

---

# Probabilistische kontextfreie Grammatiken und Parsing

---

Sebastian Pado  
18.01.2005

---

## Das Ziel

- Freien Text parsen („Robust Parsing“)
    - Freier Dialog
    - Große Textmengen (Internet)
  - Dazu notwendig:
    1. Große und flexible Grammatik
    2. Effizientes Parsing-Verfahren
      - Ermittelt **alle** möglichen Analysen eines Satzes
    3. Desambiguierungsverfahren
      - Ermittelt die **wahrscheinlichste** Analyse
-

---

# Die Rollen von Baumbanken

- Erzeugung einer grossen Grammatik
  - Von Hand schreiben ☹
  - Stattdessen: Maschinelles Lernen aus Baumbank
- Baumbank: Sammlung von Sätzen mit syntaktischer Analyse (Parse-Baum)
  - Manuell hergestellt (oder semi-automatisch mit manueller Prüfung)
    - Hohe Genauigkeit
- Bekannte **kontextfreie** Baumbanken
  - Englisch: WSJ (Wall Street Journal / Penn Treebank), 1M Worte
  - Deutsch: TIGER (Frankfurter Rundschau), derzeit 0.7M Worte

---

Einführung in die Coli WS 04/05

---

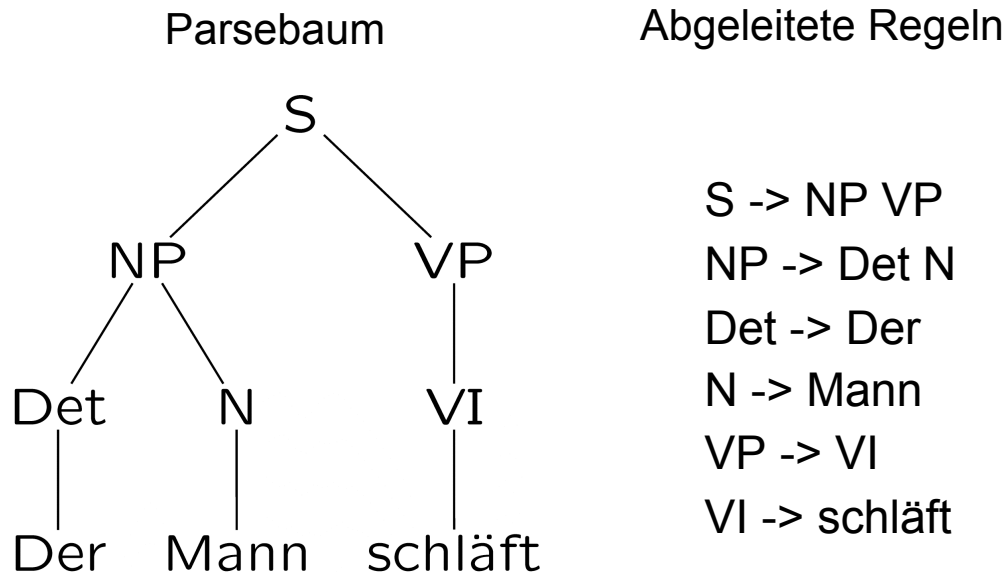
# Extraktion einer Baumbank-Grammatik

- In Baumbanken steckt implizites linguistisches Wissen
- **Naiver** Extraktionsalgorithmus einer kontextfreien Baumbank-Grammatik:
  - Initialisierung: „leere Grammatik“
  - Eine (kontextfreie) Baumbank nehmen
  - Alle Parsebäume in „lokale Bäume“ aufteilen
  - Jeden lokalen Baum als Regel zur Grammatik hinzufügen
- Vorteile gegenüber direktem Grammatikschreiben?

---

Einführung in die Coli WS 04/05

# Naive Baumbank-Grammatiken



Einführung in die Coli WS 04/05

## Probleme von Baumbank-Grammatiken

Einführung in die Coli WS 04/05

---

# 1. Abdeckung und seltene Regeln

- Es gibt viele seltene Konstruktionen
  - Bessere Abdeckung = mehr Regeln = mehr Mehrdeutigkeit
- Beispiel:
  - Kleine Gruppe von Verben nimmt viele PP-Objekte
    - Bewegungsverben
    - Ich gehe [über die Brücke] [mit Peter] [über den Fluss]
  - Grammatik enthält Regel VP -> V PP PP PP

---

Einführung in die Coli WS 04/05

---

## Übergenerierung

- Kombinationen von seltenen Regeln führt zu unerwünschten Analysen (**Übergenerierung**)
  - Eigentlich eindeutige Sätze werden mehrdeutig
    - Ich stehe auf der Brücke mit dem Schal über die Schulter
    - VP -> V PP PP; PP -> PP PP (gewollte Analyse)
      - Ich stehe [auf der Brücke] [mit dem Schal] [über die Schulter]
    - VP -> V PP PP PP (ungewollte Analyse)
      - \*Ich stehe [auf der Brücke] [mit dem Schal] [über die Schulter]
  - Ungrammatische Sätze erhalten Analysen
    - VP -> V Partikel
      - Peter sagt zu / ab
      - \*Peter sagt herab/...

