

Einführung in die Computerlinguistik Abschlussklausur

Laura Kallmeyer

Sommersemester 2018, Heinrich-Heine-Universität Düsseldorf

17.07.2018

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

Aufgabe 1 (7 Pkte) Betrachten Sie folgende CFG:

$G = (\{S, A, B, C\}, \{a, b, c, d\}, \{S \rightarrow CSB \mid CS \mid A, A \rightarrow a, B \rightarrow b \mid d, C \rightarrow c\}, S)$

1. Geben Sie alle Linksableitungen für das Wort $w = cccabd$ an.
2. Welche Sprache wird von G erzeugt?
3. Ist G eine ambige CFG?

Lösung:

1. $S \Rightarrow CS \Rightarrow cS \Rightarrow cCSB \Rightarrow ccSB \Rightarrow ccCSBB \Rightarrow cccSBB \Rightarrow cccABB \Rightarrow cccaBB \Rightarrow cccabB \Rightarrow cccabd$
 $S \Rightarrow CSB \Rightarrow cSB \Rightarrow cCSB \Rightarrow ccSB \Rightarrow ccCSBB \Rightarrow cccSBB \Rightarrow cccABB \Rightarrow cccaBB \Rightarrow cccabB \Rightarrow cccabd$
 $S \Rightarrow CSB \Rightarrow cSB \Rightarrow cCSBB \Rightarrow ccSBB \Rightarrow ccCSBB \Rightarrow cccSBB \Rightarrow cccABB \Rightarrow cccaBB \Rightarrow cccabB \Rightarrow cccabd$

3 Pkte

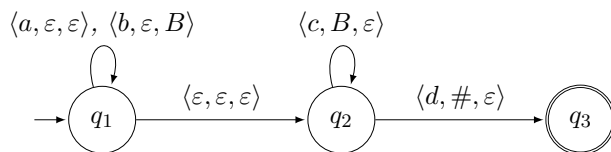
2. $\{c^n aw \mid n \geq 0, w \in \{b, d\}^*, n \geq |w|\}$

3 Pkte

3. Ja, da z.B. w drei verschiedene Linksableitungen, also auch drei verschiedene Parsbäume hat.

1 Pkt

Aufgabe 2 (7 Pkte) Betrachten Sie den Kellerautomaten (PDA) $M = (\{q_1, q_2, q_3\}, \{a, b, c, d\}, \{\#, B\}, \delta, q_1, \#, \{q_3\})$ mit der wie folgt festgelegten Überföhrungsfunktion δ :



1. Geben Sie an, welche Folge von Konfigurationen ("instantaneous descriptions") der Automat durchläuft, wenn die Eingabe $w = abbaccd$ abgearbeitet wird (nur die erfolgreichen Konfigurationen sind gefragt). Geben Sie dies in einer Tabelle wie folgt an:

Zustand	Resteingabe	Stack (Top ist links)
q_1	$abbaccd$	$\#$
q_1	$bbaccd$	$\#$
q_1	$baccd$	$B\#$
	...	

2. Welche Sprache akzeptiert der Kellerautomat bei a) Akzeptanz mit Endzustand ($L(M)$) und bei b) Akzeptanz mit leerem Stack ($N(M)$)?

Lösung:

	Zustand	Resteingabe	Stack		
	q_1	abbaccd	#		
	q_1	bbaccd	#		
	q_1	baccd	B#		
1.	q_1	accd	BB#	3 Pkte	
	q_1	ccd	BB#		
	q_2	ccd	BB#		
	q_2	cd	B#		
	q_2	d	#		
	q_3	-	-		
2.	$L(M) = N(M) = \{wc^n d \mid w \in \{a, b\}^*, w _b = n, n \geq 0\}$				4 Pkte

Aufgabe 3 (5 Pkte) Betrachten Sie nun nochmals die CFG aus Aufgabe 1. Transformieren Sie diese in eine äquivalente CFG in Chomsky Normalform, die keine nutzlosen Symbole enthält. (Tipp: Sie müssen zunächst die unäre Produktion entfernen, dann nutzlose Symbole entfernen und anschließend können Sie binarisieren.)

