

(1)

Module - 3

Context-Free Grammars (CFG)

A context-free grammar (or CFG) to be a grammar in which each rule must:

- * have a left-hand side that is a single nonterminal
- * have a right-hand side that is ϵ or terminal; nonterminal one or more.

A context grammar G is 4 tuple (or) Quad tuple

$$G = (V, T, P, S)$$

where

V is set of variables/non terminals

T is set of terminals

P is set of production

Each production is of the form $\alpha \rightarrow \beta$

where α is a string from variable & α cannot be ϵ , but β is string from $(VUT)^*$. Hence it can include ϵ also.

(Eg)

Obtain grammar to generate string consisting of any no of a's



$$\delta(S, a) = S \quad S \rightarrow aS \quad S \rightarrow \epsilon$$

$$\downarrow \quad \downarrow$$

$$S \rightarrow aS$$

$$S \rightarrow \epsilon$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

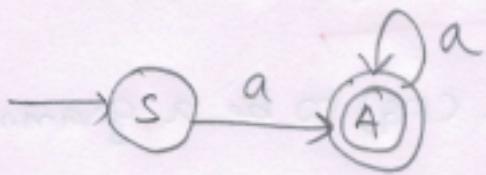
$$T = \{a\}$$

$$P = \{S \rightarrow aS, S \rightarrow \epsilon\}$$

S is start symbol.

Eg Obtain a grammar to generate string consisting of at least 1a.

Sol?



$$\delta(S, a) = A$$

$$S \rightarrow aA$$

$$\delta(A, a) = A$$

$$A \rightarrow aA$$

$$A \rightarrow \epsilon$$

$$G = (V, T, P, S)$$

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

$$T = \{a\}$$

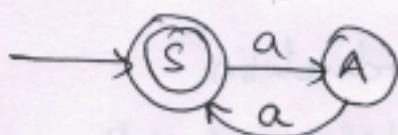
$$P = \{S \rightarrow aA\}$$

$$A \rightarrow aA / \epsilon$$

S is start symbol.

Eg Obtain a grammar to generate string consisting of even no of a's

Sol?



$$\delta(S, a) = A$$

$$S \rightarrow aA$$

$$S \rightarrow \epsilon$$

$$\delta(A, a) = S$$

$$A \rightarrow aS$$

$$G = (V, T, P, S)$$

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

$$T = \{a\}$$

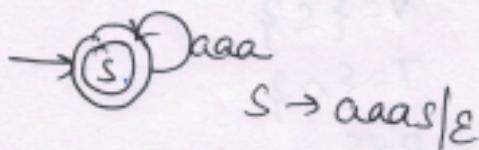
$$P = \{S \rightarrow aA / \epsilon\}$$

$$A \rightarrow aS$$

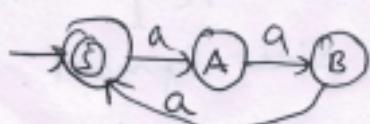
S is start symbol.

Eg Obtain a grammar to generate string consisting of multiples of 3 a's.

Sol?



or



$$G = (V, T, P, S)$$

$$V = \{S\}$$

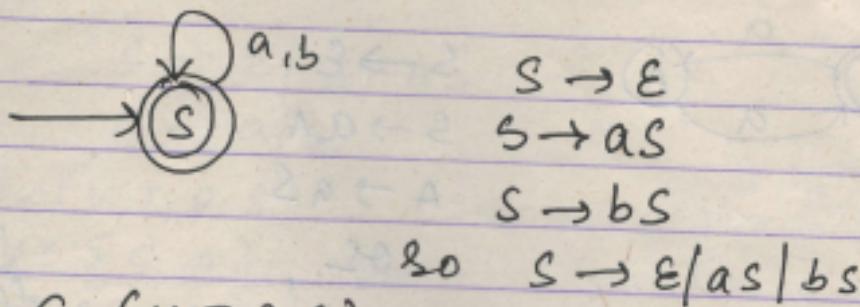
$$T = \{a\}$$

$$P = \{S \rightarrow aaa / aaaa / \epsilon\}$$

S is start symbol.

③ obtain grammar to generate string consisting of any number of a's & b's

SOTM



$$G = (V, T, P, S)$$

$$V = \{ S \}$$

$$T = \{ a, b \}$$

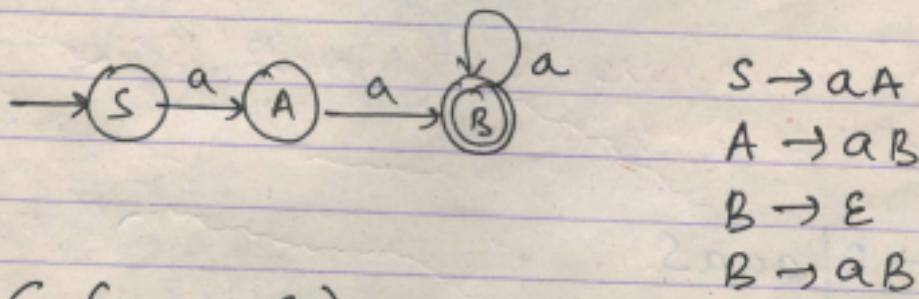
$$P = \{$$

$$S \rightarrow E/aS/bS \}$$

S is start symbol

④ obtain grammar to generate string consisting of at least two a's

SOTM



$$G = (V, T, P, S)$$

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

$$T = \{ a \}$$

$$P = \{$$

$$S \rightarrow aA$$

$$A \rightarrow aB$$

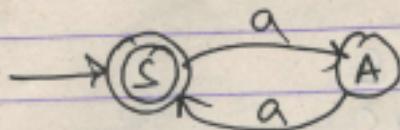
$$B \rightarrow E/aB \}$$

$$B \rightarrow aB \}$$

S. is the start symbol

(eg) obtain grammar to generate string consisting of even number of a's.

Sol:



$$S \rightarrow \epsilon$$

$$S \rightarrow aA$$

$$A \rightarrow aS$$

or

$$S \rightarrow \epsilon / aas \Rightarrow \textcircled{1}$$

$$G = (V, T, P, S)$$

$$V = \{S, a\}$$

$$T = \{a\}$$

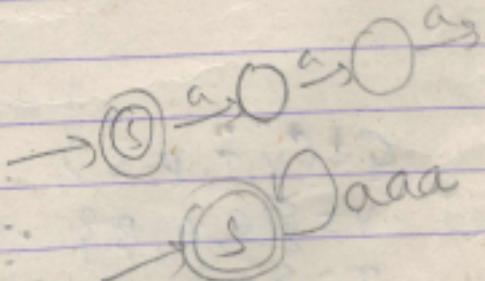
$$P = \left\{ \begin{array}{l} S \rightarrow \epsilon / aA \\ A \rightarrow aS \end{array} \right\}$$

where S is start symbol.

(eg) obtain grammar to generate string consist of multiples of three a's.

Sol:

$$S \rightarrow \epsilon / aaas$$



$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a\}$$

$$P = \left\{ \begin{array}{l} S \rightarrow \epsilon / aaas \end{array} \right\}$$

where S is start symbol.

Eg) obtain grammar to generate strings of a's & b's such that string length is multiple of 3

Soln

$$S \rightarrow \epsilon / A A A S$$

$$A \rightarrow a / b$$

$$G = (V, T, P, S)$$

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

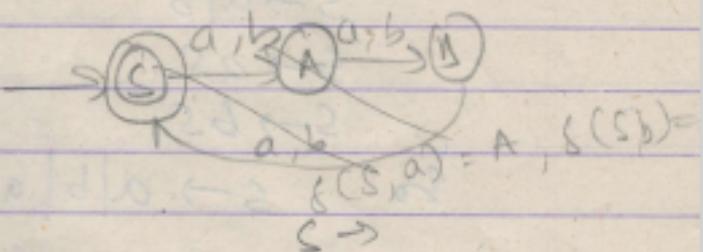
$$T = \{a, b\}$$

$$P = \{$$

$$S \rightarrow \epsilon / A A A S$$

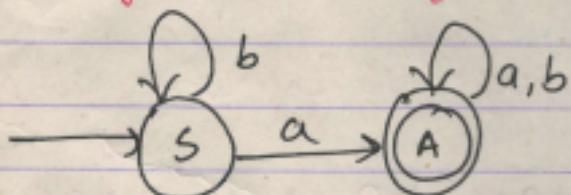
$$A \rightarrow a / b ?$$

S is the start symbol.



Eg) obtain grammar to generate string consisting of any number of a's & b's with atleast one a.

Soln



$$S \rightarrow b S$$

$$S \rightarrow a A$$

$$A \rightarrow \epsilon / a A / b A$$

$$G = (V, T, P, S)$$

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

$$T = \{a, b\}$$

$$P = \{ S \rightarrow a A / b S$$

$$A \rightarrow a A / b A / \epsilon \}$$

S is the start symbol.

(eg)

obtain grammar to generate string consisting
of any no of a's & b's with at least one a
or at least one b.

