UNIT-3

Context-Free Grammar and Languages

13/12/2021 Date — / — CFG CFL

Chomskey hierarchy :→

Chomsky hierarchy

- Type-0 → structured grammar
- Type-1 → context-sensitive grammar
- Type-2 → context-free grammar
- Type-3 → Regular grammar

Simple-sentence : subject verb object

Subject ⟶ noun

object ⟶ noun / pronoun.

Complex sentence : simple-sentence conjuction simple-sentence

conjunction : and / or / but.

Context-free grammar: A CFG ~~give a space~~ $G = (V, T, P, S)$ ~~as follow:~~ contains the following :→

where V = set of variables or non-terminals

T = set of terminals

P = set of production or rules

S = start symbol.

Variables or non terminals are Simple sentence, subject, object

Set of terminals are verb, noun and pronoun.

Set of production include all.

Simple sentence is the start symbol.

Write a CFG for any number of a's

Sol:- CFG : $S \to aS$

$S \to \epsilon$.

Page No.

Write CFG for atleast one a $a a^*$.

Sol:→ $S \rightarrow aS$

Date ___ $S \rightarrow a$

Saathi

Write CFG for any number of a's and b's including empty string $(a+b)^*$

abba

$S \Rightarrow aS$

Sol:→ $S \rightarrow aS$ $S \rightarrow aS \mid bS \mid \epsilon$ $S \Rightarrow abS$

$S \rightarrow bS$ or $S \Rightarrow abbS$

$S \rightarrow \epsilon$ $S \Rightarrow abba$

Write CFG that begin with a $a(a+b)^*$

Sol:→ ~~$A \rightarrow aS$~~ $S \rightarrow aA$

~~$S \rightarrow aS$~~ $A \rightarrow aA$

~~$S \rightarrow bS$~~ $A \rightarrow bA$

~~$S \rightarrow \epsilon$~~ $A \rightarrow \epsilon$

Write CFG that has substring ab. $(a+b)^* ab (a+b)^*$

Sol:→ $S \rightarrow AabA$

$A \rightarrow aA$

$A \rightarrow bA$

$A \rightarrow \epsilon$.

$G = \{V, T, P, S\}$

$V = \{S, A\}$

$T = \{a, b\}$

Write CFG for palindrome. Palindrome has 0's and 1's.

Sol:→ $S \rightarrow 0S0 \mid 1S1 \mid 0 \mid 1 \mid \epsilon$

Or

$S \rightarrow 0S0$

$S \rightarrow 1S1$ It works for both odd

$S \rightarrow 0$ and even palindrome.

$S \rightarrow 1$

$S \rightarrow \epsilon$.

Write CFG for equal no. of a's ~~and~~ followed by equal no. of b's

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

Date ___/___/___

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

Sol:- $\left.\begin{array}{l} S \rightarrow aSb \\ S \rightarrow \epsilon \end{array}\right\} P.$ $V = \{S\}$ $T = \{a, b\}$

S = start symbol

Write CFG for the given language.

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

sol:-

$S \rightarrow aSb$

$S \rightarrow ab.$

Write CFG that has at least 2 a's

Sol:→ RE = aaa^*

$S \rightarrow aS$ $S \rightarrow aS \mid aa$

$S \rightarrow aa$

$V = \{S\}$

$T = \{a\}$

$P = S \rightarrow aS \mid aa$

S = start symbol.

Write CFG for even no. of a's

Sol:→ RE = $(aa)^*$

$S \rightarrow aaS$

$S \rightarrow \epsilon.$

Write CFG for odd no of a's

sol:→

$S \rightarrow a$

$S \rightarrow aaS$

Write CFG for no. of a's divisible by 3

Sol:→ RE = $(aaa)^*$

$S \rightarrow \epsilon$

$S \rightarrow aaaS$

Write CFG for the following language.

$L = \{ w \mid |w| \bmod 3 = 0, \ w \in \{a, b\}^* \}$

Date ___/___/___

sol:→ $S \to \epsilon \mid abaS \mid babS \mid bbaS \mid aabS \mid aaaS \mid bbbS$

Or

$S \to AAAS$

$S \to \epsilon$

$A \to a \mid b$

Write CFG for $\{ w \mid |w| \bmod 3 > 0, \ w \in a^* \}$

sol:→ $S \to a \mid aa \mid aaaS$.

Write CFG for $n_a(w)$ divisible by 2, $w \in \{a, b\}^*$

sol:→ $RE = (b^* a b^* a b^*)^*$

$S \to SaSaS \mid bS \mid \epsilon$.

