Solution to Problem 10

Task: Transform the grammar consisting of two rules

$$S \to \varepsilon; \quad S \to +S-$$

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:

$$S \to \varepsilon; \quad S \to +S-; \quad \underline{S_0 \to S}.$$

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 two such rules: $S \to +S-$ and $S_0 \to S$.

- For the rule $S \to +S-$, if we delete the letter S from the right-hand side, we get the rule $S \to +-$ that we add to our grammar.
- 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 $S \to +-$ and $S_0 \to \varepsilon$, we get the following grammar:

$$S \to +S-; \quad S_0 \to S; \quad \underline{S \to +-}; \quad \underline{S_0 \to \varepsilon}.$$

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 two rules with S in the left-hand side: $S \to +S-$ and $S \to +-$. So, once we eliminate the rule $S_0 \to S$, we have to add rules $S_0 \to +S-$ and $S_0 \to +-$. As a result, we get the following grammar:

$$S \to +S-; \quad S \to +-; \quad S_0 \to \varepsilon; \quad \underline{S_0 \to +S-}; \quad \underline{S_0 \to +-}.$$

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 $+$ and $-$. So, we introduce two new variables V_+ and V_- and two new rules $V_+ \to +$ and $V_- \to -$.

In the rule $S \to +S-$, we replace $+$ with V_+ and $-$ with V_-, and get the new rule $S \to V_+SV_-$. 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:

$$\underline{S \to V_+SV_-}; \quad \underline{S \to V_+V_-}; \quad S_0 \to \varepsilon; \quad \underline{S_0 \to V_+SV_-}; \quad \underline{S_0 \to V_+V_-};$$

$$\underline{V_+ \to +}; \quad \underline{V_- \to -}.$$

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 $S \to V_+SV_-$ is replaced by two rules: $V_{+S} \to V_+S$ and $S \to V_{+S}V_-$. 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:

$$\underline{V_{+S} \to V_+S}; \quad \underline{S \to V_{+S}V_-}; \quad S \to V_+V_-; \quad S_0 \to \varepsilon; \quad \underline{S_0 \to V_{+S}V_-}; \quad S_0 \to V_+V_-;$$

$$V_+ \to +; \quad V_- \to -.$$

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.