

Einführung in die Computerlinguistik Abschlussklausur

Laura Kallmeyer

Sommersemester 2014, Heinrich-Heine-Universität Düsseldorf

14.07.2014

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

Aufgabe 1 1. Betrachten Sie folgende CFG:

$$G = \langle \{S\}, \{a, b\}, \{S \rightarrow aSb \mid aS \mid Sb \mid \varepsilon\}, S \rangle$$

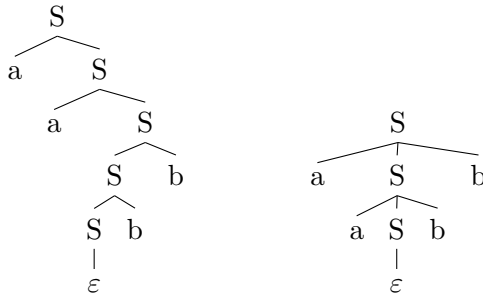
- a) Geben Sie alle möglichen Ableitungen für $w = ab$ an.
- b) Geben Sie zwei mögliche Parsbäume für $w = aabb$ an.
- c) Welche Sprache wird von G erzeugt?

2. Geben Sie eine CFG an, die die Sprache $\{(ab)^n c^{3n} \mid n \geq 1\}$ generiert.

Lösung:

1. a) $S \Rightarrow aSb \Rightarrow ab$
 $S \Rightarrow aS \Rightarrow aSb \Rightarrow ab$
 $S \Rightarrow Sb \Rightarrow aSb \Rightarrow ab$

b) Trees:



c) $\{a^n b^m \mid n, m \geq 0\}$

2. $S \rightarrow abSccc \mid abccc$

Aufgabe 2

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

Produktionen:

$$S \rightarrow aSbc \quad S \rightarrow aAB \quad A \rightarrow \varepsilon \quad B \rightarrow bc \quad B \rightarrow A$$

1. Welche Sprache wird von G erzeugt?
2. a) Berechnen Sie die Menge N_ε der Nichtterminalen, aus denen sich in G ε ableiten lässt.
 b) Geben Sie anschließend eine zu G äquivalente ε -freie CFG G' an.

Lösung:

1. $\{a^n(bc)^m \mid n > 0, n \geq m \geq n - 1\}$
2. a) $N_\varepsilon = \{A, B\}$.
b) G' enthält Produktionen $S \rightarrow aSbc, S \rightarrow a, S \rightarrow aB, B \rightarrow bc$.

Aufgabe 3

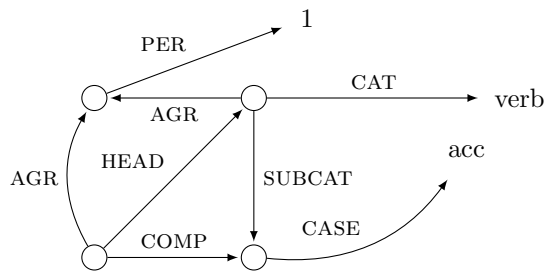
Betrachten Sie die folgenden ungetypten Merkmalsstrukturen, formuliert als Attribut-Wert Matrizen:

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

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:



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

Aufgabe 4 Betrachten Sie die CFG G mit $N = \{S, A\}$, $T = \{a, b\}$, Startsymbol S und Produktionen

$$S \rightarrow aAbb \quad A \rightarrow S \quad A \rightarrow bb$$

1. Ergänzen Sie den Anfang der Trace, die bei einem Shift-Reduce Parsing von $w = abbbb$ 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.	ε	abbbb	
2.	a	bbbb	shift(1)
3.	ab	bbb	shift(2)
4.	abb	bb	shift(3)
5.	aA	bb	reduce(4) $A \rightarrow bb$
6.	abbb	b	shift(4)
7.	aAb	b	shift(5)

2. Ergänzen Sie jetzt den Anfang der Trace, die bei einem Top-Down Parsing von $w = abbbb$ 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.

	Stack	Resteingabe	Operation
1.	S	abbbb	
2.	aAbb	abbbb	predict(1)
3.	Abb	bbbb	scan(2)

Lösung:

1. Shift reduce trace:

	Stack	Resteingabe	Operation
1.	ε	abbbb	
2.	a	bbbb	shift(1)
3.	ab	bbb	shift(2)
4.	abb	bb	shift(3)
5.	aA	bb	reduce(4) $A \rightarrow bb$
6.	abbb	b	shift(4)
7.	aAb	b	shift(5)
8.	abA	b	reduce(6) $A \rightarrow bb$
9.	abbbb	ε	shift(6)
10.	aAbb	ε	shift(7)
11.	abAb	ε	shift(8)
12.	abbA	ε	reduce(9) $A \rightarrow bb$
13.	S	ε	reduce(10) $S \rightarrow aAbb$
14.	aAA	ε	reduce(10) $A \rightarrow bb$
15.	A	ε	reduce(13) $A \rightarrow S$

