

Grammar Introduction:

A grammar of a language (G) is defined as,

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

V - finite set of objects called Variables (Non-terminal)

T - finite set of objects called Terminals

$S \in V$ - start symbol

P - finite set of productions.

Types of Grammar:

- Type 0 grammars / unrestricted grammars (Recursively Enumerable Language)
- Type 1 grammars / context sensitive grammars
- Type 2 grammars / context free grammars
- Type 3 grammars / Regular grammar.

1) Type 0 grammars

- No restrictions on the production rules
- Production rule is of the form

$$\alpha \rightarrow \beta \mid \alpha \neq \beta$$

where $\alpha, \beta \rightarrow$ can be strings composed by terminals and non-terminals.

- This grammar can be modeled using Turing Machine.

2) Type 1 grammars

- Context sensitive grammar (CSG) :
- Production rule is of the form

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

Here $A \rightarrow$ Non-terminal symbol

$\alpha, \beta, \gamma \rightarrow$ combination of terminals and non-terminals.

- This grammar can be modeled using linear bounded automata.

3) Type 2 grammar:

- Context Free Grammar (CFG).
- production rule i of the form $A \rightarrow \alpha$
 - A - Non terminal symbol
 - α - Terminal or non-terminal symbol.

- This grammar can be modeled using push down automata.

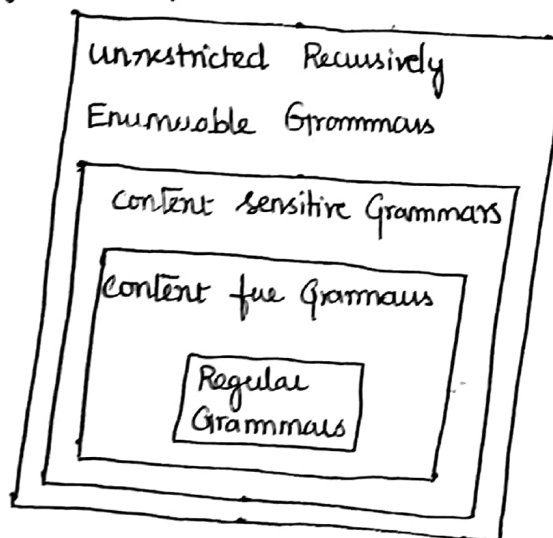
4) Type 3 grammar:

- Regular grammar that describe regular / formal languages.
- production rule consist of,
 - Only one ^{non} terminal at the left hand side.
 - Right hand side having a single terminal and may or may not be followed by non terminals.

$$A \rightarrow a, A \rightarrow aB$$

- This grammar can be modeled using finite automata.

Chomsky hierarchy.



\Rightarrow Regular grammars
 \subseteq
Context free grammars
 \subseteq
Context sensitive grammars
 \subseteq
Unrestricted grammars

Context free Language and Grammars.

Definition:

The context free grammar can be formally defined as a set denoted by $G = (V, T, P, S)$ where V and T are set of non-terminal and terminals respectively

P is set of production rules, $NT \rightarrow NT$
 $NT \rightarrow T$

S is a start symbol.

Example:

$$P = \begin{cases} S \rightarrow S+S \\ S \rightarrow S*S \\ S \rightarrow (S) \\ S \rightarrow 4 \end{cases}$$

Syntax of any English statement is,

SENTENCE \rightarrow NOUN VERB

NOUN \rightarrow Rama / Seeta / Gopal

VERB \rightarrow goes / Writes / sings.

Derived strings: "Rama sings"

Problems:

1) Construct the CFG for the regular expression $(0+1)^*$.

Solution: $CFG = (V, T, P, S)$

$$= (\{S\}, \{0, 1\}, P, S)$$

$$\text{where } P = \begin{cases} S \rightarrow 0S \mid 1S \\ S \rightarrow \epsilon \end{cases}$$

Example:

$$(0+1)^* = \{\epsilon, 0, 1, 01, 10, 00, 11, \dots\}$$

2) Construct CFG for the language L which has all the strings which are all palindromes over $\Sigma = \{a, b\}$

Solution: $G = (\{S\}, \{a, b\}, P, S)$

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

Example: abaaba

$$\begin{aligned} S &\rightarrow aSa \\ &\rightarrow abSba \\ &\rightarrow abaSaba \\ &\rightarrow aba\epsilon aba \\ &\rightarrow abaaba \end{aligned}$$

