Unit - III

1. Define Context free Grammar. State and Explain the closure properties of CFG.

→ CFG stands for context free Grammar, it is a formal grammar which is used to generate all possible patterns of strings in a given formal language.

Definition: In order to define CFG by using 4 tuples. They are $G = (V, T, P, S)$

Where V = Set of variables or Non-terminals which are represented by UPPER-CASE letters.

T = Set of Terminals (or) Symbols (or) digits is represented by LOWER-CASE letters.

P = Set of production rules (LHS → RHS)

Ex : $A \rightarrow \alpha$

$A \rightarrow id$

S = Start-symbol where

$A \in V$

$\alpha \in (V \cup T)^*$

Closure properties of CFG: CFL are closed under union, concatenation, kleen closure and not closed under Intersection and complementation.

1. **Union:** Let L_1 and L_2 be two CFL's $L_1 \cup L_2$ is also a CFL

eg1: $L_1 = \{a^n / n \geq 0\}$, $L_2 = \{b^n / n \geq 0\}$

$L_1 = \{\epsilon, a, aa \dots\}$ $L_2 = \{\epsilon, b, bb, \dots\}$

$L_1 \cup L_2 = \{\epsilon, a, b, aa, bb \dots\}$

eg2: $S_1 \rightarrow aS_1 | \epsilon$

$S_2 \rightarrow aS_2 | \epsilon$

$S = S_1 \cup S_2 \rightarrow aS_1 | aS_2 | \epsilon$

2. <u>Concatenation</u>: Let L_1 and L_2 be two CFL's then L_1 <u>Re</u>
is also a CFL

Eg1: $L_1 = \{a^n \mid n \geq 0\}$, $L_2 \{b^n / n \geq 0\}$

$L_1 = \{\epsilon, a, aa, \dots\}$ $L_2 = \{\epsilon, b, bb, \dots\}$

$L_1 \cdot L_2 = \{\epsilon, ab, aabb, \dots\}$

Eg2: $S_1 \rightarrow aS_1 / \epsilon$

$S_2 \rightarrow aS_2 / \epsilon$

$S = S_1 \cdot S_2 \rightarrow aS_1 \cdot bS_2 / \epsilon$.

3. <u>Kleen Closure</u>: Let L is a CFL then L^* is also a CFL

eg: $L = \{a\}$

$L^* = \{\epsilon, a, aaa, \dots\}$

2. Discus various steps in ~~signification~~ Simplification of context free grammar.

<u>Simplification of CFG</u>: A CFG can be simplified in three ways

1. Removal of unit productions
2. Removal of epsilon(ϵ) productions
3. Removal of useless symbols

1. <u>Removal of unit productions</u>: Unit production means having only one non terminal on left hand side & right hand side is known as unit production.

eg: $A \rightarrow B$

To eliminate unit production

$A \rightarrow B$

$B \rightarrow x_1, x_2, x_3 \dots x_n$

$A \rightarrow x_1, x_2, x_3 \dots x_n$

Removal of ε- productions: The removal of ε-productions in a context free grammar involves eliminating rules that can derive the empty string.

1. Identify epsilon productions: These are rules where the R.H.S can derive the empty string (epsilon).
2. Eliminate epsilon productions:

→ For each production $A \rightarrow \varepsilon$, remove it from the grammar

→ For each occurrence of A on the RHS of a production, add new productions without A and with A replaced by epsilon.

Removal of useless symbols:

1. Identify useless symbols: Start by finding all non-terminals that cannot derive any terminal string. These are considered useless symbols.
2. Eliminate useless productions:
   → Remove all productions involving the useless symbols.
   → Update the remaining productions by removing any occurrences of the useless symbols

Example Given CFG:

$S \rightarrow AB$

$A \rightarrow a \mid \varepsilon$

$B \rightarrow b \mid A$

Removal of unit Productions:

$S \rightarrow AB$

$A \rightarrow a \mid \varepsilon$

$B \rightarrow b \mid a \mid \varepsilon$

Removal of ε- productions:

$S \rightarrow AB \mid B$

$A \rightarrow a$

$B \rightarrow b \mid a$

Removal of useless - Symbols:

$S \rightarrow AB | B$

$A \rightarrow a$

$B \rightarrow b | a$

3. Design the CFG for the expressions $\{a^n b^n$ where $n>=1$

Given $\Sigma = \{a^n b^n$ where $n>=1\}$

$L = \{ab, aabb, aaabbb. \ldots \}$

Regular Expression $= a(a+b)^* b$

Production Rules (P):

