

Maschinelles Übersetzen natürlicher Sprachen

Aufgabe 1

Wir betrachten noch einmal die Übersetzung von arithmetischen Ausdrücken in Infix-Notation in umgekehrte polnische Notation (vgl. letztes Übungsblatt, Aufgabe 1). Es sei also wieder die kontextfreie Grammatik $G = (N, \Sigma, S, R)$ gegeben mit $N = \{S, M, E, L, D\}$, Terminalen

$$\Sigma = \{+, -, \times, \div, (,), 0, \dots, 9\},$$

und der Regelmenge

$$\begin{aligned} R = \{ & S \rightarrow S + M, \quad S \rightarrow S - M, \quad S \rightarrow M, \\ & M \rightarrow M \times E, \quad M \rightarrow M \div E, \quad M \rightarrow E, \\ & E \rightarrow (S), \quad E \rightarrow L, \\ & L \rightarrow DL, \quad L \rightarrow D\} \\ & \cup \{D \rightarrow i \mid i \in \{0, \dots, 9\}\}. \end{aligned}$$

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 2

Gegeben seien die Alphabete

$$\Gamma = \{S, NP, VP, ADJ, NN, VB, DT\},$$

sowie

$$\Sigma = \Gamma \cup \{\text{el, perro, bravo, escucha, coge, hombre}\}$$

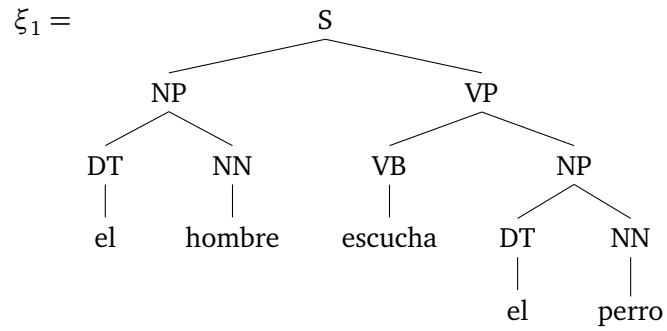
und

$$\Delta = \Gamma \cup \{\text{der, den, Hund, wilde, wilden, Mann, fängt, hört}\}.$$

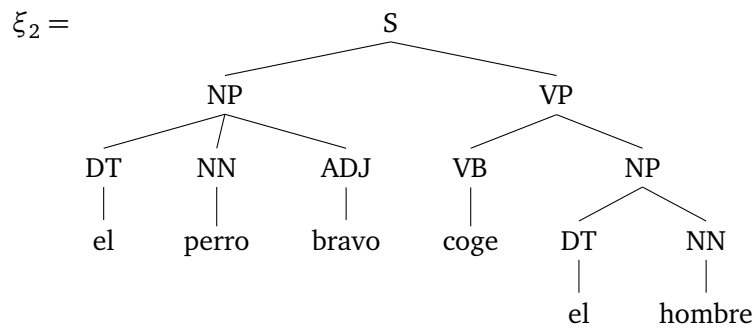
Wir nehmen einen extended tree transducer $\mathcal{M} = (Q, \Sigma, \Delta, q, R)$ an mit der Zustandsmenge $Q = \{q, n, a\}$ und der Regelmenge R wie folgt:

$q(S(x_1 : NP, VP(x_2 : VB, x_3 : NP))) \rightarrow S(n(x_1), VP(q(x_2), a(x_3)))$	$q(NN(\text{perro})) \rightarrow NN(\text{Hund})$
$n(NP(DT(\text{el}), x_1 : NN, x_2 : ADJ)) \rightarrow NP(DT(\text{der}), n(x_2), q(x_1))$	$q(NN(\text{hombre})) \rightarrow NN(\text{Mann})$
$a(NP(DT(\text{el}), x_1 : NN, x_2 : ADJ)) \rightarrow NP(DT(\text{den}), a(x_2), q(x_1))$	$q(VB(\text{escucha})) \rightarrow VB(\text{hört})$
$n(NP(DT(\text{el}), x_1 : NN)) \rightarrow NP(DT(\text{der}), q(x_1))$	$q(VB(\text{coge})) \rightarrow VB(\text{fängt})$
$a(NP(DT(\text{el}), x_1 : NN)) \rightarrow NP(DT(\text{den}), q(x_1))$	$a(ADJ(\text{bravo})) \rightarrow ADJ(\text{wilden})$
	$n(ADJ(\text{bravo})) \rightarrow ADJ(\text{wilde})$

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



und



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