

# Einführung in die Computerlinguistik

## Übungsaufgaben zur Abschlussklausur

Laura Kallmeyer

Sommersemester 2013, Heinrich-Heine-Universität Düsseldorf

Erlaubte Hilfsmittel in der Klausur: Eine Din-A4 Seite mit Notizen. Kein Taschenrechner.

**Aufgabe 1** Betrachten Sie die folgende CFG  $G : N = \{S, A, B, C\}$ ,  $T = \{a, b, c\}$ , Startsymbol  $S$ .

Produktionen:

$$\begin{array}{l} S \rightarrow AcB \quad S \rightarrow C \quad C \rightarrow cC \\ A \rightarrow aAb \quad A \rightarrow \varepsilon \quad B \rightarrow bB \quad B \rightarrow A \end{array}$$

1. Welche Symbole in  $G$  sind nutzlos? Geben Sie eine zu  $G$  äquivalente CFG  $G'$  ohne nutzlose Symbole an.
2. Berechnen Sie die Menge  $N_\varepsilon$  der Nichtterminalen, aus denen sich in  $G'$   $\varepsilon$  ableiten lässt.
3. Geben Sie eine zu  $G'$  äquivalente  $\varepsilon$ -freie CFG  $G''$  an.

Lösung:

1. Nutzlos sind  $c$  und  $C$ .

$G'$  enthält Produktionen

$$S \rightarrow AcB \quad A \rightarrow aAb \quad A \rightarrow \varepsilon \quad B \rightarrow bB \quad B \rightarrow A$$

2.  $N_\varepsilon = \{A, B\}$ .

3.  $G''$  enthält Produktionen

$$\begin{array}{l} S \rightarrow AcB \quad S \rightarrow cB \quad S \rightarrow Ac \quad S \rightarrow c \\ A \rightarrow aAb \quad A \rightarrow ab \quad B \rightarrow bB \quad B \rightarrow b \quad B \rightarrow A \end{array}$$

**Aufgabe 2** Betrachten Sie die folgende CFG  $G : N = \{S, A, B, C\}$ ,  $T = \{a, b\}$ , Startsymbol  $S$ .

Produktionen:

$$\begin{array}{l} S \rightarrow AB \quad S \rightarrow aCb \quad A \rightarrow aa \\ C \rightarrow aCb \quad C \rightarrow ab \quad B \rightarrow bb \end{array}$$

Transformieren Sie die Grammatik in eine äquivalente Grammatik in Chomsky Normal Form.

Lösung:

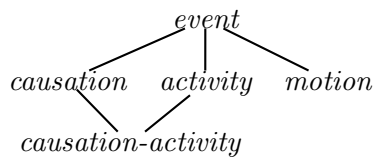
$$\begin{array}{l} S \rightarrow AB \quad S \rightarrow T_a X \quad A \rightarrow T_a T_a \\ C \rightarrow T_a X \quad C \rightarrow T_a T_b \quad B \rightarrow T_b T_b \\ T_a \rightarrow a \quad T_b \rightarrow b \quad X \rightarrow CT_b \end{array}$$

**Aufgabe 3** Betrachten Sie die folgenden Merkmalsstrukturen formuliert als Attribut-Wert Matrizen:

$$S_1 = \left[ \begin{array}{l} \text{activity} \\ \text{ACTOR} \left[ \begin{array}{l} \text{person} \\ \text{NAME Mary} \end{array} \right] \\ \text{THEME} \left[ \begin{array}{l} \text{object} \\ \text{TYPE ball} \\ \text{SHAPE round} \end{array} \right] \end{array} \right] \quad S_2 = \left[ \begin{array}{l} \text{causation} \\ \text{CAUSE} \left[ \begin{array}{l} \text{activity} \\ \text{ACTOR} \left[ \begin{array}{l} \text{person} \\ \text{NAME Paul} \end{array} \right] \\ \text{THEME} \boxed{2} \end{array} \right] \\ \text{EFFECT} \left[ \begin{array}{l} \text{event} \\ \text{THEME} \boxed{2} \left[ \begin{array}{l} \text{object} \\ \text{TYPE ball} \end{array} \right] \end{array} \right] \end{array} \right]$$

