

Lecture #5: Nonregular Languages

Key Concepts

Part One: Using the Pumping Lemma

Theorem 1 (Pumping Lemma). *Let Σ be an alphabet and let $A \subseteq \Sigma^*$.*

*If A is a regular language, then there is a number $p \geq 1$ (called the **pumping length** for A) — which only depends on A — such that if s is any string in A with length at least p , then s can be divided into three pieces $s = xyz$ (for $x, y, z \in \Sigma^*$), satisfying the following three conditions.*

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

Note: y^i is the concatenation of i copies of the string y .

A Process to Prove that a Language $L \subseteq \Sigma^*$ is not Regular:

1. Assume — to obtain a contradiction — that L is a regular language.
2. Observe that the conditions in the Pumping Lemma (with L used as the language called “ A ”) are satisfied.
3. Introduce the “pumping length” — which is generally named p . Note that **you do not get to choose this value** — you can only assume that it exists (and is a positive integer).
4. Introduce a string in Σ — which is usually named s , because that is what is called in the statement of the Pumping Lemma.

You do get to choose this string — and you will need to think about the language, L , that you are working with — and its definition will almost always depend on the (unknown) pumping length p .

5. Show that $s \in L$ and $|s| \geq p$.
6. Conclude that there exist strings $x, y, z \in \Sigma^*$ such that $s = xyz$ and properties #1, #2 and #3, given in the statement of the Pumping Lemma (again, with L used in place of A) all hold.
 Note that **you do not get to choose** the strings x , y , and z — you may only assume that these strings exist.
7. Use properties #1, #2 and #3 to get a contradiction **without making any other assumptions**.
8. Conclude that — since only one assumption was made — this assumption must be false. In other words, L is *not* a regular language.

More About This Process

- When carrying out this process, you *do* get to choose the value for the integer i , mentioned in property (a). It is frequently — but not always — possible to choose i to be either 0 or 2, when you are using this process. However, it is sometimes necessary to think more about the language L that is being considered, and to choose i to be something else.
- Note how the details in this process (including the order in which things are introduced and which values you get to choose, along with which value you *do not* get to choose) reflect the structure of the Pumping Lemma.

In the lecture notes, this process was used to show that if $\Sigma = \{a, b\}$ and

$$L = \{a^n b^n \mid n \in \mathbb{Z} \text{ and } n \geq 0\}$$

then L is *not* a regular language.

Part Two: Using Closure Properties

The Big Idea

If P and Q are Boolean statements (that is, statements that are either *true* or *false*) and

$$P \implies Q$$

then

$$\neg Q \implies \neg P$$

as well. Thus the **closure properties** that were introduced in the lecture slides on “Regular Operations and Regular Expressions” (and closure properties like them) — that were used to show that if *one* language was **regular**, then *another* was regular too — can also be used that if *one* language is **not regular**, then *another* language is not regular, either.

Closure Properties, How They Were Used Before, and How They Can Be Used Now

Closure Under Union

Theorem 2. *Suppose that $L_1, L_2 \subseteq \Sigma^*$ for an alphabet Σ . If L_1 and L_2 are both regular languages, then $L_1 \cup L_2$ is also a regular language.*

How This Was Used Before: Show that some language $L \subseteq \Sigma^*$ is regular, by showing that $L = L_1 \cup L_2$, where $L_1, L_2 \subseteq \Sigma^*$ are regular languages (and this has already been proved, or is being proved now).

How This Is Also Being Used Now: Show that a language $L \subseteq \Sigma^*$ is *not* regular by using the fact that the language \hat{L} is not regular, where $\hat{L} = L \cup \tilde{L}$, for some other language $\tilde{L} \subseteq \Sigma^*$ that *is* a regular language. (If it has not already been shown that \hat{L} is not a regular language then you must prove this too. Similarly, if \tilde{L} has not already been proved to *be* a regular language then you must do this as well.)

This can be used as the argument in a **proof by contradiction** — in which you start by assuming that L is a regular language, and obtain a contradiction without making any other assumptions.

Closure Under Concatenation

Theorem 3. *Suppose that $L_1, L_2 \subseteq \Sigma^*$ for an alphabet Σ . If L_1 and L_2 are both regular languages, then $L_1 \circ L_2$ is also a regular language.*

How This Was Used Before: Show that some language $L \subseteq \Sigma^*$ is regular, by showing that $L = L_1 \circ L_2$, where $L_1, L_2 \subseteq \Sigma^*$ are regular languages (and this has already been proved, or is being proved now).

How This Is Also Being Used Now: Show that a language $L \subseteq \Sigma^*$ is *not* regular by using the fact that the language \hat{L} is not regular, where *either* $\hat{L} = L \circ \tilde{L}$ or $\hat{L} = \tilde{L} \circ L$, for some other

language $\tilde{L} \subseteq \Sigma^*$ that *is* a regular language. (If it has not already been shown that \hat{L} is not a regular language then you must prove this too. Similarly, if \tilde{L} has not already been proved to be a regular language then you must do this as well.)

Once again the easiest way to use this in a proof, that is easy for someone else to read and understand, is probably to write a **proof by contradiction** in which you establish a contradiction after assuming that L is a regular language (without assuming anything else).

Closure Under Kleene Star

Theorem 4. *Suppose that $L \subseteq \Sigma^*$ for an alphabet Σ . If L is a regular language, then L^* is also a regular language.*

How This Was Used Before: Show that some language $L \subseteq \Sigma^*$ is regular, by showing that $L = \hat{L}^*$, where $\hat{L} \subseteq \Sigma^*$ is a regular language (and either this has already been proved, or this is being proved now).

How This is Also Being Used Now: Show that a language $L \subseteq \Sigma^*$ is *not* regular by using the fact that the language $\hat{L} \subseteq \Sigma^*$ is not regular, where $\hat{L} = L^*$ (and when either it has already been proved that \hat{L} is not a regular language, or this is also being proved now).

Again, it is probably easiest to use this, in a proof that someone else can read and understand, by writing a **proof by contradiction** that uses the assumption that L is a regular language.

A New Closure Property: Closure Under Complementation

Recall that if $L \subseteq \Sigma^*$, for an alphabet Σ , then the **complement** of L is the language¹

$$L^C = \{\omega \in \Sigma^* \mid \omega \notin L\}.$$

Note that $(L^C)^C = L$, for every language $L \subseteq \Sigma^*$.

Theorem 5. *Suppose that $L \subseteq \Sigma^*$ for an alphabet Σ . If L is a regular language, then L^C is also a regular language.*

How This Could Have Been Used, Before: Show that some language $L \subseteq \Sigma^*$ is regular, by showing that $L = \hat{L}^C$, where $\hat{L} \subseteq \Sigma^*$ is a regular language (and either this has already been proved, or this is being proved now).

How This is Also Being Used Now: Show that a language $L \subseteq \Sigma^*$ is *not* regular by using the fact that the language $\hat{L} \subseteq \Sigma^*$ is not regular, where $\hat{L} = L^C$ (and either it has already been proved that \hat{L} is not a regular language, or this is also being proved now).

¹This language is sometimes represented as \bar{L} , instead.

Once again, it is probably easiest to use this, in a proof that someone else can read and understand, by writing a ***proof by contradiction*** that uses the assumption that L is a regular language.

More To Note

If *additional* closure properties for the set of regular languages are discovered (and proved to be correct) then this will automatically yield new ways, both prove that some languages *are* regular, and also to prove that (some other) languages are *not* regular.

This approach can also be applied to other interesting sets of languages besides the set of *regular* languages. Indeed, this will be a significant part of the middle of this course.