

Lec-31: Pumping lemma for regular languages in TOC with examples



If L is an infinite language then there exists some positive integer ' n ' (Pumping length) such that any string $w \in L$ has length greater than equal to ' n ' i.e. $|w| \geq n$ then w can be divided into three parts, $w = xyz$ satisfy following conditions

i) for each $i \geq 0$, $xy^i z \in L$ ✓

ii) $|y| > 0$ and $(bb)^{i+1}$

iii) $|xy| \leq n$

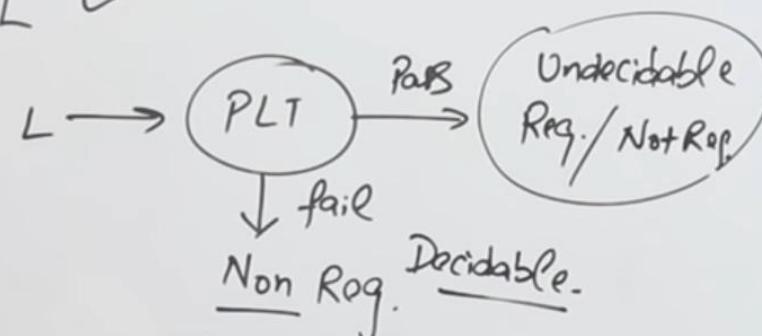
$$w = xyz$$

x y z

$$a^n b^{2n} \quad n \geq 0$$

$$\frac{aa}{x} \frac{bbbb}{y} z \in L$$

$$\frac{aa}{x} \frac{bbbb}{y} \frac{bb}{z} \notin L$$

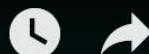


$$\begin{array}{c|c|c} x & y & z \\ \hline w & & \end{array}$$

$$\frac{aa}{x} \frac{bbbb}{y} \frac{bb}{z}$$

$$\frac{a}{x} \frac{qbqb}{y} \frac{bbb}{z} \notin L$$

Pumping Lemma (For Regular Languages)



» Pumping Lemma is used to prove that a Language is NOT REGULAR

» It cannot be used to prove that a Language is Regular

If A is a Regular Language, then A has a Pumping Length ' P ' such that any string ' S ' where $|S| \geq P$ may be divided into 3 parts $S = x y z$ such that the following conditions must be true:

- (1) $x y^i z \in A$ for every $i \geq 0$
- (2) $|y| > 0$
- (3) $|xy| \leq P$

To prove that a language is not Regular using PUMPING LEMMA, follow the below steps:

(We prove using Contradiction)

-> Assume that A is Regular

►-> It has to have a Pumping Length (say P)



Pumping Lemma (For Context Free Languages)

Pumping Lemma (for CFL) is used to prove that a language is NOT Context Free

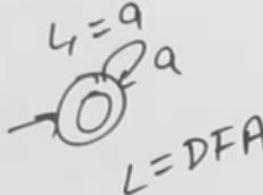
If A is a Context Free Language, then, A has a Pumping Length ' P ' such that any string ' S ', where $|S| \geq P$ may be divided into 5 pieces $S = uvxyz$ such that the following conditions must be true:

- (1) uv^ixy^iz is in A for every $i \geq 0$
- (2) $|vy| > 0$
- (3) $|vxy| \leq P$



Closure Properties of Regular Languages

1) Union ($L_1 \cup L_2$)



2) Concatenation ($L_1 \cdot L_2$)

3) Closure (*) L^*

Complementation $\bar{L} = \Sigma^* - L$

Intersection $L_1 \cap L_2 = \overline{\overline{L}_1 \cup \overline{L}_2}$

6) Difference $L_1 - L_2 = L_1 \cap \overline{L_2}$

Reversal $(L)^R$

Homomorphism

Reverse Homomorphism

Quotient operation

INIT

Substitution

\vdash Finite Union *

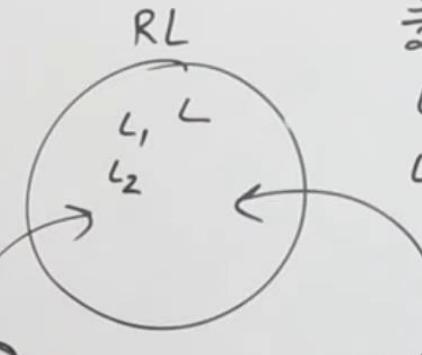
Integer $\mathbb{Z} = \dots -5 -4 -3 -2 -1 0 1 2 \dots$

$$1+2=3$$

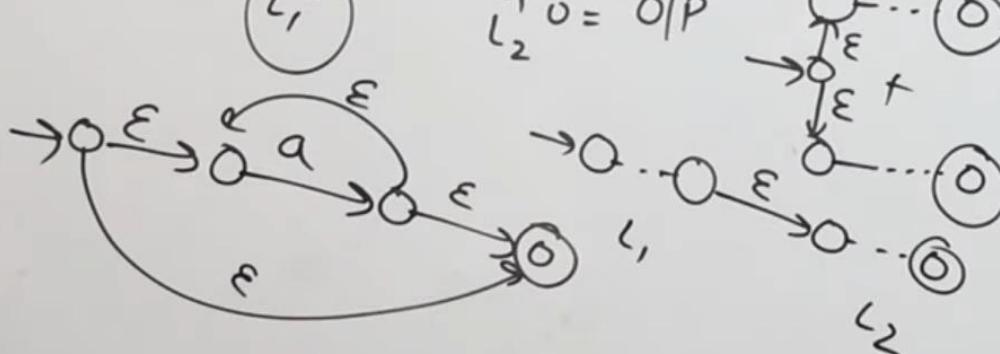
$$\frac{3}{2} = 1.5$$

RE . RE

$$\begin{cases} L_1 \rightarrow RE \\ L_2 \rightarrow RE \end{cases}$$

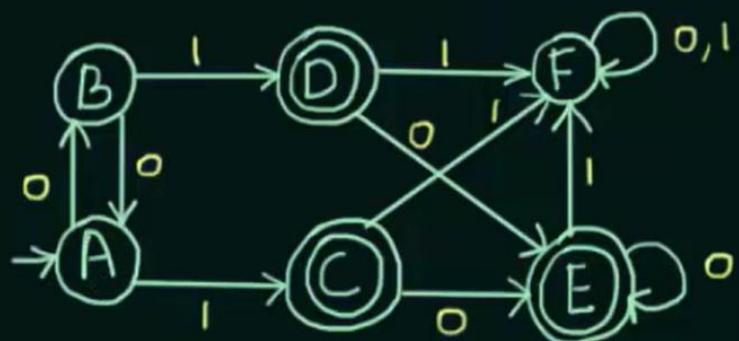


$L_1^* = O/P$



Minimization of DFA - Table Filling Method

(Myhill-Nerode Theorem)



A B C D E F

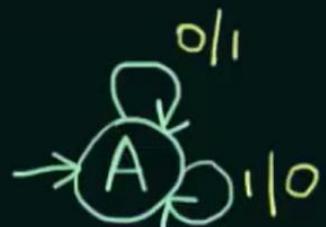
A					
B					
C	✓	✓			
D	✓	✓			
E					
F					

Steps:

- 1) Draw a table for all pairs of states (P, Q)
- 2) Mark all pairs where $P \in F$ and $Q \notin F$
- 3) If there are any Unmarked pairs (P, Q) such that $[\delta(P, x), \delta(Q, x)]$ is marked, then mark $[P, Q]$ where ' x ' is an input symbol
REPEAT THIS UNTIL NO MORE MARKINGS CAN BE MADE
- 4) Combine all the Unmarked Pairs and make them a single state in the minimized DFA

