CSE 322: Formal Models in Computer Science Winter 2005

PROBLEM SET 7

Due Friday, March 4, 2005, in class

Reading assignment: Section 2.3 and Theorem 7.14 of Sipser's text.

There are **FIVE** questions. Each question is worth **12 points**.

---

1. Use the pumping lemma for context-free languages to prove that the following languages are not context-free:

   (a) $L_1 = \{0^p \mid p \text{ is a prime}\}$.

   (b) $L_2 = \{w\#x \mid w \text{ is a substring of } x, \text{ where } w, x, \in \{a, b\}^*\}$.

2. Let G be a context-free grammar in Chomsky normal form that contains k variables. Show that if $L(G)$ contains a string of length greater than 2^{k-1}, then $L(G)$ is infinite.

3. *(Tricky!)* Prove the following strengthening of the pumping lemma, where we require that **both** substrings v and y be nonempty when the string w is broken up as $uvxyz$: If L is a CFL, then there exists an integer $p \geq 1$ such that $\forall\ w \in L$, $|w| \geq p$, there exists a way to break down w as $w = uvxyz$, satisfying the conditions:

   (i) For each $i \geq 0$, $uv^ixy^iz \in L$

   (ii) $v \neq \epsilon$ **and** $y \neq \epsilon$

   (iii) $|vxy| \leq p$

   (the second condition is the stronger one compared to the version proved in class).

4. Prove that the language $A = \{a^ib^jc^kd^\ell \mid i = 0 \text{ or } i = 1 \text{ or } j = k = \ell\}$ is not context-free.

5. Consider the grammar G that is in Chomsky Normal Form with the following rules:

$$\begin{aligned} S &\rightarrow AB \mid BC \\ A &\rightarrow BA \mid a \\ B &\rightarrow CC \mid b \\ C &\rightarrow AB \mid a \end{aligned}$$

   Your task in this problem is to run the Cocke-Younger-Kasami (CYK) algorithm to determine whether a specific string belongs to $L(G)$. Recall that the algorithm proceeds by filling out a table T whose (i, j)'th entry $T[i, j]$ for $j \geq i$ consists of the non-terminals that can derive the substring $w_iw_{i+1} \ldots w_j$.

   (a) Run the CYK algorithm to determine if the string *ababa* belongs to $L(G)$. Specifically, fill in the entries of the table $T[i, j]$, for $1 \leq i \leq j \leq 5$.

   (b) What is the verdict of the algorithm? That is, does *ababa* $\in L(G)$? Why, or why not?