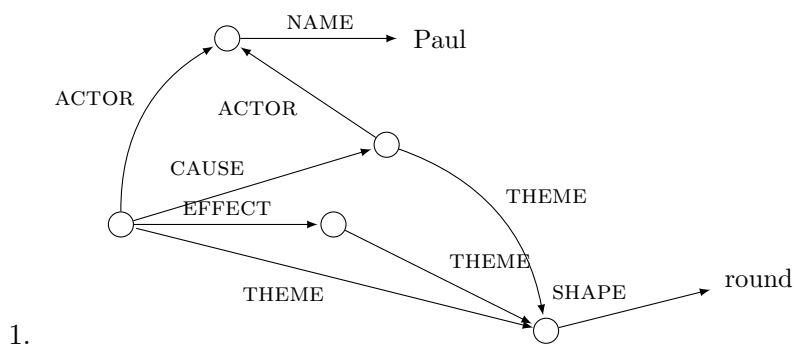
$$S_3 = \left[ \begin{array}{l} \text{causation-activity} \\ \text{ACTOR} \boxed{1} \\ \text{THEME} \boxed{2} \left[ \begin{array}{l} \text{object} \\ \text{SHAPE round} \end{array} \right] \\ \text{CAUSE} \left[ \begin{array}{l} \text{activity} \\ \text{ACTOR} \boxed{1} \left[ \begin{array}{l} \text{person} \\ \text{NAME Paul} \end{array} \right] \\ \text{THEME} \boxed{2} \end{array} \right] \\ \text{EFFECT} \left[ \begin{array}{l} \text{motion} \\ \text{THEME} \boxed{2} \end{array} \right] \end{array} \right]$$

Typenhierarchie:



1. Geben Sie für  $S_3$  den entsprechenden Graphen an.
2. Berechnen Sie  $S_1 \sqcup S_2$ ,  $S_1 \sqcup S_3$  und  $S_2 \sqcup S_3$ . In den Fällen, in denen keine Unifikation möglich ist (Ergebnis  $\perp$ ) begründen Sie, warum die Unifikation scheitert.  
Tipp: Achten Sie auch darauf, dass das Unifikationsergebnis den richtigen Typ hat.

Lösung:



$$2. S_1 \sqcup S_2 = \left[ \begin{array}{l} \textit{causation-activity} \\ \text{ACTOR} \left[ \begin{array}{l} \textit{person} \\ \text{NAME Mary} \end{array} \right] \\ \text{THEME} \left[ \begin{array}{l} \textit{object} \\ \text{TYPE ball} \\ \text{SHAPE round} \end{array} \right] \\ \text{CAUSE} \left[ \begin{array}{l} \textit{activity} \\ \text{ACTOR} \left[ \begin{array}{l} \textit{person} \\ \text{NAME Paul} \end{array} \right] \\ \text{THEME} \boxed{2} \end{array} \right] \\ \text{EFFECT} \left[ \begin{array}{l} \textit{event} \\ \text{THEME} \boxed{2} \left[ \begin{array}{l} \textit{object} \\ \text{TYPE ball} \end{array} \right] \end{array} \right] \end{array} \right]$$

$S_1 \sqcup S_3 = \perp$ , da sich unterschiedliche Werte für den Pfad ACTOR NAME ergeben und (aufgrund der Identität) auch für den Pfad CAUSE ACTOR NAME.

$$S_2 \sqcup S_3 = \left[ \begin{array}{l} \textit{causation-activity} \\ \text{ACTOR} \boxed{1} \\ \text{THEME} \boxed{2} \left[ \begin{array}{l} \textit{object} \\ \text{TYPE ball} \\ \text{SHAPE round} \end{array} \right] \\ \text{CAUSE} \left[ \begin{array}{l} \textit{activity} \\ \text{ACTOR} \boxed{1} \left[ \begin{array}{l} \textit{person} \\ \text{NAME Paul} \end{array} \right] \\ \text{THEME} \boxed{2} \end{array} \right] \\ \text{EFFECT} \left[ \begin{array}{l} \textit{motion} \\ \text{THEME} \boxed{2} \end{array} \right] \end{array} \right]$$

**Aufgabe 4** Betrachten Sie die folgende CFG  $G : N = \{S, A, B, C\}$ ,  $T = \{a, b\}$ , Startsymbol  $S$ .

Produktionen:  $S \rightarrow AB \quad S \rightarrow C \quad A \rightarrow aa$   
 $C \rightarrow aCb \quad C \rightarrow ab \quad B \rightarrow bb$

1. Geben Sie die Trace an, die bei einem Top-Down Parsing von  $w = aabb$  entsteht. D.h., alle Paare aus Stack und verbleibender Eingabe (ohne Analysestack), die sich ergeben, jeweils mit einer Angabe darüber, durch welche Operation ein Paar aus welchem anderen Paar entsteht.
2. Woran erkennt man, dass das Wort zu der von  $G$  generierten Sprache gehört?

Lösung:

	Resteingabe	Stack	
1.	aabb	S	
2.	aabb	AB	predict(1)
3.	aabb	C	predict(1)
4.	aabb	aaB	predict(2)
5.	aabb	aCb	predict(3)
6.	aabb	ab	predict(3)
7.	abb	aB	scan(4)
8.	abb	Cb	scan(5)
1. 9.	abb	b	scan(6)
10.	abb	aCbb	predict(8)
11.	abb	abb	predict(8)
12.	bb	B	scan(7)
13.	bb	Cbb	scan(10)
14.	bb	bb	scan(11) und predict(12)
15.	bb	aCbbb	predict(13)
16.	bb	abbb	predict(13)
17.	b	b	scan(14)
18.	$\epsilon$	$\epsilon$	scan(17)

2. Dass  $aabb$  zu  $L(G)$  gehört, erkennt man daran, dass der Parser mit leerem Stack und leerer Resteingabe endet.

**Aufgabe 5** Betrachten Sie die folgende PCFG:  $N = \{S, A, B, C, D, T_a, T_b, T_c\}$ ,  $T = \{a, b, c, d\}$ , Startsymbol  $S$ .

