Simplification of CFG :-

The process of detection and elimination of useless symbols, unit productions and null productions is called as simplification of CFG.

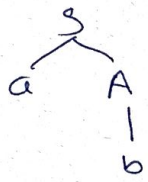
① Useless Symbols :-

The non-terminal which does not involve in the derivation of a string is called as useless symbol.

Ex :- Note :-

① Every unreachable symbol is useless symbol

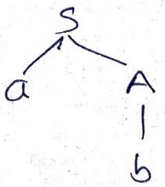
Ex :- ① $S \to aA$
$A \to b$
useless symbol → $B \to a$

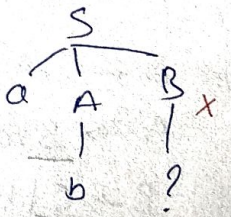


$\Rightarrow$ $S \to aA$
$A \to b$

② The symbol which is reachable from start symbol but does not derive any terminal is called as useless symbol

Ex :- ② $S \to aA \mid aAB$
$A \to b$





$\Rightarrow$ ~~$S \to aA \mid a$~~
$S \to aA$
$A \to b$

③ $S \to aAB \mid aBC$
$A \to aC \mid bB$
$B \to aB \mid bA \mid a$

$\Rightarrow$

$S \to aAB \mid aB$
$A \to bB$
$B \to aB \mid bA \mid a$

} reduced CFG

Note:- The grammer which is free from useless symbols is called as reduced CFG.

② Unit Production :-

The production of the form $A \to B$, where $A, B \in V$ is called as unit production.

Note:- Remove the unit productions & replace the equivalent derivations.

Ex:- ① $S \to aA \mid bB$

$A \to B \mid bA \mid a$ ($A \to B$)

$B \to bB \mid b$

$\Downarrow$

$S \to aA \mid bB$

$A \to bA \mid a \mid bB \mid b$

$B \to bB \mid b$

② $S \to aA$

$A \to B \mid a$

$B \to C \mid b$

$C \to D \mid e$

$D \to \in$

$\Rightarrow$

$S \to aA$

$A \to a \mid b \mid c \mid e$

unreachable symbols: $B \to b \mid c \mid e$, $C \to c \mid e$, $D \to \in$

$\Rightarrow$

$S \to aA$

$A \to a \mid b \mid c \mid e$

(3) Null Production (or) ϵ-production :-

The production of the form $A \to \epsilon$ is called as null production. where $A \in V$.

Note:-

Remove all null production & replace the equivalent derivations.

- However if a CFG has an ϵ-prod. then a^* the CFL does not necessarily contain ϵ.

a^*

Ex:- ① $S \to as \mid \epsilon$

$\Rightarrow S \to as \mid a$

② $S \to AB$
$A \to aA \mid \epsilon$
$B \to bB \mid \epsilon$

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

③ $S \to AaB$
$A \to abA \mid b$
$B \to BaA \mid bB \mid \epsilon$

$\Rightarrow S \to AaB \mid aA$
$A \to abA \mid b$
$B \to BaA \mid aA \mid bB \mid b$

④ $S \to XY$
$X \to a \mid \epsilon$
$Y \to b \mid \epsilon$

$\Rightarrow S' \to S \mid \epsilon$ ⟵ Augumented production for maintaing the same language.
$S \to XY \mid X \mid Y$
$X \to a$
$Y \to b$

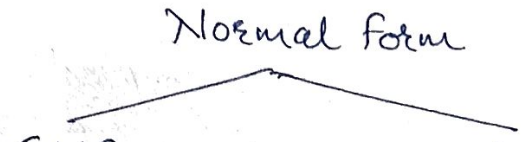
Note:-

The order of process of simplification:

Imp

1. Elimination of ϵ-production
2. Elimination of Unit production
3. Elimination of useless symbols.

Normal forms :-



$A \to BC \mid a$

⇒ CNF – Chomsky Normal Form. (derivation trees of a CFG in Chomsky Normal form are always binary trees)

⇒ GNF – Greibach Normal Form

⇒ The To convert into normal form the grammer must be free from ϵ-production.

* CNF :-

The grammer G is said to be in CNF if every production is in the form –

$A \to BC \mid a$ where $A, B, C \in V$ & $a \in T$

- The grammer should have no useless symbols

Ex:- $S \to aSb \mid \epsilon$

⇒ $S \to aSb \mid ab$

⇒ $S \to ASB \mid AB$

$A \to a$

$B \to b$

⇒ $S \to AC \mid AB$

$C \to SB$

$A \to a$

$B \to b$