Construction of Mealy Machine

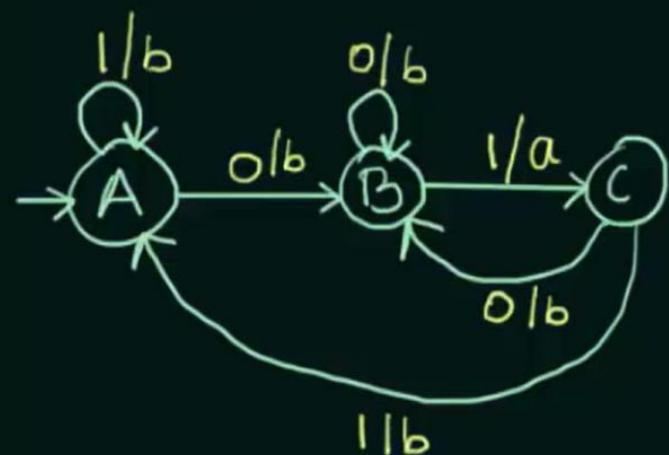
Ex-1) Construct a Mealy Machine that produces the 1's Complement of any binary input string.



10100
01011

Ex-2) Construct a Mealy Machine that prints 'a' whenever the sequence '01' is encountered in any input binary string.

$$\Sigma = \{0, 1\} \quad \Delta = \{a, b\}$$



0110
b a bb



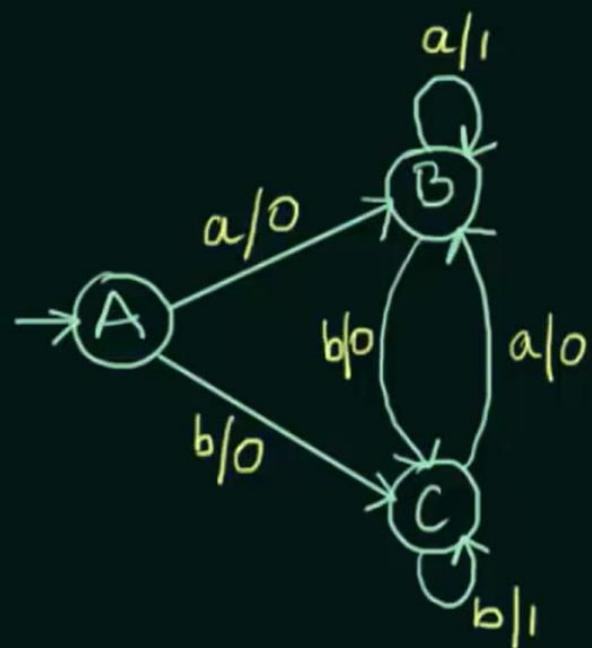
1000
b b b b

Construction of Mealy Machine - Examples (Part-1)

Design a Mealy Machine accepting the language consisting of strings from Σ^* , where $\Sigma = \{a,b\}$ and the strings should end with either aa or bb

aa - 1

bb - 1



ab**b**

baa

ba
oo

aa
01

Construction of Mealy Machine -Examples (Part-2)

Construct a Mealy Machine that gives 2's Complement of any binary input. (Assume that the last carry bit is neglected)

$$2' \text{ complement} = 1^s \text{ complement} + 1$$

M>B ← LSB

$$\begin{array}{r} \text{Eg. } 10100 \\ 1^s. \quad 01011 \end{array}$$

$$\begin{array}{r} +1 \\ \hline \end{array}$$

$$2^s = 01100$$

$$\begin{array}{r} \text{Eg. } 11(00 \\ 1^s. \quad 00011 \end{array}$$

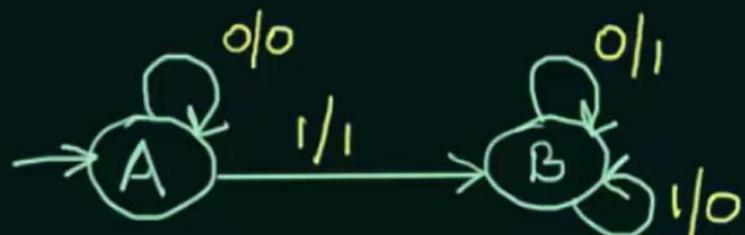
$$\begin{array}{r} +1 \\ \hline \end{array}$$

$$2^s = 00100$$

$$\begin{array}{r} \text{Eg. } 111(1 \\ 1^s. \quad 00000 \end{array}$$

$$\begin{array}{r} +1 \\ \hline \end{array}$$

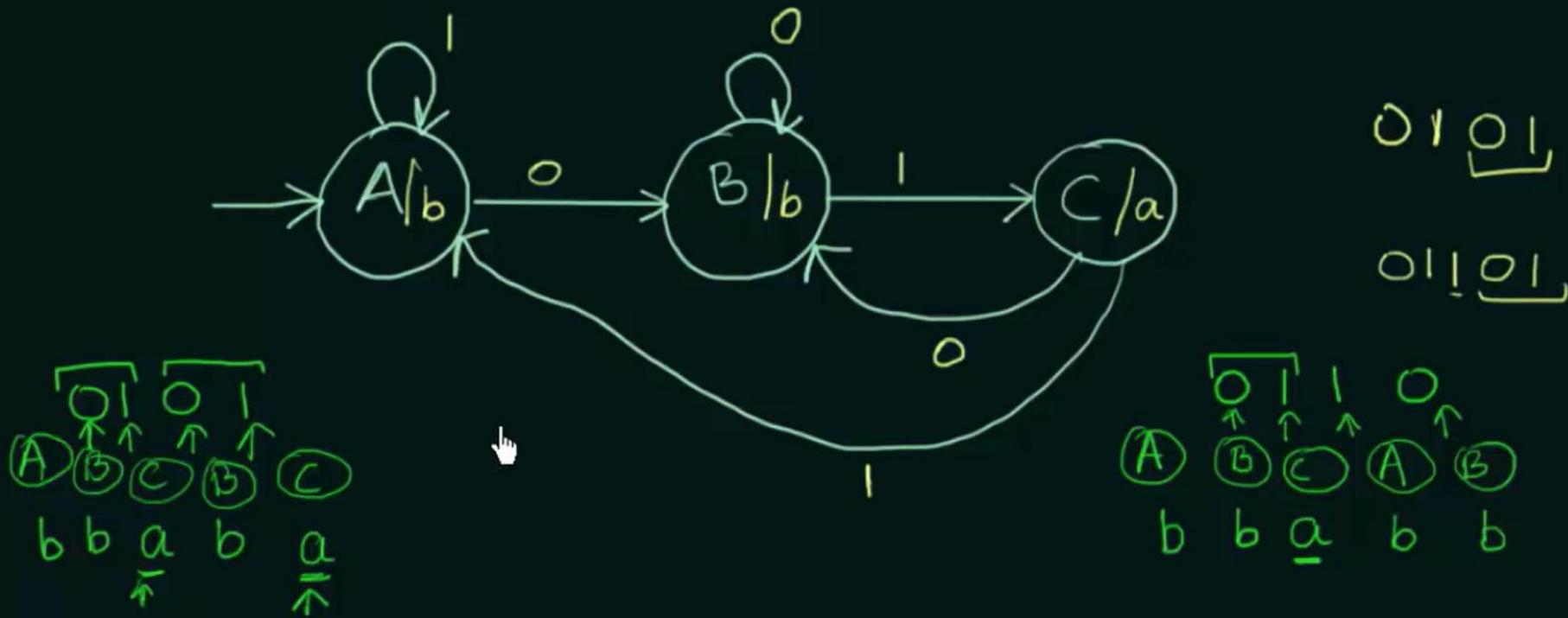
$$2^s = 00011$$



Construction of Moore Machine

Construct a Moore Machine that prints 'a' whenever the sequence '01' is encountered in any input binary string

