

Probabilistic Context-Free Grammars (PCFGs)

Katja Markert

Institut für Computerlinguistik
Uni Heidelberg
markert@cl.uni-heidelberg.de
mit einigen Folien von Yannick Versley

December 17, 2019

- 1 PCFGs
- 2 Probabilistischer CYK
- 3 Woher bekomme ich die Wahrscheinlichkeiten?
- 4 PCFG-Verbesserungen
 - PCFG Eigenschaften
 - Problem 1: Unabhängigkeit der Regelwahrscheinlichkeiten vom Kontext
 - Problem 2: Fehlende Lexikalisierung
- 5 Parserevaluation

- 1 PCFGs
- 2 Probabilistischer CYK
- 3 Woher bekomme ich die Wahrscheinlichkeiten?
- 4 PCFG-Verbesserungen
 - PCFG Eigenschaften
 - Problem 1: Unabhängigkeit der Regelwahrscheinlichkeiten vom Kontext
 - Problem 2: Fehlende Lexikalisierung
- 5 Parserevaluation

- Bis jetzt können wir via CYK alle möglichen Parses eines Satzes finden (in Laufzeit polynomial zur Eingabelänge)
- Jetzt: Wie finde ich den **besten (=wahrscheinlichsten)** Parse?
- Probabilistische kontextfreie Grammatiken (PCFGs)
- Probabilistischer CYK
- Probleme von PCFGs sowie (angedeutete) Verbesserungen
- Parserevaluation

Globale syntaktische Ambiguität

Ein Satz hat mehr als einen Parse (laut einer gegebenen Grammatik).

- CYK gibt alle Parses aus
- Präferenzen:
 - 1 Temporale PPs an Verben oder Ereignis-NPs: *John saw the president during the campaign.*
 - 2 Of-PPs eher an NPs: *He saw the president of the USA.*
 - 3 With-NPs? *He ate the caviar with a spoon. She met the actress with the new film.*
- Lösung: Grammatik als Wahrscheinlichkeitsmodell. Einfachste Lösung: (unlexikalisierte) PCFG

PCFG

Grammatik $G = \langle NT, \Sigma, R, S \rangle$ wobei

- NT : endliche Menge von Nichtterminalen
- Σ : endliche Menge von Terminalen
- S Startsymbol aus NT
- R endliche Menge von Regeln der Form $A \rightarrow \alpha[p]$,
 - wobei A ein Nichtterminal und α eine Sequenz von Symbolen
 $\alpha \in (NT \cup \Sigma)^*$
 - $p \in [0, 1]$ die Wahrscheinlichkeit $p(\alpha|A)$, also

$$\sum_{\alpha} p(\alpha|A) = 1$$

Schreibweisen:

$p(\alpha|A)$; $p(A \rightarrow \alpha|A)$; $p(RHS|LHS)$; $p(A \rightarrow |\alpha)$

Aus Manning und Schütze (1999):

S → NP VP 1.0

PP → IN NP 1.0

VP → VB NP 0.7

VP → VP PP 0.3

NP → NP PP 0.4

NP → astronomers 0.2

NP → women 0.2

NP → telescopes 0.2

IN → with 1.0

VB → meet 1.0

Wahrscheinlichkeit eines Parsebaumes T für Satz s

definiert als das Produkt der Wahrscheinlichkeiten aller Regeln, die zum Aufbau von T verwendet wurden.

$$p(T, s) = \prod_{i=1}^n p(RHS_i | LHS_i)$$

wobei n die Anzahl der in T verwendeten Regeln ist.

- Unabhängigkeitsannahmen: alle verwendeten Regeln sind voneinander unabhängig

Gegeben ein Satz s (und PCFG G):

- 1 Berechne alle möglichen Parses für s
- 2 Berechne für jeden Parsetree T von s seine Wahrscheinlichkeit (Regelmultiplikation)
- 3 Wähle dann den wahrscheinlichsten Baum T für s

Aus Manning und Schütze (1999):

