# CS420

## Introduction to the Theory of Computation

### Lecture 11: Pumping Lemma for Context-Free Languages

Tiago Cogumbreiro

# Today we will learn...

- The Pumping Lemma for Context-Free Languages
- Using the Pumping Lemma to identify non-context-free languages

> Section 2.3 Non-Context-Free Languages
> Supplementary material:
>
> - Professor Harry Porter's video *

# Exercise 1

# Exercise 1

$$L_1 = \{w \mid w \in \{a,b\}^\star \land |w| \text{ is divisible by } 3\}$$

- (i) Regular? Give a REGEX/NFA/DFA
- (ii) Context-free (and not regular)? Give a CFG/PDA. Prove using the pumping lemma.
- (ii) Not context-free

# Exercise 1

$L_1 = \{w \mid w \in \{a,b\}^\star \wedge |w| \text{ is divisible by } 3\}$

- (i) Regular? Give a REGEX/NFA/DFA
- (ii) Context-free (and not regular)? Give a CFG/PDA. Prove using the pumping lemma.
- (ii) Not context-free

**(i) Regular:** $\big((a+b)(a+b)(a+b)\big)^\star$

# Exercise 2

# Exercise 2

$L_2 = \{z \mid z \text{ has the same number of a's and b's}\}$

- (i) Regular? Give a REGEX/NFA/DFA
- (ii) Context-free (and not regular)? Give a CFG/PDA. Prove using the pumping lemma.
- (ii) Not context-free

# Exercise 2

$$L_2 = \{z \mid z \text{ has the same number of a's and b's}\}$$

- (i) Regular? Give a REGEX/NFA/DFA
- (ii) Context-free (and not regular)? Give a CFG/PDA. Prove using the pumping lemma.
- (ii) Not context-free

**(ii) Context-free:**

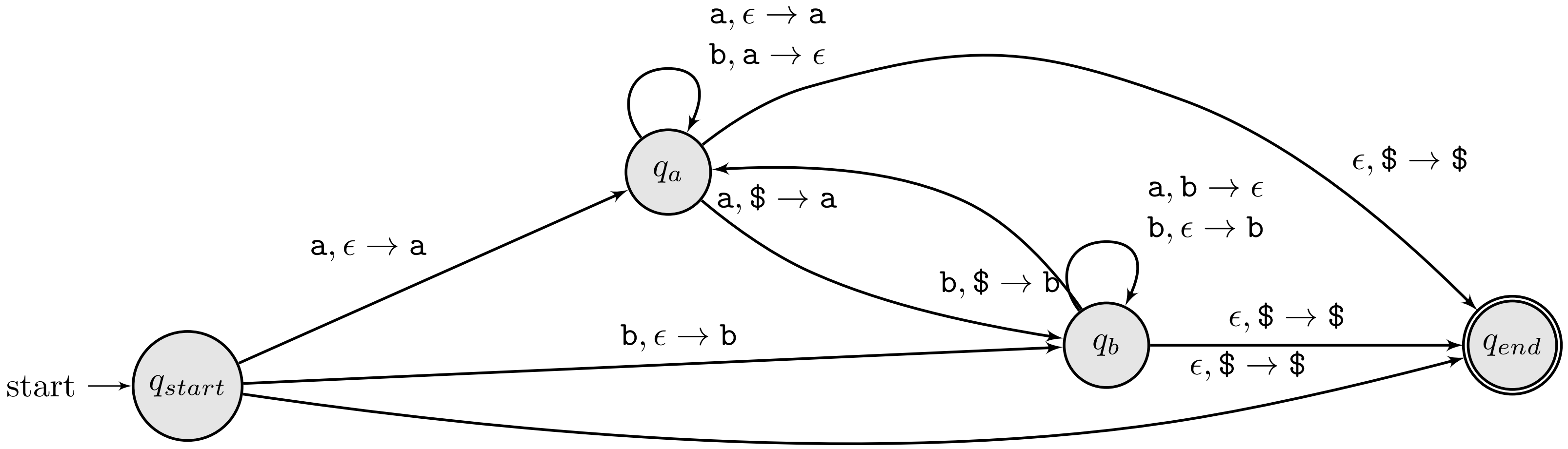
$S \to aB \mid bA \mid Xa \mid Yb \mid \epsilon$

