Chomsky Normal form

In Chomsky Normal form (CNF) we have a restriction on the length of RHS; which is; element in RHS should either be two variables or a Terminal.

A CFG is in Chomsky Normal form if the productions are in the following forms:

$$A \to a$$

$$A \to BC$$

where A, B and C are non-terminals and a is a terminal.

Steps to convert a given CFG to 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 $S' \to S$

Step 2: Remove Null Productions. (Using the Null Production Removal)

Step 3: Remove Unit Productions (Using the Unit Production Removal)

Step 4: Replace each production $A \to B_1 \dots B_n$ where $n > 2$, with $A \to B_1 C$ where $C \to B_2 \dots B_n$. Repeat this step for all productions having two or more symbols on the right side.

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

Conversion of CFG to Chomsky Normal Form

Convert the following CFG to CNF :

$P: S \to ASA \mid aB,\ A \to B \mid S,\ B \to b \mid \epsilon$

Step 1: Since S appears in RHS, we add a new state S' and $S' \to S$ is added to the production.

$P: S' \to S,\ S \to ASA \mid aB,\ A \to B \mid S,\ B \to b \mid \epsilon$

Step 2: Remove the Null productions : $B \to \epsilon$ and $A \to \epsilon$:

After Removing $B \to \epsilon$: $P: S' \to S,\ S \to ASA \mid aB \mid a,\ A \to B \mid S \mid \epsilon,$
$B \to b.$

After Removing $A \to \epsilon$: $P: S' \to S,\ S \to ASA \mid aB \mid a \mid AS \mid SA \mid S,\ A \to B \mid S$
$B \to b$

Step 3: Remove the Unit Productions: $S \to S$, $S' \to S$, $A \to B$, and $A \to S$:

After Removing $S \to S$: $P: S' \to S,\ S \to ASA \mid aB \mid a \mid AS \mid SA,\ A \to B \mid S,$
$B \to b.$

After Removing $S' \to S$: $P: S' \to ASA \mid aB \mid a \mid AS \mid SA,$
$S \to ASA \mid aB \mid a \mid AS \mid SA,$
$A \to B \mid S,\ B \to b$

After Removing $A \to B$:
$P: S' \to ASA \mid aB \mid a \mid AS \mid SA$
$S \to ASA \mid aB \mid a \mid AS \mid SA$
$A \to b \mid S,\ B \to b$

After Removing $A \to S$: $P: S' \to ASA \mid aB \mid a \mid AS \mid SA,$
$S \to ASA \mid aB \mid a \mid AS \mid SA,$
$A \to b \mid ASA \mid aB \mid a \mid AS \mid SA,$
$B \to b$

Step4: Now find out the productions that has more than Two variables in RHS

$S' \to ASA$, $S \to ASA$, and $A \to ASA$

After removing these, we get: P: $S' \to AX \mid aB \mid a \mid AS \mid SA$,

$S \to AX \mid aB \mid a \mid AS \mid SA$,

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

$B \to b$

$X \to SA$.

Step5: Now change the productions $S' \to aB$, $S \to aB$ and $A \to aB$

finally we get: P: $S' \to AX \mid YB \mid a \mid AS \mid SA$,

$S \to AX \mid YB \mid a \mid AS \mid SA$,

$A \to b \mid AX \mid YB \mid a \mid AS \mid SA$,

$B \to b$

$X \to SA$

$Y \to a$