Unit - 3 13/12/21

CONTEXT FREE GRAMMARS AND LANGUAGES

Chomsky Hierarchy

- Type 0 - Structured
- Type 1 - Content
- Type 2 - Content free
- Type 3 - Regular

Compound sentence → Simple sentence Conjunction Simple sentence
(Conjunction → and/or/but)

Simple - sentence → <subject> <verb> <object>

Subject - noun
object - noun/pronoun

Content Free Grammar

A CFG $G = (V, T, P, S)$ has following

V is set of variables or non terminals

T is set of terminals

P is set of Productions or rules

S is Start symbol

for a Simple Sentence V includes: simple sentence, subject, object

T includes: Noun, Pronoun and verbs

P includes: Simple sentence, subject(noun), Object(noun/pronoun)

S includes: Simple sentence

Write a CFG for any number of a's

$S \to aS$

$S \to \epsilon$

Write a CFG for atleast one a

$S \to aS$

$S \to a$

Write a CFG for any number of a's and b's

$S \rightarrow aS$
$S \rightarrow bS$ or $S \rightarrow aS \mid bS \mid \epsilon$
$S \rightarrow \epsilon$

Write a CFG beginning with a

$S \rightarrow aA$
$A \rightarrow aA$
$A \rightarrow bA$
$A \rightarrow \epsilon$

Write a CFG ~~beginning~~ for substring with a

$S \rightarrow A\,ab\,A$
$A \rightarrow aA \mid bA \mid \epsilon$

$G = (V, T, P, S)$
$V = \{S, A\}$
$T = \{a, b\}$
$P = \{S \rightarrow AabA \quad A \rightarrow aA \mid bA \mid \epsilon\}$
S = start symbol

Write CFG for palindrome, assume that palindrome has 0's an

$S \rightarrow 0S0 \mid 1S1$
$S \rightarrow 0$
$S \rightarrow 1$
$S \rightarrow \epsilon$

for 10101 2nd rule, 1st rule, 4th rule

CFG for equal no of a's followed by equal number of b's

$S \rightarrow aSb$ $\quad V = \{S\}$ $\quad P = \{S \rightarrow aSb \quad S \rightarrow \epsilon\}$
$S \rightarrow \epsilon$ $\quad T = \{a, b\}$ $\quad S$ = start symbol

CFG for $a^n b^n$ $n \geq 1$ $\quad L = \{a^n b^n \mid n \geq 1\}$

$S \rightarrow aSb$
$S \rightarrow ab$

① Write the CFG that has atleast 2 a's

→ $S \rightarrow aS$

$S \rightarrow aa$

③ Even no of a's

→ $S \rightarrow aaS$

$S \rightarrow \epsilon$

④ odd no of a's

⇒ $S \rightarrow aaS$

$S \rightarrow a$

⑤ no of a's is divisible by 3 n_a – no of a's

$L = \{w \mid n_a(w) \bmod 3 = 0, w \in a^*\}$

→ $S \rightarrow aaaS$

$S \rightarrow \epsilon$

⑥ $L = \{w \mid |w| \bmod 3 = 0, w \in \{a, b\}^*\}$ $|w|$ = length of a, b divisible by 3

