

# Einführung in die Computerlinguistik

## Hausaufgabe zu PCFG, Abgabe 25.06.2019

Laura Kallmeyer

Sommer 2019, Heinrich-Heine-Universität Düsseldorf

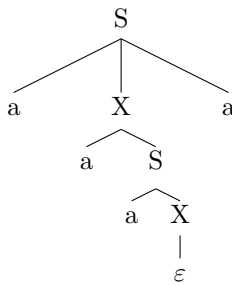
**Aufgabe 1** Betrachten Sie folgende PCFG:

$$G = \langle \{S, X\}, \{a, b\}, \{ \begin{array}{l} 0.6 : S \rightarrow bSb, \\ 0.1 : S \rightarrow aX, \\ 0.3 : S \rightarrow aXa, \\ 0.4 : X \rightarrow aS, \\ 0.3 : X \rightarrow a, \\ 0.3 : X \rightarrow \varepsilon \end{array} \}, S \rangle$$

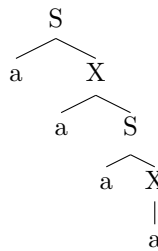
1. Geben Sie für die Eingabe  $w = aaaa$  alle Parsbäume mit ihrer jeweiligen Wahrscheinlichkeit an.
2. Wie ist demnach die Wahrscheinlichkeit  $P(w)$  des Wortes  $w = aaaa$ ?

Lösung:

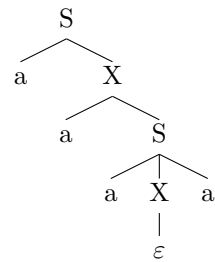
1.  $T_1$ :



$T_2$ :



$T_3$ :



Wahrscheinlichkeiten:  $P(T_1) = P(T_3) = 0.3 \cdot 0.4 \cdot 0.1 \cdot 0.3 = 36 \cdot 10^{-4} = 0.0036$   
 $P(T_2) = 0.1 \cdot 0.1 \cdot 0.4 \cdot 0.3 = 12 \cdot 10^{-4} = 0.0012$

2.  $P(aaaa) = 0.0036 + 0.0036 + 0.0012 = 0.0084$

**Aufgabe 2** Betrachten Sie folgende PCFG:  $G = \langle \{S, B, C\}, \{a, b, c\}, \{$

$$\begin{array}{l} 0.6 : S \rightarrow aCb, \\ 0.4 : S \rightarrow aBb, \\ 0.3 : B \rightarrow BB, \\ 0.7 : B \rightarrow B, \\ 1 : C \rightarrow c \end{array} \}, S \rangle$$

Ist diese PCFG konsistent? Begründen Sie Ihre Antwort.

Lösung:

Nein, denn die Kettensprache ist  $\{acb\}$ , und es gilt  $P(S \Rightarrow_l acb) = 0.6$ .

**Aufgabe 3** Betrachten Sie die PCFG  $G$  mit Nichtterminalen  $\{S, A, B\}$ , Terminalen  $\{a, b\}$ , Startsymbol  $S$  und Produktionen

$$\{ \begin{array}{l} 0.5 \quad S \rightarrow XZ, \\ 0.2 \quad S \rightarrow ZY, \\ 0.3 \quad S \rightarrow XY, \\ 0.8 \quad X \rightarrow XS, \\ 0.2 \quad X \rightarrow a, \\ 0.3 \quad Y \rightarrow SY, \\ 0.7 \quad Y \rightarrow b, \\ 1 \quad Z \rightarrow c \end{array} \}$$

(Vor jeder Produktion steht ihre Wahrscheinlichkeit.)

1. Geben Sie die Inside Chart für die Eingabe  $w = aabb$ .

Vorsicht: Spaltenindex = Anfangsposition, Zeilenindex = Endposition des jeweiligen Spans.

Geben Sie insbesondere den Rechenweg für die Kategorie  $S$  im Feld mit Indizes 1,4 an.

2. Geben Sie die Viterbi chart für ein probabilistisches CYK parsing der Eingabe  $w = aabb$ .

(Spaltenindex = Anfangsposition, Zeilenindex = Länge!)

Solution:

1. Inside Chart:

4	(S, 0.0019404)	(Y, 0.00882)	(Y, 0.7)
3	(X, 0.00672)	(S, 0.042)	(Y, 0.7)
2		(X, 0.2)	
1	(X, 0.2)		
	1	2	3

Feld 1,4: zwei Möglichkeiten:  $0.3 \cdot 0.2 \cdot 0.00882 + 0.3 \cdot 0.00672 \cdot 0.7 = 0.0019404$

2. Viterbi Chart:

4	0.0014112 : S		
3	0.00672 : X	0.00882 : Y	
2		0.042 : S	
1	0.2 : X	0.2 : X	0.7 : Y
	1	2	3

**Aufgabe 4** Betrachten Sie folgende PCFG:

$G = \langle \{N, A\}, \{car, camping, red, blue\}, P, N \rangle$  mit den folgenden Produktionen in  $P$ :

$0.4 : N \rightarrow NN$     $0.2 : N \rightarrow car$     $0.1 : N \rightarrow red$     $0.6 : A \rightarrow blue$

$0.2 : N \rightarrow AN$     $0.1 : N \rightarrow camping$     $0.4 : A \rightarrow red$

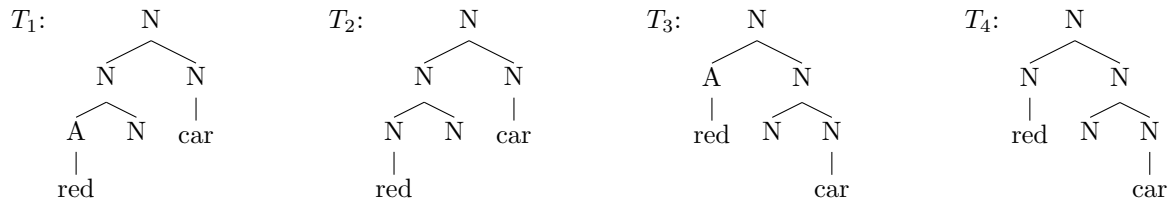
Betrachten Sie die Eingabe  $w = red\ blue\ camping\ car$ .

Geben Sie die Outsidewahrscheinlichkeit von  $N$  mit Indizes 2, 3 bezüglich dieser Eingabe an.

Sie müssen keine Outside Chart berechnen, es reicht, über Wahrscheinlichkeiten möglicher Bäume zu summieren.

Lösung:

Außerhalb von der Teilkette der Eingabe mit Indizes 2, 3 (= *blue camping*) gibt es davor *red* und danach das Wort *car*. Die Outsidewahrscheinlichkeit ergibt sich also als Summe aller Parsbäume mit Wurzel  $N$  (Startsymbol) und String *red N car* in den Blättern.



$$P(N \xrightarrow{*} red\ N\ car) = 2 \cdot 0.4 \cdot 0.2 \cdot 0.4 \cdot 0.2 + 2 \cdot 0.4 \cdot 0.4 \cdot 0.1 \cdot 0.2 = 128 \cdot 10^{-4} + 64 \cdot 10^{-4} = 192 \cdot 10^{-4} = 0.0192$$