---

Einführung in die Coli WS 04/05

# Abdeckung und Baumbank-Grammatiken

- Alternativen:
  - Seltene Phänomene ignorieren
    - Weniger Abdeckung, weniger Übergenerierung
  - Seltene Phänomene behandeln
    - Mehr Abdeckung, mehr Übergenerierung
- Baumbanken: Behandlung aller Phänomene
  - Jeder Satz erhält eine Analyse
    - Zeitungstext: komplexer Satzbau
  - Resultat: Riesige Grammatiken
    - Starke Übergenerierung
    - Sätze mit 40 Wörtern: bis zu 100 000 Lesarten
  - Verarbeitungsproblem
  - Interpretationsproblem

Einführung in die Coli WS 04/05

## 2. Form der Baumbank-Regeln

Gelernte Regeln sind vom **Annotationsschema** der annotierten Baumbank abhängig

- Tiefere Bäume
  - mehr Nonterminale
  - weniger Regeln
- Flachere Bäume
  - weniger Nonterminale
  - mehr Regeln

Kontextfreie Baumbank-Grammatiken sind **sehr flach**

Einführung in die Coli WS 04/05

---

# Tiefe und flache Grammatiken

- Flache Regeln: „kombinatorische Explosion“
  - NP: Det N (der Mann) oder N (Häuser)
  - VP: V (intran.) oder V NP (trans.) oder V NP NP (ditrans.)

## Traditionelle Grammatik

VP -> V

VP -> V NP

VP -> V NP NP

NP -> Det N

NP -> N

## Baumbank-Grammatik

VP -> V

VP -> V Det N

VP -> V N

VP -> V Det N Det N

VP -> V Det N N

VP -> V N Det N

VP -> V N N

---

Einführung in die Coli WS 04/05

---

# Grammatikgröße und linguistische Relevanz

- Kleine Grammatik = besseres Modell
  - Gegenbeispiel: Aufzählungsgrammatik
    - keine Generalisierung (Vorhersage für neue Sätze)
  - Je kleiner, desto bessere Generalisierung
  - Occam's Razor
- Baumbank-Grammatiken: viele oder wenige Regeln?
  - 19.000 Regeln für 18.000 Sätze NEGRA-Korpus
    - (Vorläufer von TIGER)
  - Fast alle Regeln kommen nur in einem Satz vor
    - Was heisst das für neue Sätze (Generalisierung)?

---

Einführung in die Coli WS 04/05

---

### 3. Grammatikalität

- **Naive** Baumbank-Grammatiken haben kein Konzept von **gradueller** Grammatikalität
  - Vgl. letzte Vorlesung
  - Keine Möglichkeit, von der „besten Analyse zu sprechen“
  - Besonders schlimm wegen hoher Mehrdeutigkeit
    - Grosse Anzahl an Analysen

---

Einführung in die Coli WS 04/05

---

### Gegenentwurf: Probabilistische kontextfreie Grammatiken (PCFGs)

---

Einführung in die Coli WS 04/05

---

# Probabilistische Grammatiken (PCFGs)

- Jede Regel erhält eine Wahrscheinlichkeit
  - $S \rightarrow NP VP$  (0.6)
  - $S \rightarrow NP VP PP$  (0.4)
    - „60% aller Sätze haben die Form NP VP, 40% die Form NP VP PP“
- Wahrscheinlichkeiten für Parsebäume
  - Berechnung aus Regelwahrscheinlichkeiten
- Vorteile:
  - Graduelle Grammatikalität
    - Beste Analyse = wahrscheinlichster Parsebaum
  - Einschränkung des Suchraumes
    - Verfolge vor allem „wahrscheinliche“ Parsebäume

---

Einführung in die Coli WS 04/05

---

## Parsing mit PCFGs

1. **Lerne Regeln mit Wahrscheinlichkeiten**
2. Berechne Wahrscheinlichkeiten für Bäume aus Regelwahrscheinlichkeiten
3. Finde effizientes Verfahren, um wahrscheinlichste Bäume zu konstruieren

