

# Formal Languages-Course 7.

Géraud Sénizergues

Bordeaux university

25/05/2020

Master computer-science MINF19, IEI, 2019/20

# contents

- 1 Proofs : induction on derivation-length
- 2 Proofs : induction on word-length
  - The fundamental lemma
  - Using the fundamental lemma
- 3 Proofs : non-ambiguity
- 4 Proofs : grammar equivalence

# Proofs on formal languages

We are interested in :

- proving that all words  $w \in L$  fulfill property  $P(w)$
- proving that a language  $L$  has a property (prefix-freeness for example)
- proving that a grammar  $G$  has a property (non-ambiguity for example)
- proving that  $G$  generates language  $L$  (specified in a different way)

What **methods** are available?

- 1- Induction on **derivation-length**
- 2- Induction on **word-length**

# Induction on derivation-length

# Induction on derivation-length

Let  $G_1 = \langle A, N, R, S \rangle$  with

$$A = \{a, b\}, \quad N = \{S\}, \quad S \longrightarrow aSS \mid b,$$

Let us define , for every word  $u \in A^*$ ,  $\|u\| = |u|_a - |u|_b$ . Let us prove that :

$$\forall u \in L(G_1, S), \quad \|u\| = -1.$$

We can extend the norm notation to words  $w \in (A \cup N)^*$  by :

$$\|w\| := |w|_a - |w|_{S,b}.$$

We prove, by induction over the integer  $n$ , the following property  $P(n)$  :

$$\forall w \in (A \cup N)^*, \forall m \leq n, \quad S \xrightarrow{m}_R w \Rightarrow \|w\| = -1.$$

**Basis** :  $n = 0$

Assume that  $S \xrightarrow{0}_R w$ . Then  $w = S$  and  $\|S\| = -1$ .

# Induction on derivation-length

**Induction step** : we assume  $P(n)$ .

Assume  $S \xrightarrow{n+1}_R w$ . The derivation can be decomposed as :

$$S \xrightarrow{n}_R w_1 = \alpha S \beta \xrightarrow{1}_R \alpha v \beta$$

for some  $\alpha, \beta \in (A \cup N)^*$  and some rule  $S \rightarrow v$ .

Since  $\|b\| = -1$  and  $\|aSS\| = 1 - 1 - 1 = -1$ ,  $\|v\| = -1$ . By (IH)

$\|w_1\| = -1$ . Hence

$$\|w\| = \|w_1\| + \|v\| - \|S\| = -1 + (-1) - (-1) = -1. \quad \square$$

# Induction on word-length

# Fundamental lemma : version 1

## Lemma

*(fundamental lemma) Let  $u_1, u_2, v \in (X \cup V)^*$ . If  $u_1 u_2 \xrightarrow{k} v$  then, there exist  $v_1, v_2 \in (X \cup V)^*$  such that*

$$v = v_1 v_2, u_1 \xrightarrow{k_1} v_1, u_2 \xrightarrow{k_2} v_2 \text{ and } k_1 + k_2 = k.$$

We prove the lemma by induction over  $k$ .

**Base 0** :  $k = 0$ .

Obvious.

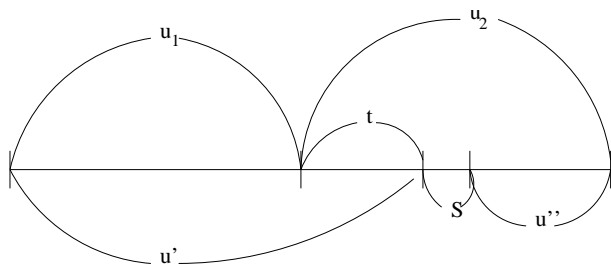


# Fundamental lemma : version 1

**Base 1** :  $k = 1$  :  $u_1 u_2 \rightarrow v$ . By definition  $u_1 u_2 = u' S u''$ ,  $v = u' m u''$  and  $S \rightarrow m \in P$ .