S → NP VP 1.0

PP → IN NP 1.0

VP → VB NP 0.7

VP → VP PP 0.3

NP → NP PP 0.4

NP → astronomers 0.2

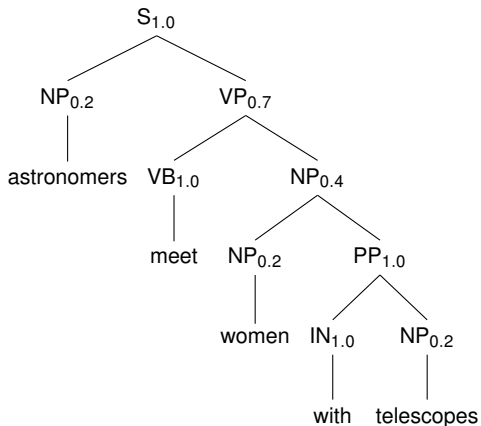
NP → women 0.2

NP → telescopes 0.2

IN → with 1.0

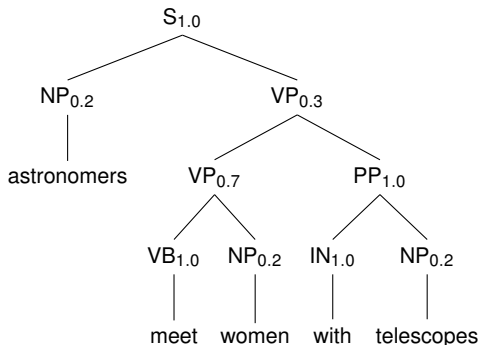
VB → meet 1.0

Beispiel 1: Lesart 1 mit Baum T_1



$$P(T_1) = 1.0 \cdot 0.2 \cdot 0.7 \cdot 1.0 \cdot 0.4 \cdot 0.2 \cdot 1.0 \cdot 1.0 \cdot 0.2 = 0.00224$$

Beispiel 1: Lesart 2 mit Baum T_2



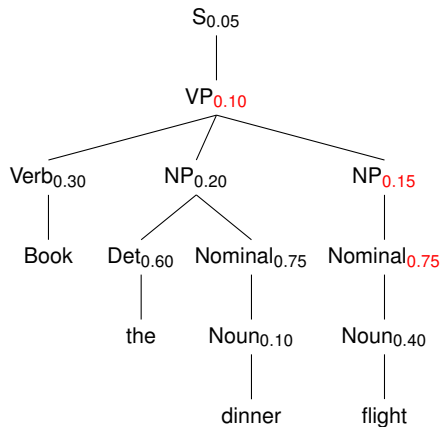
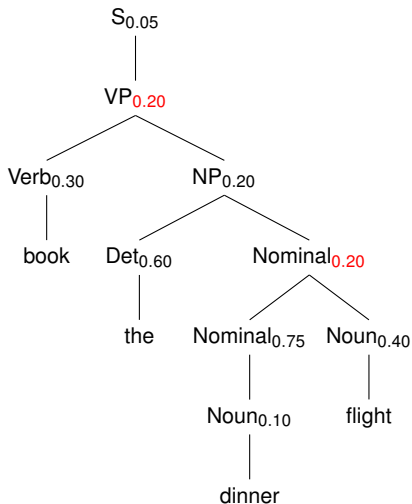
$$P(T_2) = 1.0 \cdot 0.2 \cdot 0.3 \cdot 0.7 \cdot 1.0 \cdot 0.2 \cdot 1.0 \cdot 1.0 \cdot 0.2 = 0.00168$$

T_1 wahrscheinlicher als T_2 : Präferenz für lokales attachment

Beispiel 2 (Jurafsky und Martin, Figure 14.1)

Grammar		Lexicon	
$S \rightarrow NP VP$	[.80]	$Det \rightarrow that$ [.10] a [.30] the [.60]	
$S \rightarrow Aux NP VP$	[.15]	$Noun \rightarrow book$ [.10] $flight$ [.30]	
$S \rightarrow VP$	[.05]	$meal$ [.05] $money$ [.05]	
$NP \rightarrow Pronoun$	[.35]	$flight$ [.40] $dinner$ [.10]	
$NP \rightarrow Proper-Noun$	[.30]	$Verb \rightarrow book$ [.30] $include$ [.30]	
$NP \rightarrow Det Nominal$	[.20]	$prefer$ [.40]	
$NP \rightarrow Nominal$	[.15]	$Pronoun \rightarrow I$ [.40] she [.05]	
$Nominal \rightarrow Noun$	[.75]	me [.15] you [.40]	
$Nominal \rightarrow Nominal Noun$	[.20]	$Proper-Noun \rightarrow Houston$ [.60]	
$Nominal \rightarrow Nominal PP$	[.05]	NWA [.40]	
$VP \rightarrow Verb$	[.35]	$Aux \rightarrow does$ [.60] can [.40]	
$VP \rightarrow Verb NP$	[.20]	$Preposition \rightarrow from$ [.30] to [.30]	
$VP \rightarrow Verb NP PP$	[.10]	on [.20] $near$ [.15]	
$VP \rightarrow Verb PP$	[.15]	$through$ [.05]	
$VP \rightarrow Verb NP NP$	[.05]		
$VP \rightarrow VP PP$	[.15]		
$PP \rightarrow Preposition NP$	[1.0]		