Lösung:

Ursprungsgrammatik:

$$G = \langle \{S, A, B, C\}, \{a, b, c, d\}, \{S \rightarrow CSB \mid CS \mid A, A \rightarrow a, B \rightarrow b \mid d, C \rightarrow c\}, S \rangle$$

Entfernen von $S \rightarrow A$ führt zu

$$G' = \langle \{S, A, B, C\}, \{a, b, c, d\}, \{S \rightarrow CSB \mid CS \mid a, A \rightarrow a, B \rightarrow b \mid d, C \rightarrow c\}, S \rangle \quad 2 \text{ Pkt}$$

anschließendes Entfernen des nutzlosen A:

$$G'' = \langle \{S, A, B, C\}, \{a, b, c, d\}, \{S \rightarrow CSB \mid CS \mid a, B \rightarrow b \mid d, C \rightarrow c\}, S \rangle \quad 1 \text{ Pkt}$$

Binärisierung:

$$G''' = \langle \{S, A, B, C, X\}, \{a, b, c, d\}, \{S \rightarrow CX \mid CS \mid a, X \rightarrow SB, B \rightarrow b \mid d, C \rightarrow c\}, S \rangle \quad 2 \text{ Pkte}$$

Aufgabe 4 (8 Pkte)

1. Geben Sie für folgende als Attribut-Wert Matrix (AVM) notierte ungetypte Merkmalsstruktur den entsprechenden Graphen an.

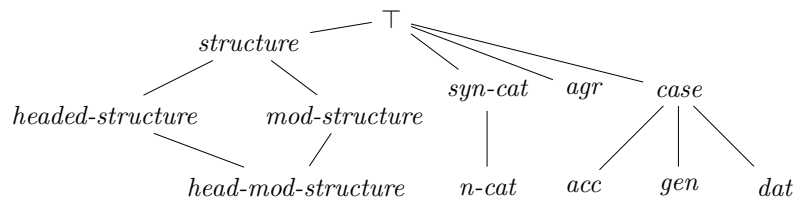
$$\left[\begin{array}{l} \text{HEAD} \\ \text{MOD} \end{array} \left[\begin{array}{l} \left[\begin{array}{l} \text{CAT} \quad \text{PRO} \\ \text{PHON} \quad \text{BÜCHER} \\ \text{AGR} \quad \boxed{2}[\text{CASE ACC}] \end{array} \right] \\ \left[\begin{array}{l} \text{CAT} \quad \text{ADJ} \\ \text{PHON} \quad \text{SPANNENDE} \\ \text{AGR} \quad \boxed{2}[\text{NUM PL}] \end{array} \right] \end{array} \right]$$

2. Betrachten Sie folgende getypte Merkmalsstrukturen:

$$S_1 = \left[\begin{array}{l} \text{headed-structure} \\ \left[\begin{array}{l} \text{n-cat} \\ \text{HEAD} \left[\begin{array}{l} \text{CAT } \boxed{1} n \\ \text{PHON } \textit{Bücher} \\ \text{AGR } \boxed{2} \left[\begin{array}{l} \text{NUM } pl \\ \text{CASE } case \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{AGR } \boxed{2} \\ \text{CAT } \boxed{1} \end{array} \right]$$

$$S_2 = \left[\begin{array}{l} \text{mod-structure} \\ \left[\begin{array}{l} \text{syn-cat} \\ \text{HEAD} \left[\begin{array}{l} \text{CAT } \boxed{4} \\ \text{AGR } \boxed{3} \end{array} \right] \\ \text{AGR } \boxed{3} \\ \text{modifier} \\ \text{CAT } adj \\ \text{PHON } \textit{spannender} \\ \text{MOD} \left[\begin{array}{l} \text{AGR } \boxed{3} \left[\begin{array}{l} \text{agreement} \\ \text{NUM } pl \\ \text{CASE } gen \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

