CSC236 tutorial exercises, Week #9
sample solutions

1. Consider the languages NOBBB $= \{s \in \{a, b\}^* \mid$ s does not contain substring $bbb\}$, and $L = \{a, ab, ba\}^*$. Show that they are not equal by finding a string that belongs to one but not the other.

   Solution:
   b, bb, and bab are a few examples of strings which belong to NOBB but not L.

2. Recall the language $\mathcal{S} \subseteq \{y, u, h\}^*$ introduced in assignment 1, which was defined as the smallest set such that:

   - $u \in \mathcal{S}$
   - if $s \in \mathcal{S}$ then $ys \in \mathcal{S}$
   - if $s \in \mathcal{S}$ then $sh \in \mathcal{S}$
   - if $s_1, s_2 \in \mathcal{S}$ then $s_1 s_2 \in \mathcal{S}$

   Write a regular expression for $\mathcal{S}$.

   Solution:
   $(y^*uh^*)(y^*uh^*)^*$
   As you probably surmised during the assignment, strings in this language have one or more u's, where each u is preceded by some number of y's (possibly zero), and followed by some number of h's. Note that the RE $(y^*uh^*)^*$ alone will not work, specifically because the empty string, ε, is not an element of $\mathcal{S}$.

3. Consider the language AA, consisting of all strings in $\{a, b\}^*$ that contain substring aa.

   (a) Give a recursive definition of AA.

   Solution:

   i. $aa \in$ AA
   ii. if $s \in$ AA, then the following strings are also in AA: as, bs, sa, sb

   (b) Write a regular expression for AA.

   Solution:
   $(a + b)^* aa(a + b)^*$

(c) Write a regular expression for $\overline{AA}$, i.e. the language of strings which *don't* contain aa.

Solution:
$b^*(abb^*)^*(a+\varepsilon)$

The intuition behind this RE is that, if s is a string in $\overline{AA}$, then every a in s (unless it's the last character) must be followed by one or more b's.

Working in the other direction (every a except for the first character must be preceded by one or more b's), gives the following parallel RE, which also captures the language:

$(a+\varepsilon)(bb^*a)^*b^*$

4. Describe a sufficient condition on languages S and T such that $ST = TS$. (This is not generally true - i.e. concatenation of languages is not commutative.)

Optional challenge: How many more conditions can you think of? Can you describe conditions that are *necessary* and sufficient?

Solution:
This problem is surprisingly non-obvious! Below are some (sometimes overlapping) conditions under which $ST = TS$, in approximately increasing order of non-obviousness:

(a) $S = T$

(b) One of the sets is $\{\varepsilon\}$. e.g. if $S = \{\varepsilon\}$ then $ST = T = TS$ (akin to multiplication by 1).

(c) One of the sets is $\emptyset$. Concatenated with any other set, this produces the empty set (akin to multiplication by 0).

(d) The alphabet of T and S has just a single symbol.

(e) $\exists k \in \mathbb{N}, S = T^k$ (or vice versa). Then $ST = T^kT = T^{k+1} = TT^k = TS$. (The $\{\varepsilon\}$ case is a special case of this, setting $k = 0$.)

(f) Generalization of the above: there exists a 'common factor' language L such that $S = L^j, T = L^k$ for some j, k.

(g) Even further generalization: there exists a common factor language L such that S and T are each the union of powers of L. e.g. $S = L^2, T = L^0 \cup L^3 \cup L^5$. (This subsumes all the conditions above, including the case of $\emptyset$ and the single-symbol alphabet).

We've come pretty far, but it's still not clear that the last condition in the list is *necessary* and sufficient. (If you can find a case that's not covered by the last condition, I'd be curious to see. Post it on Piazza!)

Postscript: Osama, who leads the tutorials in UC330/UC261, pointed out this additional condition not covered by the list above: $T = S^*AS^*$ (or vice-versa) for some language A. For example, $S = \{0\}^*$, $T = \{0\}^*\{0,1\}^*\{0\}^* = \{0,1\}^*$. (But there may well be more!)