NORMAL FORMS OF CONTEXT FREE GRAMMAR

Dr. A. Beulah

AP/CSE

$NT \rightarrow \underline{(N \cup T)^*}$ CFG

LEARNING OBJECTIVE

- To Understand the need of formal languages, and grammars (K3)
  - To Understand normal forms of CFG

NORMAL FORMS OF CONTEXT FREE GRAMMAR

- Chomsky normal form
- Greibach normal form

CHOMSKY NORMAL FORM

- The Chomsky normal form places restrictions on the *length* and the *composition* of the **right-hand side** of a rule

- A context free grammar G is in CNF if every production is of the form A→*a or A→BC and* S→ ε is in G if ε ∈L(G). When ε is in L(G) we assume that S does not appear on the R.H.S. of any other production.
  - where *B*, *C* ∈ NT

LHS → RHS

NT

CFG

(NUT)*

C.

A → a

A → BC

S → ε

S ↛ RHS

EXAMPLE

Ex:

$S \rightarrow AB$ CNF

$S \rightarrow \varepsilon$ SxRHS

$A \rightarrow \underline{a}$ T

$B \rightarrow \underline{b}$ T

CHOMSKY NORMAL FORM

- **Algorithm Step 1**

- Make sure that the following are satisfied:
  - No ε-productions (other than $S \rightarrow \varepsilon$)
  - No unit productions
  - No useless symbols

CHOMSKY NORMAL FORM

- **<u>Algorithm Step 2</u>**
- Eliminate terminals from RHS of productions
  - For each production $A \rightarrow X_1X_2\ldots X_m$
    - where $X_i \in NT \cup T$
  - If $m > 1$, replace each **terminal** $a \in$ RHS of A
    - Add (if needed) $C_a \rightarrow a$ for each $a \in T$, where each C_a is new non-terminal.
    - In production A, replace terminal a with corresponding C_a

CHOMSKY NORMAL FORM

- **Algorithm Step 3**

- Eliminate productions with long RHS:
  - For each production:
    - $A \rightarrow B_1B_2 \ldots B_m$, $m > 2$, where $B_i \in NT$
  - replace with productions
    - $A \rightarrow B_1D_1$
    - $D_1 \rightarrow B_2D_2$
    - … $D_2 \rightarrow B_3 D_3$ …
    - $D_{m-2} \rightarrow B_{m-1}B_m$
  - where $D_1 \ldots D_{m-2}$ are new non-terminals.

