

Lexikalisiertes Parsing

Sharon Friedrich, Katharina Wäschle

Gliederung

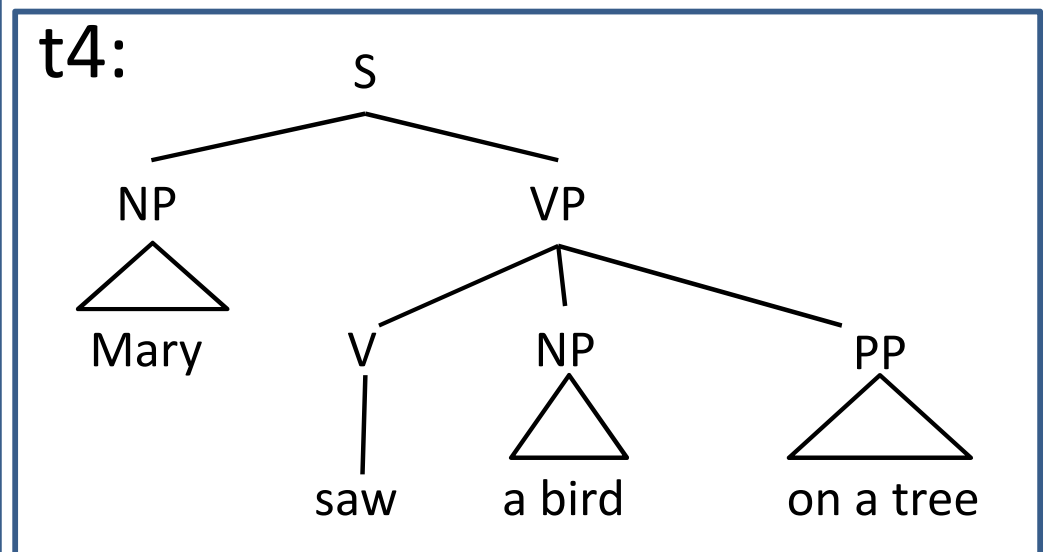
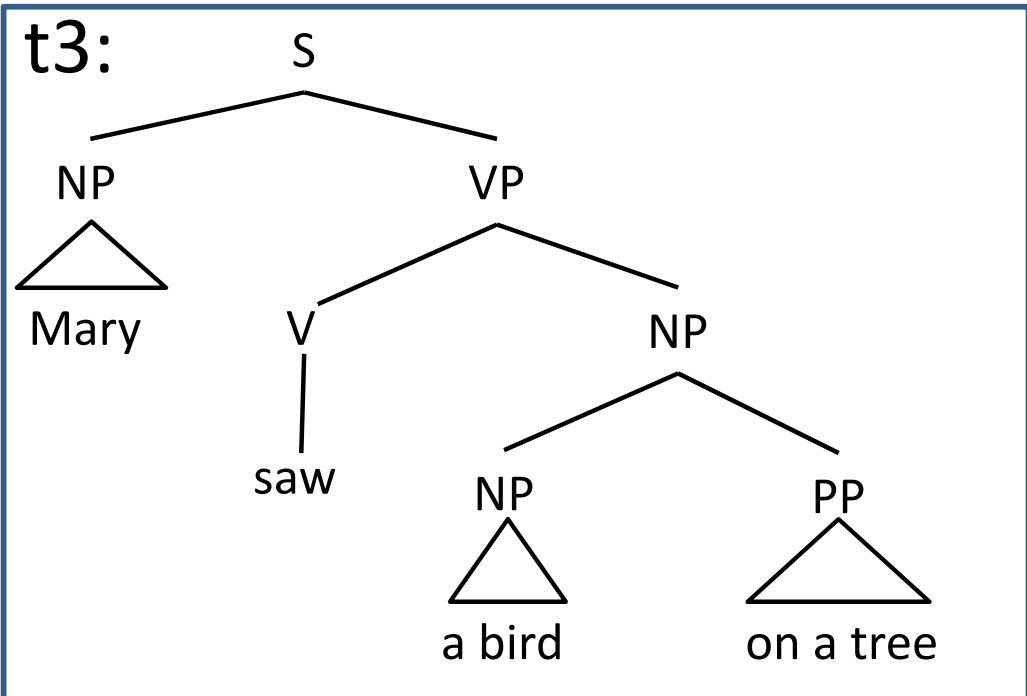
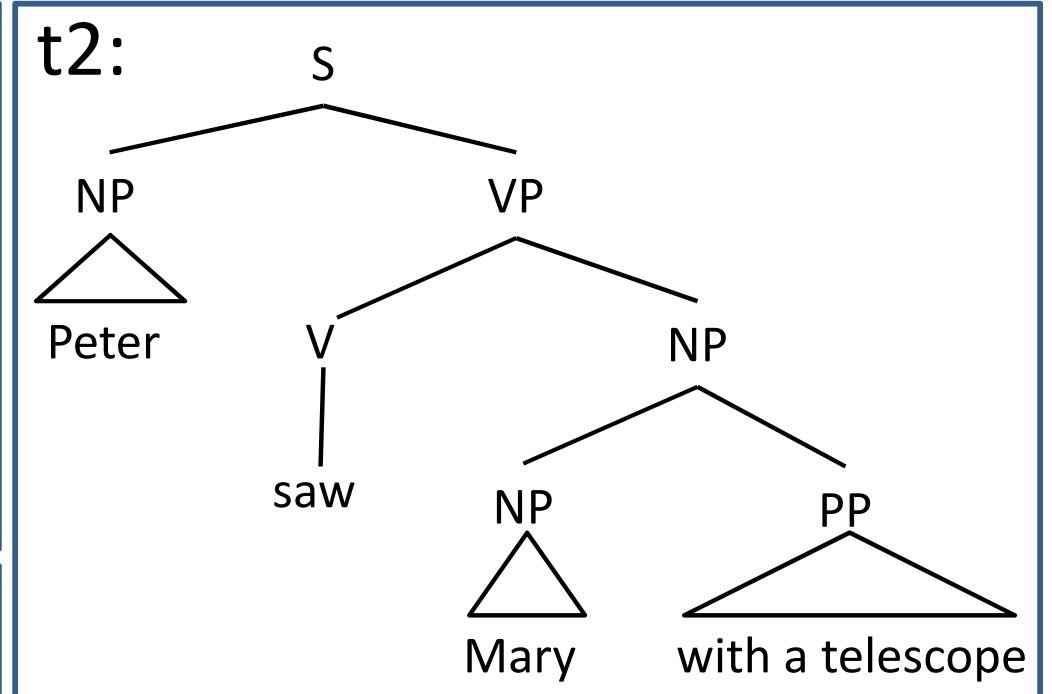
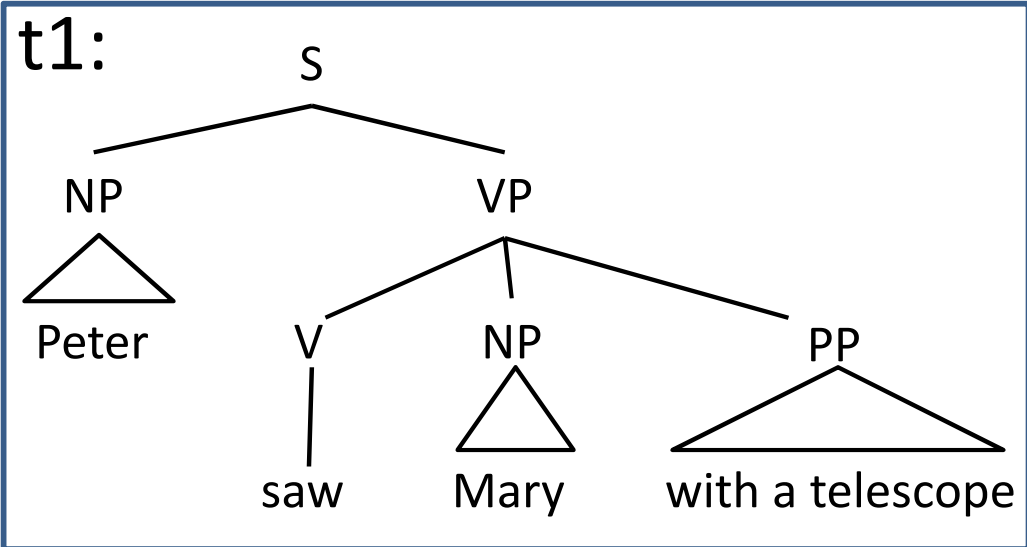
- Motivation
- Lexikalisierung
- Charniak's Wahrscheinlichkeitsmodell
 - Dependencies
 - Lexikalisierte Regeln
 - Smoothing mit Deleted Interpolation
 - Bestimmung d. Wahrscheinlichkeiten
- Auswertung des Parsing Systems
 - Parseval Scores
 - Ergebnisse
- Vergleich Charniak - Collins
- Ausblick
 - Grammatiktransformation

