

## Sprachorientierte KI Syntax und Parsing

Wolfgang Menzel  
menzel@informatik.uni-hamburg.de

Fachbereich Informatik  
Universität  $\frac{1}{2}$  Hamburg



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 1

## Sprachorientierte KI: Syntax und Parsing

- Syntax als Untersuchungsgegenstand
- Wortartendisambiguierung
- Phrasenstrukturgrammatiken
- Parsing mit Phrasenstrukturgrammatiken
- Unifikationsgrammatiken
- Constraint-basierte Grammatiken
- Parsing als Constraint-Satisfaction



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 2

## Literatur: Linguistik

**Günther Grewendorf, Fritz Hamm und Wolfgang Sternefeld.**

*Sprachliches Wissen. Eine Einführung in die Theorie der grammatischen Beschreibung, Suhrkamp Taschenbuch Wissenschaft, Band 695. Suhrkamp, Frankfurt/Main, 1987.*

**Gisbert Fanselow und Sascha W. Felix.** *Sprachtheorie, Bd. 2*

*Die Rektions- und Bindungstheorie, UTB Linguistik, Band 1442. A. Francke, Tübingen, 1987.*

**Arnim von Stechow und Wolfgang Sternefeld.** *Bausteine syntaktischen Wissens. Ein Lehrbuch der generativen Grammatik. Opladen, 1987.*

**Carl Pollard und Ivan A. Sag.** *Head Driven Phrase Structure Grammar, University of Chicago Press, Chicago 1994.*



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 3

## Literatur: Computerlinguistik

**Stuart M. Shieber.** *An Introduction to Unification-Based Approaches to Grammar. CSLI Lecture Notes. Stanford 1986.*

**Gerald Gazdar und Christopher Mellish.** *Natural Language Processing in Prolog. Addison Wesley, Wokingham, 1989.*

**Daniel Jurafsky und James Martin.** *Speech and Language Processing. Prentice Hall, Upper Saddle River, 2000.*

**Christopher Manning und Hinrich Schtze.** *Foundations of statistical natural language processing. MIT-Press, Cambridge 1999.*

**K.-U. Carstensen, Ch. Ebert, C. Endriss, S. Jekat, u.a.**

*Computerlinguistik und Sprachtechnologie. Spektrum, Heidelberg 2001*



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 4

## Sprachorientierte KI: Syntax und Parsing

- Syntax als Untersuchungsgegenstand
- Wortartendisambiguierung
- Phrasenstrukturgrammatiken
- Parsing mit Phrasenstrukturgrammatiken
- Restrikingierte Phrasenstrukturgrammatiken
- Unifikationsgrammatiken
- Constraint-basierte Grammatiken
- Robustes Parsing



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 5

## Syntax als Untersuchungsgegenstand

- Das Umfeld
- Wozu Syntax?
- Universalgrammatik
- Kompetenz vs. Performanz
- Sprachliche Beobachtungsdaten
- Strukturtests
- Syntaktische  $\Phi_{\frac{1}{2}}$ omene
- Strukturbeschreibungen



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 6

## Das Umfeld

- Sprachorientierte Künstliche Intelligenz
  - Modellierung und Simulation der menschlichen Sprachverarbeitung
  - Entwicklung "intelligenter" Sprachverarbeitungssysteme
  - Integration mit wissensbasierten Inferenzkomponenten
- Computerlinguistik
  - Repräsentation von Sprache mit formalen Mitteln
  - Verarbeitung von Sprache mit Computern
  - Anwendungen (Übersetzung, Sprachverstehen, Informationssuche, ...)
- Kognitive Linguistik
  - psychologisch plausible Modellierung des sprachlichen Wissens



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 7

## Das Umfeld

- Phonologie
  - Kombination elementarer lautsprachlicher Einheiten zu komplexen Äußerungen
  - Aussprache
    - das Haus, die Häuser, häuslich, Behausung*
  - Betonung
    - Generali'tät, \*'bearbeiten*
  - lautliche Gliederung
    - Der gute Mann denkt an sich selbst zuletzt.*
- Phoneme/Prosodeme: kleinste *bedeutungsdifferenzierende* Einheiten



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 8

## Das Umfeld

- Morphologie
  - Kombination elementarer bedeutungstragender Einheiten zu Wörtern und Wortformen
  - Wortbildung
    - treffen, Treffer, zutreffen, \*Zutreffler*
    - der Hausbau, beim Hausbauen, \*ich hausbaue*
  - Flexion
    - warte, du wartest, ...*
    - ich baue, du baust, ...*
    - ich laufe, du läufst, ...*



