

328 | Introduction to Automata Theory, Formal Languages and Computation

$$\begin{aligned} E &\rightarrow FE \\ E &\rightarrow GE \\ E &\rightarrow id \\ C &\rightarrow + \\ D &\rightarrow * \\ F &\rightarrow EC \\ G &\rightarrow ED \end{aligned}$$

In the previous grammar, all the productions are in the form of non-terminal → string of exactly two non-terminals or non-terminal → single terminal. So, the grammar is in CNF.

Example 6.40 Convert the following grammar into CNF.

$$\begin{aligned} S &\rightarrow abAB \\ A &\rightarrow bAB / \epsilon \\ B &\rightarrow bAa / \epsilon \end{aligned}$$

Solution: In the grammar, there are two null productions $A \rightarrow \epsilon$ and $B \rightarrow \epsilon$. First, these productions must be removed, and after that the grammar can be converted into CNF. After removing the null production $A \rightarrow \epsilon$, the modified grammar becomes

$$\begin{aligned} S &\rightarrow abAB / abB \\ A &\rightarrow bAB / bB \\ B &\rightarrow Ba \end{aligned}$$

After removing the null production $B \rightarrow \epsilon$, the modified grammar becomes

$$\begin{aligned} S &\rightarrow abAB / abA / abB / ab \\ A &\rightarrow bAB / bA / bB / b \\ B &\rightarrow Ba / Aa / a \end{aligned}$$

Now, this grammar can be converted into CNF. In the grammar, except $A \rightarrow b$ and $B \rightarrow a$, all the productions are not in CNF. Let us take two non-terminals C_a and C_b which will replace a and b , respectively. So, two new productions $C_a \rightarrow a$ and $C_b \rightarrow b$ will be added to the grammar. After replacing and adding new productions, the modified grammar is

$$\begin{aligned} S &\rightarrow CaC_bAB / C_aC_bA / C_aC_bB / C_aC_b \\ A &\rightarrow C_bAB / C_bA / C_bB / b \\ B &\rightarrow BC_a / ACa / a \\ C_a &\rightarrow a \\ C_b &\rightarrow b \end{aligned}$$

In this modified grammar, $S \rightarrow CaC_bAB$, $S \rightarrow CaC_bA$, $S \rightarrow CaC_bB$, and $A \rightarrow C_bAB$ are not in CNF. Let us take two other non-terminals D and E , which will be replaced in the place of CaC_b and AB , respectively. So, two new productions $D \rightarrow CaC_b$ and $E \rightarrow AB$ will be added to the grammar. By replacing and adding two new productions, the modified grammar is

$$\begin{aligned} S &\rightarrow DE / DA / DB / C_aC_b \\ A &\rightarrow C_bE / C_bA / C_bB / b \\ B &\rightarrow BC_a / ACa / a \end{aligned}$$