MOTIVATION

Aufgabenblatt 2

Aufgabe 1

Motivation



Motivation

- **Problem:**
Wahrscheinlichkeiten sind gekoppelt
 $p(t1) > p(t2) \Leftrightarrow p(t3) < p(t4)$
- **Lösungsansatz:**
Lexikalisierung der Regeln

LEXIKALISIERUNG

Kopf einer Phrase

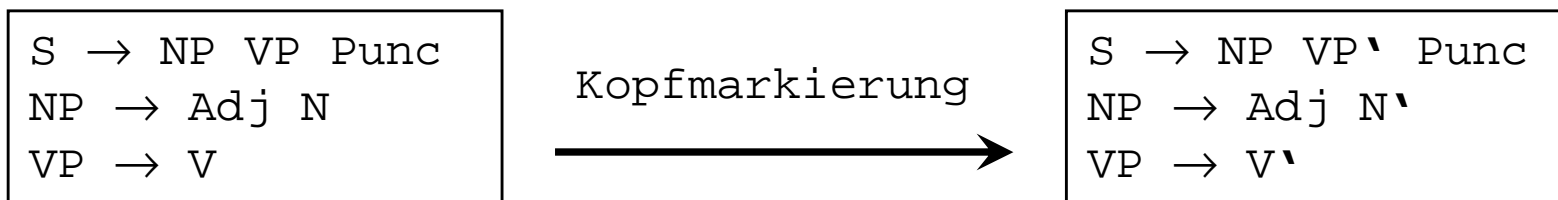
Definition:

- Wichtigstes lexikalisches Element in einer Konstituente
- z.B. V in VP, N in NP
- Blatt (Terminalknoten) im Baum

Lexikalisierung

- **Schritt (1):**

Jede Regel der Grammatik wird mit Kopfmarkierungen versehen.