Beispiel 2: Lesarten



Linker Baum $T_1: p(T_1) = 2.2 \cdot 10^{-6}$

Rechter Baum $T_2: p(T_2) = 6.1 \cdot 10^{-7}$

- Probabilistische kontextfreie Grammatiken
 - Kontextfreie Regeln mit Wahrscheinlichkeiten
 - Wahrscheinlichkeiten aller Regeln mit gleicher linker Seite summieren zu Eins
 - Wahrscheinlichkeit eines Parses ist das Produkt der Wahrscheinlichkeiten aller Regeln im Parse
- Disambiguierung mit PCFGs
- Naive Berechnung: wieder viele doppelte Berechnungen. Integriere in CYK.

- 1 PCFGs
- 2 Probabilistischer CYK**
- 3 Woher bekomme ich die Wahrscheinlichkeiten?
- 4 PCFG-Verbesserungen
 - PCFG Eigenschaften
 - Problem 1: Unabhängigkeit der Regelwahrscheinlichkeiten vom Kontext
 - Problem 2: Fehlende Lexikalisierung
- 5 Parserevaluation

Probabilistischer CYK für $s = w_1, \dots, w_n$

- 1 Konvertiere probabilistische PCFG in probabilistische PCFG in Chomsky-Normalform (CNF)
- 2 Initialisiere Chart auf der Diagonale mit POS Tags mit Wahrscheinlichkeiten $p(w_i|t_j)$ für alle $i \in \{1, \dots, n\}$ und alle mögliche Tags t_j
- 3 Wende CYK an wie nicht-probabilistisch. Entsteht ein neues Chart-item aus der Regel $R: C \rightarrow A B$ $p(R)$, dann setze

$$p(C) = p(A) \cdot p(B) \cdot p(R)$$

- 4 Bei mehreren Ableitungen für **gleichen Konstituententyp** behalten wir diejenige mit der größten Wahrscheinlichkeit
- 5 Benutze Backpointer wie üblich
- 6 Behalte wahrscheinlichste Lesart mit Konstituentente S aus letzter Zelle $[0, n]$

S → NP VP 1.0

PP → IN NP 1.0

VP → VB NP 0.7

VP → VP PP 0.3

Ist schon in CNF!

NP → NP PP 0.4

NP → astronomers 0.2

NP → women 0.2

NP → telescopes 0.2

IN → with 1.0

VB → meet 1.0

Beispiel 1: Prob CYK

astronomers	meet	women	with	telescopes
NP 0.2				
	VB 1.0			
		NP 0.2		
			IN 1.0	
				NP 0.2

Beispiel 1: Prob CYK

astronomers	meet	women	with	telescopes
NP 0.2	-	S $0.2 \cdot 0.14 \cdot 1.0 = 0.028$	-	
	VB 1.0	VP $1.0 \cdot 0.2 \cdot 0.7 = 0.14$	-	
		NP 0.2	-	NP $0.2 \cdot 0.2 \cdot 0.4 = 0.016$
			IN 1.0	PP $1.0 \cdot 0.2 \cdot 1.0 = 0.2$
				NP 0.2

Beispiel 1: Prob CYK

astronomers	meet	women	with	telescopes
NP 0.2	-	S $0.2 \cdot 0.14 \cdot 1.0 = 0.028$	-	
	VB 1.0	VP $1.0 \cdot 0.2 \cdot 0.7 = 0.14$	-	VP1 $1.0 \cdot 0.016 \cdot 0.7 = 0.0112$ VP2 $0.14 \cdot 0.2 \cdot 0.3 = 0.0084$
		NP 0.2	-	NP $0.2 \cdot 0.2 \cdot 0.4 = 0.016$
			IN 1.0	PP $1.0 \cdot 0.2 \cdot 1.0 = 0.2$
				NP 0.2

Wir müssen nur VP1 behalten (gleiche Konstituente), wenn wir nur den besten Parse wollen. Wollten wir alle Parses mit allen Wahrscheinlichkeiten, müssten wir mit beiden VPs weiterrechnen. ▶

Beispiel 1: Prob CYK

astronomers	meet	women	with	telescopes
NP 0.2	-	S $0.2 \cdot 0.14 \cdot 1.0 = 0.028$	-	S1 $0.2 \cdot 0.0112 \cdot 1.0 = 0.00224$
	VB 1.0	VP $1.0 \cdot 0.2 \cdot 0.7 = 0.14$	-	VP1 $1.0 \cdot 0.016 \cdot 0.7 = 0.0112$
		NP 0.2	-	NP $0.2 \cdot 0.2 \cdot 0.4 = 0.016$
			IN 1.0	PP $1.0 \cdot 0.2 \cdot 1.0 = 0.2$
				NP 0.2

präferiert lokales attachment (*women have the telescopes*) und gibt gleiche Lösung wie die naive Berechnung.

