

# Wiederholungsaufgaben

1	Syntaktische Ambiguität	2
2	Syntaktische Kategorien	4
3	Konstituentenstruktur	5
4	Parsingalgorithmen	6
5	Dependenzstruktur	8
6	Dependency Parsing	10
7	Komplexe Sätze	11
8	Grammatische Merkmale, Unifikation und Typhierarchie	14
9	Unifikationsgrammatiken	18
10	Statistisches Parsing	20
11	Datengestützte Syntaxanalyse	21
12	Partielles Parsing	23

# 1 Syntaktische Ambiguität

Arten syntaktischer Ambiguität .....

- (a)
- Welche Art struktureller Ambiguität liegt in folgendem Satz vor?
  - Erläutern Sie und geben Sie die Dependenz-Regeln an, die die Ambiguität verursachen.

(1) *The horse raced past the barn fell.*

- (b)
- Welche Art struktureller Ambiguität liegt in folgendem Satz vor?
  - Erläutern Sie und geben Sie die CFG-Regeln an, die die Ambiguität verursachen (Label: NP, PP, VP, Det, N, V).

(2) *I shot an elephant in my pajamas.*

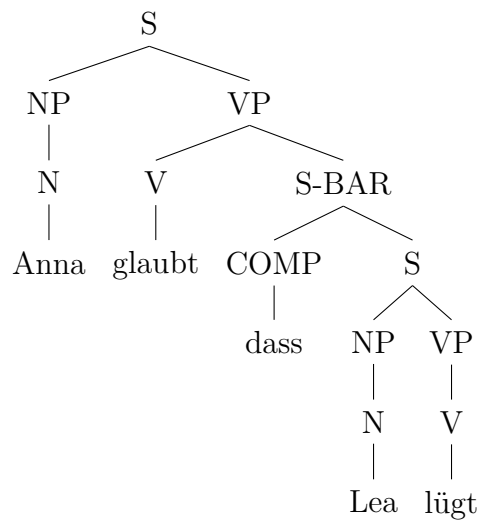
- (c)
- Welche Art syntaktischer Ambiguität liegt bei folgender NP vor?
  - Geben Sie für beide Analysen den Klammerausdruck an.

(3) *große Hunde und Katzen*

**Rekursionstypen** .....

- (a) Welche Art der Rekursion ermöglicht die verbale PP-Attachment-Regel  
 VP  $\rightarrow$  VP PP?

- (b) Erklären Sie anhand der in folgender Ableitung verwendeten Regeln, wie die rekursive Einbettung von Satzkomplementen zustande kommt?



- (c) • Welcher Art ist die Rekursion in (b)?  
 • Liegt hier eine *center-embedding*-Konstruktion vor?

## 2 Syntaktische Kategorien

### Konstituententests .....

- Identifizieren Sie in folgendem Satz in fünf Schritten alle Konstituenten (oberhalb der Wortebene).
- Pro Schritt darf ein Test angewandt werden, pro Schritt dürfen mehrere Konstituenten gleichzeitig identifiziert werden.
- Geben Sie zu jedem Schritt den verwendeten Test an.

(4) *Das Auto hält an der rot leuchtenden Ampel.*

Schritt 1:

Schritt 2:

Schritt 3:

Schritt 4:

Schritt 5:

- Geben Sie zum Schluss das daraus folgende Kastendiagramm an; verändern Sie den Satz hierbei nicht!

Kastendiagramm:

	Das		Auto		hält		an		der		roten		Ampel		
	Das		Auto		hält		an		der		rot		leuchtenden	Ampel	
	Das		Auto		hält		an		der		rot		leuchtenden	Ampel	
	Das		Auto		hält		an		der		rot		leuchtenden	Ampel	
	Das		Auto		hält		an		der		rot		leuchtenden	Ampel	
	Das		Auto		hält		an		der		rot		leuchtenden	Ampel	

### 3 Konstituentenstruktur

#### Kontextfreie Grammatiken und Syntaxbäume .....

- Erstellen sie zu dem folgenden Satz eine kontextfreie Grammatik und den dazugehörigen Syntaxbaum.
- Verwenden Sie nur folgende syntaktische Kategorien: S, NP, VP, N, DET, ADJ, V

(5) *Der Hund jagt den schreienden Briefträger.*

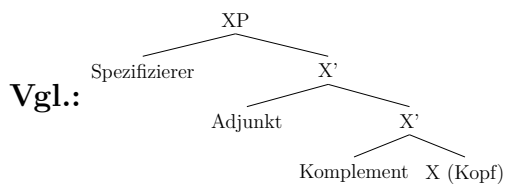
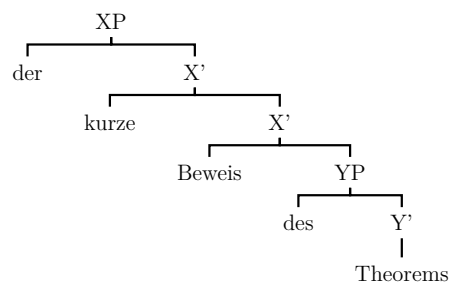
#### Erweiterung von kontextfreien Grammatiken .....

- Die Grammatik aus der vorherigen Aufgabe soll so erweitert werden, dass rekursiv PP-Adjunkte an NPs auftreten können (der Art *Franz aus München in Bayern ...*).
- Ergänzen Sie dazu im Folgenden die fehlenden rechten Seiten (PP-Regel ist gegeben; es dürfen keine neuen Kategorien eingeführt werden).

- |   |           |
|---|-----------|
| 1 | PP → P NP |
| 2 | NP →      |
| 3 | NOM →     |
| 4 | NOM →     |

#### X-Bar-Konstituentenstruktur .....

Was sind in folgendem Syntaxbaum im X-Bar-Strukturschema **Spezifizierer**, **Komplemente**, **Adjunkte** und **Köpfe**?



## 4 Parsingalgorithmen

### Shift-Reduce-Parser .....

Gegeben sei folgende Grammatik:

- |   |  |       |   |         |    |
|---|--|-------|---|---------|----|
| 1 |  | S     | → | NP      | VP |
| 2 |  | NP    | → | PROPN   |    |
| 3 |  | NP    | → | DET     | N  |
| 4 |  | VP    | → | V       | NP |
| 5 |  | VP    | → | V       |    |
| 6 |  | DET   | → | "der"   |    |
| 7 |  | N     | → | "Hund"  |    |
| 8 |  | PROPN | → | "Max"   |    |
| 9 |  | V     | → | "kennt" |    |

Mit dieser soll ein Shift-Reduce-Parser den Satz *der Hund kennt Max* analysieren. Während des Analysevorgangs wurden bereits die ersten drei Wörter eingelesen; dies ergab den folgenden Stack-Zustand:

V
NP

Der Stack-Zustand ist hier wie in der Vorlesung graphisch so dargestellt, dass neue Elemente immer oben auf den Stack kommen.

(a) Welche Parsing-Strategie verfolgt dieser Parser?

(b) Welche Operation wird als nächstes ausgeführt?

(c) Wie sieht der Stack nach Ausführen der Operation aus?

### Recursive-Descent-Parser .....

Gegeben sei folgende Grammatik: wie oben

Mit dieser soll ein Recursive-Descent-Parser den Satz *der Hund kennt Max* analysieren.

(a) Welche der beiden grundlegenden Parsing-Strategien verfolgt dieser Parser?

- (b) Welche zwei Operationen stehen diesem Parser zur Verfügung?

- (c) Nennen Sie eine Folge von Regelanwendungen, die zu Backtracking führt.

**Earley-Parser** .....

Gegeben sei folgende Grammatik: wie oben.

Mit dieser soll ein Earley-Parser den Satz *der Hund kennt Max* analysieren. Während des Analysevorgangs wird folgende Zustandsmenge  $Q_1$  erzeugt:

(DET  $\rightarrow$  der  $\cdot$ , 0) (1)

(NP  $\rightarrow$  DET  $\cdot$  N, 0) (2)

(N  $\rightarrow \cdot$  Hund, 1) (3)

- (a) Welche Operation wird als nächstes ausgeführt?

- (b) Welcher Zustand wird dabei hinzugefügt?

- (c) Welcher Zustandsmenge wird der Zustand hinzugefügt?