---

Einführung in die Coli WS 04/05



## Schritt 1:

### Regelwahrscheinlichkeiten lernen

- **Nicht naive** Baumbank-Grammatik
  - Baumbank nehmen
  - Leere Grammatik initialisieren
  - Alle Parsebäume in „lokale Bäume“ aufteilen
  - Jeden neuen lokalen Baum als Regel hinzufügen
    - Zähle Häufigkeit von Regeln
  - Terminologie:
    - Linke Seite (LS): Nonterminal
    - Rechte Seite (RS): Expansion

LS -> RS.....	freq
S -> NP VP	50
S -> S PP	30

Einführung in die Coli WS 04/05

## Regelwahrscheinlichkeiten lernen (II)

- Wahrscheinlichkeit, dass Nonterminal NT mit Expansion E expandiert wird, ist Anteil von E an allen Auftreten von NT
  - Bedingte Wahrscheinlichkeit (vgl. letzte Vorlesung)

$$P(RS|LS) = \frac{f(LS \rightarrow RS)}{f(LS)} = \frac{f(LS \rightarrow RS)}{\sum_{RS} f(LS \rightarrow RS)}$$

- Beispiel:  $f(S \rightarrow NP VP) = 50$ ,  $f(S \rightarrow S PP) = 30$ 
  - $P(NP VP | S) = 50 / (50+30) = 5/8$
  - $P(S PP | S) = 30 / (50+30) = 3/8$

Einführung in die Coli WS 04/05

## „Echte“ Regelfrequenzen in NEGRA

...		\$,: Komma
1	NP ADJA \$*LRB* \$, ADJA NN	ADJA: Attributives Adjektiv
12	NP ADJA \$*LRB* NN	CARD: Kardinalzahl
1	NP ADJA \$*LRB* NN \$*LRB* PP	CNP: Koordinierte NP
2	NP ADJA \$*LRB* NN PP	\$*LRB*: Klammer
8	NP ADJA \$, ADJA NN	NE: Eigennamen
1	NP ADJA \$, ADJA NN PP	NN: Normales Nomen
3	NP ADJA \$, AP NN	NP: Nominalphrase
1	NP ADJA ADJA ADJA NN	PP: Präpositionalphrase
1	NP ADJA ADJA CARD	
2	NP ADJA ADJA CNP	
1	NP ADJA ADJA NN \$*LRB* NE	Wenn dies alle NP-Regeln wären,
1	NP ADJA ADJA NN \$*LRB* NP	dann wäre z.B.
4	NP ADJA ADJA NN NP	$P(\text{ADJA LRB NN} \mid \text{NP}) = 12/89$
51	NP ADJA ADJA NN	$P(\text{ADJA ADJA NN} \mid \text{NN}) = 51/89$
...		$P(\text{ADJA ADJA CARD} \mid \text{NP}) = 1/89$

Einführung in die Coli WS 04/05

## Parsing mit PCFGs

1. Lerne Regeln mit Wahrscheinlichkeiten
2. Berechne Wahrscheinlichkeiten für Bäume aus Regelwahrscheinlichkeiten
3. Finde effizientes Verfahren, um wahrscheinlichste Bäume zu konstruieren

Einführung in die Coli WS 04/05

---

## Schritt 2: Baumwahrscheinlichkeiten berechnen

- Baum = Kombination von Regel(anwendungen)
- $P(\text{Baum}) = P(\text{Regel}_1, \text{Regel}_2, \dots, \text{Regel}_n)$ 
  - Gemeinsame Wahrscheinlichkeit (vgl. letzte VL)
- Allgemeine Formel zur Berechnung von gemeinsamen Wahrscheinlichkeiten (Kettenregel):
  - $P(\text{Regel}_1, \text{Regel}_2, \text{Regel}_3) = P(\text{Regel}_1) * P(\text{Regel}_2 | \text{Regel}_1) * P(\text{Regel}_3 | \text{Regel}_2, \text{Regel}_1)$

Viel zu kompliziert zum Lernen

---

Einführung in die Coli WS 04/05

---

## Unabhängigkeit von Regeln

- Starke Unabhängigkeitsannahme:  
Alle Regeln sind unabhängig
- Erinnerung: Wenn A,B unabhängig,
  - dann beeinflussen sich A und B nicht
  - dann gilt  $P(A|B) = P(A)$  und  $P(B|A) = P(B)$
- $P(\text{Regel}_1, \text{Regel}_2, \text{Regel}_3)$  vereinfacht sich zu  
 $P(R_1) * P(R_2) * P(R_3) = \prod_i P(R_i)$

Einfach zu berechnen --  
aber ist die Annahme gerechtfertigt?