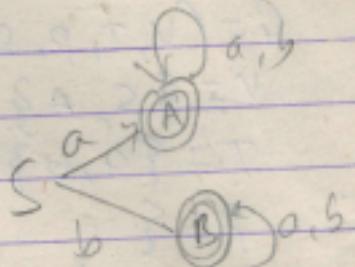
sol

$$S \rightarrow a \mid b$$

$$S \rightarrow aS$$

$$S \rightarrow bS$$

$$\text{so } S \rightarrow a \mid b \mid aS \mid bS$$



$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{ S \rightarrow a \mid b \mid aS \mid bS \}.$$

S is the start symbol.

$$S \rightarrow aA \mid bB$$

$$\begin{aligned} S &\rightarrow aA \\ A &\rightarrow aA \mid bA \mid \epsilon \end{aligned}$$

$$\begin{aligned} S &\rightarrow bB \\ B &\rightarrow aB \mid bB \mid \epsilon \end{aligned}$$

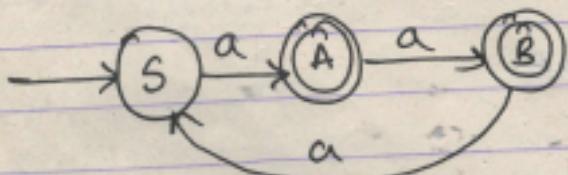
(eg)

Obtain grammar to accept the following language

$$L = \{ w : |w| \bmod 3 > 0 \text{ where } w \in \{a\}^*\}$$

sol:

$$S \rightarrow a \mid aa \mid aaaS$$



$$S \rightarrow aA$$

$$A \rightarrow aB \mid \epsilon$$

$$B \rightarrow aS \mid \epsilon$$

$$1 \bmod 3 = 1$$

$$G = (V, T, P, S)$$

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

$$T = \{ a \}$$

$$P = \{ S \rightarrow aA \\ A \rightarrow aB / \epsilon \\ B \rightarrow aS / \epsilon \}$$

S is the start symbol.

Grammars from Regular Expressions

- (Q) obtain grammar to generate strings of a's & b's having a substring ab.
- Soln: The RE representing strings of a's & b's having a substring is given by

$$(a+b)^* ab (a+b)^*$$

↓ ↓ ↓
 A ab A

$$S \rightarrow AabA$$

$$A \rightarrow \epsilon | aA | bA$$

$$G = (V, T, P, S)$$

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

$$T = \{ ab \}$$

$$P = \{$$

$$S \rightarrow AabA$$

$$A \rightarrow \epsilon | aA | bA \}$$

S is start symbol.

- (Q) obtain grammar to generate string of a's ending with string ab.

Soln

$$(a+b)^* ab$$

↓ ↓
 A ab

$$S \rightarrow Aab$$

$$A \rightarrow \epsilon | aA | bA$$

$$G = (V, T, P, S)$$

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

$$T = \{a, b\}$$

$$P = \{$$

$$S \rightarrow Aab$$

$$A \rightarrow aA | bA | \epsilon\}$$

• S is the start symbol.

Q. obtain grammar to generate strings of a's & b's starting with ab.

Soln

$$ab(a+b)^*$$

$$\Downarrow \quad \Downarrow$$

$$ab \quad A$$

$$S \rightarrow abA$$

$$A \rightarrow \epsilon | aA | bA$$

$$G = (V, T, P, S)$$

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

$$T = \{a, b\}$$

$$P = \{ S \rightarrow abA$$

$$A \rightarrow \epsilon | aA | bA \}$$

S is the start symbol.

Q. obtain the grammar to generate the following language

$$L = \{w : n_a(w) \bmod 2 = 0 \text{ where } w \in \{a, b\}^*\}$$

Sol1

$$\begin{array}{ccc}
 b^* a b^* a b^* & b^* a b^* a b^* \\
 \downarrow \downarrow \downarrow \downarrow \downarrow & S \Rightarrow A a A + b A + C \\
 s a s a s & A \rightarrow \epsilon / b A \\
 S \rightarrow s a s a s & \xleftarrow{\quad} \xrightarrow{\quad} \\
 S \rightarrow \epsilon / b S & \\
 S \rightarrow \epsilon / b S / s a s a s
 \end{array}$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{ S \rightarrow \epsilon | b S | s a s a s \}$$

s is start symbol.

Grammars for other languages

- (Q) Obtain a grammar to generate the following language

Sol2

$$\begin{array}{ccc}
 L = \{a^n b^n | n \geq 0\} & S \Rightarrow \epsilon \\
 & S \Rightarrow a S b \\
 S \rightarrow \epsilon | a S b & \Rightarrow ab
 \end{array}$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{ S \rightarrow \epsilon | a S b \}$$

s is the start symbol

$$\begin{array}{c}
 S \Rightarrow a S b \\
 \Rightarrow a a S b b \\
 \Rightarrow a a a b b b
 \end{array}$$

(Q)

obtain a grammar to generate the following language

$$L = \{a^n b^n : n \geq 1\}$$

Soln.

$$L = \{ab, aabb, aaabbb, \dots\}$$

$$S \rightarrow ab | aSb$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{S \rightarrow ab | aSb\}$$

S is the start symbol.

(Q) obtain a grammar to generate the following lang

$$L = \{a^{n+1} b^n : n \geq 0\}$$

Soln.

$$L = \{a^1 ab, a^2 abb, a^3 aabb, \dots\}$$

$$S \rightarrow a | aSb$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{S \rightarrow a | aSb\}$$

S is start symbol.

(Ex) obtain a grammar to generate the following lang.

$$L = \{a^nb^{n+1} : n \geq 0\}$$

Sol:

$$S \rightarrow b \mid aSb$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{S \rightarrow b \mid aSb\}$$

S is start symbol.

(Ex) obtain a grammar to generate the following lang.

$$L = \{a^n b^{n+2} : n \geq 0\}$$

Sol: Two extra b's should be generated. So, we find grammar to generate given lang.

$$S \rightarrow bb \mid aSb$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{S \rightarrow bb \mid aSb\}$$

S is start symbol.

(e) obtain a grammar to generate the following language

$$L = \{a^n b^{2n} : n \geq 0\}$$

Sol:

$$S \rightarrow \epsilon | aSbb$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{S \rightarrow \epsilon | aSbb\}$$

S is start symbol.

(e) Let $\Sigma = \{a, b\}$ obtain a grammar G generating set of all palindromes over Σ .

Sol:

$$S \rightarrow \epsilon | a | b | aSa | bSb$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{S \rightarrow \epsilon$$

$$S \rightarrow a | b$$

$$S \rightarrow aSa | bSb\}$$

S is the start symbol.

Made by

SAMSAM | nsa

(eg) obtain a grammar to generate a lang consisting of all non-palindromes over $\{a, b\}$

Q. Sol?

$$G = (V, T, P, S)$$

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

$$T = \{a, b\}$$

$$\begin{aligned} P = \{ & \quad S \rightarrow aSa \mid bSb \\ & \quad S \rightarrow A \\ & \quad A \rightarrow aBb \mid bBa \\ & \quad B \rightarrow aB \mid bB \mid \epsilon \} \end{aligned}$$

S is start symbol.

(eg) obtain a grammar to generate the following language

$$L = \{WW^R \text{ where } W \in \{a, b\}^*\}.$$

$$L = \{aa, bb, abba, baab, \dots\}$$

Q. Sol?

$$S \rightarrow \epsilon \mid aSa \mid bSb$$

$$G = (V, T, P, S)$$

$$V = \{S\}$$

$$T = \{a, b\}$$

$$P = \{S \rightarrow \epsilon \mid aSa \mid bSb\}$$

S is start symbol.

(eg) Obtain the grammar to generate the language

$$L = \{0^m 1^n 2^n \mid m \geq 1 \text{ & } n \geq 0\}$$

Soln

$$L = \{ \underbrace{0^m 1^n}_{} 2^n \mid m \geq 1 \text{ & } n \geq 0\}$$

$$S \rightarrow A B$$

The variable A should produce any no of 0's followed by n no of 1's.

$$A \rightarrow 01 \mid 0A1$$

B should produce any no of 2's

$$B \rightarrow \epsilon \mid 2B$$

$$S \rightarrow A B$$

$$A \rightarrow 01 \mid 0A1$$

$$B \rightarrow \epsilon \mid 2B$$

$$G = (V, T, P, S)$$

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

$$T = \{ 0, 1, 2 \}$$

$$P = \{ S \rightarrow A B, \\ A \rightarrow 01 \mid 0A1, \\ B \rightarrow \epsilon \mid 2B \}$$

S is start symbol.

(Eg) obtain the grammar to generate the lang

$$L = \{ w \mid n_a(w) = n_b(w) \}$$

Soln $S \rightarrow \epsilon \mid aSb \mid bSa \mid SS$ string starting ie with same str

Eg: abba, bhab

$$G = (V, T, P, S)$$

$$V = \{ S \}$$

$$T = \{ a, b \}$$

$$P = \{$$

$$S \rightarrow \epsilon \mid aSb \mid bSa \mid SS$$

}

S is start symbol.

