Ausgangsfrage Syntax des Deutschen + Negra Probabilistische Parser Experiment 1 Experiment 2 Schlussfolgerung I literatur

# PROBABILISTIC PARSING FOR GERMAN USING SISTER-HEAD DEPENDENCIES

Irina Gossmann Carine Dombou

9. Juli 2007



#### **INHALT**

- AUSGANGSFRAGE
- 2 SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** Experiment 1
- **S** EXPERIMENT 2
- **6** SCHLUSSFOLGERUNG
- LITERATUR



#### **INHALT**

- AUSGANGSFRAGE
- SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** EXPERIMENT 1
- **SEXPERIMENT 2**
- 6 SCHLUSSFOLGERUNG
- 1 LITERATUR

#### AUSGANGSFRAGE

- Lexikalisierte Parser sind um etwa 10% besser, als unlexikalisierte Parser
- Aber: die meisten lexikalisierten Parser wurden für Englisch/Penn Treebank entwickelt
- Frage: was wären die Ergebnisse bei anderen Sprachen/Annotationsstilen?

# Lexikalisiertes Parsing bei anderen Sprachen

Language				
English	40,000	87.4%	88.1%	(Collins, 1997)
Chinese	3,484	69.0%	74.8%	(Bikel and Chiang, 2000)
Czech				(Collins et al., 1999)

Table 1: Results for the Collins (1997) model for various languages (dependency precision for Czech)

#### **INHALT**

- AUSGANGSFRAGE
- 2 SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** EXPERIMENT 1
- **SEXPERIMENT 2**
- 6 SCHLUSSFOLGERUNG
- 1 LITERATUR

#### **INHALT**

- AUSGANGSFRAGE
- SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** EXPERIMENT 1
- **S** EXPERIMENT 2
- 6 SCHLUSSFOLGERUNG
- 1 LITERATUR

1. Unlexikalisierte PCFG (Charniak 1993)

Die lexikalisierten Parser werden gegen einen einfachen PCFG-Parser geprüft (Baseline).

$$p(LHS o RHS) = p(RHS|LHS) = rac{ ext{alle Vorkommen von rule}}{ ext{Regeln mit gleicher linken Seite}}$$

Die Summe der W'keiten aller Regeln mit derselben LHS ist 1.  $p(Baum\ T) = Produkt\ der\ W'keiten\ aller in\ T\ angewandten\ Regeln.$ 

2. Carrol und Rooth's lexikalisiertes Modell (1998)

Entspricht Charniaks Modell (1997).

$$p(rule) = p(RHS|LHS) = p_{rule}(C_1...C_n|P, I(P)) *$$

$$\prod_{i=1}^{n} p_{choice}(I(C_i)|C_i, P, I(P))$$

für eine Regel der Form P  $\rightarrow$   $C_1...Cn$ .

I(P) ist Kopf der Eltern-Kategorie,  $I(C_i)$  ist Kopf einer der Tochter-Kategorien



#### 3. Collins' lexikalisiertes Modell

#### Regel:

$$\boxed{P} \rightarrow \boxed{L_m \dots L_1} \boxed{H} \boxed{R_1 \dots R_n}$$

#### Regelw'keit

#### 3. Collins' lexikalisiertes Modell

#### Regel:

$$\boxed{P} \rightarrow \boxed{L_m}...\boxed{L_1} \boxed{H} \boxed{R_1}...\boxed{R_n}$$

#### Regelw'keit:

#### 3. Collins' lexikalisiertes Modell

#### Regel:

$$\boxed{P} \rightarrow \boxed{L_m}...\boxed{L_1} \boxed{H} \boxed{R_1}...\boxed{R_n}$$

#### Regelw'keit:

$$= P\left(\begin{bmatrix} H \\ t_P \\ t_P \\ w_P \end{bmatrix}\right)$$

$$\cdot \left[\prod_{i=0}^m P\left(\begin{bmatrix} L_i \\ t_{L_i} \\ w_{L_i} \end{bmatrix}, d(i) \middle| \begin{bmatrix} P \\ t_P \\ w_P \end{bmatrix} \middle| \end{bmatrix}\right]$$

$$\cdot \left[\prod_{i=0}^n P\left(\begin{bmatrix} R_i \\ t_{R_i} \\ w_{R_i} \end{bmatrix}, d(i) \middle| \begin{bmatrix} P \\ t_P \\ w_P \end{bmatrix} \middle| \end{bmatrix}\right]$$

w ist der Kopf der Konstituente,t ist Tag des Kopfs der Konstituente,d misst die Entfernung vom Kopf

