

Computational Linguistics II: Parsing

Formal Languages: Context Free Languages III

Frank Richter & Jan-Philipp Söhn

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

November 13th, 2006

Once Again: The Big Picture

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

DFA: Deterministic finite state automaton

(D)PDA: (Deterministic) Pushdown automaton

CFG: Context-free grammar

CSG: Context-sensitive grammar

LBA: Linear bounded automaton

The Chomsky Normal Form: CFGs in CNF

Two types of rules allowed:

$$A \rightarrow BC$$

$$A \rightarrow a$$

How to do a transformation:

- 1 Add new variable for each terminal symbol.
- 2 Recursively add one new variable for each pair of variables $B_j B_k$ in $A \rightarrow B_i B_j B_k \dots$

The Chomsky Normal Form: CFGs in CNF

Transform the following Grammar into CNF:

$$S \rightarrow AB$$

$$A \rightarrow aAb$$

$$A \rightarrow ab$$

$$B \rightarrow cB$$

$$B \rightarrow c$$

The Greibach Normal Form: CFGs in GNF I

One type of rules allowed:

$$A \rightarrow aB_1B_2\dots B_k \quad (k \geq 0)$$

The Greibach Normal Form: CFGs in GNF I

One type of rules allowed:

$$A \rightarrow aB_1B_2\dots B_k \quad (k \geq 0)$$

The Greibach Normal Form: CFGs in GNF II

We transform the following grammar into GNF:

$$S \rightarrow AB$$

$$A \rightarrow BS \mid b$$

$$B \rightarrow SA \mid a$$

The Greibach Normal Form: CFGs in GNF III

Step 1: Rename it!

$$A_1 \rightarrow A_2A_3$$

$$A_2 \rightarrow A_3A_1 \mid b$$

$$A_3 \rightarrow A_1A_2 \mid a$$

The Greibach Normal Form: CFGs in GNF IV

Step 2: $A_i \rightarrow A_j \alpha \mid i < j$

$A_1 \rightarrow A_2 A_3$

$A_2 \rightarrow A_3 A_1 \mid b$

$A_3 \rightarrow A_1 A_2 \mid a$

$A_3 \rightarrow A_2 A_3 A_2 \mid a$

The Greibach Normal Form: CFGs in GNF IV

Step 2: $A_i \rightarrow A_j \alpha \mid i < j$

$$A_1 \rightarrow A_2 A_3$$

$$A_2 \rightarrow A_3 A_1 \mid b$$

$$A_3 \rightarrow A_1 A_2 \mid a$$

$$A_3 \rightarrow A_2 A_3 A_2 \mid a$$

$$A_3 \rightarrow A_3 A_1 A_3 A_2 \mid b A_3 A_2 \mid a$$

The Greibach Normal Form: CFGs in GNF IV

Step 2: $A_i \rightarrow A_j\alpha \mid i < j$

$A_1 \rightarrow A_2A_3$

$A_2 \rightarrow A_3A_1 \mid b$

$A_3 \rightarrow A_1A_2 \mid a$

$A_3 \rightarrow A_2A_3A_2 \mid a$

$A_3 \rightarrow A_3A_1A_3A_2 \mid bA_3A_2 \mid a$

$A_3 \rightarrow bA_3A_2B_3 \mid aB_3 \mid bA_3A_2 \mid a$

$B_3 \rightarrow A_1A_3A_2 \mid A_1A_3A_2B_3$

The Greibach Normal Form: CFGs in GNF V

Step 3: Replace $A_i \rightarrow A_j\alpha$ with $A_i \rightarrow \beta\alpha$ for each $A_j \rightarrow \beta$

$$A_1 \rightarrow A_2A_3$$

$$A_2 \rightarrow A_3A_1 \mid b$$

$$A_3 \rightarrow bA_3A_2B_3 \mid aB_3 \mid bA_3A_2 \mid a$$

$$B_3 \rightarrow A_1A_3A_2 \mid A_1A_3A_2B_3$$

The Greibach Normal Form: CFGs in GNF V

Step 3: Replace $A_i \rightarrow A_j\alpha$ with $A_i \rightarrow \beta\alpha$ for each $A_j \rightarrow \beta$

$$A_1 \rightarrow A_2A_3$$

$$A_2 \rightarrow bA_3A_2B_3A_1 \mid aB_3A_1 \mid bA_3A_2A_1 \mid aA_1 \mid b$$

$$A_3 \rightarrow bA_3A_2B_3 \mid aB_3 \mid bA_3A_2 \mid a$$

$$B_3 \rightarrow A_1A_3A_2 \mid A_1A_3A_2B_3$$

The Greibach Normal Form: CFGs in GNF V

Step 3: Replace $A_i \rightarrow A_j\alpha$ with $A_i \rightarrow \beta\alpha$ for each $A_j \rightarrow \beta$

$A_1 \rightarrow bA_3A_2B_3A_1A_3 \mid aB_3A_1A_3 \mid bA_3A_2A_1A_3 \mid aA_1A_3 \mid bA_3$

$A_2 \rightarrow bA_3A_2B_3A_1 \mid aB_3A_1 \mid bA_3A_2A_1 \mid aA_1 \mid b$

$A_3 \rightarrow bA_3A_2B_3 \mid aB_3 \mid bA_3A_2 \mid a$

$B_3 \rightarrow A_1A_3A_2 \mid A_1A_3A_2B_3$

The Greibach Normal Form: CFGs in GNF V

Step 3: Replace $A_i \rightarrow A_j\alpha$ with $A_i \rightarrow \beta\alpha$ for each $A_j \rightarrow \beta$

$A_1 \rightarrow bA_3A_2B_3A_1A_3 \mid aB_3A_1A_3 \mid bA_3A_2A_1A_3 \mid aA_1A_3 \mid bA_3$

$A_2 \rightarrow bA_3A_2B_3A_1 \mid aB_3A_1 \mid bA_3A_2A_1 \mid aA_1 \mid b$

$A_3 \rightarrow bA_3A_2B_3 \mid aB_3 \mid bA_3A_2 \mid a$

$B_3 \rightarrow bA_3A_2B_3A_1A_3A_3A_2 \mid aB_3A_1A_3A_3A_2 \mid bA_3A_2A_1A_3A_3A_2 \mid$
 $aA_1A_3A_3A_2 \mid bA_3A_3A_2 \mid bA_3A_2B_3A_1A_3A_3A_2B_3 \mid$
 $aB_3A_1A_3A_3A_2B_3 \mid bA_3A_2A_1A_3A_3A_2B_3 \mid aA_1A_3A_3A_2B_3 \mid$
 $bA_3A_3A_2B_3$