Greibach Normal form

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

$A \rightarrow b$

$A \rightarrow bC_1 C_2 \ldots C_n$ (Terminal symbol followed by Non-Terminal)

Where $A, C_1, \ldots, C_n$ are Non-Terminals and b is Terminal.

Steps to convert a given CFG to GNF:

Step1: Check if the given CFG has any Unit Productions or Null Productions and Remove if there are any (using the Unit & Null Productions removal techniques discussed).

Step2: Check whether the CFG is already in Chomsky Normal form (CNF) and convert it to CNF if it is not. (Discussed in previous)

Step3: Change the names of Non-Terminal Symbols into some A_i in ascending order of i.

Example:

$S \rightarrow CA \mid BB$

$B \rightarrow b \mid SB$

$C \rightarrow b$

$A \rightarrow a$

Replace: S with A_1, C with A_2, A with A_3, B with A_4

We get:

$A_1 \rightarrow A_2 A_3 \mid A_4 A_4$

$A_4 \rightarrow b \mid A_1 A_4$

$A_2 \rightarrow b$

$A_3 \rightarrow a$

Step4: After the rules so that Non-Terminals are in ascending order such that, If the Production is of the form $A_i \rightarrow A_j x$ then, $i < j$ and should never be $i \geq j$.

$A_4 \rightarrow b \mid A_1 A_4$ (Rule is correct for form)

$A_4 \rightarrow b \mid A_2 A_3 A_4 \mid A_4 A_4 A_4$ ($A_1 \rightarrow A_2 A_3 \mid A_4 A_4$)

Now check the A_4 rule.

$A_4 \rightarrow b \mid bA_3A_4 \mid A_4A_4A_4$ ($A_2 \rightarrow b$)

↓

Left Recursion

Step5: Remove Left Recursion

$A_4 \rightarrow b \mid bA_3A_4 \mid A_4A_4A_4$

↓

Left Recursion

Introduce a New Variable to remove the Left Recursion

$A_4 \rightarrow b \mid bA_3A_4 \mid A_4A_4A_4$

$Z \rightarrow A_4A_4Z \mid A_4A_4$

$A_4 \rightarrow b \mid bA_3A_4 \mid bZ \mid bA_3A_4Z$

Now, the grammar is :

$A_1 \rightarrow A_2 A_3 \mid A_3 A_4$

$A_4 \rightarrow b \mid bA_3A_4 \mid bZ \mid bA_3A_4Z$

$Z \rightarrow A_4A_4 \mid A_4A_4Z$

$A_2 \rightarrow b$

$A_3 \rightarrow a$

$A_1 \rightarrow bA_3 \mid bA_4 \mid bA_3A_4A_4 \mid bZA_4 \mid bA_3A_4ZA_4$

Now terminal ~~production~~ symbol in first of each and every production

$A_4 \rightarrow b \mid bA_3A_4 \mid bZ \mid bA_3A_4Z$

$Z \rightarrow bA_4 \mid bA_3A_4A_4 \mid bZA_4 \mid bA_3A_4ZA_4 \mid bA_4Z \mid bA_3A_4A_4Z \mid bZA_4Z \mid bA_3A_4ZA_4Z$

Now every ~~term~~ production is having terminal at the beginning.

$A_2 \rightarrow b$

$A_3 \rightarrow a$