**Case 1** :  $|u'| \geq |u_1|$ .

In this case  $u' = u_1 t$  and  $u_2 = t S u''$



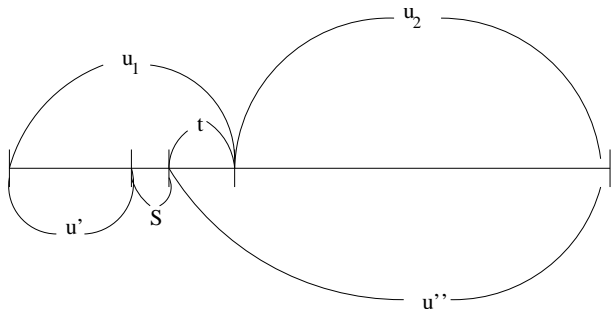
Choosing  $v_1 = u_1$ ,  $v_2 = t m u''$  we get that

$u_1 \xrightarrow{0} v_1$ ,  $u_2 \xrightarrow{1} v_2$ ,  $v = v_1 v_2$  and  $0 + 1 = 1$ .

## Fundamental lemma : version 1

**Case 2** :  $|u'| < |u_1|$ .

Symmetrically we have,  $u_1 = u'St$  and  $u'' = tu_2$



Choosing  $v_1 = u'mt$  and  $v_2 = u_2$  we get :

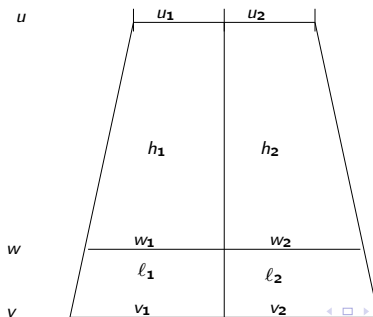
$$u_1 \xrightarrow{1} v_1, u_2 \xrightarrow{0} v_2, v = v_1 v_2 \text{ and } 1 + 0 = 1.$$

## Fundamental lemma : version 1

**Induction step** : Let  $k \geq 2$ . We assume the lemma is true for all derivations of length  $< k$ . Suppose  $u_1 u_2 \xrightarrow{k} v$ . Then

$$u_1 u_2 \xrightarrow{k-1} w \rightarrow v$$

By (IH) ,  $w = w_1 w_2$  with  $u_1 \xrightarrow{h_1} w_1$ ,  $u_2 \xrightarrow{h_2} w_2$ ,  $h_1 + h_2 = k - 1$  and  $w_1 w_2 \rightarrow v$ .



# Fundamental lemma : version 1

By the proof for order 1 derivations,  $v = v_1 v_2$  with  $w_1 \xrightarrow{\ell_1} v_1$ ,

$w_2 \xrightarrow{\ell_2} v_2$ ,  $\ell_1 + \ell_2 = 1$ .

Hence  $u_1 \xrightarrow{h_1+\ell_1} v_1$ ,  $u_2 \xrightarrow{h_2+\ell_2} v_2$ ,  $h_1 + \ell_1 + h_2 + \ell_2 = k$

□

## Fundamental lemma : version 2

## Lemma

For every  $p \geq 2$ ,  $u_1, u_2, \dots, u_p \in (A \cup N)^*$ , if  $u_1 u_2 \cdots u_p \xrightarrow{k} v$  then  $v = v_1 v_2 \cdots v_p$  with  $\forall i \in [1, p]$ ,  $u_i \xrightarrow{k_i} v_i$  and  $\sum_{i=1}^p k_i = k$ .

# Using the fundamental lemma

Let  $G_1 = \langle A, N, R, S \rangle$  with

$$A = \{a, b\}, \quad N = \{S\}, \quad S \longrightarrow aSS \mid b,$$

For every word  $u \in A^*$ ,  $\|u\| := |u|_a - |u|_b$ . Let us prove that :

$$\forall u \in L(G_1, S), \quad \|u\| = -1.$$

We prove, by induction on  $n$ , the following property  $Q(n)$  :

$$\forall w \in A^*, (S \xrightarrow{*}_R w \text{ and } |w| \leq n) \Rightarrow \|w\| = -1.$$

**Basis :**  $|w| = 1$

If  $S \xrightarrow{*}_R w$  and  $|w| = 1$ , then  $w = b$ , and  $\|b\| = -1$ .

# Using the fundamental lemma

**Induction step** : we assume  $Q(n)$ .

Assume  $S \xrightarrow{n+1}_R w$  and  $|w| \leq n + 1$ . The derivation can be decomposed as :

$$S \xrightarrow{1}_R aSS \xrightarrow{n}_R w$$

By the fundamental lemma,  $w = v_1 v_2 v_3$  with

$$a = v_1, S \xrightarrow{*} v_2, S \xrightarrow{*} v_3.$$

Since  $|a| = 1$ ,  $|v_2| \leq n$ ,  $|v_3| \leq n$ . By (IH),  $\|v_2\| = \|v_3\| = -1$ .