(Eg) obtain a grammar to generate a string balanced parentheses.

Soln:

$$G = (V, T, P, S)$$

$$S \rightarrow (S) \mid \epsilon \mid SS$$

$$V = \{ S \}$$

$$T = \{ (,) \}$$

$$P = \{ S \rightarrow (S) \mid SS \mid \epsilon \}$$

S is start symbol.

(eg)

obtain a grammar to generate the lang:

$$L = \{a^i b^j \mid i \neq j, i \geq 0 \text{ & } j \geq 0\}$$

Soln

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

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

$$A \rightarrow a \mid aa$$

$$B \rightarrow b \mid bb$$

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

$$S \rightarrow OS1 \mid A \mid B$$

$$A \rightarrow O \mid OA$$

$$B \rightarrow 1 \mid 1B$$

$$G = (V, T, P, S)$$

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

$$T = \{a, b\}$$

$$P = \left\{ \begin{array}{l} S \rightarrow aSb \mid A \mid B \\ A \rightarrow a \mid aa \\ B \rightarrow b \mid bb \end{array} \right\}$$

S is start symbol.

$$G = (V, T, P, S)$$

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

$$T = \{0, 1\}$$

$$P = \{S \rightarrow OS1 \mid A \mid B\}$$

$$A \rightarrow O \mid OA$$

$$B \rightarrow 1 \mid 1B$$

(eg)

obtain a grammar to generate the language

$$L = \{a^{n+2} b^m \mid n \geq 0 \text{ & } m > n\}$$

Soln

$$S \rightarrow aAB$$

$$A \rightarrow aAb \mid ab$$

$$B \rightarrow bB \mid \epsilon$$

$$G = (V, T, P, S)$$

$$V = \{S, A, B\}, T = \{a, b\}, P = \left\{ \begin{array}{l} S \rightarrow aAB, \\ A \rightarrow aAb \mid ab \\ B \rightarrow bB \mid \epsilon \end{array} \right\}$$

S is start symbol

Q

Obtain a grammar to generate the language

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

Sol:

$$n=0 \quad n=1 \quad n=2$$

$$m \geq 1 \quad m \geq 2 \quad m \geq 3$$

$$L = \{\boxed{\epsilon} bb^*, \boxed{ab} bb^*, \boxed{aa} \boxed{bb} bb^*, \dots\}$$

~~SOULD APP~~

$S \rightarrow aSb/B$ [Generates $a^n b^n \mid n \geq 0$ followed by at least one b]

$B \rightarrow bB/b$ [Generates one or more b 's]

$$G = (V, T, P, S)$$

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

$$T = \{a, b\}$$

$$P = \{S \rightarrow aSb/B, B \rightarrow bB/b\}$$

S is start symbol.

Q

Obtain a grammar to generate the language

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

Sol:

$$n=3 \quad n=4 \quad n=5$$

$$L = \{aaa, aaaab, aaaaaabb, \dots\}$$

$$P = \{S \rightarrow aaat \\ \rightarrow aAb \mid \epsilon\}$$

Derivation.

Defⁿ: Let $A \rightarrow \alpha B \gamma$ & $B \rightarrow \beta$ are production in grammar G, where α, β and γ are strings of terminal and/or non-terminals. A and B are non-terminals. The non-terminal A produces the string $\alpha \beta \gamma$ by replacing the non-terminal B in $\alpha B \gamma$ by the string β by applying the production $B \rightarrow \beta$, can be written as

$$A \Rightarrow \alpha \beta \gamma$$

This process of obtaining string of terminals &/or non-terminals from the start symbol by applying some or all productions is called derivation.

If a string is obtained by applying only one prodⁿ, then it is called one-step derivation & is denoted by the symbol ' \Rightarrow '. If one or more productions are applied to get the string $\alpha \beta \gamma$ from A then we write

$$A \Rightarrow^+ \alpha \beta \gamma$$

If two or more prodⁿ are applied to get the string $\alpha \beta \gamma$ from A, then we write

$$A \xrightarrow{*} \alpha \beta \gamma$$

(Eg) Consider the grammar shown below from which any arithmetic expr can be obtained.

$$E \rightarrow E + E$$

$$E \rightarrow E - E$$

$$E \rightarrow E * E$$

$$E \rightarrow E / E$$

$$E \rightarrow id.$$

E is considered to be the start symbol. obtain the string id+id*id & show the derivation for the same

Sol:

$$E \Rightarrow E + E$$

$$\Rightarrow id + E$$

$$\Rightarrow id + E * E$$

$$\Rightarrow id + id * E$$

$$\underline{E \Rightarrow id + id * id}$$

Sentence or Sentential form

Defⁿ: Let $G = (V, T, P, S)$ be a grammar. The string w obtained from the grammar G such that $S \Rightarrow^* w$ is called Sentence of grammar G . Here w is string of terminals.

In the above (eg) id+id*id is the sentence of the grammar.

If there is a derivation $S \xrightarrow{*} \alpha$, where α contains

String of terminals and/or non-terminals, then α is called sentential form of G .
In the derivation shown in above eg.

$E+E, id+E, id+E*E, id+id*E, id+id*id$
are all sentential forms of UT grammar.

Language

Defn: Let $G = (V, T, P, S)$ be a grammar. The language $L(G)$ generated by the grammar G is

$$L(G) = \{ w \mid S \xrightarrow{*} w \text{ & } w \in T^* \}$$

i.e., w is a string of terminals (may be ϵ) obtained from the start symbol S by applying an arbitrary number of productions.

Leftmost Derivation

Defⁿ: In the derivation process if a left most variable is replaced at every step, then the derivation is said to be left most.

(Eg) obtain the leftmost derivation for the string $id + id * id$ using the following grammar

$$E \rightarrow E + E$$

$$E \rightarrow E - E$$

$$E \rightarrow E * E$$

$$E \rightarrow E | E$$

$$E \rightarrow E^{\wedge} E$$

\wedge : denotes exponent of

$$E \rightarrow id$$

For

$$E \xrightarrow[lm]{+} E + E$$

$$\Rightarrow id + E$$

$$\Rightarrow id + E * E$$

$$\Rightarrow id + id * E$$

$$\Rightarrow id + id * id$$

The string $id + id * id$ is obtained from the start symbol E by applying leftmost derivation. It can be written as

$$E \xrightarrow[lm]{+} id + id * id$$

Eg ✓ obtain the leftmost derivation for the string aaabbabbba using the following grammar

$$S \rightarrow aB | bA$$

$$A \rightarrow aS | bAA | a$$

$$B \rightarrow bS | aBB | b$$

$$\underset{lm}{\begin{matrix} S \\ \Rightarrow \end{matrix}}$$

(Applying $S \rightarrow aB$)

$$\Rightarrow aAB \quad (B \rightarrow aBB)$$

$$\Rightarrow aaABBB \quad (B \rightarrow aBB)$$

$$\Rightarrow aaabBB \quad (B \rightarrow b)$$

$$\Rightarrow aaabbB \quad (B \rightarrow b)$$

$$\Rightarrow aaabbAB \quad (B \rightarrow aBB)$$

$$\Rightarrow aaabbabB \quad (B \rightarrow b)$$

$$\Rightarrow aaabbabbS \quad (B \rightarrow bS)$$

$$\Rightarrow aaabbabbAA \quad (S \rightarrow bA)$$

$$\Rightarrow aaabbabbba \quad (A \rightarrow \omega)$$

Rightmost Derivation

Defⁿ: In the derivation process if a right most variable is replaced at every step, then the derivation is said to be rightmost.

(Eg)

Obtain the rightmost derivation for the string id+id*id using following grammar.

$$E \rightarrow E+E$$

$$E \rightarrow E * E$$

$$E \rightarrow E - E$$

$$E \rightarrow E/E$$

$$E \rightarrow id$$

RD?

$$E \xrightarrow{rm} E+E$$

$$\Rightarrow E+E * E$$

$$\Rightarrow E+E * id$$

$$\Rightarrow E+id+id$$

$$\Rightarrow id+id+id$$

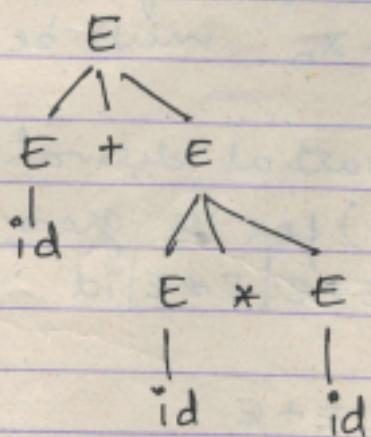
Derivation Tree (parse tree)

The derivation can be shown in the form of a tree. Such trees are called derivation or parse trees.

Defⁿ (Parse tree or derivation tree): Let $G = (V, T, P, S)$ be a CFG. The tree is derivation tree (parse tree) with the following properties.