## Das Umfeld

- Morphologie (2)
  - lexikalische Semantik
    - richtig, unrichtig*
    - handeln, Handlung*
    - Sonnenschutz, Arbeitsschutz*
  - Morpheme: kleinste, selbständig *bedeutungstragende* Einheiten



## Das Umfeld

- Syntax
  - Kombination von Wörtern zu Wortgruppen und Sätzen
    - Ich lese das Buch.*
    - \*Das Buch ich lese.*
    - \*Mir lese das Buch.*
  - gegenseitige Beeinflussung von Wortformen im Satz
    - Hans liebt [sehr schnelle] Autos.*
    - \*Hans [liebt sehr] schnelle Autos.*



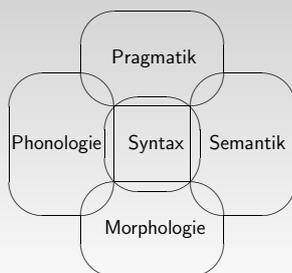
## Das Umfeld

- Semantik
  - Komposition von Satzbedeutungen aus Wortbedeutungen
  - Aufbau einer Bedeutungsrepräsentation (logische Form)
    - Hans liebt Maria* → *liebt (Hans, Maria)*
- Pragmatik
  - angemessene Verwendung von Äußerungen zum Erreichen einer kommunikativen Absicht



## Das Umfeld

- Überschneidungen und Grenzgebiete



## Wozu Syntax?

- Beschreibung sprachlich korrekter Sätze durch kompositionale Regeln
  - Anwendungen:
    - Grammatiken für den Sprachunterricht
    - Tutorielle Systeme für den Sprachunterricht
    - Spracherkennung: Reduzierung der Erkennungsunsicherheit durch syntaktische Einschränkungen



## Wozu Syntax?

- Interface zu Semantik und Pragmatik
  - Vorstrukturierung als Grundlage für
    - semantisch-konzeptuelle Interpretation und
    - Einbettung in den Diskurszusammenhang
  - Gerüst für die Ermittlung der Satzbedeutung
  - Gliederung in die Grundeinheiten einer Bedeutungsrepräsentation
- Syntax ist Abstraktion von semantischen und pragmatischen Faktoren



## Wozu Syntax?

- Ziele linguistischer Beschreibung
  - Was ist natürliche Sprache?
  - Was unterscheidet verschiedene natürliche Sprachen?
  - Wie wird sprachliches Wissen bei Produktion und Rezeption verwendet?
  - Wie wird Sprache erworben?



## Wozu Syntax?

- Ziele maschineller Sprachverarbeitung
  - Welches sprachliche Wissen ist für eine Aufgabe relevant?
  - Welche Aufgaben kann man auch mit partiellem sprachlichen Wissen lösen?
  - Welche Verarbeitungsprozeduren sind für bestimmte Aufgaben erforderlich?
  - Wie kann man sprachliches Wissen repräsentieren, so dass eine effiziente Verarbeitung unterstützt wird?
  - Was sind die minimalen Anforderungen an eine aufgabenadäquate Systemarchitektur?
  - Aus welchen Quellen und mit welchen Techniken kann sprachliches Wissen zusammengetragen werden?



## Wozu Syntax?

- Syntax - the rich sister of poor semantics.
  - traditioneller Schwerpunkt der linguistischen Forschung
  - einer Formalisierung leicht zugänglich
  - für viele Anwendungen unverzichtbar (Generierung ohne Syntax ist gegenstandslos)
  - viele Anwendungen auf syntaktischer Grundlage gut lösbar (z.B. Maschinelle Übersetzung)



## Wozu Syntax?

- Syntax as last resort. (BEVER 1970, SCHANK UND RIESBECK 1978)
  - Totaler Verzicht auf Syntax: Conceptual Dependency
    - "linguistischer Minimalismus" (TANENHAUS 1988)
  - pro: Mensch versteht auch regelwidrige Sätze  
*Gestern ihm gehen Kino.*
  - contra: sinnleere Sätze sind auch syntaktisch analysierbar  
*Knuffer mumpeln Gicklern Bota.*  
deutlicher: *Die Knuffer mumpeln den Gicklern ein Bota.*



## Wozu Syntax?

- ad hoc-Ansatz: Semantische Grammatiken
  - stark anwendungsabhängige, nicht verallgemeinerbare Lösungen
- Synthese: Modelle und Verfahren zur systematischen Interaktion von Syntax, Semantik, Diskursmodell und Weltwissen
  - Pipeline-Architekturen
  - Ziel: Integrierte Architekturen



