

Pumping Lemma for Context-Free Languages

CSCI 3130 Formal Languages and Automata Theory

Siu On CHAN

Fall 2021

Chinese University of Hong Kong



$$L_1 = \{a^n b^n \mid n \geq 0\}$$

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

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

$$L_4 = \{zz^R \mid z \in \{a, b\}^*\}$$

$$L_5 = \{zz \mid z \in \{a, b\}^*\}$$

These languages are not regular

Are they context-free?

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

Let's try to design a CFG or PDA

$$S \rightarrow aBc \mid \varepsilon$$

$$B \rightarrow ???$$

read a / push x

read b / pop x

???

Suppose we could construct some CFG G for L_3

e.g.

$S \rightarrow CC \mid BC \mid a$

$B \rightarrow CS \mid b$

$C \rightarrow SB \mid c$

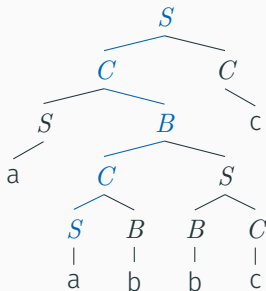
How does a long
derivation look like?

$S \Rightarrow CC$
 $\Rightarrow SBC$
 $\Rightarrow SCSC$
 $\Rightarrow SSBSC$
 $\Rightarrow SSBBC$
 $\Rightarrow aSBBC$
 $\Rightarrow aaBBCC$
 $\Rightarrow aabBCC$
 $\Rightarrow aabbCC$
 $\Rightarrow aabbcC$
 $\Rightarrow aabbcc$

Repetition in long derivations

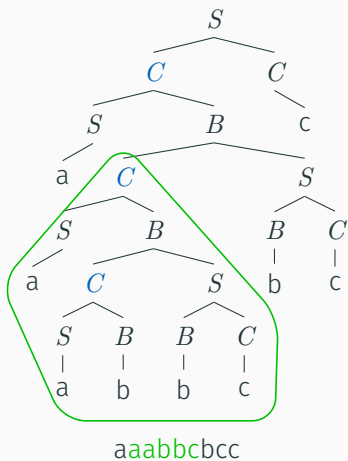
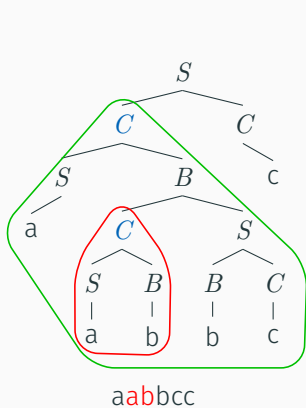
If a derivation is long enough, some variable must appear **twice on the same root-to-leave path** in a parse tree

$S \Rightarrow CC$
 $\Rightarrow SBC$
 $\Rightarrow SCSC$
 $\Rightarrow SSBSC$
 $\Rightarrow SSBBCC$
 $\Rightarrow aSBBCC$
 $\Rightarrow aaBBCC$
 $\Rightarrow aabBCC$
 $\Rightarrow aabbCC$
 $\Rightarrow aabbcC$
 $\Rightarrow aabbcc$



Pumping example

Then we can “cut and paste” part of parse tree



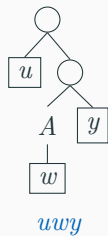
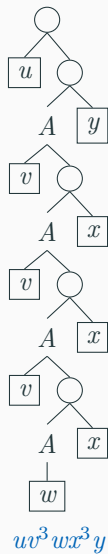
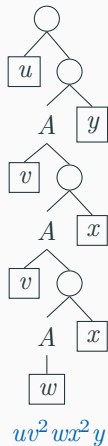
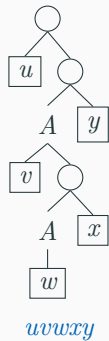
Pumping example

We can repeat this many times

$$\begin{aligned} \text{a}ab\text{bcc} &\Rightarrow \text{aa}ab\text{bcbcc} \Rightarrow \text{aaa}abb\text{cbcbcc} \Rightarrow \dots \\ &\Rightarrow (a)^i ab(bc)^i c \end{aligned}$$

Every sufficiently large derivation will have a middle part that can be repeated indefinitely

Pumping in general



Example

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

If L_3 has a context-free grammar G , then for any sufficiently long $s \in L(G)$

s can be split into $s = uvwxy$ such that $L(G)$ also contains uv^2wx^2y ,
 uv^3wx^3y, \dots

