Pumping Lemmas

The main usefulness of the two pumping lemmas is to prove that a particular language is not regular, or context-free, as the case may be. Each lemma states that every language in the class has a certain property, and thus if we can prove that a given language L does not have that property, L is not in the class.

If w is a string and a is a symbol, we write $\#_a(w)$ to be the number of instances of the symbol a in the string w.

Lemma 1 (Pumping Lemma for Regular Languages) *If L is a regular language, there exists a positive integer p, called the pumping length of L, such that for any string $w \in L$ whose length is at least p, there exist strings x, y, z such that the following conditions hold.*

1. $w = xyz$
2. $|y| \geq 1$
3. $|xy| \leq p$
4. *for any* $i \geq 0$, $xy^iz \in L$.

Note that the the value of p is not unique: if p is a pumping length of L, so is every integer larger than p. There is a minimum pumping length.

Example

Let L be the language of all base 2 numerals for multiples of 5, where leading zeros are not allowed. The minimum pumping length is 5. We won't prove that, but for example, if $w = 11001$, which means 25, we let $x = 1$, $y = 10$, and $z = 01$. The first three conditions obviously hold. If we let $i = 0$, we get $xz = 101$, which means 5, while if $i = 2$ or $i = 3$, we get $xy^2z = 1101001$ which means 105, or $xy^3z = 110101001$ which means 425. The pumping length cannot be 4, since 1111, which means 15, does not have a pumpable substring. Thus, 5 is minimum.

Another example is $w = 1110011$, which means 115. Let $x = 11$, $y = 100$, and $z = 11$.

The Pumping Lemma and Finite Automata

Can you prove the following lemma?

Remark 1 *A regular language accepted by an NFA with n states has pumping length n.*

Lemma 1 follows from Remark 1. Do you see why?

Lemma 2 (Pumping Lemma for Context-Free Languages) *If L is a context-free language, there exists a positive integer p, called the pumping length of L, such that for any string $w \in L$ whose length is at least p, there exist strings u, v, x, y, z such that the following conditions hold.*

1. $w = uvxyz$
2. $|v| + |y| \geq 1$
3. $|vxy| \leq p$
4. *for any* $i \geq 0$, $uv^ixy^iz \in L$.

Note that the the value of p is not unique: if p is a pumping length of L, so is every integer larger than p. There is a minimum pumping length.

Example

Let L be the language consisting of all palindromes over $\{a, b\}$. The following is an unambiguous grammar for L.

$S- > aSa|bSb|a|b|\lambda$

What is the minimum pumping length of L?

The answer is 3. If a palindrome w has even length, the substring aa or bb in the middle of the string. That is, $w = uaau^R$ or $w = ubbu^R$. Suppose $w = uaau^R$. We let $u = u$, $v = a$, $x = \lambda$, $y = a$, and $z = u^{\mathrm{R}}$. The first three conditions are obviously satisfied. For any $i \geq 0$, $uv^ixy^iz = ua^ia^iu^R \in L$. The case that $w = ubbu^R$ is similar.

If w has odd length, then there are four possibilities:
$w = uaaau^R$
$w = uabau^R$
$w = ubabu^R$
$w = ubbbu^R$
In the first case, we let $u = u$, $x = a$, $y = a$, and $z = u^R$. In the second case, we let $u = u$, $x = a$, $y = b$, and $z = u^R$. The four conditions are satisfied. The other two cases are similar.

The minimum pumping length cannot be 2, because $w = aba \in L$, and the four conditions cannot be fulfilled for w with $p = 2$.

Using the Pumping Lemmas

Lemma 1 states a property that all regular languages have. Hence, if a language fails to satisfy that property, it is not regular. Similarly, if language fails to satisfy the property given by Lemma 2, it is not context-free. We use the lemmas to show a language is not regular and to show another language to be not context-free.

Let $L_1 = \{a^n b^n \,:\, n \geq 0\}$, and let $L_2 = \{a^n b^n c^n \,:\, n \geq 0\}$.

Theorem 1 *L_1 is not regular.*

Proof: By contradiction. We assume L_1 is regular. Let p be a pumping length of L. (We usually say, *the* pumping length, despite the fact that it is not unique.) Let $w = a^p b^p$. Note that $|w| = 2p \geq p$, hence there exist strings x, y, z such that
1. $w = xyz$
2. $|xy| \leq p$
3. $|y| > 0$
4. For any $i \geq 0$, $xy^i z \in L_1$.

By 1. and 2., xy is a prefix of w of length no greater than p. Since the first p symbols of w are a's, that implies xy is a string of a's, hence y is also a string of a's. Write $y = a^j$. By 3., $j > 0$. Let $i = 0$. By 4., $xy^0 z = xz \in L_1$. But $xz = a^{p-j} b^p \notin L_1$ since $\#_a(xy \neq \#_b(xy)$, contradiction. ∎

Theorem 2 *L_2 is not context-free.*

Proof: By contradiction. We assume L_2 is context-free. Let p be the pumping length of L. Let $w = a^p b^p c^p$. Note that $|w| = 3p \geq p$, hence there exist strings x, y, z, u, v such that
1. $w = uvxyz$
2. $|vxy| \leq p$
3. $|v| + |y| > 0$
4. For any $i \geq 0$, $uv^i xy^i z \in L_2$.

Let $j = |v|$ and $k = |y|$. By 3., $j + k > 0$. Let $i = 0$. By 4., Then $uxz = uv^0 xy^0 z \in L_2$.

By 1., $|u| + |vxy| + |z| = |w| = 3p$. Subtracting $|vxy|$, which is less than or equal to than p, by 2., We have $|u| + |z| \geq 2p$, hence those two numbers cannot both be less than p. That is, either $|u| \geq p$ or $|z| \geq p$.

We have two cases: $|u| \geq p$, and $|z| \geq p$. Suppose $|u| \geq p$. Note that $\#_a(u) + \#_a(vxy) + \#_a(z) = \#_a(w) = p$. Since u is a prefix of $w = a^p b^p c^p$ of length it least p, it must contain all of a^p, that is, $\#_a(u) = p$. Thus $\#_a(uxz) = p$, and $uxz \in L_2$, hence $|uxz| = 3p$. But $|uxz| = 3p - j - k < 3p$, contradiction.

The case $|z| \geq p$ is similar. ∎