Produktkonstruktionen für das Maschinelle Übersetzen

Matthias Büchse
TECHNISCHE UNIVERSITÄT DRESDEN

Computerlinguistisches Kolloquium
Universität Potsdam
2011-12-19

Gliederung

Statistisches Maschinelles Übersetzen

Produktkonstruktion

Gliederung

Statistisches Maschinelles Übersetzen

Produktkonstruktion

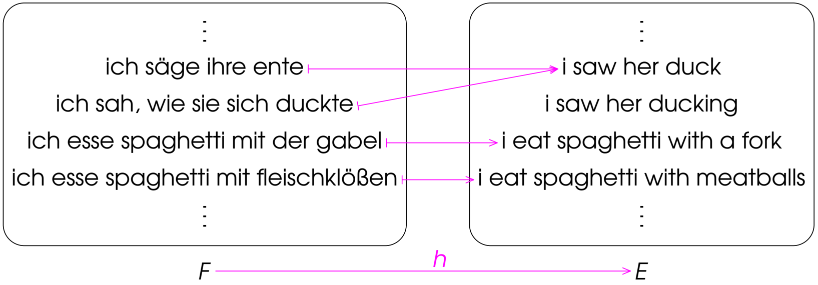
Übersetzen lernen

Liang et al. (2006); Collins (2005)

F	E
⋮	⋮
ich säge ihre ente	i saw her duck
ich sah, wie sie sich duckte	i saw her ducking
ich esse spaghetti mit der gabel	i eat spaghetti with a fork
ich esse spaghetti mit fleischklößen	i eat spaghetti with meatballs
⋮	⋮

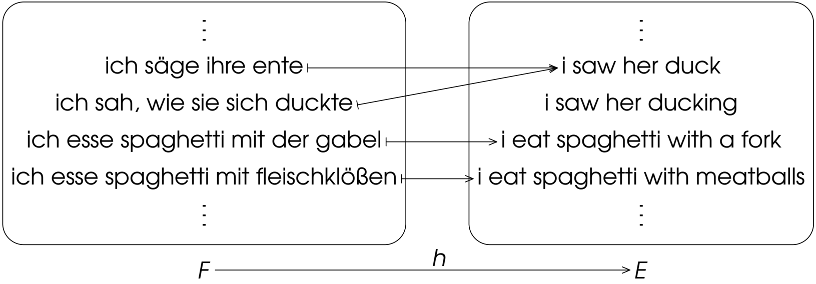
Übersetzen lernen

Liang et al. (2006); Collins (2005)

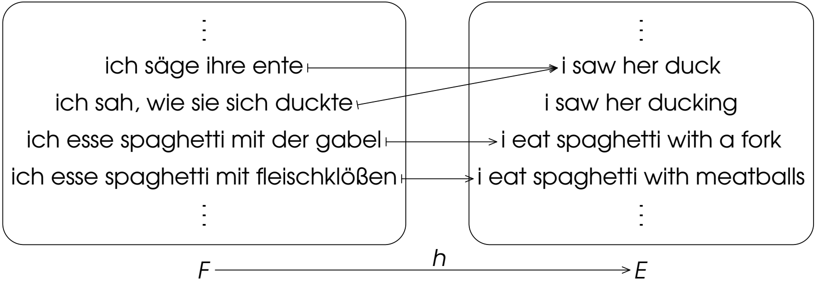


Übersetzen lernen

Liang et al. (2006); Collins (2005)



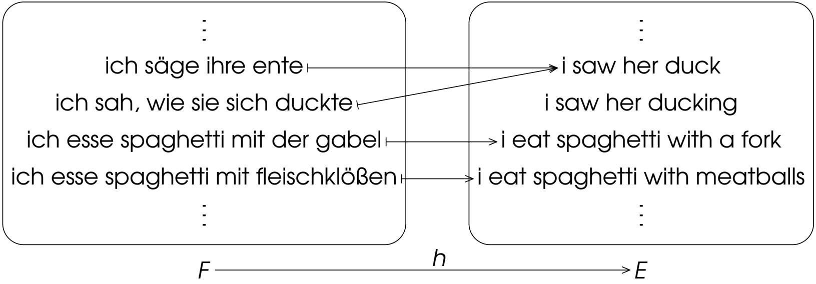
Übersetzen lernen Liang et al. (2006); Collins (2005)



gegeben Hypothesenraum $\mathcal{H} \subseteq E^F$

Übersetzen lernen

Liang et al. (2006); Collins (2005)



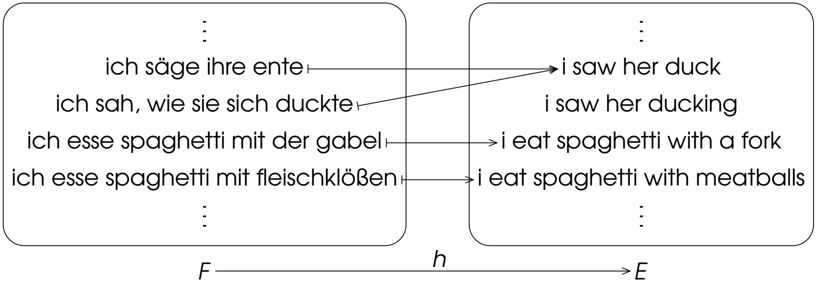
gegeben Hypothesenraum $\mathcal{H} \subseteq E^F$

Trainingsdaten

f_1	ich sah, wie sie sich duckte	i saw her duck	$e_{1,1}$
		i saw her ducking	$e_{1,2}$
f_2	ich esse spaghetti mit der gabel	i eat spaghetti with a fork	$e_{2,1}$
		i eat spaghetti with the fork	$e_{2,2}$

Übersetzen lernen

Liang et al. (2006); Collins (2005)



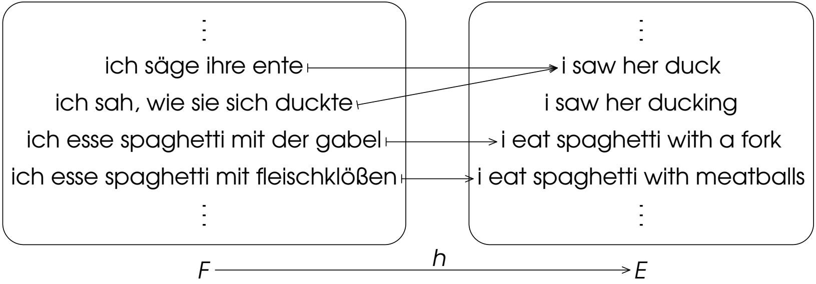
gegeben Hypothesenraum $\mathcal{H} \subseteq E^F$

Trainingsdaten

f_1	ich sah, wie sie sich duckte	i saw her duck	$e_{1,1}$
		i saw her ducking	$e_{1,2}$
f_2	ich esse spaghetti mit der gabel	i eat spaghetti with a fork	$e_{2,1}$
		i eat spaghetti with the fork	$e_{2,2}$

Fehlerfunktion $L\colon E \times E^* \to \mathbb{R}$

Übersetzen lernen   Liang et al. (2006); Collins (2005)



gegeben Hypothesenraum $\mathcal{H} \subseteq E^F$

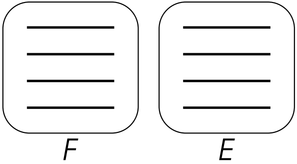
Trainingsdaten

f_1	ich sah, wie sie sich duckte	i saw her duck	$e_{1,1}$
		i saw her ducking	$e_{1,2}$
f_2	ich esse spaghetti mit der gabel	i eat spaghetti with a fork	$e_{2,1}$
		i eat spaghetti with the fork	$e_{2,2}$

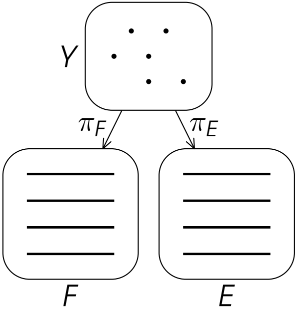
Fehlerfunktion $L\colon E \times E^* \to \mathbb{R}$

finde $\hat{h} = \operatorname{argmin}_{h\in\mathcal{H}} \sum_i L(h(f_i), e_{i,1} e_{i,2} \ldots)$

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$

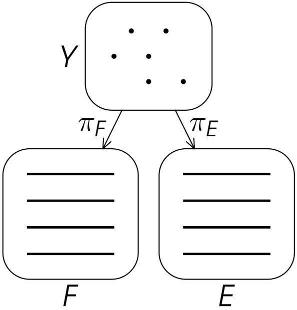


Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



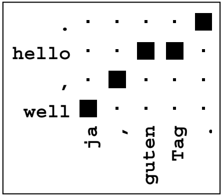
$Y \ldots$ Menge der Korrespondenzstrukturen

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



z. B.

Y = Alignments

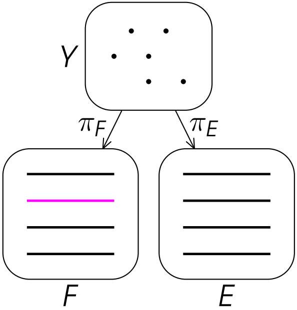


$\pi_F(y)$ = ja, guten Tag.

$\pi_E(y)$ = well, hello.

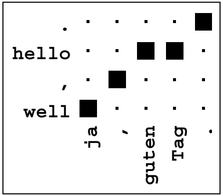
Y … Menge der Korrespondenzstrukturen

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



z. B.

Y = Alignments

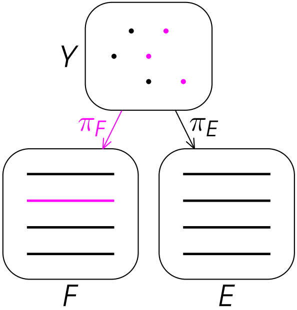


$\pi_F(y)$ = ja, guten Tag.

$\pi_E(y)$ = well, hello.

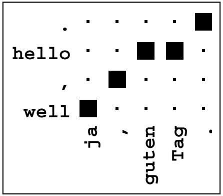
$Y \ldots$ Menge der Korrespondenzstrukturen

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



z. B.

Y = Alignments

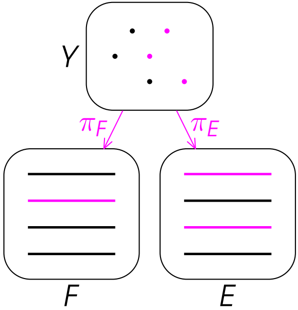


$\pi_F(y)$ = ja, guten Tag.

$\pi_E(y)$ = well, hello.

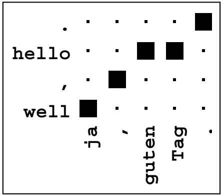
Y … Menge der Korrespondenzstrukturen

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



z. B.

Y = Alignments

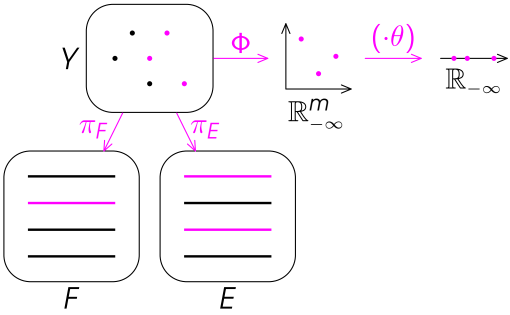


$\pi_F(y)$ = ja, guten Tag.

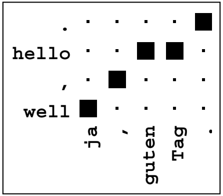
$\pi_E(y)$ = well, hello.

Y … Menge der Korrespondenzstrukturen

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



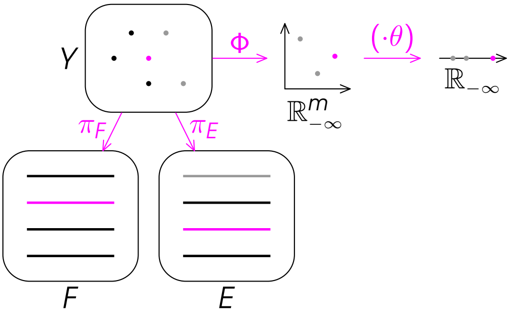
z. B.
Y = Alignments



$\pi_F(y)$ = ja, guten Tag.
$\pi_E(y)$ = well, hello.

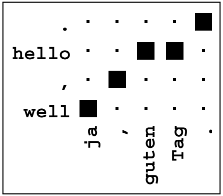
Y … Menge der Korrespondenzstrukturen
Φ … Repräsentationsfunktion $\Phi\colon Y \to \mathbb{R}^m_{-\infty}$
$\Phi(y)$ … Merkmalsvektor θ … Gewichtsvektor

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



z. B.

Y = Alignments



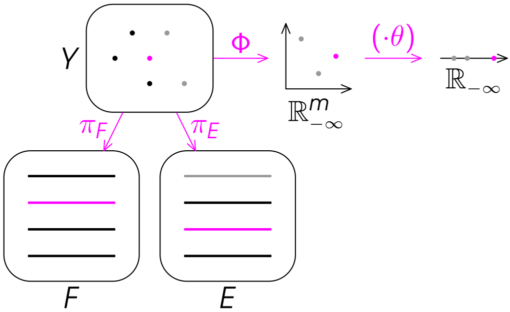
$\pi_F(y)$ = ja, guten Tag.
$\pi_E(y)$ = well, hello.

Y … Menge der Korrespondenzstrukturen

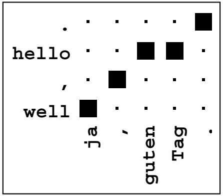
Φ … Repräsentationsfunktion $\Phi\colon Y \to \mathbb{R}^m_{-\infty}$

$\Phi(y)$ … Merkmalsvektor $\qquad \theta$ … Gewichtsvektor

Konstruieren eines Hypothesenraums $\mathcal{H} \subseteq E^F$



z. B.
Y = Alignments



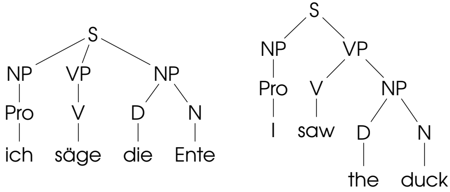
$\pi_F(y)$ = ja, guten Tag.
$\pi_E(y)$ = well, hello.

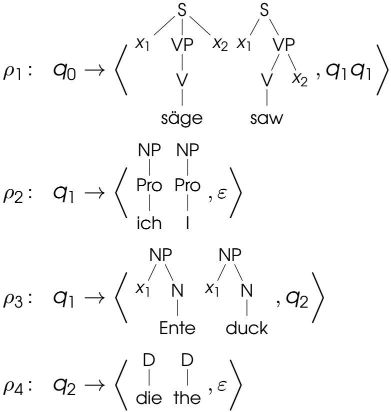
Y ... Menge der Korrespondenzstrukturen
Φ ... Repräsentationsfunktion $\Phi\colon Y \to \mathbb{R}^m_{-\infty}$
$\Phi(y)$ ... Merkmalsvektor θ ... Gewichtsvektor

$$\mathcal{H} = \{h_\theta \mid \theta \in \mathbb{R}^m\}$$

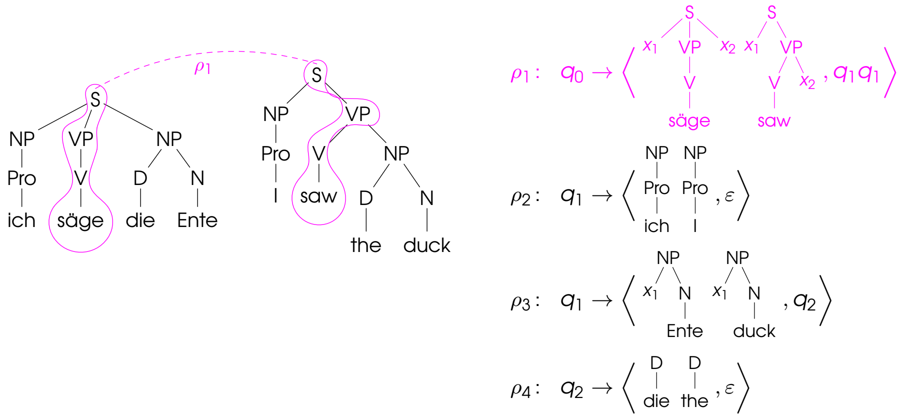
$$h_\theta\colon F \to E\colon f \mapsto \pi_E\Big(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y) = f}} \Phi(y) \cdot \theta\Big)$$

Korrespondenzstrukturen durch Grammatiken

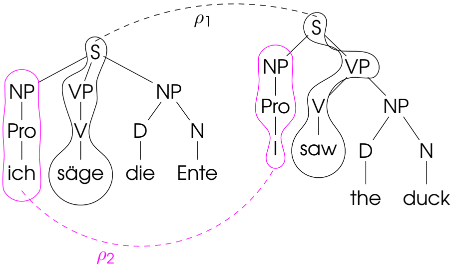


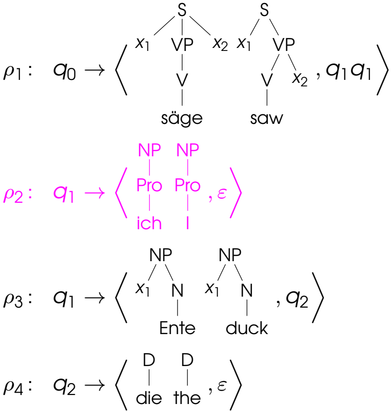


Korrespondenzstrukturen durch Grammatiken

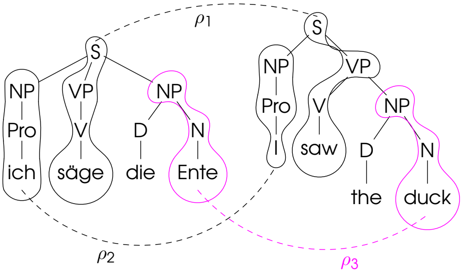


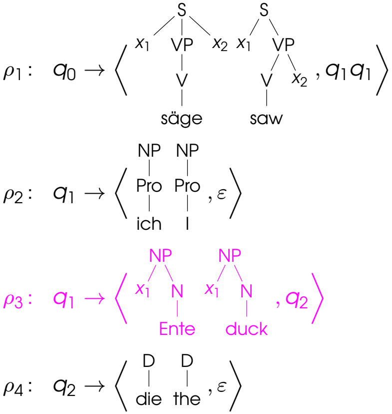
Korrespondenzstrukturen durch Grammatiken



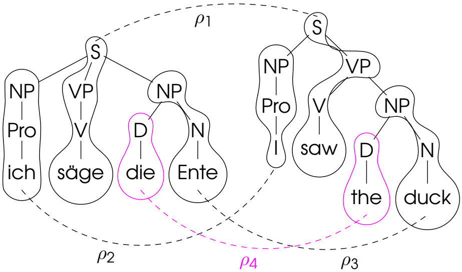


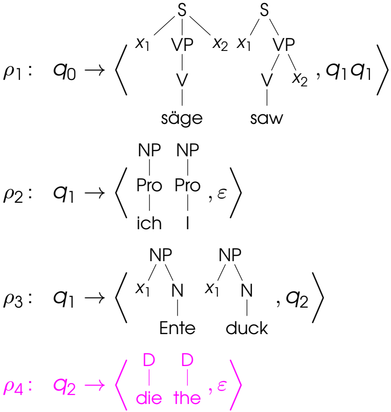
Korrespondenzstrukturen durch Grammatiken



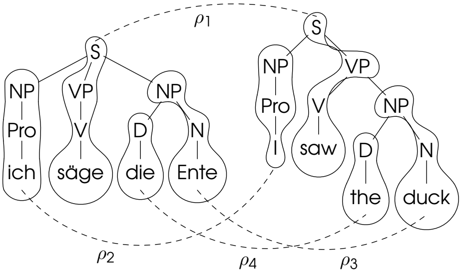


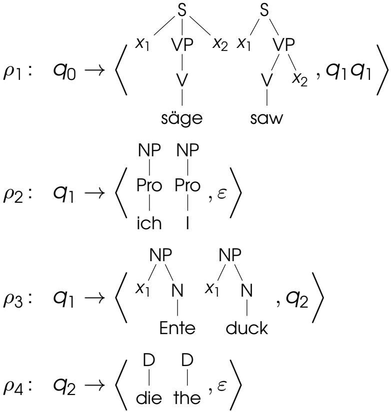
Korrespondenzstrukturen durch Grammatiken

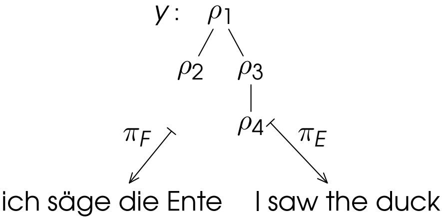




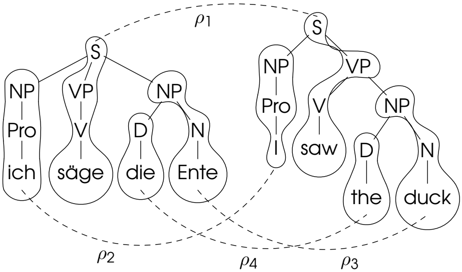
Korrespondenzstrukturen durch Grammatiken

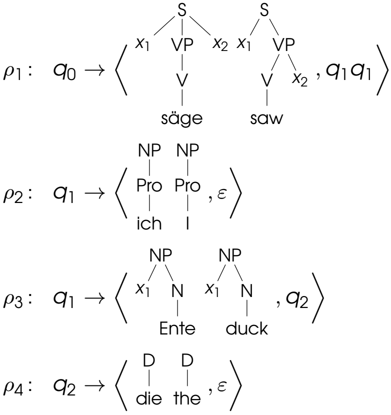


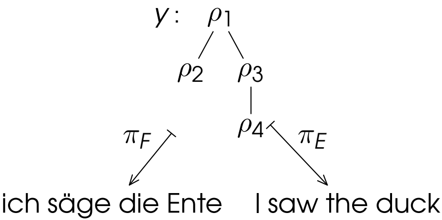




Korrespondenzstrukturen durch Grammatiken







$$q_0 \to \rho_1(q_1, q_1)$$
$$q_1 \to \rho_2()$$
$$q_1 \to \rho_3(q_2)$$
$$q_2 \to \rho_4()$$

Repräsentationsfunktion – Merkmale

$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix} \quad \begin{matrix} \text{Regelvorkommen} \\ \text{Länge} \\ \text{Sprachmodell} \\ \text{Übersetzungsmodell} \end{matrix}$$

Repräsentationsfunktion – Merkmale

$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix} \quad \begin{matrix} \text{Regelvorkommen} \\ \text{Länge} \\ \text{Sprachmodell} \\ \text{Übersetzungsmodell} \end{matrix}$$

Repräsentationsfunktion – Merkmale

$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen
Länge
Sprachmodell
Übersetzungsmodell

Repräsentationsfunktion – Merkmale

$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

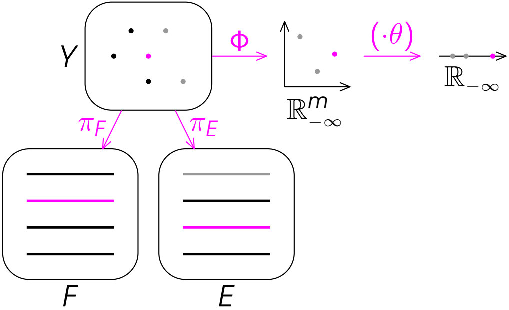
Regelvorkommen
Länge
Sprachmodell
Übersetzungsmodell

Repräsentationsfunktion – Merkmale

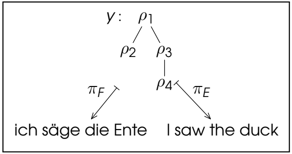
$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix} \quad \begin{matrix} \\ \text{Regelvorkommen} \\ \text{Länge} \\ \text{Sprachmodell} \\ \text{Übersetzungsmodell} \\ \\ \end{matrix}$$

Suche



z. B. $Y =$ Ableitungsbäume



Y ... Menge der Korrespondenzstrukturen

Φ ... Repräsentationsfunktion $\Phi\colon Y \to \mathbb{R}^m_{-\infty}$

$\Phi(y)$ ... Merkmalsvektor $\qquad \theta$ ... Gewichtsvektor

$$\mathcal{H} = \{h_\theta \mid \theta \in \mathbb{R}^m\}$$

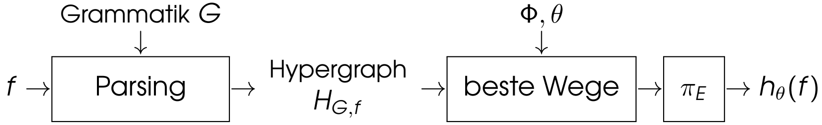
$$h_\theta\colon F \to E\colon f \mapsto \pi_E\Big(\operatorname{argmax}_{\substack{y \in Y\colon \\ \pi_F(y) = f}} \Phi(y) \cdot \theta\Big)$$

Suche

$$h_\theta \colon f \mapsto \pi_E\Big(\operatorname{argmax}_{\substack{y \in Y: \\ \pi_F(y) = f}} \Phi(y) \cdot \theta\Big)$$

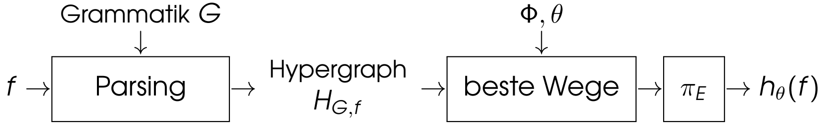
Suche

$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y)=f}} \Phi(y) \cdot \theta\right)$$



Suche

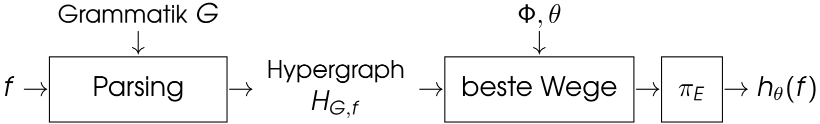
$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y)=f}} \Phi(y) \cdot \theta\right)$$



Existenz keine positiven Zyklen

Suche

$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y)=f}} \Phi(y)\cdot\theta\right)$$

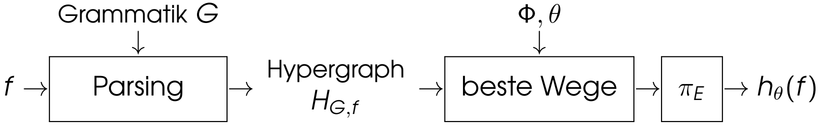


Existenz keine positiven Zyklen

Effizienz Monotonie (Bellmann-Eigenschaft)
$\rightsquigarrow$ dyn. Programmierung

Suche

$$h_\theta\colon f \mapsto \pi_E\Big(\operatorname{argmax}_{\substack{y\in Y:\\ \pi_F(y)=f}} \Phi(y)\cdot\theta\Big)$$



Existenz keine positiven Zyklen

Effizienz Monotonie (Bellmann-Eigenschaft)
$\rightsquigarrow$ dyn. Programmierung

$$O(|E| + n\cdot|V|\cdot\log|V|)$$

Monotonie von Merkmalen

$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\mathrm{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen
Länge
Sprachmodell
Übersetzungsmodell

Monotonie von Merkmalen

$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\mathrm{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix} \quad \begin{matrix} \text{Regelvorkommen} \\ \text{Länge} \\ \text{Sprachmodell} \\ \text{Übersetzungsmodell} \end{matrix}$$

with meatballs	$<$	with a fork
i eat with meatballs	$<$	i eat with a fork

Monotonie von Merkmalen

$\Phi \colon Y \to \mathbb{R}^m_{-\infty}$

$$\Phi \colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen
Länge
Sprachmodell
Übersetzungsmodell

monoton

with meatballs	$<$	with a fork
i eat with meatballs	$<$	i eat with a fork

Monotonie von Merkmalen

$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen
Länge
Sprachmodell
Übersetzungsmodell

monoton

Monotonie von Merkmalen

$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\mathrm{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen
Länge — monoton
Sprachmodell
Übersetzungsmodell

York	$<$	Berlin
New York	$>$	New Berlin

Monotonie von Merkmalen

$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\mathrm{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen	
Länge	monoton
Sprachmodell	nicht monoton
Übersetzungsmodell	

York	$<$	Berlin
New York	$>$	New Berlin

Monotonie von Merkmalen

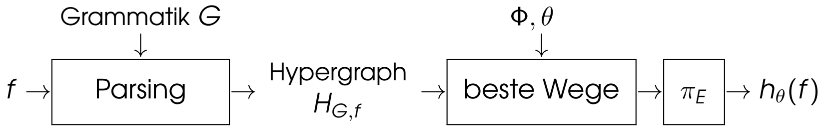
$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\mathrm{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen	monoton
Länge	monoton
Sprachmodell	nicht monoton
Übersetzungsmodell	monoton(?)

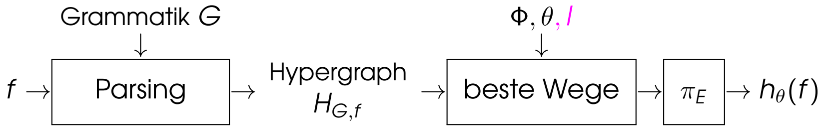
Approximation bei nichtmonotonen Merkmalen

$$h_\theta : f \mapsto \pi_E\Big(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y) = f}} \Phi(y) \cdot \theta\Big)$$



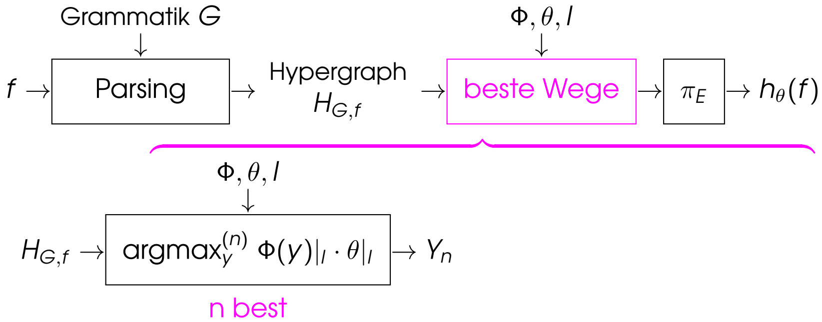
Approximation bei nichtmonotonen Merkmalen

$$h_\theta\colon f \mapsto \pi_E\Big(\operatorname{argmax}_{\substack{y\in Y:\\ \pi_F(y)=f}} \Phi(y)\cdot\theta\Big)$$



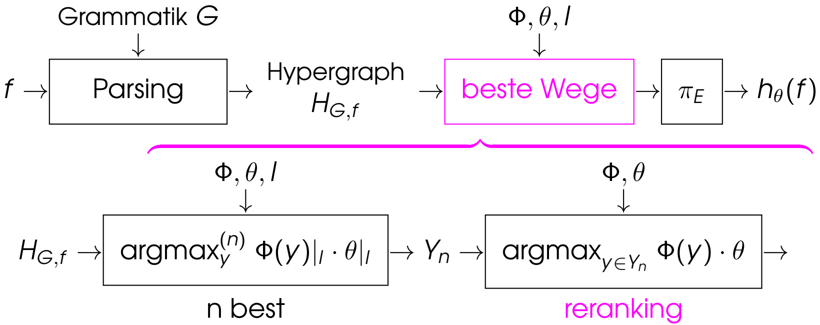
Approximation bei nichtmonotonen Merkmalen

$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y)=f}} \Phi(y)\cdot\theta\right)$$



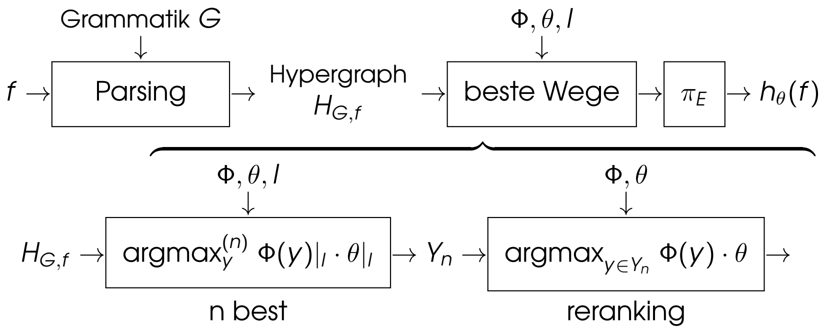
Approximation bei nichtmonotonen Merkmalen

$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y)=f}} \Phi(y) \cdot \theta\right)$$



Approximation bei nichtmonotonen Merkmalen

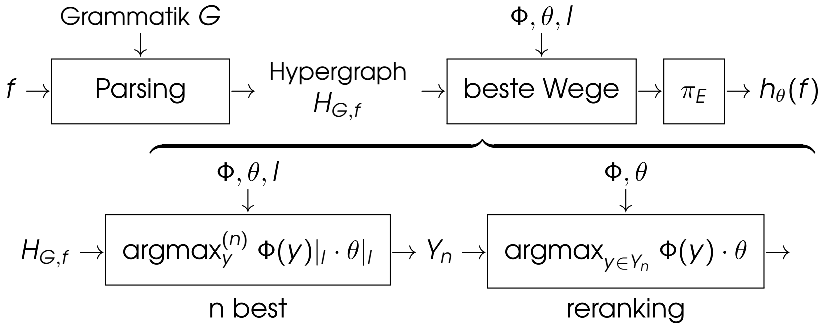
$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y)=f}} \Phi(y) \cdot \theta\right)$$



Problem: Approximation z. T. schlecht

Approximation bei nichtmonotonen Merkmalen

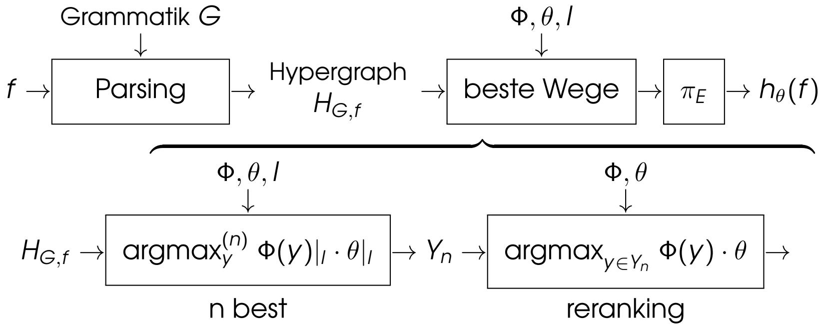
$$h_\theta : f \mapsto \pi_E\Big(\operatorname{argmax}_{\substack{y \in Y:\\ \pi_F(y)=f}} \Phi(y)\cdot\theta\Big)$$



Problem: Approximation z. T. schlecht
Alternative: Struktur von $H_{G,f}$ verändern

Approximation bei nichtmonotonen Merkmalen

$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{\substack{y \in Y : \\ \pi_F(y) = f}} \Phi(y) \cdot \theta\right)$$



Problem: Approximation z. T. schlecht
Alternative: Struktur von $H_{G,f}$ verändern
Parsing und Produktkonstruktionen!

Gliederung

Reguläre Baumgrammatiken

$q_0 \rightarrow \mathrm{NP}(q_1\ q_2)$

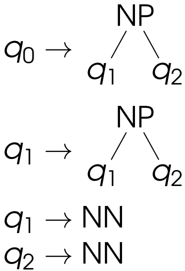
$q_1 \rightarrow \mathrm{NP}(q_1\ q_2)$

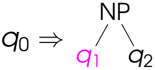
$q_1 \rightarrow \mathrm{NN}$

$q_2 \rightarrow \mathrm{NN}$

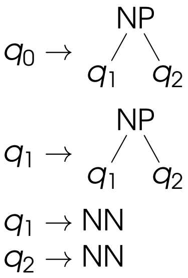
q_0

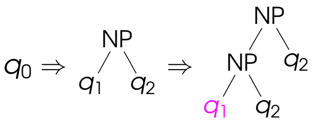
Reguläre Baumgrammatiken



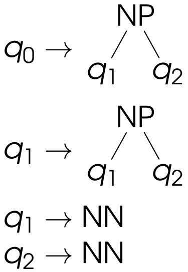


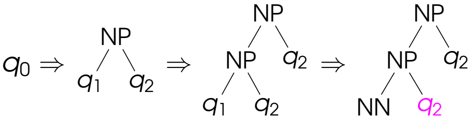
Reguläre Baumgrammatiken



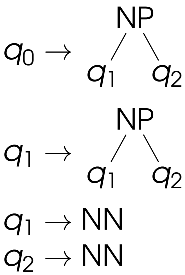


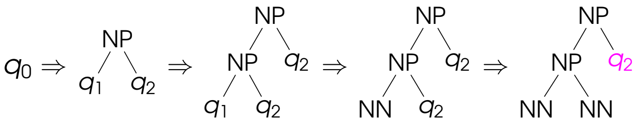
Reguläre Baumgrammatiken



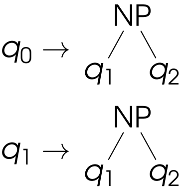


Reguläre Baumgrammatiken

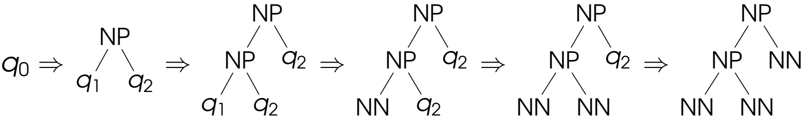




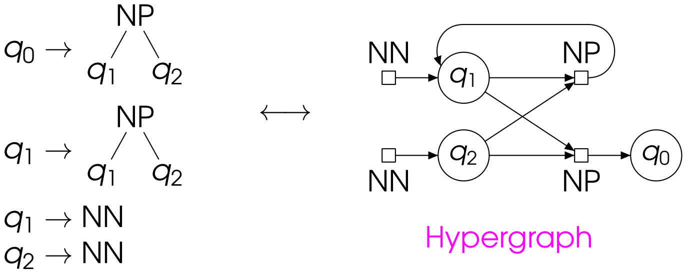
Reguläre Baumgrammatiken

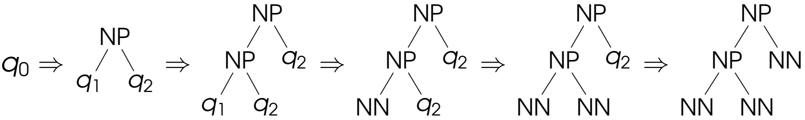


$q_1 \rightarrow \text{NN}$
$q_2 \rightarrow \text{NN}$

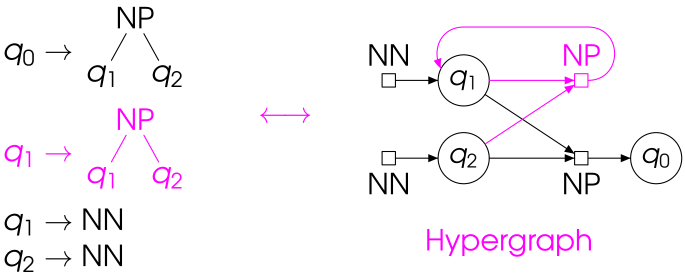


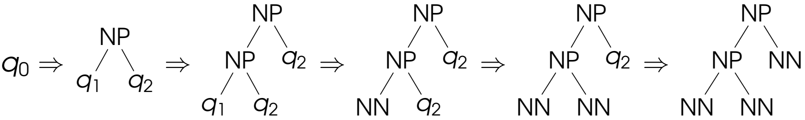
Reguläre Baumgrammatiken





Reguläre Baumgrammatiken



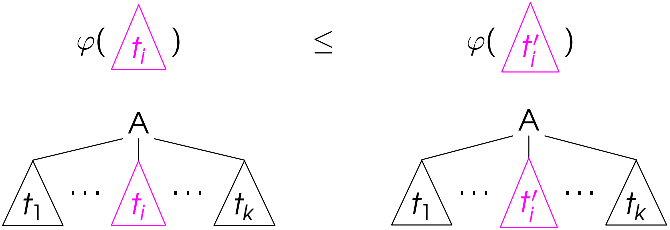


Monotonie

Funktion $\varphi\colon T \to \mathbb{R}_{-\infty}$ monoton, wenn

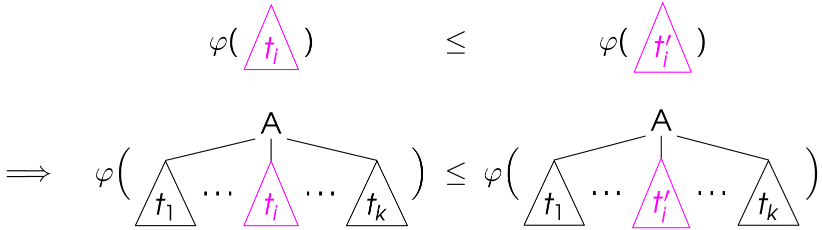
Monotonie

Funktion $\varphi\colon T \to \mathbb{R}_{-\infty}$ monoton, wenn



Monotonie

Funktion $\varphi\colon T \to \mathbb{R}_{-\infty}$ monoton, wenn



Monotonie von Merkmalen

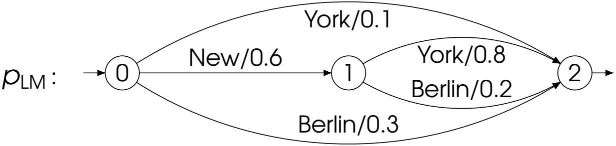
$$\Phi\colon Y \to \mathbb{R}^m_{-\infty}$$

$$\Phi\colon y \mapsto \begin{pmatrix} \vdots \\ \#(\rho, y) \\ |\pi_E(y)| \\ \log p_{\text{LM}}(\pi_E(y)) \\ \log p(y) \\ \vdots \end{pmatrix}$$

Regelvorkommen	
Länge	monoton
Sprachmodell	nicht monoton
Übersetzungsmodell	

York	$<$	Berlin
New York	$>$	New Berlin

Konkret



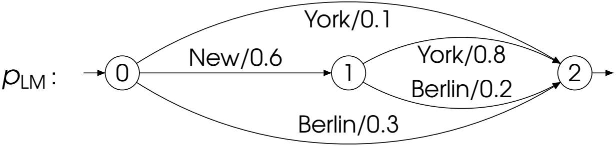
ρ_1 : $S \to \text{New } T$

ρ_2 : $S \to T$

ρ_3 : $T \to \text{York}$

ρ_4 : $T \to \text{Berlin}$

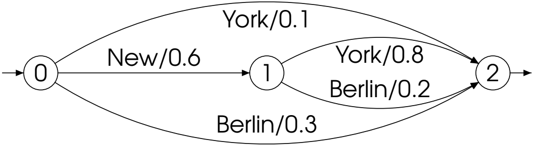
Konkret



ρ_1 :	$S \to \text{New } T$	$S \to \rho_1(T)$
ρ_2 :	$S \to T$	$S \to \rho_2(T)$
ρ_3 :	$T \to \text{York}$	$T \to \rho_3()$
ρ_4 :	$T \to \text{Berlin}$	$T \to \rho_4()$

Konkret

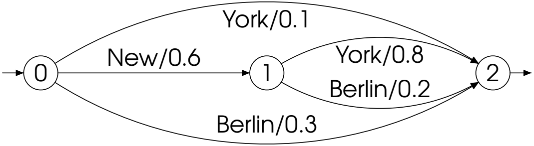
p_{LM} :



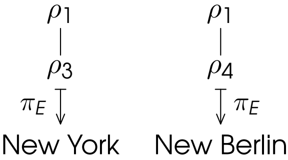
ρ_1 :	$S \to$ New T		$S \to \rho_1(T)$
ρ_2 :	$S \to T$		$S \to \rho_2(T)$
ρ_3 :	$T \to$ York		$T \to \rho_3()$
ρ_4 :	$T \to$ Berlin		$T \to \rho_4()$

Konkret

p_{LM} :

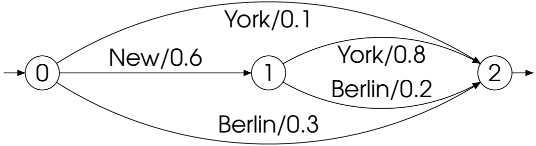


ρ_1 :	$S \rightarrow \text{New } T$	$S \rightarrow \rho_1(T)$
ρ_2 :	$S \rightarrow T$	$S \rightarrow \rho_2(T)$
ρ_3 :	$T \rightarrow \text{York}$	$T \rightarrow \rho_3()$
ρ_4 :	$T \rightarrow \text{Berlin}$	$T \rightarrow \rho_4()$

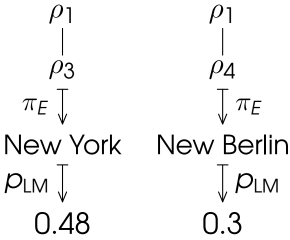


Konkret

p_{LM} :

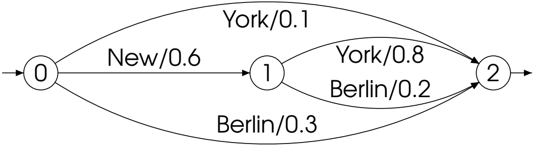


ρ_1 :	$S \to \text{New}\ T$		$S \to \rho_1(T)$
ρ_2 :	$S \to T$		$S \to \rho_2(T)$
ρ_3 :	$T \to \text{York}$		$T \to \rho_3()$
ρ_4 :	$T \to \text{Berlin}$		$T \to \rho_4()$



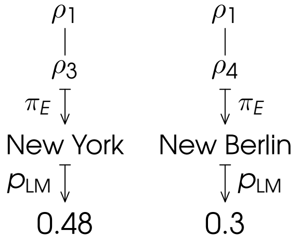
Konkret

p_{LM} :

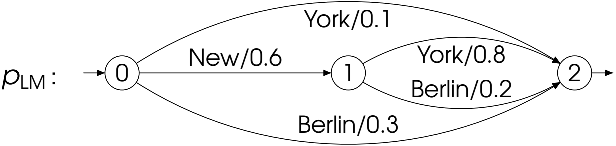


ρ_1 :	$S \to \text{New } T$	$S \to \rho_1(T)$
ρ_2 :	$S \to T$	$S \to \rho_2(T)$
ρ_3 :	$T \to \text{York}$	$T \to \rho_3()$
ρ_4 :	$T \to \text{Berlin}$	$T \to \rho_4()$

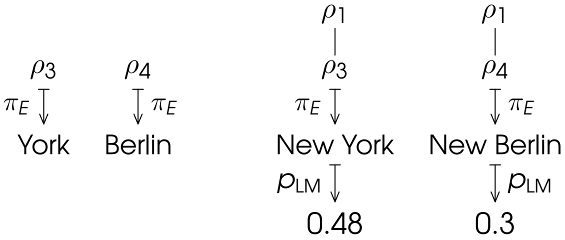
ρ_3 ρ_4



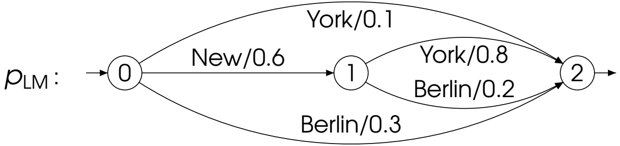
Konkret



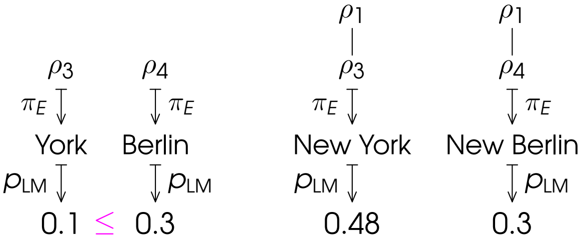
ρ_1:	$S \to \text{New } T$	$S \to \rho_1(T)$
ρ_2:	$S \to T$	$S \to \rho_2(T)$
ρ_3:	$T \to \text{York}$	$T \to \rho_3()$
ρ_4:	$T \to \text{Berlin}$	$T \to \rho_4()$



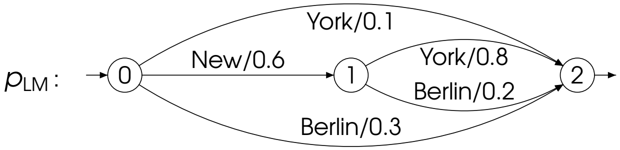
Konkret



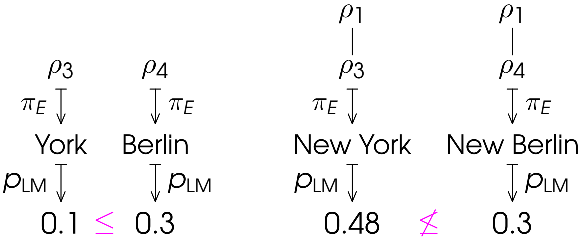
ρ_1 :	$S \to \text{New } T$	$S \to \rho_1(T)$
ρ_2 :	$S \to T$	$S \to \rho_2(T)$
ρ_3 :	$T \to \text{York}$	$T \to \rho_3()$
ρ_4 :	$T \to \text{Berlin}$	$T \to \rho_4()$



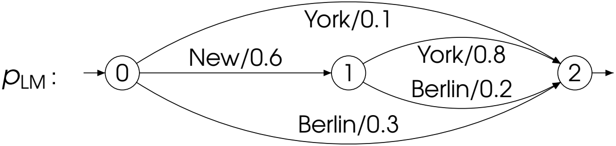
Konkret



$$
\begin{array}{llll}
\rho_1: & S \to \text{New } T & \qquad & S \to \rho_1(T) \\
\rho_2: & S \to T & & S \to \rho_2(T) \\
\rho_3: & T \to \text{York} & & T \to \rho_3() \\
\rho_4: & T \to \text{Berlin} & & T \to \rho_4()
\end{array}
$$



Konkret



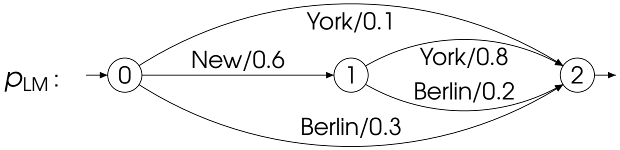
ρ_1 : $S \to \text{New } T$ $\qquad S \to \rho_1(T)$

ρ_2 : $S \to T$ $\qquad S \to \rho_2(T)$

ρ_3 : $T \to \text{York}$ $\qquad T \to \rho_3()$

ρ_4 : $T \to \text{Berlin}$ $\qquad T \to \rho_4()$

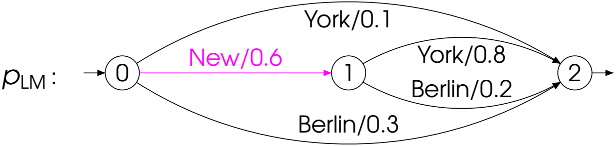
Konkret



ρ_1 :	$S \to \text{New } T$	$S \to \rho_1(T)$
ρ_2 :	$S \to T$	$S \to \rho_2(T)$
ρ_3 :	$T \to \text{York}$	$T \to \rho_3()$
ρ_4 :	$T \to \text{Berlin}$	$T \to \rho_4()$

$(0, S, 2)$

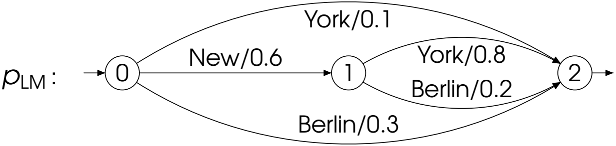
Konkret



ρ_1 :	$S \to$ New T	$S \to \rho_1(T)$
ρ_2 :	$S \to T$	$S \to \rho_2(T)$
ρ_3 :	$T \to$ York	$T \to \rho_3()$
ρ_4 :	$T \to$ Berlin	$T \to \rho_4()$

$(0, S, 2)$

Konkret



ρ_1 : $S \to \text{New } T$ $\qquad S \to \rho_1(T)$

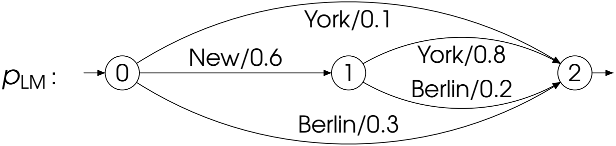
ρ_2 : $S \to T$ $\qquad S \to \rho_2(T)$

ρ_3 : $T \to \text{York}$ $\qquad T \to \rho_3()$

ρ_4 : $T \to \text{Berlin}$ $\qquad T \to \rho_4()$

ρ'_1 : $(0, S, 2) \to \text{New } (1, T, 2)$

Konkret



ρ_1 : $S \to$ New T $\quad$ $S \to \rho_1(T)$

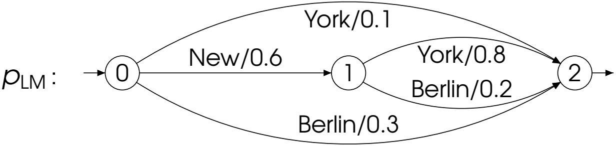
ρ_2 : $S \to T$ $\quad$ $S \to \rho_2(T)$

ρ_3 : $T \to$ York $\quad$ $T \to \rho_3()$

ρ_4 : $T \to$ Berlin $\quad$ $T \to \rho_4()$

ρ'_1 : $(0, S, 2) \to$ New $(1, T, 2)$

Konkret



ρ_1 : $S \to \text{New } T$ $\qquad S \to \rho_1(T)$

ρ_2 : $S \to T$ $\qquad S \to \rho_2(T)$

ρ_3 : $T \to \text{York}$ $\qquad T \to \rho_3()$

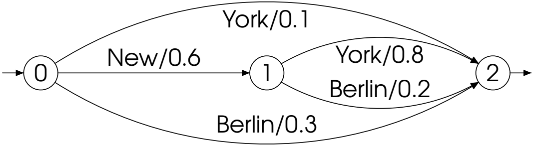
ρ_4 : $T \to \text{Berlin}$ $\qquad T \to \rho_4()$

ρ'_1 : $(0, S, 2) \to \text{New } (1, T, 2)$

ρ'_2 : $(0, S, 2) \to (0, T, 2)$

Konkret

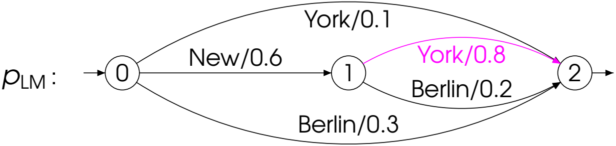
p_{LM} :



$$
\begin{array}{llll}
\rho_1 : & S \to \text{New } T & \qquad & S \to \rho_1(T) \\
\rho_2 : & S \to T & & S \to \rho_2(T) \\
\rho_3 : & T \to \text{York} & & T \to \rho_3() \\
\rho_4 : & T \to \text{Berlin} & & T \to \rho_4()
\end{array}
$$

$$
\begin{array}{ll}
\rho_1' : & (0, S, 2) \to \text{New } (1, T, 2) \\
\rho_2' : & (0, S, 2) \to (0, T, 2) \\
 & (1, T, 2)
\end{array}
$$

Konkret



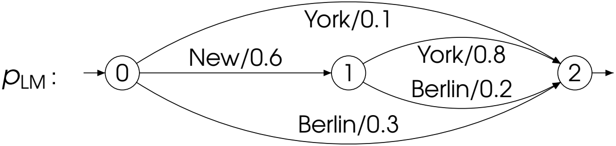
ρ_1 :	$S \rightarrow \text{New } T$	$S \rightarrow \rho_1(T)$
ρ_2 :	$S \rightarrow T$	$S \rightarrow \rho_2(T)$
ρ_3 :	$T \rightarrow \text{York}$	$T \rightarrow \rho_3()$
ρ_4 :	$T \rightarrow \text{Berlin}$	$T \rightarrow \rho_4()$

ρ'_1 : $(0, S, 2) \rightarrow \text{New } (1, T, 2)$

ρ'_2 : $(0, S, 2) \rightarrow (0, T, 2)$

$(1, T, 2)$

Konkret



ρ_1 : $S \to \text{New } T$

ρ_2 : $S \to T$

ρ_3 : $T \to \text{York}$

ρ_4 : $T \to \text{Berlin}$

$S \to \rho_1(T)$

$S \to \rho_2(T)$

$T \to \rho_3()$

$T \to \rho_4()$

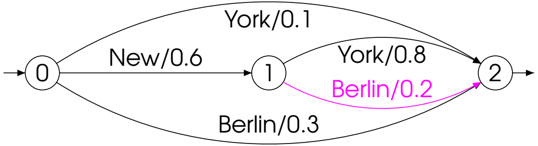
ρ'_1 : $(0, S, 2) \to \text{New } (1, T, 2)$

ρ'_2 : $(0, S, 2) \to (0, T, 2)$

ρ'_3 : $(1, T, 2) \to \text{York}$

Konkret

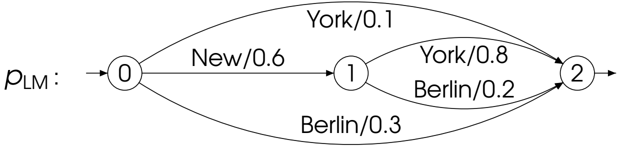
p_{LM} :



$$\begin{aligned}
\rho_1 &: \quad S \rightarrow \text{New } T \qquad & S \rightarrow \rho_1(T) \\
\rho_2 &: \quad S \rightarrow T \qquad & S \rightarrow \rho_2(T) \\
\rho_3 &: \quad T \rightarrow \text{York} \qquad & T \rightarrow \rho_3() \\
\rho_4 &: \quad T \rightarrow \text{Berlin} \qquad & T \rightarrow \rho_4()
\end{aligned}$$

$$\begin{aligned}
\rho'_1 &: \quad (0, S, 2) \rightarrow \text{New } (1, T, 2) \\
\rho'_2 &: \quad (0, S, 2) \rightarrow (0, T, 2) \\
\rho'_3 &: \quad (1, T, 2) \rightarrow \text{York}
\end{aligned}$$

Konkret

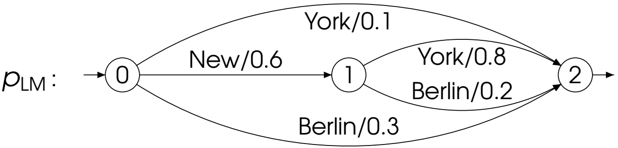


$$\begin{aligned}
\rho_1 &: \quad S \to \text{New } T \\
\rho_2 &: \quad S \to T \\
\rho_3 &: \quad T \to \text{York} \\
\rho_4 &: \quad T \to \text{Berlin}
\end{aligned}$$

$$\begin{aligned}
S &\to \rho_1(T) \\
S &\to \rho_2(T) \\
T &\to \rho_3() \\
T &\to \rho_4()
\end{aligned}$$

$$\begin{aligned}
\rho'_1 &: \quad (0, S, 2) \to \text{New } (1, T, 2) \\
\rho'_2 &: \quad (0, S, 2) \to (0, T, 2) \\
\rho'_3 &: \quad (1, T, 2) \to \text{York} \\
\rho'_4 &: \quad (1, T, 2) \to \text{Berlin}
\end{aligned}$$

Konkret



ρ_1 : $S \to \text{New}\ T$ $\qquad S \to \rho_1(T)$

ρ_2 : $S \to T$ $\qquad S \to \rho_2(T)$

ρ_3 : $T \to \text{York}$ $\qquad T \to \rho_3()$

ρ_4 : $T \to \text{Berlin}$ $\qquad T \to \rho_4()$

ρ'_1 : $(0, S, 2) \to \text{New}\ (1, T, 2)$

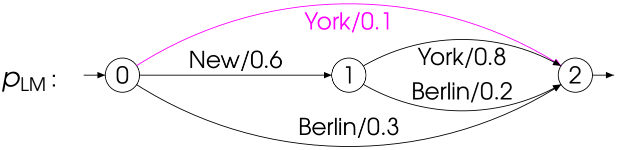
ρ'_2 : $(0, S, 2) \to (0, T, 2)$

ρ'_3 : $(1, T, 2) \to \text{York}$

ρ'_4 : $(1, T, 2) \to \text{Berlin}$

$(0, T, 2)$

Konkret

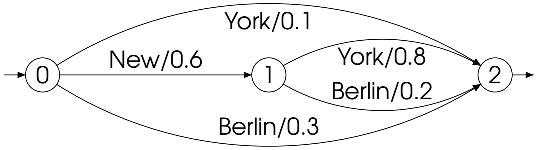


$$\begin{array}{lll}
\rho_1: & S \to \text{New } T & \qquad S \to \rho_1(T) \\
\rho_2: & S \to T & \qquad S \to \rho_2(T) \\
\rho_3: & T \to \text{York} & \qquad T \to \rho_3() \\
\rho_4: & T \to \text{Berlin} & \qquad T \to \rho_4()
\end{array}$$

$$\begin{array}{ll}
\rho_1': & (0, S, 2) \to \text{New } (1, T, 2) \\
\rho_2': & (0, S, 2) \to (0, T, 2) \\
\rho_3': & (1, T, 2) \to \text{York} \\
\rho_4': & (1, T, 2) \to \text{Berlin} \\
 & (0, T, 2)
\end{array}$$

Konkret

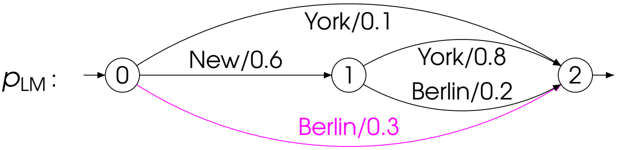
p_{LM} :



$$\begin{array}{lll} \rho_1: & S \to \text{New } T & S \to \rho_1(T) \\ \rho_2: & S \to T & S \to \rho_2(T) \\ \rho_3: & T \to \text{York} & T \to \rho_3() \\ \rho_4: & T \to \text{Berlin} & T \to \rho_4() \end{array}$$

$$\begin{array}{ll} \rho'_1: & (0, S, 2) \to \text{New } (1, T, 2) \\ \rho'_2: & (0, S, 2) \to (0, T, 2) \\ \rho'_3: & (1, T, 2) \to \text{York} \\ \rho'_4: & (1, T, 2) \to \text{Berlin} \\ \rho''_3: & (0, T, 2) \to \text{York} \end{array}$$

Konkret



ρ_1: $S \to \text{New } T$ $\qquad S \to \rho_1(T)$

ρ_2: $S \to T$ $\qquad S \to \rho_2(T)$

ρ_3: $T \to \text{York}$ $\qquad T \to \rho_3()$

ρ_4: $T \to \text{Berlin}$ $\qquad T \to \rho_4()$

ρ'_1: $(0, S, 2) \to \text{New } (1, T, 2)$

ρ'_2: $(0, S, 2) \to (0, T, 2)$

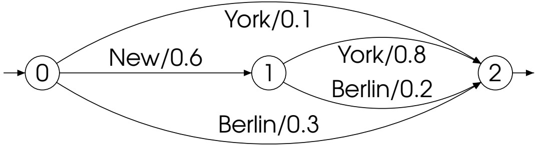
ρ'_3: $(1, T, 2) \to \text{York}$

ρ'_4: $(1, T, 2) \to \text{Berlin}$

ρ''_3: $(0, T, 2) \to \text{York}$

Konkret

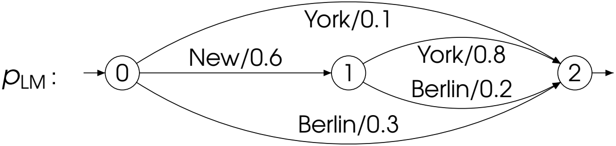
p_{LM}:



$$\begin{array}{lll} \rho_1: & S \to \text{New } T & S \to \rho_1(T) \\ \rho_2: & S \to T & S \to \rho_2(T) \\ \rho_3: & T \to \text{York} & T \to \rho_3() \\ \rho_4: & T \to \text{Berlin} & T \to \rho_4() \end{array}$$

$$\begin{array}{ll} \rho_1': & (0,S,2) \to \text{New } (1,T,2) \\ \rho_2': & (0,S,2) \to (0,T,2) \\ \rho_3': & (1,T,2) \to \text{York} \\ \rho_4': & (1,T,2) \to \text{Berlin} \\ \rho_3'': & (0,T,2) \to \text{York} \\ \rho_4'': & (0,T,2) \to \text{Berlin} \end{array}$$

Konkret



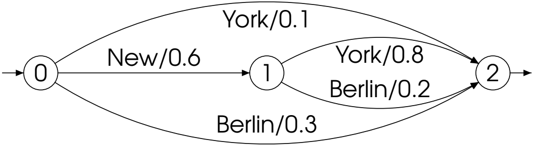
ρ_1 :	$S \to \text{New } T$	$S \to \rho_1(T)$
ρ_2 :	$S \to T$	$S \to \rho_2(T)$
ρ_3 :	$T \to \text{York}$	$T \to \rho_3()$
ρ_4 :	$T \to \text{Berlin}$	$T \to \rho_4()$

ρ'_1 :	$(0, S, 2) \to \text{New } (1, T, 2)$	$(0, S, 2) \to \rho'_1((1, T, 2))$
ρ'_2 :	$(0, S, 2) \to (0, T, 2)$	$(0, S, 2) \to \rho'_2((0, T, 2))$
ρ'_3 :	$(1, T, 2) \to \text{York}$	$(1, T, 2) \to \rho'_3()$
ρ'_4 :	$(1, T, 2) \to \text{Berlin}$	$(1, T, 2) \to \rho'_4()$
ρ''_3 :	$(0, T, 2) \to \text{York}$	$(0, T, 2) \to \rho''_3()$
ρ''_4 :	$(0, T, 2) \to \text{Berlin}$	$(0, T, 2) \to \rho''_4()$

Bar-Hillel et al. (1961)

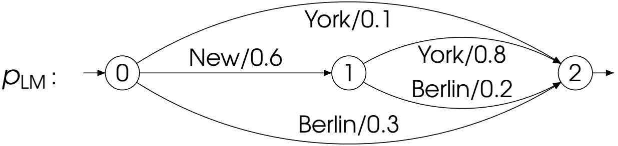
Konkret

p_{LM} :

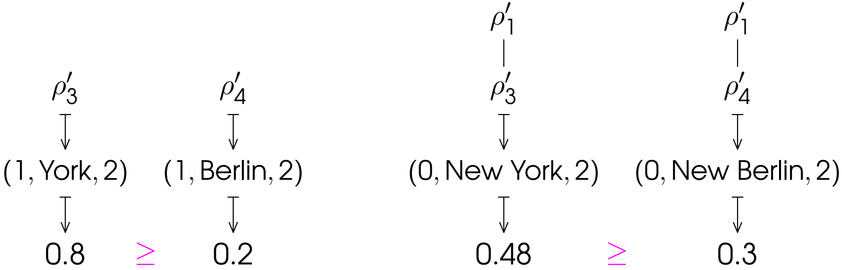


$$
\begin{array}{lll}
\rho'_1 : & (0,S,2) \to \text{New}\,(1,T,2) & (0,S,2) \to \rho'_1((1,T,2)) \\
\rho'_2 : & (0,S,2) \to (0,T,2) & (0,S,2) \to \rho'_2((0,T,2)) \\
\rho'_3 : & (1,T,2) \to \text{York} & (1,T,2) \to \rho'_3() \\
\rho'_4 : & (1,T,2) \to \text{Berlin} & (1,T,2) \to \rho'_4() \\
\rho''_3 : & (0,T,2) \to \text{York} & (0,T,2) \to \rho''_3() \\
\rho''_4 : & (0,T,2) \to \text{Berlin} & (0,T,2) \to \rho''_4()
\end{array}
$$

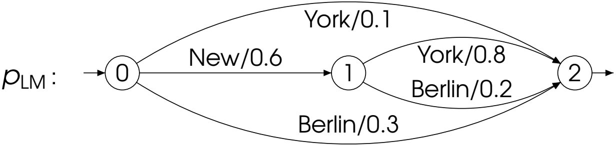
Konkret



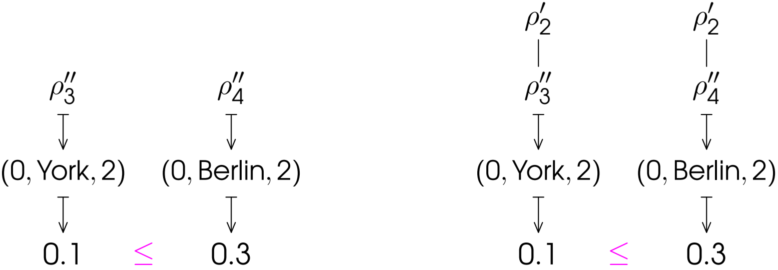
$$
\begin{array}{lll}
\rho'_1: & (0,S,2) \to \text{New}\,(1,T,2) & (0,S,2) \to \rho'_1((1,T,2)) \\
\rho'_2: & (0,S,2) \to (0,T,2) & (0,S,2) \to \rho'_2((0,T,2)) \\
\rho'_3: & (1,T,2) \to \text{York} & (1,T,2) \to \rho'_3() \\
\rho'_4: & (1,T,2) \to \text{Berlin} & (1,T,2) \to \rho'_4() \\
\rho''_3: & (0,T,2) \to \text{York} & (0,T,2) \to \rho''_3() \\
\rho''_4: & (0,T,2) \to \text{Berlin} & (0,T,2) \to \rho''_4()
\end{array}
$$



Konkret

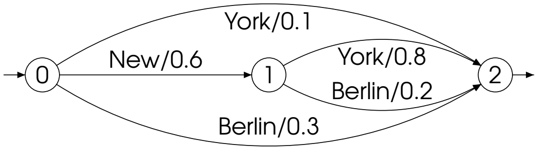


$$
\begin{array}{lll}
\rho'_1: & (0,S,2) \to \text{New}\,(1,T,2) & (0,S,2) \to \rho'_1\,((1,T,2)) \\
\rho'_2: & (0,S,2) \to (0,T,2) & (0,S,2) \to \rho'_2\,((0,T,2)) \\
\rho'_3: & (1,T,2) \to \text{York} & (1,T,2) \to \rho'_3() \\
\rho'_4: & (1,T,2) \to \text{Berlin} & (1,T,2) \to \rho'_4() \\
\rho''_3: & (0,T,2) \to \text{York} & (0,T,2) \to \rho''_3() \\
\rho''_4: & (0,T,2) \to \text{Berlin} & (0,T,2) \to \rho''_4()
\end{array}
$$



Konkret

p_{LM} :

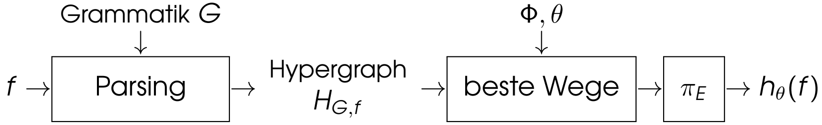


$$
\begin{array}{lll}
\rho'_1: & (0,S,2) \to \text{New}\ (1,T,2) & (0,S,2) \to \rho'_1((1,T,2)) \\
\rho'_2: & (0,S,2) \to (0,T,2) & (0,S,2) \to \rho'_2((0,T,2)) \\
\rho'_3: & (1,T,2) \to \text{York} & (1,T,2) \to \rho'_3() \\
\rho'_4: & (1,T,2) \to \text{Berlin} & (1,T,2) \to \rho'_4() \\
\rho''_3: & (0,T,2) \to \text{York} & (0,T,2) \to \rho''_3() \\
\rho''_4: & (0,T,2) \to \text{Berlin} & (0,T,2) \to \rho''_4()
\end{array}
$$

Chiang (2007)

Parsing

$$h_\theta : f \mapsto \pi_E \left(\operatorname*{argmax}_{\substack{y \in Y : \\ \pi_F(y) = f}} \Phi(y) \cdot \theta \right)$$



Parsing als Merkmal

zum Beispiel

- $\Phi(y)_{m+1} = \begin{cases} 0 & \text{falls } \pi_F(y) = f \\ -\infty & \text{sonst} \end{cases}$

Parsing als Merkmal

zum Beispiel

- $\Phi(y)_{m+1} = \begin{cases} 0 & \text{falls } \pi_F(y) = f \\ -\infty & \text{sonst} \end{cases}$
- $\Phi(y)_{m+1} = \log p(\pi'_F(y) \mid f)$

  $\pi'_F(y)$ … Syntaxbaum der Quellsprache zu y

Parsing als Merkmal

zum Beispiel

- $\Phi(y)_{m+1} = \begin{cases} 0 & \text{falls } \pi_F(y) = f \\ -\infty & \text{sonst} \end{cases}$
- $\Phi(y)_{m+1} = \log p(\pi'_F(y) \mid f)$

  $\pi'_F(y)$ ... Syntaxbaum der Quellsprache zu y

dann

$$\operatorname*{argmax}_{\substack{y \in Y: \\ \pi_F(y) = f}} \Phi(y) \cdot \theta = \operatorname{argmax}_{y \in Y} \Phi(y) \cdot \theta$$

Parsing als Merkmal

zum Beispiel

- $$\Phi(y)_{m+1} = \begin{cases} 0 & \text{falls } \pi_F(y) = f \\ -\infty & \text{sonst} \end{cases}$$
- $\Phi(y)_{m+1} = \log p(\pi'_F(y) \mid f)$

  $\pi'_F(y)$ … Syntaxbaum der Quellsprache zu y

dann

$$\operatorname{argmax}_{\substack{y \in Y: \\ \pi_F(y) = f}} \Phi(y) \cdot \theta = \operatorname{argmax}_{y \in Y} \Phi(y) \cdot \theta$$

monoton?:

$$\varphi(\text{York}) = \varphi(\text{Berlin}) \implies \varphi(\text{New York}) = \varphi(\text{New Berlin})$$

Parsing als Merkmal

zum Beispiel

- $\Phi(y)_{m+1} = \begin{cases} 0 & \text{falls } \pi_F(y) = f \\ -\infty & \text{sonst} \end{cases}$
- $\Phi(y)_{m+1} = \log p(\pi'_F(y) \mid f)$

  $\pi'_F(y)$ …Syntaxbaum der Quellsprache zu y

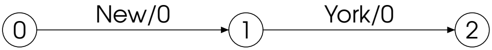
dann

$$\operatorname{argmax}_{\substack{y \in Y: \\ \pi_F(y)=f}} \Phi(y) \cdot \theta = \operatorname{argmax}_{y \in Y} \Phi(y) \cdot \theta$$

nicht monoton!:

$$\varphi(\text{York}) = \varphi(\text{Berlin}) \implies \varphi(\text{New York}) = \varphi(\text{New Berlin}) \quad ↯$$

Beispiel



$$
\begin{array}{lll}
\rho_1: & S \rightarrow \text{New } T & S \rightarrow \rho_1(T) \\
\rho_2: & S \rightarrow T & S \rightarrow \rho_2(T) \\
\rho_3: & T \rightarrow \text{York} & T \rightarrow \rho_3() \\
\rho_4: & T \rightarrow \text{Berlin} & T \rightarrow \rho_4()
\end{array}
$$

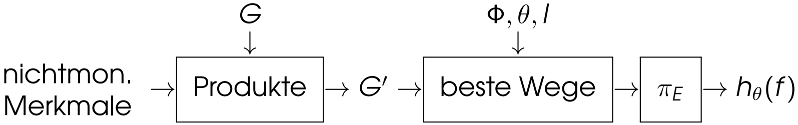
$$
\begin{array}{lll}
\rho'_1: & (0, S, 2) \rightarrow \text{New } (1, T, 2) & (0, S, 2) \rightarrow \rho'_1((1, T, 2)) \\
\rho'_3: & (1, T, 2) \rightarrow \text{York} & (1, T, 2) \rightarrow \rho'_3()
\end{array}
$$

Ergebnis

$$h_\theta \colon f \mapsto \pi_E\Big(\operatorname{argmax}_{y \in Y} \Phi_f(y) \cdot \theta\Big)$$

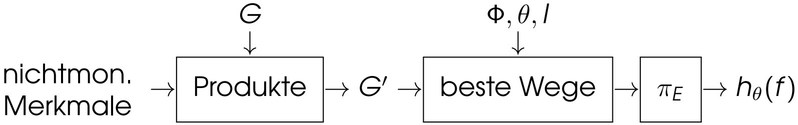
Ergebnis

$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{y \in Y} \Phi_f(y) \cdot \theta\right)$$



Ergebnis

$$h_\theta : f \mapsto \pi_E\left(\operatorname{argmax}_{y \in Y} \Phi_f(y) \cdot \theta\right)$$



- Kombinierbarkeit Büchse et al. (2011)
- Toolbox

WATA 2012:
Weighted Automata –
Theory and Applications

29. Mai – 2. Juni 2012
Dresden

Kevin Knight wird da sein!
+ Band in TCS
(Theoretical Computer Science)

Literatur

Bar-Hillel, Y., M. Perles, and E. Shamir (1961). “On Formal Properties of Simple Phrase Structure Grammars”. In: Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 14, pp. 143–172.

Büchse, Matthias, Mark-Jan Nederhof, and Heiko Vogler (2011). “Tree parsing with synchronous tree-adjoining grammars”. In: Proc. IWPT 2011, pp. 14–25.

Chiang, D. (2007). “Hierarchical Phrase-Based Translation”. In: Comp. Ling. 33.2, pp. 201–228. ISSN: 0891-2017.

Collins, Michael (2005). “Parameter Estimation for Statistical Parsing Models: Theory and Practice of Distribution-Free Methods”. In: New Developments in Parsing Technology. Vol. 23. Text, Speech and Language Technology. Springer Netherlands. Chap. 2, pp. 19–55.

Liang, Percy et al. (2006). “An End-to-End Discriminative Approach to Machine Translation”. In: Proc. COLING-ACL 2006, pp. 761–768.