Non-regular Languages

Outline

Organisational

Next week is different:

- Test on part I (lectures 1-3)
- Make sure you are registered for the course in Osiris
- Time: 8:30 – 10:30, make sure you are there at 8:15.
- Locations:
  - LIN2 and LIN3
  - Extra time: HG00.065, HG00.616

Equivalence of definitions

Theorem

Let $L \subseteq \Sigma^*$. Then the following are equivalent

1. $L = \mathcal{L}(M)$ *for some DFA* M *(or NFA, NFA-λ)*
2. L *is* ***regular****, i.e.* $L = \mathcal{L}(e)$ *for some regular expression* e*.*

So:

- To show that a language is regular we can give a regular expression or a (non-)deterministic automaton (with λ-steps).
- To show closure properties of the class of regular languages, we can use regular expressions, deterministic automata, non-deterministic automata, ...

Closure properties of the class of regular languages

If L, L_1 and L_2 over Σ are regular then so are

- $\overline{L}$ (NB. $\overline{L} = \{w \in \Sigma^* \mid w \notin L\}$)
- L^*
- $L_1 \cup L_2$
- $L_1 \cap L_2$
- $L_1 L_2$
- L^R (NB. $L^R = \{w \in \Sigma^* \mid w^R \in L\}$)
- Prefix(L)
  NB. $\text{Prefix}(L) := \{w \in \Sigma^* \mid \exists v \in L(w \text{ is a prefix of } v)\}$
  w is a prefix of v if $v = wu$ for some $u \in \Sigma^*$.

A famous language

Is the language

$$L := \{a^n b^n \in \Sigma^* \mid n \geq 0\}$$

regular?

No.

How to prove this? Showing that there is no regular expression that describes L? Showing that there is no DFA (NFA, NFA-λ) accepting L?

Proof.

Suppose the DFA $M = (Q, q_0, \delta, F)$ accepts L.

Then for all $n, m \in \mathbb{N}$, if $n \neq m$, then $\delta^*(q_0, a^n) \neq \delta^*(q_0, a^m)$.

[Why?

Because, if $\delta^*(q_0, a^n) = \delta^*(q_0, a^m) = q$, then $\delta^*(q, b^m) \in F$, but then $a^n b^m$ is also accepted, while it shouldn't be.]

So M must have infinitely many states, which is not the case.

So there is no DFA accepting L, so L is not regular. □

Non regular languages

Let $\Sigma = \{a, b\}$. We will develop a general technique that can be used to show that languages are not regular.

This technique will be applied to show (in a different way) that

$$\{a^n b^n \in \Sigma^* \mid n \geq 0\}$$

is not regular
and to show that

$$\{w \in \Sigma^* \mid w \text{ is a palindrome}\}$$

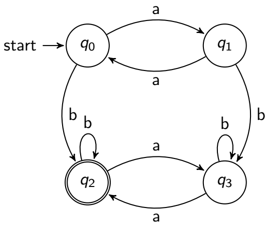
is not regular.
A *palindrome* is a word w such that $w^R = w$.
Remember that w^R is the *reverse of* w, defined by

$$\begin{aligned} \lambda^R &:= \lambda \\ (s\,w)^R &:= w^R\,s \end{aligned}$$

A general method to show that a language is *not* regular

Regular languages can be **pumped**!

Example: Consider $\Sigma = \{a, b\}$ and the automaton



What happens if a word of length $4, 5, 6, 7, \ldots$ is accepted?
It has made a cycle which can be repeated arbitrarily often!
For example, *baaaa* is accepted, and also all $baa(aa)^n$ are accepted.
We say that *aa* is a substring that can be pumped.

Pumping Lemma for Regular Languages

Lemma (Pumping Lemma.)
Let $L \subseteq \Sigma^*$ be a regular language. Then there exists a number $k \geq 1$ (pumping number) such that for every $w \in L$ with $|w| \geq k$:

1. *w can be split in three parts, $w = uvz$,*
2. *with $|uv| \leq k$ and $|v| \geq 1$,*
3. *such that for all $n \geq 0$ one has $uv^n z \in L$.*

Corollary
The language $L = \{a^n b^n \mid n \geq 0\}$ is not regular

Proof.
Suppose L is regular. (Towards a contradiction.) Let $k \geq 1$ be as in the Pumping Lemma. We take $w = a^k b^k$. Then $w \in L$ and $|w| \geq k$. Therefore there are u, v, z such that $a^k b^k = uvz$, with $|uv| \leq k$, $|v| \geq 1$ and $uv^n z \in L$ for all $n \geq 0$. Then $v = a^q$, for some $q \geq 1$. We take $n = 2$ and conclude that $uv^2 z = a^{k+q} b^k \in L \ldots$. But this wrong: $a^{k+q} b^k \notin L$! Contradiction. So L is not regular. □