## 5    Abhängigkeitsstruktur

### Tests zur Komplement/Adjunkt-Unterscheidung .....

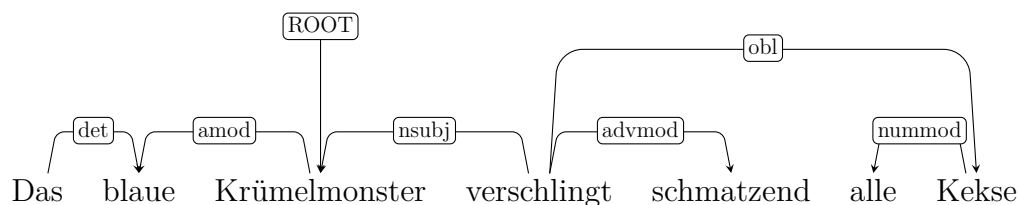
- (a) Nennen Sie die drei Tests zur Komplement/Adjunkt-Unterscheidung. Was wird geprüft? Wann handelt es sich jeweils um ein Adjunkt?

- (b) Bestimmen Sie für zwei verschiedene Konstituenten des folgenden Satzes, ob es sich jeweils um ein Komplement oder ein Adjunkt handelt. Verwenden Sie hierzu zwei der eben genannten Tests (für jede Konstituente einen).

(6)    *Das blaue Krümelmonster verschlingt schmatzend alle Kekse im Park.*

### Abhängigkeitsbäume .....

- Folgender Abhängigkeitsbaum enthält drei Fehler; diese können sich sowohl auf die Relationen als auch die Relationslabel beziehen; der Baum soll die UD-Konventionen erfüllen.
- Finden Sie die drei Fehler; markieren Sie dazu die falsche Relation bzw. schreiben Sie das korrekte Label über das falsche.
- (siehe letzte Seite für eine Übersicht der UD-Dependency-Labels)





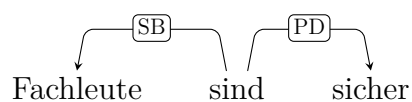
**Komplement vs. Modifikator** .....

- Bestimmen sie für den folgenden Satz die Abhängigkeiten. Markieren Sie, ob es sich bei der jeweiligen Abhängigkeit um ein Komplement oder einen Modifikator handelt.

(7) *Das Auto hält an der roten Ampel*

**UD-Schema vs. TIGER-Schema** .....

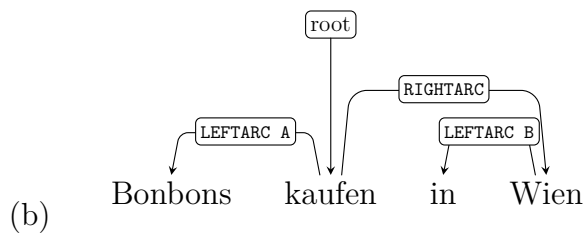
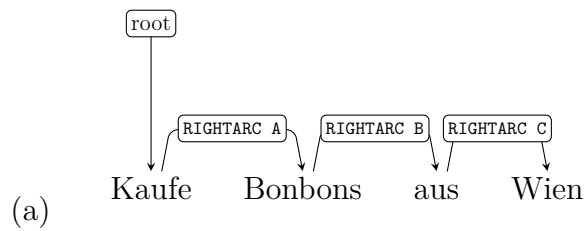
- Wandeln Sie folgenden Dependenzbaum im TIGER-Dependenz-Schema um in das *Universal-Dependencies-Schema* (*primacy of content word*-Maxime).
- (siehe letzte Seite für eine Übersicht der UD-Dependency-Labels)



## 6 Dependency Parsing

### Übergangsbasierter Shift-Reduce-Dependency-Parser .....

- In welcher Reihenfolge werden im Folgenden jeweils die angegebenen REDUCE-Übergänge durchgeführt?
- Begründen Sie.
- Wie unterscheiden sich die beiden Syntaxbäume?



## 7 Komplexe Sätze

### Konstituentenstruktur komplexer Sätze .....

- Geben Sie zu folgenden Regeln bzw. Regelgruppen **für komplexe Sätze des Englischen** einer Zeile jeweils deren Funktion an - welcher Typ komplexer Sätze wird jeweils erzeugt?
- Geben Sie zusätzlich auch die entsprechenden UD-Kantenlabel an.