Beispiel 2: Grammatik

Grammar		Lexicon	
$S \rightarrow NP VP$	[.80]	$Det \rightarrow that$ [.10] a [.30] the [.60]	
$S \rightarrow Aux NP VP$	[.15]	$Noun \rightarrow book$ [.10] $flight$ [.30]	
$S \rightarrow VP$	[.05]	$meal$ [.05] $money$ [.05]	
$NP \rightarrow Pronoun$	[.35]	$flight$ [.40] $dinner$ [.10]	
$NP \rightarrow Proper-Noun$	[.30]	$Verb \rightarrow book$ [.30] $include$ [.30]	
$NP \rightarrow Det Nominal$	[.20]	$prefer$ [.40]	
$NP \rightarrow Nominal$	[.15]	$Pronoun \rightarrow I$ [.40] she [.05]	
$Nominal \rightarrow Noun$	[.75]	me [.15] you [.40]	
$Nominal \rightarrow Nominal Noun$	[.20]	$Proper-Noun \rightarrow Houston$ [.60]	
$Nominal \rightarrow Nominal PP$	[.05]	NWA [.40]	
$VP \rightarrow Verb$	[.35]	$Aux \rightarrow does$ [.60] can [.40]	
$VP \rightarrow Verb NP$	[.20]	$Preposition \rightarrow from$ [.30] to [.30]	
$VP \rightarrow Verb NP PP$	[.10]	on [.20] $near$ [.15]	
$VP \rightarrow Verb PP$	[.15]	$through$ [.05]	
$VP \rightarrow Verb NP NP$	[.05]		
$VP \rightarrow VP PP$	[.15]		
$PP \rightarrow Preposition NP$	[1.0]		

Großbuchstaben stehen für Nichtterminale.

- ➊ $A \rightarrow B C D [p]$ wird zu $A \rightarrow B X_1 [p], X_1 \rightarrow C D [1.0]$
- ➋ $A \rightarrow B [p]$ sowie $B \rightarrow C D [q]$ wird zu $A \rightarrow C D [p \cdot q]$
Dies muss eventuell rekursiv angewendet werden.
- ➌ $A \rightarrow B \gamma [p]$ mit Terminal γ wird zu $A \rightarrow B X_2 [p]; X_2 \rightarrow \gamma [1.0]$

Bitte an Beispiel 2 üben (vollständige Lösung online und als Handout)

Großbuchstaben stehen für Nichtterminale.

- 1 $A \rightarrow B C D [p]$ wird zu $A \rightarrow B X_1 [p], X_1 \rightarrow C D [1.0]$
- 2 $A \rightarrow B [p]$ sowie $B \rightarrow C D [q]$ wird zu $A \rightarrow C D [p \cdot q]$
Dies muss eventuell rekursiv angewendet werden.
- 3 $A \rightarrow B \gamma [p]$ mit Terminal γ wird zu $A \rightarrow B X_2 [p]; X_2 \rightarrow \gamma [1.0]$

Bitte an Beispiel 2 üben (vollständige Lösung online und als Handout)

- 1 PCFGs
- 2 Probabilistischer CYK
- 3 Woher bekomme ich die Wahrscheinlichkeiten?**
- 4 PCFG-Verbesserungen
 - PCFG Eigenschaften
 - Problem 1: Unabhängigkeit der Regelwahrscheinlichkeiten vom Kontext
 - Problem 2: Fehlende Lexikalisierung
- 5 Parserevaluation

Baumbank

Korpus, das für jeden Satz die korrekte syntaktische Analyse enthält (wie die Penn Treebank).

```
((S
  (NP-SBJ (DT That)
    (JJ cold) (, ,)
    (JJ empty) (NN sky) )
  (VP (VBD was)
    (ADJP-PRD (JJ full)
      (PP (IN of)
        (NP (NN fire)
          (CC and)
          (NN light) ))))
  (. .) ))

((S
  (NP-SBJ The/DT flight/NN )
  (VP should/MD
    (VP arrive/VB
      (PP-TMP at/IN
        (NP eleven/CD a.m/RB
          (NP-TMP tomorrow/NN ))