→ $S \rightarrow AAAS$

$S \rightarrow \epsilon$

$A \rightarrow a \mid b$

⑦ Write CFG for ~~[illegible]~~

$L = \{w \mid |w| \bmod 3 > 0, w \in a^+\}$

→ $S \rightarrow a \mid aa \mid aaaS$ Multiple of 3 not valid; ϵ also not valid

S gets replaced by a or aa

⑧ write CFG no of a's is divisible by 2 and even b's

→ $L = \{w \mid n_a(w) \bmod 2 = 0, w \in \{a, b\}^*\}$

RE → $(b^* a b^* a b^*)^*$

$S \rightarrow SaSaS$

$S \rightarrow bS$

$S \rightarrow \epsilon$

(9) no of a's equal to no of b's

$S \rightarrow SS$

$\longrightarrow S \rightarrow aSb$

$S \rightarrow bSa$

$S \rightarrow \epsilon$

(10) Write a CFG for a balanced parenthesis

4A $S \rightarrow (S) \mid \{S\} \mid [S]$

$S \rightarrow SS$

$S \rightarrow \epsilon$

(11) Write a CFG

4A $L = \{0^m 1^m 2^n \mid m \geq 1, n \geq 0\}$

$S \rightarrow \underset{A}{} \; \underset{B}{}$

$S \rightarrow 0A1 \mid 01$

$B \rightarrow 2B \mid \epsilon$

(12) $L = \{a^n b^m \mid n \geq 0, m > n\}$

$\rightarrow$ $S \rightarrow aSb \mid B$

$B \rightarrow bB \mid b$

(13) $L = \{a^n b^{n-3} \mid n \geq 3\}$

$\Rightarrow S \rightarrow aSb \mid aaa$

(14) Write a CFG for the following

$L = \{0^i 1^j \mid i \neq j, i \geq 0, j \geq 0\}$

$S \rightarrow 0S1 \mid A \mid B$

$A \rightarrow 0A \mid 0$

$B \rightarrow 1B \mid 1$

* (15) $L = \{a^n b^m c^k \mid n + 2m = k \text{ for } n \geq 0, m \geq 0\}$

$a^n b^m c^k$

$a^n b^m c^{n+2m}$

$a^n b^m c^n c^{2m}$

$a^n b^m \; c^{2m} \; c^n$

$S \rightarrow aSc \,/\, A$

$A \rightarrow bAcc \,/\, \epsilon$

(16) $L = \{ a^n b^m c^k \,/\, m = n + k$ for $n \geq 0 \;\; k \geq 0 \}$

$a^n b^m c^k$

$a^n b^{n+k} c^k$

$a^n b^n b^k c^k$

$S \rightarrow AB$

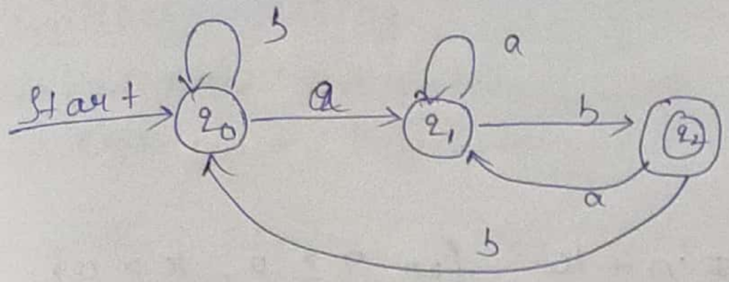
$A \rightarrow aAb \,/\, \epsilon$

$B \rightarrow bBc \,/\, \epsilon$

(17) $L = \{ w \,/\, |w| \bmod 3 \neq |w| \bmod 2 \}. \; w \in \{a, b\}^* \}$

$\rightarrow$ 2, 3, 4, 5, 8, 9, 10, 11, 14 invalid - 1, 6, 7, 12, 13

$S \rightarrow aa \,|\, aaa \,|\, aaaa \,|\, aaaaa \,|\, aaaaaaS$

① 

②

	0	1
$\{q_0\}$	$\{q_0, q_1\}$	$\{q_1\}$
$\{q_0, q_1\}$	$\{q_0, q_1\}$	$\{q_1, q_2\}$
$\{q_1\}$	$\{q_1\}$	$\{q_2\}$
$\{q_1, q_2\}$	$\{q_1\}$	$\{q_2\}$
$\{q_2\}$	ϕ	$\{q_2\}$

③ i) $aaaaa^*(b+c)(b+c)(b+c)$

ii) $(b^*ab^*ab^*ab^*)^*$

Write a CFG for the following

4.A $L = \{0^i 1^j \mid i \neq j,\ i \geq 0,\ j \geq 0\}$

$S \rightarrow 0S1 \mid A \mid B$

$A \rightarrow 0A \mid 0$

$B \rightarrow 1B \mid 1$

4.A $L = \{a^n b^m c^k \mid n + 2m = k \text{ for } n \geq 0,\ m \geq 0\}$

this kind of prob we have to solve like

$a^n b^m c^k$

[Note: Here $k = n + 2m$ so we should break c^k

$a^n b^m c^{n+2m}$

$a^n b^m c^n c^{2m}$

(Note: we can't take c directly towards left because it is like a followed by b followed by c

$a^n \underbrace{b^m c^{2m}} c^n$

$S \rightarrow aSc \mid A$

$A \rightarrow bAcc \mid \epsilon$

(3) $L = \{a^n b^m c^k \mid m = n + k$ for $n \geq 0, k \geq 0\}$.

