Midterm #2

Mark all correct answers in each of the following questions.
Σ denotes an arbitrary alphabet unless otherwise specified.

1. (a) There exists an infinite regular language L for which there exists a unique triplet of words $u, v, w \in \Sigma^*$ with $v \neq \varepsilon$ and $uv^n w \in L$ for every $n \geq 0$.

   (b) For every infinite regular language L there exist three non-empty mutually distinct words u, v, w such that $uv^n w \in L$ for every $n \geq 0$.

   (c) There exists a finite regular language L for which there exist words $u, v, w \in \Sigma^*$ with $v \neq \varepsilon$ and $uv^n w \in L$ for every $n \geq 0$. However, this property does not hold for every finite regular language.

   (d) The language $L = \{w \in \{a, b\}^* : |w|_a - |w|_b = 10\}$ is not regular.

   (e) The language $L = \{w \in \{a, b\}^* : |w|_a + |w|_b = 10\}$ is not regular.

   (f) If $a, b \in \Sigma$, then the language $L = \{w \in \Sigma^* : |w|_a + |w|_b = 10\}$ is regular.

   (g) None of the above.

2. (a) If the grammar G_1 is given by the rules
      - $S \to aSa \mid B$,
      - $B \to bB \mid \varepsilon$,

      then $L(G_1)^R = L(G_1)$.

   (b) If the grammar G_2 is given by the rules
   $S \to aSa \mid bSb \mid \varepsilon$,
   then $L(G_2)$ is the language of all palindromes over $\{a, b\}$.

(c) If the grammar G_3 is given by the rules
$S \to aSa \mid bSb \mid a \mid b \mid \varepsilon$,
then $L(G_3)$ is the language of all palindromes over $\{a, b\}$. However, $L(G_2)$ is strictly contained in $L(G_3)$.

(d) If the grammar G_4 is given by the rules
$S \to SS \mid a^3 \mid a^4$,
then the difference $\{a\}^* - L(G_4)$ consists of exactly three words.

(e) If the grammars G_5 and G_6 are given by the rules

- $S \to \varepsilon \mid aSb \mid SS$,
- $S \to \varepsilon \mid aSb \mid SSS$,

then $L(G_5) = L(G_6)$.

(f) Let $G_7,\ G_8,\ G_9$ the three grammars obtained from G_5 by omitting one of the three grammar rules of G_5. Then two of the languages $L(G_7),\ L(G_8),\ L(G_9)$ are regular and one is not.

(g) None of the above.

3. Let $G = (N, \Sigma, R, S)$ be a context-free grammar.

(a) If $R \subseteq N \times ((N \cup \Sigma)^2)^*$, then every word in $L(G)$ is of even length.

(b) If $R \subseteq N \times (N^*\Sigma N^*\Sigma N^*)^*$, then every word in $L(G)$ is of even length.

(c) If $R \subseteq N \times ((N \cup \{\varepsilon\})\,\Sigma\,(N \cup \{\varepsilon\}))$, then there exists a constant C such that $|w| \leq C$ for every $w \in L(G)$.

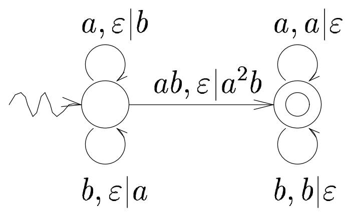
(d) If $L(G) = L\,(a^*b^*c^*)$, then G is a regular grammar.

(e) If $L(G)$ is regular, then there exists a context-free grammar $G_1 = (N_1, \Sigma, R_1, S_1)$ with $R_1 \subseteq N_1 \times (\Sigma N_1 \cup \{\varepsilon\})$ such that $L(G_1) = L(G)$.

(f) Let $G_2 = (N, \Sigma, R_2, S)$ be another context-free grammar. If $R_2 = R$ then $L(G_2) = L(G)$, if $R_2 \subseteq R$ then $L(G_2) \subseteq L(G)$. and if $R_2 \subsetneqq R$ then $L(G_2) \subsetneqq L(G)$.

(g) None of the above.

4. Let $M = (Q, \{a, b\}, \{a, b\}, \Delta, s, \{f\})$ be the pushdown automaton below.



(a) If $w \in L(M)$, then $|w| \neq 1000$.

(b) The language $L(M) \cap L(a^*b^*a^*b^*)$ is infinite.