```

- Penn III Treebank (WSJ): 1m Tokens
- OntoNotes (2007-2013): News, broadcast, Web, Telephone. ca. 1.5m Tokens
- English Web Treebank: 254K Tokens aus Blogs, Newsgroups, emails, reviews, FAQ

- Negra-Korpus (1997-1999, deutsch): Zeitung, 355K Tokens
- Tiger-Korpus (2002/2003, deutsch): Zeitung, 900K Tokens
- Tüba/DZ (2002/2013, deutsch): Zeitung, 1.6 m Tokens
- French treebank (1997-2004): Zeitung, 1m Tokens
- Penn Chinese/Arabic treebanks
- Ancora (Spanisch/Katalan)
- ...

Oft erhältlich über das Linguistic Data Consortium (LDC)

Extrahiere aus den Baumstrukturen alle Teilbäume der Tiefe 1 und kodiere sie als kontextfreie Regeln.

```
((S
  (NP-SBJ (DT That)
    (JJ cold) (, ,)
    (JJ empty) (NN sky) )
  (VP (VBD was)
    (ADJP-PRD (JJ full)
      (PP (IN of)
        (NP (NN fire)
          (CC and)
          (NN light) ))))
  (. .) ))
```

```
S → NP VP
NP → DT JJ , JJ N
VP → VBD ADJP
ADJP → JJ PP
PP → IN NP
NP → NN CC NN
plus Terminalregeln
```

- Große Regelmengen (ca. 17 500 Regeln ohne Terminalregeln aus Penn treebank)
- Hohe Kombinatorik
- Sehr ambig

Beispiel (Regeln für Adjektiphrasen in TIGER):

- AP → ADV ADJD (1007)
- AP → ADV CARD (914)
- AP → ADV NM (516)
- AP → ADJD ADJD (372)
- ...
- AP → ADV ADJD PP (23)
- ...
- AP → ADJD ADV ADJD (3)

Maximum-Likelihoodschätzung für Regelwahrscheinlichkeiten:

$$p(A \rightarrow \alpha | A) = \frac{f(A \rightarrow \alpha)}{f(A)}$$

wobei $A \in NT$ N und $\alpha \in (\Sigma \cup NT)^*$

- Baumbanken erlauben Grammatikinduktion, inklusive Wahrscheinlichkeitschätzungen für Regeln
- Diese können dann beim Parsing von neuen Sätzen benutzt werden
- CYK kann leicht um eine probabilistische Komponente erweitert werden

- 1 PCFGs
- 2 Probabilistischer CYK
- 3 Woher bekomme ich die Wahrscheinlichkeiten?
- 4 PCFG-Verbesserungen**
 - PCFG Eigenschaften
 - Problem 1: Unabhängigkeit der Regelwahrscheinlichkeiten vom Kontext
 - Problem 2: Fehlende Lexikalisierung
- 5 Parserevaluation

- Einige PCFG Eigenschaften
- Probleme mit PCFGs: Verbesserungen durch Splitting und Lexikalisierung
- Demos

PCFGs sind robust und können zur Disambiguierung (oder als Sprachmodell) verwendet werden. Aber:

- Wahrscheinlichkeit eines Sub-Baums **unabhängig von Stelle im Satz** (d.h. $NP \rightarrow PRP$ gleiche Wahrscheinlichkeit, ob Anfang oder Ende des Satzes)
- **Kontextfrei**: Wahrscheinlichkeit eines Sub-Baums unabhängig von Worten außerhalb des Baumes
- Wahrscheinlichkeit eines Sub-Baums **unabhängig von Knoten außerhalb** des Baumes.

Wahrscheinlichkeit eines Sub-Baums unabhängig von externen Knoten

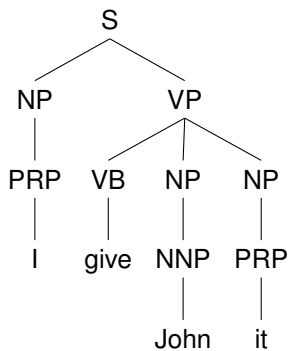
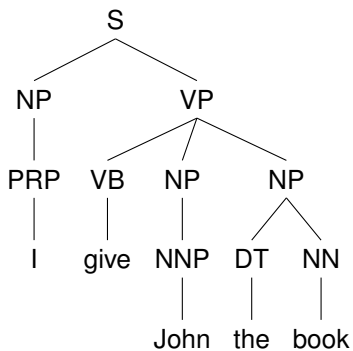
Wahr: Externe Umgebung beeinflusst interne Expansion eines Knoten.

