UNIT-III

Context Free Grammar

Simplification of C.F.G

→ In CFG, some times all the production rules and symbols are not needed for the derivation of strings.

→ Besides this, there may also be some NULL productions and UNIT productions.

→ Elimination of these productions and symbols is called simplification of CFG.

① Reduction of CFG (Elimination of useless symbols)

② Removal of Unit Productions

③ Removal of NULL Productions

Reduction of CFG

Phase 1: Derivation of equivalent Grammar G', from the CFG, G, such that each variable derives some terminal string.

step 1: Include all symbols W_1, that derives some terminal and initialize $i=1$.

step 2: Include symbols W_{i+1}, that derives W_i

Step 3: Increment i and repeat step 2, until $W_{i+1} = W_i$

Step 4: Include all production rules that have W_i in it.

Phase 2:- Derivation of an equivalent Grammar G'', from the CFG, G', such that each symbol appears in sentential form.

Step 1: Include start symbol in Y_1 and initialize $i=1$

Step 2: Include all symbols Y_{i+1}, that can be derived from Y_i and include all production rules that have been applied.

Step 3: Increment i and repeat step 2, until $Y_{i+1} = Y_i$

Eg: Find a reduced Grammar equivalent to Grammar 'G', having production rules

$P: S \to AC \mid B,\ A \to a,\ C \to c \mid BC,\ E \to aA \mid e$

Phase 1: $W_1 = \{A, C, E\}$

$W_2 = \{S, AC, E\}$

$W_3 = \{S, A, C, E\}$

$G' = \{(S, A, C, E), (a, c, e), P, S\}$

$P: S \to AC,\ A \to a,\ C \to c,\ E \to aA \mid e$

Phase 2: $Y_1 = \{S\}$

$Y_2 = \{S, A, C\}$

$Y_3 = \{S, A, C\}$

$G'' = \{ \{A, C, S\}, \{a, c\}, P, \{S\} \}$

$P: S \to AC, A \to a, C \to c$

Removal of unit productions

Any production Rule of the form $A \to B$ where $A, B \in$ Non Terminals is called unit Production.

Procedure for Removal

step 1: To remove $A \to B$, add Production $A \to x$ to the grammar rule whenever $B \to x$ occurs in the grammar [$x \in$ Terminal, x can be NULL]

step 2: Delete $A \to B$ from grammar

step 3: Repeat step-1 until all unit productions are removed.

Eg: $P: S \to XY, X \to a, Y \to Z|b, Z \to M, M \to N, N \to a$

$Y \to Z, Z \to M, M \to N$

1) $N \to a$, we can add $M \to a$

$P: S \to XY,\ X \to a,\ Y \to Z|b,\ Z \to M,\ M \to a,\ N \to a$

2) Since $M \to a$ then we add $Z \to a$ [$Z \to M$ Removed]

$P: S \to XY,\ X \to a,\ Y \to Z|b,\ Z \to a,\ M \to a,\ N \to a$

3) since $Z \to a$ Then we add $Y \to a$ [remove $Y \to Z$]

$P: S \to XY,\ X \to a,\ Y \to a|b,\ Z \to a,\ M \to a,\ N \to a$

Remove unreachable symbols

$P: S \to XY,\ X \to a,\ Y \to a|b$

Simplification of C.F.G

Removal of NULL productions

→ In a C.F.G a Non-Terminal symbol 'A' is a nullable variable if there is a production $A \to \epsilon$ or there is a derivation that starts at 'A' and leads to ϵ (Like $A \to \cdots \epsilon$)

Procedure for Removal

step 1: To remove $A \to \epsilon$, look for all productions whose right side contains A

step 2: Replace each occurence of 'A' in each of these productions with ϵ

Step 3: Add the resultant productions to the Grammar.

Eg: $S \rightarrow ABAC$, $A \rightarrow aA \mid \epsilon$, $B \rightarrow bB \mid \epsilon$, $C \rightarrow c$

$A \rightarrow \epsilon$ & $B \rightarrow \epsilon$ must be eliminated.