3) Construct CFG for $\{0^m 1^n \mid 1 \leq m \leq n\}$

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

$$V = \{S, A\}, T = \{0, 1\}$$

$$P = \begin{cases} S \rightarrow 0S1 \mid 0A1 \mid 01 \\ A \rightarrow 1A1 \mid 1 \end{cases}$$

Example: 00111

$$\begin{aligned} S &\rightarrow 0S1 \\ &\rightarrow 00S1 \\ &\rightarrow 001A1 \\ &\rightarrow 00111 \end{aligned}$$

4) Construct CFG for $L = \{a^m b^n c^p \mid m+n=p \text{ and } p \geq 1\}$

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

$$V = \{S, A\} \quad T = \{a, b, c\}$$

$$\begin{aligned} P = \begin{cases} S \rightarrow aSc \mid bAc \mid ac \mid bc \\ A \rightarrow bc \mid bc \end{cases} \end{aligned}$$

5) Consider the alphabet $\Sigma = \{a, b, (,), +, *, \cdot, /, \dots, \epsilon\}$. Construct a grammar that generates all strings in E^* that are RE over $\{a, b\}$.

Solution:

$$\begin{aligned} E &\rightarrow E + E \\ E &\rightarrow E * E \\ E &\rightarrow E \cdot E \\ E &\rightarrow E / E \\ E &\rightarrow a | b | \epsilon \end{aligned}$$

Problems: Grammar to Language

1) If $S \rightarrow aSb | aAb$, $A \rightarrow bAa$, $A \rightarrow ba$ is a CFG then determine CFL.

Solution:

$$\begin{aligned} S &\rightarrow a \underline{S} b \\ &\rightarrow a a \underline{S} b b \\ &\rightarrow a a a \underline{A} b b b \\ &\rightarrow \underbrace{a a a}_{a^n} \underbrace{b b b}_{b^n} \end{aligned}$$

$$L(G) = \{a^n b^m a^m b^n \mid m, n \geq 1\}$$

2) If $S \rightarrow aSa | bSb | \epsilon$ is CFG. Find $L(G)$.

Solution:

$$\begin{aligned} S &\rightarrow a \underline{S} a \\ &\rightarrow aa \underline{S} aa \\ &\rightarrow aa b \underline{S} b aa \\ &\rightarrow \underbrace{aa}_{w} \underbrace{bbaa}_{w^R} \end{aligned}$$

$$\therefore L(G) = \{w w^R \mid w \in (a, b)^*\}$$

3) Find the context free language for the following grammars.

(i) $S \rightarrow aSbS | bSaS | \epsilon$

(ii) $S \rightarrow aSb | ab$.

Solution

(i) $S \rightarrow a \underline{S} b S$ $\therefore L$ containing equal number of a's and b's

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

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

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

(ii) $S \rightarrow aSb | ab$

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

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

L containing equal number of a's followed by equal number of b's

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

Derivations, Ambiguity, Derivation tree

(3)

Derivations: Use the productions from head to body (i.e.) from start symbol expanding till reaches the given string.

Two types of derivations are,

(i) Left Most Derivation (LMD)

(ii) Right Most Derivation (RMD)

⇒ LMD is a derivation in which the leftmost non-terminal is replaced first from the sentential form.

(i.e.) $S \xrightarrow[lm]{*} \alpha$, then α is left sentential form.

⇒ RMD is a derivation in which rightmost non-terminal is replaced first from the sentential form.

(i.e.) $S \xrightarrow{rm}{*} \alpha$, then α is right sentential form.

Derivation tree (parse tree)

- It is a graphical representation for the derivation of the given production rules for a given CFG.

Properties

i) Root node is always a node indicating start symbol.

ii) Derivation is read from left to right

iii) Leaf nodes are always terminal nodes.

iv) Interior nodes are always non-terminal nodes.