1. The root has the label S .
2. Every vertex has a label which is in $(V \cup T \cup \epsilon)$.
3. Every leaf node has label from T & an interior vertex has a label from V .
4. If a vertex is labeled $A \in V$ if $x_1, x_2, x_3, \dots, x_n$ are all children of A from left, then $A \rightarrow x_1 x_2 x_3 \dots x_n$ must be a production in P .

Parse tree for above eg.



The string $\text{id} + \text{id} * \text{id}$ is called yield of the tree.

Yield of UI tree

The yield of a tree is the string of symbols obtained by only reading the leaves of the tree from left to right with considering the ϵ -symbols. The yield of a tree is always a terminal string.

Def^b (partial parse tree or partial derivation tree): Let $G = (V, T, P, S)$ be a CFG. The tree is partial derivation tree (parse tree) with the following properties:

- ① The ^{root} has the label S
- ② Every vertex has a label which is in $(V \cup T)^*$
- ③ Every leaf node has a label from $(V \cup T)^*$.
- ④ If a vertex is labeled $A \in V$ if $x_1, x_2, x_3, \dots, x_n$ are all children of A from left. Then $A \rightarrow x_1 x_2 x_3 \dots x_n$ must be a prod^g in P .

- Ex consider the partial derivation (by applying right most der^r) for UI grammar.

$$E \rightarrow E+E \mid E * E \mid id$$

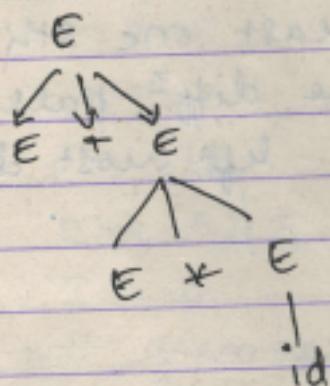
is shown below

$$E \rightarrow E+E$$

$$\Rightarrow E+E * E$$

$$\Rightarrow E+E * id$$

for this partial right most derivation, the partial derivation tree is shown below



It is clear from the parse tree the partial parse tree that all the leaves in parse tree are the symbols from (TUE) whereas in partial parse tree the leaves will be from (VUTUE).

Ambiguous Grammar

Defⁿ: Let $G = (V, T, P, S)$ be a context free grammar. A grammar G is ambiguous if & if there exists at least one string $w \in T^*$ for which two or more diff^u parse trees exists by applying either the left most derivation or right most derivation.

Note :

- 1) obtain the leftmost derivation to get a string w . obtain the rightmost derivation to get a string w . for both the derivations construct the parse tree. If there are two diff^u parse trees, then the grammar is ambiguous.
- 2) obtain the string w by applying leftmost derivation twice & construct the parse tree. If the two parse trees are different, the grammar is ambiguous.
- 3) obtain the string w by applying rightmost derivation twice & construct the parse tree. If the two parse trees are diff^u, the grammar is ambiguous.

(g) consider the grammar shown below

$$E \rightarrow E+E$$

$$E \rightarrow E-E$$

$$E \rightarrow E * E$$

$$E \rightarrow E/E$$

$$E \rightarrow (E) | I$$

$$I \rightarrow id$$

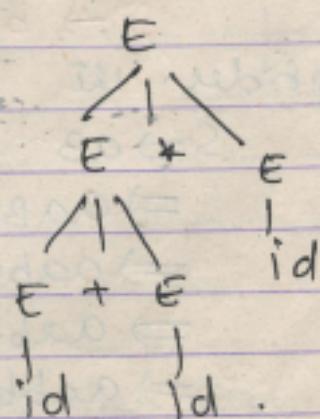
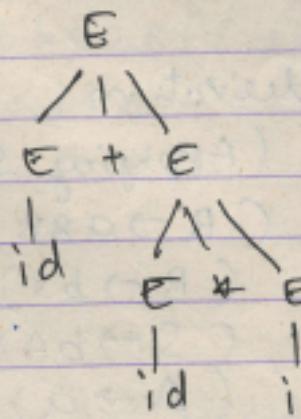
Show that the grammar is ambiguous.

The sentence $id + id * id$ can be obtained from leftmost derivation in two ways as shown below

$$\begin{aligned} E &\Rightarrow E+E \\ &\Rightarrow id+E \\ &\Rightarrow id+E*E \\ &\Rightarrow id+id*E \\ &\Rightarrow id+id*id \end{aligned}$$

$$\begin{aligned} E &\Rightarrow E * E \\ &\Rightarrow E+E*E \\ &\Rightarrow id+E*E \\ &\Rightarrow id+id*E \\ &\Rightarrow id+id*id \end{aligned}$$

The corresponding derivation trees for the two leftmost ~~derivation~~ derivation are shown below



Since the two parse trees are diffⁿ for the same sentence $id + id * id$ by applying leftmost derivation, the grammar is ambiguous.

(eg)

Is the following grammar ambiguous?

$$S \rightarrow aS \mid x$$

$$x \rightarrow ax \mid a$$

Sol?

Consider the two leftmost derivation for the string aaaa.

$$S \Rightarrow aS$$

$$\Rightarrow aaS$$

$$\Rightarrow aaaS$$

$$\Rightarrow aaax$$

$$\Rightarrow aaaa$$

$$S \Rightarrow x$$

$$\Rightarrow ax$$

$$\Rightarrow aax$$

$$\Rightarrow aaax$$

$$\Rightarrow aaaa$$

Since there are two leftmost derivations for the same sentence aaaa, the given grammar is ambiguous.

Show that the grammar is ambiguous for the string

(eg)

Is the following grammar ambiguous? aabbab

$$S \rightarrow aB \mid bA$$

$$A \rightarrow aS \mid bAA \mid a$$

$$B \rightarrow bS \mid aBB \mid b$$

Sol? Consider the leftmost derivation

$$S \Rightarrow aB \quad (\text{Applying } S \rightarrow aB)$$

$$\Rightarrow aABB \quad (B \rightarrow aBB)$$

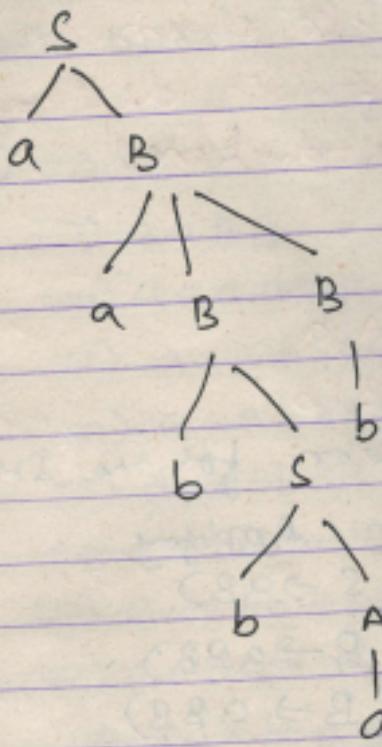
$$\Rightarrow aabsB \quad (B \rightarrow bS)$$

$$\Rightarrow aabbAB \quad (S \rightarrow bA)$$

$$\Rightarrow aabbAB \quad (A \rightarrow a)$$

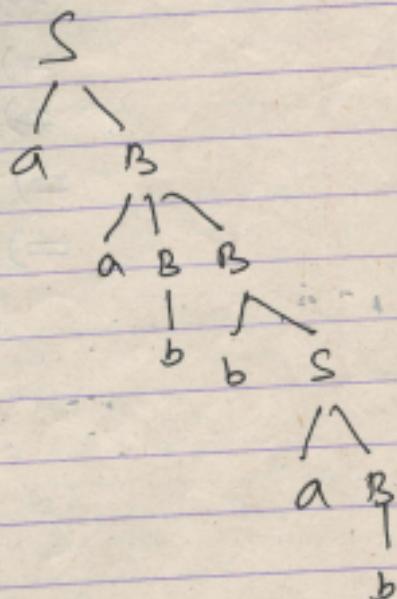
$$\Rightarrow aabbab \quad (B \rightarrow b)$$

parse tree



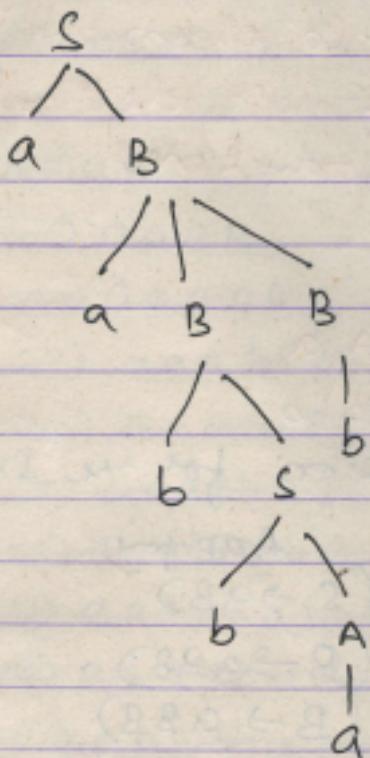
The same string aabbab can be obtained again by applying leftmost derivation as shown below.