Beispiel Switchboard Korpus:

	Pronomen	Andere NP
Subjekt	91%	9%
Objekt	34%	66%

oder

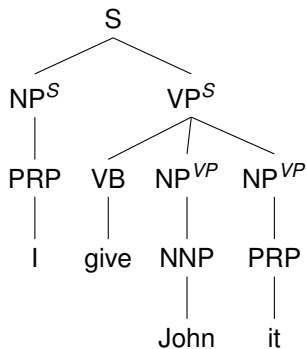
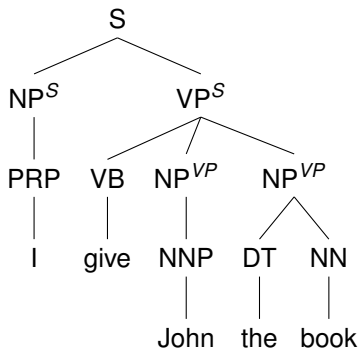
Regel	indirektes Objekt	direktes Objekt
NP → NNS	7.5%	0.2%
NP → PRP	13.4%	0.9%
NP → NP PP	12.2%	14.4%
NP → DT NN	10.4%	13.3%
NP → JJ NN	1.1%	10.4%



Lösung Verfeinerung/splitting Teil I

Johnson, 1998: PCFG models of linguistic tree representations

Parent annotation (nur für Phrasen-Nicht-Terminale):



Klein und Manning, 2003: Accurate unlexicalised parsing

Für alle Nicht-Terminale (also auch POS tags)

Beispiel: Häufigste Adverbien unter folgenden Knoten:

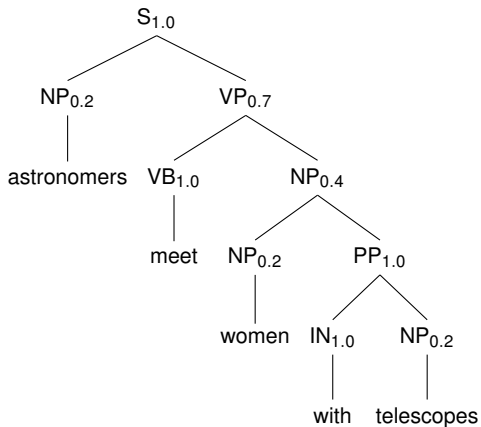
ADVP *also, now*

VP *not, n't*

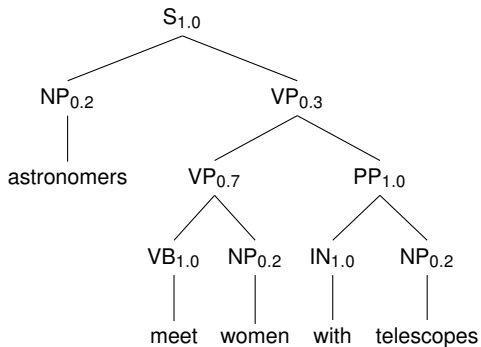
NP *only, just*

Beispiel (Manning:Schuetze:99)

S → NP VP	1.0	NP → NP PP	0.4
PP → IN NP	1.0	NP → astronomers	0.2
VP → VB NP	0.7	NP → women	0.2
VP → VP PP	0.3	NP → telescopes	0.2
IN → with	1.0	VB → meet	1.0



$$P(t_1) = 1.0 \times 0.2 \times 0.7 \times \\ 1.0 \times 0.4 \times 0.2 \times 1.0 \times \\ 1.0 \times 0.2 = 0.0024$$



$$P(t_2) = 1.0 \times 0.2 \times 0.3 \times 0.7 \times 1.0 \times 0.2 \times 1.0 \times 1.0 \times 0.2 = 0.00168$$

t_1 wahrscheinlicher als t_2 : Präferenz für lokales attachment

Was passiert, wenn das Verb sich zu "observe" ändert?

PCFG: Wahrscheinlichkeit eines Sub-Baums unabhängig von Worten ausserhalb oder innerhalb des Baumes

Richtig: Verben haben Subkategorisierungsrahmen (subcat frames).
Manche dieser frames sind präferiert.

- The women discussed the dogs on the beach.
 - The women discussed the dogs which were on the beach. (90%)
 - The women discussed them (the dogs) while on the beach. (10%)
- The women kept the dogs on the beach.
 - The women kept the dogs which were on the beach. (5%)
 - The women kept them (the dogs) while on the beach. (95%)

Results from Ford:ea:82 rating study.

- John wanted the shirt on the rack.
- John positioned the shirt on the rack.