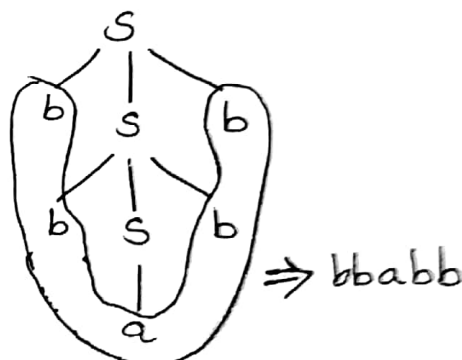
Example:

Consider the grammar G has the production

$S \rightarrow bsb|a|b$ and string "bbabb"

Derivation:

$$\begin{aligned} S &\rightarrow b \underline{S} b \\ &\rightarrow b \underline{b} \underline{S} b \\ &\rightarrow bbabb \end{aligned}$$



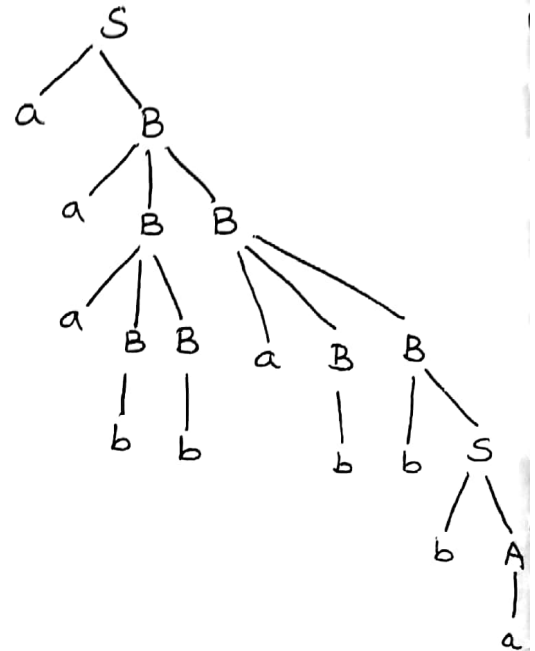
Problems: Construct the derivation tree for the string "aaabbabbbba" using LMD and RMD. using $S \rightarrow aB|bA$, $A \rightarrow a|aS|bAA$, $B \rightarrow b|bS|aBB$

Solution:

LMD:

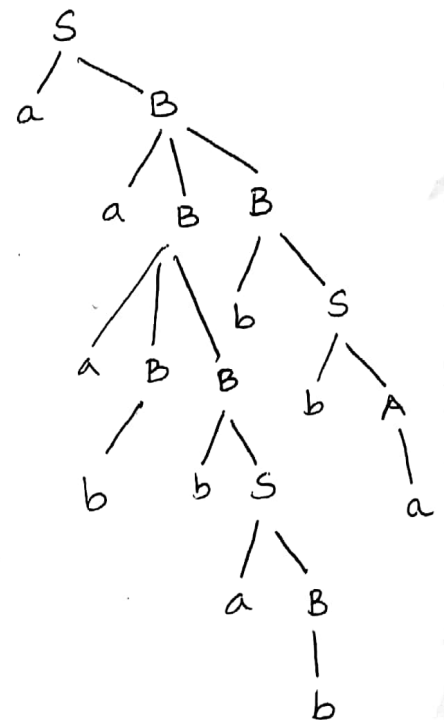
$S \Rightarrow aB$
 $\Rightarrow aaBB \quad (B \rightarrow aBB)$
 $\Rightarrow aaaBBB \quad (B \rightarrow aBB)$
 $\Rightarrow aaabbBB \quad (B \rightarrow b)$
 $\Rightarrow aaabbbB \quad (B \rightarrow b)$
 $\Rightarrow aaabbbaBB \quad (B \rightarrow aBB)$
 $\Rightarrow aaabbbaabB \quad (B \rightarrow b)$
 $\Rightarrow aaabbbaabbs \quad (B \rightarrow bs)$
 $\Rightarrow aaabbbaabbbA \quad (s \rightarrow bA)$
 $\Rightarrow aaabbbaabbbba \quad (A \rightarrow a)$

Parse tree:



RMD:

$S \Rightarrow aB$
 $\Rightarrow aaBB \quad (B \rightarrow aBB)$
 $\Rightarrow aaBbs \quad (B \rightarrow bs)$
 $\Rightarrow aaBbbA \quad (s \rightarrow bA)$
 $\Rightarrow aaBbbba \quad (A \rightarrow a)$
 $\Rightarrow aaaBBbbba \quad (B \rightarrow aBB)$
 $\Rightarrow aaaBbsbbba \quad (B \rightarrow bs)$
 $\Rightarrow aaaBbaBbbba \quad (s \rightarrow aB)$
 $\Rightarrow aaaBbaabbba \quad (B \rightarrow b)$
 $\Rightarrow aaabbaabbba \quad (B \rightarrow b)$



