, 8
Grad, 9
Kasus, 8
Kern, 15
Komparation, *siehe* Grad
Konstituente, 15
Konstituentenanalyse, 15
Konstituenz, 15
Koordination, 16
Kopf, 15
Lemma, 6
Lexem, 6
Lexemverband, 6
Merkmal, morphologisch, 7
Modus, 9
Numerus, 8, 10
Person, 9
Probe, linguistisch, 12
Satz, 15
Syntaktische Funktion, 17
Tempus, 9
Token, 5
Umformungsprobe, 14
Verschiebeprobe, 14
Weglassprobe, 13
Wort, 4
Wort,syntaktisch, 5
Wortarten, 6
Wortform, 5