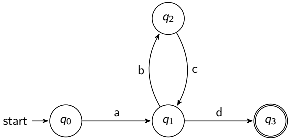
Pumping Lemma: Example

Lemma (Pumping Lemma.)
Let $L \subseteq \Sigma^*$ be a regular language. Then there exists a number $k \geq 1$ (pumping number) such that for every $w \in L$ with $|w| \geq k$:

1. *w can be split in three parts, $w = uvz$,*
2. *with $|uv| \leq k$ and $|v| \geq 1$,*
3. *such that for all $n \geq 0$ one has $uv^nz \in L$.*

Example

What can we choose for k in the automaton below?



What are u, v, z for an accepted word w?

Proof of the Pumping Lemma

Lemma (Pumping Lemma.)
Let $L \subseteq \Sigma^*$ be a regular language. Then there exists a number $k \geq 1$ (pumping number) such that for every $w \in L$ with $|w| \geq k$:

1. *w can be split in three parts, $w = uvz$,*
2. *with $|uv| \leq k$ and $|v| \geq 1$,*
3. *such that for all $n \geq 0$ one has $uv^n z \in L$.*

Proof.
Let L be regular. Let M be a DFA that accepts L. Take k to be the number of states of M.
Let $w \in L$ with $|w| \geq k$. Then, reading word w, we must pass some state more than once.
Say that q is the first state that we pass twice (reading w).
Then $w = uvz$, where we read u to go to q, read v to loop at q, read z to go to a final node. Note that indeed $|uv| \leq k$ and $|v| \geq 1$.
Then $uv^n z$ is accepted for all n. □

Pigeonhole principle

source: https://nl.wikipedia.org/wiki/Duiventilprincipe

Applying the Pumping Lemma

To show that L is not regular we try to follow a recipe:

1. Assume that L is regular.
2. By the Pumping Lemma, there is a 'pumping number' k.
3. **Give a word w such that $|w| \geq k$.**
4. The Pumping Lemma gives u, v, z such that
   a. $w = uvz$, $|uv| \leq k$ and $|v| \geq 1$, and
   b. for all $n \in \mathbb{N}$: $uv^n z \in L$.
5. **Try to show $uv^n z \notin L$ for some n.**
6. Conclude that there is a contradiction: hence L can not be regular.

PL application

Claim:

$$L = \{w \in \Sigma^* \mid w \text{ is a palindrome}\}$$

is not regular.

Proof. We follow the procedure above.

Let $k \geq 1$ (arbitrary)

Take $w = a^k b a^k$. Then $w \in L$ (**check**) and $|w| \geq k$ (**check**)

Let u, v, z (arbitrary) be so that $a^k b a^k = uvz$, with $|uv| \leq k$ and $|v| \geq 1$.

(Say $|v| = p$, so $p \geq 1$.)

Take $n = 0$. Then $uv^n z = uv^0 z = a^{k-p} b a^k \notin L$ (**check**).

So, L is not regular.

Some other non-regularity results

Let $\Sigma := \{a, b\}$. We know that $L = \{a^n b^n \mid n \geq 0\}$ is not regular.

Is $L' := \{w \in \Sigma^* \mid \forall n \in \mathbb{N}\,(w \neq a^n b^n)\}$ regular?

Answer: No it is not. If L' is regular, then $\overline{L'}$ would also be regular, but this is just L and L is not regular! So L' is not regular.

Lemma
If L is not *regular, then also $\overline{L}$ and L^R are not regular.*

Let $\Sigma := \{a, b, c\}$.
Is $L'' := \{a^n c^p b^n \in \Sigma^* \mid n \geq 0, p \geq 0\}$ regular?

Answer: No it is not. $L = L'' \cap \mathcal{L}(a^* b^*)$. If L'' is regular, then L would be regular as well, but it is not!

Lemma
If L is not *regular and $L = L_1 \cap L_2$, with L_1 regular, then L_2 is not regular.*

In proving non-regularity of a language L:

You may use

- The Pumping Lemma: Assume that L is regular, so it satisfies the Pumping Lemma, and derive a contradiction.
- Closure properties of the class of regular languages, with for instance
  - The fact that $\{a^n b^n \in \Sigma^* \mid n \geq 0\}$ is not regular.
  - The fact that $\{w \in \Sigma^* \mid w \text{ is a palindrome}\}$ is not regular.