Solution to Problem 10

Task: Transform the grammar consisting of rules

$$S \to \varepsilon; \quad S \to ABS; \quad S \to ASB; \quad S \to SAB; \quad S \to SBA; \quad S \to BSA;$$

$$S \to SBA; \quad S \to SS$$

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 $S_0 \to S$, where S is the starting variable of the original grammar. So, the grammar takes the following form:

$$S \to \varepsilon; \quad S \to ABS; \quad S \to ASB; \quad S \to SAB; \quad S \to SBA; \quad S \to BSA;$$

$$S \to SBA; \quad S \to SS; \quad \underline{S_0 \to S}.$$

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: $S \to \varepsilon$. 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:

$$S \to ABS; \quad S \to ASB; \quad S \to SAB; \quad S \to SBA; \quad S \to BSA;$$

$$S \to SBA; \quad S \to SS; \quad S_0 \to S.$$

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

- For the rule $S \to BSA$, if we delete the letter S from the right-hand side, we get the same rule $S \to BA$ that we have already added to our grammar.
- For the rule $S \to BAS$, if we delete the letter S from the right-hand side, we get the same rule $S \to BA$ that we have already added to our grammar.
- For the rule $S \to SS$, if we delete the letter S from the right-hand side, we get the rules $S \to S$ – which means no change is done, and the rule $S \to \varepsilon$ – that we are eliminating. So, no new rules are added here.
- For the rule $S_0 \to S$, if we delete the letter S from the right-hand side, we get the rule $S_0 \to \varepsilon$ that we add to our grammar

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

$$S \to ABS; \quad S \to ASB; \quad S \to SAB; \quad S \to SBA; \quad S \to BSA;$$

$$S \to SBA; \quad S \to SS; \quad S_0 \to S; \quad \underline{S \to AB}; \quad \underline{S \to BA}; \quad \underline{S_0 \to \varepsilon}.$$

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: $S_0 \to S$. To eliminate this rule, for each rule $S \to w$ that has the variable S is the left-hand side (for any right-hand side w), we add a rule $S_0 \to w$.

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

$$S \to ABS; \quad S \to ASB; \quad S \to SAB; \quad S \to SBA; \quad S \to BSA;$$

$$S \to SBA; \quad S \to SS; \quad S_0 \to S; \quad S \to AB; \quad S \to BA.$$

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

$$S \to ABS; \quad S \to ASB; \quad S \to SAB; \quad S \to SBA; \quad S \to BSA;$$

$$S \to SBA; \quad S \to SS; \quad S \to AB; \quad S \to BA; \quad S_0 \to \varepsilon;$$

$$\underline{S_0 \to ABS}; \quad \underline{S_0 \to ASB}; \quad \underline{S_0 \to SAB}; \quad \underline{S_0 \to SBA}; \quad \underline{S_0 \to BSA};$$

$$\underline{S_0 \to SBA}; \quad \underline{S_0 \to SS}; \quad \underline{S_0 \to AB}; \quad \underline{S_0 \to BA}.$$

Step 2. On this step:

- For each terminal symbol a, we introduce an auxiliary variable V_a and a rule $V_a \to a$.
- 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 $V_A \to A$ and $V_B \to B$.

In the rule $S \to ABS$, we replace A with V_A and B with V_B, and get the new rule $S \to V_A V_B S$. 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:

$$\underline{S \to V_A V_B S};\quad \underline{S \to V_A S V_B};\quad \underline{S \to S V_A V_B};\quad \underline{S \to S V_B V_A};\quad \underline{S \to V_B S V_A};$$

$$\underline{S \to S V_B V_A};\quad S \to SS;\quad \underline{S \to V_A V_B};\quad \underline{S \to V_B V_A};\quad S_0 \to \varepsilon;$$

$$\underline{S_0 \to V_A V_B S};\quad \underline{S_0 \to V_A S V_B};\quad \underline{S_0 \to S V_A V_B};\quad \underline{S_0 \to S V_B V_A};\quad \underline{S_0 \to V_B S V_A};$$

$$\underline{S_0 \to S V_B V_A};\quad \underline{S_0 \to SS};\quad \underline{S_0 \to V_A V_B};\quad \underline{S_0 \to V_B V_A}.$$

Step 3. At this step, we replace each rule of the type $V \to ABC$ with two rules: $V_{AB} \to AB$ for a new variable V_{AB} and $V \to V_{AB} C$. According to this algorithm, the rule $S \to V_A V_B S$ is replaced by two rules: $V_{AB} \to V_A V_B$ and $S \to V_{AB} S$. 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:

$$\underline{V_{AB} \to V_A V_B};\quad \underline{S \to V_{AB} S};\quad \underline{V_{AS} \to V_A S};\quad \underline{S \to V_{AS} V_B};\quad \underline{V_{SA} \to S V_A};$$

$$\underline{S \to V_{SA} V_B};\quad \underline{V_{SB} \to S V_B};\quad \underline{S \to V_{SB} V_A};\quad \underline{V_{BS} \to V_B S};\quad \underline{S \to V_{BS} V_A};$$

$$\underline{S \to V_{SB} V_A};\quad S \to SS;\quad \underline{S \to V_A V_B};\quad \underline{S \to V_B V_A};\quad S_0 \to \varepsilon;$$

$$\underline{S_0 \to V_{AB} S};\quad \underline{S_0 \to V_{AS} V_B};\quad \underline{S_0 \to V_{SA} V_B};\quad \underline{S_0 \to V_{SB} V_A};\quad \underline{S_0 \to V_{BS} V_A};$$

$$\underline{S_0 \to V_{SB} V_A};\quad \underline{S_0 \to SS};\quad \underline{S_0 \to V_A V_B};\quad \underline{S_0 \to V_B V_A}.$$

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

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