Der **Inside-Outside-Algorithmus**

- berechnet effizient die **erwarteten Regelhäufigkeiten** beim EM-Training
- und entspricht damit dem Forward-Backward-Algorithmus bei den HMMs.
- Er berechnet bottom-up **Inside**-Wahrscheinlichkeiten (ähnlich dem Viterbi-Algorithmus)
- und top-down **Outside**-Wahrscheinlichkeiten.
- Aus den Inside- und Outside-Wahrscheinlichkeiten berechnet er die **erwarteten Häufigkeiten**.

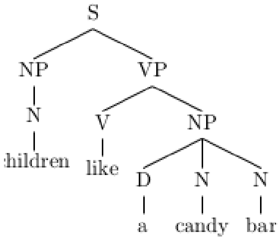
Ansatz 3: (Forts.)

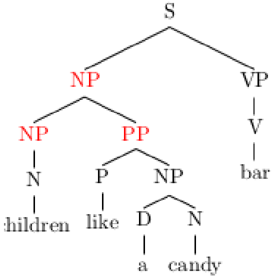
EM-Training

1. Initialisierung der Regelwahrscheinlichkeiten
2. Berechnung der Parsebaumgewichte $p(t|s)$
3. Extraktion der gewichteten Regelhäufigkeiten
4. Neuschätzung der Regelwahrscheinlichkeiten
5. Weiter mit Schritt 2

Problem: Wie können die gewichteten Regelhäufigkeiten effizient für hochambige Sätze berechnet werden?

Lösung: Inside-Outside-Algorithmus (wird etwas später vorgestellt)





Regel	p_0	f_1	p_1	f_2
S → NP VP	1.00	2.00	1.00	2.00
NP → D N	0.25	0.50	0.11	0.10
NP → D N N	0.25	0.50	0.11	0.90
NP → N	0.25	3.00	0.67	3.00
NP → NP PP	0.25	**0.50**	0.11	**0.10**
VP → V	0.50	0.50	0.25	0.10
VP → V NP	0.50	1.50	0.75	1.90
PP → P NP	1.00	0.50	1.00	0.10
D → a	0.50	1.00	1.00	1.00
D → the	0.50	0.00	0.00	0.00
N → bar	0.25	0.50	0.11	0.90
N → candy	0.25	1.00	0.22	1.00

Wie berechnet man die erwartet Häufigkeit einer Regel mit Inside-Outside-Algorithmus?

Neuschätzung der Regel-Wahrscheinlichkeiten

EM-Algorithmus: wiederholte Ausführung der beiden Schritte

1. E-Schritt
   - Jeder Satz des Trainingskorpus wird geparst und ein Parsewald ausgegeben.
   - Die erwartete Häufigkeit jeder Parsewaldregel wird berechnet.
   - Die erwarteten Häufigkeiten werden für jede CFG-Regel über alle Trainingssätze summiert.
2. M-Schritt
   Die Regel-Wahrscheinlichkeiten werden aus den erwarteten Häufigkeiten neu geschätzt:

$$p(A \rightarrow \delta) = \frac{\gamma(A \rightarrow \delta)}{\sum_{\delta'} \gamma(A \rightarrow \delta')}$$

Neuschätzung der Regel-Wahrscheinlichkeiten

EM-Algorithmus: wiederholte Ausführung der beiden Schritte

1. E-Schritt
   - Jeder Satz des Trainingskorpus wird geparst und ein Parsewald ausgegeben.
   - Die erwartete Häufigkeit jeder Parsewaldregel wird berechnet.
   - Die erwarteten Häufigkeiten werden für jede CFG-Regel über alle Trainingssätze summiert.
2. M-Schritt
   Die Regel-Wahrscheinlichkeiten werden aus den erwarteten Häufigkeiten neu geschätzt:

$$p(A \to \delta) = \frac{\gamma(A \to \delta)}{\sum_{\delta'} \gamma(A \to \delta')}$$

Z.B Für alle Sätze(Parsewälder) berechnen wir die erwartete-HF für Regel NP → N PP , dann summieren wir diese HFs und damit bekommen wir (finale) erwartete HF. Dann können wie p(regel) aus dieser HF schätzen.

inside-wk für eine Regel

- erwartete Regelhäufigkeit

$$\gamma(A \to \delta) = \alpha(A \to \delta)\ \beta(A)/\alpha(S)$$

outside-wk für ein non-terminal Symbol A

inside-wk von Startsymbol S

Inside-Algorithmus

$$\alpha(a) = 1$$

Was ist a ?