Write CFG for no. of a's equal to no. of b's

sol:- $S \to ASB \mid BSA \mid \epsilon \mid SS$

$A \to a$

$B \to b$

Or $S \to SS \mid aSb \mid bSa \mid \epsilon$

Write a CFG for a balanced parenthesis

sol: $S \to (S)$

$S \to SS$

$S \to \epsilon$

Write a CFG $L = \{ 0^m 1^m 2^n \mid m \geq 1, n \geq 0 \}$

sol: $S \to AB$

$A \to 01 \mid 0A1$

$B \to 2B \mid \epsilon$.

Write a CFG for following language

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

sol.- $S \to aSB \mid B$
$B \to bB \mid b.$

Saathi

Date __/__/____

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

sol: $S \to aSb \mid aaa.$

**

Write CFG for the following language:
$L = \{0^i 1^j \mid i \neq j, i \geq 0, j \geq 0\}$

Sol:→ $S \to 0S1 \mid A \mid B$
$A \to 0A \mid 0$
$B \to 1B \mid 1$

**

Write CFG for following language:
$L = \{a^n b^m c^k \mid n + 2m = k \text{ for } n \geq 0, m \geq 0\}$

Sol:→ $a^n b^m c^k$
$a^n b^m c^{n+2m}$
$a^n b^m c^n c^{2m}$
$a^n b^m c^{2m} c^n$
$S \to aSc \mid A$
$A \to bAcc \mid \epsilon.$

Write CFG for following language:
$L = \{a^n b^m c^k \mid m = n + k \text{ for } n \geq 0, k \geq 0\}$
$a^n b^m c^k$
$a^n b^{n+k} c^k$
$a^n b^n b^k c^k$
$S \to \cancel{aSb}AB$
$A \to aAb \mid \epsilon$
$B \to bBc \mid \epsilon$

Write CFG for following language:
$L = \{w \mid |w| \bmod 3 \neq |w| \bmod 2, w \in \{a\}^*\}$

Sol:→ Valid :→ 2, 3, 4, 5, 8, 9, 10, 11, 14 - - - - -
~~Bllaz~~

$S \to aa \mid aaa \mid aaaa \mid aaaaa \mid aaaaaaS$ ~~aaa~~

Page No.

Write CFG for following language:
$L = \{ w \mid |w| \bmod 3 \geq |w| \bmod 2,\ w \in \{a\}^* \}$

Date ___/___/___

Sol: Valid: 0, 1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14. ...

$S \rightarrow \epsilon \mid a \mid aa \mid aaaa \mid aaaaa \mid aaaaaaS$

Derivations using a Grammar

→ Leftmost Derivation (LMD)
→ Rightmost Derivation (RMD)

Derive the string 3a's followed by 3b's
Sol:-
$S \Rightarrow aSb \quad \because S \rightarrow aSb$
$\Rightarrow aaSbb \quad \because S \rightarrow aSb$
$\Rightarrow aaaSbbb \quad \because S \rightarrow aSb$
$\Rightarrow aaabbb \quad \because S \rightarrow \epsilon$

~~Derive the string~~ ababba.
Sol:→
~~$S \Rightarrow aSb$~~ ~~$S \rightarrow SS$~~
~~$S \rightarrow aSb$~~
~~$S \rightarrow$~~

Derive the string for equal no. of a's and equal no of b's
Sol:→
$S \rightarrow SS$
$S \rightarrow aSb$
$S \rightarrow bSa$
$S \rightarrow \epsilon$.

Derive the string ababba using LMD & RMD.
Sol: LMD:
~~$S \Rightarrow SS$~~
~~$S \Rightarrow SSS$~~ $\quad \because$ ~~$S \rightarrow SS$~~
$S \Rightarrow aSb\,aSb\,bSa \quad \because$ ~~$S \Rightarrow aSb\ aSb\ bSa$~~
$S \Rightarrow$ ~~$ababba$~~ $\quad$ ~~$S \Rightarrow \epsilon$~~

$$
\begin{aligned}
S &\underset{lm}{\Rightarrow} SS && \because S \to SS \\
S &\underset{lm}{\Rightarrow} SSS && \because S \to aSb \\
&\underset{lm}{\Rightarrow} aSbSS && \because S \to \epsilon \\
&\underset{lm}{\Rightarrow} abSS && \because aS \to aSb \\
&\underset{lm}{\Rightarrow} abaSbS && \because S \to \epsilon \\
&\underset{lm}{\Rightarrow} ababS && \because S \to bSa \\
&\underset{lm}{\Rightarrow} ababbSa && \because S \to \epsilon \\
&\underset{lm}{\Rightarrow} ababba
\end{aligned}
$$