## HEAD-HEAD VS. SISTER-HEAD

#### W'keit einer Konstituente

$$\prod_{i=0}^{n} p_{l}(L_{i}, t(L_{i}), I(L_{i})|P, H, t(H), d(H), d(i))$$

$$\prod_{i=0}^{n} p_{I}(L_{i}, t(L_{i}), I(L_{i})|P, L_{i-1}, t(L_{i-1}), d(L_{i-1}), d(i))$$

#### **INHALT**

- AUSGANGSFRAGE
- 2 SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** Experiment 1
- **5** EXPERIMENT 2
- 6 SCHLUSSFOLGERUNG
- 1 LITERATUR



#### VERGLEICH DER 3 PARSER

#### Hypothese:

- Lexikalisierung verbessert die Ergebnisse.
- weitere Verbesserungen der Ergebnisse durch Berürcksichtigung der zusätzlichen Informationen in Negra (grammatical function labels, coordinate categories, flache PPs).

#### VORBEREITUNG DES KORPUS

- Es wurde der Treebank-Format von Negra genommen.
- Das Korpus wurde aufgeteilt: Training (18602 Bäume),
   Development (1000 Bäume), Testing (1000 Bäume).
- In Development- und Test-Abschnitte wurden nur Sätze unter 40 Wörtern aufgenommen.

#### **GRAMMAR INDUCTION**

#### Baseline:

- Grammatik + Lexikon: wurden aus dem Korpus ausgelesen, nachdem alle grammatical function labels und Spuren entfernt wurden (nicht lexikalisiert).
- Parser: probabilistischer left-corner-Parser Lopar (unlexicalized mode).

#### **GRAMMAR INDUCTION**

#### Carrol und Rooth:

- Grammatik + Lexikon: wurden aus dem Korpus ausgelesen; die Köpfe der Kategorien S, VP, AP und AVP sind in Negra bereits markiert, für alle anderen Kategorien mussten die Köpfe manuell annotiert werden.
- Parser: Lopar (lexicalized mode).

#### **GRAMMAR INDUCTION**

#### Collins:

 Grammatik + Lexikon: wie bei CARROL UND ROOTH.

Abweichungen vom Original-Modell von Collins: alle leeren Kategoien (Spuren) wurden entfernt (das Modell der Autoren kann damit nicht umgehen), für nicht-rekursive NPs wurden head-head-W'keiten genommen.

 Parser: implementiert von einem der Autoren.

# Training + Testing (1)

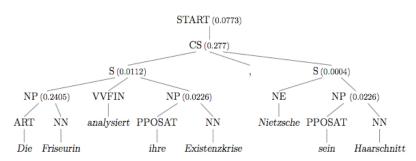
Bei den lexikalisierten Modellen hat man mit Smoothing nachgeholfen.

#### Variationen:

- Baseline und Carrol und Rooth: ± grammatical function labels
- $\bullet$  Carrol und Rooth:  $\pm$  PARAMETER POOLING (Berücksichtigung von coordinate categories und flachen PPs in Negra)

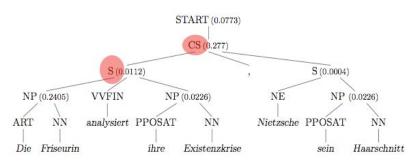
# Training + Testing (1) - Parameter Pooling

Pooling: Zusammenlegen von Kategorien (dabei wächst die W'keit der einzelnen Regeln).



# Training + Testing (1) - Parameter Pooling

Pooling: Zusammenlegen von Kategorien (dabei wächst die W'keit der einzelnen Regeln).



# Training + Testing (2)

#### Verschiedene Part-of-Speech Tags:

- Der test-set wurde für alle Parser mit dem POS-Tagger TnT getaggt, der zuvor an dem training-set trainiert wurde.
   Die Genauigkeit, gemessen am development-set, war 97,12%.
- Perfect tagging mode:
   Es wurden auch Tests mit korrekten POS-Tags (gold standard)
   durchgeführt, um die obere Performance-Grenze der Parser zu bestimmen.