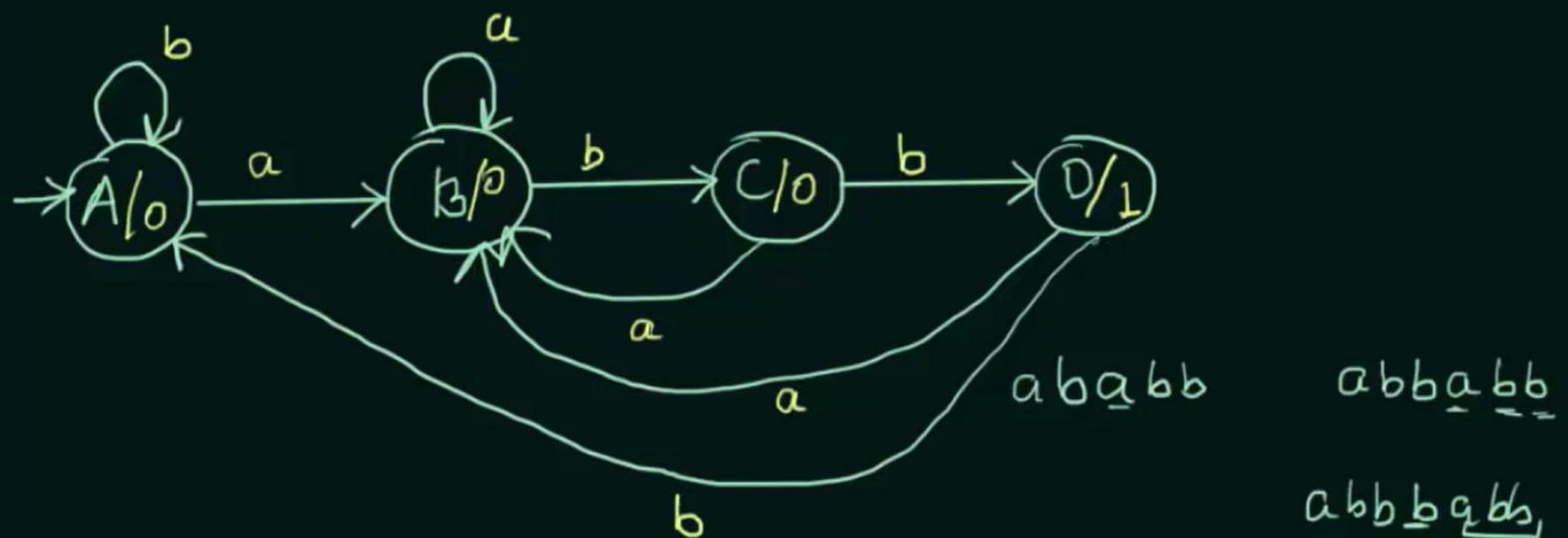
$$\begin{aligned}\Sigma &= \{0, 1\} \\ \Delta &= \{a, b\}\end{aligned}$$



Construction of Moore Machine - Examples (Part-1)

Construct a Moore Machine that counts the occurrences of the sequence 'abb' in any input strings over {a,b}

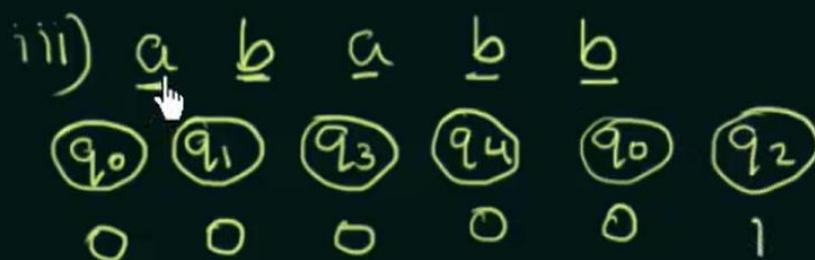
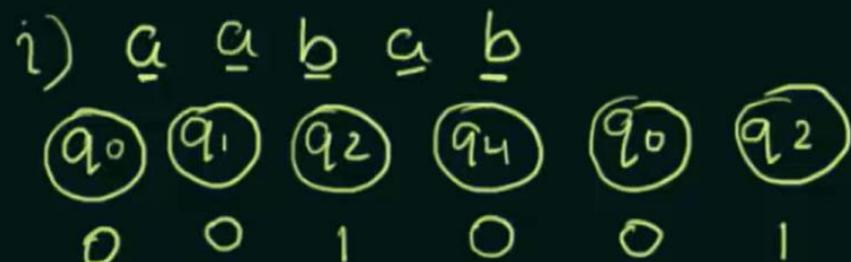
$$\Sigma = \{a, b\} \quad \Delta = \{0, 1\}$$



For the following Moore Machine the input alphabet is $\Sigma = \{a,b\}$ and the output alphabet is $\Delta = \{0,1\}$. Run the following input sequences and find the respective outputs:

(i) aabab (ii) abbb (iii) ababb

States	a	b	Outputs
$\rightarrow q_0$	q_1	q_2	0
q_1	q_2	q_3	0
q_2	q_3	q_4	1
q_3	q_4	q_4	0
q_4	q_0	q_0	0



Lec-53: Closure Properties of CFL (Context Free Languages) with explanation in Hindi



Union ✓

Concatenation ✓

Kleene closure ✓

Intersection X

CFL ✓

Complementation X

$$L_1 \cap L_2 \\ a^n b^n c^m n a^m b^n c^n \\ \underline{a^n b^n c^n} \rightarrow \underline{CSL}$$

$$L_1 \rightarrow G_1 \quad L_2 \rightarrow G_2 \quad S_1 \rightarrow aS_1b / \lambda \\ L_1^* \rightarrow G_1 \rightarrow S_1 \quad S \rightarrow S_1 / S_2 \\ S \rightarrow \underline{S_1 S_2} / \lambda \quad = \quad S \rightarrow S_1 \cdot S_2 \\ S_1, \quad S_2 \\ a^* = 0, a, \overline{aa}, \overline{aaa}, \dots$$

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}} \\ = \overline{\overline{CFL} \cup \overline{CFL}} \\ = \overline{CFL} = \text{CFL}$$

Lec-47: What is Context free grammar in TOC | Formal Definition



Formal Definition of CFG

CFG : (V, T, P, S)

V: finite set of Variables (Non Terminal)

T: finite set of Terminals ($V \cap T = \emptyset$)

P: Production Rules (Substitution Rules)

S: Start Variable

$$\begin{aligned}S &\rightarrow OA \\A &\rightarrow AO \mid O\end{aligned}$$

$$S \rightarrow OS_1 \mid \epsilon$$

Type 2:

$$\begin{array}{c}\alpha \rightarrow \beta \\ \text{Only one } (\text{Variable}) \quad (V \cup T)^* \quad \{S\} \\ \downarrow \qquad \qquad \qquad \{S\} \\ \{0, 1, \epsilon\}\end{array}$$

$$\begin{array}{c}OS_1 \\ \downarrow \\ OS_1 \\ \downarrow \\ OS_1 \\ \downarrow \\ OS_1\end{array}$$