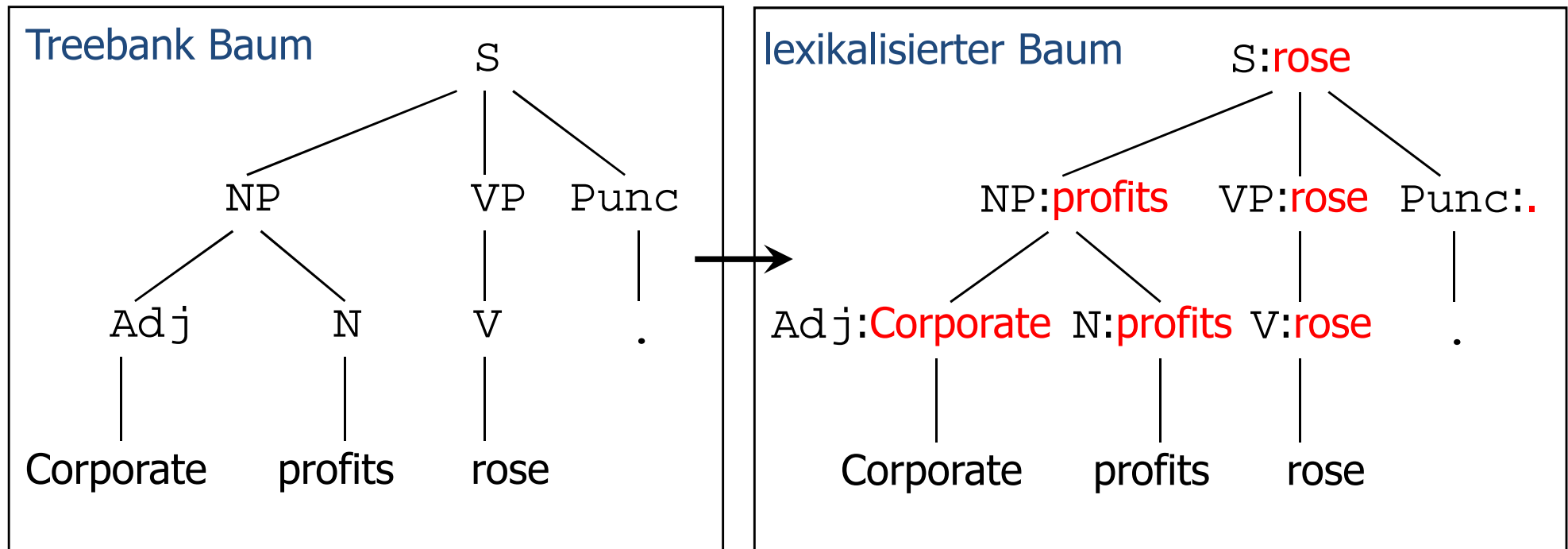


- **Schritt (2):**

Modifizieren des Originalbaumes

- mit Hilfe der markierten Regeln
- Wandern des Kopfes Bottom-Up bis zur maximalen Phrase (**head percolation**)

→ **Lexikalisierte Baum**



Effektivität von Lexikalisierung

	LR	LR2	LP	LP2	CB	OCB	2CB
PCFG	71,2	71,7	75,3	75,8	2,03	39,5	68,1
Minimal	82,9	83,4	83,6	84,1	1,40	53,2	79,0
No Classes	86,2	86,8	85,8	86,4	1,14	59,9	83,4
Basic	86,3	86,8	86,6	87,1	1,09	60,7	84,0
Full	86,9	87,5	86,8	87,4	1,00	62,1	86,1

Sätze mit ≤ 40 Wörtern
Insgesamt 2245 Sätze

CHARNIAKS WAHRSCHEINLICHKEITSMODELL

Das Wahrscheinlichkeitsmodell

- Entwickelt von Eugene Charniak
- Grundlage für ein **probabilistisches Parsingsystem**
 - Ausgabe:
Analysebaum π eines Satzes s mit $p(\pi | s)$ maximal
 - Verwendung kontextfreier Grammatik

Gesamtwahrscheinlichkeit

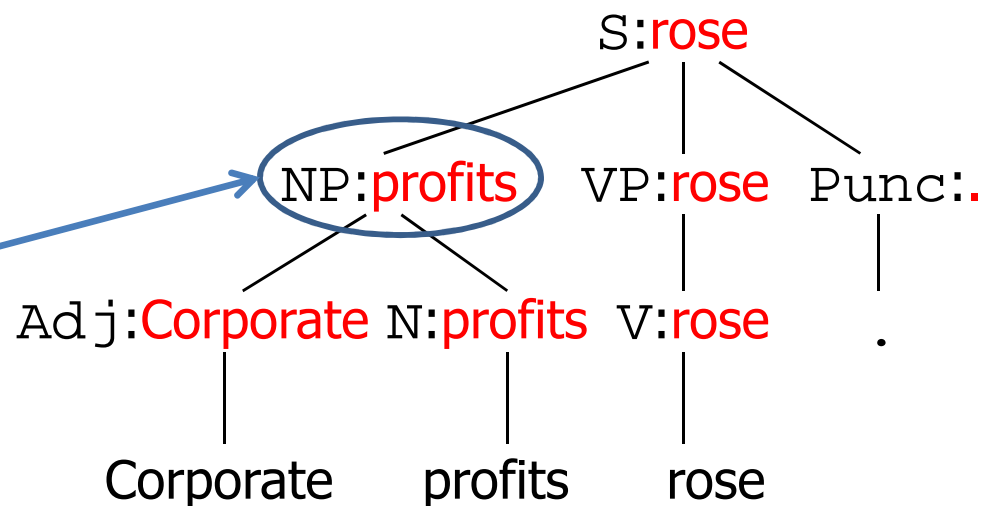
- Produkt der Wahrscheinlichkeiten aller im Baum angewandter Regeln
- **Regelwahrscheinlichkeiten:**
 - W'keit einer internen Regel = W'keit für die Form der Regel * W'keiten für die Köpfe der Subkonstituenten (außer für den Kopf selbst)
 - Wahrscheinlichkeit für den Kopf: **Dependency**
 - Wahrscheinlichkeit für die Form der Regel: **Lexikalisierte Regel**

Dependencies

- **Sei:** **d** der Kopf einer Konstituente
D der Typ dieser Konstituente
C der Typ der Mutterkonstituente
h der Kopf der Mutterkonstituente
- **Es gilt:** W'keit von d hängt nur von D, C und h ab
→ $p(d \mid D, C, h)$