Prob 2: Write a grammar G_1 to recognize all prefix expressions involving all binary arithmetic operators. Construct the parse tree for the sentence $'- * + a b c d e'$

Solution:

$G_1 = (V, T, P, S)$ where

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

$T = \{+, -, *, /, a, b, c, d, e\}$

S is a start symbol

Productions are,

$S \rightarrow A | B | C | D | E$

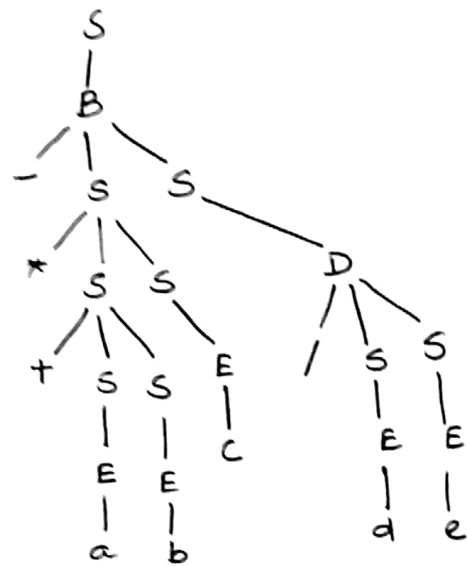
$A \rightarrow + SS$

$B \rightarrow - SS$

$C \rightarrow * SS$

$D \rightarrow / SS$

$E \rightarrow a | b | c | d | e$



Ambiguity:

If there exists more than one parse trees for a given grammar, that means there could be more than one leftmost or rightmost derivation possible and then that grammar is said to be ambiguous grammar.

Problems

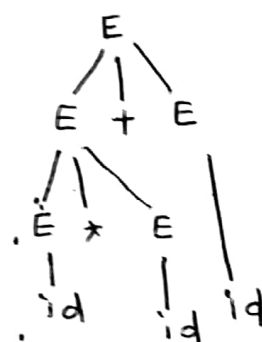
1) The CFG is given by $G_1 = (V, T, P, S)$ where $V = \{E\}$, $T = \{id\}$, $S = \{E\}$

$P = \{E \rightarrow E + E, E \rightarrow E * E, E \rightarrow id\}$. Is the grammar ambiguous?

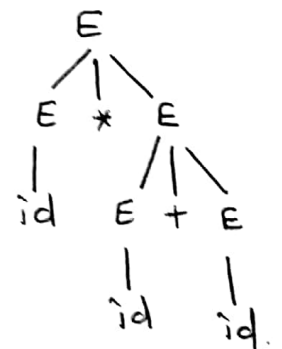
Solution:

Consider the string $id * id + id$.

$E \Rightarrow E + E$
 $\Rightarrow E * E + E$
 $\Rightarrow id * E + E$
 $\Rightarrow id * id + E$
 $\Rightarrow id * id + id$



$E \Rightarrow E * E$
 $\Rightarrow E * E + E$
 $\Rightarrow id * E + E$
 $\Rightarrow id * id + E$
 $\Rightarrow id * id + id$



Here, we obtain two different parse tree for the string $id * id + id$.

\therefore The given grammar is ambiguous.

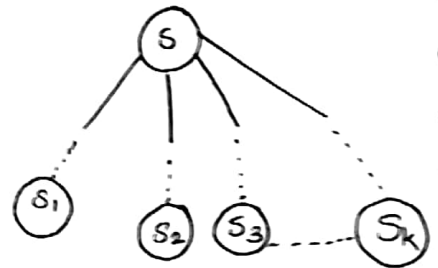
Relationship between derivation and derivation trees.

Theorem: Let $G = (V, T, P, S)$ be a context free grammar. Then $S \xRightarrow{*} \alpha$ if and only if there is a derivation tree in grammar G which gives the string α .

Proof: For a non-terminal S there exists $S \xRightarrow{*} w$ if and only if there is a derivation tree starting from root S and yielding w .

Basis of induction:

Assume that there is only one interior node S .
The derivation tree yielding $S_1, S_2, S_3, \dots, S_n$.
From S it means $S \xRightarrow{*} S_1, S_2, \dots, S_n$
 $\xRightarrow{*} a$ is input string.



