

## Sprachorientierte KI: Syntax und Parsing

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



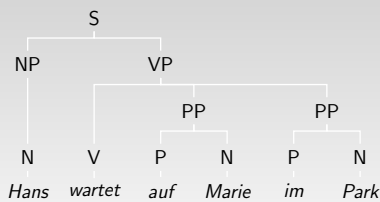
## Phrasenstrukturgrammatiken

- Kontextfreie Grammatiken
- Phrasale Kategorien
- ID/LP-Regeln
- Relationale Kategorien
- Subkategorisierung
- Rektion und Kongruenz
- Thematische Rollen
- Adäquatheit



## Kontextfreie Grammatiken

- projektive Phrasenstrukturbäume (P-Marker) können durch kontextfreie Grammatiken erzeugt werden



## Kontextfreie Grammatiken

- Regelstruktur
  - Nichtterminalsymbole: S, NP, VP, PP, ...
  - Terminalsymbole: wartet, auf, im, Hans, Marie, Park
  - NT-Symbol  $\rightarrow$  {T-Symbol | NT-Symbol}\*
- Regelverwendung
  - generativ
  - analytisch



## Kontextfreie Grammatiken

- Phrasenstrukturbaum ist Ableitungsprodukt der Grammatikanwendung
  - $\rightarrow$  enger Zusammenhang zwischen
    - Regelstruktur
    - Strukturbeschreibung
    - Regelverwendung (Analyse/Generierung)
- Regeln lassen sich aus Phrasenstrukturbaum extrahieren



## Kontextfreie Grammatiken

- lexikalische Einsetzungsregeln, präterminale Regeln, Lexikon
  - N  $\rightarrow$  Marie
  - N  $\rightarrow$  Hans
  - N  $\rightarrow$  Park
  - P  $\rightarrow$  auf
  - P  $\rightarrow$  im
  - V  $\rightarrow$  wartet



## Kontextfreie Grammatiken

- strukturbildende Regeln, Grammatik
  - S  $\rightarrow$  NP VP
  - VP  $\rightarrow$  V PP
  - VP  $\rightarrow$  V PP PP
  - PP  $\rightarrow$  P NP
  - NP  $\rightarrow$  N
- erste Beschränkung für die Form der Regeln
  - Lexikon
  - PT-Symbol  $\rightarrow$  T-Symbol
  - Grammatik
  - NT-Symbol  $\rightarrow$  {NT-Symbol | PT-Symbol}\*



## Kontextfreie Grammatiken

- rekursive Regeln: potentiell unendlich viele Sätze können generiert werden
  - $\rightarrow$  Kreativität der Sprachkompetenz
- (noch) keine weiteren Beschränkungen für die Form der Regeln
  - $\rightarrow$  Ziel linguistischer Modellierung!



## Phrasale Kategorien

- strukturelle Perspektive: Kategorie ist distributiver Typ
- Phrasale Kategorien entstehen aus lexikalischen Kategorien durch Ergänzen mit zusätzlichen Konstituenten
  - N ⇒ NP
  - V ⇒ VP
  - A ⇒ AP
  - ADV ⇒ ADVP
  - P ⇒ PP



## Phrasale Kategorien

- Kriterien für phrasale Kategorien (RADFORD 1988)
1. Morphologische Evidenz
    - Phrasale Flexion im Englischen (nur für Nominalgruppen) possessiver Genitiv
      - This crown is the [N king]'s.*
      - This crown is [NP the king]'s.*
      - This crown is [NP the [N king] of England]'s.*
      - \* *This crown is [NP the [N king]'s of England].*
      - \* *This crown is [AP very handsome]'s.*



## Phrasale Kategorien

### 2. Semantische Evidenz

- Erklärung struktureller Mehrdeutigkeiten durch unterschiedliche Kategoriezugehörigkeit
- z.B. Skopusambiguität
  - The President could not ratify the treaty.*
  - The President [M could not] ratify the treaty.*
  - The President could [VP not ratify the treaty].*
  - The President [M simply could not] ratify the treaty.*
  - The President could [VP simply not ratify the treaty].*