## **AUSWERTUNG MIT PARSEVAL SCORES**

- LR: korrekte Klamern im Parse alle Klammern im Baum
- LP: korrekte Klammern im Parse alle Klammern im Parse
- CBs: Durchschnitt der sich kreuzenden Klammern pro Baum.
- 0CB: Prozentsatz der Bäume ohne sich kreuzende Klammern.
- ≤2CB : Prozentsatz der Bäume mit zwei oder weniger sich kreuzenden Klammern.
- Cov.: Prozentsatz der Sätze, die geparst werden konnten.



		TnT tagging						Perfect tagging					
	LR	LP			≤2CB	Cov	LR	LP	CBs	0CB	≤2CB	Cov	
Baseline		66.69			84.46		72.99	70.00	0.88	60.30	87.42	95.25	
Baseline + GF	70.45	65.49	1.07	58.02	85.01	79.24	81.14	78.37	0.46	74.25	95.26	65.39	
C&R	68.04	60.07	1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25	
C&R + pool C&R + GF	69.07	61.41	1.28	53.06	80.09	94.42	71.74	64.73	1.11	56.40	83.08	95.25	
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39	
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23	

		TnT tagging LR LP CBs 0CB ≤2CB Cov						Perfect tagging					
	LR	LP	CBs	0ČB		Cov	LR	LP	CBs	0CB	≤2CB		
Baseline		66.69					72.99	70.00	0.88	60.30	87.42	95.25	
Baseline + GF	70.45	65.49	1.07	58.02	85.01	79.24	81.14	78.37	0.46	74.25	95.26	65.39	
C&R	68.04	60.07	1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25	
C&R + pool C&R + GF	69.07	61.41	1.28	53.06	80.09	94.42	71.74	64.73	1.11	56.40	83.08	95.25	
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39	
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23	

			m.r	P 4			Desferred to a significant						
				Γ taggir			Perfect tagging						
	LR	LP	CBs	0ČB	≤2CB	Cov	LR	$_{ m LP}$			≤2CB		
Baseline	70.56	66.69	1.03	58.21	84.46	94.42	72.99	70.00	0.88	60.30	87.42	95.25	
Baseline + GF					85.01						95.26		
C&R	68.04	60.07	1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25	
C&R + pool C&R + GF		61.41			80.09						83.08		
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39	
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23	

		TnT tagging LR LP CBs 0CB ≤2CB Cov						Perfect tagging					
	LR	LP	CBs	0ČB		Cov	LR	LP	CBs	0CB	≤2CB	Cov	
Baseline		66.69					72.99	70.00	0.88	60.30	87.42	95.25	
Baseline + GF	70.45	65.49	1.07	58.02	85.01	79.24	81.14	78.37	0.46	74.25	95.26	65.39	
C&R	68.04	60.07	1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25	
C&R + pool C&R + GF	69.07	61.41	1.28	53.06	80.09	94.42	71.74	64.73	1.11	56.40	83.08	95.25	
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39	
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23	

			Tn	Γ taggiı	ıg		Perfect tagging						
	LR	LP			≤2CB	Cov	LR	LP	CBs	0CB	≤2CB		
Baseline		66.69					72.99	70.00	0.88	60.30	87.42	95.25	
Baseline + GF					85.01						95.26		
C&R	68.04	60.07	1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25	
C&R + pool	69.07	61.41	1.28	53.06	80.09	94.42	71.74	64.73	1.11	56.40	83.08	95.25	
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39	
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23	

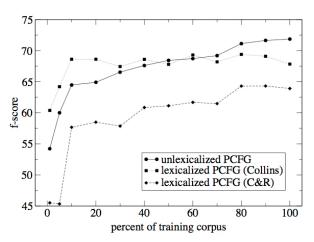
		TnT tagging						Perfect tagging						
	LR	LP	CBs	0CB	≤2CB	Cov	LR	LP	CBs	0CB	°≤2CB	Cov		
Baseline	70.56	66.69	1.03	58.21	84.46	94.42	72.99	70.00	0.88	60.30	87.42	95.25		
Baseline + GF	70.45	65.49	1.07	58.02	85.01	79.24	81.14	78.37	0.46	74.25	95.26	65.39		
C&R	68.04	60.07	1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25		
C&R + pool	69.07	61.41	1.28	53.06	80.09	94.42	71.74	64.73	1.11	56.40	83.08	95.25		
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39		
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23		

