Solution to Problem 10

Task. Transform the grammar from Homework 7 into Chomsky normal form.

Solution. The grammar from Homework 7 has the following rules:

$$I \to +U; \quad I \to -U; \quad I \to U; \quad U \to DU; \quad U \to D; \quad D \to 0;$$

$$D \to 1$$

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. In our grammar, the starting variable is I, so we end up with the following rules:

$$I \to +U; \quad I \to -U; \quad I \to U; \quad U \to DU; \quad U \to D; \quad D \to 0;$$

$$D \to 1; \quad \underline{S_0 \to I}$$

Step 0. On this step, 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 are no such rules, so we do not do anything on this step.

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 are several such rules, we will eliminate them one by one.

The first such rule is $I \to U$. To eliminate this rule, for each rule $U \to w$ that has the variable U is the left-hand side (for any right-hand side w), we add a rule $I \to w$. In the current grammar, we have two such rules: $U \to DU$ and $U \to D$, so we add rules $I \to DU$ and $I \to D$. As a result, we get the following grammar:

$$I \to +U; \quad I \to -U; \quad U \to DU; \quad U \to D; \quad D \to 0; \quad D \to 1;$$

$$S_0 \to I; \quad \underline{I \to DU}; \quad \underline{I \to D}$$

Next rule that need to be eliminated on this stage is $U \to D$. To eliminate this rule, for each rule $D \to w$ that has the variable D is the left-hand side (for any right-hand side w), we add a rule $U \to w$. In the current grammar, we have two such rules: $D \to 0$ and $D \to 1$, so we add rules $U \to 0$ and $U \to 1$. As a result, we get the following grammar:

$$I \to +U; \quad I \to -U; \quad U \to DU; \quad D \to 0; \quad D \to 1; \quad S_0 \to I;$$

$$I \to DU; \quad I \to D; \quad \underline{U \to 0}; \quad \underline{U \to 1}$$

Next rule that need to be eliminated on this stage is $S_0 \to I$. To eliminate this rule, for each rule $I \to w$ that has the variable I 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 four such rules: $I \to +U$, $I \to -U$, $I \to DU$, and $I \to D$, so we add the rules $S_0 \to +U$, $S_0 \to -U$, $S_0 \to DU$, and $S_0 \to D$. As a result, we get the following grammar:

$$I \to +U; \quad I \to -U; \quad U \to DU; \quad D \to 0; \quad D \to 1; \quad I \to DU;$$

$$I \to D; \quad U \to 0; \quad U \to 1; \quad \underline{S_0 \to +U}; \quad \underline{S_0 \to -U}; \quad \underline{S_0 \to DU};$$
$$\underline{S_0 \to D}$$

Next rule that need to be eliminated on this stage is $I \to D$. To eliminate this rule, for each rule $D \to w$ that has the variable D is the left-hand side (for any right-hand side w), we add a rule $I \to w$. In the current grammar, we have two such rules: $D \to 0$ and $D \to 1$, so we add the rules $I \to 0$ and $I \to 1$. As a result, we get the following grammar:

$$I \to +U; \quad I \to -U; \quad U \to DU; \quad D \to 0; \quad D \to 1; \quad I \to DU;$$

$$U \to 0; \quad U \to 1; \quad S_0 \to +U; \quad S_0 \to -U; \quad S_0 \to DU; \quad S_0 \to D;$$
$$\underline{I \to 0}; \quad \underline{I \to 1}$$

The last rule that need to be eliminated on this stage is $S_0 \to D$. To eliminate this rule, for each rule $D \to w$ that has the variable D 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 such rules: $D \to 0$ and $D \to 1$, so we add the rules $S_0 \to 0$ and $S_0 \to 1$. As a result, we get the following grammar:

$$I \to +U; \quad I \to -U; \quad U \to DU; \quad D \to 0; \quad D \to 1; \quad I \to DU;$$

$$U \to 0; \quad U \to 1; \quad S_0 \to +U; \quad S_0 \to -U; \quad S_0 \to DU; \quad I \to 0;$$
$$I \to 1; \quad \underline{S_0 \to 0}; \quad \underline{S_0 \to 1}$$

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 four terminal symbols $+$, $-$, 0 and 1. So, we introduce four new variables V_+, V_0, V_0, and V_1 and four new rules $V_+ \to +$, $V_- \to -$, $V_0 \to 0$, and $V_1 \to 1$. In the rule $I \to +U$, we replace $+$ with V_+ and get the new rule $I \to V_+U$. 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:

$$\underline{I \to V_+U}; \quad \underline{I \to V_-U}; \quad U \to DU; \quad D \to 0; \quad D \to 1; \quad I \to DU;$$

$$U \to 0; \quad U \to 1; \quad \underline{S_0 \to V_+U}; \quad \underline{S_0 \to V_-U}; \quad S_0 \to DU; \quad I \to 0;$$

$$I \to 1; \quad S_0 \to 0; \quad S_0 \to 1; \quad \underline{V_+ \to +}; \quad \underline{V_- \to -}; \quad \underline{V_0 \to 0};$$

$$\underline{V_1 \to 1}$$

Step 3. At this step, we deal with the rules in which the right-hand side has length 3 or larger. In our grammar, there are not such rules, so the grammar that we obtained after Step 2 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.