---

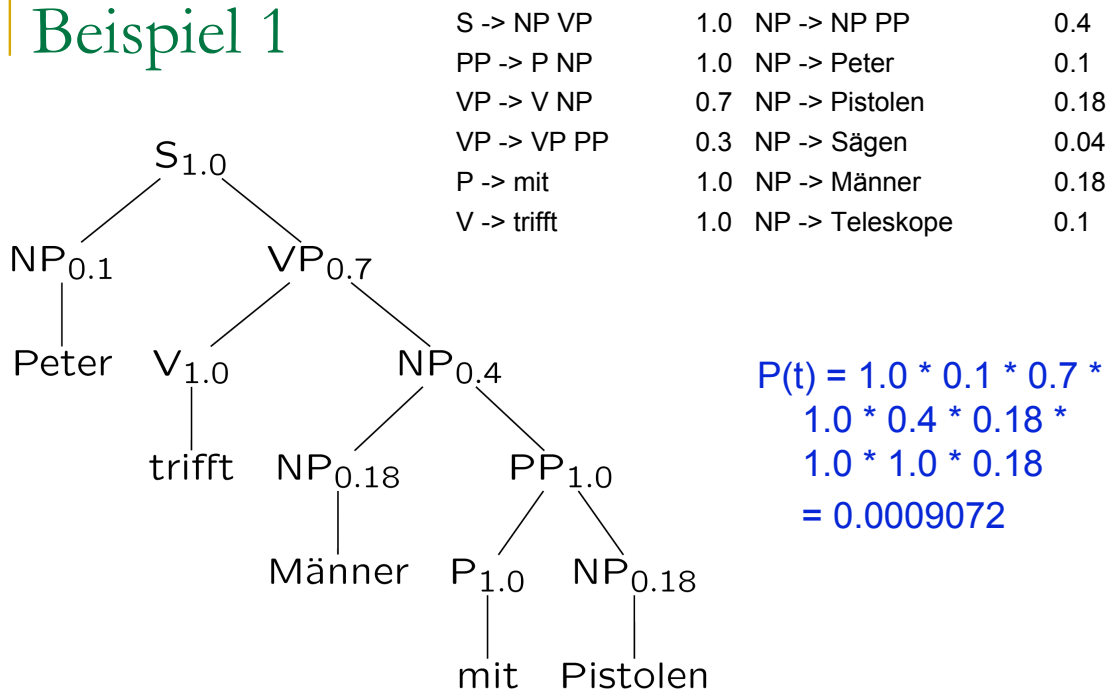
Einführung in die Coli WS 04/05

# Beispielgrammatik

S -> NP VP	1.0	NP -> NP PP	0.4
PP -> P NP	1.0	NP -> Peter	0.1
VP -> V NP	0.7	NP -> Pistolen	0.18
VP -> VP PP	0.3	NP -> Sägen	0.04
P -> mit	1.0	NP -> Männer	0.18
V -> trifft	1.0	NP -> Teleskope	0.1

Einführung in die Coli WS 04/05

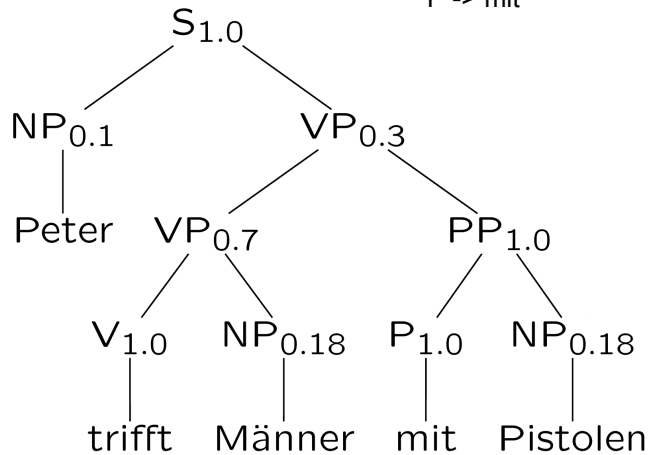
## Beispiel 1



Einführung in die Coli WS 04/05

## Beispiel 2

S -> NP VP	1.0	NP -> NP PP	0.4
PP -> P NP	1.0	NP -> Peter	0.1
VP -> V NP	0.7	NP -> Pistolen	0.18
VP -> VP PP	0.3	NP -> Sägen	0.04
P -> mit	1.0	NP -> Männer	0.18
	1.0	NP -> Teleskope	0.1



$$\begin{aligned}
 P(t) &= 1.0 * 0.1 * 0.3 * \\
 &\quad 0.7 * 1.0 * 0.18 * \\
 &\quad 1.0 * 1.0 * 0.18 \\
 &= 0.0006804
 \end{aligned}$$