S-BAR  $\rightarrow$  COMP S : \_\_\_\_\_

NP  $\rightarrow$  NOM,    NOM  $\rightarrow$  NOM S-BAR : \_\_\_\_\_

VP  $\rightarrow$  V S-BAR : \_\_\_\_\_

S  $\rightarrow$  S-BAR VP : \_\_\_\_\_

S  $\rightarrow$  S-BAR NP VP : \_\_\_\_\_

- Erstellen Sie die **Phrasenstrukturbäume** zu den folgenden beiden Sätzen; verwenden Sie dazu die dem eingebetteten Satztyp entsprechenden Regeln von oben, sowie zusätzlich folgende Regeln:

S  $\rightarrow$  NP VP,    NOM  $\rightarrow$  N,    VP  $\rightarrow$  V

(8)    *Max (N) sieht (V) dass (COMP) Franz (N) geht (V)*

(9)    *Max (N), den (COMP) Franz (N) biss (V), schrie (V)*

- Verwenden Sie die in Klammern angegebenen lexikalischen Kategorien!

**Dependenzstruktur komplexer Sätze 1** .....

- Erstellen Sie zu den Sätzen (8) und (9) auch die entsprechenden Dependenzbäume (Kopf von S = VP-Kopf; Kopf von S-BAR = Kopf von S).
- Verwenden Sie die im Anhang auf der letzten Seite angegebenen UD-Dependency-Label.

**Dependenzstruktur komplexer Sätze 2** .....

- Bestimmen Sie alle Dependenzrelationen im folgenden Satz.
- Bestimmen Sie außerdem, um welche grammatische Relation es sich jeweils handelt.

(10)    *Wenn das Auto an der roten Ampel hält, wird die Fußgängerampel bald grün.*

**Infinite Konstruktionen: Subjekt- vs. Objektkontrolle .....**

Liegt in folgendem Satz Subjekt- oder Objektkontrolle vor? Begründen Sie.

(11) *Sie überredeten ihn, noch zu bleiben*

## 8 Grammatische Merkmale, Unifikation und Typ-hierarchie

### Kodierungstypen syntaktischer Funktionen .....

Wie werden in den unteren transitiven Satzpaaren einer *Kunstsprache mit englischem Lexikon* die grammatischen Relationen **Subjekt** bzw. **Objekt** jeweils kodiert?

- Geben Sie jeweils eine der folgenden drei Kodierungsarten an:
  - **Kasusmorphologie** (*morphologisch* = substantielle Kodierung am Dependenten)
  - **Verbale Kongruenz** (*morphologisch* = substantielle Kodierung am Verb)
  - **Wortstellung** (strukturelle Kodierung)

(12) bird cat eat  
'Die Katze frisst den Vogel.'

(13) cat bird eat  
'Der Vogel frisst die Katze.'

(14) bird-fe cat eat  
'Die Katze frisst den Vogel.'

(15) bird cat-fe eat  
'Der Vogel frisst die Katze.'

(16) bird-i cat-o eat-o  
'Die Katze frisst den Vogel.'

(17) bird-i cat-o eat-i  
'Der Vogel frisst die Katze.'

**Kasusreaktion, Agreement und Subkategorisierung** .....

Welche der folgenden morphosyntaktischen Beschränkungen ist in den unteren Sätzen jeweils verletzt?

- **Kasusreaktion**
- **Nominale Kongruenz**
- **Verbale Kongruenz** (*Agreement*)
- **Subkategorisierung** (Anzahl und Art der verbalen Argumente)

(18) *Der Auto fährt schnell.*

(19) *Den Auto fährt schnell.*

(20) *Die Autos fährt schnell.*

(21) *Das Auto fährt, dass es ankommt.*

**Wortstellung: Stellungsfeldermodell** .....

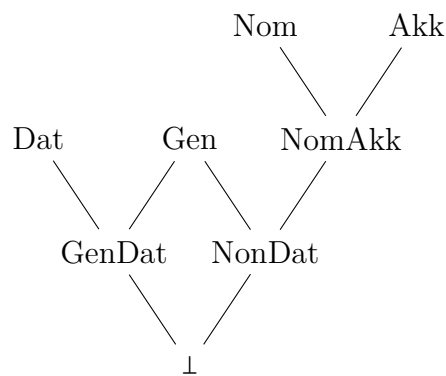
(a) Identifizieren Sie in den folgenden Sätzen durch Unterstreichen das Mittelfeld und, falls vorhanden, das Vorfeld:

(22) *Es hat Kuchen gegeben.*

(23) *Hat es Kuchen gegeben?*

(b) Um welches *Es* handelt es sich in (22)?