$a^n b^m c^k$

$a^n b^{n+k} c^k$

$a^n b^n b^k c^k$

Note: we should break b^m bcz $m = n+k$

$S \rightarrow AB$

$A \rightarrow aAb \mid \epsilon$

$B \rightarrow bBc \mid \epsilon$

(4) $L = \{w \mid |w| \bmod 3 \neq |w| \bmod 2, w \in \{a, b\}^*\}$

→ Not valid : 0, 1, 6, 7, 12

valid : 2, 3, 4, 5, 8, 9, 10, 11

$S \rightarrow aa \mid aaa \mid aaaa \mid aaaaa \mid aaaaaaS$

(5) $L = \{w \mid |w| \bmod 3 \geq |w| \bmod 2, w \in \{a\}^*\}$

→ Not valid 3, 9

valid : 0, 1, 2, 4, 5

$S \rightarrow \epsilon \mid a \mid aa \mid aaaa \mid aaaaa \mid aaaaaaS$

Derivations using a grammar

↳ Leftmost Derivation (LMD)

↳ Rightmost Derivation (RMD)

derive the string 3 a's followed by 3 b's

$S \rightarrow aSb \mid \epsilon$

$S \Rightarrow aSb$	$\because S \rightarrow aSb$
$\Rightarrow aaSbb$	$\because S \rightarrow aSb$
$\Rightarrow aaaSbbb$	$\because S \rightarrow aSb$
$\Rightarrow aaabbb$	$\because S \rightarrow \epsilon$

Derive ababba

$S \rightarrow SS$

$S \rightarrow aSb$

$S \rightarrow bSa$

$S \rightarrow \epsilon$

ababba

$S \Rightarrow SS$

$S \Rightarrow SSS$

~~asbasbbsa~~

~~ababba~~

$S \underset{lm}{\Rightarrow} aSbSS$

$S \underset{lm}{\Rightarrow} abSS$

$S \underset{lm}{\Rightarrow} abaSbS$

$S \underset{lm}{\Rightarrow} ababS$

$S \underset{lm}{\Rightarrow} ababbSa$

$S \underset{lm}{\Rightarrow} ababba$

Note: $S \rightarrow SS$ replace by SS becoz $S \rightarrow aSb$ 2 three time

each step replace only one production

$S \Rightarrow SS$ → Right, left

if we replace left S then its LMD, if right its RMD

$S \underset{rm}{\Rightarrow} SS$

$S \underset{rm}{\Rightarrow} SSS$

$S \underset{rm}{\Rightarrow} SSbSa$

$S \underset{rm}{\Rightarrow} SSba$

$S \underset{rm}{\Rightarrow} SaSbba$

$S \underset{rm}{\Rightarrow} Sabba$

$S \underset{rm}{\Rightarrow} aSbabba$

$S \underset{rm}{\Rightarrow} ababba$

20/12/21

Defination

At Each step we replace the left most variable by one of its production bodies such a derivation is called a left most derivation

At Each step we replace the Right most variable by one of its production bodies such a derivation is called Right most derivation