## Universalgrammatik

- Grammatik = sprachliches Wissen
  1. einzelsprachliche Grammatik
  2. sprachübergreifende Theorie
- deskriptive vs. präskriptive Grammatik



## Universalgrammatik

- Universalgrammatik
  - biologisch orientierter Ansatz
  - Bezug zur menschlichen Sprachverarbeitung:  
"Alle Sprachen haben die und die Eigenschaft, weil das Sprachvermögen die und die Eigenschaft hat."  
(VON STECHOW UND STERNEFELD 1988)



## Universalgrammatik

- Sprachmodul als mentales Organ
  - language acquisition device
  - Sprachfähigkeit ist nicht erlernbar!
  - aber: biologische Fundierung im einzelnen noch völlig unklar und umstritten



## Universalgrammatik

- Prinzipien
  - sehr allgemeine Grundsätze für die Strukturierung sprachlicher Ausdrücke
  - reflektieren die Arbeitsweise des menschlichen Sprachmoduls
- Beispiel: syntaktisches Autonomieprinzip  
keine syntaktische Regel darf sich auf pragmatische, phonologische oder semantische Information beziehen



## Universalgrammatik

- Beispiel: Prinzip der strukturellen Determiniertheit
  - Alle grammatischen Regeln sind strukturabhängig
  - Beispiel: Fragesatzbildung (Subjekt-Auxiliar-Inversion)  
*Hans wird nach Berlin fahren.*  
*Wird Hans nach Berlin fahren?*
  - Regel (1. Version)  
Bewege die zweite Wortform an den Satzanfang  
*Der Chef wird nach Berlin fahren.*  
*Der neue Chef wird nach Berlin fahren.*
  - Regel (2. Version)  
Vertausche die Reihenfolge von Nominalgruppe und Hilfsverb (AUX-Inversion)  
*[Der Chef] wird nach Berlin fahren.*  
*[Der neue Chef] wird nach Berlin fahren.*



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 25

## Universalgrammatik

- Parameter
  - Bündel von Eigenschaften, die immer gemeinsam auftreten
  - Beispiel: Pro-drop
    - in manchen Sprachen (Italienisch, Griechisch, Spanisch) kann die Subjektsposition im Satz unbesetzt bleiben.  
Ital.: *Giovanni canta.*  
*Canta.*  
Engl.: *John sings.*  
*\*Sings.*
  - Beobachtung (PERLMUTTER UND RIZZI):  
Pro-drop-Sprachen weisen immer weitere Eigenschaften auf, die Non-pro-drop-Sprachen nicht besitzen



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 26

## Universalgrammatik

- Tendenz zur Konvergenz linguistischer Theorien
  - Strukturhypothese
  - Prinzipienorientierte Beschreibung
  - Lexikalisierung



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 27

## Universalgrammatik

- Idealisierungen
  - Abstraktion von individuell unvollkommenen Sprachfähigkeiten
  - unbegrenzte Verarbeitungskapazität
  - Abstraktion von Kontexteinflüssen
  - homogene Sprachgemeinschaft



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 28

## Universalgrammatik

- drei Ansätze zur syntaktischen Modellierung
  1. vom Problem zum Formalismus: Theoriebildung
  2. vom Formalismus zum Problem: Theorieverifizierung
  3. Anwendungsperspektive



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 29

## Universalgrammatik

- bei 1) und 2) Konzentration auf ausgewählte Erscheinungen
  - linguistisch interessante Probleme
  - vollständige Abdeckung der Sprache meist nicht angestrebt
  - zentrale Zielstellung: Suche nach Erklärungen für crosslinguistische Phänomene (Universalien)
  - für praktische Anwendungen (selbst konfrontative Maschinelle Übersetzung) oftmals recht uninteressant.



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 30

## Universalgrammatik

- Anwendungsperspektive
  - alle Probleme müssen gelöst werden
  - vollständige Abdeckung eines Sprachbereichs gefordert
  - Modellierung stärker unter Verarbeitungsgesichtspunkten
- schwere Vorwürfe gegen eine theoretisch orientierte Computerlinguistik
  - TOMITA (1988): Linguistisch vs. praktisch interessante Sätze
  - HEYER (1992): Forderung nach anwendungsbezogener CL
- empirische Wende
  - corpusbasierte Entwicklung, trainierbare Modelle
  - Evaluation auf standardisierten Testdaten



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 31

## Universalgrammatik

- dennoch: theoretische Syntaxforschung ist von großem Interesse auch für die praktische Computerlinguistik
  - hohe Modelladäquatheit in Teilgebieten
  - Probleme gut herausgearbeitet
  - "saubere" Lösungen, keine ad hoc-Tricks
  - redundanzfreie Beschreibungen
  - Eleganz