- **Beispiel:**

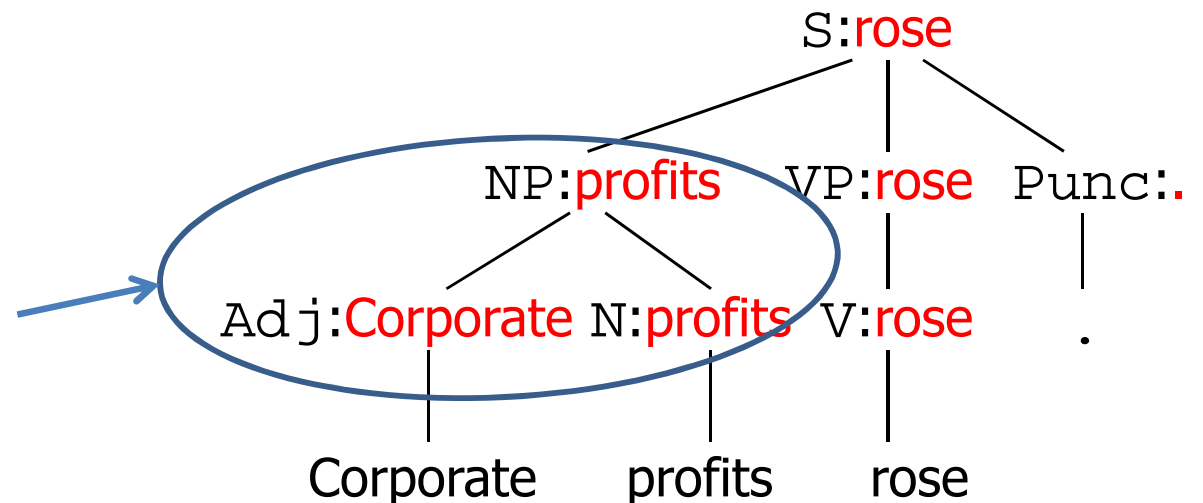
$p(\text{profits} \mid \text{NP}, \text{S}, \text{rose})$



Lexikalisierte Regeln

- **Sei:** r die Regel
 C der Typ der Konstituente
 h der Kopf der Konstituente
 C_p der Typ der Mutterkonstituente
- **Es gilt:** W'keit von r hängt nur von C , h und C_p ab:
 $\rightarrow p(r \mid C, h, C_p)$
- **Beispiel:**

$p(\text{NP} \rightarrow \text{adj N} \mid \text{NP}, \text{profits}, \text{S})$



- **Problem:**

Trainingsdaten können nicht alle Möglichkeiten überdecken.

→ **Gefahr:** Satz bzw. Analysebaum befindet sich nicht in der PTB

- **Grundregel:**

„Wenn man ein Ereignis nicht beobachtet hat, kann man daraus nicht schließen, dass es niemals vorkommt!“

(Quelle: <http://www.phil.uni-passau.de/linguistik/lehre/mlnlp/LM.pdf>)

Smoothing

- **Problem:**

Abschätzung der Wahrscheinlichkeiten von Fällen, die in der PTB nicht vorkommen

- **Lösung:**

Smoothing

– Verschiebe Wahrscheinlichkeitsmasse von den gesehenen Fällen zu den ungesehenen Fällen

→ Methode: Deleted Interpolation

Deleted Interpolation

- Methode zur Abschätzung der Wahrscheinlichkeit von ungesehenen Fällen
- Lineare Interpolation von mehreren Wahrscheinlichkeiten
→ Jede betrachtet eine Bedingung weniger.

$$p(w_i | w_{i-2}, w_{i-1}) = \lambda_1 * p'(w_i | w_{i-2}, w_{i-1}) + \lambda_2 * p'(w_i | w_{i-1}) + \lambda_3 * p'(w_i)$$

- p' Distribution, die man empirisch aus den Trainingsdaten erhält
- λ_i Gewichtungen
- $\sum_i \lambda_i = 1$ (Bestimmung der λ_i mit zweitem Trainingskorpus)

Charniak's modified formula for Deleted Interpolation:

- **Dependencies**

$$\begin{aligned} p(d \mid D, C, h) = & \lambda_1(e) * p'(d \mid D, C, h) \\ & + \lambda_2(e) * p'(d \mid D, C, \text{class}(h)) \\ & + \lambda_3(e) * p'(d \mid D, C) \\ & + \lambda_4(e) * p'(d \mid D) \end{aligned}$$

(1)

Charniaks modifizierte Formel für Deleted Interpolation:

- **Lexikalisierte Regeln**

$$\begin{aligned} p(r \mid C, h, C_p) = & \lambda_1(e) * p'(r \mid C, h, C_p) \\ & + \lambda_2(e) * p'(r \mid C, h) \\ & + \lambda_3(e) * p'(r \mid C, \text{class}(h)) \\ & + \lambda_4(e) * p'(r \mid C, C_p) \\ & + \lambda_5(e) * p'(r \mid C) \end{aligned}$$

(2)

- **Beispiel für Smoothing:**

	p(corporate ADJ, NP, profits)	p(profits NP, S, rose)
$p'(d D, C, h)$	0,2449	0
$p'(d D, C, \text{class}(h))$	0,0149821	0,00352223
$p'(d D, C)$	0,00533	0,0006274
$p'(d D)$	0,004179	0,000556527

	p(NP -> ADJ N NP, profits, S)
$p'(r C, h, C_p)$	0,1707
$p'(r C, h)$	0,1875
$p'(r C, \text{class}(h))$	0,1192
$p'(r C, C_p)$	0,0176
$p'(r C)$	0,0255