Induction hypothesis:

\Rightarrow We assume that for $k-1$ nodes the derivation tree can be drawn. We then prove that for k vertices also we can have a derivation tree.

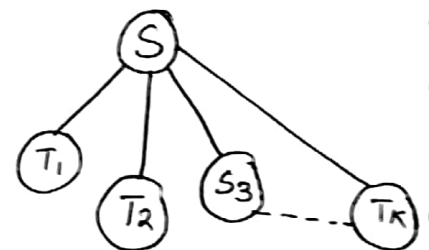
\Rightarrow That means the input string can be derived as $S \rightarrow S_1 S_2 S_3 \dots S_k$.

\Rightarrow There are two cases,

(i) S_i may be a leaf variable

(ii) S_i may be an interior node yielding a .

\Rightarrow The S derives a by fewer number of k steps then $a \in S_1 S_2 S_3 S_4 \dots S_k$.



If $a_i = S_i$ then S_i is leaf node (terminal) and if $S_i \xRightarrow{*} a_i$ then S_i is an interior node.

This proves that $S \xRightarrow{*} S_1, S_2, S_3, \dots, S_n \xRightarrow{*} a$ can be obtained.

Simplification of CFG.

Simplification of grammar means reduction of grammar by removing useless symbols.

- Elimination of useless symbols.
- Elimination of unit productions.
- Elimination of Null production (ϵ).

Elimination of useless symbols.

Any symbol is useful when it appears on the right hand side, in the production rule and generates some terminal string. If no such derivation exists then it is supposed to be an useless symbol.

Example: Eliminate the useless symbol from the following grammar.

$$S \rightarrow aS | A | C$$

$$A \rightarrow a$$

$$B \rightarrow aa$$

$$C \rightarrow aCb \quad aA^b$$

Solution: Production with terminal symbols are

$$A \rightarrow a$$

$$B \rightarrow aa$$

Start symbol $S \rightarrow aS | A | C$

Here, there is no production for B

\therefore B is useless symbol.

$$\text{III}^{\text{ly}} \quad S \rightarrow C \rightarrow aCb \rightarrow aacbb \rightarrow aaaCbbb \rightarrow \dots$$

Here, no terminating symbol for C

\therefore C is useless symbol.

\therefore Eliminate B and C, we get

$\begin{array}{l} S \rightarrow aS A \\ A \rightarrow a \end{array}$

Elimination of ϵ production:

If there is ϵ production, remove it, without changing the meaning of the grammar.

Example: Eliminate ϵ -productions from the CFG.

$$A \rightarrow 0B1|1B1$$

$$B \rightarrow 0B|1B|\epsilon$$

Solution:

$$A \rightarrow 0B1|1B1$$

$$B \rightarrow \epsilon, A \rightarrow 01|11$$

$$B \rightarrow 0B|1B$$

$$B \rightarrow \epsilon, B \rightarrow 01|11$$

\therefore After Elimination,

$$A \rightarrow 0B1|1B1|01|11$$

$$B \rightarrow 0B|1B|01|11$$

Removing Unit productions

The unit productions are the productions in which one non-terminal gives another non-terminal.

$$X \rightarrow Y$$

Example: Eliminate the unit production from following grammar.

$$S \rightarrow AB$$

$$A \rightarrow a$$

$$B \rightarrow C|b$$

$$C \rightarrow D$$

$$D \rightarrow E|bC$$

$$E \rightarrow d|Ab$$

$$\text{Here, } B \rightarrow C$$

$$C \rightarrow D$$

$$D \rightarrow E \text{ are unit productions.}$$

$$\therefore D \rightarrow E|bC \text{ can be written as } D \rightarrow d|Ab|bC$$

$$\text{III}^{\text{rd}} \text{ } C \rightarrow E|bC, B \text{ becomes } B \rightarrow d|Ab|bC|b$$

After removing unit productions.

$$S \rightarrow AB$$

$$A \rightarrow a$$

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

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

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

$$E \rightarrow d|Ab \quad \times$$

Chomsky normal form (CNF)