Einführung in die Coli WS 04/05

## Evaluation von Baumbank-Grammatiken

- Standardverfahren:
  - Trainingskorpus (Regelextraktion)
  - Testkorpus (Regelanwendung)
- **Abdeckung:** Kann die Grammatik alles parsen?
  - Wenn Regeln fehlen, können Testsätze nicht geparkt werden
  - **Ergebnis: kein Problem (Übergenerierung)**
    - Aber keine Garantier fuer vernuenftige Analysen?
- **Korrektheit:** Ist der wahrscheinlichste Baum der gelernten Grammatik (GB) auch der Baum in der Baumbank (BB)?
  - Precision: Anteil der Regeln im GB, die auch im BB vorkommen
  - Recall: Anteil der Regeln im BB, die auch im GB vorkommen

Einführung in die Coli WS 04/05

---

## Korrektheit: Ergebnisse (Englisch)

- Standard-Baumbank-Grammatik mit starker Unabhängigkeitsannahme
  - Penn Treebank / WSJ

Satzlänge	Ø Länge	Precision	Recall
2-16 Test	11.4	85.0	87.7
2-25 Test	16.3	82.0	84.0
2-40 Test	21.9	78.8	80.4

(Charniak 1993)

---

Einführung in die Coli WS 04/05

---

## Korrektheit: Ergebnisse (Deutsch)

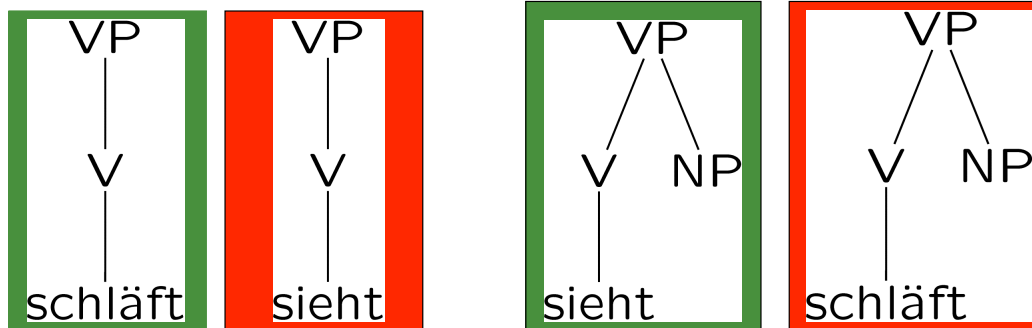
- Standard-Baumbank-Grammatik mit starker Unabhängigkeitsannahme
  - NEGRA: 72.99 Recall, 70.00 Precision
- Hypothesen:
  - Deutsch ist schwieriger
  - Verschiedene Annotationsschemata
    - Deutsche Parsebäume haben weniger Knoten

---

Einführung in die Coli WS 04/05

## Probleme der Unabhängigkeitsannahme

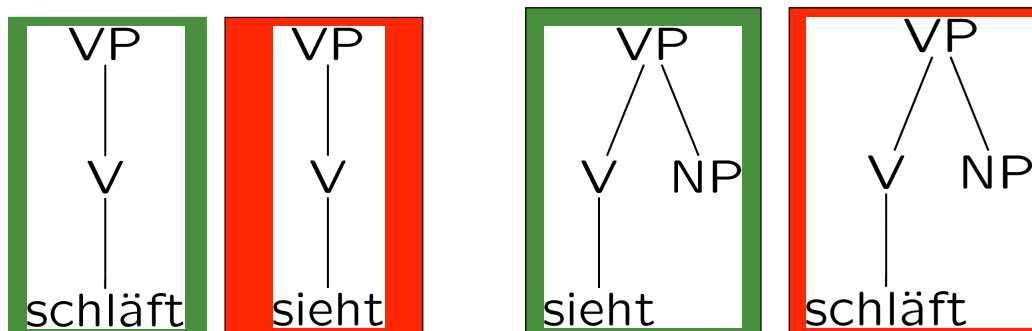
- Erinnerung: Annahme ist, daß alle Regeln voneinander **unabhängig** sind
- Gegenbeispiel: VP
  - Bei transitivem Verb soll  $VP \rightarrow V NP$  wahrscheinlicher sein
  - Bei intransitivem Verb soll  $VP \rightarrow V$  wahrscheinlicher sein