EXAMPLE

- $S \rightarrow aAbB$
- $A \rightarrow aA \mid a$
- $B \rightarrow bB \mid b$

(i) ε-pr ×
Unit ×
Useless ×

(ii) $\boxed{A \rightarrow a \quad B \rightarrow b}$ CNF

$A \rightarrow aA$
$B \rightarrow bB$
$S \rightarrow aAbB$

$A \rightarrow a$
$A \rightarrow BC$
$S \rightarrow \varepsilon$
SX RHS

$\boxed{C_a \rightarrow a \quad C_b \rightarrow b}$ ✓ CNF

$S \rightarrow C_a A C_b B$ ←

$\boxed{A \rightarrow C_a A \quad B \rightarrow C_b B}$ ✓ CNF

(iii) $S \rightarrow C_a A C_b B$ CNF

$S \rightarrow C_a D_1$
$D_1 \rightarrow A D_2$
$D_2 \rightarrow C_b B$

2 NT
$S \rightarrow C_a A$
$S \rightarrow C_b B$ ×

EXAMPLE

$S \rightarrow C_a D_1$

$D_1 \rightarrow A D_2$

$D_2 \rightarrow C_b B$

$A \rightarrow C_a A \;/ a$

$B \rightarrow C_b B \;/ b$ CNF

$C_a \rightarrow a$

$C_b \rightarrow b$

$A \rightarrow a$

$\cancel{B \rightarrow \epsilon}$

EXAMPLE

1. Original grammar (no chain rules, useless symbols, or ε-productions):

   $S \rightarrow X\,a\,Y \mid Y\,b$

   $X \rightarrow Y\,X\,a\,Y \mid a$

   $Y \rightarrow S\,S \mid a\,X \mid b$

$A \rightarrow aB \mid B$

$B \rightarrow C$

$C \rightarrow D$

$D \rightarrow xyz$

$A \rightarrow xyz$

$B \rightarrow xyz$

$C \rightarrow xyz$

$D \rightarrow xyz$

EXAMPLE

1. Original grammar (no chain rules, useless symbols, or ε-productions):

   $S \to X\, a\, Y \mid Y\, b$

   $X \to Y\, X\, a\, Y \mid a$

   $Y \to S\, S \mid a\, X \mid b$

2. Grammar after eliminating terminals from RHSs:

   $S \to X\, C_a\, Y \mid Y\, C_b$ $\qquad C_a \to a$

   $X \to Y\, X\, C_a\, Y \mid a$ $\qquad C_b \to b$

   $Y \to S\, S \mid C_a\, X \mid b$

EXAMPLE

1. Original grammar (no chain rules, useless symbols, or ε-productions):

$S \to X\, a\, Y \mid Y\, b$

$X \to Y\, X\, a\, Y \mid a$

$Y \to S\, S \mid a\, X \mid b$

2. Grammar after eliminating terminals from RHSs:

$S \to X\, C_a\, Y \mid Y\, C_b$ $\quad C_a \to a$

$X \to Y\, X\, C_a\, Y \mid a$ $\quad C_b \to b$

$Y \to S\, S \mid C_a\, X \mid b$

3. Grammar after eliminating long RHSs:

$S \to X\, D_1 \mid Y\, C_b$ $\quad D_1 \to C_a\, Y$ $\quad C_a \to a$

$X \to Y D_2 \mid a$ $\quad D_2 \to X\, D_3$ $\quad C_b \to b$

$Y \to S\, S \mid C_a\, X \mid b$ $\quad D_3 \to C_a\, Y$

EXAMPLE

- $S \rightarrow abAB$
- $A \rightarrow bAa \mid \varepsilon$
- $B \rightarrow BAa \mid A \mid \varepsilon$

EXAMPLE

- $S \rightarrow aSa \mid bSb \mid c$

EXAMPLE

Original grammar (no chain rules, useless symbols, or ε-productions):

$S \to aXYZ \mid a$ $\quad X \to aX \mid a$

$Y \to bcY \mid bc$ $\quad Z \to cZ \mid c$

1. Grammar after eliminating terminals from RHSs:

$S \to AXYZ \mid a$ $\quad A \to a$

$X \to AX \mid a$ $\quad B \to b$

$Y \to BCY \mid BC$ $\quad C \to c$

$Z \to CZ \mid c$

2. Grammar after eliminating long RHSs:

$S \to AF \mid a$ $\quad A \to a$ $\quad F \to XG$

$X \to AX \mid a$ $\quad B \to b$ $\quad G \to YZ$

$Y \to BH \mid BC$ $\quad C \to c$ $\quad H \to CY$

$Z \to CZ \mid c$

GREIBACH NORMAL FORM

- A context-free grammar is in **Greibach Normal Form** if every production is of the form $A \rightarrow aX$
  - where $A \in$ NT, $X \in \text{NT}^*$, and $a \in \Sigma$

- Examples:
  - $G_1 = (\{S, A\}, \{a, b\}, S, \{S \rightarrow aSA \mid a, A \rightarrow aA \mid b\})$
    - GNF
  - $G_2 = (\{S, A\}, \{a, b\}, S, \{S \rightarrow AS \mid AAS, A \rightarrow SA \mid aa\})$
    - not GNF
- This grammar $S \rightarrow AB \; A \rightarrow aA \mid bB \mid b \quad B \rightarrow b$

  is not in GNF
- This grammar $S \rightarrow aAB \mid bBB \mid bB$

  $A \rightarrow aA \mid bB \mid bB \rightarrow b$

  is in GNF

SUMMARY

- Different normal forms of Context free grammar
  – Chomsky normal form
  – Greibach normal form

TEST YOUR KNOWLEDGE

- The Grammar can be defined as: G=(V, ∑, p, S)
  In the given definition, what does S represents?
  a) Accepting State
  b) Starting Variable
  c) Sensitive Grammar
  d) None of these

LEARNING OUTCOME

On successful completion of this topic, the student will be able to:

- Understand simplification of CFG (K3)

REFERENCE

- Hopcroft J.E., Motwani R. and Ullman J.D, “Introduction to Automata Theory, Languages and Computations”, Second Edition, Pearson Education, 2008