⑧

A context free grammar $G_1 = (V, T, P, S)$ is said to be in CNF if each production in G_1 is of the form

$$X \rightarrow YZ$$

$$X \rightarrow \alpha, \text{ where } X, Y, Z \in V, \text{ and } \alpha \in T$$

$$\begin{bmatrix} NT \rightarrow NT \cdot NT \\ NT \rightarrow T \end{bmatrix}$$

Problems:

1) Convert the given CFG to CNF $S \rightarrow aSa | bSb | a | b$

Solution:

Productions are

$$S \rightarrow aSa$$

$$S \rightarrow bSb$$

$$S \rightarrow a$$

$$S \rightarrow b$$

Here, the productions which are already in CNF is

$$S \rightarrow a$$

$$S \rightarrow b$$

Apply CNF rule to other productions,

$$S \rightarrow aSa \quad | C_a \rightarrow a$$

$$S \rightarrow C_a S C_a$$

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

$$S \rightarrow bSb$$

$$S \rightarrow C_b S C_b$$

$$| C_b \rightarrow b$$

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

The resultant productions are,

$$S \rightarrow C_a A | C_b B | a | b$$

$$A \rightarrow S C_a$$

$$B \rightarrow S C_b$$

$$C_a \rightarrow a$$

$$C_b \rightarrow b$$

2) Reduce the following grammar to Chomsky normal form.

$$S \rightarrow a | AAB$$

$$A \rightarrow ab | aB | \epsilon$$

$$B \rightarrow aba | \epsilon$$

Solution:

Productions are,

$$S \rightarrow a$$

$$S \rightarrow AAB$$

$$A \rightarrow ab$$

$$A \rightarrow aB$$

$$A \rightarrow \epsilon$$

$$B \rightarrow aba$$

$$B \rightarrow \epsilon$$

Here, Eliminate ϵ production.

$$A \rightarrow \epsilon \text{ and } B \rightarrow \epsilon$$

\Rightarrow After elimination, productions are

$$S \rightarrow a | AAB | AA | AB | B$$

$$A \rightarrow ab | aB | a$$

$$B \rightarrow aba$$

\Rightarrow Eliminate unit production $S \rightarrow A$

$$S \rightarrow a | AAB | AA | AB | aba$$

$$A \rightarrow ab | aB | a$$

$$B \rightarrow aba$$

Productions already in CNF is

$$S \rightarrow a$$

$$S \rightarrow AA$$

$$S \rightarrow AB$$

$$A \rightarrow a$$

Apply CNF rules to other productions,

$$S \rightarrow \underbrace{AAB}_S$$

$$\boxed{S \rightarrow SB}$$

$$S \rightarrow aba$$

$$S \rightarrow Ca \underbrace{CbCa}_{C_1}$$

$$\boxed{\begin{array}{l} S \rightarrow CaC_1 \\ C_1 \rightarrow CbCa \end{array}}$$

$$\boxed{\begin{array}{l} A \rightarrow CaCb \\ A \rightarrow CaB \end{array}}$$

$$B \rightarrow Ca \underbrace{CbCa}_{C_1}$$

$$\boxed{B \rightarrow CaC_1}$$

$$\begin{array}{l} Ca \rightarrow a \\ Cb \rightarrow b \end{array}$$

\therefore The resultant productions are,

$$S \rightarrow SB | CaC_1 | AA | AB | a$$

$$C_1 \rightarrow CbCa$$

$$A \rightarrow CaCb | a | CaB | CaC_1$$

$$B \rightarrow CaC_1$$

3) Convert the given CFG to CNF.

$$\begin{array}{ll} S \rightarrow aB & A \rightarrow bAA \\ S \rightarrow bA & B \rightarrow b \\ A \rightarrow a & B \rightarrow bS \\ A \rightarrow aS & B \rightarrow aBB \end{array}$$

Solution:

Productions already in CNF is,

$$\begin{array}{l} A \rightarrow a \\ B \rightarrow b \end{array}$$

Apply CNF rules to other productions.

