

# Non-Context-Free Languages and the Pumping Lemma

Chanathip Namprempre

Computer Science  
Reed College

# Outline

- 1 Non-Context-Free Languages
  - Examples
  - Pumping Lemma

Some languages are **not** context-free! For example,

$$\{0^n 1^n 0^n \mid n \geq 0\}$$

or the language

$$\{ww \mid w \in (0 \cup 1)^*\}$$

Intuitively, these language are problematic for PDAs because the latter only has access to a stack.

# Pumping Lemma

## Theorem (Pumping Lemma)

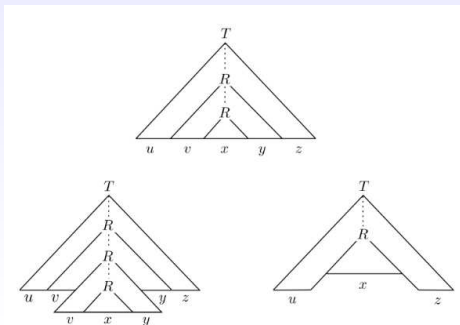
If  $L$  is a context-free language, then

there is a number  $p \geq 0$

so that for all  $s \in L$  with  $|s| \geq p$

there is a parse of  $s = uvxyz$  with  $|vy| \geq 1$  and  $|vxy| \leq p$

such that for any  $i \geq 0$ ,  $uv^i xy^i z \in L$



# Applying pumping lemma to a CFL: Example

S	→	NP VP
NP	→	Det Nom
VP	→	V NP
Nom	→	Adj Nom
Nom	→	N

Try deriving

- the angry bear chased the frightened little squirrel
- the angry crazy bear chased the frightened little squirrel
- the bear chased the frightened little squirrel

Example from Stuart Shieber

<https://dash.harvard.edu/bitstream/handle/1/2026618/Shieber.EvidenceAgainst.pdf?sequence=2>

# Pumping Lemma

## Theorem (Pumping Lemma)

If  $L$  is a context-free language, then  
there is a number  $p \geq 0$   
so that for all  $s \in L$  with  $|s| \geq p$   
there is a parse of  $s = uvxyz$  with  $|vy| \geq 1$  and  $|vxy| \leq p$   
such that for any  $i \geq 0$ ,  $uv^i xy^i z \in L$

We can use the Pumping Lemma to prove languages **not** context-free.

## Theorem (Contrapositive of Pumping Lemma)

If for any number  $p \geq 0$   
there exists a string  $s \in L$  with  $|s| \geq p$   
so that for any parse of  $s = uvxyz$  with  $|vy| \geq 1$  and  $|vxy| \leq p$   
there exists some  $i \geq 0$  such that  $uv^i xy^i z \notin L$ ,  
then  $L$  is not context-free.

# Pumping Lemma: Intuition

Consider the following language:

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

- ❶ Suppose towards a contradiction that  $L$  was context-free.
- ❷ Suppose that  $L$  is recognized by a grammar  $G$ .
- ❸ Suppose that  $G$  has  $|V|$  variables and the most number of symbols on the right hand side of any rules is  $b \geq 2$ .
- ❹ Let  $p = b^{|V|+1}$ .
- ❺ Consider the string  $w = 0^p 1^p 0^p$ .
- ❻ Since  $p = b^{|V|+1} \geq b^{|V|} + 1$ , the parse tree must be at least of height  $|V| + 1$ .
- ❼ There must be a **repeat** of at least one variable when  $G$  derives  $w$ .
- ❽ We can show that this means that  $G$  would derive a string  $w' \notin L$ .
- ❾ Thus, we have a contradiction. So  $L$  is not context-free.

# Examples

We can use the contrapositive of the Pumping Lemma to prove these languages **non-context-free**.

- ❶  $L_1 = \{0^n 1^n 0^n \mid n \geq 0\}$
- ❷  $L_2 = \{ww \mid w \in \{0,1\}^*\}$
- ❸  $L_3 = \{r\#s \mid r, s \in \{0,1\}^* \text{ and } r \text{ is not a substring of } s\}$

Notice that we **cannot** use the Pumping Lemma (or its contrapositive) to prove that a language is **context-free**.