$$\alpha(A \to X_1 \ldots X_n) = p(A \to X_1 \ldots X_n) \prod_{i=1}^{n} \alpha(X_i) \text{ für Parsewaldregeln}$$

$$\alpha(A) = \sum_{A \to \gamma} \alpha(A \to \gamma)$$

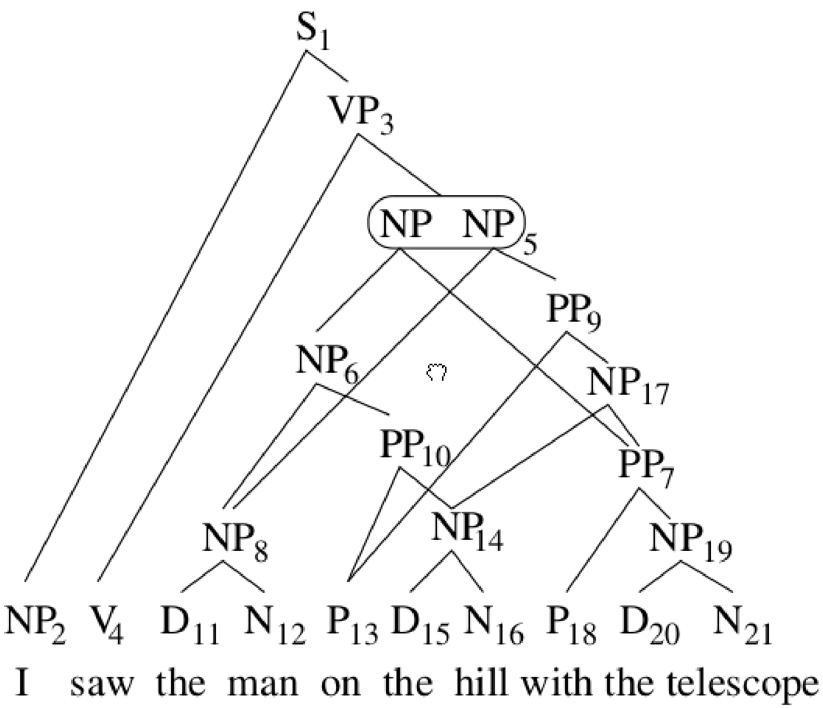
Was ist A ?

Was berechnet Inside-Algorithmus?

Inside-Algorithmus

$$\begin{aligned}
\alpha(a) &= 1 \quad \text{für Terminalsymbol a} \\
\alpha(A \rightarrow X_1 ... X_n) &= p(A \rightarrow X_1 ... X_n) \prod_{i=1}^{n} \alpha(X_i) \text{ für Parsewaldregeln} \\
\alpha(A) &= \sum_{A \rightarrow \gamma} \alpha(A \rightarrow \gamma) \quad \text{für Nichtterminale A}
\end{aligned}$$

- Der **Viterbi**-Algorithmus berechnet die Wahrscheinlichkeit der **besten Analyse** jedes Parsewaldknotens
- Der **Inside**-Algorithmus berechnet die Gesamtwahrscheinlichkeit **aller Analysen** für jeden Parsewaldknoten.



Wir haben hier alle Analyse dieses Satzes in einem Baum (Parswald).

Berechne inside-wk folgender Knoten (Symbole) und Regeln.

Inside-Algorithmus

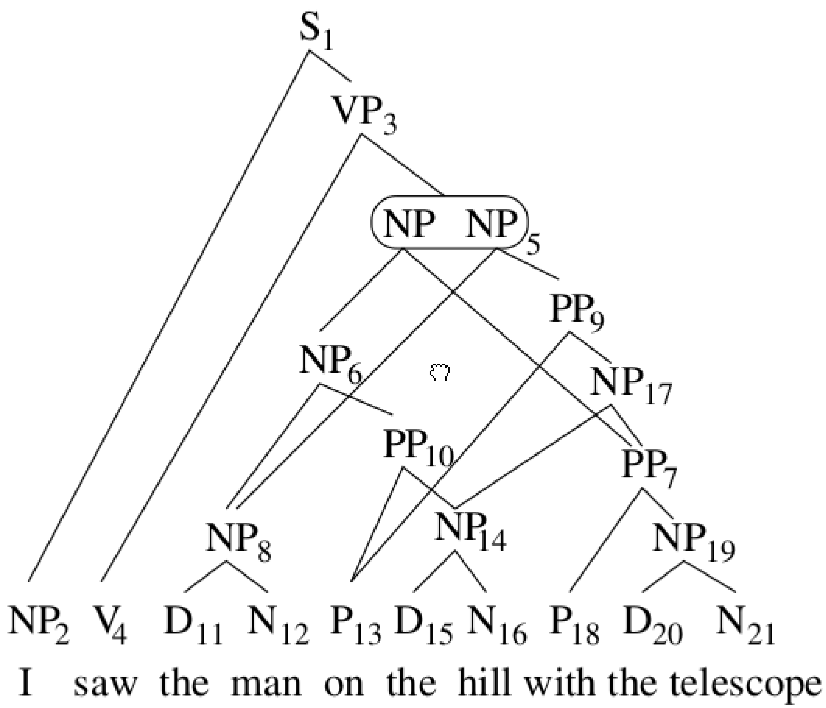
$$\alpha(a) = 1 \quad \text{für Terminalsymbol a}$$