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 32

## Kompetenz vs. Performanz

- Sprachfähigkeit = grammatische Kompetenz  
CHOMSKY (1965)
  - (z.T. unbewusstes) Wissen eines Sprechers über die Gesetzmäßigkeiten und Bestandteile seiner Sprache
  - Fähigkeit, korrekte Sätze zu bilden und mit ihnen Inhalte (Bedeutungen) zu übermitteln



## Kompetenz vs. Performanz

- kompetenzbasierte sprachliche Fähigkeiten (nach GREWENDORF ET AL. 1987)
  - Entscheidung über die Identität zweier Äußerungen
  - Äußerungen korrekt segmentieren (Laute, Silben, Wörter, Wortgruppen)
  - Entscheidung über die Grammatikalität einer Äußerung



## Kompetenz vs. Performanz

- kompetenzbasierte sprachliche Fähigkeiten (2)
  - Entscheidung über die Bedeutungsäquivalenz von Äußerungen
    - Was wir brauchen, ist ein selbsttragender Aufschwung.*
    - Was wir brauchen, ist ein Aufschwung, der sich selbst trägt.*
  - Feststellen von Ambiguitäten
    - In Dacca wurde ein Anhänger der Volkspartei von Benazir Bhutto auf offener Straße erschossen.*
    - Der Vater lässt die Kinder für sich sorgen.*
    - Er hat liebe Genossen. / Er hat Liebe genossen.*



## Kompetenz vs. Performanz

- kompetenzbasierte sprachliche Fähigkeiten (3)
  - Feststellen von graduellen Abweichungen in der Grammatikalität von Äußerungen
    - Er schneidet sich ins eigene Fleisch.*
    - Er schneidet mir den Rückweg ab.*
    - Er schneidet sich ins eigene Fleisch und mir den Rückweg ab.*
    - Hans liest mir das Buch.*
    - Ich wurde nach Hause gegangen.*
    - Er hat aus Berlin gestammt.*
    - Ich habe gestürzt.*
    - Er sagte, dass du hast in Italien gelebt.*
    - Mancher in Deutschland wollen gehen in Italien.*
    - Huming la burbu loris singen vorn.*



## Kompetenz vs. Performanz

- kompetenzbasierte sprachliche Fähigkeiten (4)
  - Unterscheidung verschiedener Typen von abweichenden Äußerungen
  - Erkennen von Unterschieden in den strukturellen Beziehungen innerhalb von Sätzen
    - Die Zwillinge sind schwer zu unterscheiden.*
    - Die Zwillinge sind fähig zu arbeiten.*



## Kompetenz vs. Performanz

- Performanz = Umsetzung der Kompetenz in Sprechhandlungen
  - unvollkommene Realisierung unter Ressourcenbeschränkungen
  - aber: Performanz ist nicht nur fehlerbehaftete Kompetenz
  - basiert auf pragmatischer Kompetenz



## Kompetenz vs. Performanz

- pragmatische Kompetenz (nach GREWENDORF ET AL. 1987)
  - Erkennen von Sprecherintentionen (Versprechen, Vorschlägen, Zustimmung, Ablehnen, ...)
  - Erkennen von emotional gefärbten Äußerungen
    - Hilfst du mir beim Aufräumen?*
    - Hilfst du mir wenigstens beim Aufräumen?*
  - Erkennen indirekter Sprecherintentionen
    - Kannst du mir sagen, wie spät es ist?*



## Kompetenz vs. Performanz

- pragmatische Kompetenz (2)
  - Erkennen versteckter Sprecherintentionen
    - Der Steuermann ist heute betrunken.*
    - Der Kapitän ist heute nicht betrunken.*
  - Erkennen verletzter Präsuppositionen
    - Hans weiß, dass Peter Maria liebt, aber Peter liebt Maria nicht*



## Kompetenz vs. Performanz

- pragmatische Kompetenz (nach GREWENDORF ET AL. 1987)
  - Erkennen inkonsistenter Sprecherintentionen  
*Hans ist ganz bestimmt ein Lügner, aber ich glaube nicht, dass er ein Lügner ist.*
- pragmatische Kompetenz bleibt bei der syntaktischen Modellierung unberücksichtigt



## Sprachliche Beobachtungsdaten

- Datengewinnung für die Modellierung
  - in der Naturwissenschaft: Beobachten, Messen, Experimentieren
  - in der Psychologie (Sprache als mentales Produkt!): Reaktionszeiten
- (idealisierte) sprachliche Äußerungen
  - Elementareinheiten: Wörter
  - komplexe Einheiten: Sätze
  - Begriffe heftig umstritten
  - externe Definition praktisch unmöglich



