Maschinelles Übersetzen natürlicher Sprachen

8. Übungsblatt

2015-01-08

Aufgabe 1 (Extraktion von SCFG-Regeln)
Bestimmen Sie in folgendem alinierten Satzpaar die initial phrase pairs.

	der	Hund	hat	den	Knochen	heute	noch	nicht	gefunden
today						×			
the	×								
dog		×							
has			×						
not								×	
found									×
the				×					
bone					×				
yet							×		

Geben Sie einige der SCFG-Regeln an, die aus diesen Phrasenpaaren extrahiert werden.

Aufgabe 2 (Training von SCFGs)
Betrachten Sie die folgenden beiden alinierten Satzpaare.

$e_1 \backslash f_1$	Kinder	spielen
children	×	
play		×

$e_2 \backslash f_2$	lasst	uns	spielen
let's	×	×	
play			×

Ihre Korpushäufigkeiten seien gegeben durch $h(e_1, f_1) = 1$ und $h(e_2, f_2) = 2$ (und für alle Satzpaare $(e, f) \notin \{(e_1, f_1), (e_2, f_2)\}$ soll $h(e, f) = 0$ gelten).

Extrahieren Sie aus dem so gegebenen Korpus SCFG-Regeln und trainieren Sie deren Wahrscheinlichkeiten mittels der in der Vorlesung vorgestellten Methode!

Aufgabe 3
Wir betrachten noch einmal die Übersetzung von arithmetischen Ausdrücken in Infix-Notation in umgekehrte polnische Notation (vgl. 7. Übungsblatt, Aufgabe 3). 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 4
Gegeben seien die Alphabete

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

sowie

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

und

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

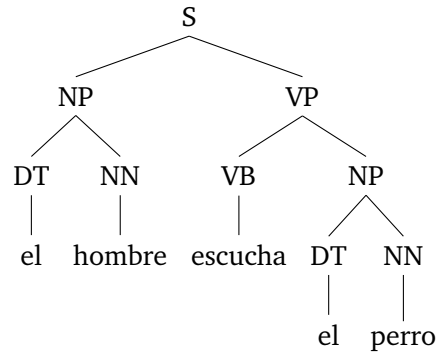
Wir nehmen einen extended tree transducer $\mathcal{M} = (Q, \Sigma, \Delta, q, R)$ an mit der Zustandsmenge $Q = \{q, n, a\}$ und der Regelmenge R wie folgt:

$$\begin{aligned}
q(\text{S}(x_1 : \text{NP}, \text{VP}(x_2 : \text{VB}, x_3 : \text{NP}))) &\to \text{S}(n(x_1), \text{VP}(q(x_2), a(x_3))) \\
n(\text{NP}(\text{DT}(\text{el}), x_1 : \text{NN}, x_2 : \text{ADJ})) &\to \text{NP}(\text{DT}(\text{der}), n(x_2), q(x_1)) \\
a(\text{NP}(\text{DT}(\text{el}), x_1 : \text{NN}, x_2 : \text{ADJ})) &\to \text{NP}(\text{DT}(\text{den}), a(x_2), q(x_1)) \\
n(\text{NP}(\text{DT}(\text{el}), x_1 : \text{NN})) &\to \text{NP}(\text{DT}(\text{der}), q(x_1)) \\
a(\text{NP}(\text{DT}(\text{el}), x_1 : \text{NN})) &\to \text{NP}(\text{DT}(\text{den}), q(x_1))
\end{aligned}$$

$$\begin{aligned}
q(\text{NN}(\text{perro})) &\to \text{NN}(\text{Hund}) \\
q(\text{NN}(\text{hombre})) &\to \text{NN}(\text{Mann}) \\
q(\text{VB}(\text{escucha})) &\to \text{VB}(\text{hört}) \\
q(\text{VB}(\text{coge})) &\to \text{VB}(\text{fängt}) \\
a(\text{ADJ}(\text{bravo})) &\to \text{ADJ}(\text{wilden}) \\
n(\text{ADJ}(\text{bravo})) &\to \text{ADJ}(\text{wilde})
\end{aligned}$$

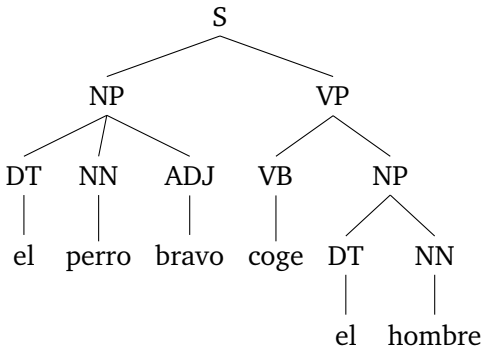
Bestimmen Sie, unter Angabe der entsprechenden Ableitungen, die Übersetzungen der Parsebäume

$\xi_1 =$



und

$\xi_2 =$



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