SUBSCRIBE



6:45 / 7:56 • Context free grammar >



Lec-49: Left Most & Right Most Derivation in CFG | TOC



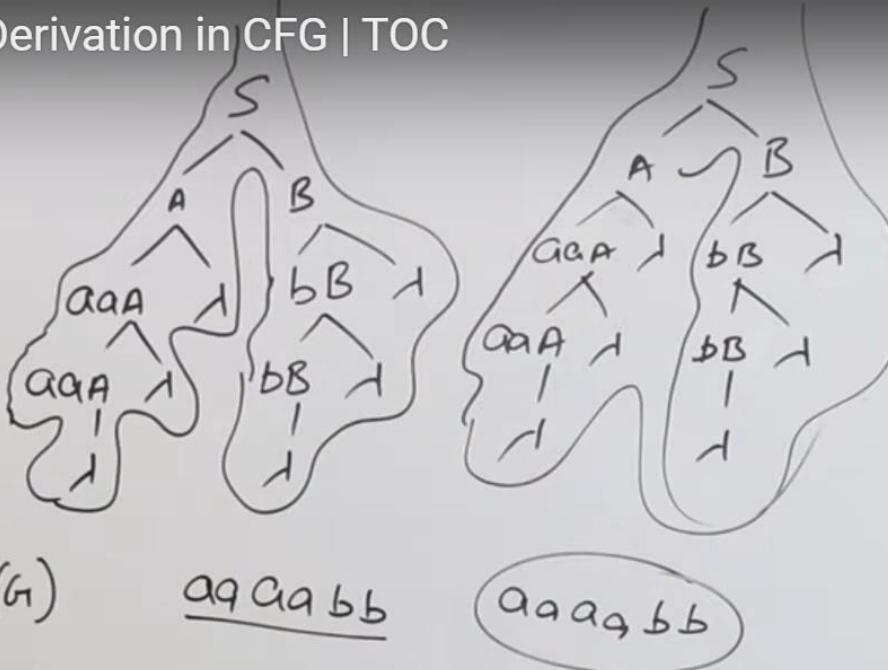
Derivation

LMD RMD

$$\begin{aligned} S &\rightarrow AB \\ A &\rightarrow aaA \mid \lambda \\ B &\rightarrow bB \mid \lambda \end{aligned}$$

Check whether

$$w = \underline{aaaabb} \in L(G)$$



◀ ▶ 🔍 SUBSCRIBE



5:52 / 6:21 • Derivation >



Lec-54: Remove Null Production from CFG (Context Free Grammar) with example in ...



Elimination of Nullable Production

$$S \rightarrow aS / a / \epsilon$$

$$a^n \ n \geq 0 \quad \{A, S\} \quad \begin{matrix} aS \\ \text{as} \end{matrix}$$
$$a^n \ n > 0 \quad L(G) = \epsilon$$

$$S \rightarrow ABC / BC / AC / C$$

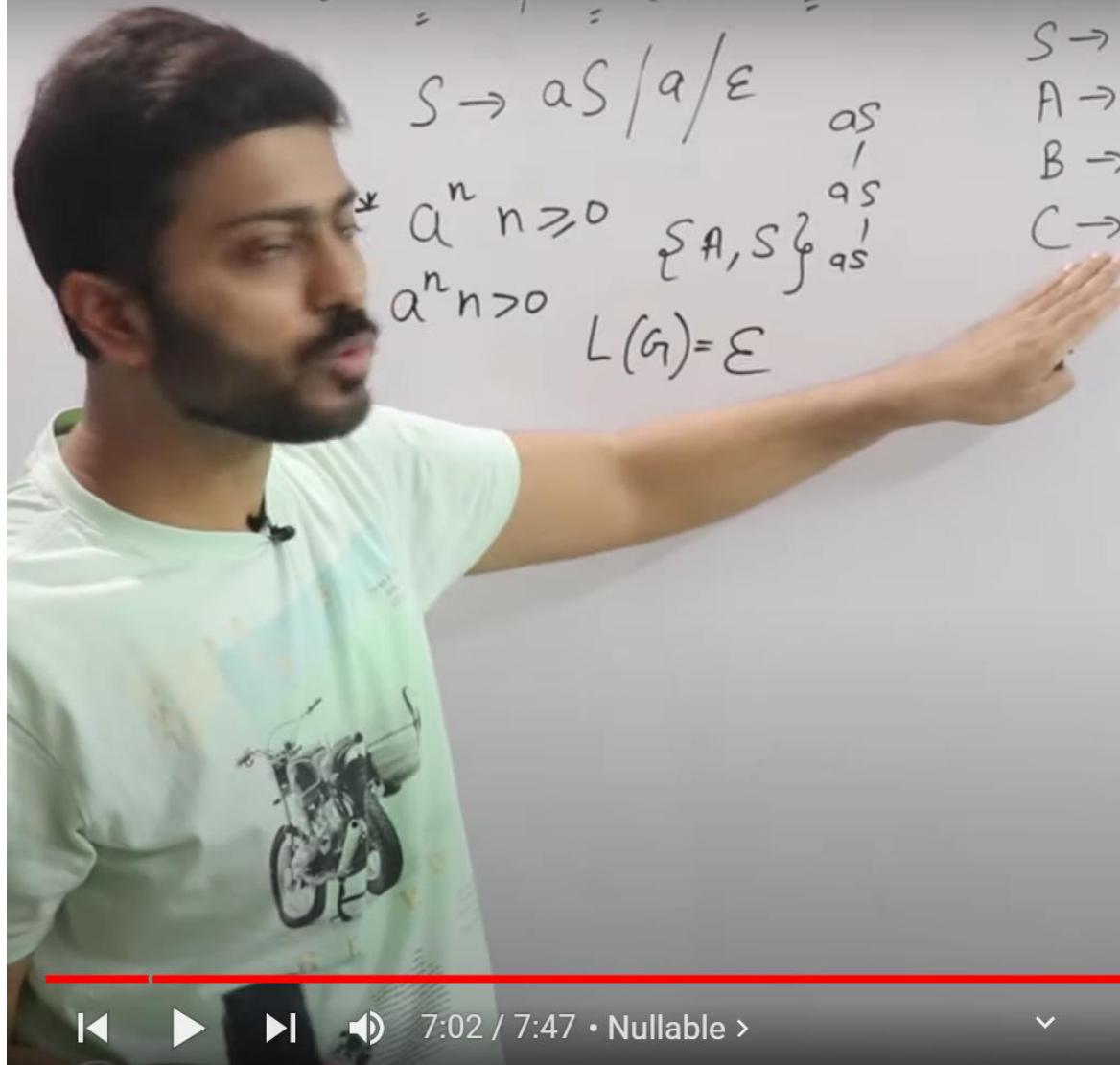
$$A \rightarrow aA / a$$

$$B \rightarrow bB / b$$

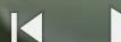
$$C \rightarrow c$$

Nullable

$$\{A, B\}$$



SUBSCRIBE



7:02 / 7:47 • Nullable >



Course: CONT_20CST-353 :: THEO X Lec-55: Remove Unit Production X toc - Google Drive X

youtube.com/watch?v=BLt4pJRBZdA&list=PLxCzCOWd7aiFM9Lj5G9G_76adtyb4ef7i&index=55

Elimination of Unit Production

$S \rightarrow AB$
 $A \rightarrow a$
 $B \rightarrow a/b$

$B \rightarrow C$
 $C \rightarrow D$
 $D \rightarrow E$

$\alpha \rightarrow \beta$
 $\alpha, \beta \in V$

$S \rightarrow aA/bba/ba/bb$
 $A \rightarrow ba/bb$

$B \rightarrow A$
 $ba \quad bb$

TOC(Theory of Computation)
900K+ Views
Syllabus explanation for GATE, NTA NET & other Competitive exam
68 videos