- **Topik-Es** = Vorfeld-Platzhalter
- **Subjekt-Expletiv**
- **pronominaler Ersatz** für NP

**Unifikationsparsing und getypte Merkmalstrukturen** .....

Gegeben sei folgende Typhierarchie:

**Typhierarchien** .....

Unifizieren Sie die folgenden Paare von Typen. Typen, die nicht unifizieren, markieren Sie als *undefiniert*.

- (a)  $\text{Nom} \sqcup \text{Akk} =$  \_\_\_\_\_
- (b)  $\text{GenDat} \sqcup \text{NonDat} =$  \_\_\_\_\_
- (c)  $\text{GenDat} \sqcup \text{Gen} =$  \_\_\_\_\_
- (d)  $\text{Nom} \sqcup \perp =$  \_\_\_\_\_
- (e)  $\text{Akk} \sqcup \text{NomAkk} =$  \_\_\_\_\_

**Subsumption** .....

Gegeben seien nun zusätzlich folgende Merkmalstrukturen mit  $\theta(\text{FS1}) = \theta(\text{FS2}) = \theta(\text{FS3}) = \perp$ .

$$\text{FS1} = \begin{bmatrix} \text{CAS} & \text{NomAkk} \\ \text{GEN} & \text{mask} \end{bmatrix} \quad \text{FS2} = \begin{bmatrix} \text{CAS} & \text{Nom} \\ \text{GEN} & \text{mask} \\ \text{PER} & 3 \end{bmatrix} \quad \text{FS3} = \begin{bmatrix} \text{CAS} & \text{Akk} \\ \text{PER} & 3 \end{bmatrix}$$

Entscheiden Sie jeweils mit *ja* oder *nein*:

- (a)  $\text{FS2} \sqsubseteq \text{FS3} ?$  \_\_\_\_\_
- (b)  $\text{FS1} \sqsubseteq \text{FS2} ?$  \_\_\_\_\_
- (c)  $\text{FS2} \sqsubseteq \text{FS1} ?$  \_\_\_\_\_
- (d)  $\text{FS3} \sqsubseteq \text{FS2} ?$  \_\_\_\_\_



(e)  $\text{FS3} \sqsubseteq \text{FS3}$  ? \_\_\_\_\_

**Unifikation** .....

Unifizieren FS2 und FS3? Falls ja, unifizieren Sie; falls nein, begründen Sie.

**Bedingungen** .....

Entscheiden Sie jeweils mit *ja* oder *nein*:

(a)  $\text{FS2} \models \text{CAS}:\perp$  ? \_\_\_\_\_

(b)  $\text{FS2} \models \text{GEN}:\text{neut}$  ? \_\_\_\_\_

(c)  $\text{FS3} \models \text{PER}:3$  ? \_\_\_\_\_

(d)  $\text{FS1} \models \text{NomAkk} \wedge \text{mask}$  ? \_\_\_\_\_

(e)  $\text{FS2} \models \text{CAS}:\text{NomAkk}$  ? \_\_\_\_\_

## 9 Unifikationsgrammatiken

### Modellierung von Subkategorisierung, Rektion und Agreement .....

- Betrachten Sie folgenden Ausschnitt aus einer merkmalsbasierten Grammatik für einen kleinen Ausschnitt des Englischen (s. NLTK/book\_grammars/feat0.fcfg).
- Beantworten Sie untenstehende Fragen und geben Sie jeweils die Zeilennummern an, auf die sich ihre Antwort bezieht.

```
1 | S -> NP [NUM=?n] VP [NUM=?n]
2 |
3 | NP [NUM=?n] -> N [NUM=?n]
4 | NP [NUM=?n] -> PropN [NUM=?n]
5 | NP [NUM=?n] -> Det [NUM=?n] N [NUM=?n]
6 | NP [NUM=p1] -> N [NUM=p1]
7 |
8 | VP [TENSE=?t, NUM=?n] -> IV [TENSE=?t, NUM=?n]
9 | VP [TENSE=?t, NUM=?n] -> TV [TENSE=?t, NUM=?n] NP
```

- (a) Wie wird hier Subkategorisierung modelliert?

