## Problem Set 5

**Due:** Thursday, March 12, 2009 at 12:30 in class (i.e., SC 1105) Version: **1.0** 

Submission instructions: Submit each problem on a separate sheet of paper, put your name on each sheet, and write your discussion section time and day (e.g. Tuesday 10am) in the upper righthand corner. These details may sound picky, but they make the huge pile of homeworks much easier to grade quickly and more importantly, since we return them in the discussion sections, easier for you to get them back.

Also, write on each exercise the name/netid of your group members.

## (Q1) What is in, what is out?

[Category: Understanding., Points: 10]

For each of the following grammars answer the following:

- (a) Is abcd in the language of the grammar? If so give an accompanying derivation and parse tree.
- (b) Is acaada in the language of the grammar, if so give an accompanying derivation and parse tree.
- (c) What is the language generated by the grammar and explain your answer.

(Note: assume  $\Sigma = \{a, b, c, d\}$  and the start symbol is S for both grammars.)

$\mathcal{G}_1$ :	$S \rightarrow aSd   A   C$ $A \rightarrow aAc   B$ $B \rightarrow bBc   \epsilon$ $C \rightarrow bCd   B$	$\mathcal{G}_2$ :	$\begin{array}{ll} S \to B \mid \ AA \\ A \to cA \mid dB \\ B \to aSa \mid \epsilon. \end{array}$
	$C \rightarrow bCd \mid B$		1

## (Q2) Building grammars.

[Category: Construction., Points: 10]

Show that the following languages are context-free by giving a context-free grammar for each of them.

(a)  $L = \left\{ \mathbf{a}^{i} \mathbf{b}^{j} \mid 2i \leq j \leq 3i, i, j \in N \right\}.$ 

(Hint: Build a grammar for the case that j = 2i and the case j = 3i, and think how to fuse these two grammars together to a single grammar.)

(b)  $L' = \left\{ w \in \{0,1\}^* \mid \text{w contains equal number of occurrences of substring 01 and 10} \right\}.$ 

Thus 101 contains a single 01 and a single 10 and as such belongs to the language, while 1010 does not as it contains two 10's and one 01.

(Q3) Prove this.

[Category: Proof, Points: 10]

Consider the following proof.

**Lemma 0.1** If  $L_1$  and  $L_2$  are both context-free languages, then  $L_1 \cup L_2$  is a context-free language.

*Proof:* Let  $\mathcal{G}_1 = (\mathcal{V}_1, \Sigma, \mathcal{R}_1, \mathsf{S}_1)$  and  $\mathcal{G}_2 = (\mathcal{V}_2, \Sigma, \mathcal{R}_2, \mathsf{S}_2)$  be context free grammars for  $L_1$  and  $L_2$ , respectively, where  $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ . Create a new grammar

$$\mathcal{G} = (\mathcal{V}_1 \cup \mathcal{V}_2, \Sigma, \mathcal{R}, \mathsf{S}),$$

where  $\mathsf{S} \notin \mathcal{V}_1 \cup \mathcal{V}_2$  and  $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \{\mathsf{S} \to \mathsf{S}_1, \mathsf{S} \to \mathsf{S}_2\}.$ 

We next prove that  $L(\mathcal{G}) = L_1 \cup L_2$ .

 $L(\mathcal{G}) \subseteq L_1 \cup L_2$ :

Consider any  $w \in L(\mathcal{G})$ , and any derivation of w by  $\mathcal{G}$ . It must be of the following form:

$$\mathsf{S} \to \mathsf{S}_i \to X_1 X_2 \to \ldots \to w,$$

where i is either 1 or 2. Assume, without loss of generality, that i = 1, and observe that if we remove the first step, this derivation becomes

$$S_1 \to X_1 X_2 \to \ldots \to w.$$

Namely,  $S_1 \stackrel{*}{\Longrightarrow} w$  using grammar rules only from  $\mathcal{R}_1$ . We conclude that  $w \in L(\mathcal{G}_1) = L_1$ , as  $S_1$  is the start symbol of  $\mathcal{G}_1$ .

The case i = 2 is handled in a similar fashion.