Lec-56: Introduction to Turing Machine and its Definition in Hindi...
Introduction to Turing Machine 9:03

THEORY OF COMPUTATION

52 53 54 55 56

Gate Smashers

Closure pro... of 8

Remove N... Protection

Remove U... Protection

400K+ Views Introduction to Turing

Lec-55: Remove Unit Production from CFG(Context Free Grammar) in Hindi



Type here to search



22:54
17-04-2023

Lec-64: CNF Vs GNF | Chomsky vs Greibach Normal Form | CFG in TOC



CNF

(Chomsky Normal Form)

$$\begin{aligned} 1) \quad A &\rightarrow BC \\ \text{OR} \\ A &\rightarrow a \end{aligned}$$

$$\left. \begin{array}{l} S \rightarrow AB \\ A \rightarrow BB/a \\ B \rightarrow AB/b \\ \{ A, B, C \in \text{Variables} \\ a \in \text{Terminal} \} \end{array} \right\}$$

2) No. of steps required to generate a string of length 'n' is " $2|n|-1$ "

$2 \times 2 - 1 = 3$ ab A → AB
 3) Used in Membership Algo. CYK. A → a
B → b

4) Derivation or Parse tree obtained from CNF is always Binary Tree.

5) Length of each Production Restricted



A
↓
 aB
↓
 aB
↓
 a_b

GNF

(Greibach Normal Form)

$$\begin{aligned} 1) \quad S &\rightarrow aABCDEF \\ A &\rightarrow a \end{aligned}$$

where $\left\{ \begin{array}{l} x \in V^* \\ a \in T \end{array} \right\}$

2) No. of steps required to generate a string of length 'n' is 'n'

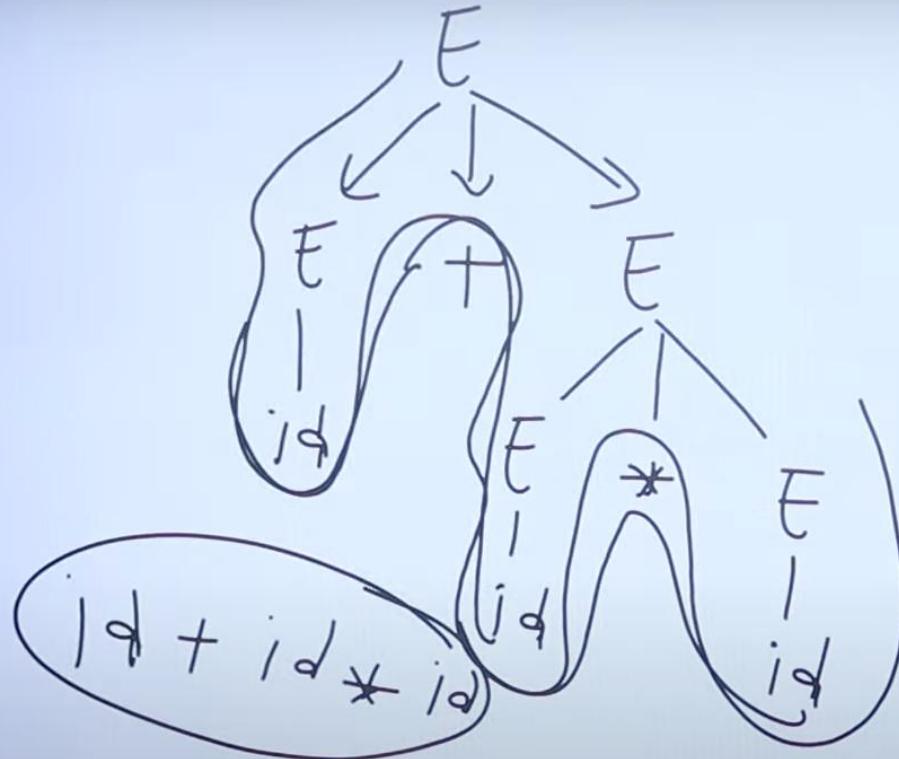
3) Used to Convert CFG to PDA

4) Not always $A \rightarrow aB$
 $B \rightarrow b$

5) Not Restricted A
 \downarrow
 aB
 \downarrow
 b

Settings SUBSCRIBE

Lec-65: Derivation Tree Parse Tree with example in TOC & Compiler design



SUBSCRIBE

Recursive CFG generates Infinite number of strings.

for example:

$$1) \underline{S} \rightarrow \underline{S}a$$

$$\begin{array}{c} S \\ \downarrow \\ S-a \\ \downarrow \\ S-a \\ \downarrow \\ S-a \end{array}$$

$\{ b, ba, \dots \}$

$$2) \underline{S} \rightarrow a\underline{S}$$

$$\begin{array}{c} S \\ \downarrow \\ S-b \\ \downarrow \\ S-b \\ \downarrow \\ S-b \end{array}$$

$\{ b, ba, bba, \dots \}$

$$S \rightarrow b$$

$$\begin{array}{c} S \\ \downarrow \\ S-b \\ \downarrow \\ S-b \\ \downarrow \\ S-b \end{array}$$

$\{ b, ab, aab, \dots \}$

$$3) \underline{S} \rightarrow \underline{A}a$$

$$\underline{A} \rightarrow \underline{A}b | c$$

$$\begin{array}{c} A \\ \downarrow \\ A-a \\ \downarrow \\ A-a \\ \downarrow \\ A-a \end{array}$$

$$\begin{array}{c} A \\ \downarrow \\ A-b \\ \downarrow \\ A-b \\ \downarrow \\ A-b \end{array}$$

$\{ a, ab, \dots \}$

$$\begin{array}{c} A \\ \downarrow \\ A-c \\ \downarrow \\ A-c \\ \downarrow \\ A-c \end{array}$$

$\{ a, ac, \dots \}$

- Non-Recursive CFG generates finite number of strings.

For example:

$$\underline{S} \rightarrow \underline{A}a$$

$$\underline{A} \rightarrow b | c$$

$$S \rightarrow \begin{array}{c} A \\ \downarrow \\ b \\ \downarrow \\ c \end{array}$$

