Lecture #10: Nonregular Languages, Part Two
Completion of Example

The Problem To Be Solved

Let $\Sigma = \{\mathtt{a}, \mathtt{b}\}$ and consider the languages

$$L_1 = \{\mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{Z} \text{ and } n \geq 0\} \subseteq \Sigma^\star$$

and

$$L_2 = \{\mathtt{a}^n\mathtt{b}^m \mid n, m \in \mathbb{Z},\ n, m \geq 0, \text{ and } n \neq m\} \subseteq \Sigma^\star.$$

It was proved in the material for Lecture #9, using the "Pumping Lemma for Regular Languages", that L_1 is *not* a regular language. In this document a proof (outlined in the preparatory material for Lecture #10), that L_2 is *also* not a regular language, is given in detail.

Three More Languages of Interest

To begin, let us consider languages

$$L_3 = \{\omega \in \Sigma^\star \mid \mathtt{ba} \text{ is a substring of } \omega\} \subseteq \Sigma^\star,$$

$$L_4 = L_2 \cup L_3 \subseteq \Sigma^\star,$$

and

$$L_5 = \{\mathtt{a}^n\mathtt{b}^m \mid n, m \in \mathbb{Z}, \text{ and } m, n \geq 0\} \subseteq \Sigma^\star.$$

Proving that $L_1 = L_4^C$

The proof of this claim *does not* involve the use of closure properties to prove that languages are not regular — so that students, who are interested in examples of this, can skip this section and continue the proof, after it, that L_4 is not a regular language, on page 4.

In order to prove that $L_1 = L_4^C$ it is necessary, and sufficient, to prove both that $L_1 \subseteq L_4^C$ and that $L_4^C \subseteq L_1$.

Proving that $L_1 \subseteq L_4^C$

Note that, since $L_4 = L_2 \cup L_3$, it follows by DeMorgan's laws that $L_4^C = L_2^C \cap L_3^C$.

- Let $\omega \in \Sigma^\star$ such that $\omega \in L_1$.

  Suppose — to obtain a contradiction — that $\omega \in L_2$ as well. Now, since $\omega \in L_1$ it follows, by the definition of this language that the number of copies of "a" in ω is equal to the number of copies of "b" in ω. However, it follows by the definition of L_3 that, since $\omega \in L_2$, the number of copies of "a" in ω must be *different* from the number of copies of "b" in ω.

  Since both of these conditions cannot be satisfied at the same time, a *contradiction* has been obtained — so that the only assumption that has been made must be false: $\omega \notin L_2$. Thus $\omega \in L_2^C$.

  Since ω was arbitrarily chosen from L_1, it follows that $L_1 \subseteq L_2^C$.

- Once again, $\omega \in \Sigma^\star$ such that $\omega \in L_1$.

  Then $\omega = \mathtt{a}^n\mathtt{b}^n$ for some non-negative integer n. The only pairs of consecutive symbols that can appear in ω are "aa", "ab" or "bb" — so that "ba" is not a substring of ω.

  Suppose, to obtain a contradiction, that $\omega \in L_3$ as well. Then it follows, by the definition of L_3, that "ba" *is* a substring of ω. Since this property, and the above property that "ba" is *not* a substring, cannot both be satisfied at the same time, a *contradiction* has been obtained — so that the only assumption that has been made must be false: $\omega \notin L_3$. Thus $\omega \in L_3^C$

  Since ω was arbitrarily chosen from L_1, it follows that $L_1 \subseteq L_3^C$.

- Since $L_1 \subseteq L_2^C$ and $L_1 \subseteq L_3^C$, $L_1 \subseteq L_2^C \cap L_3^C$. That is, $L_1 \subseteq L_4^C$ — as claimed.

Proving that $L_4^C \subseteq L_1$

Let $\omega \in \Sigma^\star$ such that L_4^C. As noted above, $L_4^C = L_2^C \cap L_3^C$ — so $\omega \in L_2^C$ and $\omega \in L_3^C$.

- Either $\omega = \lambda$, or $|\omega| \geq 1$, in which case either ω begins with "a" or ω begins with "b". All of these cases are considered below.

  - If $\omega = \lambda$ then $\omega \in L_5$, since $\omega = \mathtt{a}^0\mathtt{b}^0$.

- Consider the case that $|\omega| \geq 1$ and ω begins with "a". Either ω also includes a copy of "b" or it does not; each of these subcases is considered below.
  * Consider, first, the case that ω also includes a copy of "b" — so a string $\mathtt{a}^k\mathtt{b}$ is a prefix of ω, for some positive integer k.
    Suppose (to obtain a contradiction) that there is at least one copy of "a" in ω, somewhere after the first copy of "b". Consider the *first* such copy of "a": The symbol before it must be a copy of "b", so that "ba" is a substring of ω. Thus $\omega \in L_3$ — *contradicting* the fact that ω was chosen from L_4^C (so that $\omega \in L_3^C$ as well). It follows, from this contradiction, that the only assumption that was made must be false: There are no copies of "a" after the first copy of "b" in ω.
    It now follows that (since all the symbols after the first copy of "b" must also be copies of "b" as well) $\omega = \mathtt{a}^k\mathtt{b}^\ell$ for non-negative integers k and ℓ: $\omega \in L_5$ in this case, as well.
  * Consider, next, the case that ω *does not* include any copies of "b". Then all the symbols in ω must be copies of "a", so that $\omega = \mathtt{a}^k = \mathtt{a}^k\mathtt{b}^0$ for some positive integer k — and $\omega \in L_5$ in this case too.