Wahrscheinlichkeit einer Regel

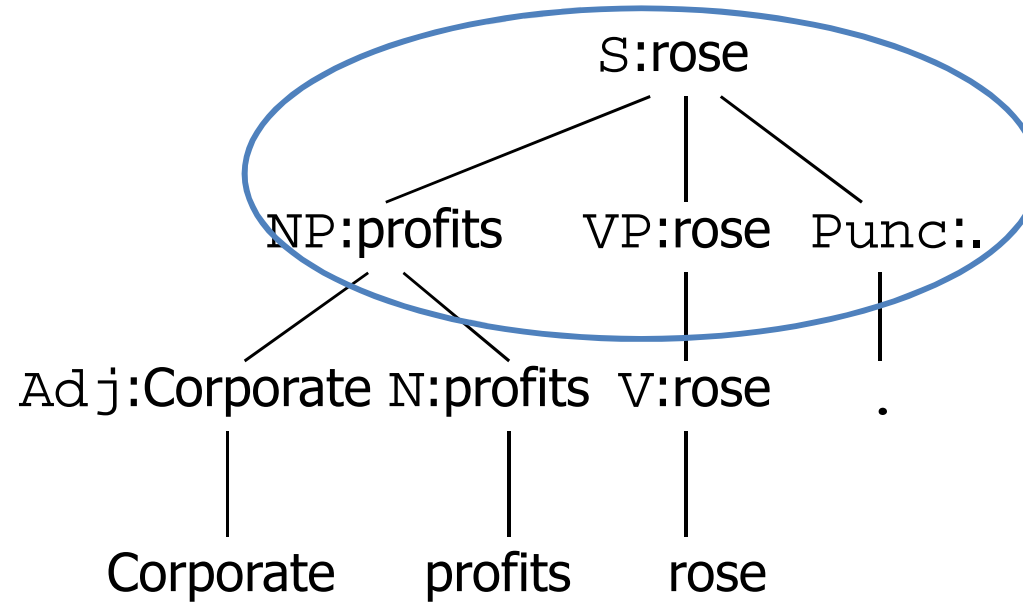
- Wiederholung:
 - W'keit einer internen Regel = W'keit für die Form der Regel * W'keiten für die Köpfe der Subkonstituenten (außer für den Kopf selbst)
 - **Dependencies:** $p(d \mid D, C, h)$
 - **Lexikalisierte Regeln:** $p(r \mid C, h, C_p)$

$$\rightarrow p(\text{Regel}) = p(r \mid C, h, C_p) * \prod_{i=1}^{n+m} p(d_i \mid D_i, C, h)$$

- m: Anzahl der Geschwister links v. Kopf H
- n: Anzahl der Geschwister rechts v. Kopf H

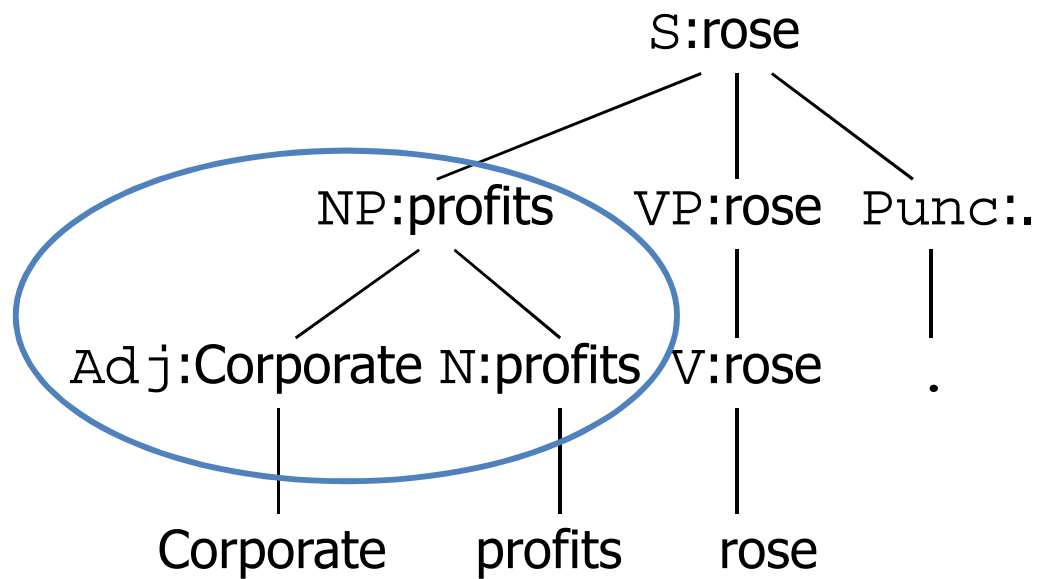
Anwendung

- Baum:



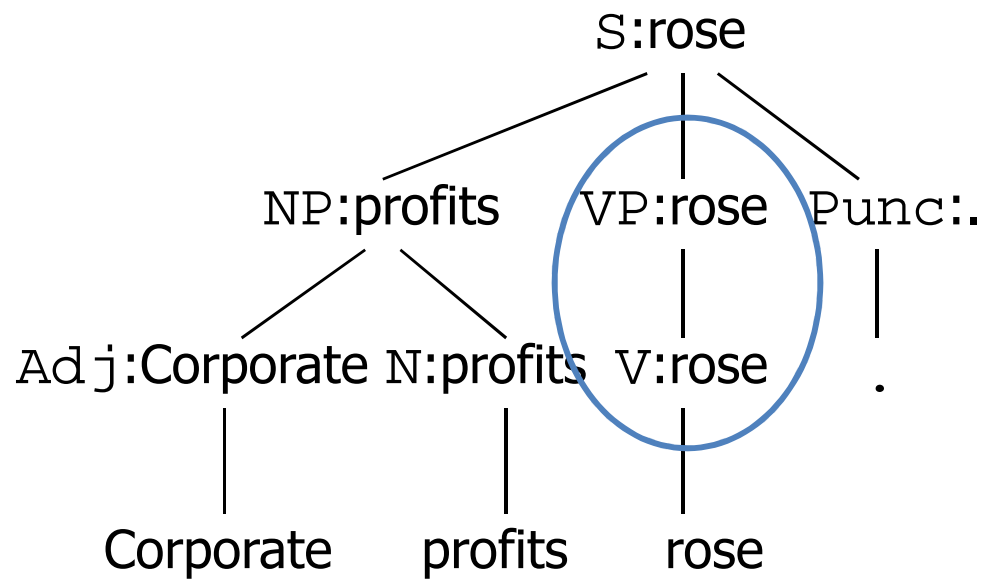
Anwendung

- Baum:



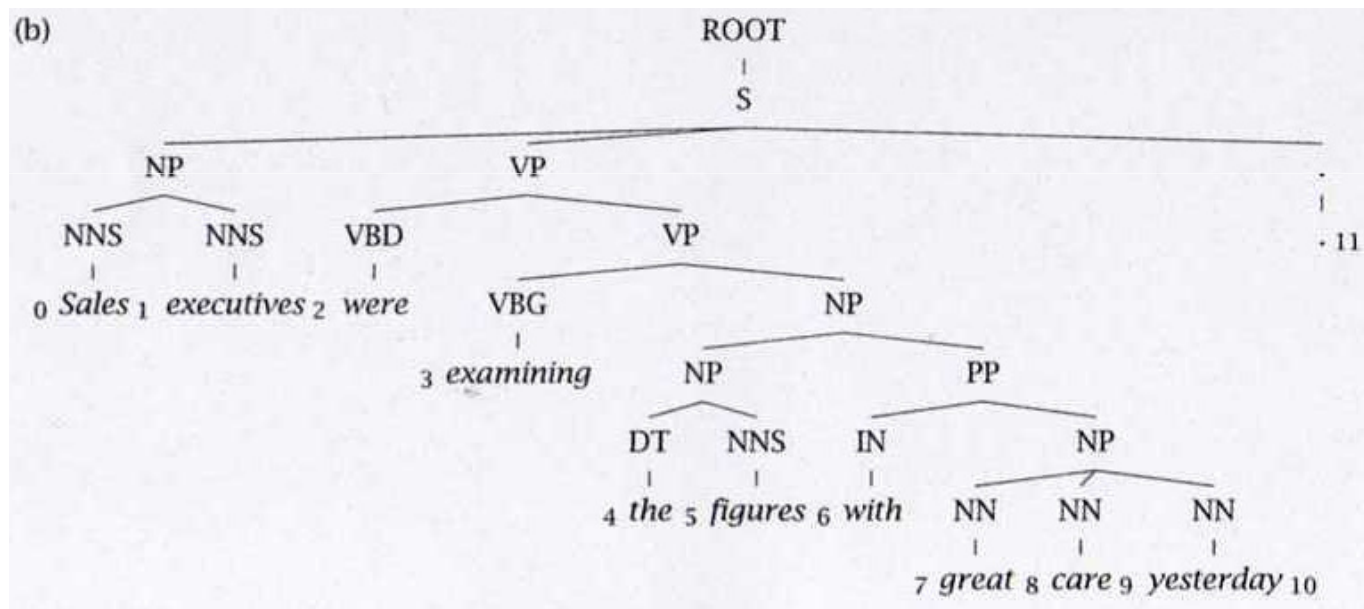
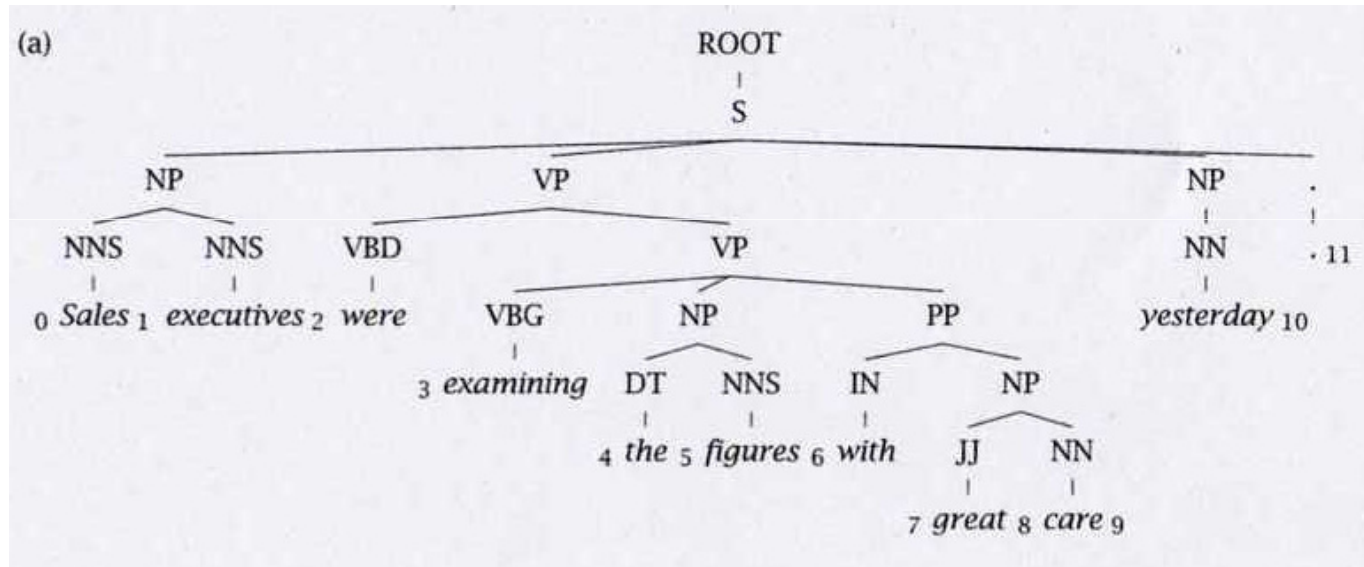
Anwendung

- Baum:

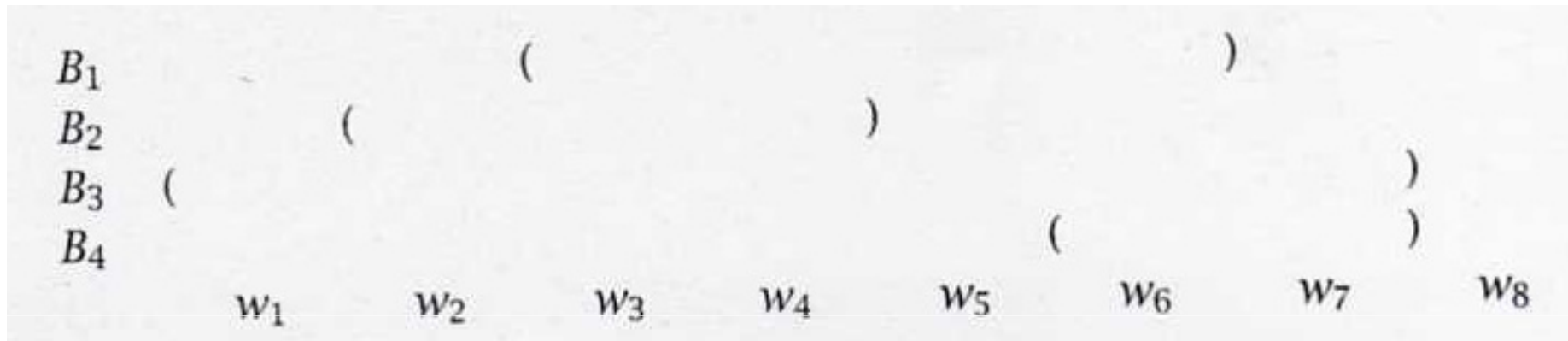


AUSWERTUNG DES PARSING SYSTEMS

Parseval Scores



Syntaxbaum in Klammerschreibweise



Parseval Scores

- **Precision (LP):** Prozentsatz der Klammern im Parse, die mit denen im korrekten Baum übereinstimmen
- **Recall (LR):** Prozentsatz der Klammern des korrekten Baums, die im Parse sind
- **Cross-Bracketing (CB):** Durchschnitt der sich überschneidenden Klammern pro Baum

Auswertung des Parsing Systems

- Training des Parsers: Sektion 02-21 (~ 1 Million Worte) der Penn Wall Street Journal Treebank
- Testen des Parsers: Sektion 23 (50000 Worte)
- Vorbereitendes Testen: Sektion 24
 - Vermeidung von wiederholtem Testen auf Sektion 23
 - Gefahr: unbewusstes Passen des Modells auf das Test Beispiel
 - gleiches Vorgehen wie Collins

Parsen des Testkorpus'

- Verwendung von **5 verschiedenen Versionen des Parsingsystems**
 - **PCFG**: nur $p(r | C)$
 - **Minimal**: zusätzlich $p'(r | C, h, C_p)$
 - **No Classes**: alle Wahrscheinlichkeiten aus den Gleichungen (1) & (2) \rightarrow außer $p'(d | D, C, \text{class}(h))$ & $p'(r | C, \text{class}(h))$
 - **Basic**: Verwendung der Gleichungen (1) und (2)
 - **Full**: Basic + Statistiken basierend auf unüberwachtem Training (30 Millionen Worte des Wall Street Journals)

- 7 verschiedene **Bewertungskriterien:**
 - **LR**
 - **LR2:** LR + Verwendung einer spezifischen Definition von Korrektheit
 - **LP**
 - **LP2:** LP + Verwendung einer spezifischen Definition von Korrektheit
 - **CB**
 - **0CB:** prozentualer Anteil an Sätzen mit 0 überschneidenden Klammern
 - **2CB:** prozentualer Anteil an Sätzen mit ≤ 2 überschneidenden Klammern

Spezifische Definition von Korrektheit

- unterscheidet sich von der Standarddefinition wie folgt:
 - der Konstituententyp ADVP entspricht PRT
 - Interpunktionsfehler haben keine Auswirkung auf die Korrektheit

Ergebnisse

	LR	LR2	LP	LP2	CB	OCB	2CB
PCFG	71,2	71,7	75,3	75,8	2,03	39,5	68,1
Minimal	82,9	83,4	83,6	84,1	1,40	53,2	79,0
No Classes	86,2	86,8	85,8	86,4	1,14	59,9	83,4
Basic	86,3	86,8	86,6	87,1	1,09	60,7	84,0
Full	86,9	87,5	86,8	87,4	1,00	62,1	86,1

Sätze mit ≤ 40 Wörtern
Insgesamt 2245 Sätze

VERGLEICH CHARNIAK - COLLINS



Eugene Charniak

Professor der Informatik und Erkenntniswissenschaften an der Brown
Universität, USA



Michael J. Collins

Professor der Informatik und Electrical Engineering an der MIT

Unüberwachtes Training vs. überwachtes Training

kein explizites Tagging
vs.
Tagging Phase

explizite Grammatik
vs.
neue Grammatik

Charniaks
Wahrscheinlichkeitsformeln
vs.
Collins' Raw-Word-Statistiken

Fazit

- Beide Wahrscheinlichkeitsmodelle weisen eine ähnlich hohe Performanz auf
- Charniak's Modell ist mächtiger als Collins'
 - Unterschiede durch Wahl der Statistik- und Smoothing-Verfahren
 - Collins vernachlässigt:
→ $p'(d | D, C, \text{class}(h))$, $p'(r | C, h, C_p)$, $p'(r | C, \text{class}(h))$

	LR2	LP2	CB	OCB	2CB
Collins	85,8	86,3	1,14	59,9	83,6
Charniak	87,5	87,4	1,0	62,1	86,1

AUSBLICK

Grammatiktransformation

Heutiges Verfahren

- Lexikalisierung bedeutend
 - brachte deutliche Verbesserungen für die Parsing-Systeme
- heutiges Verfahren zur Lexikalisierung:
Grammatiktransformation
 - ermöglicht Anwendung des Standard PCFG-Verfahrens
 - effizienter
 - liefert Ergebnisse wie Charniaks Modell

Grammatiktransformation

- Umwandlung der internen Regeln der Grammatik

→ **Start (neu)**

$S \rightarrow S:h$

→ **lexikalisierte Regeln (modifiziert)**

$C:h \rightarrow D_1:C:h \dots D_m:C:h H:h D_{m+1}:C:h \dots D_{m+n}:C:h$

→ **Dependencies (neu)**

$D:C:h \rightarrow D:d$

→ **lexikalische Regeln (bleibt gleich)**

$C:w \rightarrow w$

Lexikalisierung (Charniak)

Interne Regeln:

- S:rose → NP:profits VP:rose Punc:.
- NP:profits → Adj:Corporate N:profits
- VP:rose → V:rose

Lexikalische Regeln:

- Adj:Corporate → Corporate
- N:profits → profits
- V:rose → rose
- Punc: . → .