		TnT tagging						Perfect tagging					
	LR	LP	CBs	0ČB	≤2CB			LP	CBs	0CB	≤2CB		
Baseline					84.46							95.25	
Baseline + GF	70.45	65.49	1.07	58.02	85.01	79.24	81.14	78.37	0.46	74.25	95.26	65.39	
C&R	68.04	60.07	1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25	
C&R + pool C&R + GF	69.07	61.41	1.28	53.06	80.09	94.42	71.74	64.73	1.11	56.40	83.08	95.25	
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39	
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23	

			Γ taggir	ıg		Perfect tagging						
	LR	LP	CBs	Γ taggir 0CB	≤2CB	Cov	LR	$_{ m LP}$	CBs	0CB	$\leq 2CB$	Cov
Baseline	70.56	66.69			84.46		72.99				87.42	
Baseline + GF	, 0				85.01						95.26	
C&R			1.31	52.08	79.54	94.42	70.79	63.38	1.17	54.99	82.21	95.25
C&R + pool	69.07	61.41	1.28	53.06	80.09	94.42	71.74	64.73	1.11	56.40	83.08	95.25
C&R + GF	67.66	60.33	1.31	55.67	80.18	79.24	81.17	76.83	0.48	73.46	94.15	65.39
Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23

- 1. Berücksichtigung von grammatical functions hat die Ergebnisse sogar verschlechtert.
- 2. Parameter pooling verbesserte die Ergebnisse etwas.
- 3. Lexikalisierung hat die Ergebnisse verschlechtert (wie die Lernkurven zeigen, ist der kleine Korpusumfang nicht die Ursache dafür).

## LERNKURVEN



Ausgangsfrage
Syntax des Deutschen + Negra
Probabilistische Parser
Experiment 1
Experiment 2
Schlussfolgerung
Literatur

Also: Die Hypothese ist falsch!

#### **INHALT**

- AUSGANGSFRAGE
- SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** EXPERIMENT 1
- **S** EXPERIMENT 2
- 6 SCHLUSSFOLGERUNG
- LITERATUR



#### **INHALT**

- AUSGANGSFRAGE
- 2 SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** EXPERIMENT 1
- **SEXPERIMENT 2**
- **6** SCHLUSSFOLGERUNG
- LITERATUR

#### **SCHLUSSFOLGERUNG**

- Flache Regeln in Negra entsprechen der syntaktischen Struktur des Deutschen, insbesondere der semi-freien Wortstellung.
- Für flache Strukturen ist das sister-head-Modell besser geeignet als die head-head-Modelle.

Weiterführend: wird dieses Modell für andere Sprachen ebenfalls gute Ergebnisse liefern?

#### **INHALT**

- AUSGANGSFRAGE
- SYNTAX DES DEUTSCHEN + NEGRA
- PROBABILISTISCHE PARSER
- **4** EXPERIMENT 1
- **SEXPERIMENT 2**
- 6 SCHLUSSFOLGERUNG
- LITERATUR



#### LITERATUR

- Amit Dubey and Frank Keller, Probabilistic Parsing for German Using Sister-Head Dependencies, ACL 2003
- Amit Dubey, Statistical Parsing for German, 2004

# Deutsch parsen

## Syntaktischen Eigenschaften

- -Modellierung mit Hilfe CFG ist schwierig
- -Semi-freie Wortstellung vs. fixe Wortstellung im Englischen
- -Stellung der Komplemente (Subjekte, Objekte, Adjunkte) im Deutschen ist relativ frei
- -Verbposition zwar fix, aber abhängig vom Satztyp:

- a. Weil er gestern Musik komponiert hat. because er yesterday music composed has 'Because he has composed music yesterday.'
  - b. Hat er gestern Musik komponiert?
  - c. Er hat gestern Musik komponiert.