RMD:

$$
\begin{aligned}
S &\underset{rm}{\Rightarrow} SS && \because \\
&\underset{rm}{\Rightarrow} SSS && \because S \to SS \\
&\underset{rm}{\Rightarrow} SSbSa && \because S \to bSa \\
&\underset{rm}{\Rightarrow} SSba && \because S \to \epsilon \\
&\underset{rm}{\Rightarrow} SaSbba && \because S \to aSb \\
&\underset{rm}{\Rightarrow} Sabba && \because S \to \epsilon \\
&\underset{rm}{\Rightarrow} aSbabba && \because S \to aSb \\
&\underset{rm}{\Rightarrow} ababba && \because S \to \epsilon
\end{aligned}
$$

LMD :- At each step, we replace the leftmost variable by one of its production bodies, such a derivation is called a leftmost derivation.

RMD :- At each step, we replace the ~~different~~ rightmost variable by one of its production bodies, such a derivation is called a rightmost derivation

Give leftmost and rightmost derivation of following strings:

a) 00101
b) 1001
c) 00011

$$
\begin{aligned}
S &\to A1B \\
A &\to 0A \mid \epsilon \\
B &\to 0B \mid 1B \mid \epsilon .
\end{aligned}
$$

Sol :→ a) 00101.

LMD :

$S \underset{lm}{\Rightarrow} A1B$

$\underset{lm}{\Rightarrow} 0A1B$ $A \to 0A$

$\underset{lm}{\Rightarrow} 00A1B$ $A \to 0A$

$\underset{lm}{\Rightarrow} 001B$ $A \to \epsilon$

$\underset{lm}{\Rightarrow} 0010B$ $B \to 0B$

$\underset{lm}{\Rightarrow} 00101B$ $B \to 1B$

$\underset{lm}{\Rightarrow} 00101$ $B \to \epsilon$.

RMD :

$S \underset{rm}{\Rightarrow} A1B$

$S \underset{rm}{\Rightarrow} A10B$ $B \to 0B$

$S \underset{rm}{\Rightarrow} A101B$ $B \to 1B$

$S \underset{rm}{\Rightarrow} A101$ $B \to \epsilon$

$S \underset{rm}{\Rightarrow} 0A101$ $A \to 0A$

$S \underset{rm}{\Rightarrow} 00A101$ $A \to 0A$

$S \underset{rm}{\Rightarrow} 00101$ $A \to \epsilon$.

b) 1001.

Sol :-

LMD :→

$S \underset{lm}{\Rightarrow} A1B$

$S \underset{lm}{\Rightarrow} \cancel{0A1B}\ A1B$ $A \to \epsilon$.

$S \underset{lm}{\Rightarrow} 10B$ $B \to 0B$

$S \underset{lm}{\Rightarrow} 100B$ $B \to 0B$

$S \underset{lm}{\Rightarrow} 1001B$ $B \to \cancel{0}1B$

$S \underset{lm}{\Rightarrow} 1001$ $B \to \epsilon$.

RMD :→

$S \underset{rm}{\Rightarrow} A1B$

$\underset{rm}{\Rightarrow} \cancel{A}\ \cancel{0}\ A10B$ $B \to \cancel{1}0B$

$\underset{rm}{\Rightarrow} A100B$ $B \to 0B$

$\underset{rm}{\Rightarrow} A1001B$ $B \to 1B$

$\underset{rm}{\Rightarrow} A1001$ $\quad \because B \rightarrow \epsilon$

$\underset{rm}{\Rightarrow} \cancel{A}1001$ $\quad \cancel{B}\, A \rightarrow \epsilon.$

