Homework 5
Cpt S 317, Spring 2018
Due Date: March 28, 2018

Total points: 48

1. (5 points)
   Let $L$ be the set of all strings with an equal number of 0’s and 1’s. A CFG for $L$ is as follows:
   
   $$G : S \rightarrow 0S1S \mid 1S0S \mid \epsilon$$
   
   Give parse trees for deriving the strings 000111 and 10011010 from $G$.

2. (6 points) (from Exercise 5.1.7 in book):
   Consider the CFG $G$ defined by the productions:
   
   $$S \rightarrow aS \mid Sb \mid a \mid b$$
   
   Prove by induction that no string in $L(G)$ has $ba$ as a substring. Hint: To show this, you will have to do induction on the length of the string, for strings that be generated by the grammar.

3. (12 points)
   Give simple English language descriptions for the following grammars:
   a) $G_1 : S \rightarrow 0S \mid 0$
   b) $G_2 : S \rightarrow 0S0 \mid 00 \mid 0$
   c) $G_3 : S \rightarrow S0S \mid 0$
   d) $G_4 : S \rightarrow 0S0 \mid 0T1 \mid 1T0 \mid 1T1$
      $T \rightarrow 0T \mid 1T \mid \epsilon$
4. (5 points) Is the following grammar ambiguous? Justify your answer.

\[ G : S \rightarrow S0S | 0 | 1 \]

5. (5 points) The following grammar generates postfix expressions with operands \( x \) and \( y \) and binary operators \( +, -, \ast \) and \( / \):

\[ G : E \rightarrow EE + | EE - | EE \ast | EE/ | x | y \]

For the string \( xyxyx - \ast + / \), give a parse tree, a leftmost derivation and a rightmost derivation.

6. (5 points) Give a right-linear CFG for the language of all strings that are of the form \( a^i \) where \( i \mod 3 = 2 \). Note that the alphabet here is \( \{a\} \).

7. (10 points)

In the natural world, there are some instances where biological molecules follow rules like in a CFG. Consider the following example:

Cells have RNA molecules which can be represented as strings over alphabet \( \{a, c, g, u\} \). An RNA molecule (≡“RNA string”) \( w \) is said to have a stem-loop structure if \( w \) can be broken down into three parts: \( w = xyz \), s.t., string \( x \) is a “reverse complement” of string \( z \), and \( |x| \geq 1, |y| \geq 1, |z| \geq 1 \). “Reverse complement” of any string \( x \) is equal to another string \( z \) if \( z \) is obtained by first reversing \( x \) and replacing every \( a \) with \( u \), \( u \) with \( a \), \( c \) with \( g \), and \( g \) with \( c \).

For example, an RNA molecule \( w \) that reads \( aacauggucaacggscauguu \) has a stem-loop structure, because there exists a combination \( x, y, z \) where \( x = aacaugg, y = ucaacggs, \) and \( z = ccauguu \). Note that \( x \)’s reverse complement is \( z \). The reason why this is called “Stem-Loop” structure should be clear from the figure below, which shows the string
bonding with itself in 2D (a prefix to a suffix) as defined by the reverse complement property.

RNA Stem-Loop for the DNA $aacauggucaacgggccauguu$:

```
  a -> u
  a -> u
  c -> g
  a -> u
  u -> a
  g -> c
  g -> c

Legend:

- a always binds with u
- c always binds with g
```

a) Write a grammar for recognizing the language of RNA molecules (i.e., the set of strings over alphabet \{a,c,g,u\} that have a stem-loop structure).

b) Is your grammar ambiguous? If yes, give an example string with more than derivation. If not, just briefly justify what makes it unambiguous.