Homework 1

Email to `alexsclark@gmail.com`

1. Consider the infinite language $\{ab, aabb, aaabbb, \ldots\}$ which consists of any number of as followed by the same number of bs. When we write $u \equiv v$ this means that u and v are completely mutually substitutable.

   Is $ab \equiv_L aabb$?

   Yes. Every where that ab is found we can substitute $aabb$ and vice versa. More formally we can see that the distribution of ab

   $$ab^{\triangleright} = \{\Box, a\Box b, aa\Box bb, \ldots\} = aabb^{\triangleright}$$

   Is $a \equiv aab$?

   No because a can occur in the context $\Box abb$, but aab cannot.

2. Consider the substitutable learner that we looked at on Monday: When the learner receives the two examples::

   - the cats
   - the red cats

   What language does it generalize to?

   The learner generalizes to the language which contains "the cats", "the red cats", "the red red cats" etc.

   More formally we can see that "the" and "the red" occur in the same context as do, more reasonably, "cats" and "red cats". So we have a grammar which looks like:

   - $S \rightarrow DN$
   - $D \rightarrow the$
   - $N \rightarrow cats$
   - $A \rightarrow red$
   - $D \rightarrow DA$
   - $N \rightarrow AN$

   This generates a regular language, defined by the regular expression $tr * c$ using obvious abbreviations.

3. What happens when the learner receives the following three examples?

   - cats
   - the cats
   - the red cats

Now we have that “the cats” is also in the same context as “cats” and “the red cats”. So we have one fewer category. This gives us 3 categories rather than 4, which we can write S, D and A and a grammar:

- $S \rightarrow DS$
- $D \rightarrow the$
- $S \rightarrow cats$
- $A \rightarrow red$
- $D \rightarrow DA$

So this generates in addition to the previous example: “the red the red red cats”, “the the the cats” and so on. As a regular expression $(t(r^*))^*c$.