

# Einführung in die Computerlinguistik

## Hausaufgabe 10, Abgabe 02.07.2012

Laura Kallmeyer

SS 2012, Heinrich-Heine-Universität Düsseldorf

**Aufgabe 1** Betrachten Sie eine PCFG mit folgenden Produktionen und mit Startsymbol  $S$ :

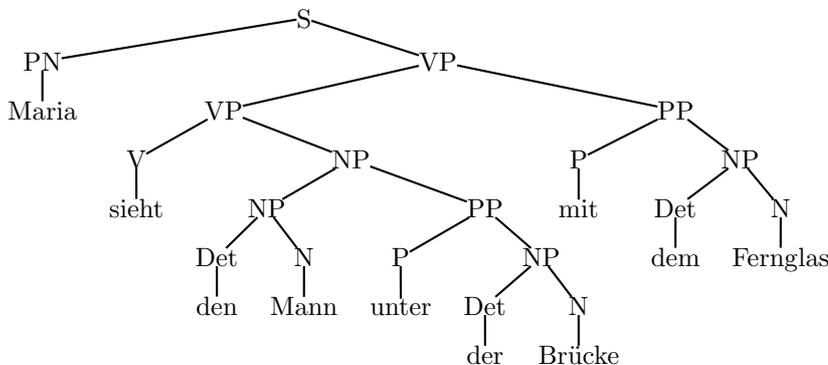
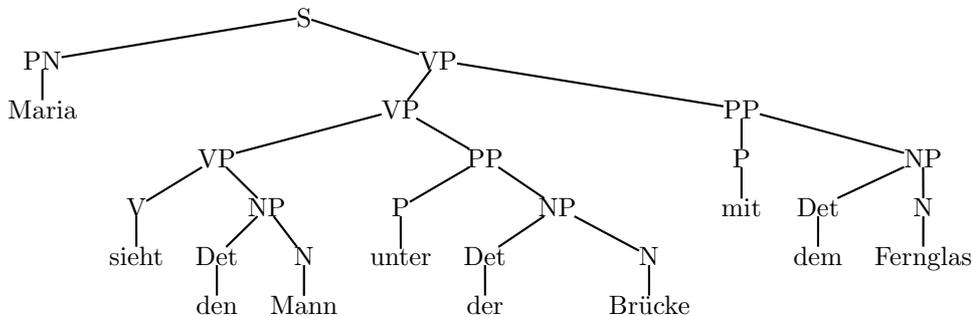
- |                           |                               |   |
|---------------------------|-------------------------------|---|
| 1 $S \rightarrow PN VP$   |                               |   |
| .7 $NP \rightarrow Det N$ | .3 $NP \rightarrow NP PP$     |   |
| .6 $VP \rightarrow V NP$  | .4 $VP \rightarrow VP PP$     |   |
| .7 $PP \rightarrow P NP$  | .3 $PP \rightarrow P PP$      |   |
| .5 $V \rightarrow sieht$  | .4 $V \rightarrow beobachtet$ | .1 $V \rightarrow vergisst$                           |
| .4 $P \rightarrow mit$    | .2 $P \rightarrow unter$      | .4 $P \rightarrow auf$                                |
| .2 $Det \rightarrow den$  | .4 $Det \rightarrow der$      | .4 $Det \rightarrow dem$                              |
| .4 $N \rightarrow Mann$   | .2 $N \rightarrow Brücke$     | .1 $N \rightarrow Fernglas$ .3 $N \rightarrow Straße$ |
| .4 $PN \rightarrow Maria$ | .6 $PN \rightarrow Paul$      |   |