1) $A \rightarrow \epsilon$

$S \rightarrow ABAC \mid BAC \mid ABC \mid BC$

$A \rightarrow a \mid aA$

$B \rightarrow bB \mid \epsilon$

$C \rightarrow c$

(2) $B \rightarrow \epsilon$ must be eliminated

$S \rightarrow ABAC \mid BAC \mid ABC \mid BC \mid AC \mid C \mid AAC$

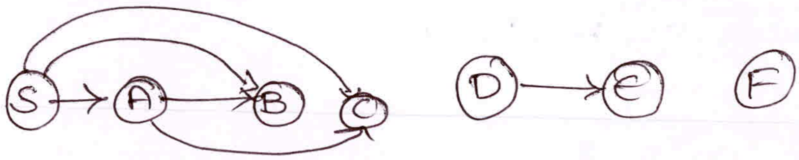
$A \rightarrow a \mid aA$

$B \rightarrow b \mid bB$

$C \rightarrow c$

Examples on Eliminating useless symbols :-

1) $S \rightarrow ABa \mid BC$, $A \rightarrow ac \mid BCC$, $C \rightarrow a$, $B \rightarrow bcc$, $D \rightarrow E$, $E \rightarrow d$, $F \rightarrow e$



- DEF are non-reachable states from S.

- $S \rightarrow ABa \mid BC$

  $A \rightarrow ac \mid BCC$

  $C \rightarrow a$

  $B \rightarrow bcc$

2) $S \rightarrow BC \mid AB \mid CA$ $\quad A \rightarrow a$ $\quad C \rightarrow aB \mid b$

- B is not defined.

  $S \rightarrow CA$

  $A \rightarrow a$

  $C \rightarrow b$

3) $S \rightarrow aAa$ $\quad A \rightarrow bBB$ $\quad B \rightarrow ab$ $\quad C \rightarrow ab$

- C is useless & not reachable from S.

  $S \rightarrow aAa$

  $A \rightarrow bBB$

  $B \rightarrow ab$.

4) $S \rightarrow aS \mid A \mid BC$ $\quad A \rightarrow a$ $\quad B \rightarrow aa$ $\quad C \rightarrow aCb$

- C is useless as it is not deriving any string.

  $S \rightarrow aS \mid A$

  $A \rightarrow a$

  $B \rightarrow aa$

B is not reachable. $\quad S \rightarrow aS \mid A$

$A \rightarrow a$

Elimination of ε-Production:-

1) $S \to ABaC \quad A \to BC \quad B \to b \mid \varepsilon \quad C \to D \mid \varepsilon \quad D \to d.$

$S \to ABaC \mid BaC \mid AaC \mid ABa \mid aC \mid Aa \mid Ba \mid a$

$A \to BC \mid C \mid B$

$B \to b$

$C \to D$

$D \to d$

2) $S \to aA \quad A \to BB \quad B \to aBb \mid \varepsilon$

$S \to aA \mid a$

$A \to BB \mid B$

$B \to aBb \mid ab$

3) $S \to ABC \quad A \to aB \mid B \quad B \to bC \mid C \quad C \to cC \mid \epsilon$

$C \to cC \mid c$

$B \to bC \mid C \mid b$

$A \to aB \mid B \mid a$

$S \to ABC \mid BC \mid AC \mid AB \mid A \mid B \mid C$

4) $S \to ABAC, \quad A \to aA \mid \varepsilon, \quad B \to bB \mid \varepsilon, \quad C \to c$

$C \to c$

$B \to bB \mid b$

$A \to aA \mid a$

$S \to ABAC \mid BAC \mid AAC \mid ABC \mid AC \mid AB \mid BC \mid C$

Elimination of unit Productions :-

1) $S \to A \mid bb \quad A \to B \mid b \quad B \to S \mid a$

unit productions are $S \to A$, $A \to B$, $B \to S$ - Eliminate

$S \to bb \mid a \mid b$

$A \to a \mid b \mid bb$.

$B \to a \mid bb \mid b$

2) $S \to Aa \mid B \quad B \to A \mid bb \quad A \to a \mid bc \mid B$.

