

## 8- Properties of Context Free Languages

### 8.1 Useless Symbols:

Variables or Terminals that do not appear in any derivation of a terminal string from the start symbol.

Omitting Useless symbols from the grammar will not change the language of the Grammar.

#### 8.1.1 Useful Symbols:

The symbols must be both generating and reachable.

Generating Symbols: Symbol  $X$  is generating if  $X \Rightarrow^* w$  for some terminal string  $w$ .

Reachable Symbols: Symbol  $X$  is reachable if  $S \Rightarrow^* \alpha X \beta$  for some sentential forms  $\alpha$  and  $\beta$ .

Useless symbol is not generating or not reachable.

#### 8.1.2 Eliminating Useless Symbols:

- Eliminate symbols that are not generating.
- Eliminate symbols that are not reachable.

**Example:** Consider the grammar

$$\begin{aligned} S &\rightarrow AB \mid a \\ A &\rightarrow b \end{aligned}$$

**Solution:**

$S$  is reachable and also generating.  $\therefore S$  is a useful symbol.

$A$  is generating but not reachable.  $\therefore A$  is a useless symbol.

$\therefore$  After eliminating  $A$  (the useless symbol) the grammar is

$$S \rightarrow AB \mid a$$

**Example:** Eliminate useless symbols from the grammar



$$S \rightarrow ABC \mid AC \mid AB$$
$$A \rightarrow ab$$
$$B \rightarrow c$$
$$D \rightarrow cc \mid d$$

### Solution:

S is reachable and also generating.  $\therefore$  S is a useful symbol.

A is reachable and also generating.  $\therefore$  A is a useful symbol.

B is reachable and also generating.  $\therefore$  B is a useful symbol.

C is reachable but not generating.  $\therefore$  C is a useless symbol.

D is not reachable but generating.  $\therefore$  D is a useless symbol.

$\therefore$  After eliminating C, D (the useless symbols) the grammar is

$$S \rightarrow AB$$
$$A \rightarrow ab$$
$$B \rightarrow c$$

## 8.2 Simplification of Grammars.

### 8.2.1 Chomsky Normal Form (CNF):

A Grammar where all productions are of the form

- $A \rightarrow BC$  OR
- $A \rightarrow b$  is called Chomsky Normal Form.

Where A, B, C are Non Terminals (Variables) and b is a terminal.

To convert a CFG to its equivalent CNF

- Eliminate useless symbols.
- Eliminate  $\varepsilon$  productions i.e., productions of the form  $A \rightarrow \varepsilon$ .
- Eliminate unit productions i.e., productions of the form  $A \rightarrow B$ .

### 8.2.2 Eliminating $\varepsilon$ productions



### Nullable Symbols:

The Symbol A is nullable if  $A \Rightarrow \varepsilon$ .

Finding Nullable Symbols (Algorithm):

- **BASIS:** If  $A \rightarrow \varepsilon$  is in G, then A is nullable.
- **INDUCTION:** If  $B \rightarrow C_1 C_2 \dots C_k$  is a production in G, where each  $C_i$  is nullable then B is nullable.

### Eliminating $\varepsilon$ productions

- Find Nullable Symbols.
- If  $A \rightarrow X_1 X_2 \dots X_{i-1} X_i X_{i+1} \dots X_k$  is a production in G and  $X_i$  is a nullable symbol then rewrite the production as

$$A \rightarrow X_1 X_2 \dots X_{i-1} X_i X_{i+1} \dots X_k \mid X_1 X_2 \dots X_{i-1} X_{i+1} X_{i+2} \dots X_k$$

**Example:** Consider the Grammar

1.  $S \rightarrow AB$
2.  $A \rightarrow aAA \mid \varepsilon$
3.  $B \rightarrow bBB \mid \varepsilon$

Eliminate  $\varepsilon$  productions from the above grammar.

**Solution:**

**Finding nullable symbols.**

**Basis:**

- A is a nullable symbol. ( $\because A \rightarrow \varepsilon$  is a production in the Grammar)
- B is a nullable symbol. ( $\because B \rightarrow \varepsilon$  is a production in the Grammar)

**Induction:**

- $S \rightarrow AB$  is in the grammar and both A and B are nullable.
- $\therefore S$  is a nullable symbol.

The set N of nullable symbols  $N = \{S, A, B\}$

**Eliminating  $\varepsilon$  productions:** If a production contains a nullable symbol then include the productions with and without nullable symbols in the grammar.