$A \to Sb$

$B \to Sa$

$X \to bS$

$Y \to aS$



# Exercise 3

# Exercise 3

$$L_3 = \{a^n b^n c^n \mid n \geq 0\}$$

- (i) Regular? Give a REGEX/NFA/DFA
- (ii) Context-free (and not regular)? Give a CFG/PDA. Prove using the pumping lemma.
- (ii) Not context-free

# Exercise 3

$L_3 = \{a^n b^n c^n \mid n \geq 0\}$

- (i) Regular? Give a REGEX/NFA/DFA
- (ii) Context-free (and not regular)? Give a CFG/PDA. Prove using the pumping lemma.
- (ii) Not context-free

Not context-free

How do we prove that a language is **not** context free?

# The Pumping Lemma for CFL

# Intuition

If we have a string that is long enough, then we will need to repeat a non variable, say $R$, in the parse tree.

## Example

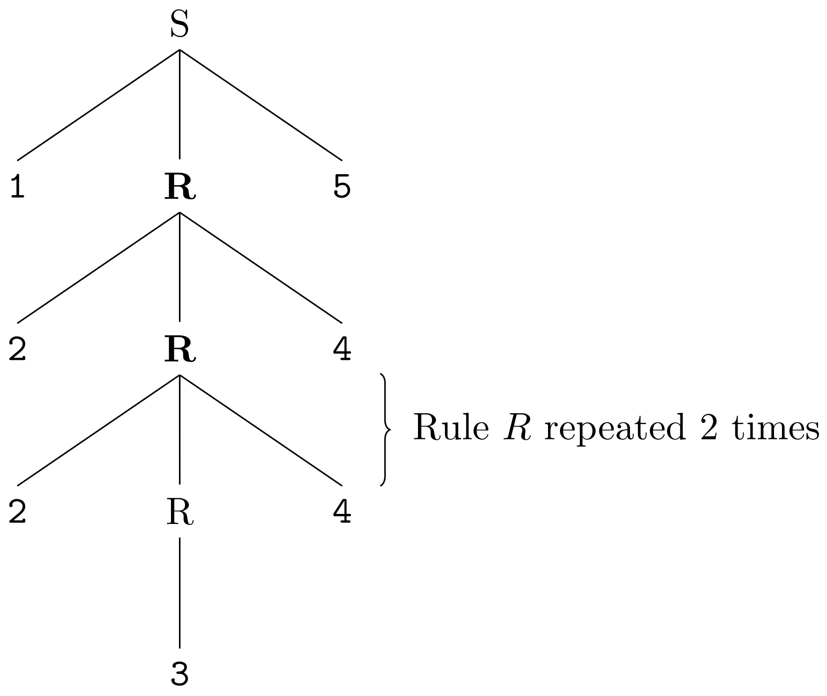
$$S \to 1R5$$

$$R \to 2R4 \mid 3$$

If we vary the number of times $R \to 2R4$ appears we note that:

- 1223445 is accepted (repeat 2 $\times$)
- 135 is accepted (repeat 0 $\times$)
- 12345 is accepted (repeat 1 $\times$)
- 122234445 is accepted (repeat 3 $\times$)