$$C_a \rightarrow a, C_b \rightarrow b$$

$$S \rightarrow C_a B$$

$$S \rightarrow C_b A$$

$$A \rightarrow C_a S$$

$$A \rightarrow C_b \underbrace{AA}_{C_1}$$

$$A \rightarrow C_b C_1$$

$$C_1 \rightarrow AA$$

$$B \rightarrow C_b S$$

$$B \rightarrow C_a \underbrace{BB}_{C_2}$$

$$B \rightarrow C_a C_2$$

$$C_2 \rightarrow BB$$

The resultant productions are,

$$S \rightarrow C_a B$$

$$S \rightarrow C_b A$$

$$A \rightarrow C_a S$$

$$A \rightarrow a$$

$$A \rightarrow C_b C_1$$

$$C_1 \rightarrow AA$$

$$B \rightarrow b$$

$$B \rightarrow C_b S$$

$$B \rightarrow C_a C_2$$

$$C_2 \rightarrow BB$$

4) Convert the grammar with productions into CNF

$$A \rightarrow bAB | \lambda, B \rightarrow BAa | \lambda.$$

5) Convert the grammar $S \rightarrow AB | aB, A \rightarrow aab | \epsilon, B \rightarrow bBA$ into CNF

6) Convert to chomsky Normal Form.

$$S \rightarrow A | CB$$

$$A \rightarrow C | D$$

$$B \rightarrow |B|_1$$

$$C \rightarrow 0C|_0$$

$$D \rightarrow 2D|_2$$

Greibach Normal Form (GNF) ③

A grammar $G = (V, T, P, S)$ is said to be in GNF if every production rule is of the form.

$$X \rightarrow \cdot a \alpha$$

where $a \in T, X \in V$
 $\alpha \in V^*$

Right hand side of every productions starts with a terminal, followed by a string of variables of zero or more length.

Problems:

1) Convert the given CFG to GNF

$$S \rightarrow ABA$$

$$A \rightarrow aA \mid \epsilon$$

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

Solution:

Simplify the CFG, Eliminate ϵ production $A \rightarrow \epsilon, B \rightarrow \epsilon$.

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

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

Eliminate unit productions,

$$S \rightarrow ABA \mid AB \mid BA \mid AA \mid aA \mid a \mid bB \mid b$$

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

Apply GNF rules,

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

$$S \rightarrow aABA \mid aBA$$

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

$$S \rightarrow aAB \mid aB$$

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

$$S \rightarrow bBA \mid bA$$

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

$$S \rightarrow aAA \mid aA$$

\therefore The resultant productions are,

$$S \rightarrow aABA \mid aBA \mid aAB \mid aB \mid bBA \mid bA$$

$$S \rightarrow aAA \mid aA \mid a \mid b$$

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

2) Convert given CFG to GNF where $V = \{S, A\}$, $T = \{0, 1\}$ and P is

$$S \rightarrow AA \mid 0$$

$$A \rightarrow SS \mid 1$$

Solution: Replace S as A_1 and A as A_2

CFG becomes,

$$A_1 \rightarrow A_2 A_2 \mid 0$$

$$A_2 \rightarrow A_1 A_1 \mid 1$$

start with $A_2 \rightarrow A_1 A_1 \mid 1$

$$A_2 \rightarrow A_2 A_2 A_1 \mid 0 A_1 \mid 1$$

left recursion.

Introduce B_2
to eliminate left
recursion

$$A_2 \rightarrow A_2 A_2 A_1, A_2 \rightarrow 0 A_1 \mid 1$$

$$B_2 \rightarrow A_2 A_1 \mid A_2 A_1 B_2, A_2 \rightarrow 0 A_1 B_2 \mid 1 B_2$$

The productions are

$$\begin{aligned} A_2 &\rightarrow 0 A_1 \mid 1 \\ A_2 &\rightarrow 0 A_1 B_2 \mid 1 B_2 \end{aligned} \Rightarrow A_1 \rightarrow A_2 A_2 \mid 0$$

$$A_1 \rightarrow 0 A_1 A_2 \mid 1 A_2 \mid 0 A_1 B_2 A_2 \mid 1 B_2 A_2 \mid 0$$

(GNF)

$$B_2 \rightarrow A_2 A_1$$

$$B_2 \rightarrow 0 A_1 \mid 1 A_1 \mid 0 A_1 B_2 A_1 \mid 1 B_2 A_1 \quad (\text{GNF})$$

$$B_2 \rightarrow A_2 A_1 B_2$$

$$B_2 \rightarrow 0 A_1 A_1 B_2 \mid 1 A_1 B_2 \mid 0 A_1 B_2 A_1 B_2 \mid 1 B_2 A_1 B_2 \quad (\text{GNF})$$