## Sprachliche Beobachtungsdaten

- Korrektheitsurteile
  - unabhängig von der Verstehbarkeit (CHOMSKY 1957)  
*Colourless green ideas sleep furiously.*  
*\*Furiously sleep ideas green colourless.*
  - Rolle von Grammatikalitätsurteilen ist umstritten: menschliche Sprachfähigkeit dient der Kommunikation, nicht der Unterscheidung von richtig und falsch
  - aber: kein Selbstzweck, sondern linguistisch-empirische Technik um strukturelle Zusammenhänge aufzuklären
  - eröffnen ein gutes Beobachtungsfenster auf die Auswirkungen der zugrundeliegenden syntaktischen Phänomene



## Sprachliche Beobachtungsdaten

- Korrektheitsurteile (2)
    - Grammatikalität ist u. U. kontextabhängig.  
*\*Peter thinks bananas.*  
*What are Mary's favourite fruits?*  
*Peter thinks bananas.*
- elliptische Bildung von  
*Peter thinks bananas are Mary's favourite fruits.*



## Sprachliche Beobachtungsdaten

- graduierte Akzeptabilitätsurteile (z.B. USZKOREIT 1986)
  - ..., weil er ihm das Buch gegeben hat.
  - ..., weil er das Buch ihm gegeben hat.
  - ..., weil ihm das Buch er gegeben hat.
  - ..., weil das Buch er ihm gegeben hat.
  - ..., weil ihm er das Buch gegeben hat.
  - ..., weil das Buch ihm er gegeben hat.

Akzeptabilität ist abhängig von der Verwendungssituation  
→ CHOMSKY (1965): Performanzkriterium

■ Präferenzen

*Er sucht den Mann mit der Taschenlampe.*



## Sprachliche Beobachtungsdaten

- Spracherwerb
  - *Hanni pullert.*
  - *Hanni Ball spielt.*
  - *Hanni spielt Ball.*



## Sprachliche Beobachtungsdaten

- Sprechfehler
  - Hypothese: mögliche Versprecher sind durch die Struktur der Sprache eingeschränkt  
→ lassen Rückschlüsse auf die zugrundeliegenden mentalen Fähigkeiten und Repräsentationen zu
  - Kontaminationen (Verschmelzungen)
  - Anakoluthe (plötzlicher Wechsel im Satzbau)  
*deswegen, weil wir im Augenblick eine große Wandlung sich vollzieht.*  
*eine große Wandlung durchmachen + eine Wandlung vollzieht sich*



## Sprachliche Beobachtungsdaten

- Sprachstörungen
  - (erworbene) Beeinträchtigungen im zentralen Nervensystem durch Hirnschaden oder psycho-emotionale Störungen
  - BROCA-Aphasie (motorische, expressive, nicht-flüssige Aphasie)
    - verlangsamer, mühsamer Sprechstil
    - weitgehende Beschränkung auf Inhaltswörter
    - vereinfachte morphologische Markierungen





## Strukturtests

- Koordinationstest:
  - Welche Teile eines Satzes lassen sich koordinativ verknüpfen?
  - Vor Mayer und seinen endlosen Reden graust es mir!*
  - Er zertritt die Raupe und schämt sich nicht mal dafür.*
  - Er bringt viele wichtige und sehr interessante Neuigkeiten mit.*
  - \* *Hans und läuft auf die Straße.*



## Strukturtests

- Verschiebeprobe:
  - Welche Teile eines Satzes lassen sich an eine andere Satzposition verschieben?
  - Fragesatzbildung, Topikalisierung
    - Susi liest die ganze Nacht.*
    - Liest Susi die ganze Nacht?*
    - Die ganze Nacht liest Susi.*



## Strukturtests

- Verschiebeprobe (2)
  - Ausklammerung
    - Uns hat der Wein, den du so toll findest, nicht geschmeckt.*
    - Uns hat der Wein nicht geschmeckt, den du so toll findest.*
    - Uns hat der Wein mit dem roten Etikett nicht geschmeckt.*
    - \* *Uns hat der Wein nicht geschmeckt mit dem roten Etikett.*



## Strukturtests

- reguläre Transformierbarkeit (Passivbildung)
  - Susi liest das ganze Buch.*
  - Susi liest die ganze Nacht.*
  - Liest Susi das ganze Buch?*
  - Liest Susi die ganze Nacht?*
  - Das ganze Buch wird von Susi gelesen.*
  - \* *Die ganze Nacht wird von Susi gelesen.*