Thus, we conclude that  $w \in L_1 \cup L_2$ , implying that  $L(\mathcal{G}) \subseteq L_1 \cup L_2$ .

## $L_1 \cup L_2 \subseteq L(\mathcal{G})$ :

Consider any word  $w \in L_1 \cup L_2$ , and assume without limiting generality, that  $w \in L_1$ . As such, we have that  $\mathsf{S}_1 \xrightarrow{*}_{\mathcal{G}_1} w$ . But  $\mathsf{S} \to \mathsf{S}_1$  is a rule in  $\mathcal{G}$ , and as such we have that

$$\mathsf{S} \to \mathsf{S}_1 \xrightarrow[\mathcal{G}_1]{*} w.$$

Namely,  $S \xrightarrow{*}_{\mathcal{G}} w$ , since all the rules of  $\mathcal{G}_1$  are in  $\mathcal{G}$ . We conclude that  $w \in L(\mathcal{G})$ .

Putting the above together, implies the lemma.

Provide a detailed formal proof to the following claim, similar in spirit and structure to the above proof.

**Lemma 0.2** If  $L_1$  and  $L_2$  are both context-free languages, then  $L_1L_2$  is a context-free language.

(Q4) Edit with some mistakes.

[Category: Construction., Points: 10]

The *edit distance* between two strings w and w', is the minimal number of edit operations one has to do to modify w into w'. We denote this distance between two strings x and yby  $\mathsf{EditDist}(x, y)$ . We allow the following edit operations: (i) insert a character, (ii) delete a character, and (iii) replace a character by a different character. For example, the edit distance between shalom and halo is 2. The edit distance between har-peled and sharp\_eyed is 4:

```
har-peled \implies shar-peled \implies sharp_eled \implies sharp_eled \implies sharp_eyed.
```

For the sake of simplicity, assume that  $\Sigma = \{a, b, \$\}$ . For a parameter k, describe a CFG for the language

$$L_k = \left\{ x \$ y^R \mid x, y \in \{\mathtt{a}, \mathtt{b}\}^*, \mathsf{EditDist}(x, y) \le k \right\}.$$

For example, since  $\mathsf{EditDist}(\mathsf{aba}, \mathsf{bab}) = 2$ , we have that  $\mathsf{aba}$   $\mathsf{bab} \in L_2$ , but  $\mathsf{aba}$   $\mathsf{bab} \notin L_1$ . Similarly,  $\mathsf{EditDist}(\mathsf{aaaa}, \mathsf{abb}) = 3$ , and as such  $\mathsf{aaaa}$   $\mathsf{bba} \in L_3$ , but  $\mathsf{aaaa}$   $\mathsf{bba} \notin L_2$ .

(Hint: What is the language  $L_0$ ? Try to give a grammar to  $L_1$  before solving the general case.)

Provide a short argument why your CFG works.

(Q5) Speedup theorem for CFGs.[Category: Proof., Points: 10]

Assume you are given a CFG  $\mathcal{G} = (\mathcal{V}, \Sigma, \mathcal{R}, \mathsf{S})$ , such that any word  $w \in L(\mathcal{G})$  has a derivation of w with f(n) steps, where n = |w|. Here f(n) is some function.

Prove the following claim.

**Claim 0.3** There exists a grammar  $\mathcal{G}'$  such that  $L(\mathcal{G}) = L(\mathcal{G}')$ , and furthermore, any derivation of a word  $w \in L(\mathcal{G})$  of length n requires at most  $\lceil f(n)/2 \rceil$  derivation steps.

For example, consider the grammar  $S \rightarrow aSb \mid \epsilon$ .

A word  $\mathbf{a}^n \mathbf{b}^n$  can be derived using n+1 steps. Its speeded up version is

 $S' \rightarrow aaS'bb \mid ab \mid \epsilon.$ 

Now,  $\mathbf{a}^n \mathbf{b}^n$  has derivation of length  $\lceil (n+1)/2 \rceil$  by the grammar of S'.

(Hint: Consider a parse tree of height h in the grammar, and think how to modify the grammar, so that you get a parse tree for the same word with height only  $\lceil h/2 \rceil$ .)