Computational Linguistics II: Parsing

The CYK Parser

Frank Richter & Jan-Philipp Söhn

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

ϵ-free Type 2 Grammars

Lemma 1
There is an algorithm which for each context free grammar $G = \langle N, T, P, S \rangle$ produces an equivalent ϵ-free context free grammar $G' = \langle N', T', P', S' \rangle$.

Sketch of the procedure: Let
$W_1 = \{A \in N | A \rightarrow \epsilon \in P\}$,
$W_{i+1} = \{A \in N | A \rightarrow x \in P$ with $x \in W_i^*\} \cup W_i$.

1. $W_i \subseteq W_{i+1}$ ($i \geq 1$),
2. If $W_i = W_{i+1}$ then $W_i = W_{i+m}$ for $m \geq 0$,
3. $W_n = W_{n+1}$, $n = |N|$,
4. $W_n = \{A \in N | A \Rightarrow_G^* \epsilon\}$

ϵ-free Type 2 Grammars

Let
$N' = N \cup S'$, S' a new start symbol,
$T' = T$,
$P' = \{S' \rightarrow S\} \cup$
$\{A \rightarrow A_1 \dots A_k | k \geq 1, A_i \in N \cup T$
and there are $x_1 \dots x_{k+1} \in W_n^*$ with
$A \rightarrow x_1 A_1 x_2 \dots A_k x_{k+1} \in P\} \cup P_\epsilon$,
with $P_\epsilon = \{\}$ in case $\epsilon \notin L(G)$, else $P_\epsilon = S' \rightarrow \epsilon$.

Type 2 Grammars without Chain Rules

Lemma 2
There is an algorithm which for each ϵ-free context free grammar $G = \langle N, T, P, S \rangle$ produces an equivalent ϵ-free context free grammar $G' = \langle N', T', P', S' \rangle$ without chain rules.

Sketch of the procedure: For each $A \in N$, let
$W_0(A) = \{A\}$,
$W_{i+1}(A) = \{B \in N | C \rightarrow B \in P$ for some $C \in W_i(A)\} \cup W_i(A)$.

1. $W_i(A) \subseteq W_{i+1}(A)$ ($i \geq 1$),
2. If $W_i(A) = W_{i+1}(A)$ then $W_i(A) = W_{i+m}(A)$ for $m \geq 0$,
3. $W_n(A) = W_{n+1}(A)$, $n = |N|$,
4. $W_n(A) = \{B \in N | A \Rightarrow_G^* B\}$

Type 2 Grammars without Chain Rules

Let
$N' = N$,
$T' = T$,
$S' = S$,
$P' = \{A \rightarrow x | x \notin N, B \rightarrow x \in P$ for some $B \in W_n(A)\}$

Grammars w/o Useless Symbols

Lemma 3

There is an algorithm which for each ϵ-free context free grammar $G = \langle N, T, P, S\rangle$ with $L(G) \neq \{\}$ produces an equivalent ϵ-free context free grammar $G' = \langle N', T', P', S'\rangle$ such that for each $A \in N$ there is some $w \in T^*$ for which $A \Rightarrow^*_G w$.

Sketch of the procedure: Let
$W_1 = \{A \in N | A \rightarrow w \in P, w \in T^*\}$,
$W_{i+1} = \{A \in N | A \rightarrow x \in P$ with $x \in (W_i \cup T)^*\} \cup W_i$.

1. $W_i \subseteq W_{i+1}$ ($i \geq 1$),
2. If $W_i = W_{i+1}$ then $W_i = W_{i+m}$ for $m \geq 0$,
3. $W_n = W_{n+1}$, $n = |N|$,
4. $W_n = \{A \in N | A \Rightarrow^*_G w,$ with $w \in T^*\}$

Grammars w/o Useless Symbols

Let
$N' = W_n \cup \{S\}$,
$T' = T$,
$S' = S$,
$P' = \{A \rightarrow x \in P | A, x \in (T' \cup N')^*\}$

Grammars w/o Non-reachable Symbols

Lemma 4

There is an algorithm which for each context free grammar $G = \langle N, T, P, S \rangle$ produces an equivalent context free grammar $G' = \langle N', T', P', S' \rangle$ such that for each $A \in N' \cup T'$ there are $x, y \in (N' \cup T')^*$ for which $S \Rightarrow_G^* xAy$. If G is ϵ-free, has no unit rules, and for each $A \in N$ there is a $w \in T^*$ s.t. $A \Rightarrow_G^* w$, then the same properties hold of G'.

Grammars w/o Non-reachable Symbols

Sketch of the procedure: Let
$W_1(S) = \{S\}$,
$W_{i+1}(S) = \{B \in (N \cup T) |$ there are $x, y \in (N \cup T)^*$,
$A \in W_i(S)$ such that $A \rightarrow xBy \in P\} \cup W_i(S)$.

1. $W_i(S) \subseteq W_{i+1}(S)$ $(i \geq 1)$,
2. If $W_i(S) = W_{i+1}(S)$ then $W_i(S) = W_{i+m}(S)$ for $m \geq 0$,
3. $W_n(S) = W_{n+1}(S)$, $n = |N|$,
4. $W_n(S) = \{A \in (N \cup T) | S \Rightarrow^*_G xAy$ for some $x, y \in (N \cup T)^*\}$

Grammars w/o Non-reachable Symbols

Let
$N' = N \cap W_n(S)$,
$T' = T \cap W_n(S)$,
$S' = S$,
$P' = \{A \to x \in P | A \in N', x \in (N' \cup T')^*\}$