Lecture #8: Nonregular Languages, Part One

Proving That a Language is Not Regular — Another Example

Let $\Sigma = \{0, 1\}$. The ***reversal*** of a string

$$\omega = \sigma_1\sigma_2 \ldots \sigma_n$$

is the string

$$\omega^R = \sigma_n \ldots \sigma_2\sigma_1$$

obtained by listing the symbols in ω in reverse order. Thus $\lambda^R = \lambda$, $0^R = 0$, and $001^R = 100$. Note that if $\omega \in \Sigma^\star$ then $\omega^R \in \Sigma^\star$ as well. Furthermore ω and ω^R always have the same length. They also contain the number of 0's and the same number of 1's.

Suppose that $L_2 = \{\omega \cdot \omega^R \mid \omega \in \Sigma^\star\}$. Then L_2 includes $\lambda, 00, 001100$ and lots of other strings, but not 10, 0010, or any string with odd length.

Claim. *L_2 is not a regular language.*

Proof. Suppose, to obtain a contradiction, that L_2 is not a regular language.

Then it follows, by the Pumping Lemma for Regular Languages, that there is a number $p \geq 1$ such that if s is any string in L_2 with length at least p, then s can be divided into three pieces $s = xyz$ (for $x, y, z \in \Sigma^\star$), satisfying the following three conditions.

1. $xy^iz \in L_2$ for every integer i such that $i \geq 0$.
2. $|y| > 0$ (so that $y \neq \lambda$).
3. $|xy| \leq p$.

Consider the string $s = 0^p110^p$.

- $s \in L_2$ since $s = \omega \cdot \omega^R$ for the string $\omega = 0^p1 \in \Sigma^\star$.
- $|s| = 2p + 2 \geq p$.

It follows that $s = xyz$ for strings $x, y, z \in \Sigma^\star$ for strings $x, y, z \in \Sigma^\star$ that satisfy properties #1–#3, above.

- Since xy is a prefix of s with length at most p (by property #3), and the first p symbols of s are all copies of “0”, $xy = 0^k$ for some integer k such that $0 \leq k \leq p$.

  Since $s = 0^p110^p = xyz = 0^k z$, it follows that $z = 0^{p-k}110^p$.

- By property #2, $|y| > 0$, so that $y \neq \lambda$. Thus (since $|xy| = |x| + |y| = k$, $|x| = h$ and $|y| = \ell$ for integers h and ℓ such that $h \geq 0$, $\ell \geq 1$, and $h + \ell = k$.

  Since $xy = 0^k$ it now follows that $x = 0^h$ and $y = 0^\ell$.

- Let $i = 2$. Then

$$xy^iz = xy^2z = xyyz = 0^h0^\ell0^\ell0^{p-k}110^p = 0^{p+\ell}110^k, \tag{1}$$

  since $h + \ell = k$ — and it follows, by property #1, above, that this string is in the language L_2.

Now, if ℓ is odd then it is impossible for xy^2z to belong to L_2, because the length $(2p+2+\ell)$ of this string is odd and, as noted above, L_2 only contains strings with even length. The integer ℓ must, therefore, be even. Since $\ell \geq 1$, it now follows that $\ell \geq 2$.

Furthermore, one can see by the equation at line (1) that the string xy^2z has length $2p+\ell+2 = 2(p + \ell/2 + 1)$. Since $xy^2z \in L_2$, it must be the case that

$$xy^22 = \omega \cdot \omega^R \tag{2}$$

for some string $\omega \in \Sigma^\star$. Since ω and its reversal have the same length, the length of ω must be one-half the length of xy^2z, that is, $p + \ell/2 + 1$. Thus ω is the ***prefix*** of xy^2z with this length.

Now, since $\ell \geq 2$, $1 \leq \ell/2$, and the length of ω is $p + \ell/2 + 1 \leq p + \ell/2 + \ell/2 = p + \ell$. Since ω is a prefix of xy^2z and (again, by the equation at line (1)) the first $p + \ell$ symbols in sy^2z are all copies of “0”, it must be the case that $\omega = 0^{p+\ell/2+1}$.

However, ω^R must be the string $0^{p+\ell/2+1}$ as well, so that

$$\omega \cdot \omega^R = 0^{p+\ell/2+1} \cdot 0^{p+\ell/2+1} = 0^{2(p+\ell/2+1)} = 0^{2p+\ell+2} \neq 0^p110^p = xy^2z.$$

This ***contradicts*** the equation at line (2), above, and it follows, by this contradiction, that the assumption in this proof must be false. Thus L_2 is not a regular language, as claimed. $\square$