- (b) Nennen Sie eine alternative Modellierung von Subkategorisierung.

- (c) Wie wird hier das Subjekt-Verb-Agreement modelliert?

- (d) Geben Sie das entsprechende Constraint für das Subjekt-Verb-Agreement mittels einer Gleichung an (z.B. als Pfadgleichung der Form `CAT1@Feat = CAT2@Feat`).

**Erweiterung von kontextfreien Grammatiken um Merkmale .....**

- Gegeben sei folgende Grammatik:

1	-> ADV VP
2	VP -> V NP PP
3	PP -> P NP
4	NP -> N
5	NP -> PROPN
6	ADV -> "gestern"
7	ADV -> "heute"
8	V -> "ging"
9	V -> "geht"
10	PROPN -> "Fritz"
11	N -> "Arbeit"
12	P -> "zur"

- Mit dieser können die folgenden Sätze hergeleitet werden:  
*gestern ging Fritz zur Arbeit*  
*heute ging Fritz zur Arbeit*  
*heute geht Fritz zur Arbeit*
- Passen sie die Grammatik so an, dass Imperfekt und Präsens als Merkmale in der Grammatik unterschieden werden können, so dass der folgende Satz nicht mehr erkannt wird:

(26) *\*gestern geht Fritz zur Arbeit*

1		-> ADV	VP		
2	VP	-> V		NP	PP
3	PP	-> P		NP	
4	NP	-> N			
5	NP	-> PROPN			
6	ADV	-> "gestern"			
7	ADV	-> "heute"			
8	V	-> "ging"			
9	V	-> "geht"			
10	PROPN	-> "Fritz"			
11	N	-> "Arbeit"			
12	P	-> "zur"			

## 10    Statistisches Parsing

### PCFG: Gewichte und Ableitungswahrscheinlichkeit .....

Betrachten Sie folgendes PCFG-Parsing (\*\* = unkenntlich gemacht):

```
1 | grammar = nltk.PCFG.fromstring("""
2 |     S      -> NP VP          [1.0]
3 |     VP     -> TV NP          [0.4]
4 |     VP     -> IV            [**]
5 |     VP     -> DatV NP NP     [0.3]
6 |     TV     -> 'saw'          [1.0]
7 |     IV     -> 'ate'          [1.0]
8 |     DatV   -> 'gave'         [1.0]
9 |     NP     -> 'telescopes'   [0.8]
10 |    NP     -> 'Jack'          [0.2]
11 |    """)
12 | viterbi_parser = nltk.ViterbiParser(grammar)
13 | for tree in viterbi_parser.parse(['Jack', 'saw', 'telescopes']):
14 |     print(tree)
15 | (S (NP Jack) (VP (TV saw) (NP telescopes))) (p=0.064)
```

- (a)    Geben Sie die Berechnung für die Ableitungswahrscheinlichkeit in Zeile 15 an?

- (b)    Welchen Wert muss das Gewicht für die Regel **VP** → **IV** haben?

## 11 Datengestützte Syntaxanalyse

### Datengestützte Methoden: Abschätzung Regelwahrscheinlichkeiten .....

- (a) Für ein großes Korpus sei lediglich *part of speech* (POS) annotiert; Syntaxbäume stehen nicht zur Verfügung.

Welche Arten von Regeln einer kontextfreien Grammatik kann man mit diesen Daten automatisch generieren?

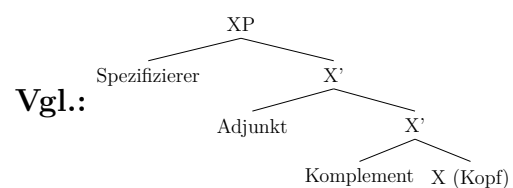
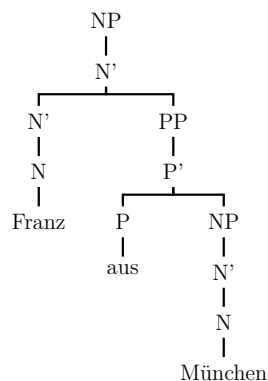
- (b) Folgende Häufigkeiten wurden aus einem Datensatz gezählt:

$count(VP \rightarrow V) = 200$ ,  $count(VP \rightarrow V NP) = 100$ ,  $count(VP \rightarrow \backslash *) = 300$ .