$S \longrightarrow aA\ b$

$B \longrightarrow aA$

$A \longrightarrow bA$

$A \longrightarrow \epsilon$

input string : ab

$S \longrightarrow aA\ b$

$S \longrightarrow ab \quad (A \rightarrow \epsilon)$

C.F.G: $G = \{ \{S, A\}, \{a, b\}, S, P\}$

4. Simplify the following grammar with the following production

$S \rightarrow Aa / B | cA$

$B \rightarrow A / bb / \epsilon$

$A \rightarrow bc / B$

Given CFG $S \rightarrow Aa | B | cA$

$B \rightarrow A / bb | \epsilon$

$A \rightarrow bc / B$

Step-1: Removal of useless symbols

(i) Eliminate non-generating symbols

$V = \{S, A, B\}$

(i) S

$S \longrightarrow Aa$

$S \longrightarrow bca$

(ii) A

$A \longrightarrow bc$

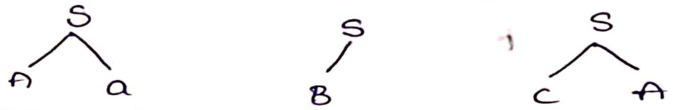
(iii) B

$B \rightarrow A$

$B \rightarrow bb$

2) Eliminate non-Recursive Symbols

$V = \{S, A, B\}$



Simplified CFG after step-1:

$S \longrightarrow Aa \mid B \mid cA$

$B \longrightarrow A \mid bb \mid \epsilon$.

$A \longrightarrow bc \mid B$

<u>Step-2:</u> Removal of ϵ-productions

$B \longrightarrow \epsilon \longrightarrow$ null production

$S \longrightarrow Aa \mid \epsilon \mid cA$

$B \longrightarrow A \mid bb$.

$A \longrightarrow bc \mid \epsilon$.

Simplified CFG after step-2

$S \longrightarrow Aa \mid cA$

$B \longrightarrow A \mid bb$

$A \longrightarrow bc$

<u>Step-3:</u> Removal of unit productions

$B \longrightarrow A$ $\quad S \rightarrow A$

$B \longrightarrow bc$ $\quad S \longrightarrow bc$

Simplified CFG after step 3:

$S \longrightarrow bca \mid cbc$

$B \longrightarrow bc \mid bb$

$A \longrightarrow bc$

5. Derive the left most and the right most derivation for the CFG $S \rightarrow aAS \mid a$, $A \rightarrow SbA \mid SS \mid ba$ with the input string aabbaa.

Left Most Derivation (LMD):

$S \rightarrow aAS$

$\rightarrow aSbAS$ ($A \rightarrow SbA$)

$\rightarrow aabAS$ ($S \rightarrow a$)

$\rightarrow aabbaS$ ($A \rightarrow ba$)

$\rightarrow aabbaa$ ($S \rightarrow a$)

Right Most Derivation:

$S \rightarrow aAS$

$\rightarrow aAa$ ($S \rightarrow a$)

$\rightarrow aSbAa$ ($A \rightarrow SbA$)

$\rightarrow aSbbaa$ ($A \rightarrow ba$)

$\rightarrow aabbaa$ ($S \rightarrow a$)

L.M.D: In each state the left most non-terminal is to be extended

R.M.D: In each state the right-most non terminal is to be extended.

6. Eliminate unit productions and epsilon production from the grammar

$S \rightarrow Aa \mid B$,
$B \rightarrow A \mid bb$,
$A \rightarrow a \mid bc \mid B$

Step-1: Removal of useless symbols

(I) Eliminating non-generating symbols

$V = \{S, A, B\}$

(i) S

$S \rightarrow Aa$
$\rightarrow aa$

(ii) A

$A \rightarrow a$

(iii) B

$B \rightarrow bb$

2) Eliminate non-Recursive Symbols

$V = \{S, A, B\}$

Simplified CFG after step 1:

$S \longrightarrow Aa \mid B$

$B \longrightarrow A \mid bb$

$A \longrightarrow a \mid bc \mid B$

Step-2 : Removal of null productions

∴ Here no null productions

Step-3 : Removal of unit productions

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

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

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

$S \to B$ $S \longrightarrow bb \mid A$

$\longrightarrow bb \mid a \mid bc$

Simplified CFG

$S \longrightarrow bb \mid a \mid bc$

$A \longrightarrow a \mid bc \mid bb$

$B \longrightarrow a \mid bc \mid bb$

7. Convert the given grammar into CNF

$S \longrightarrow ASB \mid aB$

$A \longrightarrow B \mid S$

$B \longrightarrow b \mid \varepsilon$

Given $S \to ASB \mid aB$

$A \to B \mid S$

$B \to b \mid \epsilon$

Step-1: Here we have start symbol 's' on both

LHS & RHS. so, add a new production like

$$S' \longrightarrow S$$
$$S \longrightarrow ASB \mid aB$$
$$A \longrightarrow B \mid S$$
$$B \longrightarrow b \mid \epsilon$$