give left most & Right most derivation of the following string

a) 00101

b) 1001

c) 00011

S → A1B

A → 0A | ε

B → 0B | 1B | ε

a) 00101

$S \underset{lm}{\Rightarrow} A1B$ ∵ S → A1B

$S \underset{lm}{\Rightarrow} 0A1B$ ∵ A → 0A

$S \underset{lm}{\Rightarrow} 00A1B$ ∵ A → 0A

$S \underset{lm}{\Rightarrow} 001B$ ∵ A → ε

$S \underset{lm}{\Rightarrow} 0010B$ ∵ B → 0B

$S \underset{lm}{\Rightarrow} 00101B$ ∵ B → 1B

$S \underset{lm}{\Rightarrow} 00101$ ∵ B → ε

$S \underset{rm}{\Rightarrow} A1B$

$S \underset{rm}{\Rightarrow} A10B$ $\quad \because S \to A1B$

$S \underset{rm}{\Rightarrow} A101B$ $\quad \because B \to 0B$

$S \underset{rm}{\Rightarrow} A101$ $\quad \because B \to 1B$

$\because B \to \epsilon$

$S \underset{rm}{\Rightarrow} 0A101$ $\quad \because A \to 0A$

$S \underset{rm}{\Rightarrow} 00A101$ $\quad \because A \to 0A$

$S \underset{rm}{\Rightarrow} 00101$ $\quad \because A \to \epsilon$

(5) 1001

$S \underset{lm}{\Rightarrow} A1B$ $\quad S \to A1B$

$S \underset{lm}{\Rightarrow} 1B$ $\quad A \to \epsilon$

$S \underset{lm}{\Rightarrow} 10B$ $\quad B \to 0B$

$S \underset{lm}{\Rightarrow} 100B$ $\quad B \to 0B$

$S \underset{lm}{\Rightarrow} 1001B$ $\quad B \to 1B$

$S \underset{lm}{\Rightarrow} 1001$ $\quad B \to \epsilon$

$S \underset{rm}{\Rightarrow} A1B$ $\quad S \to A1B$

$S \underset{rm}{\Rightarrow} A10B$ $\quad S \to 0B$

$S \underset{rm}{\Rightarrow} A100B$ $\quad B \to 0B$

$S \underset{rm}{\Rightarrow} A1001B$ $\quad B \to 1B$

$S \underset{rm}{\Rightarrow} A1001$ $\quad B \to \epsilon$

$S \underset{rm}{\Rightarrow} \cancel{A}1001$ $\quad A \to \epsilon$

③ 00011

$S \xrightarrow[lm]{} A1B$ $S \rightarrow A1B$

$S \xrightarrow[lm]{} 0A1B$ $A \rightarrow 0A$

$S \xRightarrow[lm]{} 00A1B$ $A \rightarrow 0A$

$S \xrightarrow[lm]{} 000A1B$ $A \rightarrow 0A$

$S \xrightarrow[lm]{} 0001B$ $A \rightarrow \epsilon$

$S \xRightarrow[lm]{} 00011B$ $B \rightarrow 1B$

$S \xRightarrow[lm]{} 00011$ $B \rightarrow \epsilon$

$S \xrightarrow[rm]{} A1B$ $S \Rightarrow A1B$

$S \xrightarrow[rm]{} A11B$ $S \rightarrow 1B$

$S \xRightarrow[rm]{} A11$ $S \rightarrow \epsilon$

$S \xrightarrow[rm]{} 0A11$ $A \rightarrow 0A$

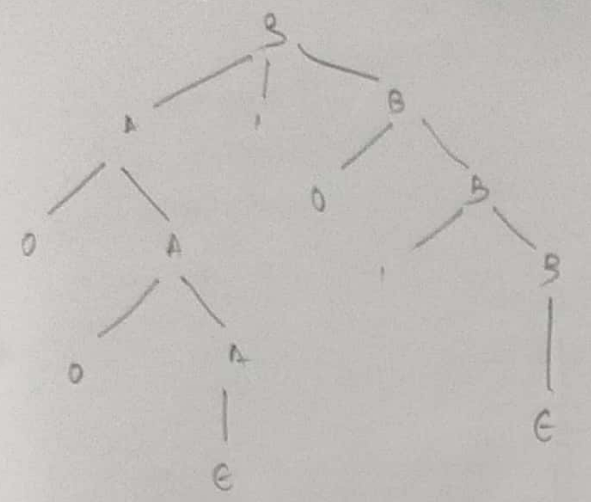
$S \xRightarrow[rm]{} 00A11$ $A \rightarrow 0A$

$S \xRightarrow[rm]{} 000A11$ $A \rightarrow 0A$

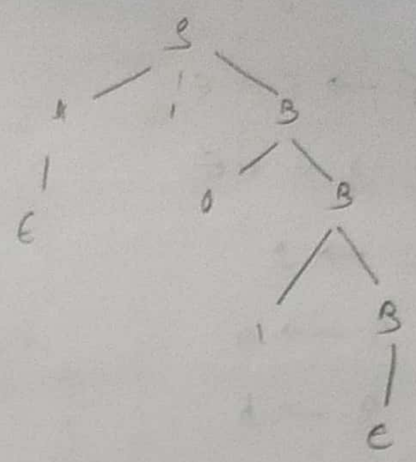
$S \xRightarrow[rm]{} 00011$ $A \rightarrow \epsilon$