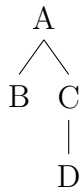
Berechnen Sie  $P(V NP | VP)$  mit der MLE-Methode.

### Methoden für lexikalisierte und *history-based* PCFGs .....

- (a) Führen Sie im linken Syntaxbaum eine Kopfannotation durch; Geben Sie anschließend die lexikalisierte Regel für den Wurzelknoten an. Orientieren Sie bei der Kopfannotation an der Strukturposition des Kopfes im X-Bar-Schema (vgl. rechter Syntaxbaum).



- (b) Führen Sie für die CFG-Regeln  $\mathbf{C} \rightarrow \mathbf{D}$  und  $\mathbf{A} \rightarrow \mathbf{B}$  in unterem Syntaxbaum *parent-annotation* durch.



--

## 12 Partielles Parsing

### Chunking .....

Markieren Sie alle Nominalphrasen (NPs), indem Sie den folgenden deutschen Satz vollständig nach dem IOB-Tagging-Schema annotieren; verwenden Sie nur folgende Label: B-NP, I-NP, O.

Token	Der	junge	Mann	gab	ihr	das	Buch	.
Tag								

### Kaskadierende Chunk-Parser .....

- (a) Mit welcher Methode kann z.B. folgende hierarchische Struktur einer Präpositionalphrase mit flachen Chunk-Parsern erzeugt werden:

[PP auf/P [NP dem/DET Baum/N ] ]

### Evaluationsmetriken .....

Berechnen Sie Accuracy, Precision und Recall für folgende korrekte Annotationen (**truth**) und folgende Hypothesen (**predict**). Geben Sie bitte jeweils Brüche an.

Sample	0	1	2	3	4	5	6	7	8	9
<b>truth</b>	PP	PP	PP	O	O	O	O	PP	O	PP
<b>predict</b>	PP	O	O	PP	O	PP	PP	O	PP	PP

Accuracy: \_\_\_\_\_

Precision (für die Klasse PP): \_\_\_\_\_

Recall (für die Klasse PP): \_\_\_\_\_

## Angabe: Hilfsmittel

### Universal Dependency Relations

	Nominals	Clauses	Modifier words	Function Words
Core arguments	<a href="#"><u>nsubj</u></a> <a href="#"><u>obj</u></a> <a href="#"><u>iobj</u></a>	<a href="#"><u>csubj</u></a> <a href="#"><u>ccomp</u></a> <a href="#"><u>xcomp</u></a>		
Non-core dependents	<a href="#"><u>obl</u></a> <a href="#"><u>vocative</u></a> <a href="#"><u>expl</u></a> <a href="#"><u>dislocated</u></a>	<a href="#"><u>advcl</u></a>	<a href="#"><u>advmod</u></a> * <a href="#"><u>discourse</u></a>	<a href="#"><u>aux</u></a> <a href="#"><u>cop</u></a> <a href="#"><u>mark</u></a>
Nominal dependents	<a href="#"><u>nmod</u></a> <a href="#"><u>appos</u></a> <a href="#"><u>nummod</u></a>	<a href="#"><u>acl</u></a>	<a href="#"><u>amod</u></a>	<a href="#"><u>det</u></a> <a href="#"><u>clf</u></a> <a href="#"><u>case</u></a>
Coordination	MWE	Loose	Special	Other
<a href="#"><u>conj</u></a> <a href="#"><u>cc</u></a>	<a href="#"><u>fixed</u></a> <a href="#"><u>flat</u></a> <a href="#"><u>compound</u></a>	<a href="#"><u>list</u></a> <a href="#"><u>parataxis</u></a>	<a href="#"><u>orphan</u></a> <a href="#"><u>goeswith</u></a> <a href="#"><u>reparandum</u></a>	<a href="#"><u>punct</u></a> <a href="#"><u>root</u></a> <a href="#"><u>dep</u></a>