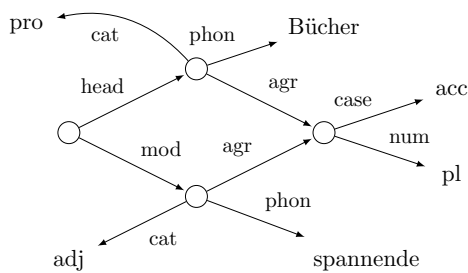
Typenhierarchie:



Berechnen Sie $S_1 \sqcup S_2$. Berücksichtigen Sie dabei auch neue Typen, die sich eventuell ergeben. Falls die Unifikation nicht möglich ist, erklären Sie, warum dies so ist.

Lösung:

1.



4 Pkte

$$2. S_1 \sqcup S_2 = \left[\begin{array}{l} \text{head-mod-structure} \\ \left[\begin{array}{l} \text{n-cat} \\ \text{HEAD} \left[\begin{array}{l} \text{CAT } \boxed{1} n \\ \text{PHON } \textit{Bücher} \\ \text{AGR } \boxed{2} \end{array} \right] \\ \text{AGR } \boxed{2} \\ \text{CAT } \boxed{1} \\ \text{modifier} \\ \text{CAT } adj \\ \text{PHON } \textit{spannender} \\ \text{MOD} \left[\begin{array}{l} \text{AGR } \boxed{2} \left[\begin{array}{l} \text{agreement} \\ \text{NUM } pl \\ \text{CASE } gen \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

4 Pkte

Aufgabe 5 (5 Pkte) Betrachten Sie nun folgende CFG:

$G = \langle \{S, A, B, C\}, \{a, b, c, d\}, \{S \rightarrow cSB \mid CSd \mid a, B \rightarrow b, C \rightarrow c\}, S \rangle$

Geben Sie die erfolgreiche Folge von Stack, Resteingabe und Analysestack an, die sich bei einem Top-Down Parsing der Eingabe $w = ccabd$ ergibt. Die Produktionen seien dabei in der Reihenfolge ihres Auftretens oben numeriert, $S \rightarrow cSB$ bekommt also z.B. den Index $\langle S, 1 \rangle$.

Lösung:

Resteingabe	predict	stack	Analysestack	
ccabd		S	-	
ccabd		CSd	$\langle S, 2 \rangle$	
ccabd		cSd	$\langle C, 1 \rangle, \langle S, 2 \rangle$	
cabd		Sd	$c\langle C, 1 \rangle, \langle S, 2 \rangle$	
cabd		cSBd	$\langle S, 1 \rangle c\langle C, 1 \rangle, \langle S, 2 \rangle$	
abd		SBd	$c\langle S, 1 \rangle c\langle C, 1 \rangle, \langle S, 2 \rangle$	5 Pkte
abd		aBd	$\langle S, 3 \rangle c\langle S, 1 \rangle c\langle C, 1 \rangle, \langle S, 2 \rangle$	
bd		Bd	$a\langle S, 3 \rangle c\langle S, 1 \rangle c\langle C, 1 \rangle, \langle S, 2 \rangle$	
bd		bd	$\langle B, 1 \rangle, a\langle S, 3 \rangle c\langle S, 1 \rangle c\langle C, 1 \rangle, \langle S, 2 \rangle$	
d		d	$b\langle B, 1 \rangle, a\langle S, 3 \rangle c\langle S, 1 \rangle c\langle C, 1 \rangle, \langle S, 2 \rangle$	
ε		ε	$db\langle B, 1 \rangle, a\langle S, 3 \rangle c\langle S, 1 \rangle c\langle C, 1 \rangle, \langle S, 2 \rangle$	

