Tutorial 4

Exercise 1: Consider the following context-free grammar:

$$\begin{aligned} S &\longrightarrow aBb \mid AB \\ A &\longrightarrow bAb \mid a \\ B &\longrightarrow \varepsilon \mid aABb \end{aligned}$$

a) Give (some) derivation of word $babaab$ in this grammar.

b) Draw the corrensponding derivation tree.

c) Write the left and right derivations corresponding to the derivation tree drawn in the previous point.

Solution:

Left derivation: $S \Rightarrow AB \Rightarrow bAbB \Rightarrow babB \Rightarrow babaABb \Rightarrow babaaBb \Rightarrow babaab$

Right derivation: $S \Rightarrow AB \Rightarrow AaABb \Rightarrow AaAb \Rightarrow Aaab \Rightarrow bAbaab \Rightarrow babaab$

Exercise 2: Construct context-free grammars for all following languages:

- $L_1 = \{w \in \{a, b, c\}^* \mid w \text{ contains subword } babb\}$

  Solution:

$$\begin{aligned} S &\longrightarrow AbabbA \\ A &\longrightarrow \varepsilon \mid aA \mid bA \mid cA \end{aligned}$$

- $L_2 = \{0^n 1^m \mid 1 \leq n < m\}$

  Solution:

$$\begin{aligned} S &\longrightarrow AB \\ A &\longrightarrow 0A1 \mid 01 \\ B &\longrightarrow 1B \mid 1 \end{aligned}$$

- $L_3 = \{a^n b^m a^{n+2} \mid m, n \in \mathbb{N}\}$

  Solution:

$$\begin{aligned} S &\longrightarrow Aaa \\ A &\longrightarrow aAa \mid B \\ B &\longrightarrow bB \mid \varepsilon \end{aligned}$$

- $L_4 = \{w \in \{0, 1\}^* \mid w = w^R\}$

  Solution:

$$S \longrightarrow 0S0 \mid 1S1 \mid 0 \mid 1 \mid \varepsilon$$

- $L_5 = \{w \in \{0,1\}^* \mid |w|_0 > 1,\ |w|_1 \leq 2\}$

  Solution:

$$\begin{array}{rcl} S & \longrightarrow & 00ABABA \mid 0AB0ABA \mid 0ABAB0A \mid AB00ABA \mid AB0AB0A \mid ABAB00A \\ A & \longrightarrow & \varepsilon \mid 0A \\ B & \longrightarrow & \varepsilon \mid 1 \end{array}$$

- $L_6 = \{0^n w w^R 1^n \mid w \in \{0,1\}^*,\ n \in \mathbb{N}\}$

  Solution:

$$\begin{array}{rcl} S & \longrightarrow & 0S1 \mid A \\ A & \longrightarrow & 0A0 \mid 1A1 \mid \varepsilon \end{array}$$

- $L_7 = \{w \in \{a,b\}^* \mid$ in w, every a is directly followed by b, or $w = b^n a^m$, where $0 \leq m \leq n\}$

  Solution:

$$\begin{array}{rcl} S & \longrightarrow & A \mid BC \\ A & \longrightarrow & \varepsilon \mid abA \mid bA \\ B & \longrightarrow & \varepsilon \mid bB \\ C & \longrightarrow & bCa \mid \varepsilon \end{array}$$

- $L_8 = \{uv^Rv \mid u,v \in \{0,1\}^*,\ |u|_0 \bmod 4 = 2,\ u$ ends with suffix 101 and v contains subword $10\}$

  Solution:

$$\begin{array}{rcl} S & \longrightarrow & A101C \\ A & \longrightarrow & B0B0B0B0BA \mid B0B \\ B & \longrightarrow & \varepsilon \mid B1 \\ C & \longrightarrow & 0C0 \mid 1C1 \mid 01D10 \\ D & \longrightarrow & 0D0 \mid 1D1 \mid \varepsilon \end{array}$$

- $L_9 = \{w \in \{a,b\}^* \mid w = w^R,\ |w| \bmod 4 = 0\}$

  Solution:

$$S \longrightarrow aaSaa \mid abSba \mid baSab \mid bbSbb \mid \varepsilon$$

- $L_{10} = \{w \in \{a,b\}^* \mid w = w^R,\ |w| \bmod 3 = 0\}$

  Solution:

$$\begin{array}{rcl} S & \longrightarrow & aTa|bTb|\varepsilon \\ T & \longrightarrow & aUa|bUb|a|b \\ U & \longrightarrow & aSa|bSb \end{array}$$