Converting Back $A_1 = S$ and $A_2 = A$

$$S \rightarrow 0 S A \mid 1 A \mid 0 S B_2 A \mid 1 B_2 A \mid 0$$

$$A \rightarrow 0 S \mid 1 \mid 0 S B_2 \mid 1 B_2$$

$$B_2 \rightarrow 0 S S \mid 1 S \mid 0 S B_2 S \mid 1 B_2 S$$

$$B_2 \rightarrow 0 S S B_2 \mid 1 S B_2 \mid 0 S B_2 S B_2 \mid 1 B_2 S B_2$$

3) Convert the grammar $S \rightarrow AB$, $A \rightarrow BS|b$, $B \rightarrow SA|a$ into Greibach Normal form. (9)

Solution:

Consider the grammar,

$$S \rightarrow AB, A \rightarrow BS|b, B \rightarrow SA|a$$

Assume $S = A_1$, $A = A_2$, and $B = A_3$.

$$A_1 \rightarrow A_2 A_3$$

$$A_2 \rightarrow A_3 A_1 | b$$

$$A_3 \rightarrow A_1 A_2 | a$$

Consider,

$$A_3 \rightarrow A_1 A_2 | a$$

$$A_3 \rightarrow A_2 A_3 A_2 | a$$

$$A_3 \rightarrow A_3 A_1 A_3 A_2 | b A_3 A_2 | a$$

Now,

$$\underbrace{A_3 \rightarrow A_3 A_1 A_3 A_2}, A_3 \rightarrow b A_3 A_2 | a$$

Introduce B_3 to
eliminate left
recursion

$$\Rightarrow B_3 \rightarrow A_1 A_3 A_2 | A_1 A_3 A_2 B_3$$

$$A_3 \rightarrow b A_3 A_2 | a | b A_3 A_2 B_3 | a B_3. (G_{NF})$$

$$A_2 \rightarrow A_3 A_1 | b$$

$$A_2 \rightarrow b A_3 A_2 A_1 | a A_1 | b A_3 A_2 B_3 A_1 | a B_3 A_1 | b (G_{NF})$$

$$A_1 \rightarrow A_2 A_3$$

$$A_1 \rightarrow b A_3 A_2 A_1 A_3 | a A_1 A_3 | b A_3 A_2 B_3 A_1 A_3 | a B_3 A_1 A_3 | b A_3$$

$$B_3 \rightarrow A_1 A_3 A_2$$

$$\rightarrow b A_3 A_2 A_1 A_3 A_3 A_2 | a A_1 A_3 A_3 A_2 | b A_3 A_2 B_3 A_1 A_3 A_3 A_2 |$$

$$a B_3 A_1 A_3 A_3 A_2 | b A_3 A_3 A_2.$$

$$B_3 \rightarrow A_1 A_3 A_2 B_3$$

$$\rightarrow b A_3 A_2 A_1 A_3 A_3 A_2 B_3 \mid a A_1 A_3 A_3 A_2 B_3 \mid b A_3 A_2 B_3 A_1 A_3 A_2 B_3 \mid a B_3 A_1 A_3 A_2 B_3 \mid b A_3 A_3 A_2 B_3$$

∴ The resultant productions are,

$$A_1 \rightarrow b A_3 A_2 A_1 A_3 \mid a A_1 A_3 \mid b A_3 A_2 B_3 A_1 A_3 \mid b A_3$$

$$A_2 \rightarrow b A_3 A_2 A_1 \mid a A_1 \mid b A_3 A_2 B_3 A_1 \mid a B_3 A_1 \mid b$$

$$A_3 \rightarrow b A_3 A_2 \mid a \mid b A_3 A_2 B_3 \mid a B_3$$

$$B_3 \rightarrow b A_3 A_3 A_2 b A_3 A_2 A_1 A_3 A_2 \mid a A_1 A_3 A_3 A_2 \mid$$

$$b A_3 A_2 B_3 A_1 A_3 A_3 A_2 \mid a B_3 A_1 A_3 A_3 A_2 \mid$$

$$b A_3 A_2 A_1 A_3 A_3 A_2 B_3 \mid a A_1 A_3 A_3 A_2 B_3 \mid$$

$$b A_3 A_2 B_3 A_1 A_3 A_3 A_2 B_3 \mid a B_3 A_1 A_3 A_2 B_3 \mid b A_3 A_3 A_2 B_3$$