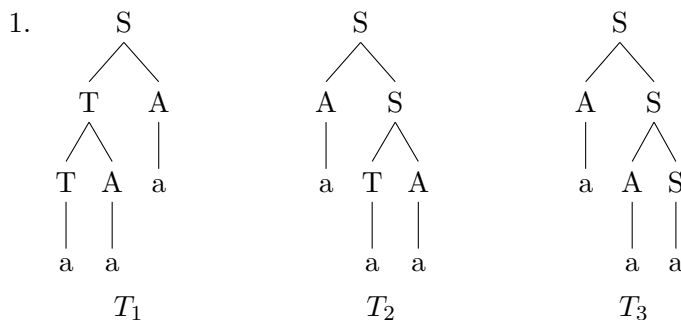
Aufgabe 6 (13 Pkte) Betrachten Sie folgende PCFG:

$G = \langle \{S, A, T, B\}, \{a, b\}, P, S \rangle$ wobei P folgende Produktionen enthält (die Zahlen davor sind die jeweiligen Produktionswahrscheinlichkeiten):

0.1 : $S \rightarrow AS$ 0.2 : $S \rightarrow TA$ 0.6 : $S \rightarrow BB$ 0.1 : $S \rightarrow a$ 1 : $A \rightarrow a$ 1 : $B \rightarrow b$
 0.2 : $T \rightarrow TA$ 0.5 : $T \rightarrow BB$ 0.3 : $T \rightarrow a$

- Geben Sie sämtliche Parsbäume für aaa an und berechnen Sie ihre Wahrscheinlichkeiten.
- Berechnen Sie die Inside Chart für aaa . Geben Sie für die Felder $\alpha_{S,2,3}$ und $\alpha_{S,1,3}$ Ihre Rechnung an.

Lösung:



3 Pkte

Wahrscheinlichkeiten:

$$T_1: 0.2 \cdot 0.2 \cdot 0.3 \cdot 1 \cdot 1 = 12 \cdot 10^{-3}$$

$$T_2: 0.1 \cdot 1 \cdot 0.2 \cdot 0.3 \cdot 1 = 6 \cdot 10^{-3}$$

$$T_3: 0.1 \cdot 1 \cdot 0.1 \cdot 1 \cdot 0.1 = 1 \cdot 10^{-3}$$

3 Pkte

2. Inside chart:

j				
3	$\langle S, 0.019 \rangle, \langle T, 0.012 \rangle$	$\langle S, 0.07 \rangle, \langle T, 0.06 \rangle$	$\langle A, 1 \rangle, \langle S, 0.1 \rangle, \langle T, 0.3 \rangle$	
2	$\langle S, 0.07 \rangle, \langle T, 0.06 \rangle$	$\langle A, 1 \rangle, \langle S, 0.1 \rangle, \langle T, 0.3 \rangle$		
1	$\langle A, 1 \rangle, \langle S, 0.1 \rangle, \langle T, 0.3 \rangle$			
	1	2	3	i

$$\alpha_{S,2,3} = 0.3 \cdot 1 \cdot 0.2 + 1 \cdot 0.1 \cdot 0.1$$

$$\alpha_{S,1,3} = 0.2 \cdot 0.06 \cdot 1 + 0.1 \cdot 1 \cdot 0.07$$

7 Pkte

Aufgabe 7 (5 Pkte) Betrachten Sie nochmals die PCFG aus Aufgabe 6. Geben Sie die CYK Chart (= Viterbi Chart) für ein Parsing von $aaan$, einschließlich Angabe von Produktionen und Zwischenlängen.

Lösung:

l				
2	$\langle 0.06 : T \rightarrow TA, 1 \rangle, \langle 0.06 : S \rightarrow TA, 1 \rangle$			
1	$\langle 1 : A \rightarrow a \rangle, \langle 0.1 : S \rightarrow a \rangle, \langle 0.3 : T \rightarrow a \rangle$			
	1	2	i	

5 Pkte