- $S \rightarrow AB \mid A \mid B$
- $A \rightarrow aAA \mid aA \mid a$
- $B \rightarrow bBB \mid bB \mid b$

**Example:** Consider the Grammar

1.  $S \rightarrow aTa$
2.  $T \rightarrow ABC$
3.  $A \rightarrow aA \mid C$
4.  $B \rightarrow Bb \mid C$
5.  $C \rightarrow c \mid \varepsilon$

Eliminate  $\varepsilon$  productions from the above grammar.

Solution:

Set of nullable symbols  $N = \{T, A, B, C\}$

Grammar after eliminating  $\varepsilon$  productions

1.  $S \rightarrow aTa \mid aa$
2.  $T \rightarrow ABC \mid AB \mid AC \mid BC \mid A \mid B \mid C$
3.  $A \rightarrow aA \mid a \mid C$
4.  $B \rightarrow Bb \mid b \mid C$
5.  $C \rightarrow c$

### 8.2.3 Eliminating Unit Productions

- A Unit production is a production of the form  $A \rightarrow B$  where both A and B are Variables or Non Terminals.
- If  $A \rightarrow B$  is a production in the grammar and  $B \rightarrow \beta$  is in G then replace the production  $A \rightarrow B$  by the production  $A \rightarrow \beta$ .

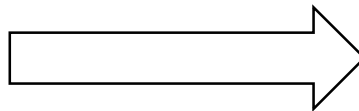
**Example:** Remove Unit Productions from the grammar

$S \rightarrow XY$



$$\begin{aligned}
 X &\rightarrow A \\
 A &\rightarrow B \mid a \\
 B &\rightarrow b \\
 Y &\rightarrow T \\
 T &\rightarrow Y \mid c
 \end{aligned}$$

**Solution:**

$$\begin{aligned}
 S &\rightarrow XY \\
 X &\rightarrow B \mid a \\
 A &\rightarrow b \mid a \\
 B &\rightarrow b \\
 Y &\rightarrow Y \mid c \\
 T &\rightarrow T \mid c
 \end{aligned}$$


$$\begin{aligned}
 S &\rightarrow XY \\
 X &\rightarrow b \mid a \\
 A &\rightarrow b \mid a \\
 B &\rightarrow b \\
 Y &\rightarrow c \\
 T &\rightarrow c
 \end{aligned}$$

### 8.2.4 Converting a Grammar to CNF

1. Eliminate  $\varepsilon$  Productions.
2. Eliminate Unit Productions.
3. Eliminate Useless Symbols.
4. If the length of the Right Hand Side of a production is  $\geq 2$  and there is a terminal  $a$  on the RHS, then replace the terminal  $a$  by a nonterminal  $T_a$  and add a new production  $T_a \rightarrow a$  to the grammar.
5. If the length of the Right Hand Side of a production is  $> 2$ , i.e., the production is of the form  $A \rightarrow X_1X_2 \dots X_k$  ( $k > 2$ ) rewrite the production as

$$\begin{aligned}
 A &\rightarrow X_1M_1 \\
 M_1 &\rightarrow X_2M_2 \\
 M_2 &\rightarrow X_3M_3 \\
 &\vdots \\
 &\vdots \\
 &\vdots \\
 M_{k-2} &\rightarrow X_{k-1}X_k
 \end{aligned}$$

Where  $M_1, M_2, \dots, M_{k-2}$  are the new non terminals.

**Example:** Convert the below CFG to CNF.

$$S \rightarrow aACa$$


$$A \rightarrow B \mid a$$

$$B \rightarrow C \mid c$$

$$C \rightarrow cC \mid \varepsilon$$

### 1. Eliminate $\varepsilon$ Productions:

$$N = \{C, B, A\}$$

$$S \rightarrow aACa \mid aAa \mid aCa \mid aa$$

$$A \rightarrow B \mid a$$

$$B \rightarrow C \mid c$$

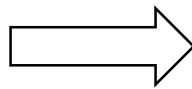
$$C \rightarrow cC$$

### 2. Eliminate Unit productions:

$$S \rightarrow aACa \mid aAa \mid aCa \mid aa$$

$$A \rightarrow B \mid a$$

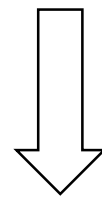
$$B \rightarrow C \mid c$$

$$C \rightarrow cC$$


$$S \rightarrow aACa \mid aAa \mid aCa \mid aa$$

$$A \rightarrow C \mid c \mid a$$

$$B \rightarrow cC \mid c$$

$$C \rightarrow cC$$


$$S \rightarrow aACa \mid aAa \mid aCa \mid aa$$

$$A \rightarrow cC \mid c \mid a$$

$$B \rightarrow cC \mid c$$

$$C \rightarrow cC$$

### 3. Eliminate Useless Symbols.

$$S \rightarrow aACa \mid aAa \mid aCa \mid aa$$


$$A \rightarrow cC \mid c \mid a$$

$$B \rightarrow cC \mid c$$

$$C \rightarrow cC$$

B is a useless symbol because it is not reachable from the start symbol.