## Syntaktische Phänomene

- Konstituenz
- Dependenz
- Valenz
- Anordnung
- Rektion/Selektion
- Kongruenz



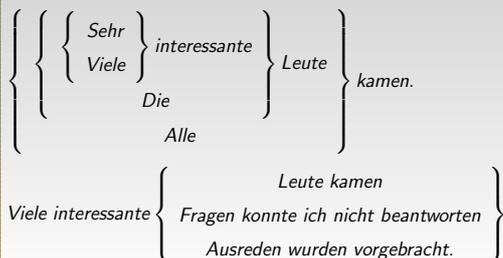
## Konstituenz

- Unmittelbare Konstituenten (immediate constituency), IC-Analyse
  - Längste Zeichenfolgen, in die man einen Satz "sinnvoll" zerlegen kann
  - vollständige Zerlegung eines Satzes durch rekursive Ermittlung der unmittelbaren Konstituenten
- Was ist eine "sinnvolle Zerlegung"?



## Konstituenz

- Ersetzungsprobe: Wortfolgen, die sich in einem gegebenen Kontext gegenseitig ersetzen können sind Kandidaten für Konstituenten



## Konstituenz

- Pronominalisierungsprobe: Wortfolgen, auf die man sich mit einem Pronomen beziehen kann, sind Konstituenten
  - Sie kamen.*
  - \* *Sie Ausreden wurden vorgebracht.*
- nicht alle Konstituenten lassen sich pronominalisieren



## Konstituenz

- Weglassprobe (II): In elliptischen Konstruktionen können nur Konstituenten weggelassen werden.

*Marie fährt [zur Arbeit] und Hans läuft zur Arbeit.*

*Marie sucht ihre [bunte Lieblingszeitung] und Hans liest seine bunte Lieblingszeitung.*



## Konstituenz

- Frageprobe: Wortfolgen, nach denen man fragen kann, sind Konstituenten.

*Wer hat die Kekse aufgegessen? – Mein Chef.*

*Wann ist Betriebsausflug? – Am Mittwoch.*

*Wo findet das Fest statt? – Bei mir.*

*Wohin ist Marie gegangen? – Zu Hans.*

*Wie geht es dir heute? – Schon besser.*

*Womit hat er das Klavier zerlegt? – Mit der Axt.*

*Was hat der liebe Gott vor der Schöpfung gemacht? – Ruten für unnütze Frager geschnitten.*

*Warum raucht Dieter so viel? – Weil er süchtig ist.*



## Konstituenz

- Koordinationstest: Was sich koordinieren lässt, ist eine Konstituente.

- relativ zuverlässig

- aber: Tilgungen in elliptische Konstruktionen

*Er mag schnelle Autos und sie Sahnetörtchen.*

*Er mag schnelle Autos und sie mag Sahnetörtchen.*

- Konstituenz als zentrale syntaktische Eigenschaft:

→ **Phrasenstrukturgrammatik**



## Dependenz

- Abhängigkeit bestimmter Wortformen von anderen im Satz.

- Weglassprobe (I):

Von Wortformen, die weglassbar sind, hängen keine weiteren Wortformen ab.

*Die Leute kamen.*

*Leute kamen.* → dep(die, nil)



## Dependenz

- Weglassprobe (I):

Ist eine Wortform *A* nur zusammen mit einer anderen Wortform *B* weglassbar, hängt *B* von *A* ab.

*Die Leute kamen.*

\* *Die kamen.* → dep(die, Leute)

*Sehr interessante Leute kamen.*

\* *Sehr Leute kamen.* → dep(sehr, interessante)

- Dependenz als zentrale syntaktische Eigenschaft:

→ **Abhängigkeitsgrammatik**



## Valenz

- Fähigkeit des Verbs, Leerstellen zu eröffnen, die durch andere Satzglieder abgesättigt werden können

→ Argumentstellen in der logischen Form

- Aktanten (Komplemente): obligatorische Ergänzungen (des Verbs)

- Subkategorisierung (der Verben)

- absolute Verben: Keine Objekte möglich (intransitiv)

*Paul schlief.*

- binäre Verben: ein Objekt erforderlich (transitiv: Akkusativ; intransitiv: Dativ, Genitiv)

*Paul betritt den Raum. / Hans dankt dem Lehrer.*



## Valenz

- Subkategorisierung (der Verben) (2)

- ternäre Verben: zwei Objekte notwendig

*Karl gab dem Mann ein Buch.*

- obligatorischer Infinitiv

*Max scheint zu kommen.*

- obligatorischer dass-Satz

*Es scheint, dass Max kommt.*

- ...