Applying
 $S \Rightarrow aB \quad (S \rightarrow aB)$
 $\Rightarrow aAB\bar{B} \quad (\bar{B} \rightarrow a\bar{B}\bar{B})$
 $\Rightarrow aab\bar{B} \quad (\bar{B} \rightarrow b)$
 $\Rightarrow aabbs \quad (B \rightarrow bS)$
 $\Rightarrow aabb\bar{a}B \quad (S \rightarrow aB)$
 $\Rightarrow aabbab \quad (B \rightarrow b)$



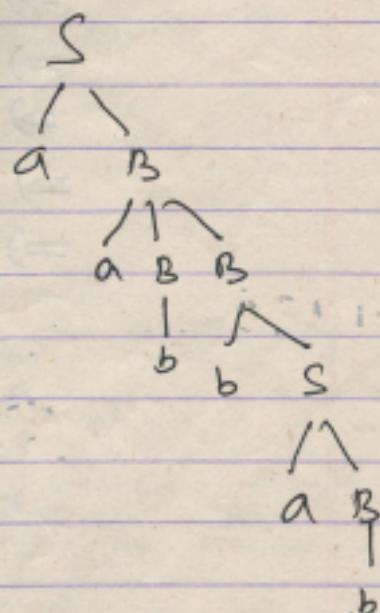
There are two parse trees for the string aabbab by applying leftmost derivation as to the given grammar is ambiguous.

Parse tree



The same string aabbab can be obtained again by applying leftmost derivation as shown below.

$$\begin{aligned}
 & S \xrightarrow{\text{apply } S \rightarrow aB} aB \\
 & \Rightarrow aAB \xrightarrow{B \rightarrow aBB} aABB \\
 & \Rightarrow aabB \xrightarrow{B \rightarrow b} aabb \\
 & \Rightarrow aabbs \xrightarrow{B \rightarrow bS} aabbaS \\
 & \Rightarrow aabbab \xrightarrow{S \rightarrow aB} aabbab \\
 & \Rightarrow aabbab \xrightarrow{B \rightarrow b} aabbab
 \end{aligned}$$



There are two parse trees for the string aabbab by applying leftmost derivation \Rightarrow so the given grammar is ambiguous.

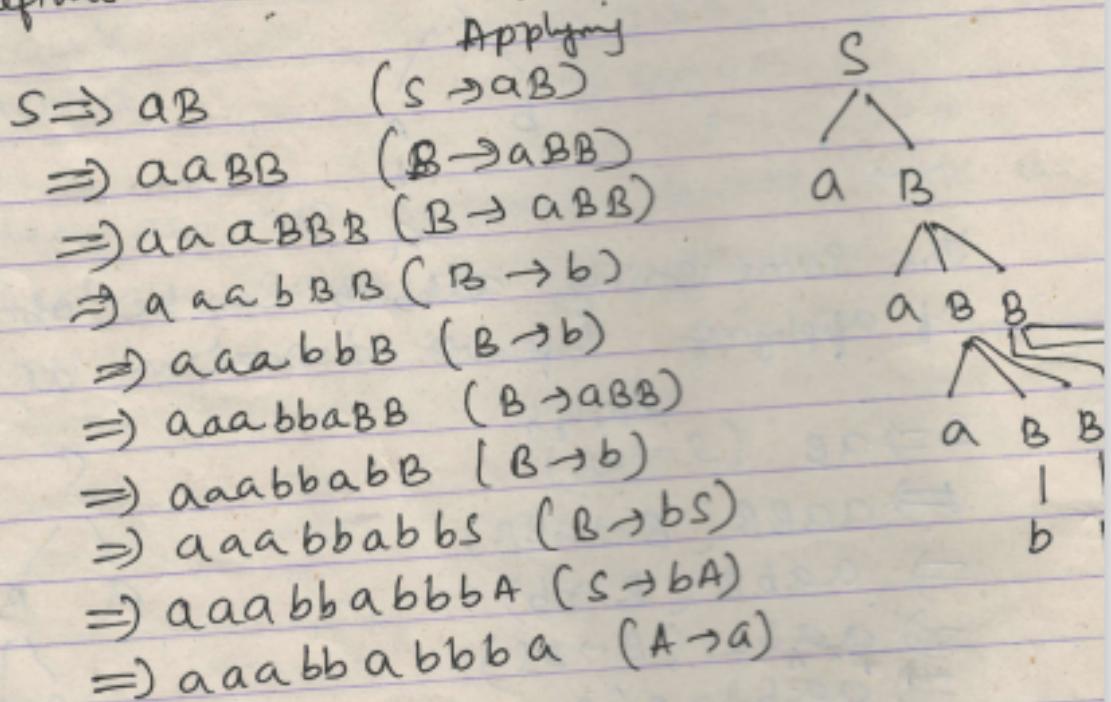
Eg) Obtain the string aaabbabbba by applying leftmost derivation & the parse tree for the grammar shown below. Is it possible to obtain the same string again by applying leftmost derivation but by selecting different productions?

$$S \rightarrow aB | bA$$

$$A \rightarrow aS | bAA | a$$

$$B \rightarrow bS | aBB | b$$

Sol2. The leftmost derivation for the string aaabbabbba

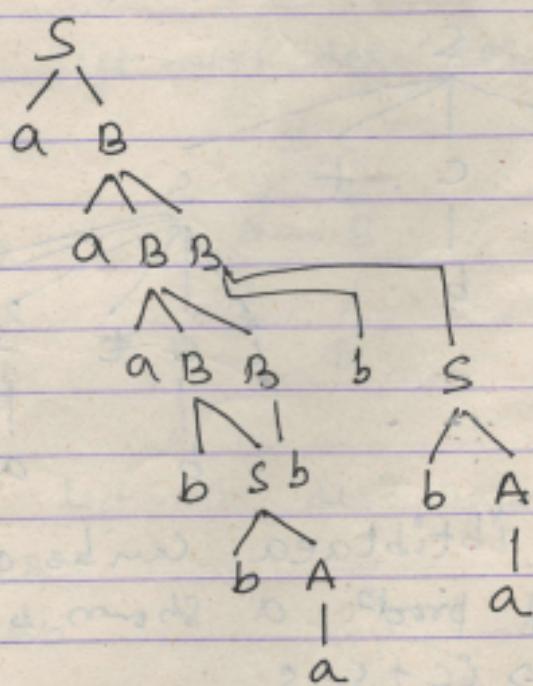


The leftmost der² for the same string
 aaabbabbba but by applying diff² set of prod²
 is shown below:

$$\begin{aligned}
 S &\Rightarrow aB & (S \rightarrow aB) \\
 &\Rightarrow a a B B & (B \rightarrow a B B) \\
 &\Rightarrow a a a B B B & (B \rightarrow a B B) \\
 &\Rightarrow a a a b S B B & (B \rightarrow b S) \\
 &\Rightarrow a a a b b A B B & (S \rightarrow b A) \\
 &\Rightarrow a a a b b a B B & (A \rightarrow a) \\
 &\Rightarrow a a a b b a b B & (B \rightarrow b) \\
 &\Rightarrow a a a b b a b b S & (B \rightarrow b S) \\
 &\Rightarrow a a a b b a b b b A & (S \rightarrow b A) \\
 &\Rightarrow a a a b b a b b b a & (A \rightarrow a)
 \end{aligned}$$

of

my



(eg)

IS THE following grammar ambiguous?

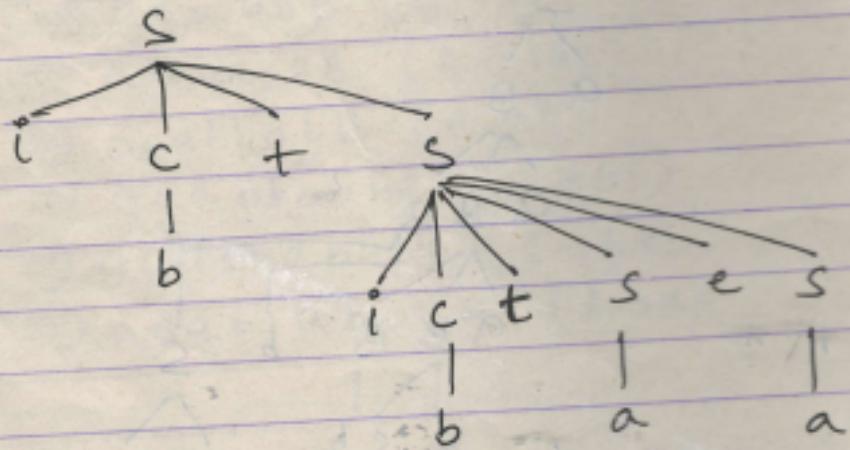
$$S \rightarrow icts \mid ictses \mid a$$

$$C \rightarrow b$$

Soln. The string ibtibtaea can be obtained by applying the leftmost deriv as shown below

$$\begin{aligned} S &\Rightarrow icts \\ &\Rightarrow ibts \\ &\Rightarrow ibticstses \\ &\Rightarrow ibtibtses \\ &\Rightarrow ibtibtaes \\ &\Rightarrow ibtibtaea \end{aligned}$$

