

## Section 6

# Context-Free and Non-Context-Free Languages

- 6 Context-Free and Non-Context-Free Languages
  - Pumping Lemma
  - Decision problems

Huiswerk. . .

Volgende week: geen werkcollege 's middags

Vragenuur: dinsdag 21 december, 14.15-16.00? **Ja!**

# Pumping lemma for regular languages

From lecture 2:

*Regular language* is language accepted by an FA.

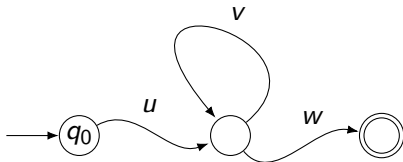
## Theorem

*Suppose  $L$  is a language over the alphabet  $\Sigma$ . If  $L$  is accepted by a finite automaton  $M$ , and if  $n$  is the number of states of  $M$ , then*

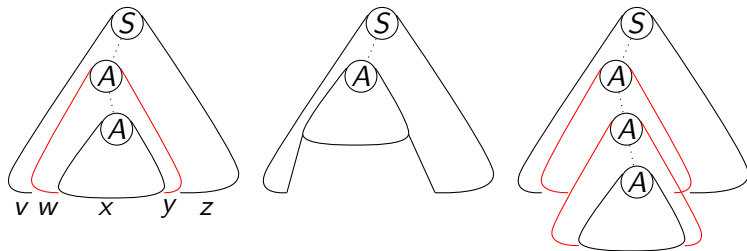
- $\forall$  for every  $x \in L$   
satisfying  $|x| \geq n$
- $\exists$  there are three string  $u$ ,  $v$ , and  $w$ ,  
such that  $x = uvw$  and the following three conditions are true:
  - (1)  $|uv| \leq n$ ,
  - (2)  $|v| \geq 1$
- $\forall$  and (3) for all  $i \geq 0$ ,  $uv^i w$  belongs to  $L$

[M] Thm. 2.29

From lecture 2:

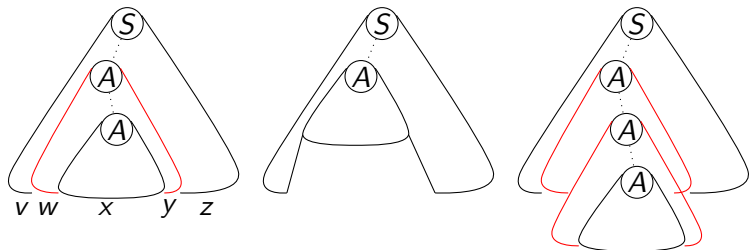


[M] Fig. 2.28



$$S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vwxyz, v, w, x, y, z \in \Sigma^*$$

$$S \underset{(1)}{\Rightarrow^*} vAz, A \underset{(2)}{\Rightarrow^*} wAy, A \underset{(3)}{\Rightarrow^*} x$$



$$S \underset{(1)}{\Rightarrow^*} vAz \underset{(3)}{\Rightarrow^*} vxz$$

$$S \underset{(1)}{\Rightarrow^*} vAz \underset{(2)}{\Rightarrow^*} vwAyz \underset{(2)}{\Rightarrow^*} vwwAyyz \underset{(3)}{\Rightarrow^*} vwwwxyyz$$

## Theorem (Pumping Lemma for context-free languages)

- ∀ for every context-free language  $L$
- ∃ there exists a constant  $n \geq 1$   
such that
- ∀ for every  $u \in L$   
with  $|u| \geq n$
- ∃ there exists a decomposition  $u = vwxyz$   
such that
  - (1)  $|wy| \geq 1$
  - (2)  $|wxy| \leq n$ ,
- ∀ (3) for all  $m \geq 0$ ,  $vw^mxy^mz \in L$

[M] Thm. 6.1



## Example

$AnBnCn$  is not context-free.