Erklärung erfordert Rückgriff auf phrasale Kategorien, z.B. VP



## Phrasale Kategorien

### 3. Phonologische Evidenz

- bei phonologischer Kontraktion entfällt die Mehrdeutigkeit
  - The President couldn't ratify the treaty.*
  - The President [M couldn't] ratify the treaty.*
  - \* *The President could[VP n't ratify the treaty].*



## Phrasale Kategorien

### 4. Syntaktische Evidenz

- Distributionskriterien
  - Verschiebeprobe
  - Frageprobe
- Koordinationstest
  - Nur Konstituenten mit der gleichen Kategorie können koordinativ verknüpft werden
    - Hans schrieb an Marie und an Susi.*
    - Hans schrieb einen Brief und eine Postkarte.*
    - \* *Hans schrieb an Marie und eine Postkarte.*
    - ? *Hans schrieb einen Brief und an Susi.*

- Pronominalisierung



## Phrasale Kategorien

- Behandlung von Einzelwortphrasen
  - Ergänzungen einer lexikalischen zu einer phrasalen Kategorie können auch leer sein
  - keine Verdopplung von Phrasenstrukturen
    - [NP Der Mann] lacht über diesen Witz.*
    - [N Hans] lacht über diesen Witz.*
    - S → NP VP
    - S → N VP
    - [NP [N Hans]] lacht über diesen Witz.*



## Phrasale Kategorien

- Behandlung von Einzelwortphrasen (2)
  - gleiches distributionelles Verhalten (Ersetzungsprobe, Fragetest, Pronominalisierungsprobe)
    - Die dicken Männer lachen über diesen Witz.*
    - Männer lachen über diesen Witz.*
    - Wer lacht über diesen Witz? - Die dicken Männer. / Männer.*
    - Sie lachen über diesen Witz.*
  - analog für Verb-, Adjektiv- und Adverbialgruppen



## ID/LP-Regeln

- anordnungsfreie CFG-Regeln (immediate dominance)
  - S → {NP,VP}
- zusätzliche Anordnungsrestriktionen durch Abfolgebedingungen (linear precedence)
  - NP < VP
- LP-Constraints meist auf der Basis von Merkmalen für komplexe Kategorien formuliert
  - N+ < V+
- Verwendung in der Generalized Phrase-Structure Grammar (GPSG, GAZDAR ET AL. 1985)



## Relationale Kategorien

- reine Dependenzrelationen
  - depends\_of( $X, Y$ )
- verfeinert: funktionale Beschreibungen (Unterordnungsrelationen)
  - subject\_of( $X, Y$ ), object\_of( $X, Y$ ), ...
- Subjekt: **Hans singt.**
- direktes Objekt: **Hans trägt die Tasche.**
- indirektes Objekt: **Hans gibt der Mutter die Tasche.**
- Präpositionalobjekt: **Hans glaubt an das Gute im Menschen.**



## Relationale Kategorien

- andere syntaktische Perspektive: Valenzorientierte Beschreibung
- Abbildung auf semantische Funktionen (Tiefenkasus): Agens, Rezipient, Patiens, Benefizient, ...



## Relationale Kategorien

- CHOMSKY 1965, 1981: funktionale Information ist redundant
- Subjekt: NP, die unmittelbar vom Satzknotten dominiert wird
- direktes Objekt: von den unmittelbar von einer VP dominierten NPs, die am weitesten links angeordnete
- indirektes Objekt: die zweite, der unmittelbar von einer VP dominierten NPs
- Präpositionalobjekt: die unmittelbar von einer PP dominierte NP



## Relationale Kategorien

- Differenzierungsprinzip für Unterordnungsrelationen
- Zwei Teilbäume stehen in der gleichen Unterordnungsrelation zu einem Baumkontext wenn ihre wechselseitige Substitution keine ungrammatischen Bildungen zur Folge hat.
- rein distributionelle Definition!