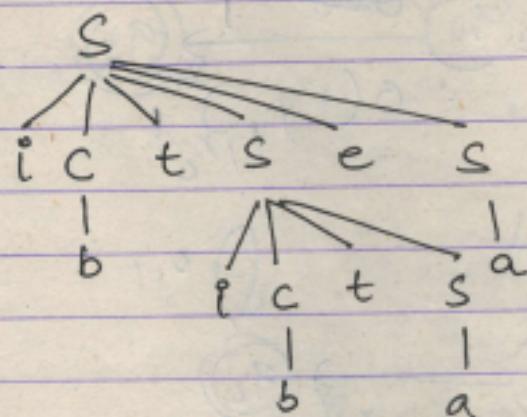
parse tree



The string ibtibtaea can be obtained by using diff set of prod as shown below -

$$\begin{aligned} S &\Rightarrow ic + ses \\ &\Rightarrow ibt + ses \\ &\Rightarrow ibtic + ses \\ &\Rightarrow ibtib + ses \\ &\Rightarrow ibtibtaes \\ &\Rightarrow ibtibtaea \end{aligned}$$

The parse tree for this is shown below



g) Is the grammar ambiguous?

$$S \rightarrow AB \mid aab$$

$$A \rightarrow a \mid Aa$$

$$B \rightarrow b.$$

$$S \rightarrow AB \quad S \Rightarrow aab$$

$$\Rightarrow Aab \quad \Rightarrow aab$$

$$\Rightarrow ab$$

s

any

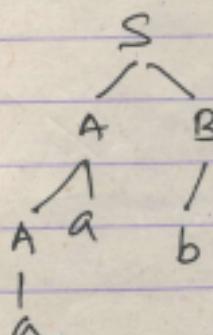
ii) consider the left most derivⁿ for the string aab

$$S \Rightarrow AB$$

$$\Rightarrow AaB$$

$$\Rightarrow aaB$$

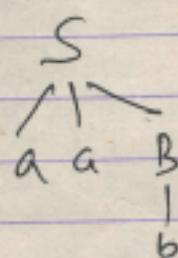
$$\Rightarrow aab$$



consider the left most derivⁿ again for the string aab

$$S \Rightarrow aAB$$

$$\Rightarrow aab$$



since there are two parse trees for the string aab, the given grammar is ambiguous.

Properties of context-free languages

The goal of this section is to show that every CFL (without ϵ) is generated by a CFG in which all prodⁿs are of the form $A \rightarrow BC$ or $A \rightarrow a$, where $A, B \in C$ are variables & a is a terminal.

This form is called Chomsky Normal form (CNF).

To get there, we need to make a number of preliminary simplifications, which are themselves useful in various ways:

- 1) We must eliminate useless symbols, those variable or terminals that do not appear in any derivation of a terminal string from the start symbol.
- 2) We must eliminate ϵ -prodⁿ, those of the form $A \rightarrow \epsilon$ for some variable A .
- 3) We must eliminate unit prodⁿ, those of the form $A \rightarrow B$ for variable $A \neq B$.

Eliminating useless symbols:

We say a symbol x is useful for a grammar $G = (V, T, P, S)$ if there is a some derivation of the form $S \xrightarrow{*} \alpha x \beta \xrightarrow{*} w$, where w is in T^* . Note that x may be in either V or T , & the sentential form $\alpha x \beta$ might be the first or last in the derivation. If x is not useful, we say it is useless.

Evidently, omitting useless symbols from a grammar will not change the language generated. So we may as well decide to eliminate all useless symbols.

our approach to eliminating useless symbols begins by identifying the two things a symbol has to be able to do to be useful:

- 1) we say x is generating if $x^* \Rightarrow w$ for some terminal string w . Note that every terminal is generating, since w can be that terminal itself, which is derived by zero steps.
- 2) we say x is reachable if there is a derivation $S^* \Rightarrow \alpha x \beta$ for some $\alpha \in \Sigma$.

surely a symbol that is useful will be both generating & reachable. If we eliminate the symbols that are not generating first & then eliminate from the remaining grammar those symbols that are not reachable we have only the useful symbols left.

Ex) Eliminate useless symbols from the following grammar

$$S \rightarrow AB | a$$

$$A \rightarrow a.$$

1) Step: find the set of vae & the prod^{nt} from which we get only string of terminals.

OV	nv	prod ^{nt}
\emptyset	S, A	$S \rightarrow a$ $A \rightarrow a$
S, A	S, A	$S \rightarrow a$ $A \rightarrow a$

The resulting grammar, from which we get only string of terminals is given by.

$$G_1 = (V_1, T_1, P_1, S)$$

where

$$V_1 = \{ S, A \}$$

$$T_1 = \{ a \}$$

$$P_1 = \{ S \rightarrow a, A \rightarrow a \}$$

Step: Eliminate the symbol ϵ production which are not reachable from the start symbol.

P'	T'	V'
-	-	S
$S \rightarrow a$	a	S

So the final grammar obtained after eliminating useless symbols ϵ productions is given by

$$G' = (V', T', P', S)$$

where

$$V' = \{ S \}$$

$$T' = \{ a \}$$

$$P' = \{ S \rightarrow a \}$$

S is the start symbol.

eliminate the useless symbols in the grammar

$$S \rightarrow aA | bB$$

$$A \rightarrow aA | a$$

$$B \rightarrow bB$$

$$D \rightarrow ab | Ea$$

$$E \rightarrow ac | d$$

OV	nv	Prod \Rightarrow
\emptyset	A, E, D	$A \rightarrow a$ $E \rightarrow d$ $D \rightarrow ab$
A, E, D	A, E, D, S	$S \rightarrow aA$ $A \rightarrow aA a$ $D \rightarrow ab Ea$
A, D, E, S	A, D, E, S	-

The resulting grammar $G_1 = (V_1, T_1, P_1, S)$ where

$$V_1 = \{A, D, E, S\}$$

$$T_1 = \{a, b, d\}$$

$$P_1 = \{S \rightarrow aA, A \rightarrow aA | a, D \rightarrow ab | Ea, E \rightarrow d\}$$

S is the start symbol.

P'	T'	V'
-	-	S
$S \rightarrow aA$	a	S, A
$A \rightarrow aA/a$	a	S, A

The resulting grammar $G' = (V', T', P', S)$
where

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

$$T' = \{a\}$$

$$P' = \{S \rightarrow aA, A \rightarrow aA/a\}$$

$\left. \begin{matrix} \\ S \end{matrix} \right\}$ is the start symbol.

⑨ Eliminate useless symbols from the following grammar

$$S \rightarrow aAB$$

$$A \rightarrow aB/B$$

$$B \rightarrow aB/b/bC$$

$$D \rightarrow EA$$

$$E \rightarrow a/aE/bC$$

Sol)

O_V	N_V	Production
\emptyset	B, E	$B \rightarrow b$ $E \rightarrow a$
B, E	A, B, D, E	$A \rightarrow aB/B$ $B \rightarrow aB$ $D \rightarrow EA$ $E \rightarrow aE$

ov	nv	prod^n.
A, B, D, E	S, A, B, D, E	$S \rightarrow aAB$
S, A, B, D, E	-	S → aAB

$$G_1 = (V_1, T_1, P_1, S)$$

$$V_1 = \{ S, A, B, D, E \}$$

$$T_1 = \{ a, b \}$$

$$P_1 = \{ S \rightarrow aAB, A \rightarrow aB|b, B \rightarrow ab|b, D \rightarrow Ea, E \rightarrow aE|a \}$$

prod^n(p')	T'	V'
-	-	S
$S \rightarrow aAB$	a	S, A, B
$A \rightarrow aB b$	a	S, A, B
$B \rightarrow ab b$	a, b	S, A, B

$$G' = (V', T', P', S)$$

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

$$T' = \{ a, b \}$$

$$P' = \{ S \rightarrow aAB, A \rightarrow aB|b, B \rightarrow ab|b \}$$

S is the start symbol.

(Eg) Eliminate useless symbols from the following grammar

$$S \rightarrow aA | a | Bb | cc$$

$$A \rightarrow AB$$

$$B \rightarrow a | Aa$$

$$C \rightarrow cCD$$

$$D \rightarrow ddd$$

Soln: Step 1:

OV	nv	prod ²
\emptyset	S, B, D	$S \rightarrow a$ $B \rightarrow a$ $D \rightarrow ddd$
S, B, D	S, A, B, D	$S \rightarrow Bb$ $A \rightarrow AB$
S, A, B, D	S, A, B, D	$S \rightarrow aA$ $A \rightarrow AB$ $B \rightarrow Aa$

$$G_1 = (V_1, T_1, P_1, S)$$

$$V_1 = \{ S, A, B, D \}$$

$$T_1 = \{ a, b, d \}$$

$$P_1 = \{ S \rightarrow a | Bb | aA, A \rightarrow AB, B \rightarrow a | Aa, D \rightarrow ddd \}$$

production (p')

$$S \rightarrow a | Bb | aA$$

$$A \rightarrow aB$$

$$B \rightarrow a | Aa$$

T'

$$a, b$$

$$a, b$$

$$a, b$$

V'

S

$$S, A, B$$

$$S, A, B$$

$$S, A, B$$

The resulting grammar is

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

$$T' = \{ a, b \}$$

$$p' = \{ S \rightarrow a | Bb | aA, A \rightarrow aB, B \rightarrow a | Aa \}$$

S is the start symbol.