[M] E 6.3

## Theorem (Pumping Lemma for context-free languages)

- ∀ for every context-free language  $L$
- ∃ there exists a constant  $n \geq 1$   
such that
- ∀ for every  $u \in L$   
with  $|u| \geq n$
- ∃ there exists a decomposition  $u = vwxyz$   
such that
  - (1)  $|wy| \geq 1$
  - (2)  $|wxy| \leq n$ ,
- ∀ (3) for all  $m \geq 0$ ,  $vw^mxy^mz \in L$

If  $L = L(G)$  with  $G$  in ChNF, then  $n = 2^{|V|}$ .

Proof...

[M] Thm. 6.1

*From lecture 9:*

## Definition

CFG in *Chomsky normal form*

productions are of the form

- $A \rightarrow BC$  variables  $A, B, C$
- $A \rightarrow \sigma$  variable  $A$ , terminal  $\sigma$

## Theorem

*For every CFG  $G$  there is CFG  $G_1$  in CNF such that  $L(G_1) = L(G) - \{\Lambda\}$ .*

[M] Def 4.29, Thm 4.30

## Theorem (Pumping Lemma for context-free languages)

[M] Thm. 6.1

### Proof

Let  $G$  be CFG in Chomsky normal form with  $L(G) = L - \{\Lambda\}$ .

Derivation tree in  $G$  is binary tree

(where each parent of a leaf node has only one child).

Height of a tree is number of edges in longest path from root to leaf node.

At most  $2^h$  leaf nodes in binary tree of height  $h$ :  $|u| \leq 2^h$ .

## Theorem (Pumping Lemma for context-free languages)

[M] Thm. 6.1

### Proof (continued)

At most  $2^h$  leaf nodes in binary tree of height  $h$ :  $|u| \leq 2^h$ .

Let  $p$  be number of variables in  $G$ ,

let  $n = 2^p$

and let  $u \in L(G)$  with  $|u| \geq n$ .

(Internal part of) derivation tree of  $u$  in  $G$  has height at least  $p$ .

Hence, longest path in (internal part of) tree contains at least  $p + 1$  (internal) nodes.

Consider final portion of longest path in derivation tree.

(leaf node +  $p + 1$  internal nodes),

with  $\geq 2$  occurrences of a variable  $A$ .

Pump up derivation tree, and hence  $u$ .

## Application of pumping lemma:

mainly to prove that a language  $L$  **cannot** be generated by a context-free grammar.

How?

Find a string  $u \in L$  with  $|u| \geq n$  that cannot be pumped up!

What is  $n$ ?

What should  $u$  be?

What can  $v$ ,  $w$ ,  $x$ ,  $y$  and  $z$  be?

What should  $m$  be?

**Suppose that there exists context-free grammar  $G$  with  $L(G) = L$ .  
Let  $n$  be the integer from the pumping lemma.**

We prove:

There exists  $u \in L$  with  $|u| \geq n$ , such that  
for every five strings  $v, w, x, y$  and  $z$  such that  $u = vwxyz$

if

1.  $|wy| \geq 1$
2.  $|wxy| \leq n$

then

3. there exists  $m \geq 0$ , such that  $vw^mxy^mz$  **does not** belong to  $L$

## Example

$AnBnCn$  is not context-free.

[M] E 6.3

$$u = a^n b^n c^n$$

$$\{ x \in \{a, b, c\}^* \mid n_a(x) = n_b(x) = n_c(x) \}$$

## Example

$XX$  is not context-free.

[M] E 6.4



## Example

$AnBnCn$  is not context-free.

[M] E 6.3

$$u = a^n b^n c^n$$

$$\{ x \in \{a, b, c\}^* \mid n_a(x) = n_b(x) = n_c(x) \}$$

## Example

$XX$  is not context-free.

[M] E 6.4

$$u = a^n b^n a^n b^n$$

$$\{ a^i b^j a^i b^j \mid i, j \geq 0 \}$$

## Example

$\{ x \in \{a, b, c\}^* \mid n_a(x) < n_b(x) \text{ and } n_a(x) < n_c(x) \}$  is not context-free.

[M] E 6.5

ABOVE

$L = \{ x \in \{a, b, c\}^* \mid n_a(x) < n_b(x) \text{ and } n_a(x) < n_c(x) \}$  is not context-free.

Proof by contradiction.