  Thus $\omega \in L_5$ if $|\omega| \geq 1$ and ω begins with "a".
- Finally, consider the case that $|\omega| \geq 1$ and that ω begins with a copy of "b". Suppose — to obtain a contradiction — that ω includes at least one copy of "a". As above, consider the *first* copy of "a" in ω: There must be a copy of "b" immediately ahead of it, so that "ba" is a substring of ω, and $\omega \in L_3$. This *contradicts* the fact that $\omega \in L_4^C$, so that $\omega \in L_3^C$ (and $\omega \notin L_3$), and it follows that the only assumption that has been made must be false: ω does not include any copies of "a".
  Thus $\omega = \mathtt{b}^k = \mathtt{a}^0\mathtt{b}^k$ for some positive integer k, so that $\omega \in L_5$ in this case, once again.

Since it is true in every possible case, it follows from the above that $\omega \in L_5$.

Since ω was arbitrarily chosen such that $\omega \in L_4^C$, it follows that $L_4^C \subseteq L_5$

• Once again, let $\omega \in \Sigma^\star$ such that $\omega \in L_4^C$: Then it follows from the above that $\omega \in L_5 = L_1 \cup L_2$. However, since $L_4^C = L_2^C \cap L_3^C$, $\omega \in L_2^C$ — that is, $\omega \notin L_2$. It must therefore, be the case that (since $\omega \in L_1 \cup L_2$) $\omega \in L_1$.

Since ω was arbitrarily chosen from L_4^C it follows that $L_4^C \subseteq L_1$, as claimed.

Finishing Up This Part

Since $L_1 \subseteq L_4^C$ and $L_4^C \subseteq L_1$ it follows that $L_1 = L_4^C$, as claimed.

Proving that L_4 is Not a Regular Language

A ***proof by contradiction*** can be used to show that L_4 is not a regular language: Assume — to obtain a contradiction — that L_4 *is* a regular language.

Then it follows, by the closure of the set of regular languages under complementation — that L_4^C is also a regular language.

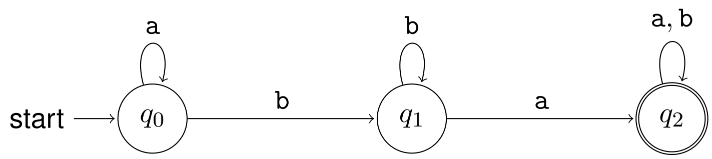
However, $L_4^C = L_1$ and it has already been proved that L_1 is *not* a regular language.

It follows, from this *contradiction*, that the only assumption made here must be false: L_4 is *not* a regular language.

Proving that L_3 is a Regular Language

There are many ways to prove that L_3 is a regular language. One of these ways is to give a deterministic finite automaton and prove that this has language L_3.

With that noted, let $M = (Q, \Sigma, \delta, q_0, F)$ where $Q = \{q_0, q_1, q_2\}$, $F = \{q_2\}$, and the transition function δ is as shown by the following diagram.



Consider the following subsets of $\Sigma^\star$

- $S_0 = \{\mathtt{a}^k \mid k \in \mathbb{Z}$ and $k \geq 0\}$.
- $S_1 = \{\mathtt{a}^k\mathtt{b}^\ell \mid k, \ell \in \mathbb{Z}$, $k \geq 0$, and $\ell \geq 1\}$.
- $S_3 = L_3$.

Exercise: Use the "Correctness of a DFA" Theorem, included in the preparatory material for Lecture #4, to prove that

$$\{\omega \in \Sigma^\star \mid \delta^\star(q_0, \omega) = q_i\} = S_i$$

for every integer i such that $0 \leq i \leq 2$. Then use this to prove that L_3 is the language of this deterministic finite automaton — so that L_3 is a regular language, as claimed.

Proving that L_2 is not a Regular Language

A ***proof by contradiction*** can be used to show that L_2 is not a regular language: Assume, to obtain a contradiction that L_2 is a regular language.

As shown above, L_3 is also a regular language, and it follows by closure of the set of regular languages under union that the language $L_2 \cup L_3 = L_4$ is also a regular language.

However, it has been shown above that L_4 is *not* a regular language: A *contradiction* has been obtained.

It follows, by this contradiction, that the only assumption that has been made here must be false: L_2 is *not* a regular language.