Computational Linguistics II: Parsing

Summing up CF Languages: Derivations

Frank Richter & Jan-Philipp Söhn

fr@sfs.uni-tuebingen.de, jp.soehn@uni-tuebingen.de

The Big Picture

hierarchy	grammar	machine	other
type 3	reg. grammar	D/NFA	reg. expressions
det. cf.	LR(k) grammar	DPDA	
type 2	CFG	PDA	
type 1	CSG	LBA	
type 0	unrestricted grammar	Turing machine	

Form of Grammars of Type 0–3

For $i \in \{0, 1, 2, 3\}$, a grammar $\langle N, T, P, S \rangle$ of Type i, with N the set of non-terminal symbols, T the set of terminal symbols (N and T disjoint, $\Sigma = N \cup T$), P the set of productions, and S the start symbol ($S \in N$), obeys the following restrictions:

T3: Every production in P is of the form $A \rightarrow aB$ or $A \rightarrow \epsilon$, with $B, A \in N$, $a \in T$.

T2: Every production in P is of the form $A \rightarrow x$, with $A \in N$ and $x \in \Sigma^*$.

T1: Every production in P is of the form $x_1 A x_2 \rightarrow x_1 y x_2$, with $x_1, x_2 \in \Sigma^*$, $y \in \Sigma^+$, $A \in N$ and the possible exception of $C \rightarrow \epsilon$ in case C does not occur on the righthand side of a rule in P.

T0: No restrictions.

Derivation Steps of Grammars

Definition

For every grammar G with $G = \langle N, T, P, S \rangle$ and $\Sigma = N \cup T$, for every $u, v \in \Sigma^*$,

if there is a rule $l \rightarrow r \in P$ with $u = w_1 l w_2$ and $v = w_1 r w_2$, where $w_1, w_2 \in \Sigma^*$ then

$u \Rightarrow_G^1 v$.

We say that u directly derives v in grammar G.

We write $\Rightarrow_G^*$ for the reflexive transitive closure of $\Rightarrow_G^1$ and omit the subscript G if the grammar is clear from the context.

Language Generated by a Grammar

Definition

For every grammar G with $G = \langle N, T, P, S \rangle$ the language $L(G)$ generated by G is

$$L(G) = \{x \in T^* | S \Rightarrow_G^* x\}.$$

More on Derivations (1)

If at each step in a derivation a production is applied to the leftmost nonterminal, then the derivation is said to be *leftmost*.

More on Derivations (1)

If at each step in a derivation a production is applied to the leftmost nonterminal, then the derivation is said to be *leftmost*.

A derivation in which the rightmost nonterminal is replaced at each step is said to be *rightmost*.