

# Einführung in die Computerlinguistik

## Hausaufgabe zu PCFG, Abgabe 06.07.2020

Laura Kallmeyer

Sommer 2020, Heinrich-Heine-Universität Düsseldorf

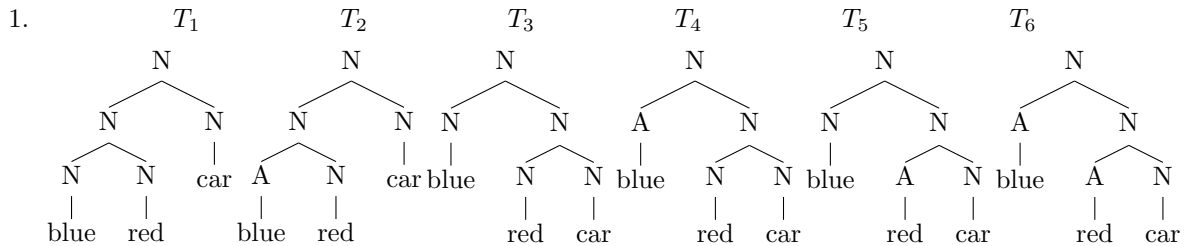
**Aufgabe 1** Betrachten Sie folgende PCFG:

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

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

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

Lösung:



$$P(T_1) = P(T_3) = 0.4 \cdot 0.4 \cdot 0.2 \cdot 0.1 \cdot 0.1 = 32 \cdot 10^{-5}$$

$$P(T_2) = P(T_4) = 4 \cdot 2 \cdot 6 \cdot 1 \cdot 2 \cdot 10^{-5} = 96 \cdot 10^{-5}$$

$$P(T_5) = 4 \cdot 2 \cdot 4 \cdot 1 \cdot 2 \cdot 10^{-5} = 64 \cdot 10^{-5}$$

$$P(T_6) = 2 \cdot 2 \cdot 4 \cdot 6 \cdot 2 \cdot 10^{-5} = 192 \cdot 10^{-5}$$

$$2. P(blue\ red\ car) = P(T_1) + P(T_2) + \dots + P(T_6) = (2 \cdot 32 + 2 \cdot 96 + 64 + 192) \cdot 10^{-5} = 512 \cdot 10^{-5}$$

**Aufgabe 2** Betrachten Sie erneut die PCFG aus Aufgabe 1.

1. Geben Sie die Inside Chart für die Eingabe  $w = red\ blue\ car$ .  
*Vorsicht: Spaltenindex = Anfangsposition, Zeilenindex = Endposition des jeweiligen Spans.*  
 Geben Sie insbesondere den Rechenweg für die Kategorie  $N$  in den Feldern mit Indizes  $\langle 1,2 \rangle$ ,  $\langle 2,3 \rangle$  und  $\langle 1,3 \rangle$  an.
2. Geben Sie die Viterbi chart für ein probabilistisches CYK parsing der Eingabe  $w = red\ blue\ car$ .  
 Erläutern Sie Ihren Rechenweg für die Kategorie  $N$  in dem Feld mit Indizes 1,3.  
 (Spaltenindex = Anfangsposition, Zeilenindex = Länge!)

Solution:

1. Inside Chart:

3	$(N, 0.0048)$	$(N, 0.032)$	$(N, 0.2)$
2	$(N, 0.012)$	$(N, 0.1), (A, 0.6)$	
1	$(N, 0.1), (A, 0.4)$		
	1	2	3

Feld 1,2, Kategorie N: zwei Möglichkeiten:  $0.4 \cdot 0.1 \cdot 0.1 + 0.2 \cdot 0.4 \cdot 0.1 = 0.012$

Feld 2,3, Kategorie N: zwei Möglichkeiten:  $0.4 \cdot 0.1 \cdot 0.2 + 0.2 \cdot 0.6 \cdot 0.2 = 0.008 + 0.024 = 0.032$

Feld 1,3, Kategorie N: drei Möglichkeiten:  $0.4 \cdot 0.1 \cdot 0.032 + 0.4 \cdot 0.012 \cdot 0.2 + 0.2 \cdot 0.4 \cdot 0.032 = 0.00128 + 0.00096 + 0.00256 = 0.0048$

2. Viterbi Chart:

3	0.00192 : N		
2	0.008 : N	0.024 : N	
1	0.1 : N, 0.4 : A	0.1 : N, 0.6 : A	0.2 : N
	1	2	3

Feld 1,3: drei Möglichkeiten:

$N \rightarrow NN: 0.4 \cdot 0.008 \cdot 0.2 = 0.00064$

$N \rightarrow NN: 0.4 \cdot 0.1 \cdot 0.024 = 0.00096$

$N \rightarrow AN: 0.2 \cdot 0.4 \cdot 0.024 = 0.00192$

Das dritte ist der größte Wert, damit also auch die Wahrscheinlichkeit des besten Parsbaums.

**Aufgabe 3** Betrachten Sie wieder die PCFG aus Aufgabe 1.

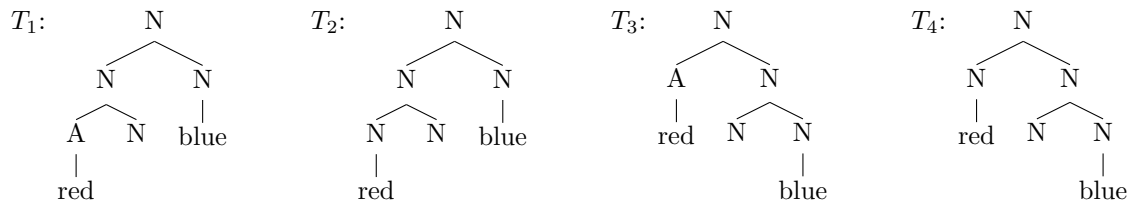
Betrachten Sie die Eingabe  $w = \text{red blue car blue}$ .

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 car*) gibt es davor *red* und danach das Wort *blue*. Die Outsidewahrscheinlichkeit ergibt sich also als Summe aller Parsbäume mit Wurzel  $N$  (Startsymbol) und String *red N blue* in den Blättern.



$$P(N \xrightarrow{*} \text{red } N \text{ car}) = 2 \cdot 0.4 \cdot 0.1 \cdot 0.4 \cdot 0.1 + 2 \cdot 0.4 \cdot 0.1 \cdot 0.2 \cdot 0.4 = 32 \cdot 10^{-4} + 64 \cdot 10^{-4} = 96 \cdot 10^{-4} = 0.0096$$