$$\alpha(A \to X_1 ... X_n) = p(A \to X_1 ... X_n) \prod_{i=1}^{n} \alpha(X_i) \text{ für Parsewaldregeln}$$

$$\alpha(A) = \sum_{A \to \gamma} \alpha(A \to \gamma) \quad \text{für Nichtterminale A}$$

α(I) =
α(saw) =
α(NP2) =
α(NP8) =
α(NP5) =
α(S1) =

α(VP3 → V4 NP5) = ...
α(NP5 → NP8 PP9) = ...



Wir haben hier alle Analyse dieses Satzes in einem Baum (Parswald).

Berechne inside-wk folgender Knoten (Symbole) und Regeln.

α(I) = 1
α(saw) = 1
α(NP2) = α(NP2 → I)
= p(NP → I) α(I)

α(NP8) = α(NP8 → D11 N12)
= p(NP → D N) α (D11) α(N12)
α(NP5) =α(NP5->NP6 PP7)+α(NP5 → NP8 PP9)
= ...
α(S1) = α(S1 →NP2 VP3)
= ..
α(VP3 → V4 NP5) = p(VP → V NP) α(V4) α(NP5)
α(NP5 → NP8 PP9) = ….

Inside-Algorithmus

$$\alpha(a) = 1 \quad \text{für Terminalsymbol a}$$

$$\alpha(A \to X_1 \dots X_n) = p(A \to X_1 \dots X_n) \prod_{i=1}^{n} \alpha(X_i) \text{ für Parsewaldregeln}$$

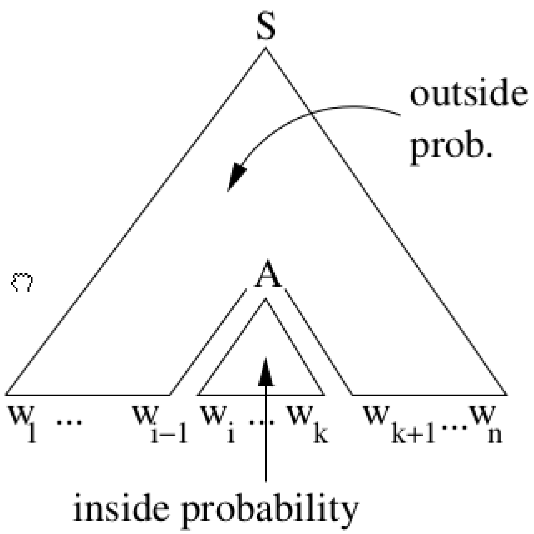
$$\alpha(A) = \sum_{A \to \gamma} \alpha(A \to \gamma) \quad \text{für Nichtterminale A}$$

Inside-Outside-Algorithmus

- Das Produkt der Inside- und Outside-Wahrscheinlichkeiten eines Parsewaldknotens ergibt die **Gesamtwahrscheinlichkeit aller Parsebäume mit diesem Knoten**.
- Durch Division dieses Produktes mit der Wahrscheinlichkeit aller Analysen erhält man die **erwartete Häufigkeit** des Parsewaldknotens.