## Subkategorisierung

- Regeln einer Phrasenstrukturgrammatik ergeben sich direkt aus dem Phrasenstrukturbaum, z.B.
  - $S \rightarrow NP VP$
  - $VP \rightarrow V NP PP$
  - $PP \rightarrow P NP$
- Beseitigung von Redundanz
  - $VP \rightarrow V NP$
  - $VP \rightarrow V$
- optionale Phrasenbestandteile
  - $VP \rightarrow V (NP)$



## Subkategorisierung

- Stelligkeit der Verben
  - intransitiv: *schlafen*,
  - transitiv: *tragen, beruhigen*
  - bitransitiv: *geben, verkaufen*
  - dass-Satz: *wissen, vermuten, glauben, denken* (Verben der propositionalen Einstellung)
- Subkategorisierung: Unterscheidung bezüglich der Anzahl und der Kategorie von Ergänzungen (Komplementen) für eine lexikalische Kategorie



## Subkategorisierung

- feinere Unterteilung des Kategorieninventars
  - $V_i, V_t, V_{bt}, V_{dass}$
- Generalisierungen gehen verloren z.B. Modifikation durch Adverbien
  - ... weil er viel schläft.
  - ... weil er den Koffer ungern trägt.
  - ... weil er ihr die Bücher günstig verkauft.
  - ... weil er sicher weiß, dass ...
- keine allgemeine Regel mehr
  - $V \rightarrow Adv V$



## Subkategorisierung

- sondern mehrere partikuläre Regeln
  - $V_i \rightarrow Adv V_i$
  - $V_t \rightarrow Adv V_t$
  - $V_{bt} \rightarrow Adv V_{bt}$
  - $V_{dass} \rightarrow Adv V_{dass}$
  - ...
- einfache distributionelle Fakten können nicht in kontextfreien Regeln dargestellt werden



## Subkategorisierung

- Alternative: Subkategorisierungsrahmen
  - sehr allgemeine Phrasenstrukturregeln  
 $VP \rightarrow V (NP) (NP) (PP) (S)$
  - Subkategorisierungsrahmen als zusätzliche Beschränkung für zulässige Phrasenstrukturbäume
    - schlafen:* \_\_\_
    - tragen:* \_\_\_ NP
    - geben:* \_\_\_ NP NP
    - wissen:* \_\_\_ S
  - werden im Wörterbuch kodiert
  - lizenzieren eine bestimmte syntaktische Umgebung



## Subkategorisierung

- Lexikalisierungsprinzip:  
Jede präterminale Kategorie kann zu einem Terminalsymbol expandiert werden, das
  - die geforderte Kategorie besitzt und
  - so subkategorisiert ist, dass es die Komplemente seiner syntaktischen Umgebung lizenziert



## Rektion und Kongruenz

- Verben bestimmen nicht nur Anzahl und Art ihrer Komplemente zusätzlich auch deren Kasus (bzw. die Präposition von Präpositionalobjekten)
  - danken:* Dativ
  - kennen:* Akkusativ
  - gedenken:* Genitiv
  - hoffen:* auf (+ Akkusativ)
  - verkaufen:* Akkusativ, Dativ



## Subkategorisierung

- Rektion
  - Modellierung durch Kategoriedifferenzierung in den Subkategorisierungsrahmen
    - tragen:* \_\_\_ NP<sub>akk</sub>
    - drohen:* \_\_\_ NP<sub>dat</sub>
    - gedenken:* \_\_\_ NP<sub>gen</sub>
    - hoffen:* \_\_\_ PP<sub>auf</sub>
    - verkaufen:* \_\_\_ NP<sub>akk</sub> NP<sub>dat</sub>