$S \to Aa \mid bb \mid a \mid bc$

$B \to bb \mid a \mid bc$

$A \to a \mid bc \mid bb \mid a \mid bc \Rightarrow a \mid bc \mid bb$.

~~3) $S \to aA \mid aBB \quad A \to aAA \mid \epsilon \quad B \to bB \mid bbC \quad C \to B$~~

~~$S \to aA \mid aBB$~~

~~$A \to$~~

3) $S \to 0A \mid 1B \mid C \quad A \to 0S \mid 00 \quad B \to 1 \mid A \quad C \to 01$

$S \to 0A \mid 1B \mid 01$

$A \to 0S \mid 00$

$B \to 1 \mid 0S \mid 00$

$C \to 01$

4) $S \to AB \quad A \to a \quad B \to C \mid b \quad C \to D \quad D \to E \quad E \to a$

$S \to AB$ $\qquad C \to a$

$A \to a$ $\qquad D \to a$

$B \to a \mid b$ $\qquad E \to a$

CNF (Chomsky Normal Form)

→ In CNF, we have a restriction on the length of R.H.S, which is elements in RHS should either be two variables or a Terminal.

→ A CFG is in CNF if productions are in the following forms:

$$\boxed{\begin{array}{l} A \rightarrow a, \\ A \rightarrow BC \end{array}}$$

where A, B, C are non-terminals and 'a' is terminal.

steps to convert to CNF:- (CFG to CNF)

step 1: If start symbol 'S' occurs on some right side, create a new symbol S' and add a new production $S' \rightarrow S$

step 2: Remove Null productions

step 3: Remove unit productions

step 4: Replace each production $A \rightarrow B_1 \dots B_n$ where $n > 2$, with $A \rightarrow B_1 C$ and $C \rightarrow B_2 \dots B_n$

Repeat this step for all productions having two (or) more symbols on the right side.

step 5: If the right side of any production is in the form $A \to aB$ where 'a' is terminal and A and B' Non terminals, then the production is replaced by $A \to XB$ and $X \to a$

Repeat this step for every production which is of form $A \to aB$.

Eg: Convert CFG to CNF:

~~(S, S, A)~~

$P: S \to ASA \mid aB$

$A \to B \mid S$

$B \to b \mid \epsilon$

1) $P: S' \to S, \ S \to ASA \mid aB, \ A \to B \mid S, \ B \to b \mid \epsilon$

2) Remove Null productions $B \to \epsilon$

$P: S' \to S, \ S \to ASA \mid aB \mid a, \ A \to \epsilon \mid S, \ B \to b$

3) Remove NULL production $A \to \epsilon$

$P: S' \to S, \ S \to ASA \mid aB \mid a \mid SA \mid AS \mid S, \ A \to B \mid S, \ B \to b$

4) Removal of unit productions $(A \to S, \ S' \to S, \ S \to S)$

$(S \to S)$ same productions

$P: S' \to S, \ S \to ASA \mid aB \mid a \mid SA \mid AS, \ A \to B \mid S, \ B \to b$

$(S' \to S)$ (Remove)

$P: S' \to ASA \mid SA \mid AS \mid a \mid aB$

$S \to ASA \mid SA \mid AS \mid a \mid aB$

$A \to B \mid S$

$B \to b$

$A \to S$ (Remove)

$P: S' \to ASA \mid SA \mid AS \mid a \mid aB$

$S \to ASA \mid SA \mid AS \mid a \mid aB$

$A \to ASA \mid SA \mid AS \mid a \mid aB \mid b$

$B \to b$

5) $P: S' \to AX \mid SA \mid AS \mid a \mid aB$

$S \to AX \mid SA \mid AS \mid a \mid aB$

$A \to AX \mid SA \mid AS \mid a \mid aB \mid b$

$X \to SA$

$B \to b$

(CFG. CNF)