## Parse tree for 1223445



# Example

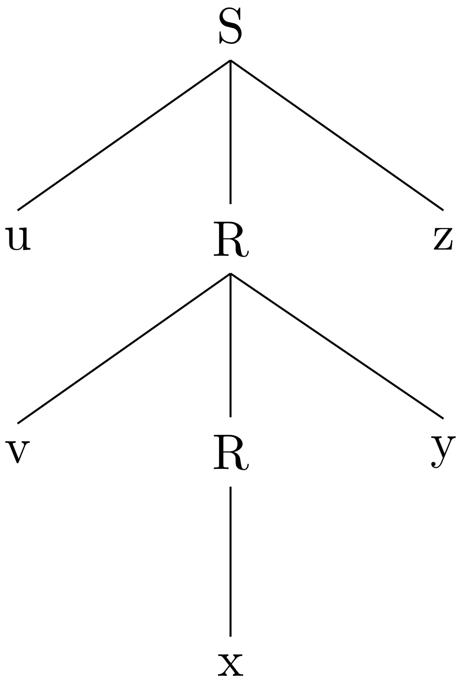
$$\boxed{\begin{aligned} S &\to 1R5 \\ R &\to 2R4 \mid 3 \end{aligned}}$$

- $\underbrace{1}_{u}\underbrace{22}_{v^2}\underbrace{3}_{x}\underbrace{44}_{y^2}\underbrace{5}_{z}$, where $i = 2$
- $\underbrace{1}_{u}\underbrace{3}_{x}\underbrace{5}_{z}$, where $i = 0$
- $\underbrace{1}_{u}\underbrace{2}_{v^1}\underbrace{3}_{x}\underbrace{4}_{y^1}\underbrace{5}_{z}$, where $i = 2$
- $\underbrace{1}_{u}\underbrace{222}_{v^3}\underbrace{3}_{x}\underbrace{444}_{y^3}\underbrace{5}_{z}$, where $i = 3$

Thus, $uv^ixy^iz$ is also in the language

# Generalizing

For a long enough string, say $uvxyz$ in the language, then $uv^i xy^i z$ is also in the language.



# Pumping Lemma for context-free languages

> The pumping lemma tells us that all regular languages (that have a loop) can be partitioned:

Every word in a context-free language, $w \in L$, can be partitioned into 5 parts $w = uvxyz$:

- an outer portion $u$ and $z$
- a repeating portion $v$ and $y$
- a non-repeating center portion $x$

Additionally, since $v$ and $y$ are a repeating portion, then $v$ and $y$ may be omitted or replicated as many times as we want and that word will also be in the given language, that is $uv^i xy^i z \in L$.

# Example

$L_2 = \{z \mid z \text{ same number of a's and b's}\}$

**You:** Give me a string of size 4.

# Example

$L_2 = \{z \mid z \text{ same number of a's and b's}\}$

**You:** Give me a string of size 4.

**Example:** abab

# Example

$L_2 = \{z \mid z \text{ same number of a's and b's}\}$

**You:** Give me a string of size 4.

**Example:** abab

**Me:** I will partition abab into 5 parts $abab = uvxyz$ such that $uv^ixy^iz$ is accepted for any $i$:

$$\underbrace{a}_{u}\underbrace{a}_{v}\underbrace{\epsilon}_{x}\underbrace{b}_{y}\underbrace{b}_{z}$$

- $|vy| > 0$, since $|ab| = 2$
- $|vxy| \leq 4$, since $|a\epsilon b| = 2$

- $uxz = ab$ is accepted
- $u\underline{v}x\underline{y}z = a\underline{a}\epsilon\underline{b}b$ is accepted

- $u\underline{vv}x\underline{yy}z = a\underline{aa}\epsilon\underline{bb}b$ is accepted
- $u\underline{vvv}x\underline{yyy}z = a\underline{aaa}\epsilon\underline{bbb}b$ is accepted

# The Pumping Lemma (Theorem 2.34)

## For context-free languages

If $L$ is **context free**, then there is a *pumping length* $p$ where, if $w \in L$ and $|s| \geq p$, then there exists $u, v, x, y, z$ such that:

1. $w = uvxyz$
2. $|vy| \geq 1$
3. $|vxy| \leq p$
4. $uv^i xy^i z \in L$ for any $i \geq 0$

```
Theorem pumping_cfl:
  forall L,
  ContextFree L →
  exists p, p ≥ 1 /\
  forall w, L w → (* w ∊ L *)
  length w ≥ p → (* |w| ≥ p *)
  exists u v x y z, (
    w = u ++ v ++ x ++ y ++ z /\ (* w = uvxyz *)
    length (v ++ y) ≥ 1 /\      (* |vy| ≥ 1 *)
    length (v ++ x ++ y) ≤ p /\ (* |vxy| ≤ p *)
    forall i,
    L (u ++ (pow v i) ++ x ++ (pow y i) ++ z)
    (* u v^i x y^i z ∊ L *)
  ).
```

# Non-context-free languages

# Theorem: non-context-free languages

## Informally

If there exist a word $w \in L$ such that for any pumping length $p \geq 1$,

- $w \in L$
- $|w| \geq p$
- $w = uvxyz, |vy| \geq 1, |vxy| \leq p$ implies $\exists i, uv^ixy^iz \in L$

then, $L$ is not context-free.

## Formally

```
Lemma not_cfl:
  forall (L:lang),
  (* Assume 0 *) (forall p, p ≥ 1 →
  (exists w,
  (* Goal 1 *) L w /\
  (* Goal 2 *) length w ≥ p /\
  forall u v x y z, (
    (* Assume 1 *) w = u ++ v ++ x ++ y ++ z →
    (* Assume 2 *) length (v ++ y) ≥ 1 →
    (* Assume 3 *) length (v ++ x ++ y) ≤ p →
    (* Goal 3 *) exists i,
     ~ L (u ++ (pow v i) ++ x ++ (pow y i) ++ z)
  ))) →
  ~ ContextFree L.
```

# Theorem: non-context-free languages

## Part 1

There exist a word $w$ such that for any pumping length $p \geq 1$

**Goal 1:** $w \in L$

**Goal 2:** $|w| \geq p$

## Part 2

Assumptions:

- $H_1$: $w = uvxyz$
- $H_2$: $|vy| \geq 1$
- $H_3$: $|vxy| \leq p$

**Goal 3:** $\exists i, uv^i xy^i z$

# Exercise 3

Show that $L_3 = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

**Proof.**

We use the theorem of non-CFL.

For any pumping length $p > 0$ we pick $w = a^p b^p c^p$.

**Goal 1:** $w \in L_3$. *Proof.* which holds since $w = a^p b^p c^p$ and $p \geq 0$ (by hypothesis).

**Goal 2:** $|w| \geq p$. *Proof.* $|w| = 3p$, thus $|w| \geq p$.

# Exercise 3

**Assumptions**

- $H_1$: $w = uvxyz$
- $H_2$: $|vy| \geq 1$
- $H_3$: $|vxy| \leq p$

**Goal 3:** $\exists i, uv^i xy^i z \notin L_3$

*Proof.* We pick $i = 2$. Let

$$w = a^p b^p c^p$$

# Exercise 3

## Assumptions

- $H_1$: $w = uvxyz$
- $H_2$: $|vy| \geq 1$
- $H_3$: $|vxy| \leq p$

**Goal 3:** $\exists i, uv^i xy^i z \notin L_3$

*Proof.* We pick $i = 2$. Let

$$w = a^p b^p c^p$$

Let $N = |vxy|$. From $(H_1)$ $a^p b^p c^p = u\underline{vxy}z$ and $(H_2)$ $|vxy| \leq p$ we can conclude that $vxy$ can match one of two cases:

1. $vxy$ has only a's (or only b's) (or only c's)
2. $vxy$ has only a's and b's (or only b's and c's)

## Proof. (Continuation...)
### Case: only contains one type of letter

1. Without loss of generality, let us consider that there are only a's.
2. We must show that $a^{p+N}b^p c^p \notin L_3$.
3. It is enough to show that there are more a's than b's, thus $p + N \neq p$. This holds because $N > 0$ (from $H_2$).

