Normal Forms

A CFG is in Chomsky Normal Form if the Productions are in the following forms −

$A \rightarrow a$
$A \rightarrow BC$
$S \rightarrow \varepsilon$
where A, B, and C are non-terminals and a is terminal.

Algorithm to Convert into Chomsky Normal Form −
Step 1 − If the start symbol S occurs on some right side, create a new start symbol S' and a new production $S' \rightarrow S$.

Step 2 − Remove Null productions. (Using the Null production removal algorithm discussed earlier)

Step 3 − Remove unit productions. (Using the Unit production removal algorithm discussed earlier)

Step 4 − Replace each production $A \rightarrow B_1 \ldots B_n$ where $n > 2$ with $A \rightarrow B_1 C$ where $C \rightarrow B_2 \ldots B_n$. Repeat this step for all productions having two or more symbols in the right side.

Step 5 − If the right side of any production is in the form $A \rightarrow aB$ where a is a terminal and A, B are non-terminal, then the production is replaced by $A \rightarrow XB$ and $X \rightarrow a$. Repeat this step for every production which is in the form $A \rightarrow aB$.

Problem
Convert the following CFG into CNF

$S \rightarrow ASA \mid aB,\ A \rightarrow B \mid S,\ B \rightarrow b \mid \varepsilon$

Solution
(1) Since S appears in R.H.S, we add a new state S_0 and $S_0 \rightarrow S$ is added to the production set and it becomes −

$S_0 \rightarrow S,\ S \rightarrow ASA \mid aB,\ A \rightarrow B \mid S,\ B \rightarrow b \mid \in$

(2) Now we will remove the null productions −

$B \rightarrow \in$ and $A \rightarrow \in$

After removing $B \rightarrow \varepsilon$, the production set becomes −

$S_0 \rightarrow S,\ S \rightarrow ASA \mid aB \mid a,\ A \rightarrow B \mid S \mid \in,\ B \rightarrow b$

After removing $A \rightarrow \in$, the production set becomes −

$S_0 \rightarrow S,\ S \rightarrow ASA \mid aB \mid a \mid AS \mid SA \mid S,\ A \rightarrow B \mid S,\ B \rightarrow b$

(3) Now we will remove the unit productions.

After removing $S \rightarrow S$, the production set becomes −

$S0 \rightarrow S$, $S \rightarrow ASA \mid aB \mid a \mid AS \mid SA$, $A \rightarrow B \mid S$, $B \rightarrow b$

After removing $S0 \rightarrow S$, the production set becomes −

$S0 \rightarrow ASA \mid aB \mid a \mid AS \mid SA$, $S \rightarrow ASA \mid aB \mid a \mid AS \mid SA$

$A \rightarrow B \mid S$, $B \rightarrow b$

After removing $A \rightarrow B$, the production set becomes −

$S0 \rightarrow ASA \mid aB \mid a \mid AS \mid SA$, $S \rightarrow ASA \mid aB \mid a \mid AS \mid SA$

$A \rightarrow S \mid b$

$B \rightarrow b$

After removing $A \rightarrow S$, the production set becomes −

$S0 \rightarrow ASA \mid aB \mid a \mid AS \mid SA$, $S \rightarrow ASA \mid aB \mid a \mid AS \mid SA$

$A \rightarrow b \mid ASA \mid aB \mid a \mid AS \mid SA$, $B \rightarrow b$

(4) Now we will find out more than two variables in the R.H.S

Here, $S0 \rightarrow ASA$, $S \rightarrow ASA$, $A \rightarrow ASA$ violates two Non-terminals in R.H.S.

Hence we will apply step 4 and step 5 to get the following final production set which is in CNF −

$S0 \rightarrow AX \mid aB \mid a \mid AS \mid SA$

$S \rightarrow AX \mid aB \mid a \mid AS \mid SA$

$A \rightarrow b \mid AX \mid aB \mid a \mid AS \mid SA$

$B \rightarrow b$

$X \rightarrow SA$

(5) We have to change the productions $S0 \rightarrow aB$, $S \rightarrow aB$, $A \rightarrow aB$

And the final production set becomes −

$S0 \rightarrow AX \mid YB \mid a \mid AS \mid SA$

$S \rightarrow AX \mid YB \mid a \mid AS \mid SA$

$A \rightarrow b\ A \rightarrow b \mid AX \mid YB \mid a \mid AS \mid SA$

$B \rightarrow b$

$X \rightarrow SA$

$Y \rightarrow a$