

# Maschinelles Übersetzen natürlicher Sprachen

## Aufgabe 1 (Übersetzung arithmetischer Ausdrücke)

Betrachten Sie die kontextfreie Grammatik  $G = (N, \Sigma, S, R)$  mit  $N = \{S, M, E, L, D\}$  und

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

Ihre Regelmenge sei gegeben durch

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

Die durch  $G$  erzeugte Sprache ist die der arithmetischen Ausdrücke in Infix-Notation. Entwerfen Sie nun eine synchrone kontextfreie Grammatik, mit der man solche Infix-Ausdrücke in die *umgekehrte polnische Notation* übersetzen kann. Beispielsweise sollte der Ausdruck  $1 + (23 + 4) \times 5$  in  $1.23.4.+5.\times+$  umgewandelt werden. Dabei soll das Symbol  $.$  das Ende einer Zahl kennzeichnen. Geben Sie für dieses Übersetzungsbeispiel eine Ableitung in der erstellten Grammatik an!

## Aufgabe 2 (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 3 (Training von SCFGs)

Betrachten Sie die folgenden beiden alinierten Satzpaare.

$e_1 \setminus f_1$	Kinder	spielen	$e_2 \setminus f_2$	lasst	uns	spielen
children	×		let's	×	×	
play		×	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!