Hence  $\|w\| = \|av_2v_3\| = 1 + (-1) + (-1) = -1$ .  $\square$

# Non-ambiguity



# Non-ambiguity

Let us prove that the following grammar is non-ambiguous :

Let  $H = \langle A, N, R, S \rangle$  with

$$A = \{a, b\}, \quad N = \{S, S_a, S_b, D_a, D_b\},$$

$$S \longrightarrow D_a S \mid D_b S \mid \varepsilon.$$

$$S_a \longrightarrow D_a S_a \mid \varepsilon, \quad S_b \longrightarrow D_b S_b \mid \varepsilon.$$

$$D_a \longrightarrow a S_a b, \quad D_b \longrightarrow b S_b a.$$

## Non-ambiguity

## Lemma

$$L(H, S_a) = (L(H, D_a))^*.$$

**Proof :**

We prove the two sided-inclusion by induction on the length of words.

1- Suppose  $w \in (L(H, D_a))^*$ .

If  $w \in L(H, D_a)^0$ ,  $S_a \rightarrow \varepsilon = w$ .

If  $w \in L(H, D_a)^1$ ,  $S_a \rightarrow D_a S_a \rightarrow D_a \xrightarrow{*} w$ .

If  $w \in L(H, D_a)^n$ , with  $n \geq 2$ , then

$$w = d \cdot w_2 \text{ with } d \in L(H, D_a), w_2 \in L(H, D_a)^{n-1}.$$

Since  $\varepsilon \notin L(H, D_a)$ ,  $|w_2| < |w|$ . By (IH),  $S_a \xrightarrow{*} w_2$ , so that

$$S_a \rightarrow D_a S_a \xrightarrow{*} dw_2 = w.$$

# Non-ambiguity

2- Suppose  $w \in L(H, S_a)$ .

2.1 If  $S_a \rightarrow \varepsilon = w$ , then  $w \in (L(H, D_a))^*$ .

2.2 Otherwise

$$S_a \rightarrow D_a S_a \xrightarrow{*} w.$$

By the fundamental lemma :  $w = w_1 \cdot w_2$  with

$D_a \xrightarrow{*} w_1, S_a \xrightarrow{*} w_2$ . By (IH)  $w_2 \in (L(H, D_a))^*$  hence

$w \in (L(H, D_a))^*$ .  $\square$

# Non-ambiguity

## Lemma

$$L(H, D_a) = a(L(H, D_a))^*b.$$

### Proof :

The two sided-inclusion can be proved using the fundamental lemma :

1- If  $D_a \xrightarrow{*}_H w$  then  $D_a \rightarrow aS_ab \xrightarrow{*}_H w$ . By the fundamental lemma  $w = avb$  for some  $v \in L(H, S_a)$ . Hence by previous lemma,  $w \in a(L(H, D_a))^*b$ .

2- If  $w \in a(L(H, D_a))^*b$ , then  $w = avb$  for some  $v \in L(H, S_a)$ .

Hence  $D_a \rightarrow aS_ab \xrightarrow{*}_H avb = w$ .  $\square$

## Non-ambiguity

## Lemma

$L(H, D_a)$  is prefix-free.

We prove, by induction on  $\max(|u|, |v|)$  that,

$$\forall u, v \in L(H, D_a), \quad u \preceq v \Rightarrow u = v.$$

Suppose that  $u, v \in L(H, D_a)$ ,  $u \preceq v$ ,  $|u| \leq n + 1$ ,  $|v| \leq n + 1$ .

By the lemma above,  $u = ad_1 \cdots d_p b$ ,  $v = ad'_1 \cdots d'_q b$  with

$p, q \geq 0$ ,  $d_i, d'_j \in L(H, D_a)$  and  $|d_i| \leq n$ ,  $|d'_j| \leq n$ .

By (IH),  $d_1 = d'_1$ , hence  $d_2 \cdots d_p b \preceq d'_2 \cdots d'_q b$ , hence

$ad_2 \cdots d_p b \preceq ad'_2 \cdots d'_q b$ . By (IH)  $ad_2 \cdots d_p b = ad'_2 \cdots d'_q b$  so

that, finally

$$u = ad_1 \cdots d_p b = ad'_1 \cdots d'_q b = v.$$

# Non-ambiguity

## Lemma

$$L(H, S_b) = (L(H, D_b))^*.$$

## Lemma

$L(H, D_b)$  is prefix-free.