Suppose  $L$  is context-free, then there exists a pumping constant  $n$  for  $L$ .

Choose  $u = a^n b^{n+1} c^{n+1}$ . Then  $u \in L$ , and  $|u| \geq n$ .

This means that we can pump  $u$  within the language  $L$ .

Consider a decomposition  $u = vwxyz$  that satisfies the pumping lemma, in particular  $|wxy| \leq n$ .

**Case 1:**  $wy$  contains a letter  $a$ . Then  $wy$  cannot contain letter  $c$  (otherwise  $|wxy| > n$ ). Now  $u_2 = vw^2xy^2z$  contains more  $a$ 's than  $u$ , so at least  $n+1$ , while  $u_2$  still contains  $n+1$   $c$ 's. Hence  $u_2 \notin L$ .

**Case 2:**  $wy$  contains no  $a$ . Then  $wy$  contains at least one  $b$  or one  $c$  (or both). Then  $u_0 = vw^0xy^0z = vxz$  has still  $n$   $a$ 's, but less than  $n+1$   $b$ 's or less than  $n+1$   $c$ 's (depending on which letter is in  $wy$ ). Hence  $u_0 \notin L$ .

These are two possibilities for the decomposition  $vwxyz$ , in both cases we see that pumping leads out of the language  $L$ .

Hence  $u$  cannot be pumped.

Contradiction; so  $L$  is not context-free.

## Example

### The Set of Legal C Programs is Not a CFL

[M] E 6.6

Choose  $u =$

```
main(){int aaa...a;aaa...a=aaa...a;}
```

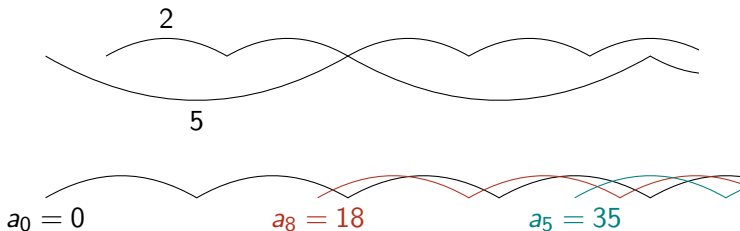
where  $aaa...a$  contains  $n + 1$  a's

# Applying the Pumping Lemma (2)

Lemma (☒)

$L \subseteq \{a\}^*$  context-free, then  $L$  regular.

[M] Exercise 6.23



This exercise does not have to be known for the exam.

*From lecture 3:*

Prove the following generalization of the pumping lemma, which can sometimes make it unnecessary to break the proof into cases.

If  $L$  can be accepted by an FA,

then there is an integer  $n$

such that for any  $x \in L$  with  $|x| \geq n$

and for any way of writing  $x$  as  $x_1x_2x_3$  with  $|x_2| = n$ ,

there are strings  $u$ ,  $v$  and  $w$  such that

a.  $x_2 = uvw$

b.  $|v| \geq 1$

c. For every  $m \geq 0$ ,  $x_1uv^mw x_3 \in L$

Generalization of pumping lemma for CFL:  
pump at distinguished positions in  $u$

Ogden's lemma does not have to be known for the exam.

From lecture 2:

FA  $M_i = (Q_i, \Sigma, q_i, A_i, \delta_i) \quad i = 1, 2$

### Product construction

construct FA  $M = (Q, \Sigma, q_0, A, \delta)$  such that

- $Q = Q_1 \times Q_2$
- $q_0 = (q_1, q_2)$
- $\delta((p, q), \sigma) = (\delta_1(p, \sigma), \delta_2(q, \sigma))$
- $A$  as needed

### Theorem (2.15 Parallel simulation)

- $A = \{(p, q) \mid p \in A_1 \text{ or } q \in A_2\}$ , then  $L(M) = L(M_1) \cup L(M_2)$
- $A = \{(p, q) \mid p \in A_1 \text{ and } q \in A_2\}$ , then  $L(M) = L(M_1) \cap L(M_2)$
- $A = \{(p, q) \mid p \in A_1 \text{ and } q \notin A_2\}$ , then  $L(M) = L(M_1) - L(M_2)$

Proof...

