

Maschinelles Übersetzen natürlicher Sprachen

Aufgabe 1

Gegeben sei ein Alphabet Δ . Die Menge der ranglosen Bäume, wie sie in der Vorlesung eingeführt wurde, bezeichnen wir als U_Δ . Es sei ein beliebiger Baum $\xi \in U_\Delta$ gegeben. Definieren Sie formal die folgenden Funktionen:

- $\text{rk}_\xi: \text{pos}(\xi) \rightarrow \mathbb{N}$, wobei $\text{rk}_\xi(\rho)$ den Rang des Knotens an der Position $\rho \in \text{pos}(\xi)$ in ξ bestimmen soll;
- $\xi(\cdot): \text{pos}(\xi) \rightarrow \Delta$, dabei soll $\xi(\rho)$ das Label von ξ an einer Position $\rho \in \text{pos}(\xi)$ berechnen;
- $\text{yield}: U_\Delta \rightarrow \Delta^*$, hierbei soll $\text{yield}(\xi)$ die Sequenz der Blattsymbole von ξ zurückgeben, in der Reihenfolge von links nach rechts.

Aufgabe 2

Die Regeln einer kontextfreien Grammatik seien wie folgt gegeben:

$S \rightarrow NP VP$
 $NP \rightarrow \text{Pronoun} \mid \text{Proper-Noun} \mid \text{Det Nominal}$
 $\text{Nominal} \rightarrow \text{Noun Nominal} \mid \text{Noun}$
 $VP \rightarrow \text{Verb} \mid \text{Verb NP} \mid \text{Verb NP PP} \mid \text{Verb PP}$
 $PP \rightarrow \text{Preposition NP}$
 $\text{Noun} \rightarrow \text{flight} \mid \text{breeze} \mid \text{trip} \mid \text{morning} \mid \dots$
 $\text{Verb} \rightarrow \text{is} \mid \text{prefer} \mid \text{like} \mid \text{need} \mid \text{want} \mid \text{fly}$
 $\text{Pronoun} \rightarrow \text{me} \mid \text{I} \mid \text{you} \mid \text{it} \mid \dots$
 $\text{Proper-Noun} \rightarrow \text{Alaska} \mid \text{Baltimore} \mid \text{Los Angeles} \mid \text{Chicago} \mid \dots$
 $\text{Det} \rightarrow \text{the} \mid \text{a} \mid \text{an} \mid \text{this} \mid \text{these} \mid \text{that} \mid \dots$
 $\text{Preposition} \rightarrow \text{from} \mid \text{to} \mid \text{on} \mid \text{near} \mid \dots$

Bestimmen Sie mithilfe dieser Grammatik eine Linksableitung des Satzes "I fly to Alaska" und geben Sie den entsprechenden abstrakten Syntaxbaum sowie Parsebaum an.

Aufgabe 3

Gegeben sei eine PCFG $G = (N, \Sigma, S, R, p)$ mit Nichtterminalen $N = \{S\}$, Terminalalphabet $\Sigma = \{a\}$ und den zwei Regeln

$$\begin{array}{ll} S \rightarrow SS & (q) \\ S \rightarrow a & (1 - q), \end{array}$$

wobei $q \in [0, 1]$.

1. Treffen Sie formal begründete Aussagen über die Konsistenz von G in Abhängigkeit von q .
2. Bestimmen Sie ausgehend von dieser Grammatik $p(a^n)$, für $n \in \mathbb{N}$.

Aufgabe 4 (Zusatzaufgabe)

Wir betrachten nun die PCFG $G' = (N, \Sigma, S, R', p')$ mit N und Σ wie in Aufgabe 3, sowie Regeln

$$S \rightarrow SS \quad (0.8)$$

$$S \rightarrow \varepsilon \quad (0.2).$$

Bestimmen Sie für G' die Wortwahrscheinlichkeit $p'(\varepsilon)$.