© Daniel Leivant 2023

Assignment 6: Context-free languages

Solutions.

1. (20%) For each grammar below describe in words the language it generates.

(a) $S \to \mathtt{a}S\mathtt{b} \mid \mathtt{b}S\mathtt{a} \mid \varepsilon$

Solution. The set of strings $w\tilde{w}$ where $\tilde{w}$ is the flip of w, i.e. with $\mathtt{a}$ and $\mathtt{b}$ interchanged.

(b) $S \to SS \mid \mathtt{a}$

Solution. $\mathcal{L}(\mathtt{a}^+)$, that is the non-empty strings over the letter $\mathtt{a}$.

(c) $S \to \mathtt{a}S \mid \mathtt{c}T, \quad T \to \mathtt{a}T \mid \mathtt{c}R, \quad R \to \mathtt{a}R \mid \varepsilon.$

Solution. The strings of the form $\mathtt{a}^i\mathtt{c}\mathtt{a}^j\mathtt{c}\mathtt{a}^k$ where $i, j, k \geqslant 0$.

(d) $S \to \mathtt{a}A \mid \mathtt{b}B, \quad A \to \mathtt{a}A \mid \mathtt{b}A \mid \mathtt{a}, \quad B \to \mathtt{a}B \mid \mathtt{b}B \mid \mathtt{b}.$

Solution. The strings that start and end with the same letter, i.e. the strings of the form $\mathtt{a} \cdot w \cdot \mathtt{a}$ and the strings of the form $\mathtt{b} \cdot w \cdot \mathtt{b}$.

2. (30%) For each of the following languages give a CFG that generates it.

(a) $L = \{\mathtt{a}^p\mathtt{b}^q\mathtt{c}^r \mid p + q = r\}$

Solution. $S \to \mathtt{a}S\mathtt{c} \mid T, \quad T \to \mathtt{b}T\mathtt{c} \mid \varepsilon.$

(b) $L = \{\mathtt{a}^n\mathtt{b}^k\mathtt{c}^n \mid k, n \geqslant 0\}$

Solution. $S \to \mathtt{a}S\mathtt{c} \mid T, \quad T \to \mathtt{b}T \mid \varepsilon.$

(c) $L = \{\mathtt{a}^p\mathtt{b}^q\mathtt{c}^r \mid p + q < r\}$

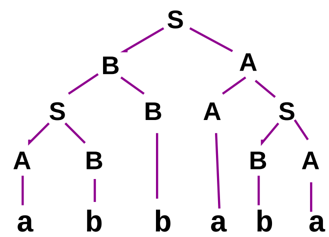
Solution. $S \to S\mathtt{c} \mid T\mathtt{c}, \quad T \to \mathtt{a}T\mathtt{c} \mid \mathtt{a}T\mathtt{c} \mid R, \quad R \to \mathtt{b}R\mathtt{c} \mid \varepsilon.$

3. (30%) Consider the CFG G over the alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}\}$, with initial nonterminal S and with productions

$$
\begin{array}{lcl}
S & \to & AB \mid BA \\
A & \to & SA \mid AS \mid \mathtt{a} \\
B & \to & SB \mid BS \mid \mathtt{b}
\end{array}
$$

(a) Give a parse tree of G for the string **abbaba**.

Solution. S generates the non-empty strings with an equal number of `a`'s and `b`'s, A the strings with an excess of one `a`, and B the strings with an excess of one `b`. So a parse-tree for **abbaba** is



(b) Give the leftmost-derivation for your parse-tree, as well as another derivation for it.

Solution. Leftmost derivation:

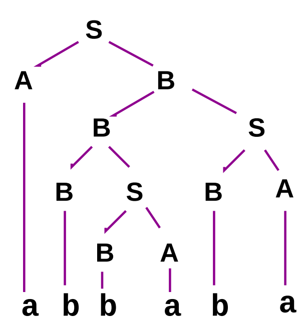
$$S \Rightarrow BA \Rightarrow SBA \Rightarrow ABBA \Rightarrow \mathtt{a}BBA \Rightarrow \mathtt{ab}BA \Rightarrow \mathtt{abb}A \Rightarrow \mathtt{abb}AS \\ \Rightarrow \mathtt{abba}S \Rightarrow \mathtt{abba}BA \Rightarrow \mathtt{abbab}A \Rightarrow \mathtt{abbaba}$$

Another derivation (rightmost chosen):

$$S \Rightarrow BA \Rightarrow BAS \Rightarrow BABA \Rightarrow BAB\mathtt{a} \Rightarrow BA\mathtt{ba} \Rightarrow B\mathtt{aba} \\ \Rightarrow SB\mathtt{aba} \Rightarrow S\mathtt{baba} \Rightarrow AB\mathtt{baba} \Rightarrow A\mathtt{bbaba} \Rightarrow \mathtt{abbaba}$$

(c) Show that G is an ambiguous grammar, by giving two different parse-trees for the same string.