Date

c) 00011

sol:→ LMD : $S \underset{lm}{\Rightarrow} A1B$

$S \underset{lm}{\Rightarrow} 0A1B$ $\quad A \rightarrow 0A$

$S \underset{lm}{\Rightarrow} 00A1B$ $\quad A \rightarrow 0A$

$S \underset{lm}{\Rightarrow} 000A1B$ $\quad A \rightarrow 0A$

$S \underset{lm}{\Rightarrow} 0001B$ $\quad A \rightarrow \epsilon$

$S \underset{lm}{\Rightarrow} 00011B$ $\quad B \rightarrow 1B$

$S \underset{lm}{\Rightarrow} 00011$ $\quad B \rightarrow \epsilon.$

RMD : $S \underset{rm}{\Rightarrow} A1B$

$S \underset{rm}{\Rightarrow} A11B$ $\quad B \rightarrow 1B$

$S \underset{rm}{\Rightarrow} A11$ $\quad B \rightarrow \epsilon$

$S \underset{rm}{\Rightarrow} 0A11$ $\quad A \rightarrow 0A$

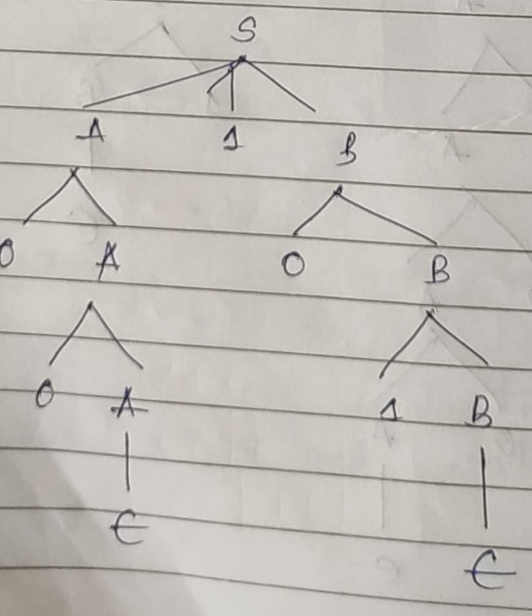
$S \underset{rm}{\Rightarrow} 00A11$ $\quad A \rightarrow 0A$

$S \underset{rm}{\Rightarrow} 000A11$ $\quad A \rightarrow 0A$

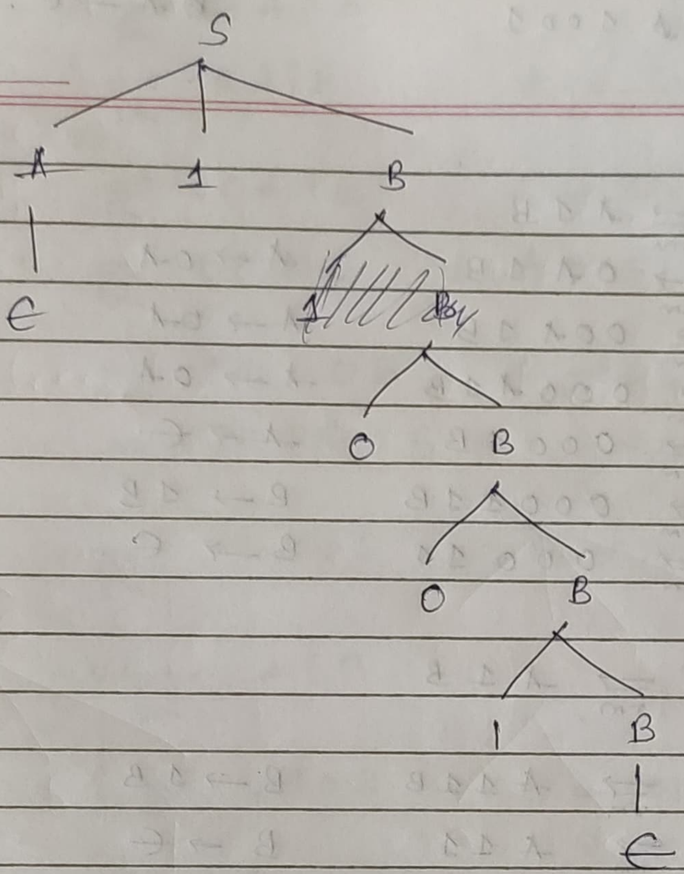
$S \underset{rm}{\Rightarrow} 00011$ $\quad A \rightarrow \epsilon$

Parse Tree (Derivation Tree)