*From lecture 6:*

Regular languages are closed under

- Boolean operations (complement, union, intersection)
- Regular operations (union, concatenation, star)
- Reverse (mirror)
- [inverse] Homomorphism



*From lecture 7:*

Using building blocks

## Theorem

If  $L_1, L_2$  are CFL, then so are  $L_1 \cup L_2$ ,  $L_1 L_2$  and  $L_1^*$ .

$G_i = (V_i, \Sigma, S_i, P_i)$ , having no variables in common.

## Construction

$G = (V_1 \cup V_2 \cup \{S\}, \Sigma, S, P)$ , new axiom  $S$   
 -  $P = P_1 \cup P_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}$      $L(G) = L(G_1) \cup L(G_2)$   
 -  $P = P_1 \cup P_2 \cup \{S \rightarrow S_1 S_2\}$      $L(G) = L(G_1) L(G_2)$

$G = (V_1 \cup \{S\}, \Sigma, S, P)$ , new axiom  $S$   
 -  $P = P_1 \cup \{S \rightarrow S S_1, S \rightarrow \Lambda\}$      $L(G) = L(G_1)^*$

[M] Thm 4.9

How about

- $L_1 \cap L_2$
- $L_1 - L_2$
- $L'_1$

for CFLs  $L_1$  and  $L_2$  ?

*From lecture 8:*

### Example

$AnBnCn$  is intersection of two context-free languages.

$$L_1 = \{a^i b^j c^k \mid i, k \geq 0\}$$

$$L_2 = \{a^i b^k c^k \mid i, k \geq 0\}$$

[M] E 6.10

Hence, CFL is not closed under intersection

## Example

$AnBnCn$  is intersection of two context-free languages.

[M] E 6.10

Hence, CFL is not closed under intersection

$$L_1 \cap L_2 = (L_1' \cup L_2)'$$

Hence, CFL is not closed under complement

$$L_1' = \Sigma^* - L_1$$

Hence, CFL is not closed under setminus

## Example

Complement of  $XX$

$= \{ x \in \{a, b\}^* \mid |x| \text{ is odd} \} \cup \{ xy \mid x, y \in \{a, b\}^*, |x| = |y|, x \neq y \}$   
is context-free

[M] E 6.11

Indeed, CFL is not closed under complement

## Example

Complement of  $AnBnCn$  is context-free.

Complement of  $\{x \in \{a, b, c\}^* \mid n_a(x) = n_b(x) = n_c(x)\}$  is context-free.

[M] E 6.12

## Example

Complement of  $AnBnCn$  is context-free.

Complement of  $\{x \in \{a, b, c\}^* \mid n_a(x) = n_b(x) = n_c(x)\}$  is context-free.

$AnBnCn = L_1 \cap L_2 \cap L_3$ , with

$$L_1 = \{a^i b^j c^k \mid i \leq j\}$$

$$L_2 = \{a^i b^j c^k \mid j \leq k\}$$

$$L_3 = \{a^i b^j c^k \mid k \leq i\}$$

[M] E 6.12

## Example

Complement of  $AnBnCn$  is context-free.

Complement of  $\{x \in \{a, b, c\}^* \mid n_a(x) = n_b(x) = n_c(x)\}$  is context-free.

$\{x \in \{a, b, c\}^* \mid n_a(x) = n_b(x) = n_c(x)\} = A_1 \cap A_2 \cap A_3$ , with

$$A_1 = \{x \in \{a, b, c\}^* \mid n_a(x) \leq n_b(x)\}$$

$$A_2 = \{x \in \{a, b, c\}^* \mid n_b(x) \leq n_c(x)\}$$

$$A_3 = \{x \in \{a, b, c\}^* \mid n_c(x) \leq n_a(x)\}$$

[M] E 6.12



## Example

$$L_1 = \{ a^{2^n} b^n \mid n \geq 1 \}^*$$

$$a^{16} b^8 a^8 b^4 a^4 b^2 a^2 b^1$$

$$L_2 = a^* \{ b^n a^n \mid n \geq 1 \}^* \{ b \}$$