Eliminate B. The grammar is

$$S \rightarrow aACa \mid aAa \mid aCa \mid aa$$

$$A \rightarrow cC \mid c \mid a$$

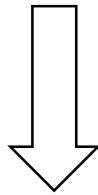
$$C \rightarrow cC$$

#### 4. If the length of the Right Hand Side of a production is $\geq 2$

$$S \rightarrow aACa \mid aAa \mid aCa \mid aa$$

$$A \rightarrow cC \mid c \mid a$$

$$C \rightarrow cC$$



$$S \rightarrow T_aACT_a \mid T_aAT_a \mid T_aCT_a \mid T_aT_a$$

$$A \rightarrow T_cC \mid c \mid a$$

$$C \rightarrow T_cC$$

$$T_a \rightarrow a$$

$$T_c \rightarrow c$$

#### 5.

$$S \rightarrow T_aACT_a \mid T_aAT_a \mid T_aCT_a \mid T_aT_a$$

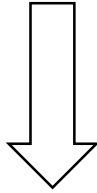
$$A \rightarrow T_cC \mid c \mid a$$

$$C \rightarrow T_cC$$

$$T_a \rightarrow a$$

$$T_c \rightarrow c$$




$$S \rightarrow T_a M_1 \mid T_a M_2 \mid T_a M_3 \mid T_a T_a$$
$$A \rightarrow T_c C \mid c \mid a$$
$$C \rightarrow T_c C$$
$$T_a \rightarrow a$$
$$T_c \rightarrow c$$

**CNF**

$$M_1 \rightarrow A M_3$$
$$M_2 \rightarrow A T_a$$
$$M_3 \rightarrow C T_a$$

**Example:** Convert the below CFG to CNF.

$$S \rightarrow aSa$$
$$S \rightarrow B$$
$$B \rightarrow bbC$$
$$B \rightarrow bb$$
$$C \rightarrow \varepsilon$$
$$C \rightarrow cC$$

**Example:** Convert the below CFG to CNF.

$$S \rightarrow ABC$$
$$A \rightarrow aC \mid D$$
$$B \rightarrow bB \mid \varepsilon \mid A$$
$$C \rightarrow Ac \mid \varepsilon \mid Cc$$
$$D \rightarrow aa$$

### 8.3 Context Free Languages:





Languages defined by the CFG are called context free languages.

Language of a CFG:

If G is a CFG then

$$L(G) = \{ w \mid S \xRightarrow{*} w \}$$

### 8.3.1 Properties of Parse Tree (Introduction)

- Let G be the CFG i.e.,  $G = (V, T, P, S)$ .
- Let m be the number of non terminals or variables in G.
- Let b be the branching factor i.e., the length of the longest right hand side of any production in the Grammar G. In other words b is the maximum number of children of any node in the parse tree.
- Let h be the height of the parse tree, i.e., the length of the longest path from root to any of the leaf nodes.

### 8.3.2 Pumping Lemma for Context Free Languages

**Statement:** Let L be a CFL. There exists a constant n such that , if z is any string in L and  $|z| \geq n$ , then we can write  $z=uvwxy$ , subject to the following conditions:

1.  $|vwx| \leq n$ . That is the middle portion is not too long.
2.  $vx \neq \varepsilon$ . Since v and x are the strings to be "pumped ", at least one of them must not be empty.
3. For all  $i \geq 0$  ,  $uv^iwx^iy$  is in L. That is two strings v and x may be pumped in any number of times, the resulting string still will be a member of L.

**Example:** Prove that the language  $L = \{0^n1^n2^n \mid n \geq 1\}$  is a non-context free language.



### Proof:

- Let  $L$  be a context Free Language.
- Let  $z = 0^n 1^n 2^n$ .
- Split  $z$  as  $z = uvwxy$ , where  $|vwx| \leq n$  and  $v$  and  $x$  are not both  $\epsilon$ .
- $vwx$  cannot contain both 0's and 2's since the last 0 and the first 2 are separated by  $n + 1$  positions.
- We shall prove  $L$  contains some string  $\notin L$ , thus contradicting the assumption.

There are two cases

1.  $vwx$  has no 2's.
  - a. Then  $vx$  consists of only 0's and 1's and has at least one of these symbols.
  - b.  $uw$  has  $n$  2's but has fewer than  $n$  0's or fewer than  $n$  1's or both.
  - c.  $\therefore uw \notin L$  and  $L$  is not a CFL.