Parse Tree (Derivation Tree)

00101

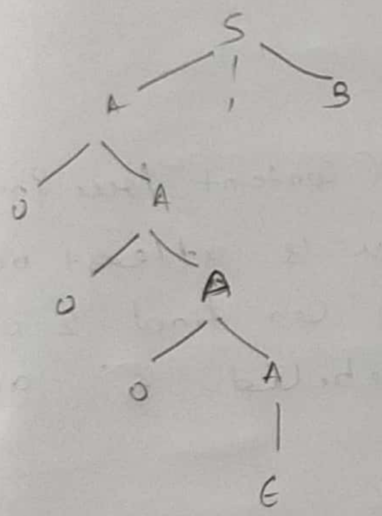


1001



When we concatenate a tree from left to right we get a tree called as <u>yield</u> of the Parse tree

00011



Parse Tree :

21/12/21

Ambiguity in Grammars & Languages

$E \rightarrow E + E$

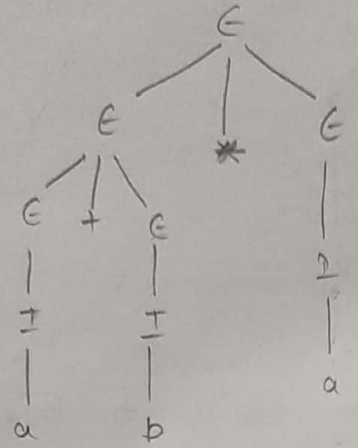
$E \rightarrow E * E$

$E \rightarrow I$

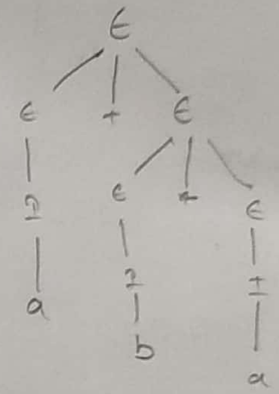
$I \rightarrow a$

$I \rightarrow b$

$a + b * a$



or



variables are also called as non terminals

a, b ~~→ terminals~~

+, * → terminals

Ambiguous grammar

Let $G = (V, T, P, S)$ be a CFG (Context free Grammar) we say G is Ambiguous if there is atleast one string w in T^* for which we can find 2 different parse Tree's each with root labeled 'S' and yield w

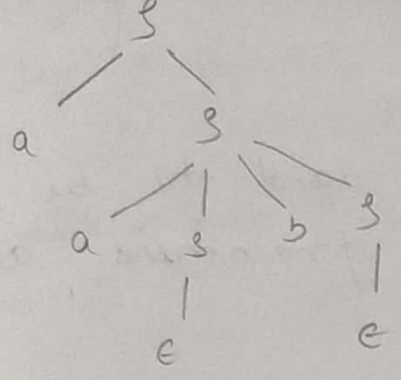
Consider the following grammar

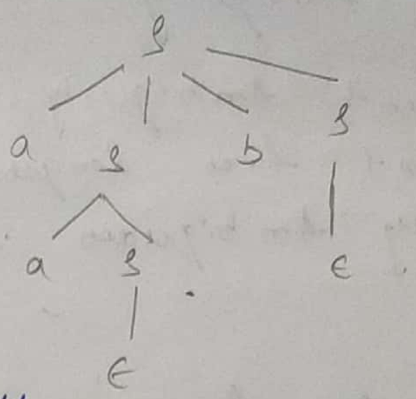
$S \rightarrow aS$

$S \rightarrow aSbS$

$S \rightarrow \epsilon$

Check whether this grammar is Ambiguous if so construct two parse tree, two left most derivations, two right most derivation