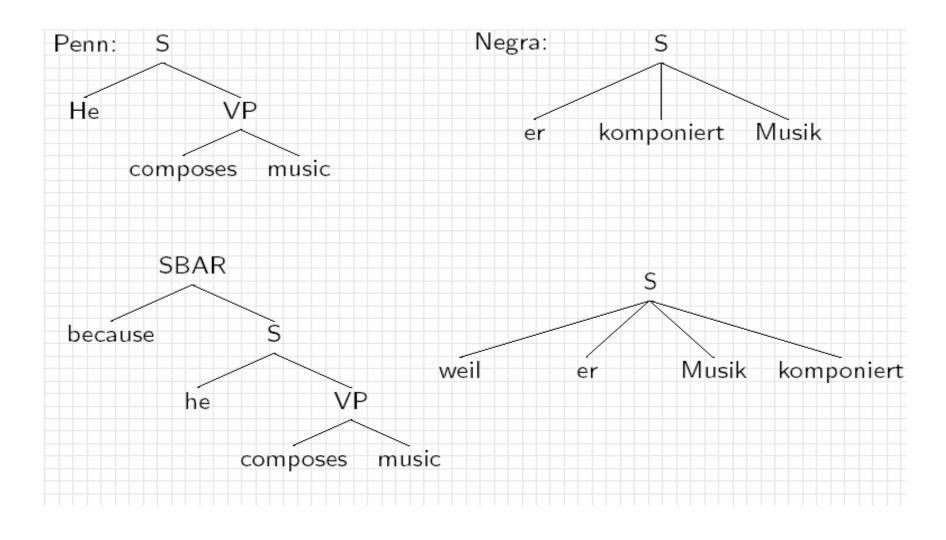
# Negra Annotationsschema (Skut et al.,1997)

- NEGRA Baumbank
  - 355.096 Tokens (20.602 handannotierte Sätze)
  - Deutsche Zeitungsartikel
- flache syntaktische Repräsentationen

```
*S \longrightarrow NP VP
S \longrightarrow NP V NP
```

- koordinierende Kategorien (CS, CNP)
- Beschriftungen für grammatische Funktionen z.B: MO(Modifier), HD(head), SB(subject), OC(clausal object)

# Penn vs Negra

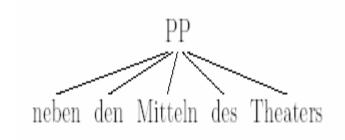


# Experiment 2

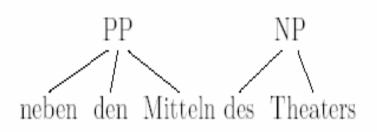
# Fehleranalyse

 Chunking Fehler aus Collins Modell

 a)[PP neben den Mitteln[NP des Theaters]]



b)[PP neben den Mitteln][NP des Theaters]



# Fehleranalyse(bis1)

- Falsche Grenzen bei PPs,Ss, und VPs
- Gründe von diesen Fehlern:
  - -Head-Head Abhängigkeit im Collins Modell
  - -Flachheit von Regeln in Negra

# Fehleranalyse(bis2)

- Negra ist flacher als Penn
- Unterschiedliche Anzahl an Töchtern

	Penn	Negra		Penn	Negra
NP	2.20	3.08	VP	2.32	2.59
PP	2.03	2.66	S	2.22	4.22

Table 3: Average number of daughters for the grammatical categories in the Penn Treebank and Negra

## Neuer Ansatz

- Hypothese:
  - -Schlechte Performanz bei lexikalisiertem Modell wegen Flachheit des Corpus

-Die Flachheit von Regeln in Negra verschlechtert die Ergebnisse

# Wie kann man diese Flachheit loswerden?

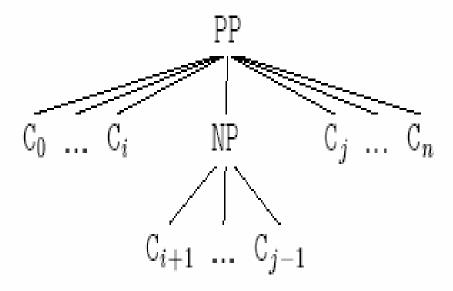
• Idee:

```
1) [PP P \dots] \rightarrow [PP P [NP \dots]]
```

### PPs bearbeiten: Algorithmus

```
for a tree node that corresponds to the rule PP \to C0 ... Cn let i = position of the last preposition, or -1 if there is no preposition let j = position of the first postposition, or n if there is postposition
```

if j-i=0 or if j-i=1 and the  $i+1^{\rm st}$  constituent is a CNP, return the rule unchanged else return



- Reduktion der Flachheit von PPs durch Einfügung von NPs
- → Denn PPs haben keine innere NPs

→ Diese Transformation wird in Negra angewendet

2) Idee 2:

Collins Modell für non- rekursive NPs und alle andere Kategorien mit Sister-Head-Abhängigkeiten testen

→Implikation binärer Zweige in der Grammatikproduktion

## Sister-head Dependendies

$$P_r(R_i, t(R_i), l(R_i)|P, R_{i-1}, t(R_{i-1}), l(R_{i-1}), d(i))$$

	C&R	Collins	Charniak	Current
Head sister category	X	X	X	
Head sister head word	X	X	X	
Head sister head tag		X	X	
Prev. sister category	X		X	X
Prev. sister head word				X
Prev. sister head tag				X