2. Top down trace:

	Stack	Resteingabe	Operation
1.	S	abbbb	
2.	aAbb	abbbb	predict(1)
3.	Abb	bbbb	scan(2)
4.	Sbb	bbbb	predict(3)
5.	bbbb	bbbb	predict(3)
6.	aAbbbb	bbbb	predict(4)
7.	bbb	bbb	scan(5)
8.	bb	bb	scan(7)
9.	b	b	scan(8)
10.	ε	ε	scan(9)

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

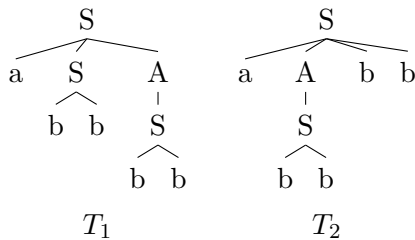
$$\begin{aligned}
 2 \cdot 10^{-1}: S &\rightarrow aAbb & 1 \cdot 10^{-1}: S &\rightarrow bb & 3 \cdot 10^{-1}: S &\rightarrow aSA & 4 \cdot 10^{-1}: S &\rightarrow a \\
 3 \cdot 10^{-1}: A &\rightarrow S & 1 \cdot 10^{-1}: A &\rightarrow cc & 6 \cdot 10^{-1}: A &\rightarrow c
 \end{aligned}$$

Betrachten Sie die Eingabe $w = abbbb$.

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

Lösung:

- zwei Parsbäume:



$$p(T_1) = 3 \cdot 1 \cdot 3 \cdot 1 \cdot 10^{-4} = 9 \cdot 10^{-4}$$

$$p(T_2) = 2 \cdot 3 \cdot 1 \cdot 10^{-3} = 6 \cdot 10^{-3}$$

- $p(abbbb) = (9 + 60) \cdot 10^{-4} = 0.0069$

Aufgabe 6 Betrachten Sie nun folgende PCFG: $N = \{S, A, B, T, U\}$, $T = \{a, b, c\}$, Startsymbol S .

Produktionen mit Wahrscheinlichkeiten:

$$\begin{aligned}
 3 \cdot 10^{-1}: S &\rightarrow AT & 2 \cdot 10^{-1}: S &\rightarrow UB & 5 \cdot 10^{-1}: S &\rightarrow c \\
 3 \cdot 10^{-1}: T &\rightarrow SB & 1 \cdot 10^{-1}: T &\rightarrow b & 6 \cdot 10^{-1}: T &\rightarrow c \\
 1 \cdot 10^{-1}: U &\rightarrow AS & 4 \cdot 10^{-1}: U &\rightarrow a & 5 \cdot 10^{-1}: U &\rightarrow c \\
 1: A &\rightarrow a & 1: B &\rightarrow b
 \end{aligned}$$

Betrachten Sie die Eingabe $w = aabb$.

- Wie hoch ist die Insidewahrscheinlichkeit von S mit Positionen 2, 3, bezogen auf diese Eingabe?

2. Wie hoch ist die Outsidewahrscheinlichkeit von S mit Positionen 2, 3, bezogen auf diese Eingabe?

Lösung:

1. Es gibt zwei Möglichkeiten, aus S den Teil der Eingabe von 2 bis 3, also ab , abzuleiten.

Daraus ergibt sich $\beta_S(2, 3) = 3 \cdot 10^{-2} + 8 \cdot 10^{-2}$.

2. Es gibt zwei Möglichkeiten, aus S die Folge aSb (also den Outside Teil) abzuleiten.

$S \Rightarrow AT \Rightarrow aT \Rightarrow aSB \Rightarrow aSb$, Wahrscheinlichkeit $3 \cdot 1 \cdot 3 \cdot 1 \cdot 10^{-2} = 9 \cdot 10^{-2}$

$S \Rightarrow UB \Rightarrow ASB \Rightarrow aSB \Rightarrow aSb$, Wahrscheinlichkeit $2 \cdot 1 \cdot 1 \cdot 1 \cdot 10^{-2} = 2 \cdot 10^{-2}$

Damit ist $\beta_S(2, 3) = 11 \cdot 10^{-2}$

Aufgabe 7 Betrachten Sie nun erneut die PCFG aus der vorhergehenden Aufgabe mit der Eingabe $w = ab$.

- Welche Chart ergibt sich bei einem probabilistischen CYK Chart Parsing von $w = ab$ mit dieser Grammatik? Verwenden Sie den Algorithmus, der auch Produktionen und Zwischenlängen festhält. Erläutern Sie Ihren Rechenweg für das Chartfeld mit Länge 2.
- Wie sieht der beste Parsbaum für $w = ab$ aus, der sich aus der Chart ergibt?

Lösung:

1. Chart:

l			
2	$8 \cdot 10^{-2} : S \rightarrow UB, 1$		
1	$1 : A \rightarrow a, -, 4 \cdot 10^{-1} : U \rightarrow a, -$	$1 : B \rightarrow b, -, 1 \cdot 10^{-1} : T \rightarrow b, -$	
	1:a	2:b	i

Länge 2, Kategorie S: $3 \cdot 10^{-2} : S \rightarrow AT, 1$ oder $8 \cdot 10^{-2} : S \rightarrow UB, 1$, letzteres ist besser

2. bester Baum:

