Maschinelles Übersetzen natürlicher Sprachen

9. Übungsblatt

2017-01-05

Aufgabe 1

Gegeben seien die Alphabete

$$\Gamma = \{\mathrm{S}, \mathrm{NP}, \mathrm{VP}, \mathrm{ADJ}, \mathrm{NN}, \mathrm{VB}, \mathrm{DT}\},$$
$$\Sigma = \Gamma \cup \{\mathrm{el}, \mathrm{perro}, \mathrm{bravo}, \mathrm{escucha}, \mathrm{coge}, \mathrm{hombre}\},$$
$$\Delta = \Gamma \cup \{\mathrm{der}, \mathrm{den}, \mathrm{Hund}, \mathrm{wilde}, \mathrm{wilden}, \mathrm{Mann}, \mathrm{fängt}, \mathrm{hört}\}.$$

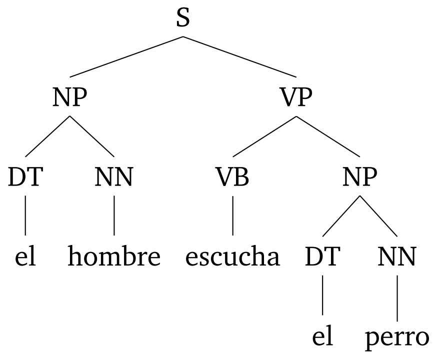
Wir nehmen einen extended tree transducer $\mathcal{M} = (Q, \Sigma, \Delta, q, R)$ an mit der Zustandsmenge $Q = \{q, n, a\}$ und der Regelmenge R wie folgt:

$$q(\mathrm{S}(x_1 : \mathrm{NP}, \mathrm{VP}(x_2 : \mathrm{VB}, x_3 : \mathrm{NP}))) \xrightarrow{\rho_1} \mathrm{S}(n(x_1), \mathrm{VP}(q(x_2), a(x_3)))$$
$$n(\mathrm{NP}(\mathrm{DT}(\mathrm{el}), x_1 : \mathrm{NN}, x_2 : \mathrm{ADJ})) \xrightarrow{\rho_2} \mathrm{NP}(\mathrm{DT}(\mathrm{der}), n(x_2), q(x_1))$$
$$a(\mathrm{NP}(\mathrm{DT}(\mathrm{el}), x_1 : \mathrm{NN}, x_2 : \mathrm{ADJ})) \xrightarrow{\rho_3} \mathrm{NP}(\mathrm{DT}(\mathrm{den}), a(x_2), q(x_1))$$
$$n(\mathrm{NP}(\mathrm{DT}(\mathrm{el}), x_1 : \mathrm{NN})) \xrightarrow{\rho_4} \mathrm{NP}(\mathrm{DT}(\mathrm{der}), q(x_1))$$
$$a(\mathrm{NP}(\mathrm{DT}(\mathrm{el}), x_1 : \mathrm{NN})) \xrightarrow{\rho_5} \mathrm{NP}(\mathrm{DT}(\mathrm{den}), q(x_1))$$

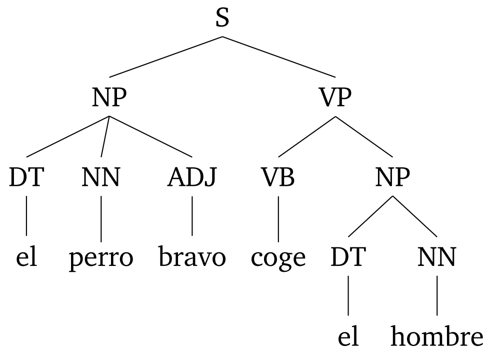
$$q(\mathrm{NN}(\mathrm{perro})) \xrightarrow{\rho_6} \mathrm{NN}(\mathrm{Hund})$$
$$q(\mathrm{NN}(\mathrm{hombre})) \xrightarrow{\rho_7} \mathrm{NN}(\mathrm{Mann})$$
$$q(\mathrm{VB}(\mathrm{escucha})) \xrightarrow{\rho_8} \mathrm{VB}(\mathrm{hört})$$
$$q(\mathrm{VB}(\mathrm{coge})) \xrightarrow{\rho_9} \mathrm{VB}(\mathrm{fängt})$$
$$a(\mathrm{ADJ}(\mathrm{bravo})) \xrightarrow{\rho_{10}} \mathrm{ADJ}(\mathrm{wilden})$$
$$n(\mathrm{ADJ}(\mathrm{bravo})) \xrightarrow{\rho_{11}} \mathrm{ADJ}(\mathrm{wilde})$$

Bestimmen Sie, unter Angabe der entsprechenden Ableitungen, die Übersetzungen der folgenden Parsebäume:

$\xi_1 =$



$\xi_2 =$



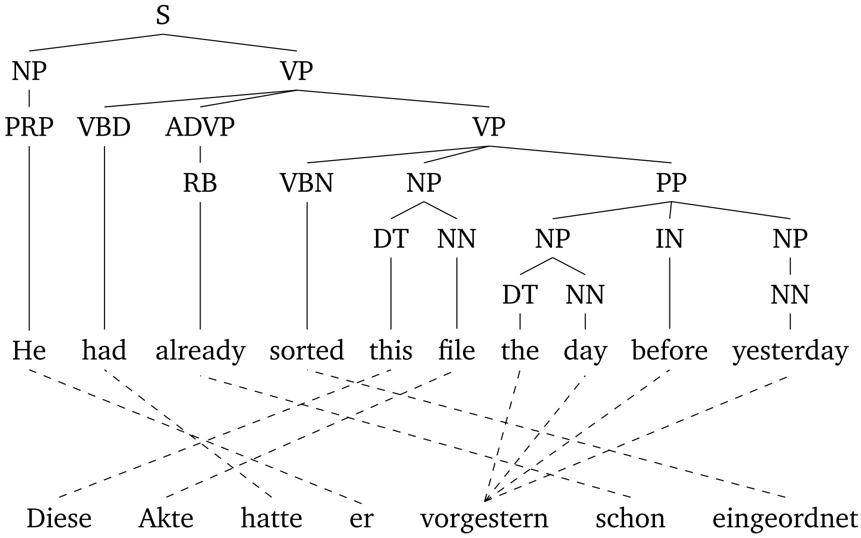
Vollziehen Sie für eine dieser Ableitungen die Arbeitsweise der Projektionsfunktionen $\pi_\Sigma \colon D_\mathcal{M} \to T_\Sigma$ sowie $\pi_\Delta \colon D_\mathcal{M} \to T_\Delta$ nach! Wie müssten die Regeln von $\mathcal{M}$ geändert werden, damit der Eingabe- bzw. Ausgabebaum einer Ableitung nicht eindeutig bestimmt ist?

Aufgabe 2

Wir betrachten noch einmal die Übersetzung von arithmetischen Ausdrücken in Infix-Notation in umgekehrte polnische Notation (vgl. 8. Übungsblatt, Aufgabe 1). Entwerfen Sie einen XTT $\mathcal{M}$, welcher einen Parsebaum $\xi \in PT_G$ eines Infix-Ausdrucks in den entsprechenden Parsebaum einer CFG G' für RPN-Ausdrücke übersetzt!

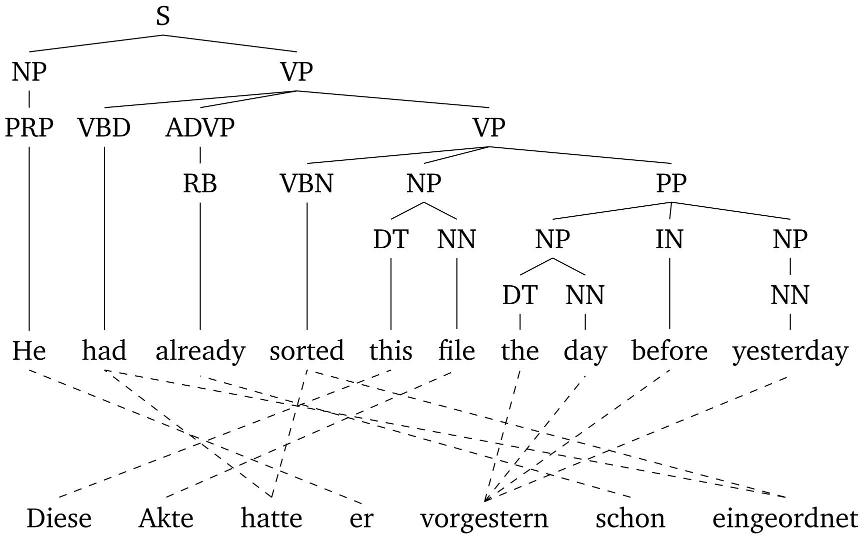
Zusatzaufgabe: Entwerfen Sie außerdem einen yXTT $\mathcal{M}'$, welcher direkt den Yield des abgeleiteten Baumes ausgibt!

Aufgabe 3 (Regelextraktion für yXTT)
Gegeben ist der folgende Alignment-Graph $G(\xi, f, A)$:



Nutzen Sie die in der Vorlesung vorgestellte Technik, um Regeln für einen yXTT zu erzeugen. Kennzeichnen Sie dabei die *frontier nodes* des Graphen!

Zusatzaufgabe: Inwiefern verändern sich die Regeln, wenn wir zum Extrahieren den folgenden Alignment-Graphen $G(\xi, f, A')$ verwenden? Welche Regeln versprechen eine bessere Übersetzung?



Aufgabe 4 (Konstruktion von RTG)
Gegeben sei das Alphabet $\Sigma = \{\sigma, \tau, \gamma, \alpha\}$. Konstruieren Sie eine RTG $\mathcal{G}$, so dass $\xi \in L(\mathcal{G})$ genau dann, wenn

- die Symbole σ und τ nur mit Rang 2 in ξ vorkommen, γ nur mit Rang 1 und α mit Rang 0, und außerdem
- das Symbol τ genau einmal in ξ enthalten ist.

Geben Sie für den Baum $\sigma\big(\alpha, \tau(\alpha, \gamma(\alpha))\big)$ eine Ableitung von $\mathcal{G}$ an!

Aufgabe 5 (Top-Down-Determinismus)
Wir nehmen das Alphabet $\Sigma = \{\sigma, \alpha, \beta\}$ an. Warum existiert für die Baumsprache $L = \{\sigma(\alpha,\beta), \sigma(\beta,\alpha)\}$ keine top-down-deterministische RTG $\mathcal{G}$ mit $L = L(\mathcal{G})$?

Wie stehen die Sprachklassen $\textsc{Reg}$ und $\textsc{Reg}_{det}$ daher also in Beziehung, wenn $\textsc{Reg}$ die Klasse aller durch reguläre Baumgrammatiken, und $\textsc{Reg}_{det}$ die Klasse aller durch top-down-deterministische reguläre Baumgrammatiken erzeugten Baumsprachen bezeichnet?

Welchen Vorteil bieten deterministische Grammatiken jedoch im Vergleich zu nicht deterministischen?

Aufgabe 6 (Input- and Output-Product)
Sei $\Sigma = \{\sigma, \gamma, \alpha, \beta\}$. Wir nehmen einen XTT $\mathcal{M} = (\{q,p\}, \Sigma, \Sigma, q, R)$ an mit den Regeln

$$q(\sigma(x_1,x_2)) \xrightarrow{\rho_1} \sigma(q(x_2), p(x_1)) \qquad q(\alpha) \xrightarrow{\rho_3} \alpha$$
$$p(\gamma(x_1)) \xrightarrow{\rho_2} \gamma(\gamma(\gamma(p(x_1)))) \qquad p(\alpha) \xrightarrow{\rho_4} \beta,$$

sowie RTGs $\mathcal{G}_1 = (\{S,A\}, \Sigma, S, R_1)$ und $\mathcal{G}_2 = (\{S,B,C\}, \Sigma, S, R_2)$, wobei die Regeln aus R_1 durch

$$\begin{array}{ll} S \to \sigma(A,S) & A \to \gamma(A) \\ S \to \alpha & A \to \alpha \end{array}$$

und die aus R_2 durch

$$\begin{array}{ll} S \to \sigma(S,B) & S \to \alpha \\ B \to \gamma(C) & B \to \beta \\ C \to \gamma(B) & \end{array}$$

gegeben sind.

Bilden Sie mittels der in der Vorlesung vorgestellten Konstruktion das *Input- and Output-Product* $\mathrm{Prod}(\mathcal{G}_1, \mathcal{M}, \mathcal{G}_2)$.

Fleißaufgabe: Berechnen Sie $\mathrm{Prod}(\mathcal{G}_1', \mathcal{M}, \mathcal{G}_2)$ für $\mathcal{G}_1' = (\{S,A,B,C\}, \Sigma, S, R_1')$ mit den Regeln

$$\begin{array}{lll} S \to \sigma(A,S) & S \to \alpha & A \to \alpha \\ A \to \gamma(B) & B \to \gamma(C) & C \to \gamma(A). \end{array}$$