Produktionen mit Wahrscheinlichkeiten:

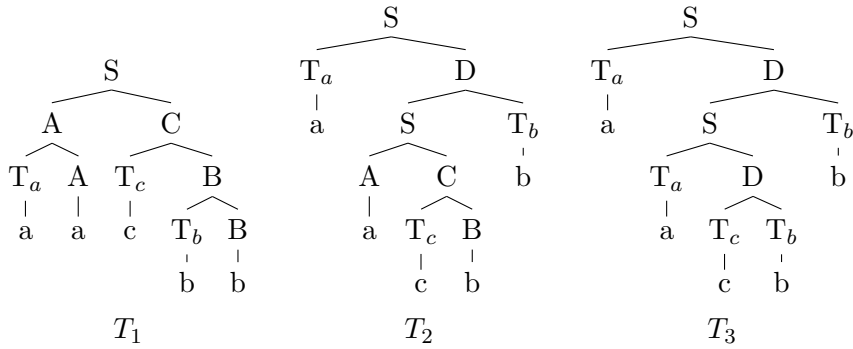
$$\begin{array}{lll}
3 \cdot 10^{-1}: S \rightarrow AC & 1: C \rightarrow T_c B & \\
1 \cdot 10^{-1}: S \rightarrow T_a D & 1 \cdot 10^{-1}: D \rightarrow S T_b & 2 \cdot 10^{-1}: D \rightarrow T_c T_b \\
2 \cdot 10^{-1}: A \rightarrow T_a A & 1 \cdot 10^{-1}: A \rightarrow a & \\
2 \cdot 10^{-1}: B \rightarrow T_b B & 1 \cdot 10^{-1}: B \rightarrow b & \\
1: T_a \rightarrow a & 1: T_b \rightarrow b & 1: T_c \rightarrow c \\
6 \cdot 10^{-1}: S \rightarrow T_c T_c & 7 \cdot 10^{-1}: D \rightarrow T_c T_c & \\
7 \cdot 10^{-1}: A \rightarrow d & 7 \cdot 10^{-1}: B \rightarrow d & 
\end{array}$$

Betrachten Sie die Eingabe  $w = aacbb$ .

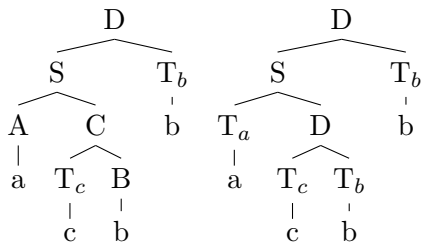
- $aacbb$  hat drei Lesarten. Geben Sie die entsprechenden Parsbäume an.  
Tipp: Die Produktionen in den letzten zwei Zeilen werden hierbei nicht benötigt.
- Wie hoch ist die Insidewahrscheinlichkeit von  $D$  mit Positionen 2, 5, bezogen auf diese Eingabe?
- Wie hoch ist die Outsidewahrscheinlichkeit von  $S$  mit Positionen 2, 4, bezogen auf diese Eingabe?
- Welche Chart ergibt sich bei einem probabilistischen CYK Chart Parsing der Eingabe  $w = acb$  mit dieser Grammatik? Verwenden Sie den Algorithmus, der auch Produktionen und Zwischenlängen festhält.

Lösung:

- drei Lesarten:



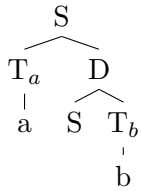
2. Es gibt zwei mögliche Ableitungsbäume mit Wurzel  $D$  und Kette  $acbb$ :



Also ergibt sich

$$\beta_D(2, 5) = 1 \cdot 3 \cdot 1 \cdot 1 \cdot 10^{-4} + 1 \cdot 1 \cdot 2 \cdot 10^{-3} = 23 \cdot 10^{-4}$$

3. Es gibt (bis auf die Reihenfolge der Ableitungsschritte) nur eine Möglichkeit, aus  $S$  die Kette  $aSb$  herzuleiten:



Daher ergibt sich

$$\alpha_S(2, 4) = 1 \cdot 10^{-2}$$

4. Chart:

$l$	$l$		
3	.02 : $S \rightarrow T_a D, 1$		
2		.2 : $D \rightarrow T_c T_b, 1$ .1 : $C \rightarrow T_c B, 1$	
1	1 : $T_a \rightarrow a, -$ .1 : $A \rightarrow a, -$	1 : $T_c \rightarrow c, -$	1 : $T_b \rightarrow b, -$ .1 : $B \rightarrow b, -$
1	a	2	3 c b i