Eliminating ϵ -productions.

A production of the form $A \rightarrow \epsilon$ is undesirable in a CFG, unless an empty string is derived from the start symbol. Suppose, the language generated from a grammar G does not derive any empty string & the grammar consists of ϵ -productions. Such ϵ -prod^{ns}s can be removed.

Defn: Let $G = (V, T, P, S)$ be a CFG. A production in P of the form

$$A \rightarrow \epsilon$$

is called an ϵ -prodⁿ or null prodⁿ. After applying the production the variable A is erased. For each A in V , if there is a derivation of the form

$$A \xrightarrow{*} \epsilon$$

then A is a nullable variable.

(Eg) In grammar if there is one prod of the form

$$B \rightarrow \epsilon$$

$$C \rightarrow \epsilon$$

are ϵ -prod^{ns} and the variable B, C are nullable variables. If there is a production

$$A \rightarrow BC$$

as both B, C are nullable variables, then A is also a nullable variable.

Eliminate all ϵ -productions from the grammar

$$S \rightarrow ABCa/bD$$

$$A \rightarrow BC/b$$

$$B \rightarrow b/\epsilon$$

$$C \rightarrow c/\epsilon$$

$$D \rightarrow d$$

$$B \rightarrow \epsilon$$

$$C \rightarrow \epsilon$$

$$A \rightarrow BC$$

$V = \{B, C, A\}$ are all nullable variables.

productions	Resulting prod ⁿ
$S \rightarrow ABCa$	$S \rightarrow ABCa/BCa/Aca/ABA/ca/Aa/Ba/a$
$S \rightarrow bD$	$S \rightarrow bD$
$A \rightarrow BC/b$	$A \rightarrow BC/B/c/b$
$B \rightarrow b/\epsilon$	$B \rightarrow b$
$C \rightarrow c/\epsilon$	$C \rightarrow c$
$D \rightarrow d$	$D \rightarrow d$

The grammar $G^1 = (V^1, T^1, P^1, S)$ where

$$V^1 = \{S, A, B, C, D\}$$

$$T^1 = \{a, b, c, d\}$$

$$P^1 = \{S \rightarrow ABCa/BCa/Aca/ABA/ca/Aa/Ba/a/bD, A \rightarrow BC/B/c/b\}$$

$$B \rightarrow b$$

$$C \rightarrow c$$

$$D \rightarrow d$$

S is the start symbol.

(Eq)

Eliminate all ϵ -productions from the grammar

$$S \rightarrow BAAB$$

$$A \rightarrow 0A2 | 2AO | \epsilon$$

$$B \rightarrow AB | 1B | \epsilon$$

Soln

$$A \rightarrow \epsilon$$

$$B \rightarrow \epsilon$$

$$S \rightarrow BAAB$$

so S, A, B are all nullable variables.

productions	resulting prod $\sim^*(P)$
$S \rightarrow BAAB$	$S \rightarrow BAAB AAB BAB BAA AB BB BA AA A B$
$A \rightarrow 0A2$	$A \rightarrow 0A2 02$
$A \rightarrow 2AO$	$A \rightarrow 2AO 2O$
$B \rightarrow AB$	$B \rightarrow AB B A$
$B \rightarrow 1B$	$B \rightarrow 1B 1$

The grammar $G^1 = (V^1, T^1, P^1, S)$ where

$$V^1 = \{S, A, B\}$$

$$T^1 = \{0, 1, 2\}$$

$$P^1 = \{ S \rightarrow BAAB | AAB | BAB | BAA | AB | BB | BA | AA | A | B \}$$

$$A \rightarrow 0A2 | 02 | 2AO | 2O$$

$$B \rightarrow AB | B | A | 1B | 1 \}$$

S is the start symbol.

Eliminating unit productions

A unit production is a prodⁿ of the form

$A \rightarrow B$, where both A & B are variables.

Eliminate all unit prodⁿ from the grammar

$$S \rightarrow AB$$

$$A \rightarrow a$$

$$B \rightarrow C \mid b$$

$$C \rightarrow D$$

$$D \rightarrow E \mid bc$$

$$E \rightarrow d \mid Ab$$

Non unit prodⁿ

unit prodⁿ

$$S \rightarrow AB$$

$$B \rightarrow C$$

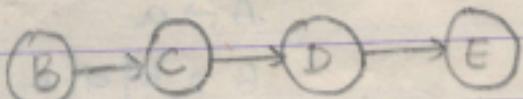
$$A \rightarrow a$$

$$C \rightarrow D$$

$$B \rightarrow b$$

$$D \rightarrow E$$

$$D \rightarrow bc$$



$$E \rightarrow d \mid Ab$$

It is clear from the dependency graph that $D \rightarrow E$, so the non-unit prodⁿ generated from 'E' can also be generated from 'D'.

$$E \rightarrow d \mid Ab$$

Since $D \Rightarrow E$

$$D \rightarrow d \mid Ab$$

The resulting D prodⁿ are

$$D \rightarrow d \mid Ab \mid bc$$

Since $C \xrightarrow{*} D \xrightarrow{*} E$

$C \rightarrow d | Ab | bC$

Since $B \xrightarrow{*} C, C \xrightarrow{*} D, D \xrightarrow{*} E$

$B \rightarrow b | d | Ab | bC$

The final grammar obtained after eliminating unit production is shown below:

$$V = \{ S, A, B, C, D, E \}$$

$$T = \{ a, b, d \}$$

$$P = \{ S \rightarrow AB \}$$

$$A \rightarrow a$$

$$B \rightarrow b | d | Ab | bC$$

$$C \rightarrow bC | d | Ab$$

$$D \rightarrow bC | d | Ab$$

$$E \rightarrow d | Ab$$

}

S is the start symbol.

(eg) Eliminate unit production from the grammar

$$S \rightarrow A0 | B$$

$$B \rightarrow A | 11$$

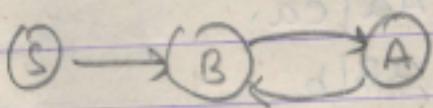
$$A \rightarrow 0 | 12 | B$$

Non-unit prodⁿ

$$\begin{aligned} S &\rightarrow A0 \\ B &\rightarrow 11 \\ A &\rightarrow 0|12 \end{aligned}$$

Unit prodⁿ

$$\begin{aligned} S &\rightarrow B \\ B &\rightarrow A \\ A &\rightarrow B \end{aligned}$$



It is clear from the dependence graph that
 $S \xrightarrow{*} B$, $S \xrightarrow{*} A$, $B \xrightarrow{*} A$ & $A \xrightarrow{*} B$ so the new
prodⁿ for S, A & B are

$$S \rightarrow A0|11|0|12$$

$$B \rightarrow 0|12|11$$

$$A \rightarrow 0|12|11$$

The resulting grammar without unit prodⁿ

$$G' = (V', T', P', S)$$

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

$$T' = \{0, 1, 2\}$$

$$P' = \{S \rightarrow A0|11|0|12,$$

$$B \rightarrow 0|12|11,$$

$$A \rightarrow 0|12|11\}$$

S is the start symbol.

(Eq)

eliminate unit prodⁿs from the grammar

$$S \rightarrow Aa | B | Ca$$

$$B \rightarrow aB | b$$

$$C \rightarrow Db | D$$

$$D \rightarrow E | d$$

$$E \rightarrow ab$$

Ans?

Non-unit prodⁿs.

$$S \rightarrow Aa | Ca$$

$$B \rightarrow aB | b$$

$$C \rightarrow Db$$

$$D \rightarrow d$$

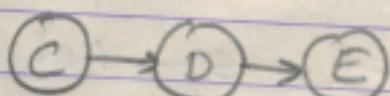
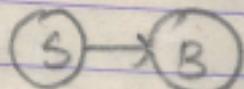
$$E \rightarrow ab$$

Unit prodⁿs

$$S \rightarrow B$$

$$C \rightarrow D$$

$$D \rightarrow \cancel{d}$$



$$S \xrightarrow{*} B$$

$$S \rightarrow Aa | aB | b | Ca$$

~~$$C \not\rightarrow D$$~~
$$D \not\rightarrow E$$

$$D \rightarrow ab | d$$

$$C \xrightarrow{1} D \quad D \xrightarrow{2} E$$

$$C \rightarrow D_b | ab | d$$

The resulting grammar

$$G' = (V', T', P', S)$$

$$V' = \{ S, A, B, C, D, E \}$$

$$T' = \{ a, b, d \}$$

$$P' = \{ S \rightarrow Aa | aB | b | ca \}$$

$$B \rightarrow aB | b$$

$$C \rightarrow D_b | ab | d$$

$$D \rightarrow d | ab$$

$$E \rightarrow ab ?$$

S is the start symbol.