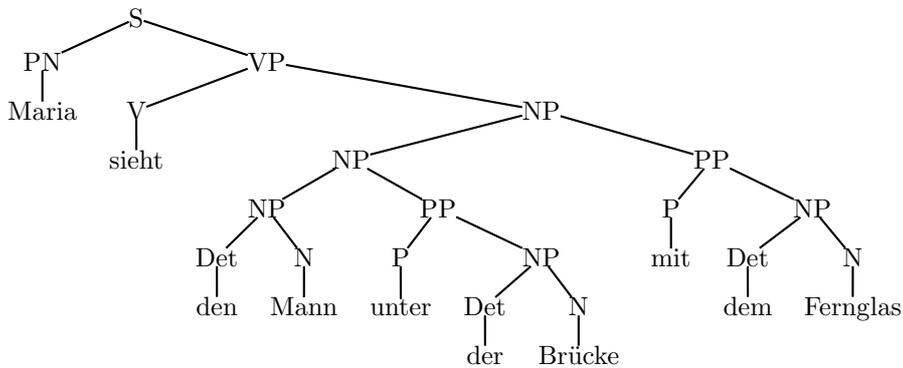
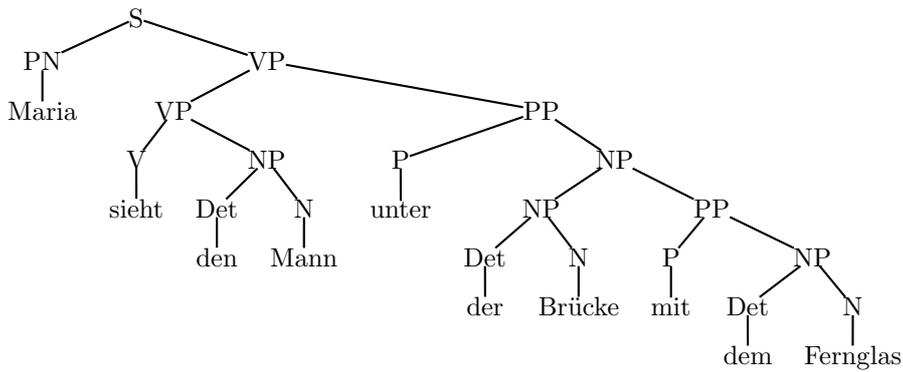
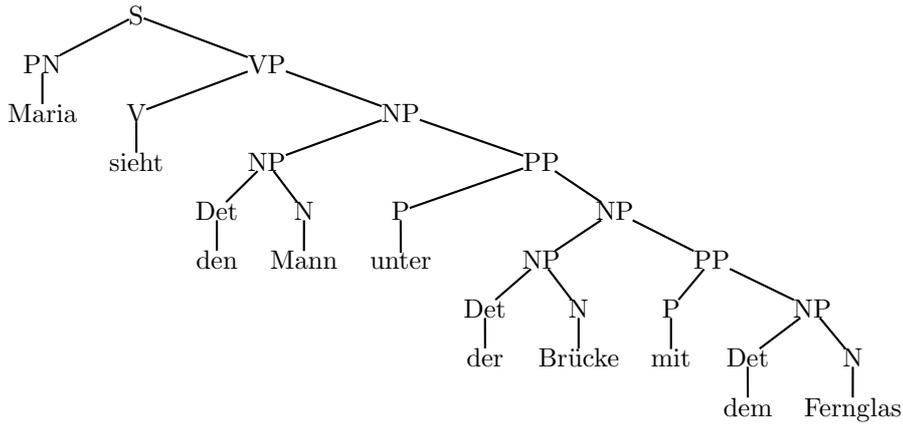
Nehmen Sie an, Sie haben den Eingabesatz  $w =$  Maria sieht den Mann unter der Brücke mit dem Fernglas.

1. Wieviel Lesarten hat diese Eingabe? Geben Sie die verschiedenen Ableitungsbäume an.
2. Geben Sie die Wahrscheinlichkeit von  $w$  an.
3. Geben Sie die Insidewahrscheinlichkeit von  $VP$  mit Indizes 2, 7 bezüglich dieses Satzes an.
4. Geben Sie die Outsidewahrscheinlichkeit von  $PP$  mit Indizes 5, 10 bezüglich dieses Satzes an.

Lösung:

1. Es gibt 5 Parsbäume, also 5 Lesarten.





2. Es gibt 5 Linksableitungen, deren Wahrscheinlichkeiten summiert werden müssen (siehe die 5 Parsbäume). Jede einzelne Ableitung hat das Produkt der Wahrscheinlichkeiten ihrer Produktionen als Wahrscheinlichkeit.

$$p(t_1) = 0,4 \cdot 0,5 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,4 \cdot 1 \cdot 0,4 \cdot 0,4 \cdot 0,6 \cdot 0,7 \cdot 0,7 \cdot 0,7 \cdot 0,7 \cdot 0,7 = 6,608781 \cdot 10^{-8}$$

$$p(t_2) = 0,4 \cdot 0,5 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,4 \cdot 1 \cdot 0,4 \cdot 0,6 \cdot 0,3 \cdot 0,7 \cdot 0,7 \cdot 0,7 \cdot 0,7 \cdot 0,7 = 4,956586 \cdot 10^{-8}$$

$$p(t_3) = 0,4 \cdot 0,5 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,4 \cdot 1 \cdot 0,6 \cdot 0,3 \cdot 0,7 \cdot 0,7 \cdot 0,3 \cdot 0,7 \cdot 0,7 \cdot 0,7 = 3,717439 \cdot 10^{-8}$$

$$p(t_4) = 0,4 \cdot 0,5 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,4 \cdot 1 \cdot 0,4 \cdot 0,6 \cdot 0,7 \cdot 0,7 \cdot 0,3 \cdot 0,7 \cdot 0,7 \cdot 0,7 = 4,956585 \cdot 10^{-8}$$

