Assignment 6: Context-free languages

Points for each problem coming up, as well as a layout revision.

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

(a) $S \to aSb \mid bSa \mid \varepsilon$

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

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

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

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

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

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

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

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**.

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

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

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

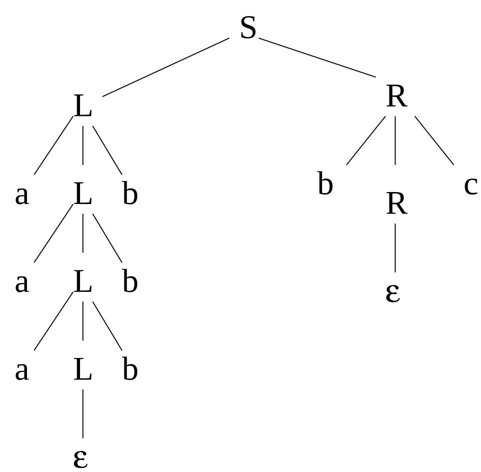
i. Give a CFG G that generates L.

Solution.

$$
\begin{aligned}
S &\rightarrow LR \\
L &\rightarrow aLb \\
L &\rightarrow \varepsilon \\
R &\rightarrow bRc \\
R &\rightarrow \varepsilon
\end{aligned}
$$

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

Solution.



iii. Give the leftmost derivation for T.

Solution.

$$
\begin{aligned}
S &\Rightarrow LR \Rightarrow \texttt{a}L\texttt{b}R \Rightarrow \texttt{aa}L\texttt{bb}R \\
&\Rightarrow \texttt{aaa}L\texttt{bbb}R \Rightarrow \texttt{aaabbb}R \Rightarrow \texttt{aaabbbb}R\texttt{c} \Rightarrow \texttt{aaabbbbc}
\end{aligned}
$$

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

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 $\{\mathtt{a}^i \# \mathtt{b}^{i+j} \$ \mathtt{a}^j\}$.

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

(c) Recall the Mahi-Mahi language $M = \{\, w \cdot w \mid w \in \{\mathtt{a}, \mathtt{b}\}^* \,\}$. Show that the ***complement*** of M is a CFL.