Einführung in die Coli WS 04/05

## Unabhängigkeitsannahmen (II)

- Wieso kann die Grammatik nicht unterscheiden?
  - Wahl der VP-Regel und Wahl des Verbs sind **unabhängig**
    - $P(\dots | VP)$  und  $P(\dots | V)$
    - Können sich nicht beeinflussen: Wahrscheinlichkeit ist **Produkt**



Einführung in die Coli WS 04/05

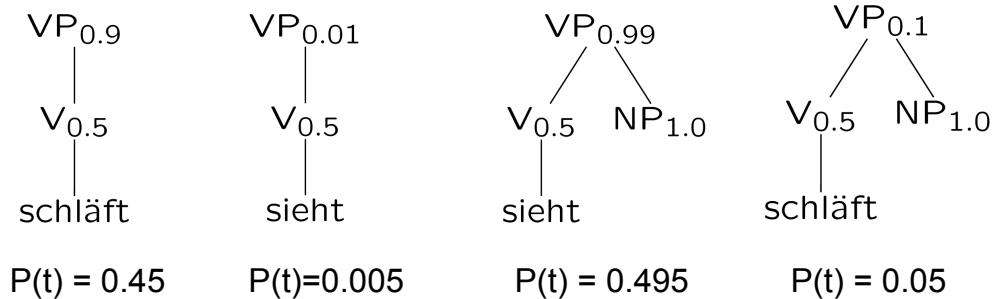
# Lexikalisierung

## ■ Lösung: Lexikalisierung

- P(Regel) auf **lexikalischen Kopf** konditionieren

- $P(V \text{ NP} \mid VP, \text{schlafen}) = 0.1$
- $P(V \mid VP, \text{schlafen}) = 0.9$
- $P(V \text{ NP} \mid VP, \text{sehen}) = 0.99$
- $P(V \mid VP, \text{sehen}) = 0.01$

- $P(\text{schlafen} \mid V) = 0.5$
- $P(\text{sehen} \mid V) = 0.5$
- $P(NP) = 1.0$



Einführung in die Coli WS 04/05

# Mehr zur Lexikalisierung

## ■ Lexikalischer Kopf hängt von Phrasentyp ab

- Verb (VP,S)
- Nomen (NP)
- Adjektiv(AP)

## ■ Berechnung aus Frequenzen:

- Kopf wird neues konditionierendes Ereignis (Feature)
- **Interaktion** zwischen Wahl von Phrasenstruktur und Kopf
  - Beispiel: Wahl des Verbs beeinflusst Entscheidung transitiv/intransitiv

$$P(RS \mid LS, K) = \frac{f(LS \rightarrow RS, K)}{f(LS, K)} = \frac{f(LS \rightarrow RS, K)}{\sum_{RS} f(LS \rightarrow RS, K)}$$

- Berechnung der Baumwahrscheinlichkeit immer noch als Produkt
  - Nur lokale Abschwächung der Unabhäenigkeitsannahme

Einführung in die Coli WS 04/05



# Korrektheit und Grammatikgröße

- Lexikalisierung erhöht **Korrektheit**
  - Modelliert wichtige Eigenschaften der Daten
    - Sinnvolle Parsebäume erhalten hohe Wahrscheinlichkeit
- Aber: Modell wird sehr groß
  - Je eine Regel für Kombination (Regel,Kopfwort)
  - Abdeckung wird ein Problem (Sparse Data)
    - Wie behandelt man ungesehene Verben?
    - Typisch: Backoff zu unlexikalisiertem Modell
  - Generalisierung wird u.U. schlechter
  - Sehr großer Suchraum

Einführung in die Coli WS 04/05

## Weitere Konsequenzen der Unabhängigkeitsannahmen

- Regeln „kennen“ ihre globale Position im Baum nicht
  - Subjekt-NPs und Objekt-NPs verwenden gleiche Wahrscheinlichkeit  $P((RS).... | NP)$

Wird von den Daten (Englisch) nicht bestätigt

Regel	% als Subjekt	% als Objekt
NP -> PRP	13.7%	2.1%
NP -> NP PP	5.6%	14.1%
NP -> NP SBAR	0.5%	2.6%
NP -> DT NN	5.6%	4.6%