Solution. An alternative parse-tree for **abbaba** is:



A. Let $L = \{\mathtt{a}^i\mathtt{b}^{i+j}\mathtt{c}^j \mid i,j \geqslant 0\}$

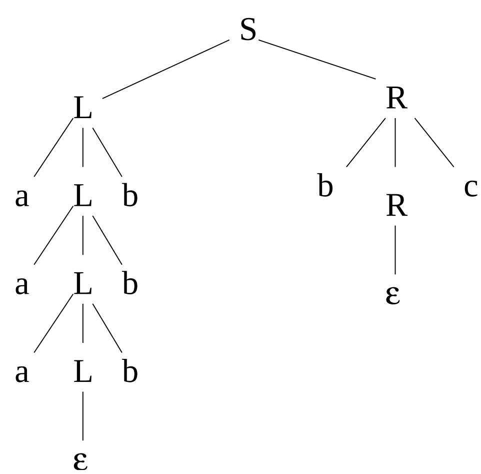
i. Give a CFG G that generates L.

Solution.

$$\begin{array}{rcl} S & \rightarrow & LR \\ L & \rightarrow & \mathtt{a}L\mathtt{b} \\ L & \rightarrow & \varepsilon \\ R & \rightarrow & \mathtt{b}R\mathtt{c} \\ R & \rightarrow & \varepsilon \end{array}$$

ii. Give a parse tree of G for the string **aaabbbbc**.

Solution.



iii. Give the leftmost derivation for T.

Solution.

$$S \Rightarrow LR \Rightarrow \mathtt{a}L\mathtt{b}R \Rightarrow \mathtt{aa}L\mathtt{bb}R$$
$$\Rightarrow \mathtt{aaa}L\mathtt{bbb}R \Rightarrow \mathtt{aaabbb}R \Rightarrow \mathtt{aaabbbb}R\mathtt{c} \Rightarrow \mathtt{aaabbbbc}$$

B. The following problem refers to the symbols $\#$ and $\$$, to draw attention to their being different.

i. Construct a CFG generating $\{\mathtt{a}^i \# \mathtt{b}^k \$ \mathtt{a}^i\}$.

Solution. $S \rightarrow \mathtt{a}S\mathtt{a} \mid \# M$, $M \rightarrow \mathtt{b}M \mid \$$.

ii. Show that the language $L = \{\, u \# w \$ u' \mid |u'| = |u| \,\}$ is CF.

Solution. This is like (a), except that the a's can be any letter.

iii. Show that no string in L can be a palindrome.

Solution. Each $w \in L$ has $\#$ and $\$$ are at symmetric positions. Since these are different symbols w can't be a palindrome.

iv. Show that the non-palindromes (over an alphabet Σ) constitute a CFL.

Solution. For each pair σ, τ of different letters, the set of strings with σ and τ in symmetric positions is a CFL: just take σ for $\#$ above and τ for $\$$. The set of non-palindromes is the union over all pairs of distinct letters of languages as above, and the union of CFL's is a CFL.

4. (20%)

(a) Construct a CFG generating $\{\mathtt{a}^i \# \mathtt{b}^{i+j} \$ \mathtt{a}^j\}$.

Solution. $S \to LR$, $L \to \mathtt{a}L\mathtt{b} \mid \#$, $R \to \mathtt{b}R\mathtt{a} \mid \$$.

(b) Show that the language $\{ u \# v \cdot u' \$ v' \mid |u'| = |u|,\ |v'| = |v| \}$ is CF. [Hint: The proof idea is (a).]

Solution. $S \to LR$, $LXLX \mid \#$, $R \to XRX \mid \$$, $X \to \mathtt{a} \mid ttb$

(c) Recall the Mahi-Mahi language $M = \{ w \cdot w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \}$.
We showed that it is not CF.
Show that the ***complement*** of M ***is*** a CFL.

Solution. For each pair σ, τ of different letters, the set of strings with σ and τ in corresponding positions is a CFL: just take σ for $\#$ above and τ for $\$$. The set of strings not of the form $w \cdot w$ is the union over all pairs of distinct letters of languages as above, and the set of strings of odd length. All these languages are CF, so their union is a CF language.