# Homework 5

## Cpt S 317, Spring 2009

## Due Date: March 25, 2009

Total points: 43

1. (5 points)

   Let $L$ be the set of all strings with an equal number of a's and b's. A CFG for L is as follows:

   $$G: \quad S \rightarrow aSbS \mid bSaS \mid \epsilon$$

   Give parse trees for deriving the strings *baababba* and *bbbaaa* using $G$.

2. (5 points)

   Exercise 5.3.1.

3. (8 points)

   Give the English language description of each of the following grammars:

   $$G_1: \quad S \rightarrow SaS \mid a$$

   $$G_2: \quad S \rightarrow aSa \mid aa \mid a$$

   $$G_3: \quad S \rightarrow aS \mid a$$

   $$G_4: \quad S \rightarrow aSa \mid bSb \qquad A \rightarrow abA \mid ab$$

4. (5 points) The following grammar is ambiguous:

   $$G: \quad S \rightarrow SS \mid a \mid b$$

Give two strings such that one string has more than one derivation, and another has exactly one derivation. For each of these two strings, show the parse trees for their respective derivation(s).

5. (5 points) The following grammar generates *prefix expressions* with operands $x$ and $y$ and binary operators $+$, $-$ and $*$:

$$G: \quad E \rightarrow +EE \mid -EE \mid *EE \mid x \mid y$$

For the string $+*-xyxy$, give a parse tree, a leftmost derivation and a rightmost derivation.

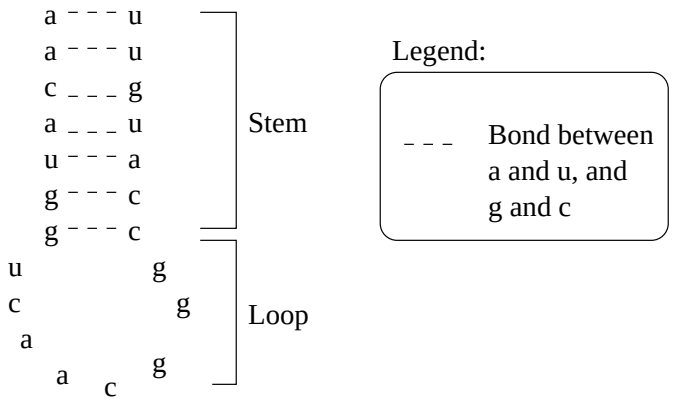
6. (5 points) Give a *right-linear CFG* for the language of all strings that are of the form $a^i$ where $(i \mod 3 = 1)$.
Hint: Note that this language is regular. So you should be able to construct a DFA and then build a right-linear CFG from that.

7. (10 points)

Cells have RNA molecules which can be written down as strings over alphabet $\{a, c, g, u\}$. An RNA molecule $w$ 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 $aacauggucaacgggccauguu$ has a stem-loop structure, because there exists a combination $x, y, z$ where $x = aacaugg$, $y = ucaacggg$, 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.

RNA Stem-Loop for the RNA *aacauggucaacgggccauguu*:




a) Write a grammar for recognizing RNA molecules (i.e., strings) 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.