**Proof. (Continuation...)**

**Case: contains two two types of letters.**

Without loss of generality, let us consider that $v$ containts a's and $y$ contains b's. Let $N = n + m$, where $n$ is the number of a's and $m$ is the number of b's.

$$\underbrace{a^p b^p c^p}_{uvxyz} = \underbrace{a^{p-n}}_{u} \underbrace{a^n b^m}_{vxy} \underbrace{b^{p-m} c^p}_{z}$$

Next, we recall that $vx$ may still contain only a's, or it may contain a's and b's (because of $H_2$ and $H_3$). In the case of the latter, then since we picked $i = 2$ the string is trivially not in $L_3$. The rest of the proof assumes that $v$ only has a's and $y$ only has b's.

Our goal is to show that

$$\underbrace{a^{p-n}}_{u} \underbrace{a^{n+|v|} b^{m+|y|}}_{v^2 x y^2} \underbrace{b^{p-m} c^p}_{z} \notin L_3$$

## Proof. (Continuation...)

Goal

$$\underbrace{a^{p-n}}_{u}\underbrace{a^{n+|v|}b^{m+|y|}}_{v^2xy^2}\underbrace{b^{p-m}c^p}_{z} \notin L_3$$

Since $(H_2)$ $|vy| \geq 1$, then either $|v| \geq 1$ or $|y| \geq 1$.

- If $|v| \geq 1$, it is enough to show that the number of a's differs from the number of c's, thus $p - n + n + |v| \neq p$, which holds because $|v| \geq 1$.
- If $|y| \geq 1$, then we must show that the number of b's differs from the number of a's. Hence, $m + |y| + p - n \neq p$, which holds because $|y| \geq 1$.

# Exercise 4

$L_4 = \{ww \mid w \in \{a,b\}^\star\}$

The language is **not** context free.

We pick $w = a^p b^p a^p b^p$

**Goal 1:** $w \in L_4$, because $a^p b^p \in \{a,b\}^\star$

**Goal 2:** $|w| \geq p$, because $|w| = 4p$.

**Goal 3:** $\exists i, uv^i xy^i z \notin L_4$.

Assumptions

- $H_1 \colon w = uvxyz$
- $H_2 \colon |vy| \geq 1$
- $H_3 \colon |vxy| \leq p$

**(Proof...)** Let $|vxy| = V$. If $a^p b^p a^p b^p = uvxyz$, then because $H_3 : |vxy| \leq p$, we have that $w$ can be divided into two cases:

**(Proof...)** Let $|vxy| = V$. If $a^p b^p a^p b^p = uvxyz$, then because $H_3 : |vxy| \leq p$, we have that $w$ can be divided into two cases:

**Case 1:** only a's/only b's.

Without loss of generality we handle the case for only a's and any portion of the string will work.

Thus, $w = \underbrace{a^{|u|}}_{u} \underbrace{a^V}_{xyz} \underbrace{b^p a^p b^p}_{z}$ and $|u| + V = p$.

**(Proof...)** Let $|vxy| = V$. If $a^p b^p a^p b^p = uvxyz$, then because $H_3 : |vxy| \leq p$, we have that $w$ can be divided into two cases:

**Case 1:** only a's/only b's.

Without loss of generality we handle the case for only a's and any portion of the string will work.

Thus, $w = \underbrace{a^{|u|}}_{u} \underbrace{a^V}_{xyz} \underbrace{b^p a^p b^p}_{z}$ and $|u| + V = p$.

**Case 2:** some a's and some b's. Let $A$ be the number of a's and $B$ be the number of b's, where $V = A + B$. Without loss of generality we handle the case where the string has some a's and some b's. Thus, $w = \underbrace{a^{p-A}}_{u} \underbrace{a^A b^B}_{xyz} \underbrace{b^{p-B} a^p b^p}_{z}$

## Why do we need only this 2 cases?

- Whatever a's and b's you pick (even in the middle), you must always show that that either you add/subtract $|x|$ non-empty and then you add/subtract $|y|$ non empty.