- I lost the ticket to Berlin.
- I mailed the ticket to Berlin.

Für *keep*:

NP AP	keep the prices reasonable
NP VP	keep his foes guessing
NP VP	keep their eyes closed
NP PRT	keep the people in
NP PP	keep his nerves from jangling

Subcat Wahrscheinlichkeiten sagen, wie wahrscheinlich ein bestimmter Frame ist. Unterscheiden sich stark per Wort.

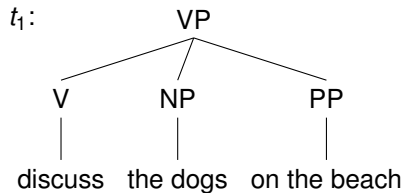
Aus Treebank.

Beispiel:

discuss	$\langle \text{NP PP} \rangle$.24
	$\langle \text{NP} \rangle$.76
keep	$\langle \text{NP PP} \rangle$.81
	$\langle \text{NP} \rangle$.19

$$p(\text{discuss}, \langle \text{NP PP} \rangle) = 0.24$$

$$\text{VP} \rightarrow \text{V NP PP} \quad 0.15$$

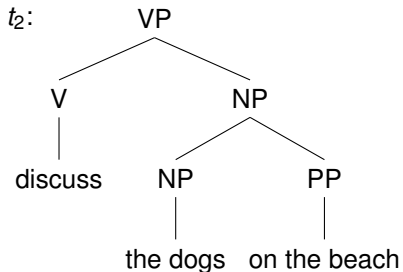


$$p(t_1) = 0.15 \times 0.24 = 0.036 \text{ (dispreferred)}$$

$$p(\text{discuss}, \langle \text{NP} \rangle) = 0.76$$

$$\text{VP} \rightarrow \text{V NP} \quad 0.39$$

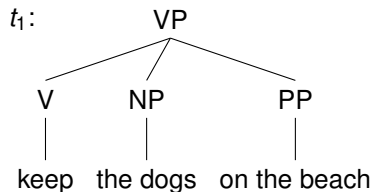
$$\text{NP} \rightarrow \text{NP PP} \quad 0.14$$



$$p(t_2) = 0.76 \times 0.39 \times 0.14 = 0.041 \text{ (preferred)}$$

$$p(\text{keep}, \langle \text{NP PP} \rangle) = 0.81$$

$$\text{VP} \rightarrow \text{V NP PP} \quad 0.15$$

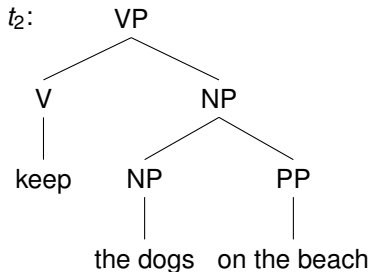


$$p(t_1) = 0.15 \times 0.81 = 0.12 \text{ (preferred)}$$

$$p(\text{keep}, \langle \text{NP} \rangle) = 0.19$$

$$\text{VP} \rightarrow \text{V NP} \quad 0.39$$

$$\text{NP} \rightarrow \text{NP PP} \quad 0.14$$



$$p(t_2) = 0.19 \times 0.39 \times 0.14 = 0.01 \text{ (dispreferred)}$$

Dies war eine stark vereinfachte Darstellung! Man kann viel weitergehend lexikalisieren. Dann: wegen Schätzproblemen markovisieren wir oder machen Unabhängigkeitsannahmen in einem generativem Modell...

Zudem sollten die subcat Wahrscheinlichkeiten mathematisch richtig in die Regelwahrscheinlichkeiten integriert werden.

Siehe auch Jurafsky und Martin, 3rd online edition, Kapitel 14.6. Collins Parser

Problematisch für PCFGs:

- NP Konjunktionen und PPs: *dogs in houses and cats vs. men in jackets and ties*
- NP Konjunktionen und Komposita: *university lecturers and administrators vs. dog collars and handcuffs*

- PCFGs gebraucht nur strukturelle Wahrscheinlichkeiten mit vielen Unabhängigkeitsannahmen
- Splitting und Parent Annotation: Bäume abhängig von externen Knoten
- Lexikalisierung: Bäume abhängig von Worten

Ein State-of-the art Parser für CFGs: benutzt PCFGs, ist lexikalisiert.
Stanford Toolkit hat auch einen neuronalen Dependenzparser.