  $\gamma(A) = \alpha(A)\beta(A)/\alpha(S)$
- erwartete Regelhäufigkeit

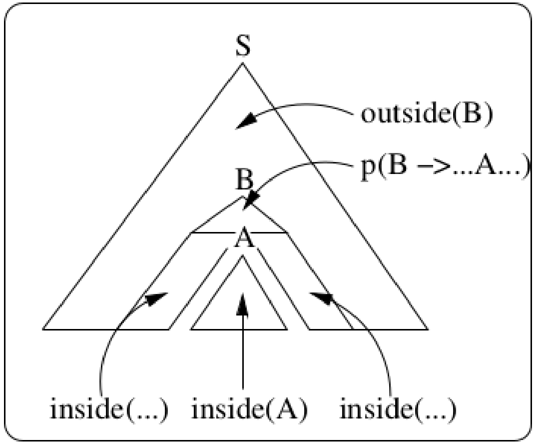
  $\gamma(A \to \delta) = \alpha(A \to \delta)\ \beta(A)/\alpha(S)$



Berechnung der Outside-Wahrscheinlichkeiten

Outside-Algorithmus

$$\beta(S) = 1 \quad \text{für Startsymbol S}$$
$$\beta(A) = \sum_{B \rightarrow \gamma A \delta} \beta(B \rightarrow \gamma \underline{A} \delta)$$



$$\beta(B \rightarrow X_1 ... X_m \underline{A} X_{m+1} ... X_n) = \beta(B) p(B \rightarrow X_1 ... X_m A X_{m+1} ... X_n) \prod_{i=1}^{n} \alpha(X_i)$$

$$\beta(B \rightarrow \gamma \underline{A} \delta) = \beta(B) \frac{\alpha(B \rightarrow \gamma A \delta)}{\alpha(A)}$$

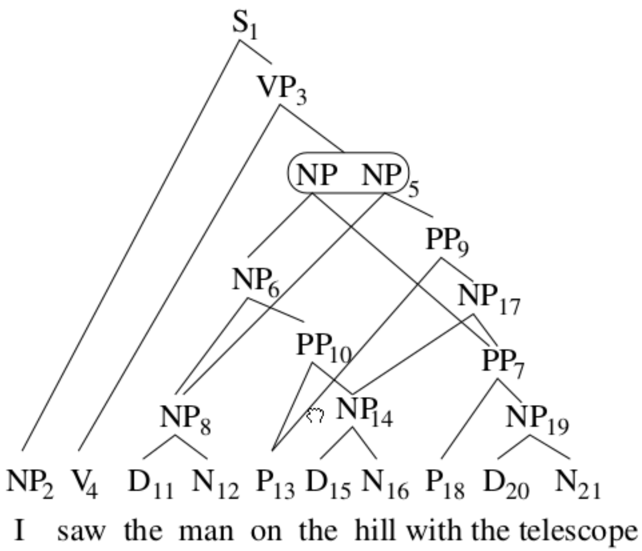
$\beta(B \rightarrow \gamma \underline{A} \delta)$: Beitrag der Regel $B \rightarrow \gamma A \delta$ zur Outside-Wahrscheinlichkeit von A

Beachte, dass $p(B \rightarrow X_1 ... X_m A X_{m+1} ... X_n) \prod_{i=1}^{n} \alpha(X_i) = \alpha(B \rightarrow X_1 ... X_m A X_{m+1} ... X_n)/\alpha(A)$

$$\beta(S) = 1 \quad \text{für Startsymbol S}$$
$$\beta(A) = \sum_{B \to \gamma A \delta} \beta(B \to \gamma \underline{A} \delta)$$

$$\beta(B \to X_1 \ldots X_m \underline{A} X_{m+1} \ldots X_n) = \beta(B) p(B \to X_1 \ldots X_m A X_{m+1} \ldots X_n) \prod_{i=1}^{n} \alpha(X_i)$$

$$\beta(B \to \gamma \underline{A} \delta) = \beta(B) \frac{\alpha(B \to \gamma A \delta)}{\alpha(A)}$$