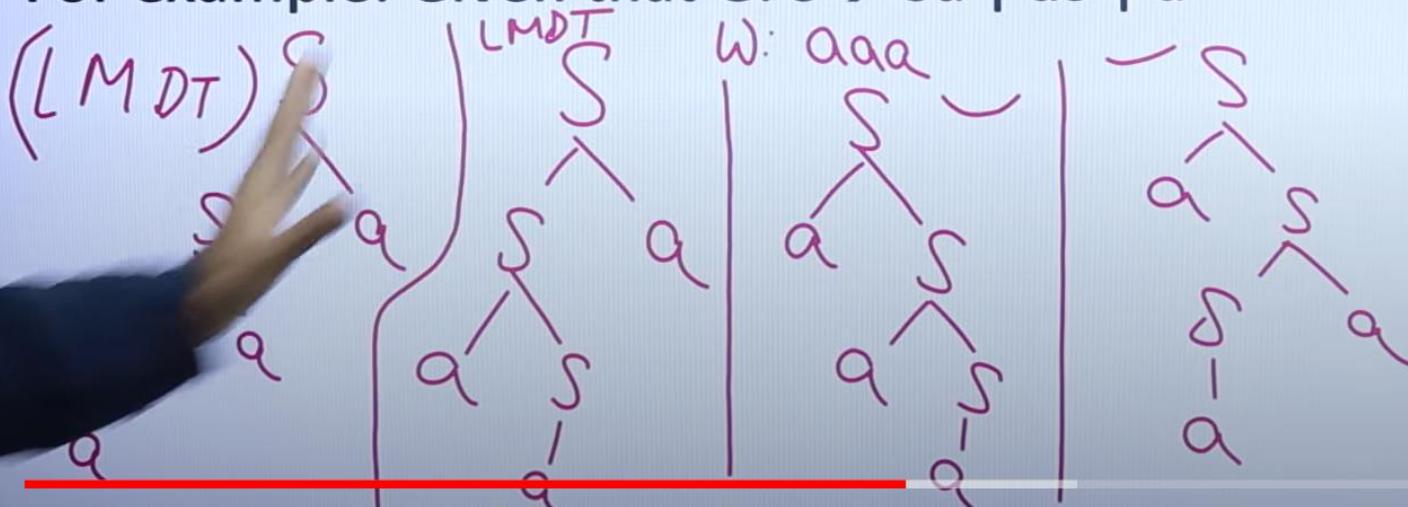
$$\{ ba, ca \}$$



Ambiguous & Unambiguous Grammar

- For ambiguous there exists **more than one** derivation for any word that belongs to Grammar.
- For unambiguous there exist **exactly one** derivation for any word that belongs to Grammar.

For example: Given that $G: S \rightarrow Sa \mid as \mid a$



More Example

G1: S → AB

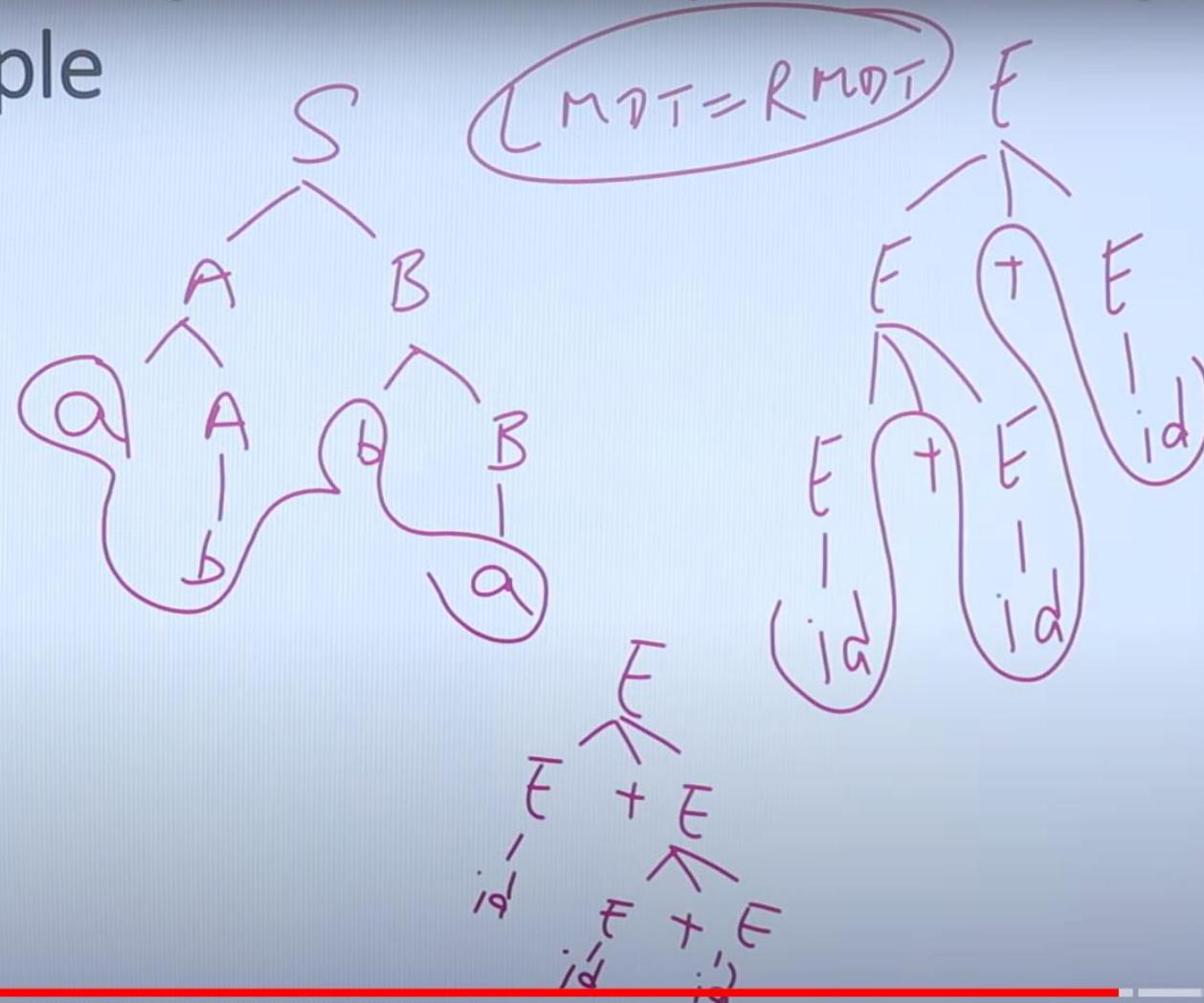
$A \rightarrow aA | b$

$B \rightarrow bB | a$

W: abba

G2: E → E + E | id

W: id+id+id



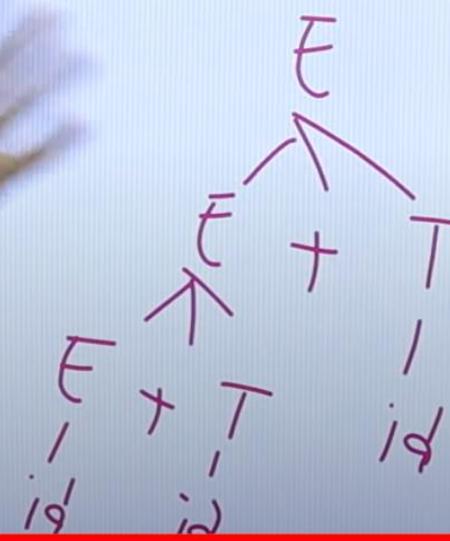


Conversion of Ambiguous to Unambiguous

$$G: E \rightarrow E + E \mid id \rightarrow E \rightarrow E + T / T$$

Where w= id+id+id

$$T \rightarrow id$$



☰ Context Free Grammar & Context Free Language



Σ = Set of Terminal Symbols

S = Start Symbol

P = Production Rule

Context Free Grammar has Production Rule of the form

$$A \rightarrow a$$

where, $a = \{V \cup \Sigma\}^*$ and $A \in V$

Example: For generating a language that generates equal number of a's and b's in the form $a^n b^n$, the Context Free Grammar will be defined as

$$G = \{ (S, A), (a, b), (S \rightarrow aAb, A \rightarrow aAb | \epsilon) \}$$

$$S \rightarrow a \underline{A} b$$

$$\rightarrow a a \underline{A} b b \quad (\text{by } A \rightarrow aAb)$$

$$\rightarrow aa a \underline{A} b b b \quad ("")$$

$$\rightarrow aaa bbb \quad (\text{by } A \rightarrow \epsilon)$$

$$\Rightarrow a^n b^n$$



☰ Context Free Grammar & Context Free Language



Σ = Set of Terminal Symbols

S = Start Symbol

P = Production Rule

Context Free Grammar has Production Rule of the form

$$A \rightarrow a$$

where, $a = \{V \cup \Sigma\}^*$ and $A \in V$

Example: For generating a language that generates equal number of a's and b's in the form $a^n b^n$, the Context Free Grammar will be defined as

$$G = \{ (S, A), (a, b), (S \rightarrow aAb, A \rightarrow aAb | \epsilon) \}$$

$$S \rightarrow a \underline{A} b$$

$$\rightarrow a a \underline{A} b b \text{ (by } A \rightarrow aAb \text{)}$$

$$\rightarrow aa a \underline{A} b b b \text{ (")}$$

$$\rightarrow aaa bbb \text{ (by } A \rightarrow \epsilon \text{)}$$

$$\Rightarrow a^n b^n$$



- 1) Start with the Start Symbol and choose the closest production that matches to the given string.
- 2) Replace the Variables with its most appropriate production. Repeat the process until the string is generated or until no other productions are left.

