Einführung in die Computerlinguistik
Hausaufgabe zu PCFG, Abgabe 05.07.2021 (35 Pkte)

Laura Kallmeyer

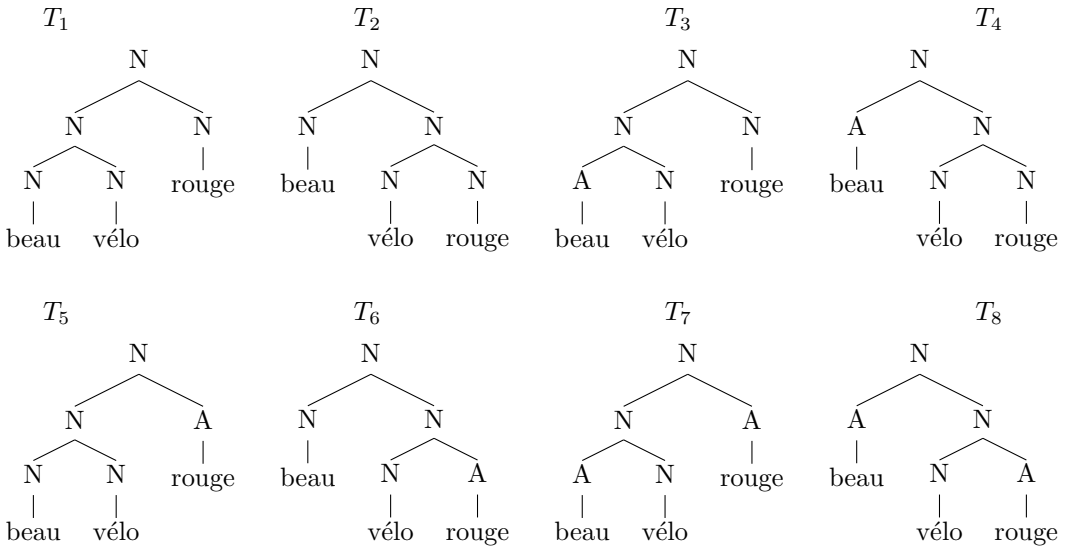
Sommer 2021, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1 *Betrachten Sie folgende PCFG:*
$G = \langle \{N, A\}, \{$*vélo, rouge, beau*$\}$[1]$P, N\rangle$ *mit den folgenden Produktionen in* P*:*

$0.3 : N \to NN$	$0.1 : N \to$ *vélo*	$0.1 : N \to$ *rouge*	$0.6 : A \to$ *beau*
$0.2 : N \to AN$	$0.2 : N \to NA$	$0.1 : N \to$ *beau*	$0.4 : A \to$ *rouge*

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

Lösung:

1. 

   $P(T_1) = P(T_2) = 0.3 \cdot 0.3 \cdot 0.1 \cdot 0.1 \cdot 0.1 = 9 \cdot 10^{-5}$
   $P(T_3) = P(T_4) = 0.3 \cdot 0.2 \cdot 0.6 \cdot 0.1 \cdot 0.1 = 36 \cdot 10^{-5}$
   $P(T_5) = P(T_6) = 0.3 \cdot 0.2 \cdot 0.4 \cdot 0.1 \cdot 0.1 = 24 \cdot 10^{-5}$
   $P(T_7) = P(T_8) = 0.2 \cdot 0.2 \cdot 0.6 \cdot 0.1 \cdot 0.4 = 96 \cdot 10^{-5}$
2. $P($*beau vélo rouge*$) = \sum_{i=1}^{8} P(T_i) = 2(9 + 36 + 24 + 96) \cdot 10^{-5} = 2 \cdot 165 \cdot 10^{-5} = 330 \cdot 10^{-5}$

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

1. *Geben Sie die Inside Chart für die Eingabe* $w =$*beau vélo rouge.*
   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.*

[1] Französisch: vélo 'Fahrrad', rouge 'red', beau 'beautiful'

2. *Geben Sie die Viterbi chart für ein probabilistisches CYK parsing der Eingabe w=beau vélo rouge. Erläutern Sie Ihren Rechenweg für die Kategorie N in dem Feld mit Indizes 1,3. (Spaltenindex = Anfangsposition, Zeilenindex = Länge!)*

Solution:

1. Inside Chart:

3	$(N, 0.0033)$	$(N, 0.011)$	$(N, 0.1)$, $(A, 0.4)$
2	$(N, 0.015)$	$(N, 0.1)$	
1	$(N, 0.1)$, $(A, 0.6)$		
	1	2	3

Feld 1,2, Kategorie N: zwei Möglichkeiten:
$0.3 \cdot 0.1 \cdot 0.1(N \to NN) + 0.2 \cdot 0.6 \cdot 0.1(N \to AN) = 0.015$

Feld 2,3, Kategorie N: zwei Möglichkeiten:
$0.3 \cdot 0.1 \cdot 0.1(N \to NN) + 0.2 \cdot 0.4 \cdot 0.1(N \to NA) = 0.011$

Feld 1,3, Kategorie N: vier Möglichkeiten:
$0.3 \cdot 0.1 \cdot 0.011(N \to NN) + 0.3 \cdot 0.015 \cdot 0.1(N \to NN)$
$+0.2 \cdot 0.6 \cdot 0.011(N \to AN) + 0.2 \cdot 0.015 \cdot 0.4(N \to NA)$
$= 33 \cdot 10^{-5} + 45 \cdot 10^{-5} + 132 \cdot 10^{-5} + 120 \cdot 10^{-5} = 330 \cdot 10^{-5} = 0.0033$

2. Viterbi Chart:

3	$0.00096 : N$		
2	$0.012 : N$	$0.008 : N$	
1	$0.1 : N$, $0.6 : A$	$0.1 : N$	$0.1 : N$, $0.4 : A$
	1	2	3

Feld 1,3: vier Möglichkeiten:

$N \to NN$: $0.3 \cdot 0.1 \cdot 0.008 = 0.00024$

$N \to NN$: $0.3 \cdot 0.012 \cdot 0.1 = 0.00036$

$N \to AN$: $0.2 \cdot 0.6 \cdot 0.008 = 0.00096$

$N \to NA$: $0.2 \cdot 0.012 \cdot 0.4 = 0.00096$

Die dritte und vierte Möglichkeit sind gleich gut, 0.00096 ist also die Wahrscheinlichkeit des besten Parsbaums.

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

Betrachten Sie die Eingabe w = beau vélo vélo rouge rouge.

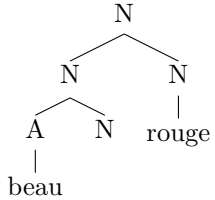
Geben Sie die Outsidewahrscheinlichkeit von N mit Indizes 2, 4 *bezüglich dieser Eingabe an.*

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

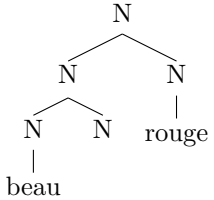
Lösung:

Außerhalb von der Teilkette der Eingabe mit Indizes 2, 4 (= *vélo vélo rouge*) gibt es davor *beau* und danach das Wort *rouge*. Die Outsidewahrscheinlichkeit ergibt sich also als Summe aller Parsbäume mit Wurzel N (Startsymbol) und String *beau N rouge* in den Blättern.

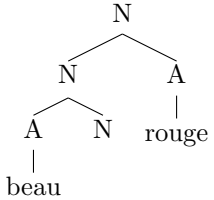
T_1:



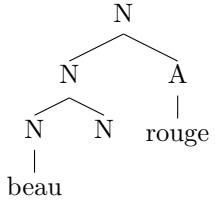
T_2:

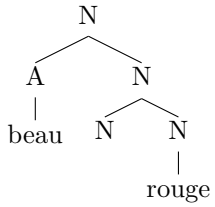


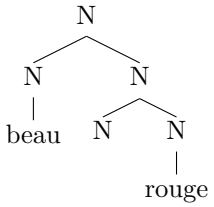
T_3:

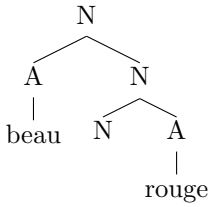


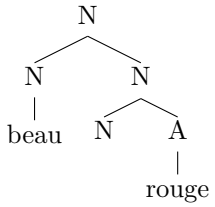
T_4:



T_5: 

T_6: 

T_7: 

T_8: 

$P(N \overset{*}{\Rightarrow} \textit{beau}\ N\ \textit{rouge}) = 2{\cdot}0.3{\cdot}0.2{\cdot}0.6{\cdot}0.1{+}2{\cdot}0.3{\cdot}0.3{\cdot}0.1{\cdot}0.1{+}2{\cdot}0.3{\cdot}0.2{\cdot}0.4{\cdot}0.1{+}2{\cdot}0.2{\cdot}0.2{\cdot}0.6{\cdot}0.4 = 0.033$