Step-2 : Remove the null productions: $B \rightarrow \epsilon$ and $A \rightarrow \epsilon$

After removing $B \rightarrow \epsilon$: $S' \rightarrow S$, $S \rightarrow ASB \mid aB \mid a \mid AS$, $A \rightarrow B \mid S \mid \epsilon$,
$B \rightarrow b$

After Removing $A \rightarrow \epsilon$: $S' \rightarrow S$, $S \rightarrow ASB \mid aB \mid a \mid AS \mid SB \mid S$, $A \rightarrow B \mid S$
$B \rightarrow b$

Step-3 : Remove the unit productions : $S \rightarrow S$, $S' \rightarrow S$,
$A \rightarrow B$ and $A \rightarrow S$

After removing $S \rightarrow S$: R: $S' \rightarrow S$, $S \rightarrow ASB \mid aB \mid a \mid AS \mid SB$, $A \rightarrow B \mid S$,
$B \rightarrow b$

After removing $S' \rightarrow S$: P: $S' \rightarrow ASB \mid aB \mid a \mid AS \mid SB$,
$S \rightarrow ASB \mid aB \mid a \mid AS \mid SB$,
$A \rightarrow B \mid S$, $B \rightarrow b$

After removing $A \rightarrow B$: P : $S' \rightarrow ASB \mid aB \mid a \mid AS \mid SB$,
$S \rightarrow ASB \mid aB \mid a \mid AS \mid SB$,
$A \rightarrow b \mid S$, ~~$B \rightarrow b$~~

After removing $A \rightarrow S$: P : $S' \rightarrow ASB \mid aB \mid a \mid AS \mid SB$,
$S \rightarrow ASB \mid aB \mid a \mid AS \mid SB$,
$A \rightarrow b \mid ASB \mid aB \mid a \mid AS \mid SB$
~~$B \rightarrow b$~~ $B \rightarrow b$

Step-4 : Now find out the productions that has more than two variables in RHS

$S' \rightarrow ASB$, $S \rightarrow ASB$ & $A \rightarrow ASB$

After removing these, we get :

P : S' ⟶ AX | aB | a | AS | SB ,

S ⟶ AX | aB | a | AS | SB ,

A ⟶ b | AX | aB | a | AS | SB

B ⟶ b ; A ⟶ b

X ⟶ SB

Step - 5 : Now change the productions S' ⟶ aB, S ⟶ aB and A ⟶ aB

finally we get :

P : S' ⟶ AX | YB | a | AS | SB

S ⟶ AX | YB | a | AS | SB

A ⟶ b | AX | YB | a | AS | SB , A ⟶ b

B ⟶ b

X ⟶ SAB

Y ⟶ a

Which is the required chomsky Normal form for the given CFG is

S' ⟶ AX | YB | AS | a | SB

S ⟶ AX | YB | AS | a | SB

A ⟶ b | AX | YB | AS | a | SB

B ⟶ b

X ⟶ SB

Y ⟶ a

8. Convert the grammar into Greibach Normal form

S ⟶ CA | BB

B ⟶ b | SB

C ⟶ a

A ⟶ b

Step - 1 : Given CFG doesn't have any unit productions or Null productions.

Step-2: Given CFG is in CNF

Step-3: Replace S with A_1

Replace C with A_2

Replace A with A_3

Replace B with A_4

$A_1 \rightarrow A_2A_3 \mid A_4A_4$

$A_4 \rightarrow b \mid A_1A_4$

$A_2 \rightarrow a$

$A_3 \rightarrow b$

Step-4: Consider the productions one by one and che

$A_1 \rightarrow A_2A_3 \rightarrow 1 < 2$

$A_1 \rightarrow A_4A_4 \rightarrow 1 < 4$

$A_4 \rightarrow b$

$A_4 \rightarrow A_1A_4 \quad 4 > 1$

Here $i > j$, so substitute A_1 in A_4

$A_4 \rightarrow b \mid A_2A_3A_4 \mid A_4A_4A_4$

$A_4 \rightarrow A_2A_3A_4$

$4 > 2$

Here $i > j$ so substitute A_2 in A_4

$A_4 \rightarrow b \mid aA_3A_4 \mid A_4A_4A_4$

Now check $A_4 \rightarrow A_4A_4A_4$

$4 = 4.$

$i = j$ where left recurssion is occured. So remove left recursion

Step-5: $\underbrace{A_4}_{A} \rightarrow \underbrace{\underbrace{A_4}_{A}A_4A_4}_{\alpha} \mid \underbrace{b \mid aA_3A_4}_{\beta}$