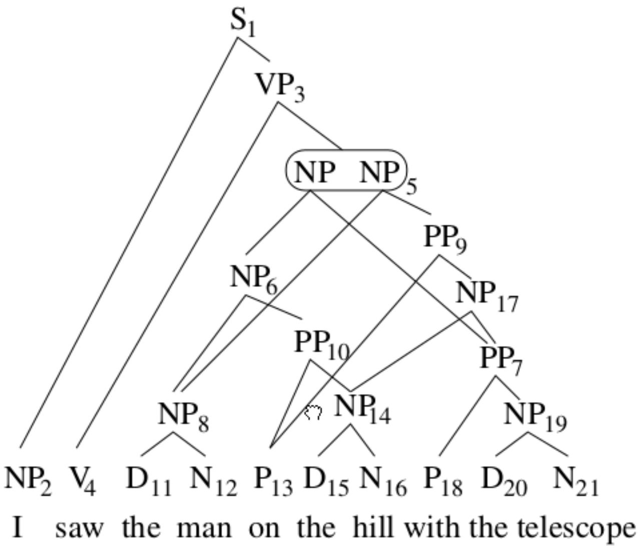
β(S) =
β(NP2) =
β(VP3) =
β(with) =

$$\beta(S) = 1 \quad \text{für Startsymbol S}$$

$$\beta(A) = \sum_{B \to \gamma A \delta} \beta(B \to \gamma \underline{A} \delta)$$

$$\beta(B \to X_1 \ldots X_m \underline{A} X_{m+1} \ldots X_n) = \beta(B) p(B \to X_1 \ldots X_m A X_{m+1} \ldots X_n) \prod_{i=1}^{n} \alpha(X_i)$$

$$\beta(B \to \gamma \underline{A} \delta) = \beta(B) \frac{\alpha(B \to \gamma A \delta)}{\alpha(A)}$$



β(S1) = 1

β(NP2) = β(S1 → NP2 VP3)

= β(S1) p(S → **NP** VP) α(VP)

β(VP3) = β(S1 → NP3 VP3)

= β(S1) p(S → NP **VP**) α(NP) --- (1. Formel)

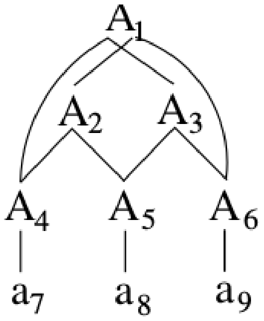
= β(S1) α(S1 → NP3 **VP3**) / α(VP3) --- (2. Formel)

β(with) = β(P18 → with)

= β(P18) p(P18 → with) α(with) --- α(terminal) = 1

Inside-Outside-Beispiel

PCFG: $\begin{array}{ll} A \rightarrow A\ A & 0.6 \\ A \rightarrow a & 0.4 \end{array}$



$$
\begin{aligned}
\alpha(a_7) &= \mathbf{1} \quad = \alpha(a_8) = \alpha(a_9) \\
\alpha(A_4) &= P(A \rightarrow a)\ \alpha(a_7) = 0.4 * 1 = \mathbf{0.4}\ = \alpha(A_5) = \alpha(A_6) \\
\alpha(A_2) &= p(A \rightarrow A\ A)\ \alpha(A_4)\ \alpha(A_5) = 0.6 * 0.4 * 0.4 = \mathbf{0.096} \\
\alpha(A_3) &= \alpha(A_2) \\
\alpha(A_1) &= p(A \rightarrow A\ A)\ \alpha(A_4)\ \alpha(A_3) + p(A \rightarrow A\ A)\ \alpha(A_2)\ \alpha(A_6) \\
&= 0.6 * 0.4 * 0.096 * 2 = \mathbf{0.04608} \\
\beta(A_1) &= 1 \\
\beta(A_2) &= \beta(A_1)\ p(A \rightarrow A\ A)\ \alpha(A_6) = 1 * 0.6 * 0.4 = \mathbf{0.24} = \beta(A_3) \\
\beta(A_4) &= \beta(A_1)\ p(A \rightarrow A\ A)\ \alpha(A_3) + \beta(A_2)\ p(A \rightarrow A\ A)\ \alpha(A_5) \\
&= 1 * 0.6 * 0.096 + 0.24 * 0.6 * 0.4 = \mathbf{0.1152} = \beta(A_6) \\
\beta(A_5) &= \beta(A_2)\ p(A \rightarrow A\ A)\ \alpha(A_4) + \beta(A_3)\ p(A \rightarrow A\ A)\ \alpha(A_6) \\
&= 0.24 * 0.6 * 0.4 * 2 = \mathbf{0.1152} \\
\beta(a_7) &= \beta(A_4)\ p(A \rightarrow a) = 0.1152 * 0.4 = \mathbf{0.04608}
\end{aligned}
$$

$$
\begin{aligned}
\gamma(A_2 \rightarrow A_4\ A_5) &= \beta(A_2) p(A \rightarrow A\ A)\ \alpha(A_4)\ \alpha(A_5) / \alpha(A_1) \\
&= 0.24 * 0.6 * 0.4 * 0.4 / 0.04608 = 0.5
\end{aligned}
$$