# Proving languages not to be regular

## 1 The pumping lemma

If $L$ is a regular language, then there is a number $p > 0$ (sometimes called the pumping length) such that if $w \in L$ and $|w| \geq p$, then $w$ can be written as $xyz$ where

1. $|xy| \leq p$
2. $y \neq \varepsilon$
3. $xy^kz \in L \quad \forall k \geq 0$

Pumping Lemma can be useful to prove that some languages are not regular.

### Example 1a:

$\Sigma = \{0, 1\}$
$L = \{0^n1^n \mid n \geq 0\}$

**Solution A:**
We will use an indirect proof to show that L is not a regular language. Suppose that L is regular, which means that any string $w \in L$ of length not smaller than $p$ should satisfy the above conditions. Let

$$w = 0^p1^p \in L$$

It can be seen that $|w| = 2p$, which satisfies the condition $|w| \geq p$. If we divide $w$ into parts $x, y$ and $z$, we know that $xy$ contains only 0 symbols; $|xy| \leq p$ states that they are at most $p$ symbol long, and our first $p$ symbols are 0s in the string.
Because $y \neq \varepsilon$, it has to contain at least one 0 symbol. Pumping $y$ ($xy^kz$) for any $k \neq 1$ would result in a new string where the number of 0s is not equal to the number of 1s. (for example, $w_2 = 0^{p-|y|}1^p$ for $k = 0$, $w_3 = 0^{p+|y|}1^p$ for $k = 2$, $w_4 = 0^{p+2|y|}1^p$ for $k = 3$, etc.). As $w_2 \notin L$, we came to a contradiction, meaning that our original assumption is not true; L is not regular.
**Solution B:**
We will again use an indirect proof to show that L is not a regular language. Suppose that L is regular, which means that any string $w \in L$ of length not smaller than $p$ should satisfy the above conditions. Let

$$w = 0^{\lceil p/2 \rceil}1^{\lceil p/2 \rceil} \in L$$

It can be seen that $|w| = p$ if $p$ is even, while $|w| = p + 2$ if $p$ is odd. Both of these satisfy the condition $|w| \geq p$. If we divide $w$ into parts $x, y$ and $z$, we will not get additional information about the composition of $x$ and $y$ as we did in Solution A. The length of our string is $p$ or $p+2$, and while $xy$ will surely start with a 0, it can end with a 1 because of $|xy| \leq p$. Moreover, $y$ is not guaranteed to start with a 0 because of this, it can happen that the leading 0s belong to $x$. This leads to the following cases, where we all have to show the contradiction:

1. $xy$ contains only 0s: this case is identical to the proof of Solution A, we will come to a contradiction the same way.

2. $xy$ contains both 0s and 1s: This leads to two further subcases:

   (a) $x$ contains all the 0s: Similarly to Solution A, we can state that $y$ has to contain at least one 1 symbol because of $y \neq \varepsilon$. Pumping $y$ ($xy^kz$) for any $k \neq 1$ would result in a new string where the number of 1s is not equal to the number of 0s. We can come to a contradiction in a similar way to Solution A.

   (b) $y$ contains both 0s and 1s: Let $y = 0^i1^j, i, j \geq 1$. Pumping $y$ ($xy^kz$) for any $k > 1$ would result in the string $0^{\lceil p/2 \rceil - i}(0^i1^j)^k1^{\lceil p/2 \rceil - j}$. This for example in the case of $k = 2$ will give $w_2 = 0^{\lceil p/2 \rceil - i}(0^i1^j)^21^{\lceil p/2 \rceil - j} = 0^{\lceil p/2 \rceil - i}01011^{\lceil p/2 \rceil - j} = 0^{\lceil p/2 \rceil - i + 1}101^{\lceil p/2 \rceil - j + 1}$, and clearly $w_2 \notin L$ (even if $i = j$), so we came to a contradiction.

As we came to a contradiction in every possible case, our original assumption is not true; $L$ is not regular.

## Example 1b:

$\Sigma = \{0, 1\}$
$L = \{0^n1^n \mid 0 \leq n \leq 1234567890\}$
**Solution:** $L$ is a finite language, thus it also has to be regular.

## Example 1c:

$\Sigma = \{0, 1\}$
$L = \{0^n10^n \mid n \geq 0\}$

## Example 2:

$\Sigma = \{0, 1\}$
$L = \{vv | v = (0|1)^*\}$

**Solution:**
The language contains strings that consist of the same substring concatenated twice. We will use an indirect proof to show that L is not a regular language. Suppose that L is regular, which means that any string $w \in L$ of length not smaller than $p$ should satisfy the above conditions. Let

$$v = 0^p1,$$

so consequently

$$w = 0^p10^p1 \in L$$

It can be seen that $|w| = 2p + 2$, which satisfies the condition $|w| \geq p$. If we divide $w$ into parts $x, y$ and $z$, we know that $xy$ contains only 0 symbols; $|xy| \leq p$ states that they are at most $p$ symbol long, and our first $p$ symbols are 0s in the string. From here, showing the contradiction is similar to Example 1 Solution A.

As a sidenote, if you choose $v = 0^p$, and $w = 0^p0^p$, finding the contradiction would again involve examining multiple cases. This is because your string is $w = 0^{2p}$, and all such strings are in $L$ (even length, containing a single symbol). You have to examine all combinations of $y$ having an even/odd length and $k$ being even/odd, and finding a contradiction this way. You will not be able to find a contradiction in cases where $|y|$ or $k$ is even, as these will result in a $w_2$ of even length.

### Example 3:

$\Sigma = \{1, \#\}$
$L = \{1^i \# 1^j \# 1^{i+j} \mid i \geq 0\}$

### Example 4:

$\Sigma = \{0, 1\}$
$L = \{v \mid |v| \text{ is prime}\}$

**Solution:**
The language contains strings whose length is a prime number. The symbol composition of the strings is irrelevant. We will use an indirect proof to show that L is not a regular language. Suppose that L is regular, which means that any string $w \in L$ of length not smaller than $p$ should satisfy the above conditions. Let $m > p$ be a prime ($m$ has to exist for any $p$, as we have an infinite number of primes). Let

$$w = 0^m \in L$$

By definition, $|w| = m$ will satisfy $|w| \geq p$. For any division of $w$ into parts $x, y$ and $z$ and $|xy| \leq p$, we know that $y = 0^i$ because of $y \neq \varepsilon$. We would like to pump $y$ by choosing an appropriate $k$ where we can show that the length of the resulting string is not a prime. Let $k = m + 1$. This will give us $w_2 = xy^{m+1}z = 0^{m+m \cdot i} = 0^{m(i+1)}$ We see that the length of $w_2$ can be factorized. If none of the factors are 1, then this cannot be a prime number. By definition, $m > p > 0$, and $i = |y| > 0$, so $m \geq 2$ and $(i+1) \geq 2$, meaning that the length of $w_2$ is not a prime. As $w_2 \notin L$, we came to a contradiction, meaning that our original assumption is not true; L is not regular.