UNIT - III

Name: Prof Anand Ghare
College: PVGCOE, Nasik
Class: TE COMPUTER
Subject: TOC

Context Free Grammars & Language

* Context Free Grammar (CFG) ≡ (Regular Grammar)

"A context free Grammar G is a quadruple (V, T, P, S)

Where, V is set of variables
T is set of terminals
P is set of productions
S is a special variable called start symbol $S \in V$.

A production is of the form
$V_i \rightarrow \alpha_i$ where $V_i \in V$ and α_i is string of terminal and variables.

Notations :

1. terminal are denoted by lower case letters a, b, c or digits 0, 1, 2 --- 9.
2. Non-terminal (variables) are denoted by capital letters A, B, C --- Z.
3. A string of terminals or a word $w \in L$ is represented using u, v, w, x, y, z.
4. Sentential form is a string of terminals and variables and it is denoted by α, β, γ etc.

e.g. Anand writes, Anand reads, Sunny watch

A sentense for above word can be written as

<sentense> → <noun> <verb>
Noun → Anand | Sunny
Verb → writes | reads | watch.

P = { sentense → noun | verb ,
noun → Anand | sunny
verb → writes | reads | watch }

Page No.
Date : / /

* Sentential form :

In sentential form, derivation starts from the start symbol through finite application of productions.

A string α is derived so far consist of terminals and non-terminals.

$$S \xRightarrow[G]{*} \alpha \mid \alpha \in (V \cup T)^*$$

- A final string consist of terminals.
- In left sentential form, leftmost symbol is picked up for expansion.
- In right sentential form, rightmost symbol is picked up for expansion.
- A string can be derived in two ways.
  1) Leftmost Derivation
  2) Rightmost Derivation.

e.g.

$S \rightarrow A1B$

$A \rightarrow 0A \mid \epsilon$

$B \rightarrow 0B \mid 1B \mid \epsilon$

where G is given by (V, T, P, S)

with

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

$T = \{0, 1\}$

$P = \{$ production $S \rightarrow A1B,\ A \rightarrow 0A \mid \epsilon,\ B \rightarrow 0B \mid 1B \mid \epsilon \}$

S = start Symbol.

"A Language of grammar can be defined by creating the production rules for the given condition.

The language generated by context free grammar is called as Context free Language (CFL)

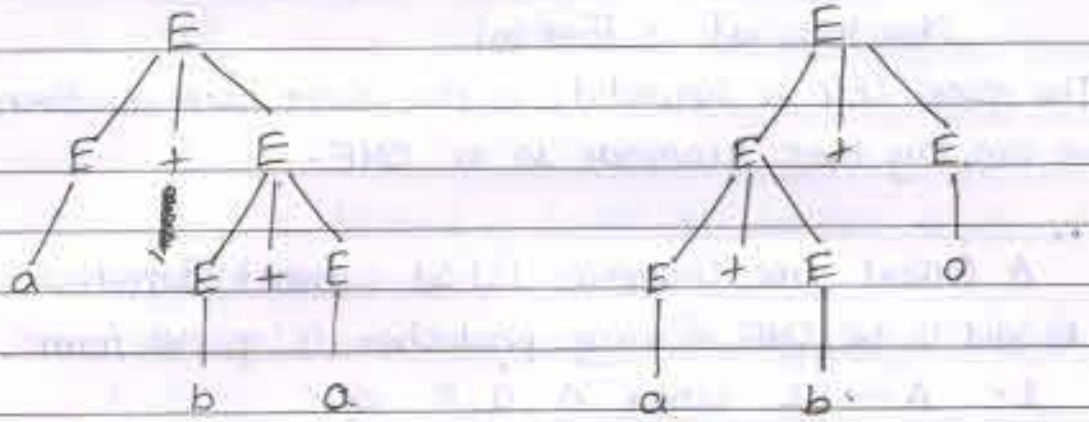
* Ambiguous Grammar :-

"A grammar is said to be ambiguous if the language generated by grammar contains some string that has two different parse tree".

e.g.

$E \rightarrow E+E \mid a \mid b$;

A string a+b+a is generated by given grammar



Parse tree 1 Parse tree 2

As above, two parse tree are generated for string a+b+a so that's why it is ambiguous grammar.

* Elimination of Useless Symbol & Production

A grammar may contain symbols & productions which are not useful for derivation of string.

Two types of symbol are useless

1. Non-generating symbols
2. Non-reachable symbols

* Normal form :

There are 2 imp normal form

1] Chomsky's Normal form

2] Greibach Normal form

1] Chomsky's Normal form (CNF) :

→ The CNF can be defined as

Non-terminal → Non-terminal • Non terminal.

Non-terminal → Terminal

The given CFG is converted in the above format then we can say that grammar is in CNF.

"A Context Free Grammar (CFG) without ε-production is said to be CNF if every production is of the form.

1. $A \rightarrow BC$ where $A, B, C \in V$
2. $A \rightarrow a$ where $A \in V$ and $a \in T$.

The grammar should have no useless symbols.

Every CFG without ε Productions can be converted into an equivalent CNF form."

* Algorithm for CFG to CNF.

→ 1) Eliminate ε-production, unit Production & useless Symbol from grammar.

2) Every variable deriving a string of length 2 or more should consist only of variable.

$A \rightarrow \alpha$ with $|\alpha|$ should consist on variable.

e.g. $A \rightarrow V_1V_2aV_3bV_4$

as $A \rightarrow V_1V_2C_aV_3C_bV_4$ and adding two prod :-

$C_a \rightarrow a$, $C_b = C_b$

3) Every prodn deriving 3 or more variables

($A \rightarrow \alpha$ with $|\alpha| \geq 3$ can be split into cascade Production with each deriving a string of two variables.

e.g $A \rightarrow X_1X_2X_3 \ldots X_n$ where $n \geq 3$ with $|\alpha| \geq 3$

e.g Consider production $A \rightarrow X_1X_2 \ldots X_n$ where $n \geq 3$ and X_i are variable. $A \rightarrow X_1C_1$

$C_1 \rightarrow X_2C_2$

$C_2 \rightarrow X_3C_3$

Date / /

* Greibach Normal form (GNF)

→ "A context free Grammar $G = (V, T, P, S)$ is said to be in GNF if every production is of the form

$$A \rightarrow a\alpha$$

where; $a \in T$ is a terminal &
α is string of zero or more variable

- The Language $L(G)$ should be without β
- RHS of each Prod should start with a terminal followed by a string.

- Removing Left Recursion :

→ Elimination of left Recursion is a important step in algo. used in conversion of a CFG into GNF form

- Left Recursive grammar:

A prod of the form $A \rightarrow A\alpha$ is called left recursive as the left hand side variable appears as the first symbol on the right-hand.