6) $P:$

$S' \to AX \mid SA \mid AS \mid a \mid YB$

$S \to AX \mid SA \mid AS \mid a \mid YB$

$A \to AX \mid SA \mid AS \mid a \mid YB$

$X \to SA$

$Y \to a$

$B \to b.$

MREC Exam Cell

1 Unit productions

$$\boxed{\begin{aligned} &S \rightarrow A\ A \\ &A \rightarrow B \mid BB \\ &B \rightarrow abB \mid b \mid bb \end{aligned}}$$

$A \rightarrow B$ is the unit production.

$$\boxed{\begin{aligned} &S \rightarrow A\ A \\ &A \rightarrow BB \mid abB \mid b \mid bb \end{aligned}}$$

2 (CNF) (CFG)

$S \rightarrow ABA \mid AB \mid BA \mid AA \mid B$

$A \rightarrow aA \mid a$

$B \rightarrow bB \mid b$

3 Derivation Tree

$S \rightarrow aB \mid bA$, $A \rightarrow a \mid aS \mid bAA$, $B \rightarrow b \mid bS \mid aBB$

R.M.D, L.M.D for "aaabbabbba"

Conversion of CFG TO CNF :-

1) $S \to ASB \mid \varepsilon$ $\quad A \to aAS \mid a$ $\quad B \to SbS \mid A \mid bb.$

Eliminate ε-production :-

$S \to ASB \mid AB$ $\quad A \to aAS \mid a \mid aA$ $\quad B \to SbS \mid A \mid bb \mid Sb \mid bS \mid b$

Eliminate unit productions :-

Eliminate $B \to A$

$S \to ASB \mid AB$ $\quad A \to aAS \mid a \mid aA$ $\quad B \to SbS \mid bb \mid Sb \mid bS \mid b \mid aAS \mid a \mid aA$

Conversion :- $P' \Rightarrow$

$S \to AB$ $\quad A \to a$ $\quad B \to b \mid a$ are in CNF

- Convert the remaining Productions to CNF.
- Eliminate terminals in RHS.

$A \to C_a A S$

$A \to C_a A$

$B \to C_b C_b$

$B \to S C_b$

$B \to C_b S$

$B \to C_a A S$

$B \to C_a A$, $C_a \to a$, $C_b \to b$

- Add productions that are in CNF to P' :-

$S \to AB$, $A \to a \mid C_a A$, $B \to b \mid a \mid C_b C_b \mid S C_b \mid C_b S \mid C_a A$,

$C_a \to a$, $C_b \to b$

- Reduce the RHS of the remaining productions with more than 2 variables to the required CNF.

$S \to ASB \Rightarrow AC_4$ & $C_4 \to SB$.

$A \to C_aAS \Rightarrow A \to C_aC_1$ & $C_1 \to AS$

$B \to SC_bS \Rightarrow B \to SC_2$ & $C_2 \to C_bS$

$B \to C_aAS \Rightarrow B \to C_aC_3$ & $C_3 \to AS$.

Adding these productions to P' the complete grammar is

$S \to AB \mid AC_4$

$A \to a \mid C_aA \mid C_aC_1$

$B \to C_bC_b \mid C_bS \mid SC_b \mid C_aA \mid b \mid a \mid SC_2 \mid C_aC_3$

$C_a \to a$, $C_b \to b$, $C_1 \to AS$, $C_2 \to C_bS$,

$C_4 \to SB$

2) $S \to bA \mid aB \quad A \to bAA \mid aS \mid a \quad B \to aBB \mid bS \mid b$.

There is no null productions nor unit productions.

P' :- $A \to a$ $B \to b$ are in CNF.

Replace terminals by variables.

$S \to C_bA \mid C_aB$

$A \to C_bAA \mid C_aS$

$B \to C_aBB \mid C_bS$, $C_a \to a$, $C_b \to b$

Add Productions which are in CNF to P'.