(1) 00101

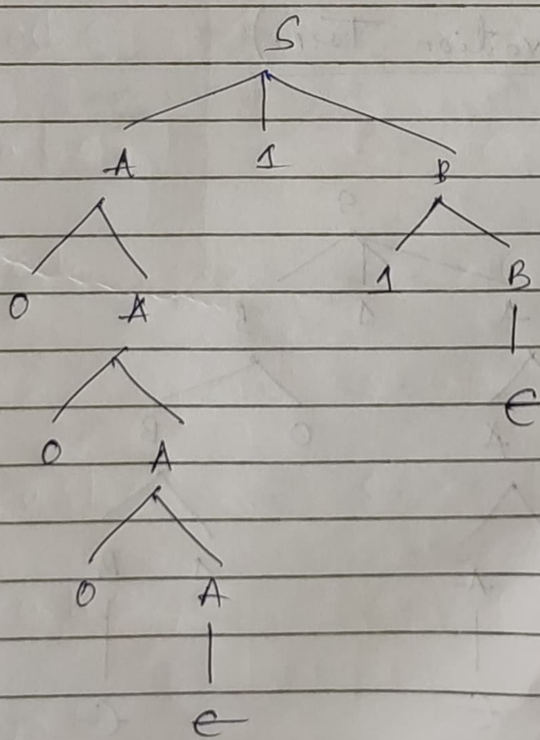


(ii) 1001



Concatinating of leaves from left to right in parse tree is called as yield of a parse tree

(iii) 00011



Ambiguity in Grammars and Languages.

$E \rightarrow E+E$

Date ___/___/___

$E \rightarrow E*E$

$E \rightarrow I$

$I \rightarrow a$

$I \rightarrow b$

i) $a+b*a$

Sol:→

$E \Rightarrow E+E$

$E \Rightarrow I+E \quad \because E \rightarrow I$

$E \Rightarrow a+E \quad \because I \rightarrow a$

$E \Rightarrow a+E*E \quad \because E \rightarrow E*E$

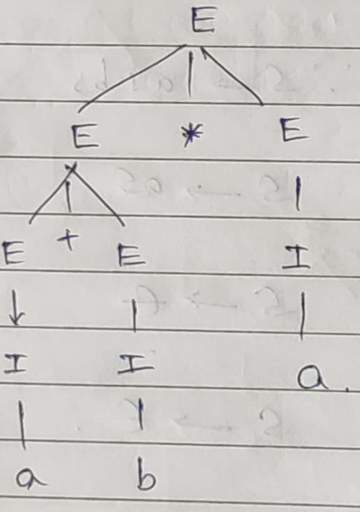
$E \Rightarrow a+I*E \quad \because E \rightarrow I$

$E \Rightarrow a+b*E \quad \because \cancel{E \rightarrow I}\ I \rightarrow b$

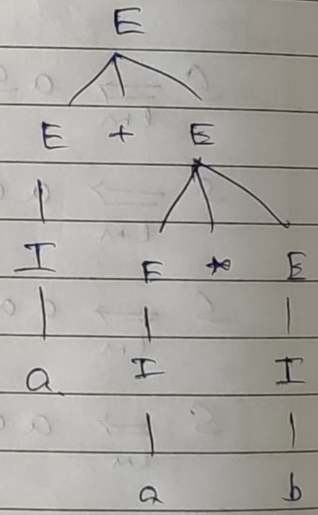
$E \Rightarrow a+b*I \quad \because E \rightarrow I$

$E \Rightarrow a+b*a \quad \because I \rightarrow a$

Parse tree :



or



***

Ambiguous grammar:→ Let $G=(V,T,P,S)$ be a CFG, we say G is ambiguous if there is atleast one string w in T^* for which we can find two different parse trees, each with root labeled S and yield w.

Page No.

Consider the following grammar

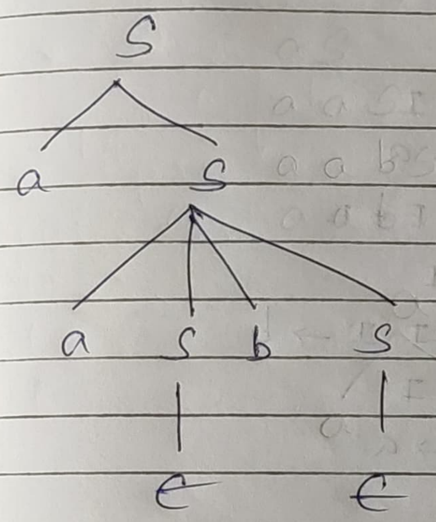
$S \rightarrow aS$

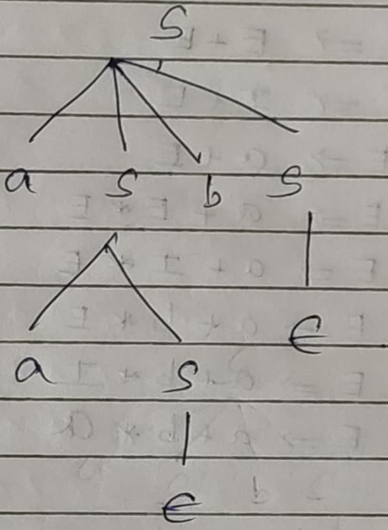
Date $S \rightarrow aSbS$

$S \rightarrow \epsilon$

Check whether this grammar is ambiguous, if so, construct two parse trees, two lmd, two rmd.