- Achtung: Terminologisches Durcheinander!

binär ~ (strikt) transitiv

ternär ~ bitransitiv



## Valenz

- fakultative Aktanten vs. zwei syntaktische Lesarten?

*Karl liest. / Karl liest ein Buch.*

- freie Adjunkte

*Paul kommt morgen.*

*Vater singt in der Badewanne.*



## Valenz

- Valenzen auch für Substantive
  - relationale Nomina
    - Vater von ...
    - Ecke von ...
  - Verbleitungen
    - Der Arzt öffnet die Wunde.
    - Die Öffnung der Wunde ...
- Valenzen lassen sich gut durch eine Abhängigkeitsgrammatik modellieren



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 73

## Anordnung

- strenge Anordnungsrestriktionen
  - innerhalb der Nominal-/Präpositionalgruppen
  - Konstituentenreihung im Englischen (S P O)
- schwache Anordnungsrestriktionen
  - Aktantenreihung im deutschen Nebensatz
  - Anordnung der freien Adjunkte
  - Partikel
- traditionell vernachlässigtes Gebiet



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 74

## Anordnung

- Einfluss der Satzperspektive
  - Topik (Thema)
    - worüber der Satz etwas aussagt
    - gegebene Information, Bezugnahme
    - Satzinitiale Stellung
  - Fokus (Rhema)
    - was der Satz über das Thema aussagt
    - neue Information
    - im Deutschen: Nachverbposition
  - Prominenzordnung
    - pragmatische Fragestellung



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 75

## Anordnung

- unmarkierte Anordnung im deutschen Aussagesatz
  - Subjekt in Topikposition
  - Objekt(e), Präpositionalgruppen, Adverbialgruppen, Nebensätze etc. im Fokus
- Topikalisierung: Bewegung einer Konstituente aus dem Fokus in die Topikposition
  - engl.: einfache Bewegung der betreffenden Konstituente
    - Nice things you've been telling me!*
  - dtsh.: Vertauschung zweier Konstituenten
    - Hübsch sieht das Mädchen aus!*



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 76

## Rektion

- Forderung nach einem morphologischen Merkmal (insbesondere Kasus) bei einem Aktanten
  - Paul betritt den Raum.* (Akkusativ)
  - Hans dankt dem Lehrer.* (Dativ)
  - Das sieht ihm wieder ähnlich!* (Dativ)
  - Marie zweifelt an seiner Glaubwürdigkeit* (an + Dativ)
  - der Anfang der Geschichte* (Genitiv)
  - die Hoffnung auf ein glückliches Ende* (auf + Akkusativ)



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 77

## Selektion

- Forderung nach einem bestimmten Merkmal bei einem Aktanten
  - \* *Die Lampe trinkt das Geld.*
- syntaktische vs. semantische Eigenschaft
  - CHOMSKY: nichtkategoriale Eigenschaft



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 78

## Kongruenz

- Übereinstimmung von Wortformen bezüglich (strukturell definierter) morphosyntaktischer Merkmale (Kasus, Genus, Numerus, Person)
- Subjekt–Verb: Person, Numerus
  - Ich stehe am Wasser.*
  - Du stehst am Wasser.*
  - ...



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 79

## Kongruenz

- Artikel–attributives Adjektiv–Substantiv: Kasus, Genus, Numerus
  - der kleine Mann*
  - (wegen) des kleinen Mannes*
  - ...
  - die kleinen Männer*
  - aber: *ein kleiner Mann*
- Relativpronomen–Bezugsnomen: Genus, Numerus
  - ... *das Mädchen, für das ...*



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 80

## Kongruenz

- Possessivpronomen–Antezedent: Genus, Person, Numerus  
*Oma hat ihre Brille verloren.*
- Reflexivpronomen–Subjekt: Person, Numerus  
*Er fühlt sich stark.*



## Strukturbeschreibungen

- syntaktische Strukturbeziehungen sind transitiv und antisymmetrisch

$\text{constituent\_of}(X, Y)$   
 $\text{depends\_of}(X, Y)$

- strikte partielle Ordnung
- erfüllt zentrale Forderung an eine (syntaktische) Repräsentation:

Kompositionalität:  $B(K') \subset B(K) \Leftrightarrow K' \subset K$

Beschreibung von Teilkomponenten soll Teil der Beschreibung für die Gesamtkomponente sein.



## Strukturbeschreibungen

- Modellierung der Konstituenz: Phrasenstrukturbäume

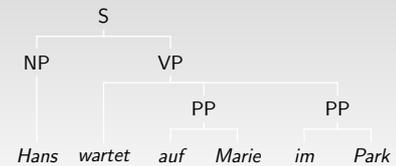
Hans wartet auf Marie im Park.					
Hans	wartet auf Marie im Park.				
	wartet	auf Marie		im Park.	
		auf	Marie	im	Park.



