Computational Models, Spring 2015 Exrecise #2

Non-Regular Languages and Context free grammars

1. We showed in class that Regular languages are closed under union. Are they closed under infinite union? Prove your answer formally.

   No, consider $\mathbf{L_i} = \{\mathbf{a^i b^i}\}$ which is regular for every i (finite language). Their union is $\{\mathbf{a^n b^n}$ s.t. $\mathbf{n} \in \mathbb{N}\}$ which is not regular

2. Let L be a regular language state for each language if it is regular or not. Give a complete formal proof:

   (a) $L_a = \{ww' : w \in L \text{ and } w' \notin L\}$

   REGULAR - this is simply LL^c

   (b) $L_b = \{y : \exists x, z \; s.t. |x| = |z| \text{ and } xyz \in L\}$
   REGULAR - ("keep middle") for each pair of states p and q such that p is reachable from q_0 in the same number of steps that a state in F is reachable from q (that is, the potential x and z) we will have an epsilon transition from a new initial state to p, and q will be the accepting state (we will have a replica for each pair of such states). Note that we don't need an algorithm to find the relevant pairs, only that such set of pairs exists (and is finite.)

   (c) $L_c = \{xy : |x| = |y| \text{ and } \exists \sigma \; s.t. \; x\sigma y \in L\}$
   NON REGULAR - choose $L = 0^*12^*$. Thus $L_c \cap 0^*2^* = 0^n 2^n$.

   (d) $\text{Pref}(L) = \{x : \exists y \; s.t. \; xy \in L\}$
   REGULAR - for each state that LEADS to an accepting state, add an epsilon transition to an accepting state.

3. Prove that the following languages are not regular

   (a) $L_a = \{(ab)^n c^n : n \geq 0\}$ above $\Sigma = \{a, b, c\}$
   Apply the homomorphism $h : \Sigma \to \Delta$ for $\Delta = \{0, 1\}$ such that, $h(a) = 0$, $h(b) = \epsilon$, $h(c) = 1$, To obtain $h(L_a) = \{0^n 1^n : n \geq 0\}$ which we know is not regular.
   Alternatively, use pumping lemma, for length p choose $\mathbf{w = xyz = (ab)^p c^p}$. The y may be either at the a, ab, b or bc part. Every part, if pumped yields a word not in the language

   (b) $L_b = \{a^i b^j c^k : i = j \text{ or } j = k\}$ above $\Sigma = \{a, b, c\}$
   Assume regularity and get contradiction by $\mathbf{L_b \cap \{a^* b^*\} = \{a^n b^n\}}$

   (c) $L_c = \{ww : w \in \{0, 1\}^*\}$
   Assume regularity and get contradiction by $\mathbf{L_c \cap \{0^* 1 0^* 1\} = \{0^n 1 0^n 1\}}$

   (d) $L_d = \{\text{balanced strings }\}$ above $\Sigma = \{(,)\}$
   A string is balanced if the number of (equals the number of) and while reading the string, the number of (is equal or greater than the number of).
   Assume regularity and get contradiction by $\mathbf{L_d \cap \{(^* \;)^*\} = \{(^n \;)^n\}}$

   (e) $L_e = \{\{0\}^n \mid n \text{ is a prime number}\}$ above $\Sigma = \{0\}$
   For the sake of contradiction, assume that L_e is regular. The Pumping Lemma must then apply; let p be the pumping length and let n be any prime number at least as large as p (such an n is guaranteed to exist since there are an infinite number of primes), and consider the string $w = 0^n \in L_e$. Since $|w| \geq p$, it must be possible to

split w into three pieces xyz satisfying the conditions of the Pumping Lemma. Now consider the string xy^iz. The string xy^iz has length $n+(i-1)|y|$. Letting $i=n+1$, we have

$$n+(i-1)|y| = n+((n+1)-1)|y| = n+n|y| = n\cdot(1+|y|)$$

which is a composite number since $|y|>0$, and hence xy^iz is not an element of the language. Thus, the Pumping Lemma is violated, and the language in question cannot be regular.

4. Prove that the following languages are not regular using the Myhill-Nerode Theorem

   (a)
   $$L = \{ww \mid w \in \Sigma^*\} \text{ for any } \Sigma^* \text{ such that } |\Sigma| > 1$$

   Let $S = \{a^n \mid n \in \mathbb{N}\}$. This set is infinite because it contains one string for each natural number. Now, consider any a $a^n, a^m \in S$ for $n \neq m$. Then $a^na^n \in L$ while $a^ma^n \notin L$, so a^n and a^m are in different equivalent classes. Thus S is an infinite set of strings that are all distinguishable relative to L. Therefore, by the Myhill-Nerode Theorem, L is not regular.