yes the grammer is Ambigoous

LMD

$S \underset{lm}{\Rightarrow} aS$ $\quad S \to aS$

$S \underset{lm}{\Rightarrow} aaSbS$ $\quad S \to aSbS$

$S \underset{lm}{\Rightarrow} aabS$ $\quad S \to \epsilon$

$S \underset{lm}{\Rightarrow} aab$

LM

$S \underset{lm}{\Rightarrow} aSbS$ $\quad S \to aSbS$

$S \underset{lm}{\Rightarrow} aaSbS$ $\quad S \to aS$

$S \underset{lm}{\Rightarrow} aabS$ $\quad S \to \epsilon$

$S \underset{lm}{\Rightarrow} aab$ $\quad S \to \epsilon$

RMD

$S \underset{rm}{\Rightarrow} aS$ $\quad \because S \to aS$

$S \underset{rm}{\Rightarrow} aaSbS$ $\quad S \to aSbS$

$S \underset{rm}{\Rightarrow} aaSb$ $\quad S \to \epsilon$

$S \underset{rm}{\Rightarrow} aab$

$S \underset{rm}{\Rightarrow} aSbS$ $\quad S \to aSbS$

$S \underset{rm}{\Rightarrow} aSb$ $\quad S \to \epsilon$

$S \underset{rm}{\Rightarrow} aaSb$ $\quad S \to aS$

$S \underset{rm}{\Rightarrow} aab$ $\quad S \to \epsilon$

3/1/2022

- Removing Ambiguity from grammer (not imp)

- Inherent Ambiguity (not imp)

a Context free language L is said to be inherently Ambiguous if all its grammers are Ambiguous

- Sentential form

if $G = (V, T, P, S)$ is a CFG then any string α in $(V \cup T)^*$ such that

$S \xrightarrow{*} \alpha$ is a sentential form

10 M

Properties of CFL (Context free language)

Normal forms for CFG (Conversion from CFG to CNF)

LHS → RHS A → BC
S → Aa A → a

→ The goal is to convert all productions of the form $A \to BC$ A, B, C – variable

$A \to a$ a → terminal

This form is called chomsky normal form (CNF)

→ To convert the gramer to CNF we need to make a number of preliminary simplifications:

1) we must eliminate useless symbols. Those variables or terminals that do not appear in any derivation of a terminal string from the start symbol

2) we must eliminate ϵ - productions those of the form $A \to \epsilon$ for some variable 'A'

3) we must eliminate unit production those of the form $A \to B$ for some variables A & B

Elimination order from

2, 3, 1 → steps

Scanned with CamScanner

Eliminating useless symbols

we say a symbol X is useful for a grammer G if there is some derivation of the form

$$S \xRightarrow{*} \alpha X \beta \xRightarrow{*} w$$

were w is in T^*

Note : X may be variable or terminal

The approach to eliminate useless symbols begins by identifying two things :

i) generating - we say X is generating if

i) $X \xRightarrow{*} w$ $w \Rightarrow$ terminal

for some terminal string w

ii) we say X is reachable if there is derivation

$$S \xRightarrow{*} \alpha X \beta$$

for some Alpha & β

SM

① Eliminate useless symbol from following grammer.

$S \rightarrow AB \mid a$

$A \rightarrow b$

$S, A, B, a, b \rightarrow$ symbols

Check whether

i) Generating symbols

S, A, a, b are generating

$S \rightarrow a$

$A \rightarrow b$

∴ we have to remove B after removing B entier grammer is removed i.e $\underline{AB}$

ii) Reachable

S, a are reachable

$S \rightarrow a$

(2) Eliminate useless symbols from following grammar

$S \to AB \mid CA$

$A \to a$

$B \to BC \mid AB$

$C \to aB \mid b$

symbols → S, A, B, C, a, b

i) generating

generating symbol → a, b, A, C, S

$S \to CA$

$A \to a$

$C \to b$

$S \to AB \mid CA$

C is generating & A is generating, so S is generating

B → not generating becz in BC B not generating & AB → B not generating

2) Reachable

S, C, A, a b

$S \to CA$

$A \to a$

$C \to b$

(3) Eliminate useless symbols from following grammar

$S \to ASB \mid \epsilon$

$A \to aAS \mid a$

$B \to SbS \mid A \mid bb$

generating → A, B, S, a, b

Reachable → S, A, B, a, b

grammar is reachable

Eliminating ϵ - productions

|4|1|2022|

Nullable Symbol

These ~~are~~ are the symbols which derive "ϵ"

