

# Lecture 21



# The Pumping Lemma for CFL



# Reminder



# Pumping Lemma for regular languages

- Idea: Let  $M=(Q,\Sigma, q_0, A, \delta)$  be a FA recognizing  $L$ . Let  $x=a_1a_2\cdots a_n \in L$  be a string with  $n>|Q|$ . Then the sequence

$$q_0 = \delta^*(q_0, \Lambda)$$

$$q_1 = \delta^*(q_0, a_1)$$

$$q_2 = \delta^*(q_0, a_1a_2)$$

...

$$q_n = \delta^*(q_0, a_1a_2\cdots a_n)$$

Must contain some state at least twice



■ **Theorem 5.2**

Suppose  $L$  is a regular language recognized by a finite automaton with  $n$  states. For any  $x \in L$  with  $|x| \geq n$ ,  $x$  may be written as  $x = uvw$  for some strings  $u$ ,  $v$ , and  $w$  satisfying

$$|uv| \leq n$$

$$|v| > 0$$

for any  $m \geq 0$ ,  $uv^m w \in L$



■ **Theorem 5.2a The Pumping Lemma for Regular Languages**

Suppose  $L$  is a regular language. Then there is an integer  $n$  so that for any  $x \in L$  with  $|x| \geq n$ , there are strings  $u$ ,  $v$ , and  $w$  so that

$$x = uvw \quad (5.1)$$

$$|uv| \leq n \quad (5.2)$$

$$|v| > 0 \quad (5.3)$$

$$\text{for any } m \geq 0, uv^m w \in L \quad (5.4)$$



## Intuition

- What happens if some derivation involves a variable, say  $A$ , more than once?
- $A \rightarrow xAy$
- $A \rightarrow a$



## Reminder

- Chomsky Normal Form
- $A \rightarrow BC$
- $A \rightarrow a$



- Tree representing derivation: Binary tree
- Path, length (number of nodes)
- Height



## Lemma

- Binary tree having more than  $2^{h-1}$  leaf nodes must have height greater than  $h$ .



### Theorem 8.1

Let  $G = (V, \Sigma, S, P)$  be a context-free grammar in Chomsky normal form, with a total of  $p$  variables. Any string  $u$  in  $L(G)$  with  $|u| \geq 2^{p+1}$  can be written as  $u = vwx^mzy$ , for some strings  $v, w, x, y,$  and  $z$  satisfying

$$|wy| > 0$$

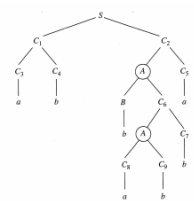
$$|wxy| \leq 2^{p+1}$$

for any  $m \geq 0, vw^mxy^mz \in L(G)$



## Proof

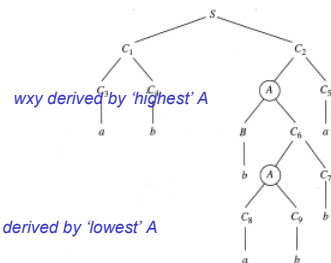
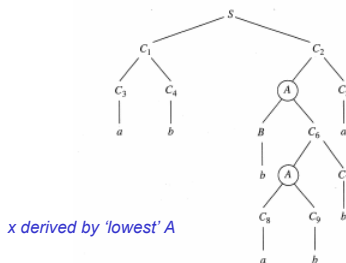
- 



- By Lemma, any derivation tree has height  $\geq p+2$
- Take a path of maximum length



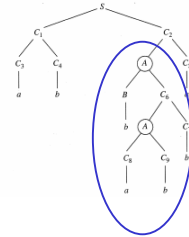
- Take a path of maximum length: its end leaf is a terminal and all other  $p+1$  nodes are variables
- Thus there is a variable  $A$  appearing at least twice in this path



- $u=vwxyz$



Take a subtree



- This is a derivation subtree for  $wxy$
- Its height is  $\leq p+2$
- Thus the number of leaves is  $\leq 2^{p+1}$
- and  $|wxy| \leq 2^{p+1}$



- The root  $A$  of the subtree has two children corresponding to variables
- Let  $B$  be the one who is not ancestor of the second  $A$
- Then the string derived from  $B$  does not overlap  $x$



- $wxy$
- $w$  or  $y$  is nonnull
- $|wy| > 0$
- $S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vwxyz$



## The pumping lemma for CFL

- Let  $L \in \text{CFL}$ . There is an integer  $n$  so that for any  $u \in L$ ,  $|u| \geq n$ , there are strings  $v, w, x, y$ , and  $z$  satisfying:
  - $u = vwxyz$
  - $|wy| > 0$
  - $|wxy| \leq n$
  - For any  $m \geq 0$ ,  $vw^mxy^mz \in L$



## Proof

- Find a CFG in Chomsky normal form that generates  $L - \{\Lambda\}$
- Let  $p$  be the number of variables and  $n = 2^{p+1}$



## Example

- $L = \{a^i b^j c^i\}$

