

# Einführung in die Computerlinguistik Abschlussklausur

Laura Kallmeyer

Sommersemester 2015, Heinrich-Heine-Universität Düsseldorf

15.07.2015

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

**Aufgabe 1 (8 Pkte)** Betrachten Sie folgende CFG:

$$G = \langle \{S, A, B\}, \{a, b\}, \{S \rightarrow aB \mid bA, A \rightarrow a \mid aS \mid bAA, B \rightarrow b \mid bS \mid aBB\}, S \rangle$$

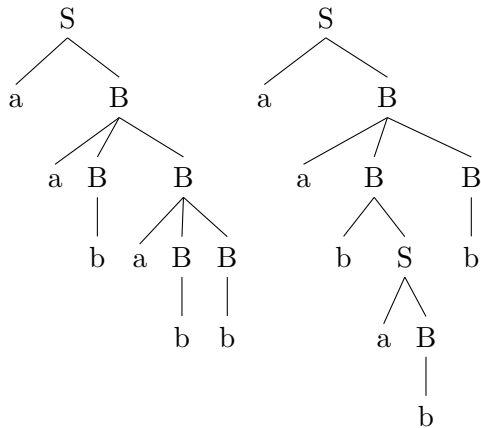
1. Ist diese Grammatik in Greibach Normalform? Begründen Sie Ihre Antwort.
2. Geben Sie eine Links- und eine Rechtsableitung für  $w = aabb$  an. (Es gibt jeweils nur eine.)
3. Geben Sie zwei mögliche Parsbäume für  $w = aababb$  an.
4. Welche Sprache wird von  $G$  erzeugt?

Lösung:

1. Ja, da die rechten Seiten aller Produktionen mit einem Terminalen beginnen, dem eine (eventuell leere) Sequenz von Nichtterminalen folgt. 1 Pkt

2. Linksableitung:  $S \Rightarrow aB \Rightarrow aaBB \Rightarrow aabB \Rightarrow aabb$  2 Pkte  
 Rechtsableitung:  $S \Rightarrow aB \Rightarrow aaBB \Rightarrow aaBb \Rightarrow aabb$

3. 4 Pkte



4.  $\{w \in \{a, b\}^+ \mid |w|_a = |w|_b\}$  1 Pkt

**Aufgabe 2 (9 Pkte)**

1. Geben Sie CFGs für folgende Sprachen an:
  - (a)  $\{a^n b^{2n} \mid n > 0\}$
  - (b)  $\{(ad)^n c^m \mid n \geq m \geq 0\}$
2. Welche Sprachen werden von folgenden CFGs erzeugt:
  - (a)  $G = \langle \{S, T\}, \{a, b\}, \{S \rightarrow aS \mid T, T \rightarrow aTb \mid \varepsilon\}, S \rangle$

$$(b) G = \langle \{S\}, \{a, b, c\}, \{S \rightarrow aSb \mid aSc \mid ab \mid ac\}, S \rangle$$

Lösung:

1. (a)  $S \rightarrow aSbb \mid abb$  (b)  $S \rightarrow adSc \mid adS \mid \varepsilon$  2+3 Pkte
2. (a)  $\{a^n a^m b^m \mid n, m \geq 0\}$  (b)  $\{a^n w \mid w \in \{b, c\}^*, |w| = n, n \geq 1\}$  2+2 Pkte

**Aufgabe 3 (4 Pkte)** Transformieren Sie die Grammatik aus Aufgabe 1 in eine äquivalente CFG in Chomsky Normalform.

Lösung:

1. Einführen von Präterminalen:  
 $\{S \rightarrow C_a B \mid C_b A, A \rightarrow a \mid C_a S \mid C_b A A, B \rightarrow b \mid C_b S \mid C_a B B, C_a \rightarrow a, C_b \rightarrow b\}$  2 Pkte
2. Binarisierung:  
 $\{S \rightarrow C_a B \mid C_b A, A \rightarrow a \mid C_a S \mid C_b X, X \rightarrow A A, B \rightarrow b \mid C_b S \mid C_a Y, Y \rightarrow B B, C_a \rightarrow a, C_b \rightarrow b\}$  2 Pkte

**Aufgabe 4 (4 Pkte)** Wir wissen schon, dass  $L_1 = \{a^n b^n c^n \mid n \geq 1\}$  keine CFL ist.

Zeigen Sie, dass dann auch  $L_2 = \{a^n b^n c^n d^n \mid n \geq 1\}$  keine CFL ist.

Lösung:

Wir nehmen an,  $L_2$  ist eine CFL. Dann muss auch das Bild von  $L_2$  unter dem Homomorphismus  $f$  mit  $f(a) = a, f(b) = b, f(c) = c, f(d) = \varepsilon$  eine CFL sein. Dieses Bild ist aber  $L_1$ , wovon wir wissen, dass es keine CFL ist.