Example: Verify whether the Grammar $S \rightarrow OB|1A$, $A \rightarrow O|OS|1AA|\hat{}$, $B \rightarrow 1|1S|OBB$ generates the string 00110101

$S \rightarrow OB \quad (S \rightarrow OB)$
 $\rightarrow OOB_B \quad (B \rightarrow OBB)$
 $\rightarrow OO1_B \quad (B \rightarrow 1)$
 $\rightarrow OO11_S \quad (B \rightarrow 1S)$
 $\rightarrow OO11_OB \quad (S \rightarrow OB)$
 $\rightarrow OO11_01_S \quad (B \rightarrow 1S)$
 $\rightarrow OO11_01_OB \quad (S \rightarrow OB)$
 $\rightarrow OO11_0101 \quad (B \rightarrow 1)$



Method to find whether a string belongs to a Grammar or not
Example: Verify whether the Grammar $S \rightarrow aAb$, $A \rightarrow aAb \mid \epsilon$ generates the string aabb

$S \rightarrow aAb$
 $\rightarrow a aAb b \quad (A \rightarrow aAb)$
 $\rightarrow a a b b \quad (A \rightarrow \epsilon)$
 $\rightarrow a a aAb b b \quad (A \rightarrow aAb)$
 $a a a b b b \quad (A \rightarrow \epsilon) \quad X$



11:18 / 11:30



A Grammar is said to be Ambiguous if there exists two or more derivation tree for a string w (that means two or more left derivation trees)

Example: $G = (\{S\}, \{a+b, +, *\}, P, S)$ where P consists of $S \rightarrow S+S | S*S | a | b$

The String $a+a*b$ can be generated as:

$$\begin{aligned} S &\rightarrow S+S \\ &\rightarrow a+\underline{S} \\ &\rightarrow a+S*\underline{S} \\ &\rightarrow a+a+\underline{S} \\ &\rightarrow a+a+\underline{a}+b \end{aligned}$$

$$\begin{aligned} S &\rightarrow \underline{S}+S \\ &\rightarrow \underline{S}+S*\underline{S} \\ &\rightarrow a+\underline{S}*\underline{S} \\ &\rightarrow a+a+\underline{S} \\ &\rightarrow a+a+\underline{a}*b \end{aligned}$$

Thus, this Grammar is Ambiguous



Example: Find a reduced grammar equivalent to the grammar G , having production rules
 $P: S \rightarrow AC|B, A \rightarrow a, C \rightarrow c|BC, E \rightarrow aA|e$



Phase 1: $T = \{a, c, e\}$

$$W_1 = \{A, C, E\}$$

$$W_2 = \{A, C, E, S\}$$

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

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

$$P: S \rightarrow AC, A \rightarrow a, C \rightarrow c, E \rightarrow aA|e$$

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

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



$$\omega_2 = \{A, C, E, S\}$$

$$\omega_3 = \{A, C, E, S\}$$

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

$$P: \quad S \rightarrow AC, \quad A \rightarrow a, \quad C \rightarrow c, \quad E \rightarrow aA|e$$

Phase 2 : $\gamma_1 = \{S\}$

$$\gamma_2 = \{S, A, C\}$$

$$\gamma_3 = \{S, A, C, a, c\}$$

$$\gamma_4 = \{S, A, C, a, c\}$$

$$Q'' = \{(A, C, S), \{a, c\}, P, \{S\}\}$$

$$P: \quad S \rightarrow AC, \quad A \rightarrow a, \quad C \rightarrow c$$



Simplification of Context Free Grammar

Removal of Unit Productions

Any Production Rule of the form $A \rightarrow B$ where $A, B \in \text{Non Terminals}$ is called Unit Production

Procedure for Removal

Step 1: To remove $A \rightarrow B$, add production $A \rightarrow x$ to the grammar rule whenever $B \rightarrow x$ occurs in the grammar. [$x \in \text{Terminal}$, x can be Null]

Step 2: Delete $A \rightarrow B$ from the grammar.

Step 3: Repeat from Step 1 until all Unit Productions are removed.

Example: Remove Unit Productions from the Grammar whose production rule is given by

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

$Y \rightarrow Z, Z \rightarrow M, M \rightarrow N$

i) Since $N \rightarrow a$, we add $M \rightarrow a$

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

Example: Remove Unit Productions from the Grammar whose production rule is given by

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

$Y \rightarrow Z$, $Z \rightarrow M$, $M \rightarrow N$

1) Since $N \rightarrow a$, we add $M \rightarrow a$

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

2) Since $M \rightarrow a$, we add $Z \rightarrow a$

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

3) Since $Z \rightarrow a$, we add $Y \rightarrow a$

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

Remove the Unreachable symbols

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

Simplification of Context Free Grammar

Removal of Null Productions

In a CFG, a Non-Terminal Symbol 'A' is a nullable variable if there is a production $A \rightarrow \epsilon$ or there is a derivation that starts at 'A' and leads to ϵ . (Like $A \rightarrow \dots \rightarrow \epsilon$)

Procedure for Removal:

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

Step 2: Replace each occurrences of 'A' in each of these productions with ϵ

Step 3: Add the resultant productions to the Grammar

Example: Remove Null Productions from the following Grammar

$S \rightarrow ABAC, A \rightarrow aA|\epsilon, B \rightarrow bB|\epsilon, C \rightarrow c$

$A \rightarrow \epsilon, B \rightarrow \epsilon$

i) To eliminate $A \rightarrow \epsilon$

$S \rightarrow ABAC$



→ 1. step

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

$$A \rightarrow aA$$

$$A \rightarrow a$$

New production: $S \rightarrow ABAC \mid ABC \mid BAC \mid BC$

$$A \rightarrow aA \mid a, \quad B \rightarrow bB \mid \epsilon, \quad C \rightarrow c$$

2) To eliminate $B \rightarrow \epsilon$

$$S \rightarrow AAC \mid AC \mid C, \quad B \rightarrow b$$

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

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

$$C \rightarrow c$$



Chomsky Normal Form



In Chomsky Normal Form (CNF) we have a restriction on the length of RHS; which is; elements in RHS should either be two variables or a Terminal.