What happens if $s = a^m b^m c^m$

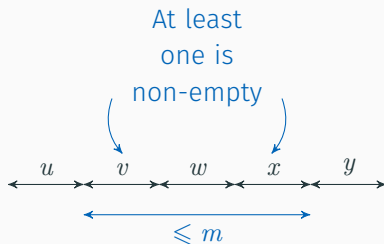
No matter how it is split, $uv^2wx^2y \notin L_3$

Pumping lemma for context-free languages

For every context-free language L

There exists a number m such that for every long string s in L ($|s| \geq m$), we can write $s = uvwxy$ where

1. $|vwx| \leq m$
2. $|vx| \geq 1$
3. For every $i \geq 0$, the string uv^iwx^iy is in L



Pumping lemma for context-free languages

To prove L is not context-free, it is enough to show that

For every m there is a long string $s \in L$, $|s| \geq m$, such that for every way of writing $s = uvwxy$ where

1. $|vwx| \leq m$
2. $|vx| \geq 1$

there is $i \geq 0$ such that uv^iwx^iy is not in L

Using the pumping lemma

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

1. for every m
2. there is $s = a^m b^m c^m$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$
($|vwx| \leq m, |vx| \geq 1$)
4. $uv^2wx^2y \notin L_3$ (but why?)

Using the pumping lemma

Case 1: v or x contains two kinds of symbols

aa aabb bccccc
 v

Then $uv^2wx^2y \notin L_3$ because the pattern is wrong

Case 2: v and x both contain (at most) one kind of symbol

aaa a b bb bccccc
 v x

Then uv^2wx^2y does not have the same number of a's, b's and c's

Conclusion: $uv^2wx^2y \notin L_3$

Which is context-free?

$$L_1 = \{a^n b^n \mid n \geq 0\} \quad \checkmark$$

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

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

$$L_4 = \{zz^R \mid z \in \{a, b\}^*\} \quad \checkmark$$

$$L_5 = \{zz \mid z \in \{a, b\}^*\}$$

Example

$$L_5 = \{zz \mid z \in \{a, b\}^*\}$$

1. for every m
2. there is $s = a^m b a^m b$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$
($|vwx| \leq m, |vx| \geq 1$)
4. Is $uv^2wx^2y \notin L_5$?

Example

$$L_5 = \{zz \mid z \in \{a, b\}^*\}$$

1. for every m
2. there is $s = a^m b a^m b$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$
($|vwx| \leq m$, $|vx| \geq 1$)
4. Is $uv^2wx^2y \notin L_5$?

$$\text{aaa} \underbrace{\text{a}}_v \text{aba} \underbrace{\text{a}}_x \text{aab}$$

Example

$$L_5 = \{zz \mid z \in \{a, b\}^*\}$$

1. for every m
2. there is $s = a^m b^m a^m b^m$ (at least m symbols)
3. no matter how the pumping lemma splits s into $uvwxy$
($|vwx| \leq m$, $|vx| \geq 1$)
4. Is $uv^iwx^iy \notin L_5$ for some i ?

Recall that $|vwx| \leq m$

Example

Three cases

Case 1 $aaa \underbrace{aabb}_{vwx} bbaaaaabbbb$

vwx is in the **first half** of $a^m b^m a^m b^m$

Case 2 $aaaaabb \underbrace{bbba}_{vwx} aaabbbb$

vwx is in the **middle part** of $a^m b^m a^m b^m$

Case 3 $aaaaabbbbbaaa \underbrace{aabb}_{vwx} bb$

vwx is in the **second half** of $a^m b^m a^m b^m$

Example

Apply pumping lemma with $i = 0$

Case 1 $aaa \underbrace{aabb}_{vwx} bbaaaaabbbb$
 uwy becomes $a^j b^k a^m b^m$ ($j < m$ or $k < m$)

Case 2 $aaaaabb \underbrace{bbba}_{vwx} aaabbbb$
 uwy becomes $a^m b^j a^k b^m$ ($j < m$ or $k < m$)

Case 3 $aaaaabbbbbaaa \underbrace{aabb}_{vwx} bb$
 uwy becomes $a^m b^m a^j b^k$ ($j < m$ or $k < m$)

Not of the form zz

This covers all cases, so L_5 is not context-free