Solution to Problem 10

Task: Transform the grammar consisting of rules

\[ S \rightarrow \varepsilon; \quad S \rightarrow ABS; \quad S \rightarrow ASB; \quad S \rightarrow SAB; \quad S \rightarrow SBA; \quad S \rightarrow SBA; \quad S \rightarrow SS \]

where \( A \) and \( B \) are terminal symbols, into Chomsky normal form.

Solution.

Preliminary step. First, we introduce a new starting variable \( S_0 \) and a rule \( S_0 \rightarrow S \), where \( S \) is the starting variable of the original grammar. So, the grammar takes the following form:

\[ S \rightarrow \varepsilon; \quad S \rightarrow ABS; \quad S \rightarrow ASB; \quad S \rightarrow SAB; \quad S \rightarrow SBA; \quad S \rightarrow SBA; \quad S \rightarrow SS \]

\[ S_0 \rightarrow S \]

Step 0. We eliminate non-Chomsky rules with right-hand side of length 0, i.e., with right-hand side an empty string and the left-hand side is not a starting variable. In the above grammar, there is one such rule: \( S \rightarrow \varepsilon \). To eliminate this rule, for each rule that has \( S \) in the right-hand side, we add another rule in which this symbol \( S \) is deleted.

In the above grammar, there are several such rules:

\[ S \rightarrow ABS; \quad S \rightarrow ASB; \quad S \rightarrow SAB; \quad S \rightarrow SBA; \quad S \rightarrow SBA; \quad S \rightarrow SS \]

- For the rule \( S \rightarrow ABS \), if we delete the letter \( S \) from the right-hand side, we get the rule \( S \rightarrow AB \) that we add to our grammar.

- For the rule \( S \rightarrow ASB \), if we delete the letter \( S \) from the right-hand side, we get the same rule \( S \rightarrow AB \) that we have already added to our grammar.

- For the rule \( S \rightarrow ABS \), if we delete the letter \( S \) from the right-hand side, we get the same rule \( S \rightarrow AB \) that we have already added to our grammar.

- For the rule \( S \rightarrow BAS \), if we delete the letter \( S \) from the right-hand side, we get the rule \( S \rightarrow BA \) that we add to our grammar.

1
• For the rule $S \rightarrow BSA$, if we delete the letter $S$ from the right-hand side, we get the same rule $S \rightarrow BA$ that we have already added to our grammar.

• For the rule $S \rightarrow BAS$, if we delete the letter $S$ from the right-hand side, we get the same rule $S \rightarrow BA$ that we have already added to our grammar.

• For the rule $S \rightarrow SS$, if we delete the letter $S$ from the right-hand side, we get the rules $S \rightarrow S$ – which means no change is done, and the rule $S \rightarrow \varepsilon$ – that we are eliminating. So, no new rules are added here.

• For the rule $S_0 \rightarrow S$, if we delete the letter $S$ from the right-hand side, we get the rule $S_0 \rightarrow \varepsilon$ that we add to our grammar

After we delete the rule $S \rightarrow \varepsilon$ and add the new rules $S \rightarrow AB$, $S \rightarrow BA$, and $S_0 \rightarrow \varepsilon$, we get the following grammar:

$S \rightarrow ABS; \ S \rightarrow ASB; \ S \rightarrow SAB; \ S \rightarrow SBA; \ S \rightarrow BSA; \\
S \rightarrow SBA; \ S \rightarrow SS; \ S_0 \rightarrow S; \ S \rightarrow AB; \ S \rightarrow BA; \ S_0 \rightarrow \varepsilon.$

**Step 1.** On this step, we eliminate non-Chomsky rules in which the right-hand side has length 1, i.e., in which the right-hand side is a variable. In the above grammar, there is only one such rule: $S_0 \rightarrow S$. To eliminate this rule, for each rule $S \rightarrow w$ that has the variable $S$ is the left-hand side (for any right-hand side $w$), we add a rule $S_0 \rightarrow w$.

In the current grammar, we have several rules with $S$ in the left-hand side:

$S \rightarrow ABS; \ S \rightarrow ASB; \ S \rightarrow SAB; \ S \rightarrow SBA; \ S \rightarrow BSA; \\
S \rightarrow SBA; \ S \rightarrow SS; \ S_0 \rightarrow S; \ S \rightarrow AB; \ S \rightarrow BA.$

So, once we eliminate the rule $S_0 \rightarrow S$, we have to add rules $S_0 \rightarrow ABS$, etc. As a result, we get the following grammar:

$S \rightarrow ABS; \ S \rightarrow ASB; \ S \rightarrow SAB; \ S \rightarrow SBA; \ S \rightarrow BSA; \\
S \rightarrow SBA; \ S \rightarrow SS; \ S \rightarrow AB; \ S \rightarrow BA; \ S_0 \rightarrow \varepsilon; \\
S_0 \rightarrow ABS; \ S_0 \rightarrow ASB; \ S_0 \rightarrow SAB; \ S_0 \rightarrow SBA; \ S_0 \rightarrow BSA; \\
S_0 \rightarrow SBA; \ S_0 \rightarrow SS; \ S_0 \rightarrow AB; \ S_0 \rightarrow BA.$

**Step 2.** On this step:

• For each terminal symbol $a$, we introduce an auxiliary variable $V_a$ and a rule $V_a \rightarrow a$.

• Then, in each rule in which the right-hand side has 2 or more symbols and at least one of them is a terminal symbol, we replace each terminal symbol with the corresponding variable.
In our grammar, we have two terminal symbols $A$ and $B$. So, we introduce two new variables $V_A$ and $V_B$ and two new rules $V_A \rightarrow A$ and $V_B \rightarrow B$.

In the rule $S \rightarrow ABS$, we replace $A$ with $V_A$ and $B$ with $V_B$, and get the new rule $S \rightarrow V_AV_BS$. We do the same replacement with all other rules in which the right-hand side has 2 or more symbols and at least one of them is a terminal symbol. As a result, we get the following grammar:

$$
S \rightarrow V_AV_BS; \\
S \rightarrow V_AV_BS; \\
S \rightarrow V_AV_BS; \\
S \rightarrow V_AV_BS;
$$

Step 3. At this step, we replace each rule of the type $V \rightarrow ABC$ with two rules: $V_{AB} \rightarrow AB$ for a new variable $V_{AB}$ and $V \rightarrow V_{AB}C$. According to this algorithm, the rule $S \rightarrow V_AV_BS$ is replaced by two rules: $V_{AB} \rightarrow V_AV_B$ and $S \rightarrow V_{AB}S$. After we perform the same replacement for all other rules that have three or more symbols in the right-hand side, we get the following grammar:

$$
V_{AB} \rightarrow V_AV_B; \\
S \rightarrow V_AV_BS; \\
V_{AS} \rightarrow V_A S; \\
S \rightarrow V_AV_BS; \\
V_{SA} \rightarrow SV_A; \\
S \rightarrow V_AV_BS; \\
S \rightarrow V_AV_BS; \\
S \rightarrow V_AV_BS; \\
S \rightarrow V_AV_BS;
$$

This grammar is already in Chomsky normal form, i.e., it only has three types of rules:

- rules of the type $S_0 \rightarrow \varepsilon$, where $S_0$ is the starting variable;
- rules of the type $V \rightarrow a$, where $V$ is a variable and $a$ is a terminal symbol; and
- rules of the type $V \rightarrow AB$, where $V$, $A$, and $B$ are variables.