Grammatiktransformation

Start: S → S:rose

Lexikalisierte Regeln:

- S:rose → NP:S:rose VP:rose Punc:S:rose
- NP:profits → Adj:NP:profits N:profits
- VP:rose → V:rose

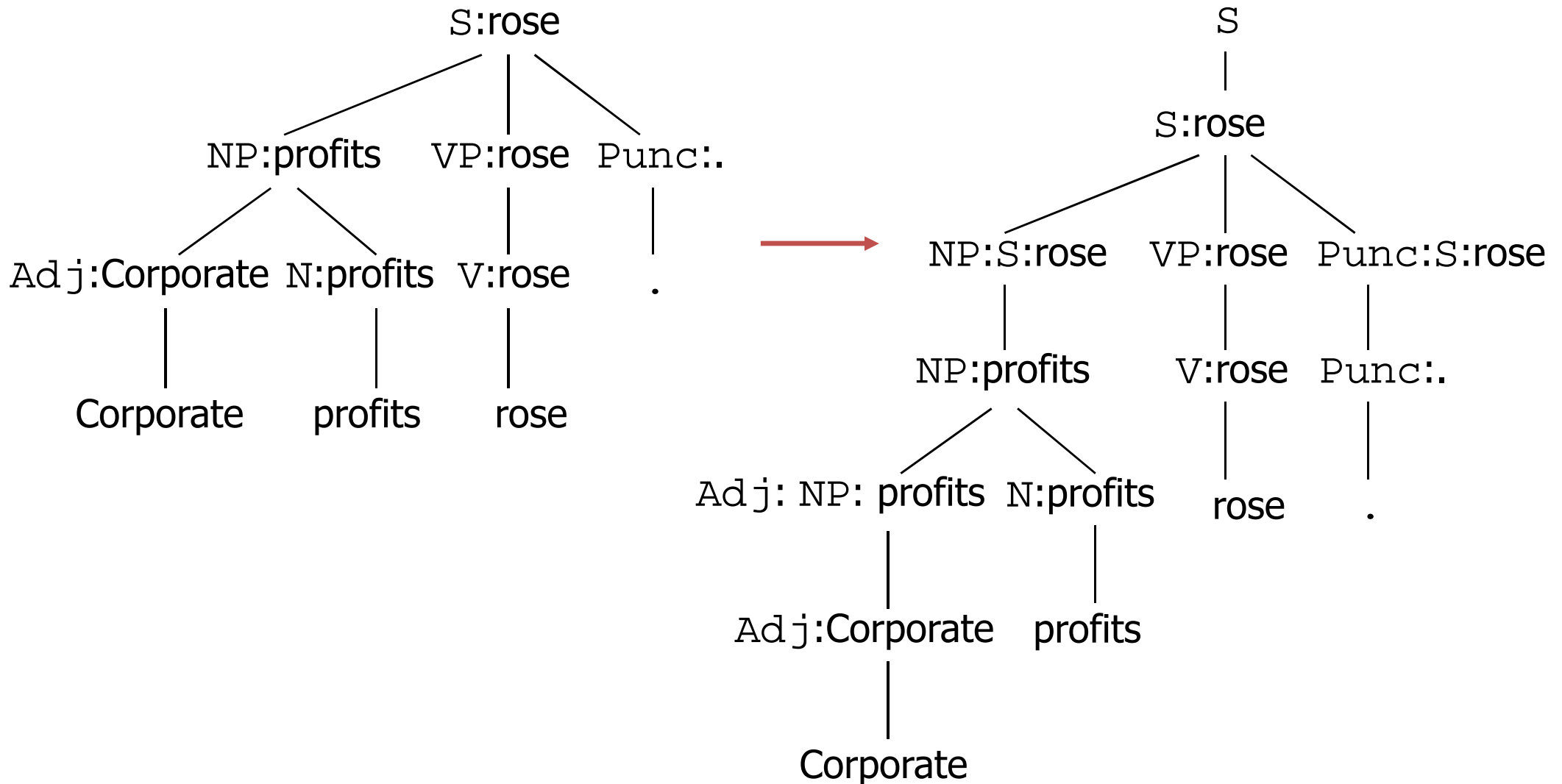
Dependencies:

- NP:S:rose → NP:profits
- Punc:S:rose → Punc:.
- Adj:NP:profits → Adj:Corporate

Lexikalische Regeln:

- Adj:Corporate → Corporate
- N:profits → profits
- V:rose → rose
- Punc: . → .

Transformierung des Parsebaums



Standard PCFG-Wahrscheinlichkeit

- $p_{\text{STANDARD}}(\text{Teilbaum})$
= $p(D_1:C:h \dots D_m:C:h \ H:h \ D_{m+1}:C:h \dots D_{m+n}:C:h \mid C:h)$
* $\prod_{i=1}^{m+n} p(D_i:d_i \mid D_i:C:h)$
= $p(D_1 h \dots D_m \ H \ D_{m+1} \dots \ D_{m+n} \mid C, h) * \prod_{i=1}^{m+n} p(d_i \mid D_i, C, h)$
= $p(r \mid C, h) * \prod_{i=1}^{m+n} p(d_i \mid D_i, C, h)$
= $p_{\text{CHARNIAK}}(\text{Interne Regel})$

Schlussfolgerung

- W' keit für den Teilbaum = W' keit für die interne Regel
 - Beide Wahrscheinlichkeitsmodelle basieren auf der gleichen Lexikalisierungs-idee
 - allerdings: unterschiedliche Korporatypen (Bäume - Sätze)
 - Verwendung des Standard PCFG-Verfahrens effizienter

Literatur

- Eugene Charniak (1997): Statistical Parsing with a Context-Free Grammar and Word Statistics, in: AAI/IAAA.
- Eugene Charniak (1996): Expected-Frequency Interpolation, Technical Report CS96-37.
- Detlef Prescher (2005): Head-Driven PCFGs with Latent-Head Statistics, in: IWPT, S. 115 – 124.
- Karl-Michael Schneider (2004): Sprachmodellierung. Maschinelles Lernen in der Sprachverarbeitung, <http://www.phil.uni-passau.de/linguistik/lehre/mlnlp/LM.pdf>.