Analogous proofs.

# Non-ambiguity

## Proposition

*H is non-ambiguous*

We prove, by induction over  $n$  the property  $UN(n)$  :

$\forall T \in N, \forall u \in A^*, \forall D_1 : T \xrightarrow{\ell}^{n_1}_R u, \forall D_2 : T \xrightarrow{\ell}^{n_2}_R u,$

$$\max(n_1, n_2) \leq n \Rightarrow D_1 = D_2.$$

**Basis** :  $n = 1$ .

$D_1 : T \rightarrow u, D_2 : T \rightarrow u$  with  $T \in \{S, S_a, S_b\}$ . Then  $u = \varepsilon$  and  $D_1 = D_2$ .

**Induction step** :

Suppose that

$$D_1 : T \xrightarrow{\ell}^{n_1}_R u, \quad D_2 : T \xrightarrow{\ell}^{n_2}_R u$$

and  $\max(n_1, n_2) = n + 1 \geq 2$ .

# Non-ambiguity

Case 1 :  $T = S_a$ .

$$D_1 : S_a \rightarrow D_a S_a \xrightarrow{\ell^{n_1-1}}_R u, \quad D_2 : S_a \rightarrow D_a S_a \xrightarrow{\ell^{n_2-1}}_R u.$$

By the fundamental lemma :

$$D_1 : S_a \rightarrow D_a S_a \xrightarrow{\ell^{p_1}}_R d_1 S_a \xrightarrow{\ell^{q_1}}_R d_1 u_1 \text{ with } p_1 + q_1 = n_1 - 1$$

$$D_2 : S_a \rightarrow D_a S_a \xrightarrow{\ell^{p_2}}_R d_2 S_a \xrightarrow{\ell^{q_2}}_R d_2 u_2 \text{ with } p_2 + q_2 = n_2 - 1$$

Since  $d_1, d_2 \in L(H, D_a)$  and  $(d_1 \preceq d_2 \text{ or } d_2 \preceq d_1)$ ,  $d_1 = d_2$ . It follows that  $u_1 = u_2$ . By (IH), the derivations  $D_a S_a \xrightarrow{\ell^{p_1}}_R d_1 S_a$  and  $D_a S_a \xrightarrow{\ell^{p_2}}_R d_2 S_a$  are equal. As well, the derivations  $d_1 S_a \xrightarrow{\ell^{q_1}}_R d_1 u_1$  and  $d_2 S_a \xrightarrow{\ell^{q_2}}_R d_2 u_2$  are equal. Hence  $D_1 = D_2$ .



# Non-ambiguity

Case 2 :  $T = D_a$ .

$$D_1 : D_a \rightarrow aS_a b \xrightarrow{\ell^{n_1-1}}_R u, \quad D_2 : D_a \rightarrow aS_a b \xrightarrow{\ell^{n_2-1}}_R u.$$

By the fundamental lemma :

$$D_1 : D_a \rightarrow aS_a b \xrightarrow{\ell^{n_1-1}}_R ad_1 b, \quad D_2 : D_a \rightarrow aS_a b \xrightarrow{\ell^{n_2-1}}_R ad_2 b.$$

Since  $ad_1 b = ad_2 b$ , we have  $d_1 = d_2$ .

By (IH), the derivations  $aS_a b \xrightarrow{\ell^{n_1-1}}_R ad_1 b$  and  $aS_a b \xrightarrow{\ell^{n_2-1}}_R ad_2 b$  are equal. Hence  $D_1 = D_2$ .

# Non-ambiguity

Case 3 :  $T = S_b$ .

Case 4 :  $T = D_b$ .

Can be treated in the same way.

Case 5 :  $T = S$ .

$D_1 : S \rightarrow D_a S \xrightarrow{\ell^{n_1-1}}_R u$ ,  $D_2 : S \rightarrow D_a S \xrightarrow{\ell^{n_2-1}}_R u$ . Using the fundamental lemma and (IH) we get that  $D_1 = D_2$ .