(c) Let $w \in L(M)$. If the second letter of w is a, then the second last letter of w is b.

(d) In $L(M)$ there is exactly one word in which both the third letter and the third letter from the end (i.e., at the $(n-2)$'nd place if the word is of length n) are a.

(e) Let u be a word over $\{a, b\}$, such that if we put

$$\Delta_1 = \Delta - \{((s, ab, \varepsilon), (f, a^2b))\} \cup \{((s, u, \varepsilon), (f, \varepsilon))\}$$

and

$$M_1 = (Q, \{a, b\}, \{a, b\}, \Delta_1, s, \{f\}),$$

then $L(M_1) = L(M)$. Then u is of even length.

(f) There exists no pushdown automaton M_2 with a single state for which $L(M_2) = L(M)$.

(g) None of the above.

Solutions

1. If L is an infinite regular language, then by the pumping lemma there exist words $u, v, w \in \Sigma^*$ with $v \neq \varepsilon$ and $uv^nw \in L$ for every $n \geq 0$. Taking $u_1 = u$, $v_1 = v^2$ and $w_1 = w$ we obtain a distinct triplet of

words with the same property. Any triplet of the form $u_2 = uv^m$, $v_2 = v$, $w_2 = v^n w$, also satisfies this property, and by selecting m and n appropriately we may ensure that the three words are mutually distinct.

If $v \neq \varepsilon$, then the set $\{uv^n w : n \geq 0\}$ is infinite. Hence no finite language contains such a set.

If the language $L = \{w \in \ a, b\}^* : |w|_a - |w|_b = 10\}$ was regular, then so would be the intersection

$$L_1 = L \cap L(a^* b^*) = \left\{a^{n+10} b^n : n \geq 0\right\},$$

the concatenation

$$L_2 = L_1 \{b\}^{10} = \left\{a^{n+10} b^{n+10} : n \geq 0\right\} = \{a^n b^n : n \geq 10\},$$

and the union (with a finite language)

$$L_3 = L_2 \cup \{a^n b^n : n < 10\} = \{a^n b^n : n \geq 0\}.$$

As the latter language is known to be non-regular, so is L.

If $a, b \in \Sigma$, then, putting $\Sigma_1 = \Sigma - \{a, b\}$, we have

$$\{w \in \Sigma^* : |w|_a + |w|_b = 10\} = \left(\Sigma_1^* \{a, b\}\right)^{10} \Sigma_1^*,$$

which is a regular language.

Thus, only (b), (d) and (f) are true.

2. Any word in $L(G_1)$ is obtained by applying some number m of times the rule $S \rightarrow aSa$, then a single application of rule $S \rightarrow B$, some number n of times the rule $B \rightarrow bB$, and finally a single application of $B \rightarrow \varepsilon$. Consequently, $L(G_1) = \{a^m b^n a^m : m, n \geq 0\}$. Thus, each word in $L(G_1)$ is palindromic, and in particular $L(G_1)^R = L(G_1)$.

   Palindromes of even length are of the form ww^R for some $w \in \Sigma^*$, and palindromes of odd length are of the form $w\sigma w^R$ for some $w \in \Sigma^*$ and $\sigma \in \Sigma$. The grammar G_2 accepts only words of the form ww^R, while G_3 accepts those of the form $w\sigma w^R$ as well.

   The language $L(G_4)$ consists of all words of the form a^n, where $n = 3k + 4l$, with k and l non-negative integers, not both 0. We see easily that n may take any non-negative value, except for 0, 1, 2 and 5. Therefore, $L(G_4) = \{a\}^* - \{\varepsilon, a, a^2, a^5\}$.

We have

$$S \Rightarrow_{G_5} SS \Rightarrow_{G_5} SSS$$

and

$$S \Rightarrow_{G_6} SSS \Rightarrow_{G_5} SS,$$

so that each rule of either grammar can effectively be applied in the other grammar by means of 1 or 2 derivations. Hence $L(G_5) = L(G_6)$.

If we omit from G_5 the rule $S \to \varepsilon$, then the resulting grammar clearly accepts the language $\emptyset$, which is regular. If we omit the rule $S \to aSb$, then the resulting grammar accepts the language $\{\varepsilon\}$ – again regular. However, when omitting the rule $S \to SS$, the resulting language is $\{a^n b^n : n \geq 0\}$, which is not regular.