Chomsky Normal form.

Every nonempty CFL with ϵ has a grammar G in which all productions are in one of two forms, either

- 1) $A \rightarrow BC$, where $A, B \in V_F$ & C are variable
 - 2) $A \rightarrow a$, where A is a variable & a is a terminal
- Further, G has no useless symbols. Such a grammar is said to be in Chomsky Normal form, or CNF.

(Eg) Convert the following grammar to CNF

- ① Remove ϵ -prod^{ing}
- ② Remove unit-p
- ③ Remove useless

$$\begin{aligned} S &\rightarrow 0A \mid 1B \\ A &\rightarrow 0AA \mid 1S \mid 1 \\ B &\rightarrow 1BB \mid 0S \mid 0 \end{aligned}$$

soln)

Std form

$$A \rightarrow 1$$

$$B \rightarrow 0$$

$$\begin{aligned} S &\rightarrow 0A \mid 1B \\ S &\rightarrow B_0A \mid B_1B \quad ; \quad A \rightarrow B_0AA \mid B_1S \quad , \quad B \rightarrow B_1BB \mid B_0S \\ B_0 &\rightarrow 0 \quad \quad \quad B_0 \rightarrow 0 \quad \quad \quad B_0 \rightarrow 0 \\ B_1 &\rightarrow 1 \quad \quad \quad B_1 \rightarrow 1 \quad \quad \quad B_1 \rightarrow 1 \end{aligned}$$

Grammar obtained

$$V_1 = \{S, A, B, B_0, B_1\}$$

$$T_1 = \{0, 1\}$$

$$\text{P1 P1} = \{S \rightarrow B_0A \mid B_1B, A \rightarrow B_0AA \mid B_1S \mid 1, B \rightarrow B_1BB \mid B_0S \mid 0, B_1 \rightarrow 1\}$$

S is the start symbol.

$$S \rightarrow B_0 A \mid B_1 B$$
$$A \rightarrow B_0 AA \mid B_1 S \mid 1$$
$$B \rightarrow B_1 BB \mid B_0 S \mid 0$$
$$B_0 \rightarrow 0$$
$$B_1 \rightarrow 1$$

CNF prod's are

$$S \rightarrow B_0 A \mid B_1 S$$
$$A \rightarrow B_1 S \mid 1$$
$$B \rightarrow B_0 S \mid 0$$
$$B_0 \rightarrow 0$$
$$B_1 \rightarrow 1$$

productions which are not in CNF form.

$$A \rightarrow B_0 AA$$
$$B \rightarrow B_1 BB$$
$$A \rightarrow B_0 AA, A \rightarrow B_0 D_1$$
$$D_1 = AA$$
~~B~~
$$B \rightarrow B_1 BB$$
$$B \rightarrow B_1 D_2$$
$$D_2 \Rightarrow BB$$
$$G = \{ S, A, B, B_0, B_1, D_1, D_2 \}$$
$$T = \{ 0, 1 \}$$
$$P_1 = \{ S \rightarrow B_0 A \mid B_1 S, A \rightarrow B_0 D_1 \mid B_1 S \mid 1, B \rightarrow B_1 D_2 \mid B_0 S \mid 0 \}$$
$$B_0 \rightarrow 0, D_1 \rightarrow AA$$
$$B_1 \rightarrow 1, D_2 \rightarrow BB \}$$

s is the start symbol

(Eq) Convert the following grammar to CNF

$$I \rightarrow C | d | I_a | I_b | I_0 | I_1$$

$$F \rightarrow I | (E)$$

$$T \rightarrow F | T * F$$

$$E \rightarrow T | E + T$$

(Eq) Find a grammar equivalent to

$$S \rightarrow AB | CA$$

$$A \rightarrow a$$

$$B \rightarrow BC | AB$$

$$C \rightarrow aB | b$$

with no useless symbols

(Eq) Begin with the grammar

$$S \rightarrow ASB | \epsilon$$

$$A \rightarrow aAS | a$$

$$B \rightarrow SbS | A | bb$$

- (a) Are there any useless symbols? Eliminate them if so.
- (b) Eliminate ϵ -productions
- (c) Eliminate unit-prod \Rightarrow
- (d) put the grammar into CNF.

Greibach Normal Form (GNF).

In GNF there is no restriction on the number of symbols on the right hand side, but there is restriction on the terminals & variables appear on the right hand side of the prodⁿ.

Defⁿ: Let $G = (V, T, P, S)$ be a CFG. The CFG G is said to be in GNF if all the productions are of the form

$$\boxed{A \rightarrow \alpha\beta}$$

where $\alpha \in T$ and $\beta \in V^*$, i.e. the first symbol on the right hand side of the prodⁿ must be a terminal & it can be followed by zero or more variables.

e.g) convert the following grammar to GNF.

$$S \rightarrow AB1 | 0$$

$$A \rightarrow 00A | 1A1$$

$$B \rightarrow 1A1$$

$$S \rightarrow ABA_1 | 0$$

$$A \rightarrow A_0 A_0 A | A_1 A A_1$$

$$B \rightarrow A_1 A A_1$$

$$A_1 \rightarrow 1$$

$$A_0 \rightarrow 0$$

Now restrict the no of variables on the right hand side of the prodⁿ to two & the resulting grammar in CNF notation is

$$S \rightarrow A D_1 | 0$$

$$A \rightarrow A_0 D_2 | A_1 D_3$$

$$B \rightarrow A_1 D_3$$

$$A_1 \rightarrow 1$$

$$A_0 \rightarrow 0$$

$$D_1 \rightarrow B A_1$$

$$D_2 \rightarrow A_0 A_1$$

$$D_3 \rightarrow A A_1$$

Now let us rename all the variables as shown below:

Let $S = A_1, A = A_2, B = A_3, A_0 = A_4, A_1 = A_5,$

$$D_1 = A_6, D_2 = A_7, D_3 = A_8$$

Now the grammar can be rewritten as:

$$A_1 \rightarrow A_2 A_6 | 0$$

$$A_2 \rightarrow A_4 A_7 | A_5 A_9$$

$$A_3 \rightarrow A_5 A_8$$

$$A_4 \rightarrow 1$$

$$A_5 \rightarrow 0$$

$$A_6 \rightarrow A_3 A_5$$

$$A_7 \rightarrow A_4 A_2$$

$$A_8 \rightarrow A_2 A_5$$

In the above prodⁿ, note that A_4 & A_5 -prodⁿ are in GNF

Consider A_3 prodⁿ: Substituting for A_5 in A_3 -prodⁿ we get

$$A_3 \rightarrow A_5 A_2 = 1 A_8 \rightarrow \text{GNF}.$$

Consider A_2 prodⁿ

$$A_2 \rightarrow A_4 A_2 | A_5 A_2 = 0 A_7 | 1 A_8 \rightarrow \text{GNF}.$$

Consider A_1 prodⁿ.

$$A_1 \rightarrow A_2 A_6 | 0 = (0 A_7 | 1 A_8) A_6 | 0 = 0 A_7 A_6 | 1 A_8 A_6 | 0 \xrightarrow{\text{GNF}}$$

Consider A_6 prodⁿ.

$$A_6 \rightarrow A_3 A_5 \Rightarrow \cancel{A_8} (A_5 A_2) A_5$$

$$A_6 \rightarrow A_5 A_8 A_5 = 1 A_3 A_5 \rightarrow \text{GNF}$$

Consider A_7 prodⁿ

$$A_7 \rightarrow A_4 A_2 = 0 A_2 \rightarrow \text{GNF}.$$

Consider A_3 prodⁿ

$$\begin{aligned} A_3 \rightarrow A_2 A_5 &= (0 A_7 | 1 A_8) A_5 \\ &= 0 A_7 A_5 | 1 A_8 A_5 \rightarrow \text{GNF} \end{aligned}$$

$$G = (V, T, P, S)$$

$$V = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$$

$$T = \{0, 1\}$$

$$P = \{$$

$$A_1 \rightarrow 0 A_7 A_6 | 1 A_8 A_6 | 0$$

$$A_2 \rightarrow 0 A_7 | 1 A_8$$

$$A_3 \rightarrow 1 A_8$$

$$A_4 \rightarrow 0$$

$$A_5 \rightarrow 1$$

$$A_6 \rightarrow 1 A_8 A_5$$

$$A_7 \rightarrow 0 A_2$$

$$A_8 \rightarrow 0 A_7 A_5 | 1 A_7 A_5$$

}

S is start symbol.

Eg

Convert the following grammar

$$A \rightarrow BC$$

$$B \rightarrow CA | b$$

$$C \rightarrow AB | a$$

into GNF.