## Strukturbeschreibungen

- Beschreibung durch eine kontextfreie Grammatik

- Zuweisung von Kategoriennamen zu den Knoten des Phrasenstrukturbau
- Interpretation als Konstituententypen



## Strukturbeschreibungen

- Grundsatz: gleiche Distribution – gleicher Kategoriename
- alternative Bezeichnung: P-Marker
- Basisrelationen: Dominanz, Präzedenz von Knoten
- alternative Darstellung als lineare Klammerstruktur

$[S [NP Hans] [VP wartet [PP auf [NP Marie] ] ] ]$   
 $[PP im [NP Wald] ] ] ]$



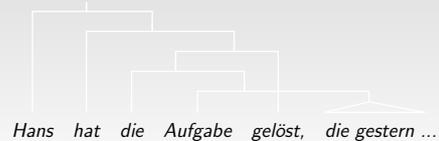
## Strukturbeschreibungen

- Beschränkung auf projektiv geordnete Bäume

- Probleme:

- diskontinuierliche Konstituenten

*Hans hat die Aufgabe gelöst, die gestern an der Tafel stand.*



## Strukturbeschreibungen

- Probleme (2):

- abtrennbare Verbzusätze

*Ich rufe meinen Vater an.*

*Weil ich meinen Vater anrufe, ...*

abtrennbare Verbzusätze sind koordinierbar!

*Er macht das Fenster jeden Morgen auf und erst am Abend wieder zu.*

→ selbständige Konstituenten



## Strukturbeschreibungen

- Probleme (3):

- Ellipsen (shared constituent coordination)

*Hans kauft und verkauft Autos.*

*Hans kauft und [verkauft Autos].*

*Hans [kauft ...] und verkauft [... Autos].*



## Strukturbeschreibungen

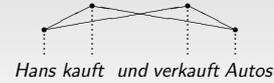
- Modellierung der Dependenz: Abhängigkeitsbäume
- Unterordnung von Wortformen unter Wortformen (bzw. Teilbäumen unter Wortformen)



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 89

## Strukturbeschreibungen

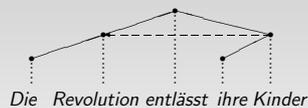
- Vereinfachung: Beschränkung auf eindeutige Unterordnungsrelationen
- hierarchische Gliederung: Bäume
- Probleme
  - Syntax: elliptische Konstruktionen bei der Koordination



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 90

## Strukturbeschreibungen

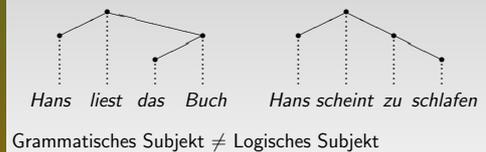
- Probleme
  - Semantik: Anaphora



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 91

## Strukturbeschreibungen

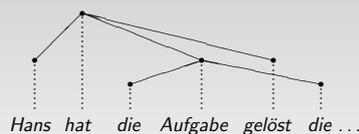
- Probleme (3)
  - unterschiedliche Gliederungen für Syntax und Semantik (Raising-Verben)



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 92

## Strukturbeschreibungen

- Vereinfachung: Projektivität
  - Problem: diskontinuierliche Konstituenten



- projektive Dependenzstrukturen induzieren eine Konstituentenstruktur,
- weisen aber keine Konstituentenkategorien zu



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 93

## Adäquatheit

- Beobachtungsadäquatheit
  - Spezifikation aller wohlgeformten Äußerungen einer Sprache
    - formal explizit
    - vollständig und korrekt
    - Semantik vernachlässigt
    - CHOMSKY 1957



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 94

## Adäquatheit

- Beschreibungsadäquatheit
  - zusätzlich: Spezifikation von Strukturbeschreibungen, die in grundsätzlicher Weise mit den Intuitionen eines Sprechers der Sprache übereinstimmen
    - Verbindung von sprachlichen Sequenzen mit Bedeutungen
    - CHOMSKY 1965



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 95

## Adäquatheit

- Erklärungsadäquatheit
  - zusätzlich: Spezifikation möglichst weniger, sprachuniverseller Prinzipien, die psychologisch plausible Verarbeitungsmechanismen des Menschen repräsentieren
    - Erlernbarkeit von Sprache
    - CHOMSKY 1981
  - Verarbeitungsadäquatheit
    - effiziente Umsetzung in Verstehens- und Generierungsprozesse
    - deklarativ vs. derivational



Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 96