Thus, (a), (c), (e) and (f) are true.

3. The language $N^*\Sigma N^*\Sigma N^*$ consists of all words over $N \cup \Sigma$ containing exactly two letters from Σ. Hence $(N^*\Sigma N^*\Sigma N^*)^*$ consists of all words containing an even number of letters from Σ. It follows that, if $R \subseteq N \times (N^*\Sigma N^*\Sigma N^*)^*$, then each time a grammatical rule is used, the number of letters from Σ in the words increases by an even number. In particular, as we always start with the letter S, any word in $L(G)$ will contain an even number of letters. However, the condition $R \subseteq N \times ((N \cup \Sigma)^2)^*$ only ensures that each derivation replaces a letter by a word of even length, which implies nothing regarding the number of letters in words in $L(G)$. For example, the grammar G given by the rules
$S \to aS \mid a^2$
satisfies the required condition, yet $L(G)$ consists of all words of length at least 2 over $\{a\}$.

   The grammar G given by the rules
   $S \to aS \mid a$
   satisfies the condition $R \subseteq N \times ((N \cup \{\varepsilon\})\,\Sigma\,(N \cup \{\varepsilon\}))$, yet $L(G) = \{a\}^+$, which contains arbitrarily long words.

   The grammar G given by the rules

- $S \to A^2$,
- $A \to aA \mid B$,
- $B \to bB \mid C$,
- $C \to Cc \mid \varepsilon$,

satisfies $L(G) = L(a^*b^*c^*)$ even though it is not regular.

Going over the proof of the theorem stating that a language is regular if and only if it is accepted by a regular grammar, we see that the regular grammar $G = (N, \Sigma, R, S)$, constructed so as to accept a regular language L, satisfies $R \subseteq N \times (\Sigma N \cup \{\varepsilon\})$.

The grammar G given by the rules
$S \to aS \mid \varepsilon$
accepts the language $\{a\}^*$, same as any grammar over $\{a\}$ consisting of the same rules and any additional ones.

Thus, only (b) and (e) are true.

4. To obtain a word in $L(G)$, one may start with any sequence of transitions from s to s. If this sequence generates the word v, then after reading v the contents of the stack is clearly $\bar{v}^R$, where $\bar{v}$ denotes the word obtained from v upon replacing each a by b and vice versa. The transition from s to f changes the word read by now to vab and the contents of the stack to $a^2b\bar{v}^R$. The only way to empty the stack is by reading now $a^2b\bar{v}^R$. Summing up, we have $L(M) = \{vaba^2b\bar{v}^R : v \in \Sigma^*\}$.

   From the above representation of $L(M)$ it is clear that each word in $L(M)$ is of odd length, and in particular cannot be of length 1000. Also

   $$L(M) \cap L(a^*b^*a^*b^*) = \left\{a^kba^2b^k : k \geq 1\right\},$$

   so that the intersection is infinite.

   Let $w = vaba^2b\bar{v}^R \in L(M)$. The second letter of w is a if and only if either $|v| = 1$ or both $|v| \geq 2$ and the second letter of v is a. In either case the second last letter of w is b. The third letter of w is a if and only if either $|v| = 0$, or $|v| = 2$, or both $|v| \geq 3$ and the third letter of v is a. The third last letter of w is a if and only if either $|v| = 0$, or $|v| = 1$, or both $|v| \geq 3$ and the third letter of v is b. Hence the only word in $L(M)$, in which both the third letter and the third letter from the end are a, is aba^2b.

By the same considerations used to find $L(M)$, we see that $L(M_1) = \{vu\bar{v}^R : v \in \Sigma^*\}$. Hence $L(M_1) = L(M)$ if and only if $u = aba^2b$.

Suppose M_2 is a pushdown automaton with a single state, accepting the same language as M. Since $L(M)$ is non-empty, the state of M_2 must be accepting. Hence any concatenation of words accepted by M_2 is also accepted by M_2. In particular, since $aba^2b \in L(M)$ we have $aba^2b \in L(M_2)$ and therefore $(aba^2b)^2 \in L(M_2)$. However, $(aba^2b)^2 \notin L(M_2)$, so that $L(M_2) \neq L(M)$.

Thus, (a), (b), (c), (d) and (f) are true.