$S \to C_bA \mid C_aB$

$A \to C_aS \mid a$

$B \to C_bS \mid b$

$C_a \to a$, $C_b \to b$

$S \rightarrow C_b A \mid C_a B$

$A \rightarrow C_b C_1 \mid C_a S$

$B \rightarrow C_a C_2 \mid C_b S$

$C_a \rightarrow a$

$C_b \rightarrow b$

$C_1 \rightarrow AA$

$C_2 \rightarrow BB$

3) $S \rightarrow AB \mid aB \quad A \rightarrow aab \mid \varepsilon \quad B \rightarrow bbA$

i) Eliminate ε-production :-

$S \rightarrow AB \mid aB \mid B \quad A \rightarrow aab \quad B \rightarrow bbA \mid bb$

ii) Eliminate unit Production

$S \rightarrow B$ need to remove

$S \rightarrow AB \mid aB \mid bbA \mid bb \quad A \rightarrow aab \quad B \rightarrow bbA \mid bb$

iii) Eliminate terminals by replacing it with non-terminals

$S \rightarrow C_a B \mid C_b C_b A \mid C_b C_b \quad A \rightarrow C_a C_a C_b \quad B \rightarrow C_b C_b A \mid C_b C_b$

$C_a \rightarrow a \quad C_b \rightarrow b$

P' :- $S \rightarrow AB \mid C_a B \mid C_b C_b \quad B \rightarrow C_b C_b \quad C_a \rightarrow a \quad C_b \rightarrow b$

iv) Restrict RHS with 2 Non-terminals

$S \rightarrow C_b C_1 \quad A \rightarrow C_a C_2 \quad B \rightarrow C_b C_1$.

P' :- $S \rightarrow AB \mid C_a B \mid C_b C_b \mid C_b C_1$

$A \rightarrow C_a C_2$, $B \rightarrow C_b C_1$, $C_a \rightarrow a$, $C_b \rightarrow b$

$C_1 \rightarrow C_b A$, $C_2 \rightarrow C_a C_b$,

Greibach Normal Form

→ A CFG is in Greibach Normal Form if the productions are in the following forms:

$A \to b$

$A \to b C_1 C_2 \ldots C_n$

where $A, C_1, \ldots C_n$ are Non-Terminals and b is a Terminal.

Steps to convert a given CFG to GNF:

Step 1: check if the given CFG has any unit productions (or) NULL productions and remove if there are any

Step 2: check whether the CFG is already in CNF and convert it to CNF if it is not.

Step 3: Change the names of the Non-Terminal symbols into some A_i in ascending order of i.

Eg: $S \to CA \mid BB$

$B \to b \mid SB$

$C \to b$

$A \to a$

Examples on Greibach Normal Form :-

$S \rightarrow AA \mid a \qquad A \rightarrow SS \mid b$

Rename variables :- $S = A_1 \quad A = A_2$

$A_1 \rightarrow A_2A_2 \mid a$ — (1) $\quad A_2 \rightarrow A_1A_1 \mid b$ — (2)

Consider eq (2)

$A_2 \rightarrow A_1A_1 \mid b$

$i = 2 \qquad j = 1$

$j < i \rightarrow$ so go for substitution

Substitute A_1 in A_2

$A_2 \rightarrow A_2A_2A_1 \mid aA_1 \mid b$

$i = 2 \qquad j = 2$

$i = j \rightarrow$ so go for left Recursion.

$Z \rightarrow A_2A_1Z \mid \varepsilon$

$A_2 \rightarrow \cancel{aA_1 b} \; aA_1Z \mid bZ$.

Left Recursion :- $A \rightarrow A\alpha \mid \beta$

$Z \rightarrow \alpha Z \mid \varepsilon$

$A \rightarrow \beta Z$

Eliminate ε-production :-

$Z \rightarrow A_2A_1Z \mid A_2A_1$

$A_2 \rightarrow aA_1Z \mid bZ \mid aA_1 \mid b$

Replace A_2 in Z

$Z \rightarrow aA_1ZA_1Z \mid bZA_1Z \mid aA_1A_1Z \mid bA_1Z \mid aA_1A_1Z \mid bZA_1 \mid aA_1A_1 \mid bA_1$

Replace A_2 in A_1

$A_1 \rightarrow aA_1ZA_2 \mid bZA_2 \mid aA_1A_2 \mid bA_2 \mid a$