$$p(t_5) = 0,4 \cdot 0,5 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,2 \cdot 0,4 \cdot 0,4 \cdot 1 \cdot 0,6 \cdot 0,3 \cdot 0,3 \cdot 0,7 \cdot 0,7 \cdot 0,7 \cdot 0,7 \cdot 0,7 = 3,717439 \cdot 10^{-8}$$

$$p(w) = p(t_1) + p(t_2) + p(t_3) + p(t_4) + p(t_5) = 6,608781 \cdot 10^{-8} + 4,956586 \cdot 10^{-8} + 3,717439 \cdot 10^{-8} + 4,956585 \cdot 10^{-8} + 3,717439 \cdot 10^{-8} = 2,3956832256 \cdot 10^{-7}$$

3. Man kann sich überlegen, dass es nur zwei Bäume mit Wurzel VP und Kette *sieht den Mann unter der Brücke* geben kann (siehe die entsprechenden VP Teilbäume in den ersten beiden Ableitungsbäumen für *w*). Für diese beiden Bäume muss man jeweils das Produkt der Wahrscheinlichkeiten der verwendeten Produktionen bilden und anschließend über beide Ergebnisse summieren. Die beiden VP Bäume unterscheiden sich nur in der Produktion, die die PP anbindet, alles andere ist gleich. Damit ergibt sich

$$\beta_{VP}(2, 7) = (0.4 + 0.3) \cdot 0.6 \cdot 0.5 \cdot 0.7 \cdot 0.2 \cdot 0.4 \cdot 0.7 \cdot 0.2 \cdot 0.7 \cdot 0.4 \cdot 0.2 = 9219840 \cdot 10^{-11} = 0.0000921984$$

Man kann auch den Inside Algorithmus so weit nötig durchrechnen und dann die gesuchte Wahrscheinlichkeit ablesen:

Schritt 1:

$$\begin{aligned} \beta_V(2, 2) &= 0,5 & \beta_{Det}(3, 3) &= 0,2 & \beta_N(4, 4) &= 0,4 \\ \beta_P(5, 5) &= 0,2 & \beta_{Det}(6, 6) &= 0,4 & \beta_N(7, 7) &= 0,2 \end{aligned}$$

Schritt 2 (nur die notwendigen Felder müssen berechnet werden, siehe die ersten beiden Bäume):

$$\beta_{NP}(6, 7) = \beta_{Det}(6, 6) \cdot \beta_N(7, 7) \cdot p(NP \rightarrow Det N) = 0,4 \cdot 0,2 \cdot 0,7 = 0,056$$

$$\beta_{PP}(5, 7) = \beta_P(5, 5) \cdot \beta_{NP}(6, 7) \cdot p(PP \rightarrow P NP) = 0,056 \cdot 0,2 \cdot 0,7 = 0,00784$$

$$\beta_{NP}(3, 4) = \beta_{Det}(3, 3) \cdot \beta_N(4, 4) \cdot p(NP \rightarrow Det N) = 0,2 \cdot 0,4 \cdot 0,7 = 0,056$$

$$\beta_{VP}(2, 4) = \beta_V(2, 2) \cdot \beta_{NP}(3, 4) \cdot p(VP \rightarrow V NP) = 0,5 \cdot 0,056 \cdot 0,6 = 0,0168$$

$$\beta_{NP}(3, 7) = \beta_{NP}(3, 4) \cdot \beta_{PP}(5, 7) \cdot p(NP \rightarrow NP PP) = 0,056 \cdot 0,00784 \cdot 0,3 = 0,000131712$$

$$\beta_{VP}(2, 7) = \beta_{VP}(2, 4) \cdot \beta_{PP}(5, 7) \cdot p(VP \rightarrow VP PP) + \beta_V(2, 2) \cdot \beta_{NP}(3, 7) \cdot p(VP \rightarrow V NP) = 0,0168 \cdot 0,00784 \cdot 0,4 + 0,5 \cdot 0,000131712 \cdot 0,6 = 0,000052684 + 0,000039513 = 0,000092197$$

(Der Unterschied entstand aus Rundungsfehlern.)

4. Es gibt zwei Möglichkeiten, aus *S Maria sieht den Mann PP* abzuleiten, die einen Teil Produktionen gemeinsam haben. Einziger Unterschied: Anbindung der PP. Daraus ergibt sich

$$0.5 \cdot 1 \cdot 0.4 \cdot 0.6 \cdot 0.7 \cdot 0.2 \cdot 0.4 \cdot (0.3 + 0.4) = 0,004704$$