## Rektion und Kongruenz

- Rektion (2)
  - Problem: Kombinatorik bei der Beschreibung komplexer Nominalgruppen
    - NP<sub>akk</sub> → NP<sub>gen</sub> NP<sub>akk</sub> *Er trägt Marias Taschen.*
    - NP<sub>akk</sub> → NP<sub>akk</sub> PP *Er trägt die Taschen mit den Streifen.*
    - NP<sub>akk</sub> → NP<sub>akk</sub> NP<sub>gen</sub> *Er trägt die Taschen der Kinder.*
    - NP<sub>akk</sub> → NP<sub>akk</sub> S *Er trägt die Taschen, die hier standen.*
    - ...
    - NP<sub>dat</sub> → NP<sub>gen</sub> NP<sub>dat</sub> *Er droht Marias Töchtern.*
    - NP<sub>dat</sub> → NP<sub>dat</sub> PP *Er droht dem Mädchen mit den Locken.*
    - NP<sub>dat</sub> → NP<sub>dat</sub> NP<sub>gen</sub> *Er droht den Töchtern der Nachbarin.*
    - NP<sub>dat</sub> → NP<sub>dat</sub> S *Er droht den Mädchen, die dort spielen.*
    - ...



## Rektion und Kongruenz

- Kongruenz
  - Subjekt-Prädikat: Numerus, Person
  - Modellierung durch Kategoriedifferenzierung in den Regeln
    - $S \rightarrow NP_{nom,sg,1} VP_{sg,1}$
    - $S \rightarrow NP_{nom,sg,2} VP_{sg,2}$
    - $S \rightarrow NP_{nom,sg,3} VP_{sg,3}$
    - $S \rightarrow NP_{nom,pl,1} VP_{pl,1}$
    - $S \rightarrow NP_{nom,pl,2} VP_{pl,2}$
    - $S \rightarrow NP_{nom,pl,3} VP_{pl,3}$
    - ...



## Rektion und Kongruenz

- Kombinatorik
- zusätzlich: Fortpflanzung der Differenzierung auf weitere Regeln
  - NP<sub>nom,sg,3</sub> → D<sub>nom,sg</sub> N<sub>nom,sg</sub>
  - ...
  - VP<sub>sg,1</sub> → V<sub>sg,1</sub> (NP) (NP) (PP) (S)
  - ...
- Explosionsartige Vermehrung der Regeln



## Rektion und Kongruenz

- Verlust von Generalisierungen
  - systematische Beziehungen gehen verloren  
z.B. zwischen NP<sub>nom,sg,1</sub> und NP<sub>nom,pl,3</sub>
  - Kategorien sind atomare Symbole, innere Struktur hat keinerlei Bedeutung innerhalb der Grammatik
- kontextfreie Grammatiken sind deskriptiv inadäquat
  - flexionsmorphologische Fakten aus der Grammatik ausschließen
  - Merkmalsbasierte Unifikationsgrammatiken



## Thematische Rollen

- Interface zur Semantik
  - Verben ~ Prädikate
    - jemand schläft ⇒ schlafen(X)
    - jemand trägt etwas ⇒ tragen(X, Y)
  - Valenzen ~ Argumentstellen
  - Argumentstellen werden durch thematische Rollen (Θ-Rollen) beschrieben
  - Stelligkeit eines semantischen Prädikats: Θ-Raster



## Thematische Rollen

- Theta-Kriterium: eindeutige Beziehungen zwischen der syntaktischen Struktur und den thematischen Rollen
- prinzipienbasierte Zuordnung
  - ein intransitives Verb weist nur seiner Subjektsposition eine Θ-Rolle zu
    - Hier schläft Hans.*
    - \* *Hier schläft.*
  - bei der Passivierung wird die Zuweisung einer thematischen Rolle an das Subjekt blockiert
    - Hier wird geschlafen.*
    - \* *Hier wird Hans geschlafen.*



## Thematische Rollen

- Thematische Rollen

THEME (patiens)	Entität, auf die der Effekt einer Aktion ausgeübt wird <i>Marie fiel hin.</i>
AGENT (actor)	Verursacher einer Aktion <i>Hans traf das Ziel.</i>
EXPERIENCER	Entität, die einen bestimmten (psychologischen) Zustand einnimmt <i>Marie ist glücklich.</i>
BENEFACTIVE	Entität, die eine Auswirkung einer Aktion erfährt <i>Marie umarmte Hans.</i> <i>Hans fiel der Hammer auf den Fuß.</i>




## Thematische Rollen

- Thematische Rollen

INSTRUMENT	Hilfsmittel für das Ausführen einer Aktion <i>Hans aß den Salat mit der Gabel.</i>
LOCATIVE	Ort an dem sich etwas befindet, oder an dem eine Aktion stattfindet <i>Marie suchte die Pralinen im Schrank.</i>
GOAL	Entität zu der sich etwas hinbewegt <i>Er legte das Schreiben in die Ablage.</i>
SOURCE	Entität von der sich etwas wegbewegt <i>Hans kommt aus der Schule.</i>




## Thematische Rollen

- selektionale Einschränkungen für die Zuweisung der Θ-Rollen
  - Der Vater singt ein Lied.*
  - ? *Der Stuhl singt ein Lied.*
  - ? *Der Vater singt eine Tasse.*

verkaufen[agent(human), theme, goal(human)]  
diskutieren[agent(human), theme(abstract)]  
schlafen[agent(animate)]

- aber:
  - übertragene Bedeutungen
  - Granularität der Beschreibungen
    - z.B. THEME bei *essen, singen, auffädeln, ...*



## Thematische Rollen

- Modellierung der Ähnlichkeit zwischen verschiedenen syntaktischen Verwendungen eines Wortes
  - transitiv: *Hans rollte den Ball den Berg hinunter.*
  - ergativ: *Der Ball rollte den Berg hinunter.*
- Objekt wird bei intrans. (ergativen) Verwendung zum Subjekt
- Zuweisung der gleichen thematischen Rolle: THEME
- gestützt durch gleiche selektionale Distribution
  - Hans rollte den Ball / den Stein / den Käse / ? die Ansprache / ? die Dämmerung den Berg hinunter.*
  - Der Ball / der Stein / der Käse / ? die Ansprache / ? die Dämmerung rollte den Berg hinunter.*



## Thematische Rollen

- Zuweisung der thematischen Rollen ist unabhängig von der syntaktischen Strukturierung
  - Die Vase zerbrach die Fensterscheibe.*
  - Die Vase zerbrach.*
- gleiche syntaktische Struktur, aber INSTRUMENT in der transitiven Verwendung gegenüber THEME in der ergativen Verwendung
- unterschiedlicher thematischer Status führt zu unterschiedlichen selektionalen Beschränkungen

*Die Vase / der Fußball / der Knall zerbrach die Fensterscheibe.*

*Die Vase / ? der Fußball / ? der Knall zerbrach.*



## Adäquatheit

- Erzeugungsadäquatheit
- Beschreibungsadäquatheit
- Erklärungsadäquatheit
- Verarbeitungsadäquatheit



## Erzeugungsadäquatheit

- sprachliche Abdeckung / Übergeneralisierung
- prinzipiell ausreichend
- nur ganz seltene Phänomene mit "milder" Kontextsensitivität
- Schwyzerdeutscher Dialekt der Züricher Region (Huybregts 1985, Shieber 1985)

*Claudia beobachtete Helmut ließ Eva helfen Hans machte Ulrike arbeitete.*

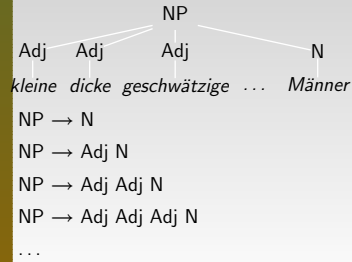
*Claudia Helmut Eva Hans Ulrike beobachtete ließ helfen machte arbeitete.*

- überlagert mit Kasusrektion:  $a^m b^n c^m d^n$  mit  $n > 0$



## Beschreibungsadäquatheit

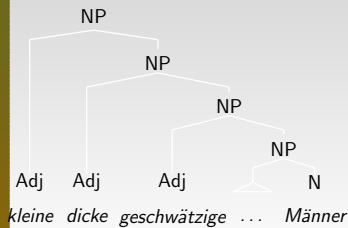
- Repräsentationsprobleme
- (unendliche) Adjektivreihenungen