Folglich ist unsere Annahme falsch und  $L_2$  ist keine CFL. 4 Pkte

**Aufgabe 5 (7 Pkte)**

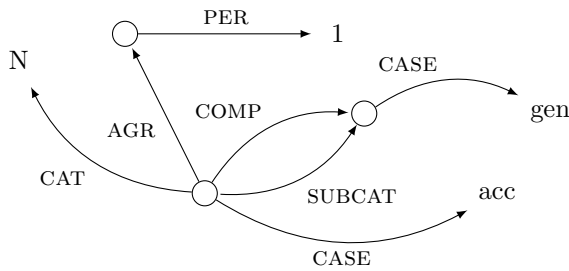
Betrachten Sie die folgenden ungetypten Merkmalsstrukturen, formuliert als Attribut-Wert Matrizen:

$$S_1 = \begin{bmatrix} \text{CAT} & \text{N} \\ \text{CASE} & \text{acc} \\ \text{SUBCAT} & \boxed{1} \\ \text{AGR} & [\text{PER } 1] \\ \text{COMP} & \boxed{1}[\text{CASE gen}] \end{bmatrix} \quad S_2 = \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CAT} & \text{V} \\ \text{COMP} & [\text{CASE dat}] \end{bmatrix} \\ \text{AGR} & [\text{NUM sg}] \\ \text{COMP} & [\text{CAT N}] \end{bmatrix}$$

1. Geben Sie für  $S_1$  den entsprechenden Graphen an.
2. Berechnen Sie  $S_1 \sqcup S_2$ . Sollte keine Unifikation möglich sein (Ergebnis  $\perp$ ) begründen Sie, warum die Unifikation scheitert.

Lösung:

1. Graph:



3 Pkte

$$2. S_1 \sqcup S_2 = \left[ \begin{array}{l} \text{CAT} \quad \text{N} \\ \text{HEAD} \quad \left[ \begin{array}{l} \text{CAT} \quad \text{V} \\ \text{COMP} \quad \left[ \text{CASE} \quad \text{dat} \right] \end{array} \right] \\ \text{CASE} \quad \text{acc} \\ \text{SUBCAT} \quad \boxed{1} \\ \text{AGR} \quad \left[ \begin{array}{l} \text{NUM} \quad \text{sg} \\ \text{PER} \quad 1 \end{array} \right] \\ \text{COMP} \quad \boxed{1} \left[ \begin{array}{l} \text{CAT} \quad \text{N} \\ \text{CASE} \quad \text{gen} \end{array} \right] \end{array} \right]$$

4 Pkte

**Aufgabe 6 (6 Pkte)** Betrachten Sie die CFG  $G$  mit  $N = \{S, A\}$ ,  $T = \{a, b\}$ , Startsymbol  $S$  und Produktionen

$$S \rightarrow aAcb \quad A \rightarrow aS \quad A \rightarrow a$$

1. Ergänzen Sie den Anfang der Trace, die bei einem Shift-Reduce Parsing von  $w = aacbcb$  entsteht, indem Sie alle Paare aus Stack und verbleibender Eingabe hinzufügen, die sich noch ergeben, wenn sämtliche Shift und Reduce Schritte ausprobiert werden (auch nach einem erfolgreichen Parse wird noch weitergemacht, bis nichts mehr geht). Geben Sie jeweils an, durch welche Operation und, falls es ein reduce Schritt war, mit welcher Produktion ein Paar aus welchem anderen Paar entsteht.

	Stack	Resteingabe	Operation
1.	$\varepsilon$	$aacb$	
2.	$a$	$acb$	$shift(1)$
3.	$A$	$acb$	$reduce(2) \quad A \rightarrow a$
4.	$aa$	$cb$	$shift(2)$
5.	$Aa$	$cb$	$shift(3)$

2. Ergänzen Sie jetzt den Anfang der Trace, die bei einem Top-Down Parsing von  $w = aacb$  entsteht, indem Sie alle Paare aus Stack und verbleibender Eingabe hinzufügen, die sich noch ergeben, wenn sämtliche predict und scan Schritte ausprobiert werden. Geben Sie auch hier jeweils an, durch welche Operation ein Paar aus welchem anderen Paar entsteht.

Wir nehmen an, dass ein Filter eingebaut ist, der verhindert, dass der Stack länger als die Resteingabe wird.