Soln → 1 Parse tree.   2 Parse tree





LMD 1)

$S \underset{lm}{\Rightarrow} aSbS$ $\because S \rightarrow aSbS$

$S \underset{lm}{\Rightarrow} aaSbS$ $S \rightarrow aS$

$S \underset{lm}{\Rightarrow} aabS$ $S \rightarrow \epsilon$

$S \underset{lm}{\Rightarrow} aab$ $S \rightarrow \epsilon$.

LMD 2)

$S \underset{lm}{\Rightarrow} aS$ $\because S \rightarrow aS$

$S \underset{lm}{\Rightarrow} aaSbS$ $\because S \rightarrow aSbS$

$S \underset{lm}{\Rightarrow} aabS$ $\because S \rightarrow \epsilon$

$S \underset{lm}{\Rightarrow} aab$ $\because S \rightarrow \epsilon$.

Date __/__/__

RMD 1)

$S \underset{rm}{\Rightarrow} aS$ $\quad S \to aS$

$S \underset{rm}{\Rightarrow} aaSbS$ $\quad S \to aSbS$

$S \underset{rm}{\Rightarrow} aaSb$ $\quad S \to \epsilon$

$S \underset{rm}{\Rightarrow} aab$ $\quad S \to \epsilon$.

RMD 2)

$S \underset{rm}{\Rightarrow} aSbS$ $\quad S \to aSbS$

$S \underset{rm}{\Rightarrow} aSb$ $\quad S \to \epsilon$

$S \underset{rm}{\Rightarrow} aaSb$ $\quad S \to aS$

$S \underset{rm}{\Rightarrow} aab$ $\quad S \to \epsilon$.

~~Removing ambiguity from grammars :-~~

Inherent ambiguity :- A context free language L is said to be inherently ambiguous if all its grammars are ambiguous

Sentential Form :- If $G = (V, T, P, S)$ is a CFG then any string α in $(V \cup T)^*$ such that $S \overset{*}{\Rightarrow} \alpha$ is a sentential form.

*** Write a note on application of CFG.

Write a lex program to count the number of words, characters and lines from a given input file.

```
%{
 #include<stdio.h>
 int w=0 c=0 l=0
%}
%%
[\n]          {l++; c++;}
[^ \t\n]+     {w++; c = c+yyleng;}
%%
```

Properties of CFL (Context Free Languages)

Saathi

Date

① Normal forms for CFG.

Conversion from CFG to CNF (Chomsky Normal Form)

Goal: To convert all productions to the form

$A \to BC$

$A \to a$

$A, B, C \to$ variables

$a \to$ Terminal

This form is called as chomsky normal form (CNF).

To convert the grammar to CNF, we need to make a number of preliminary simplifications :-

① We must eliminate useless symbols, those variables/terminals that donot appear in derivation of any terminal string from the start symbol.

② We must eliminate epsilon production those of the form $A \to \epsilon$ for some variables.

③ We must eliminate unit productions those of the form $A \to B$ for some variables A, B.

Step 1 → Eliminate ϵ production

Step 2 → Eliminate unit production

Step 3 → eliminate useless symbols

④ Eliminating useless symbols.

We say a symbol X is useful for a grammar G if there is some derivation of the form $S \xrightarrow{*} \alpha X \beta \xRightarrow{*} w$ where w is in T^*

* Note:
X may be variable / terminal

The approach to eliminate useless symbol begins by identifying 2 things:
1. We say X is reachable if there is a derivation
$S \Rightarrow \alpha X \beta$
for some α, β

eg: Eliminate useless symbols from following grammar
(1) $S \rightarrow AB \mid a$
$A \rightarrow b$

Sol:→ Symbol : S, A, B, a, b
• Generating symbols: S, A, a, b (∵ a, b generates themselves)
∴ Resulting grammar
$S \rightarrow a$
$A \rightarrow b$
• Reachable symbols: S, a
Resulting grammar is
$S \rightarrow a$
$A \rightarrow b$
• Reachable symbols: S, a
Resulting grammar is
$S \rightarrow a$

(2) $S \rightarrow AB \mid CA$
$A \rightarrow a$
$B \rightarrow BC \mid AB$
$C \rightarrow aB \mid b$.