## Beschreibungsadäquatheit

→ rekursive Regelstruktur mit Hilfsknoten

- NP → N
- NP → Adj NP



## Beschreibungsadäquatheit

- suggestiert ein Drominanzverhältnis, das nicht gegeben ist
- ähnliche Probleme bei der Behandlung freier Adjunkte
- Problem tritt bei der Dependenzgrammatik in analoger Weise auf



## Beschreibungsadäquatheit

- nichtprojektive Strukturen
  - abtrennbare Verbzusätze
  - elliptische Konstruktionen bei der Koordination
  - diskontinuierliche Konstituenten
    - Extraposition
    - Fragesatzbildung
    - Topikalisierung



## Beschreibungsadäquatheit

- Hauptproblem bei diskontinuierlichen Konstituenten: fehlende Parallelität bei strukturell verwandten Konstruktionen
 

*Uns hat der Wein, den du so toll findest, nicht geschmeckt.*

*Uns hat der Wein nicht geschmeckt, den du so toll findest.*
- Lösungsmöglichkeiten
  - Transformationen (derivationaler Ansatz)
  - Slash-Merkmal in Unifikationsgrammatiken (nichtderivationaler Ansatz)
- Kombinatorik und Generalisierungen bei flexionsmorphologischen Phänomenen



## Erklärungsadäquatheit

- Lernbarkeit
  - Unterbestimmtheit des sprachlichen Wissens durch die verfügbare sprachliche Evidenz
  - poverty of stimulus (FODOR 1983)
- quantitative Unterbestimmtheit
  - Sprachausschnitt, aus dem das Kind Hypothesen über Sprachstrukturen gewinnen muss ist sehr begrenzt
  - nur ein geringer Teil der möglichen Strukturen wird dem Kind angeboten



## Erklärungsadäquatheit

- qualitative Unterbestimmtheit
  - Erlernen von Regeln erfordert ausreichend positive und negative Evidenz
  - Elernte Regeln müssen auf ihre Adäquatheit geprüft werden
  - Erlernen von Regeln ist aufwendig und lässt die Kommunikation in den Hintergrund treten:

"Es ist jedoch hinreichend bekannt, dass Kinder Äußerungen produzieren, um ihre Gedanken und Wünsche mitzuteilen und nicht, um syntaktische Regeln auszuprobieren." (FANSELOW UND FELIX 1987)



## Erklärungsadäquatheit

- insbesondere negative Evidenz ist unzureichend
  - Nichtauftreten einer Konstruktion  $\neq$  Ungrammatikalität
  - Eltern korrigieren ihre Kinder selten und unsystematisch wenn, dann vorrangig bezüglich inhaltlicher Aspekte
  - systematische Korrekturversuche bringen nur wenig Vorteile für den Grammatikerwerb



## Erklärungsadäquatheit

- Sprache wird nicht gelernt, sondern erworben
  - Language acquisition device als mentales Organ
  - Universalgrammatik mit universalgrammatischen Prinzipien
  - beim Spracherwerb werden nur einzelsprachliche Varianten (Parameter) gelernt



## Erklärungsadäquatheit

- Strukturelle Eigenschaften der Phrasenstrukturgrammatik erklären nicht, was eine natürliche Sprache ist
- wenn es möglich sein sollte, beliebige rekursive Regelsysteme zu erlernen, dann müsste, dies auf beliebige sequentielle Zeichenanordnungen übertragbar sein
- Phrasenstrukturgrammatiken sind für sich alleine noch keine Theorie der Sprachstruktur  
keinerlei Bezug zum Sprachvermögen des Menschen



## Verarbeitungsadäquatheit

- einfacher Formalismus
- deklarativer Formalismus
- relativ effiziente Analysealgorithmen