   (b) Let L be the set of algebraic expressions involving identifiers x and y, operations $+$ and $*$ and left and right parentheses. L can be defined recursively as follows:
   Basis Clause: x and y are in L.
   Inductive Clause: If α and β are in L, then $(\alpha+\beta)$ and $(\alpha*\beta)$ are in L.
   Extremal Clause: Nothing is in L unless it is obtained from the above two clauses.

   For example, x , $(x*y)$, $((x+y)*x)$ and $(((x*y)+x)+(y*y))$ are algebraic expressions.

   Consider the set of strings $S = (^kx \mid k$ is a positive integer , that is, the set of strings consisting of one or more right parentheses followed by identifier x. This set is infinite and it can be shown to be pairwise distinguishable with respect to L as follows: Let $(^kx$ and $(^mx$ be arbitrary two strings of S, where k and m are positive integers and $k \neq m$. Select $[+x)]^k$ as a string to be appended to $(^k$ and $(^m$. For example $[+x)]^3$ is $+x)+x)+x)$. Then $(^kx+[+x)]^k$ is in L but $(m^x+[+x)]^k$ is not in L because the number of ('s is not equal to the number of)'s in the latter string. Hence S is pairwise distinguishable with respect to L which, in turn, implies that it is not regular.

5. Give an example of a non-regular language L and a homomorphism h such that $h(L)$ is regular.

   Consider $L = \{0^n1^n \mid n \geq 0\}$ $h(0)=0$ $h(1)=\varepsilon$ $h(L)=0^*$

6. Let $L = L((00+1)^*)$ and $h : \{a,b\}^* \to \{0,1\}^*$ be defined by $h(a)=01$ and $h(b)=10$.

   What is $h^{-1}(L)$? Prove your answer formally.
   Hint: You need to prove that $h(w) \in L$ iff $w = (?)^n$ where ? is some string above $\{a,b\}$. One direction (if...) is easy. For the other direction (only-if) assume that $h(w)$ is in L and show that it is in the right form. This can be done by considering all the possible conditions where a string is NOT of that form and showing that if any of them hold, then $h(w)$ is in not L

   The language is: $h^{-1}(L) = L((ba)^*)$.
   We shall prove that $h(w) \in L$ iff $w = (ba)^n$.
   (if) Suppose $w=(ba)^n$ for $n \geq 0$. Note that $h(ba)=1001$ so $h(w)$ is n repetitions of 1001. Thus $h(w) = (1001)^n \in L$.
   (Only-If) We assume that $h(w)$ is in L and show that w is of the form $baba\ldots ba$. There are four conditions under which a string is NOT of the above form, and we shall show that if any of them hold then $h(w)$ is not in L. That is, we prove the contrapositive of the statement we set out to prove.
   So, let $h(w) \in L$ and suppose $w \notin L((ab)^*)$. There are 4 cases to consider:

(a) w **begins with** a**. Then** $h(w)$ **begins with** 01 **and** $\notin L((00+1)^*)$**, since it has an isolated 0.**

(b) w **ends in** b**. Then** $h(w)$ **ends in** 10 **and** $\notin L((00+1)^*)$**.**

(c) w **has two consecutive** a**, i.e.,** $w = xaay$**. Then** $h(w) = z0101v$ **and** $\notin L((00+1)^*)$**.**

(d) w **has two consecutive** b**, i.e.,** $w = xaay$**. Then** $h(w) = z1010v$ **and** $\notin L((00+1)^*)$**.**

Thus, whenever one of the above cases hold, $h(w)$ **is not in** L**. However, unless at least one of items (1) through (4) hold, then** w **is of the form** $baba \ldots ba$**. To see why, assume none of (1) through (4) hold. Then (1) tells us** w **must begin with** b **and (2) tells us** w **must end with** a**. Statements (3) and (4) tell us that** a **and** b **must alternate in** w**. Thus the logical OR of (1) through (4) is equivalent to the statement** w **is not of the form** $baba \ldots ba$**.**

7. Given the following grammar,

$$S \to PS|\varepsilon \qquad P \to (S)$$

Show how to construct the word $(())()$. Explain (in words) what is the language described by this grammar.

This is the language of all balanced Parentheses

$S \Rightarrow PS$
$\Rightarrow PPS$
$\Rightarrow PP$
$\Rightarrow (S)P$
$\Rightarrow (S)(S)$
$\Rightarrow (PS)(S)$
$\Rightarrow (P)(S)$
$\Rightarrow ((S))(S)$
$\Rightarrow (())()$