	Stack	Resteingabe	Operation
1.	$S$	$aacbcb$	
2.	$aAcb$	$aacbcb$	$predict(1)$
3.	$Acb$	$acbc$	$scan(2)$

Lösung:

1. Shift reduce trace:

	Stack	Resteingabe	Operation
1.	$\varepsilon$	aacb	
2.	a	acb	shift(1)
3.	A	acb	reduce(2) $A \rightarrow a$
4.	aa	cb	shift(2)
5.	Aa	cb	shift(3)
6.	aA	cb	reduce(4) $A \rightarrow a$
7.	AA	cb	reduce(5) $A \rightarrow a$
8.	aAc	b	shift(6)
9.	AAc	b	shift(7)
10.	aAcb	$\varepsilon$	shift(8)
11.	AAcb	$\varepsilon$	shift(9)
12.	S	$\varepsilon$	reduce(10) $S \rightarrow aAcb$

3 Pkte

2. Top down trace. Das Wort gehört nicht zur Sprache.

	Stack	Resteingabe	Operation
1.	S	aacbc	
2.	aAcb	aacbc	predict(1)
3.	Acb	acbc	scan(2)
4.	acb	acbc	predict(3)
5.	aScb	acbc	predict(3)
6.	cb	cbcb	scan(4)
7.	Scb	cbcb	scan(5)
8.	b	cbcb	scan(6)
9.	aAcbc	cbcb	predict(7), ist aber zu lang
10.	$\varepsilon$	cb	scan(8)

3 Pkte

**Aufgabe 7 (5 Pkte)** Betrachten Sie die PCFG mit  $N = \{S, A, B\}$ ,  $T = \{a, b, c\}$ , Startsymbol  $S$  und folgenden Produktionen mit Wahrscheinlichkeiten:

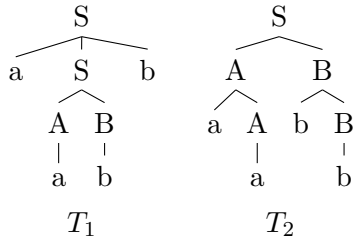
$$\begin{array}{lll}
 4 \cdot 10^{-1}: S \rightarrow aSb & 1 \cdot 10^{-1}: S \rightarrow AB & 5 \cdot 10^{-1}: S \rightarrow c \\
 3 \cdot 10^{-1}: A \rightarrow aA & 1 \cdot 10^{-1}: A \rightarrow a & 6 \cdot 10^{-1}: A \rightarrow c \\
 2 \cdot 10^{-1}: B \rightarrow bB & 2 \cdot 10^{-1}: B \rightarrow b & 6 \cdot 10^{-1}: B \rightarrow c
 \end{array}$$

Betrachten Sie die Eingabe  $w = aabb$ .

- Geben Sie die beiden Parsbäume für  $w$  mit ihren Wahrscheinlichkeiten an.
- Wie hoch ist die Wahrscheinlichkeit von  $aabb$ ?

Lösung:

- zwei Parsbäume:



$$p(T_1) = 4 \cdot 1 \cdot 1 \cdot 2 \cdot 10^{-4} = 8 \cdot 10^{-4}$$

$$p(T_2) = 1 \cdot 3 \cdot 1 \cdot 2 \cdot 2 \cdot 10^{-5} = 12 \cdot 10^{-5}$$

4 Pkte

2.  $p(aabb) = (80 + 12) \cdot 10^{-5} = 0.00092$

1 Pkt

**Aufgabe 8 (7 Pkte)** Betrachten Sie nochmals die PCFG aus der vorhergehenden Aufgabe.

Betrachten Sie die Eingabe  $w = aaabbb$ .

1. Wie hoch ist die Insidewahrscheinlichkeit von  $S$  mit Positionen 2, 5, bezogen auf diese Eingabe?
2. Wie hoch ist die Outsidewahrscheinlichkeit von  $A$  mit Positionen 2, 3, bezogen auf diese Eingabe?

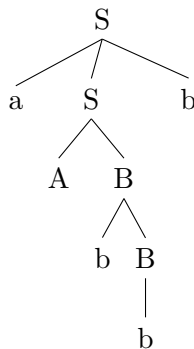
Lösung:

1. Es gibt zwei Möglichkeiten, aus  $S$  den Teil der Eingabe von 2 bis 5, also  $aabb$ , abzuleiten (siehe vorhergehende Lösung).

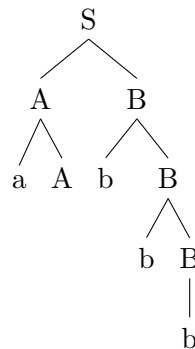
Daraus ergibt sich  $\beta_S(2, 5) = (80 + 12) \cdot 10^{-5} = 0.00092$ .

2 Pkte

2. Es gibt zwei Möglichkeiten, aus  $S$  die Folge  $aAbbb$  (also den Outside Teil  $\alpha_A(2, 3)$ ) abzuleiten.



Wahrscheinlichkeit  
 $0.4 \cdot 0.1 \cdot 0.2 \cdot 0.2$   
 $= 16 \cdot 10^{-4}$



Wahrscheinlichkeit  
 $0.1 \cdot 0.3 \cdot 0.2 \cdot 0.2 \cdot 0.2$   
 $= 24 \cdot 10^{-5}$

2 Pkte

2 Pkte

Damit ist  $\alpha_A(2, 3) = (160 + 24) \cdot 10^{-5} = 0.00184$

1 Pkt