EXERCISES ABOUT CONTEXT-FREE GRAMMARS

1 The following grammar defines the language given by the regular expression 0*1(0+1)*

```
S → A1B
A → 0A | ε
B → 0B | 1B | ε
```

Give the leftmost and the rightmost derivations and the abstract syntax trees for the following strings:

a) 1001

b) 00011.

2 Given the grammar G= (V, T, P, S), in which T = {0,1,2}, V= {S, B, C} and the productions are:

```
S → 0 S 0 0 | B
B → 1 1 B 2 2 | C
C → 1 2
```

a) Draw an abstract syntax tree for 011122200.

b) Explain in text the language accepted by the grammar.

c) Prove by induction that the statement before is correct.

3 Consider the CFG G defined by the following productions

S → aS | Sb | a | b.

a) Prove by induction in the length of the string that none string of L(G) has **ba** as substring.

b) Describe L(G) informally. Justify your answer using a).

4 Show that every regular language is a context free language. Suggestion: build a CFG by induction over the number of operators of the regular expression.

5 Given T= {0,1,(,),+,*,∅,e} the set of symbols used in regular expressions over the alphabet {0,1}, and with the empty string represented by e.

a) Define a CFG with the set of terminal symbols T that is able to generate exactly those regular expressions, using only one variable E. Draw the syntax tree for 010*+1(0+1).

b) Study the ambiguity of the grammar, trying to find a second syntax tree for the same expression. Can you obtain two different leftmost derivations for the expression? Exemplify.

c) Give a non-ambiguous version of the CFG and the syntax tree for the expression in a). Verify if the precedence and association rules are respected. Can you obtain two different leftmost derivations for the expression? Exemplify.

6 Consider the following fragment for a grammar of HTML code:

Char → a | A | …

Text → ε | Char Text

Doc → ε | Element Doc

List → ε | ItemList List

Element → Text |
<P> Doc |
<B> Doc </B> |
<OL> List </OL>

ItemList → <LI> Doc

Add the following rules to the grammar:

a) An item of a list must be closed by the tag </LI>.

b) The elements must include the non-ordered lists, with tags <UL> e </UL>.

c) Include as elements the tables, identified as <TABLE>, </TABLE>, with lines <TR>, </TR>. The first line as header components <TH>, </TH>. The other lines have data components <TD>, </TD>.

7 Suppose that G is a CFG without productions with ε. If w belongs to L(G), |w| = n and w has one derivation in m steps, show that w has a syntax tree with n+m nodes.

8 Consider the grammar G, given by: $S \rightarrow aSb \mid bSa \mid SS \mid \varepsilon$.

a) Prove by induction that it is a grammar for L_{equal}, the language that has the same number of a's than b's, independently of the order.

b) Give a rightmost derivation for the string baaabb.

c) The G grammar is ambiguous? Justify your answer using the syntax trees for a specific example.

9 Obtain CFGs for the following languages:

a) $L = \{0n1m \mid n > 2m, m \geq 0\}$.

b) $M = \{a^n b^m \mid n < 2m, n \geq 0\}$

c) $N = \{0^n 1^m 0^p 1^q \mid n < q, m > p\}$

10 How would you describe the language generated by the following CFG in English?
$S \rightarrow 0S1 \mid 1S0 \mid \varepsilon$

11 Is the following grammar ambiguous? Prove it.
$S \rightarrow 0S1S \mid 1S0S \mid \varepsilon$