Table 4: Linguistic features in the current model compared to the models of Carroll and Rooth (1998), Collins (1997), and Charniak (2000)

## Methoden

- 1)Original Collins Modell mit modifiziertem Training Test (PPs Spaltung)trainieren
- 2)Auswahl von Modellen mit sister-head dependencies für verschiedene Kategrorien testen
  - -sister-head für NPs
  - -sister-head für PPs
  - -sister-head für alle Kategorien

### Methoden

- 3)Modell mit sister-head-Abhängigkeiten für alle Kategorien außer einer.
- →Bestimmung der Sister-Head-Abhängigkeit für die beste Performanz.

# Ergebnisse

	TnT tagging			Perfect tagging								
	LR	LP	CBs	0ĈB	≤2CB	Cov	LR	LP	CBs	0CB	≤2CB	Cov
Unmod. Collins	67.91	66.07	0.73	65.67	89.52	95.21	68.63	66.94	0.71	64.97	89.73	96.23
Split PP	73.84	73.77	0.82	62.89	88.98	95.11	75.93	75.27	0.77	65.36	89.03	93.79
Collapsed PP	66.45	66.07	0.89	66.60	87.04	95.11	68.22	67.32	0.94	66.67	85.88	93.79
Sister-head NP	67.84	65.96	0.75	65.85	88.97	95.11	71.54	70.31	0.60	68.03	93.33	94.60
Sister-head PP	70.27	68.45	0.69	66.27	90.33	94.81	73.20	72.44	0.60	68.53	93.21	94.50
Sister-head all	71.32	70.93	0.61	69.53	91.72	95.92	73.93	74.24	0.54	72.30	93.47	95.21

Results for Experiment 2: performance for models using split phrases and sister-head dependencies

# Ergebnisse der Methode1

"Split PP" in training and testing tests

→Steigerung in LR(6-7%) und LP(~8%)

 By "Collapsed PP": eine leichte Senkung in LR and LP im Vergleich zu unmodifiziertem Collins' Modell

# Ergebnisse der Methode2

- Für TnT tags:
- →NPs fällt zu 67.84%LR und 68.45 LP
- unmodifiziertes Collins Modell besser
- →PPs erhöht um 70.27LR und 68.45LP
- Beste Performanz bei Sister-Head Abhängigkeit für alle Kategorien(71.32LR und 70.93%)
- Für perfect tags: Verbesserung von allen Kategorien mit Sister-Head-Abhägigkeiten

# Ergebnisse der Methode3

	TnT ta	agging ALP	Perfect tagging			
	ΔLR	$\Delta$ LP	$\Delta LR$	ΔLP		
PP	-3.45	-1.60	-4.21	-3.35		
S	-1.28	0.11	-2.23	-1.22		
Coord	-1.87	-0.39	-1.54	-0.80		
VP	-0.72	0.18	-0.58	-0.30		
AP	-0.57	0.10	0.08	-0.07		
AVP	-0.32	0.44	0.10	0.11		
NP	0.06	0.78	-0.15	0.02		

Table 6: Change in performance when reverting to head-head statistics for individual categories

- Bei der Umwandlung von PP to Head-Head Dependencing
- →Die größte Senkung ungefähr um 4%
- -Für S und coordinierende Kategorie
- →Eine Senkung um 1%
- Fast kein Effekt bei AP,AVP,und NP für perfekt Tagging

## Discussion

- Problem der Flachheit gelöst?
  - → Keine Verbesserung mit PPs Splitting auf "collapsed categories".
- Aber! Grund zur Besten Performanz bei Sister-Head-Abhängigkeiten Flachheit von Negra im Gegensatz zur Fehleranalyse
- Sister-Head Modell übertrifft sowohl original Collins Modell als auch unlexikalisierte Baseline mit 1%LP und 4%LR

# Vergleich mit früheren Arbeiten

- Keiner probabilistischer Treebanktrainierter parser für Deutsch
- Head-lexikalisiertes Modell von Carroll und Rooth angewandt für Deutsch von Beil et al.(1999,2002)
  - -Handannotierte Grammatik
  - -Unannotierter Corpus
  - -Nur auf Nebensätze und relative Sätze

Parsen von anderen Sprachen:

→ Arbeit von Collins et al.(1999) und Chiang und Bikel(2000) für Tcheschen und Chinesisch