Sol:- Solution symbols: S, A, B, C, a, b
Generating symbols: S, A, C, a, b

∴ Resulting grammar :

Saathi

Date __/__/__

$S \rightarrow CA$
$A \rightarrow a$
$C \rightarrow b$

Reachable symbols : S, A, C, a, b

∴ Resulting grammer :

$S \rightarrow CA$
$A \rightarrow a$
$C \rightarrow b$

(5) $S \rightarrow ASB \mid \epsilon$
$A \rightarrow aAS \mid a$
$B \rightarrow SbS \mid A \mid bb.$

Symbols : S, A, B, ϵ, a, b.

Generating symbols : A, B, a, b, S
Reachable symbols : A, B, a, b, S
Same grammar

* Elimination of ϵ productions

* Nullable symbol :
A variable A is nullable if
$A \xrightarrow{*} \epsilon$

If A is nullable then whenever A appears in the production body, say, $S \rightarrow BAC$, A might or might not derive ϵ

$S \rightarrow BC \longrightarrow$ derive ϵ
$S \rightarrow BAC \longrightarrow$ not derive ϵ

Basis and Induction

Basis : If $A \to \epsilon$ is a production of G then A is nullable symbol.

Induction : If there is a production $B \to C_1 C_2 \ldots C_k$ where each C_i is nullable then B is also nullable.

* (1) Eliminate ϵ productions from following grammar

$S \to AB$
$A \to aAA \mid \epsilon$
$B \to bBB \mid \epsilon$.

Sol :→ Symbols : S, A, B, a, b, ϵ
Nullable symbols : S, A, B

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

~~ϵ no~~

(2)

$S \to ASB \mid \epsilon$
$A \to aAS \mid a$
$B \to SbS \mid A \mid bb$.

Nullable symbol : S

$S \to ASB \mid AB$
$A \to aAS \mid aA \mid a$
$B \to SbS \mid bS \mid b \mid A \mid bb \mid Sb$.

* Eliminating unit productions

→ A unit production is the production of the form $A \to B$ where A, B are variable

→ Unit pair: A pair (A, B) is said to be a unit pair such that $A \overset{*}{\Rightarrow} B$ using only unit production.

Date — Eg: $A \rightarrow C$

$C \rightarrow D$

$D \rightarrow B$ $\quad A \Rightarrow C \Rightarrow D \Rightarrow B$

all are UP

Basis: (A, A) is a unit pair for any variable A, i.e., $A \overset{*}{\Rightarrow} A$ by 0 steps.

Induction: Suppose we have determined that (A, B) is a unit pair and $B \rightarrow C$ is a production where C is a variable then (A, C) is a unit pair.

* Eliminate unit productions from the following grammar.

①
$S \rightarrow ASB \mid \epsilon$
$A \rightarrow aAS \mid a$
$B \rightarrow SbS \mid A \mid bb$

Variable: S, A, B.

Sol:- (S, S) (A, A) (B, B) are unit pairs from basis step.

$B \rightarrow A$ is a unit production

(B, A) is a unit pair as (B, B) is a unit pair.

Unit pair.	Non unit production
(S, S)	$S \rightarrow ASB \mid \epsilon$
(A, A)	$A \rightarrow aAS \mid a$
(B, B)	$B \rightarrow SbS \mid bb$
(B, A)	$B \rightarrow aAS \mid a$

Page No.

Final grammar

$S \to ASB \mid \epsilon$

Date __/__/____

$A \to aAS \mid a$

$B \to SbS \mid bb \mid aAS \mid a$.

(2). $I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

$F \to I \mid (E)$

$T \to F \mid T * F$

$E \to T \mid E + T$

Sol:→ Variables : I, F, T, E

E → start symbol

(E, E) (I, I), (F, F), (T, T)

We have $E \to T$

∴ (E, T) is also a unit pair

(T, T) is a unit pair

and we have $T \to F$

∴ (T, F) is also a unit pair.

We have (F, F) a unit pair

and also $F \to I$

∴ (F, I) is also a unit pair.

∵ (T, F) & (F, I) are unit pairs (T, I) is also a unit pair.

Unit pair	Non unit productions
(E, E)	$E \to E+T$
(E, F)	$F \to (E)$
(T, T)	$T \to T*F$
(E, I)	$E \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(F, F)	$F \to (E)$
(I, I)	$I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1$
(E, T)	$E \to T*F$
(T, R)	$T \to (E)$

(F, I) $F \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

(T, I) $I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

Date ___/___/____

Resulting grammar is

$E \rightarrow E+T \mid T*F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

$T \rightarrow T*F \mid (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

$F \rightarrow (E) \mid a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$.

Chomsky Normal Form (CNF)

* * *
10M

$A \to BC$

$A \to a$.