(Manning and Schütze 1999)

Einführung in die Coli WS 04/05

## Regeln mit Geschichte (History)

- Lösung: Knoten kennen ihre „Geschichte“

- P(Regel) auf „Großvater“ konditionieren

- $P(..(RS).... | NP, \mathbf{S})$  = Subjekt

- $P(..(RS).... | NP, \mathbf{VP})$  = Objekt

Regel	% als Subjekt	% als Objekt		
NP -> PRP	13.7%	2.1%	$P(\text{PRP}   \text{NP}, \mathbf{S}) =$	0.137
			$P(\text{PRP}   \text{NP}, \mathbf{VP}) =$	0.021
NP -> NP PP	5.6%	14.1%	$P(\text{NP PP}   \text{NP}, \mathbf{S}) =$	0.056
			$P(\text{NP PP}   \text{NP}, \mathbf{VP}) =$	0.141
NP -> NP SBAR	0.5%	2.6%	$P(\text{NP SBAR}   \text{NP}, \mathbf{S}) =$	0.005
			$P(\text{NP SBAR}   \text{NP}, \mathbf{VP}) =$	0.026
NP -> DET NN	5.6%	4.6%	$P(\text{DET NN}   \text{NP}, \mathbf{S}) =$	0.056
			$P(\text{DET NN}   \text{NP}, \mathbf{VP}) =$	0.046

Einführung in die Coli WS 04/05

## Regeln mit Geschichte

- Mehr Regeln

- $P(RS | LS, GV)$ : Eine Regel pro Kombination aus Regel und Großvaterkategorie

- Nicht so schlimm wie bei Lexikalisierung

- Beispiel NP: Regelzahl verdoppelt sich (NP in VP und NP in S)

- Sparse data-Problem nicht ganz so dramatisch

- Hilft auch bei anderen Unterscheidungen

- Direktes und indirektes Objekt

- Hauptsatz und Nebensatz

- Schwierig bei flachen Annotationsschemata

- Wenn es keinen VP-Knoten gibt...?

- Mögliche Lösung: Schwesterknoten anschauen

Einführung in die Coli WS 04/05

---

## Ein Modell aus der Praxis

- Benutzt Geschichte und Lexikalisierung
  - Großvaterknoten l
  - Kopf h
  - Modell mit allen Features ist am genauesten, hat aber keine gute Abdeckung
  - Einfachere Modelle sind dümmer, aber robuster
    - Die Koeffizienten  $\lambda_i$  werden durch Optimierung festgelegt

$$P(RS|LS, h, l) = \lambda_1 p(RS|LS, h, l) + \lambda_2 p(RS|LS, h) + \lambda_3 p(RS|LS, l) + \lambda_4 p(RS|LS)$$

(Charniak 1996)

---

Einführung in die Coli WS 04/05

---

## State of the art (Modelle mit Geschichte, Lexikalisierung und weiteren Tricks)

- Englisch: Recall 90%, Precision 90%
  - Penn Treebank
  - Vergleich: naives Modell Recall 79%, Precision 80%
- Deutsch: Recall 81%, Precision 77%
  - NEGRA
  - Vergleich: naives Modell Recall 73%, Precision 70%
- Deutliche Verbesserung durch Abschwächung der Unabhaengigkeitsannahme

---

Einführung in die Coli WS 04/05

---

## Parsing mit PCFGs

1. Lerne Regeln mit Wahrscheinlichkeiten
2. Berechne Wahrscheinlichkeiten für Bäume aus Regelwahrscheinlichkeiten
3. Finde effizientes Verfahren, um wahrscheinlichste Bäume zu konstruieren

---

Einführung in die Coli WS 04/05

---

## Herkömmliches Parsing

- Top-Down: Fange an mit S, und wende **alle anwendbaren** Regeln an
  - Beispiel: Earley-Algorithmus
- Bottom-Up: Fange an mit Worten, kombiniere mit **allen anwendbaren** Regeln

## Erschöpfendes (exhaustives) Parsing

---

Einführung in die Coli WS 04/05

---

# Parsing mit Baumbank-Grammatiken

- Viele Regeln
  - Exhaustives Parsing nicht beherrschbar für Sätze mit mehr als 10 Wörtern
- Hoffnung: ungleiche Wahrscheinlichkeitsverteilung
  - Es gibt wenig „vernünftige“ Parsebäume, viel Müll
  - „Vernünftige“ Parsebäume sind viel wahrscheinlicher als (fast) alle anderen

Aufgabe: indentifiziere (konstruiere) die kleine Anzahl sehr wahrscheinlicher Parsebäume

---

Einführung in die Coli WS 04/05

---

## Folge dem wahrscheinlichen Parsebaum?

- Deterministisches Parsing (verfolge eine einzige Hypothese)
  - Beginne irgendwo
  - Finde den lokal wahrscheinlichsten Parsebaum
  - Finde wahrscheinlichste Erweiterung
  - Wiederhole
- Erhält man so immer den global wahrscheinlichsten Parsebaum?
- **Nein!**
  - Holzwegsätze (garden path sentences):  
[Peter, der viel redet,] mag ich nicht.  
[The horse [raced past the barn]] fell.