A variable "A" is Nullable if $A \xRightarrow{*} \epsilon$. If A is nullable then whenever A appears in a product in body say $S \rightarrow BAC$, A might (might not) derive Epsilon

$S \rightarrow BC$ - might

$S \rightarrow BAC$ - might not

~~Grammar~~

BASIS : If $A \rightarrow \epsilon$ is the production of G then A is nullable
Grammar

INDUCTION: If there is a production $B \rightarrow C_1 C_2 \dots C_k$ where each C_i is nullable then B is also nullable

Q) Eliminate ϵ - production from the following

$S \rightarrow AB$

$A \rightarrow aAA \mid \epsilon$

$B \rightarrow bBB \mid \epsilon$

$\Rightarrow$ S, A, B all are nullable

$S \rightarrow AB \mid A \mid B$

$A \rightarrow aAA \mid a \mid a\underbrace{A \mid aA}$

similar so keep only one

$A \rightarrow aAA \mid a \mid aA$

$B \rightarrow bBB \mid b \mid bB \mid bB$

$B \rightarrow bBB \mid b \mid bB$

there are 4 ways to remove

① $S \rightarrow AB$ keep both

② $S \rightarrow A$ remove B

③ $S \rightarrow B$ remove A

④ remove both $S \rightarrow \epsilon$

Q) Eliminate ϵ production from the following grammar

$S \to ASB \mid \epsilon$

$A \to aAS \mid a$

$B \to SbS \mid A \mid bb$

A & B are nullable because they won't produce ϵ

$\to$ only S is nullable

$S \to ASB \mid AB$

$A \to aAS \mid aA \mid a$

$B \to SbS \mid b \mid Sb \mid bS \mid A \mid bb$

Eliminating Unit Production

$A \to B$

a Unit production is the Production of the form $A \to B$ where both A & B are variables

Unit pair - A pair (A, B) is called unit pair such that A derives B ($A \overset{*}{\Rightarrow} B$) using only unit productions

BASIS: (A, A) is a unit pair for any variable A

$A \overset{*}{\Rightarrow} A$

ie A derives A by zero steps

INDUCTION: Suppose we have determine that A, B is a unit pair and $B \to C$ is a production where C is a variable then (A, C) is a unit pair

(A, B) is a unit pair

$B \to C$ is a production

(A, C) is a unit pair

① Eliminate Unit Production from the following grammar.

$S \to ASB \mid \epsilon$

$A \to aAS \mid a$

$B \to SbS \mid A \mid bb$

⇒ (S, S), (A, A), (B, B) are the unit pairs from the BASIS step

(S, S)

(A, A)

(B, B) & $B \to A$ (B, A)

Unit pair	non unit production body
(S, S)	$S \to ASB \mid \epsilon$
(A, A)	$A \to aAS \mid a$
(B, B)	$B \to SbS \mid bb$
(B, A) (B, A)	$B \to aAS \mid a$

$S \to ASB \mid \epsilon$

$A \to aAS \mid a$

$B \to SbS \mid bb \mid aAS \mid a$

② Eliminate Unit Production from the following grammar.

$I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

$F \to I \mid (E)$

$T \to F \mid T * F$

$E \to T \mid E + T$

OR

⇒ (E, E), (T, T), (F, F), (I, I)

(E, E) & $E \to T$ (E, T)

(E, T) & $T \to F$ (E, F)

(E, F) & $F \to I$ (E, I)

(T, T) & T → F (T, F)

(T, F) & F → I (T, I)

(F, F) & F → I (F, I)

Unit Pairs	Non Unit Productions
(E, E)	E → E + T
(E, T)	E → T * F
(E, F)	E → (E)
(E, I)	E → a | b | Ia | Ib | I0 | I1
(T, T)	T → T * F
(T, F)	T → (E)
(T, I)	T → a | b | Ia | Ib | I0 | I1
(F, F)	F → (E)
(F, I)	F → a | b | Ia | Ib | I0 | I1
(I, I)	I → a | b | Ia | Ib | I0 | I1

E → E + T | T * F | (E) | a | b | Ia | Ib | I0 | I1

T → T * F | (E) | a | b | Ia | Ib | I0 | I1

F → (E) | a | b | Ia | Ib | I0 | I1

I → a | b | Ia | Ib | I0 | I1

Scanned with CamScanner