1) Eliminate ϵ-productions
2) Eliminate unit productions
3) Eliminate useless symbols.

Converting CFG to CNF.

Task 1: Arrange all bodies of length 2 or more consists of only variables.

Task 2: Break bodies of length 3 or more into a cascade of productions, each with a body consisting of two variables.

Q. ~~Convert the follo~~ Begin with the grammar

$S \to ASB \mid \epsilon$

$A \to aAS \mid a$

$B \to SbS \mid A \mid bb$.

Put the grammar into CNF.

Sol:→ 1. Eliminate ϵ-production.

Nullable symbols : S

$S \to ASB \mid AB$

$A \to aAS \mid aA \mid a$

$B \to SbS \mid Sb \mid bS \mid b \mid A \mid bb$

Eliminate unit productions.

Unit pairs : (S, S) (A, A) (B, B) (B, A)

papergrid

Date: / /

Unit pairs

(S, S)

(A, A)

(B, B)

(B, A)

Non-unit production body

S → ASB | AB

A → aAS | aA | a

B → SbS | Sb | bS | b | bb

B → aAS | aA | a.

∴ Resulting grammar

S → ASB | AB

A → aAS | aA | a.

B → SbS | Sb | bS | b | aAS | aA | a | bb.

3) Eliminate useless symbol.

Symbol : S, A, B, a, b

Generating : S, A, B, a, b.

Reachable : S, A, B, a, b.

∴ Resulting grammar.

S → ASB | AB

A → aAS | aA | a

B → SbS | Sb | bS | b | aAS | aA | a | bb.

$S \to ASB \mid AB$ ⟸ Task 1
$A \to PAS \mid PA \mid a$
$B \to SQS \mid SQ \mid QS \mid b \mid PAS \mid PA \mid a \mid QQ$.

$P \to a$
$Q \to b$

Task 2 :
$S \to AC_1 \mid AB$
$A \to PC_2 \mid PA \mid a$
$B \to SC_3 \mid SQ \mid QS \mid b \mid PC_2 \mid PA \mid a \mid QQ$
$C_1 \to SB$
$C_2 \to AS$
$C_3 \to QS$.
$P \to a$
$Q \to b$

Q) Convert the given CFG to CNF
$S \to 0A0 \mid 1B1 \mid BB$
$A \to C$
$B \to S \mid A$
$C \to S \mid \epsilon$.

Sol:- (i) Eliminate ϵ production
Nullable symbols
$S \to 0A0 \mid 1B1 \mid BB \mid 00 \mid 11 \mid B$
$A \to C$
$B \to S \mid A$
$C \to S$

(ii) Eliminate unit production

unit pairs (S, S) (A, A) (B, B) (C, C)

(A, C) (~~B, S~~) ~~(B, A)~~ ~~(C, S)~~

papergrid

Date: / /

(S, S) + S → B (S, B)

(S, B) + B → A (S, A)

(S, A) + A → C (S, C)

(A, A) + A → C (A, C)

(A, C) + C → S (A, S)

(A, S) + S → B (A, B)

(B, B) + B → S (B, S)

(B, B) + B → A (B, A)

(B, A) + ~~B~~ A → C (B, C)

(C, C) + C → S (C, S)

(C, S) + S → B (C, B)

(C, B) + B → A (C, A).

Unit pair non-~~unit~~ production body

(S, S)

(A, A)

(B, B)

(C, C)

(S, A)

(S, B)

(S, C)

(A, B)

(A, S)

(A, C)

(B, A)

(B, C)

(B, S)

(C, A)
(C, B)
(C, S)

papergrid

Date: / /

Resulting grammar

$S \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB$
$A \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB$
$B \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB$
$C \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB$

3) Eliminate useless symbols.

Generating S, A, B, C
Reachable : S, A, B

$S \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB$
$A \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB$
$B \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB$.

Task 1 : $S \to PAP \mid PP \mid QBQ \mid QQ \mid BB$
$A \to PAP \mid PP \mid QBQ \mid QQ \mid BB$
$B \to PAP \mid PP \mid QBQ \mid QQ \mid BB$
$P \to 0$
$Q \to 1$.

Task 2 : $S \to PC_1 \mid PP \mid QC_2 \mid QQ \mid BB$
$A \to PC_1 \mid PP \mid QC_2 \mid QQ \mid BB$
$B \to PC_1 \mid PP \mid QC_2 \mid QQ \mid BB$
$C_1 \to AP$
$C_2 \to BQ$
$P \to 0$
$Q \to 1$.