$A_4 \longrightarrow bz \mid aA_3A_4z$

$z \longrightarrow A_4A_4z \mid \epsilon$

Remove null production

$z \longrightarrow \epsilon$

$A_4 \longrightarrow bz \mid aA_3A_4z \mid b \mid aA_3A_4$

$z \longrightarrow A_4A_4z \mid A_4A_4$

$A \longrightarrow A\alpha \mid \beta$

$\Downarrow$

$A \longrightarrow \beta A'$

$A' \longrightarrow \alpha A' \mid \epsilon$

Step-6 : We got CNF as

$A_1 \longrightarrow A_2A_3 \mid A_4A_4$

$A_4 \longrightarrow bz \mid aA_3A_4z \mid b \mid aA_3A_4$

$z \longrightarrow A_4A_4z \mid A_4A_4$

$A_2 \longrightarrow a$

$A_3 \longrightarrow b$

Here A_1 is not in GNF. so, convert it into GNF by GNF rules substitute A_2 & A_4 in A_1

$A_1 \longrightarrow aA_3 \mid bzA_4 \mid aA_3A_4zA_4 \mid bA_4 \mid aA_3A_4A_4$

and z is also not in GNF. So substitute A_4 in z

$z \rightarrow bzA_4z \mid aA_3A_4zA_4z \mid bA_4z \mid aA_3A_4A_4z \mid bzA_4 \mid aA_3A_4zA_4 \mid bA_4 \mid aA_3A_4A_4$

Finally we got CNF as

$A_1 \rightarrow aA_3 \mid bzA_4 \mid aA_3A_4zA_4 \mid bA_4 \mid aA_3A_4A_4$

$A_4 \rightarrow bz \mid aA_3A_4z \mid b \mid aA_3A_4$

$z \rightarrow \cancel{A_4A_4z \mid A_4} \; bzA_4z \mid aA_3A_4zA_4z \mid bA_4z \mid aA_3A_4A_4z \mid bzA_4 \mid aA_3A_4zA_4 \mid bA_4 \mid aA_3A_4A_4$

$A_2 \rightarrow b$

$A_3 \rightarrow a$

10. Prove that the given language $L = \{a^n b^m$ where $n > 1$, $m \neq n\}$ is not regular

10. Design the CFG for the expressions $\{WCW^R$, where W belongs to $(a+b)^+$ and W^R is the reverse of the string$\}$.

10.Ans Given $L = \{wcw^R / w \in (a+b)^+\}$

$w = ab$ $w^R = ba$

$wcw^R = abcba$

$w = ba$ $w^R = ab$

$wcw^R = bacab$

Production rules

$S \rightarrow aSa \mid bSb \mid a \mid b \mid c$

input string : abcba

$S \rightarrow aSa$

$\rightarrow abSba$ $(S \rightarrow bSb)$

$\rightarrow abcba$ $(S \rightarrow c)$

CFG :

$G = \{\{S\}, \{a, b\}, P, S, \}$

9. Prove that the given language $L = \{a^n b^m$ where $n > 1$, $m > n\}$ is not regular.

$L = \{a^n b^m \mid n > 1, m > n\}$

let L be a Regular language

$L = \{aabbb, aaabbbb, aaaabbbbb, \ldots\}$

$n=2, m=3$ $n=3, m=4$ $n=4, m=5$

$|z| \geq n$

let $n = 5$, $z = aabbb$

$|z| \geq n$

$|aabbb| \geq 5$

$5 \geq 5$

select a string aabbb, and then divide into 3 Parts

now divide $z = x \cdot y \cdot z$ ~~u.v.w~~

$z = \underbrace{aa}_{x}\underbrace{bb}_{y}\underbrace{b}_{z}$

$x = aa$
$y = bb$
$z = b$

Case 1: $|xy| \le n$

$|aabb| \le 5$

$4 \le 5$ ✓

Case 2: $|x| \ge 1$

$|aa| \ge 1$

$2 \ge 1$ ✓

Case 3: $\forall i \ge 0,\ x \cdot y^i z \in L$

$i = 0 \Rightarrow x y^0 z \Rightarrow aa(bb)^0 b$

$\Rightarrow aab \notin L$

It is not Regular Language.