`www.corenlp.run`

Berkely parser (mit automatischem Splitting, vor allem auch für
Wortarten):

`http://tomato.banatao.berkeley.edu:
8080/parser/parser.html`

- 1 PCFGs
- 2 Probabilistischer CYK
- 3 Woher bekomme ich die Wahrscheinlichkeiten?
- 4 PCFG-Verbesserungen
 - PCFG Eigenschaften
 - Problem 1: Unabhängigkeit der Regelwahrscheinlichkeiten vom Kontext
 - Problem 2: Fehlende Lexikalisierung
- 5 Parserevaluation

- “Trainiere” auf einer Baumbank (d.h. extrahiere Regeln und Wahrscheinlichkeiten)
- Teste auf ungesehenen Sätzen (mit Goldparses)

Evaluationsmaße:

- Accuracy: Ist Systemparse = Goldparse?
- Labelled Precision/Recall/F-measure: auf Konstituentenbasis

① Übereinstimmung: Kategorie, Anfangs- und Endpunkt

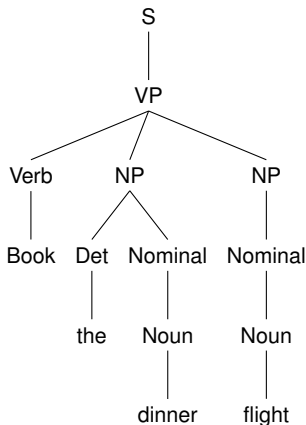
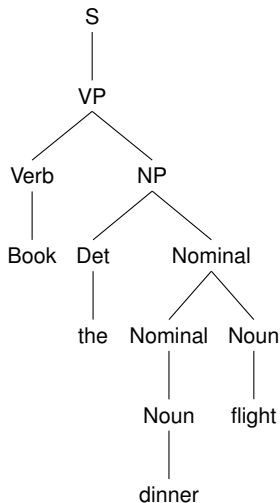
②

$$\text{LabeledPrec} := \frac{\# \text{correct constituents in system parse}}{\# \text{total constituents in system parse}}$$

③

$$\text{LabeledRec} := \frac{\# \text{correct constituents in system parse}}{\# \text{constituents in gold parse}}$$

- Unlabelled Precision/Recall: Übereinstimmung beschränkt auf Anfangs/ und Endzeitpunkt



GP (links): S (0:4), VP (0:4), NP(1:4), Nominal (2:4), **Nominal (2:3)**

SP (rechts): S (0:4), VP (0:4), NP (1:3), **Nominal (2:3)**, NP (3:4),

Nominal (3:4)

- Zuerst verwandt in PARSEVAL (1991) als unlabeled Version, mit einigen Normalisierungen und unter Vernachlässigung von unitären Branches
- Man lässt immer die POS Tags bei der Evaluation weg (Warum?)
- Je flacher die Regeln, desto einfacher das Parsing...
- Korrelation mit Anwendungen?
- Weiteres schönes Beispiel in Figure 12.6 in Manning und Schütze

	F1 (Penn treebank)
Baumbank-PCFG	0.72-0.72
Manuelle Verfeinerung	0.85-0.86
Mit Lexikalisierung	0.87 -0.89
Gelernter Verfeinerung	0.89-0.90
Dyer et al (NAACL 2016) RNN	0.92
Kitaev et al (ACL 2019) NN	0.95

- * Jurafsky und Martin, 3rd online edition, Kapitel 14 (außer 14.7)
- Manning und Schütze, 1999. *Foundations of Statistical NLP*. Kapitel 11, 12.
- Stanford Parser und Papiere: <https://nlp.stanford.edu/software/lex-parser.shtml>
- Berkely PCFG parser: <https://github.com/slavpetrov/berkeleyparser>. Demo bei: <http://tomato.banatao.berkeley.edu:8080/parser/parser.html>
- Jetziger Berkely Neural Parser: <https://github.com/nikitakit/self-attentive-parser>. Demo: <https://parser.kitaev.io/>
- Übungsblatt 9

- Ford et al (1982): A competence-based theory of syntactic closure. In Bresnan et al *The mental representation of grammatical relations*
- Black et al (1991): A procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proc. DARPA Speech and Natural Language Workshop*
- Sekine und Collins (1997): The evalb software.
<https://nlp.cs.nyu.edu/evalb/>
- Collins (1999): Head-driven Statistical Methods for Natural Language Parsing. PhD Thesis
- Klein und Manning (2003): Accurate unlexicalized parsing. In: *NAACL 2003*
- Johnson (1998): PCFG models of linguistic tree representations. In: *Computational Linguistics*
- Kitaev et al. (2019): Multilingual Constituency Parsing with Self-attention and pre-Training. In ACL.