- $L_{11} = \{w \in \{a,b,c\}^* \mid$ every sequence of a's is directly followed by a sequence of b's, which is twice as long$\}$

  Solution:

$$\begin{array}{rcl} S & \longrightarrow & bS \mid cS \mid AB \mid \varepsilon \\ A & \longrightarrow & aAbb \mid abb \\ B & \longrightarrow & cS \mid AB \mid \varepsilon \end{array}$$

- $L_{12} = \{w \in \{0,1\}^* \mid |w|_0 = |w|_1\}$

  Solution:

$$S \longrightarrow \varepsilon \mid 0S1 \mid 1S0 \mid SS$$

Exercise 3: Decide for the following pairs of grammars if both grammars generate the same language. Justify your answers.

a) $S \longrightarrow aaSbb \mid ab \mid aabb$ $\qquad$ $S \longrightarrow aSb \mid ab$

Solution: Yes

The second grammar obviously generates language $\{a^i b^i : i \geq 1\}$. We must verify that the first grammar generates the same language. This grammar also generates a language constisting of words where a sequence of as is followed with a sequence of bs. The rule $S \longrightarrow aaSbb$ allows to generate all sentential forms of the form $a^j S b^j$, where $j \geq 0$ is even. So if i in a word generated by the second grammar is odd, we finish the corresponding derivation by using rule $S \longrightarrow ab$. When we want to generate a word $a^i b^i$ for even $i \geq 2$, we apply the rule $S \longrightarrow aabb$ in the end, by which we obtain the word $a^{j+2} b^{j+2}$ with $i = j+2$. So we have shown that both grammar generate the same set of words over $\{a, b\}$.

b) $S \longrightarrow aaSbb \mid ab \mid \varepsilon$ $\qquad$ $S \longrightarrow aSb \mid ab$

Solution: No, since the second one does not generate ε.

c) $S \longrightarrow aaSb \mid ab \mid \varepsilon$ $\qquad$ $S \longrightarrow aSb \mid aab \mid \varepsilon$

Solution: No, since the first one does not generate $aaaabb$.

Exercise 4: Construct a context-free grammar for the language L over the alphabet $\Sigma = \{(,), [,]\}$ consisting of all "correctly parenthesized" expressions. As correctly parenthesized expressions we consider those sequences of symbols where each left parenthesis has a corresponding right parenthesis of the same type, and where parenthesis do not "cross" (i.e., coresponding pairs of parenthesis are composed correctly).

Solution:

$$S \longrightarrow \varepsilon \mid SS \mid (S) \mid [S]$$

Exercise 5: Is the following grammar unambiguous?

$$E \longrightarrow E + E \mid F$$
$$F \longrightarrow (E) \mid F \times F \mid a$$

Exercise 6: Propose a syntax for writing simple arithmetic expressions as words over the alphabet

$$\Sigma = \{\texttt{A}, \texttt{B}, \ldots, \texttt{Z}, \texttt{a}, \texttt{b}, \ldots, \texttt{z}, \texttt{0}, \texttt{1}, \ldots, \texttt{9}, \texttt{.}, \texttt{+}, \texttt{-}, \texttt{*}, \texttt{/}, \texttt{(}, \texttt{)}\}.$$

and describe the proposed syntax by a context-free grammar.

Exercise 7: Construct a context-free grammar generating the set of all well-formed formulas of the propositional logic. Consider the set $At = \{x_0, x_1, x_2, \ldots\}$ as the set of atomic propositions, where individual variables can be written as $\texttt{x0}, \texttt{x1}, \texttt{x2}, \ldots$

a) Find out if the grammar you have constructed is unambiguous.

b) If the grammar is ambiguous then modify it to be unambiguous.

c) Modify your grammar in such a way, which ensures that a structure of a derivation tree for an arbitrary derivation in the grammar reflects the "real" priority of logical connectives, i.e.. $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$ (from the highest to the lowest).

Solution:

$$\begin{array}{rcl}
S & \longrightarrow & A \mid A \leftrightarrow S \\
A & \longrightarrow & B \mid B \rightarrow A \\
B & \longrightarrow & C \mid C \vee B \\
C & \longrightarrow & D \mid D \wedge C \\
D & \longrightarrow & \neg D \mid (S) \mid \texttt{x}E \mid \bot \mid \top \\
E & \longrightarrow & F \mid EF \\
F & \longrightarrow & 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
\end{array}$$

This grammar in unambiguous and a structure of a derivation tree corresponds to the priority of logical connectives.