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

$$A \rightarrow a$$

$$A \rightarrow BC$$

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



Steps to convert a given CFG to Chomsky Normal Form:

- Step 1: If the Start Symbol S occurs on some right side, create a new Start Symbol S' and a new Production $S' \rightarrow S$. 
- Step 2: Remove Null Productions. (Using the Null Production Removal discussed in previous Lecture)
- Step 3: Remove Unit Productions. (Using the Unit Production Removal discussed in previous Lecture)
- Step 4: Replace each Production $A \rightarrow B_1 \dots B_n$ where $n > 2$, with $A \rightarrow B_1 C$ where $C \rightarrow B_2 \dots B_n$
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 $A \rightarrow aB$ where 'a' is a terminal and A and B are non-terminals, then the Production is replaced by $A \rightarrow XB$ and $X \rightarrow a$.
Repeat this step for every Production which is of the form $A \rightarrow aB$

Conversion of CFG to Chomsky Normal Form

Convert the following CFG to CNF: P: $S \rightarrow ASA \mid aB, A \rightarrow B|S, B \rightarrow b \mid \epsilon$

1) Since S appears in RHS, we add a new State S' and $S' \rightarrow S$ is added to the production

P: $S' \rightarrow S, S \rightarrow ASA|aB, A \rightarrow B|S, B \rightarrow b \mid \epsilon$

2) Remove the Null Productions: $B \rightarrow \epsilon$ and $A \rightarrow \epsilon$:

After Removing $B \rightarrow \epsilon$: P: $S' \rightarrow S, S \rightarrow ASA|aB|a, A \rightarrow B|S| \epsilon, B \rightarrow b$

After Removing $A \rightarrow \epsilon$: P: $S' \rightarrow S, S \rightarrow ASA|aB|a|AS|SA|S, A \rightarrow B|S, B \rightarrow b$



2) Remove the Null Productions: $B \rightarrow \in$ and $A \rightarrow \in$:

After Removing $B \rightarrow \in$: P: $S' \rightarrow S, S \rightarrow ASA|aB|a, A \rightarrow B|S|\in, B \rightarrow b$

After Removing $A \rightarrow \in$: P: $S' \rightarrow S, S \rightarrow ASA|aB|a|AS|SA|S, A \rightarrow B|S, B \rightarrow b$

3) Remove the Unit Productions: $S \rightarrow S, S' \rightarrow S, A \rightarrow B$ and $A \rightarrow S$:

After Removing $S \rightarrow S$: P: $S' \rightarrow S, S \rightarrow ASA|aB|a|AS|SA, A \rightarrow B|S, B \rightarrow b$

After Removing $S' \rightarrow S$: P: $S' \rightarrow ASA|aB|a|AS|SA,$
 $S \rightarrow ASA|aB|a|AS|SA,$
 $A \rightarrow B|S, B \rightarrow b$

After Removing $A \rightarrow B$: P: $S' \rightarrow ASA|aB|a|AS|SA,$
 $S \rightarrow ASA|aB|a|AS|SA,$
 $A \rightarrow b|S, B \rightarrow b$

After Removing $A \rightarrow S$: P: $S' \rightarrow ASA|aB|a|AS|SA,$
 $S \rightarrow ASA|aB|a|AS|SA,$
 $A \rightarrow b|ASA|aB|a|AS|SA,$
 $B \rightarrow b$



4) Now find out the productions that has more than TWO variables in RHS

$S' \rightarrow ASA$, $S \rightarrow ASA$ and $A \rightarrow ASA$

After removing these, we get: P: $S' \rightarrow AX|aB|a|AS|SA$,

$S \rightarrow AX|aB|a|AS|SA$,

$A \rightarrow b|AX|aB|a|AS|SA$,

$B \rightarrow b$,

$X \rightarrow SA$

5) Now change the productions $S' \rightarrow aB$, $S \rightarrow aB$ and $A \rightarrow aB$

Finally we get:

P: $S' \rightarrow AX|YB|a|AS|SA$,

$S \rightarrow AX|YB|a|AS|SA$,

$A \rightarrow b|AX|YB|a|AS|SA$,

$B \rightarrow b$,

$X \rightarrow SA$,

$Y \rightarrow a$

which is the required Chomsky Normal Form for the given CFG



Greibach Normal Form

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

$$A \rightarrow b$$

$$A \rightarrow bC_1C_2 \dots C_n$$

where A, C_1, \dots, 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 (using the Unit & Null Productions removal techniques discussed in the previous lecture)

Step 2: Check whether the CFG is already in Chomsky Normal Form (CNF) and convert it to CNF if it is not. (using the CFG to CNF conversion technique discussed in the previous lecture)

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



~~Step 1: Check if the given CFG has any Unit Productions or Null Productions and Remove if there are any~~ Greibach Normal Form & CFG to GNF Conversion



Step 1: Check if the given *CFG* has any Unit Productions or Null Productions and Remove if there are any (using the Unit & Null Productions removal techniques discussed in the previous lecture)

Step 2: Check whether the *CFG* is already in Chomsky Normal Form (CNF) and convert it to CNF if it is not. (using the *CFG* to CNF conversion technique discussed in the previous lecture)

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

Example: $S \rightarrow CA \mid BB$
 $B \rightarrow b \mid SB$
 $C \rightarrow b$
 $A \rightarrow a$

Replace: S with A_1
 C with A_2
 A with A_3
 B with A_4

We get:

$$A_1 \rightarrow A_2 A_3 \mid A_4 A_4$$

$$A_4 \rightarrow b \mid A_1 A_4$$

$$A_2 \rightarrow b$$

$$A_3 \rightarrow a$$



$A_2 \rightarrow D$

$A_3 \rightarrow a$

Step 4: Alter the rules so that the Non-Terminals are in ascending order, such that,

If the Production is of the form $A_i \rightarrow A_j x$, then,

$i < j$ and should never be $i \geq j$

$A_4 \rightarrow b | \underline{A_1} A_4$

$A_4 \rightarrow b | \underline{A_2} A_3 A_4 | A_4 A_4 A_4$

$A_4 \rightarrow b | b A_3 A_4 | A_4 A_4 A_4$



Left Recursion

Step 5: Remove Left Recursion



Step 5: Remove Left Recursion

Introduce a New Variable to remove the Left Recursion

$$A_4 \rightarrow b \mid b A_3 A_4 \mid A_4 A_4 A_4$$

$$Z \rightarrow A_4 A_4 Z \mid A_4 A_4$$

$$A_4 \rightarrow b \mid b A_3 A_4 \mid b Z \mid b A_3 A_4 Z$$

Now the grammar is:

$$A_1 \rightarrow A_2 A_3 \mid A_4 A_4$$

$$A_4 \rightarrow b \mid b A_3 A_4 \mid b Z \mid b A_3 A_4 Z$$

$$Z \rightarrow A_4 A_4 \mid A_4 A_4 Z$$

$$A_2 \rightarrow b$$

$$A_3 \rightarrow a$$



$bA_3A_4Z \rightarrow bA_3A_4A_4 \mid bZ \mid bA_3A_4Z$

Now the grammar is:

$$A_1 \rightarrow A_2 A_3 \mid A_4 A_4$$

$$A_4 \rightarrow b \mid b A_3 A_4 \mid bZ \mid b A_3 A_4 Z$$

$$Z \rightarrow A_4 A_4 \mid A_4 A_4 Z$$

$$A_2 \rightarrow b$$

$$A_3 \rightarrow a$$

$$A_1 \rightarrow bA_3 \mid bA_4 \mid bA_3A_4A_4 \mid bZA_4 \mid bA_3A_4ZA_4$$

$$A_4 \rightarrow b \mid bA_3A_4 \mid bZ \mid bA_3A_4Z$$

$$Z \rightarrow bA_4 \mid bA_3A_4A_4 \mid bZA_4 \mid bA_3A_4ZA_4 \mid \\ bA_4Z \mid bA_3A_4A_4Z \mid bZA_4Z \mid bA_3A_4ZA_4Z$$

$$A_2 \rightarrow b$$

$$A_3 \rightarrow a$$