2.  $vwx$  has no 0's.

Similarly  $uw$  has  $n$  0's but has fewer than  $n$  1's or fewer than  $n$  2's or both.

**Example:** Prove that the language  $L = \{0^i 1^j 2^i 3^j \mid i, j \geq 1\}$  is a non context free language.

### Proof:

- Let  $L$  be a context Free Language.
- Pick  $z = 0^n 1^n 2^n 3^n$ .
- Split  $z$  as  $z = uvwxy$ , where  $|vwx| \leq n$  and  $v$  and  $x$  are not both  $\epsilon$ .
- $vwx$  either contains single symbol or two adjacent symbols.
- We shall prove  $L$  contains some string  $\notin L$ , thus contradicting the assumption.

There are two cases

1.  $vwx$  consists of only one symbol.



- a. Then  $uwy$  consists of  $n$  occurrences of three symbols and fewer than  $n$  occurrences of the fourth symbol.
  - b.  $\therefore uwy \notin L$  and  $L$  is not a CFL.
2.  $vwxy$  consists of two adjacent symbols (say 1's and 2's).
  - a. Then  $uwy$  is missing some 1's or some 2's or both.
  - b.  $\therefore uwy \notin L$  and  $L$  is not a CFL.

### 8.3.3 Closure Properties of Context Free Languages

1. Context Free Languages are closed Under Union.
2. Context Free Languages are closed Under Concatenation.
3. Context Free Languages are closed Under Kleene Closure.
4. Context Free Languages are closed Under Reversal.

#### 1. Context Free Languages are closed Under Union.

If  $L_1$  and  $L_2$  are Context Free Languages then  $L_1 \cup L_2$  is a Context Free Language.

##### Proof:

$L_1$  is a CFL  $\therefore L_1 = L(G_1)$  where  $G_1 = (V_1, T_1, P_1, S_1)$

$L_2$  is a CFL  $\therefore L_2 = L(G_2)$  where  $G_2 = (V_2, T_2, P_2, S_2)$

Build a new Grammar  $G$  such that

$$L(G) = L_1 \cup L_2 = L(G_1) \cup L(G_2)$$

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

$$V = V_1 \cup V_2$$

$$T = T_1 \cup T_2$$

$$P = P_1 \cup P_2 \text{ and}$$

$$S \rightarrow S_1 \mid S_2$$

#### 2. Context Free Languages are closed Under Concatenation.

If  $L_1$  and  $L_2$  are Context Free Languages then  $L_1 L_2$  is a Context Free Language.



**Proof:**

$L_1$  is a CFL  $\therefore L_1 = L(G_1)$  where  $G_1 = (V_1, T_1, P_1, S_1)$

$L_2$  is a CFL  $\therefore L_2 = L(G_2)$  where  $G_2 = (V_2, T_2, P_2, S_2)$

Build a new Grammar  $G$  such that

$$L(G) = L_1 L_2 = L(G_1) L(G_2)$$

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

$$V = V_1 \cup V_2$$

$$T = T_1 \cup T_2$$

$$P = P_1 \cup P_2 \text{ and}$$

$$S \rightarrow S_1 S_2$$

**3. Context Free Languages are closed Under Closure.**

If  $L$  is a Context Free Language then  $L^*$  is a Context Free Language.

**Proof:**

$L$  is a CFL  $\therefore L = L(G)$  where  $G = (V, T, P, S)$

Build a new Grammar  $G_1$  such that

$$L(G_1) = L^* = L(G)^*$$

Where  $G_1 = (V_1, T_1, P_1, S_1)$  and

$$V_1 = V \cup S_1$$

$$T_1 = T$$

$$P_1 = P \cup \{ S_1 \rightarrow \varepsilon, S_1 \rightarrow S_1 S \} \text{ and}$$

$$S_1 \rightarrow \varepsilon \mid S_1 S$$

**4. Context Free Languages are closed Under Reversal.**

If  $L$  is a Context Free Language then  $L^R$  is a Context Free Language.



**Proof:**

$L$  is a CFL  $\therefore L = L(G)$  where  $G = (V, T, P, S)$

Build a new Grammar  $G_1$  such that

$$L(G_1) = L^R = L(G)^R$$

Where  $G_1 = (V_1, T_1, P_1, S_1)$  and

$$V_1 = V$$

$$T_1 = T$$

$P_1$  = Each production  $A \rightarrow \alpha$  is replaced by  $A \rightarrow \alpha^R$  and

$$S_1 = S$$