---

Einführung in die Coli WS 04/05

---

# Das Problem der lokalen Ambiguitäten

- Problem: lokale Ambiguitäten
    - Der lokal beste Baum muss nicht zum global besten Baum gehören
    - Mögliche Folgen:
      - Resultierender Baum ist nicht der wahrscheinlichste
      - Parsing schlägt komplett fehl („harter Holzwegsatz“)
  - Durch geschickte Formulierung der Grammatik ist deterministisches Parsing manchmal möglich
    - In diesem Fall gibt es einen Algorithmus, der garantiert den besten Baum liefert (Viterbi-Algorithmus)
  - Oft nicht möglich
    - Näherungslösung: [Beam Search \(Strahlsuche\)](#)
- 

Einführung in die Coli WS 04/05

---

## Beam Search

- Verfolge die [n](#) besten Analysen
    - Verallgemeinerung von deterministischem Parsing
  - Datenstruktur: Liste von halbfertigen Parsebäumen (Hypothesen) mit Wahrscheinlichkeiten
    - Jede Hypothese um jede anwendbare Regel erweitern
    - Wahrscheinlichkeiten für erweiterte Hypothesen ausrechnen
    - Die [n](#) besten Hypothesen behalten, die anderen vergessen
    - Wiederholen, bis n Hypothesen für ganzen Satz
- 

Einführung in die Coli WS 04/05

## Beam Search

- $n = 1$ : deterministisches Parsing
- $n = \infty$ : exhaustives Parsing
- Gefahr: für kleines  $n$  kann der global beste Parsebaum „aus dem Strahl fallen“
  - Wie groß muß  $n$  sein?

Einführung in die Coli WS 04/05

## Beam Search beim Parsen

(Ratnaparkhi 1999)

n	Zeit (s)	Precision	Recall
20	2.07	87.9	87.1
15	1.58	87.7	86.9
10	1.07	87.7	86.9
7	0.76	87.4	86.6
5	0.56	87.3	86.8
2	0.35	85.1	86.1
1	0.14	82.4	83.4

Einführung in die Coli WS 04/05

---

## Beam Search in der Praxis

- Optimales  $n$  muß empirisch ermittelt werden
- Funktioniert ziemlich gut
  - Wegen ungleicher Verteilung ist global bester Parsebaum ist fast immer in der Spitzengruppe
- Beam Search nur Methode, um Suchraum einzuschränken („Filter“)
  - Läßt offen, wie Hypothesen konstruiert werden
    - Top-down parsing: Beginne mit S
    - Bottom-up parsing: Beginne mit Worten

---

Einführung in die Coli WS 04/05

---

## Zusammenfassung

- Kontextfreie Grammatiken können aus Baumbanken gelernt werden
  - Viele Regeln
  - Kein Problem mit Abdeckung
  - Hohe Ambiguität (Übergenerierung)
- Probabilistische kontextfreie Grammatiken
  - Möglichkeit, über „besten Baum“ zu sprechen
  - Regelwahrscheinlichkeiten aus Baumbank ablesen

---

Einführung in die Coli WS 04/05



---

# Zusammenfassung

- Was ist die Wahrscheinlichkeit eines Baumes?
    - Naiv: Starke Unabhängigkeitsannahme
      - Jede Regel ist unabhängig
    - Lokale Abschwächung der Unabhängigkeit
      - Lexikalisierung: Interaktion Regel - lexik. Kopf
        - $P(\text{VP} \rightarrow \text{V NP} | \text{Trans. V}) > P(\text{VP} \rightarrow \text{V NP} | \text{Intrans. V})$
      - Geschichte: Interaktion Regel - Grossvaterkategorie
        - $P(\text{NP} \rightarrow \text{PRP} | \text{S}) > P(\text{NP} \rightarrow \text{PRP} | \text{VP})$
  - Wie finde ich den besten Baum?
    - Beam search
-