$D_1 : S \rightarrow D_b S \xrightarrow{\ell^{n_1-1}}_R u$ ,  $D_2 : S \rightarrow D_b S \xrightarrow{\ell^{n_2-1}}_R u$ . By the same arguments  $D_1 = D_2$ .

$$D_1 : S \rightarrow D_a S \xrightarrow{\ell^{n_1-1}}_R u, \quad D_2 : S \rightarrow D_b S \xrightarrow{\ell^{n_2-1}}_R u.$$

is impossible since the first letter  $\alpha$  of  $u$  determines whether the first rule is  $S \rightarrow D_a S$  (if  $\alpha = a$ ) or  $S \rightarrow D_b S$  (if  $\alpha = b$ ).

In all cases :  $D_1 = D_2$ .

□

# Grammar equivalence

# Grammar equivalence

Let  $G_2 = \langle A, N, R'_2, S \rangle$  with

$$A = \{a, b\}, \quad N = \{S\}, \quad S \longrightarrow aSbS \mid bSaS \mid \varepsilon$$

and  $H = \langle A, N, R, S \rangle$  with

$$A = \{a, b\}, \quad N = \{S, S_a, S_b, D_a, D_b\},$$

$$S \longrightarrow D_aS \mid D_bS \mid \varepsilon.$$

$$S_a \longrightarrow D_aS_a \mid \varepsilon, \quad S_b \longrightarrow D_bS_b \mid \varepsilon.$$

$$D_a \longrightarrow aS_ab, \quad D_b \longrightarrow bS_ba.$$

We shall prove that :

$$L(G_2, S) = L(H, S).$$

# Grammar equivalence

By removing the non-terminals  $D_a, D_b$  in  $H$  (see the transformation in course 5), we obtain  $H' = \langle A, N, R', S \rangle$  with

$$A = \{a, b\}, \quad N = \{S, S_a, S_b\},$$

$$S \longrightarrow aS_a bS \mid bS_b aS \mid \varepsilon.$$

$$S_a \longrightarrow aS_a bS_a \mid \varepsilon, \quad S_b \longrightarrow bS_b aS_b \mid \varepsilon.$$

We know that  $L(H', S) = L(H, S)$ . It remains to prove that

$$L(G_2, S) = L(H', S).$$

## Grammar equivalence

## Lemma

$$L(H', S) \subseteq L(G_2, S).$$

**Proof :**

Let  $\varphi : (A \cup \{S, S_a, S_b\})^* \rightarrow (A \cup \{S\})^*$  be the homomorphism defined by :

$$\varphi(S_a) = \varphi(S_b) = \varphi(S) = S \quad \varphi(a) = a, \quad \varphi(b) = b.$$

We check that : for every rule  $(T, m) \in R'$ ,  $(\varphi(T), \varphi(m)) \in R_2$ .  
It follows, by induction on the length of derivations, that

$$\forall u, v \in (A \cup \{S, S_a, S_b\})^*, \quad u \xrightarrow{*}_{G_2} v \Rightarrow \varphi(u) \xrightarrow{*}_{H'} \varphi(v).$$

In particular :

$$S \xrightarrow{*}_{G_2} w \in A^* \Rightarrow S \xrightarrow{*}_{H'} w.$$

□

## Grammar equivalence

We recall that, for every  $u \in \{a, b\}^*$ ,  $\|u\| = |u|_a - |u|_b$ . Let

$$L_{=} := \{w \in \{a, b\}^* \mid \|w\| = 0\}.$$

## Lemma

$$L(G_2, S) \subseteq L_{=}.$$

**Proof :**

By induction on the length of words, using the fundamental lemma.

□

## Lemma

Let  $u \in A^*$ . If  $\|u\| = 0$  and  $\forall v \preceq u$ ,  $\|v\| \geq 0$ , then  $S_a \xrightarrow{*}_{H'} u$ .

**Proof :**

By induction on the length of words :

**Basis :**  $|u| = 0$ . Then  $S_a \xrightarrow{1}_{H'} \varepsilon = u$ .

## Grammar equivalence

**Induction step** :  $|u| = n + 1$ .

The first letter of  $u$  must be an  $a$ . Let

$$u_1 := \min_{\preceq} \{u' \preceq u \mid u' \neq \varepsilon \text{ and } \|u'\| = 0\}$$

NB :  $u_1$  exists because the set in the rhs is non-empty.

This word  $u_1$  must end with a letter  $b$  (otherwise it would not be minimal). Hence  $u_1 = au'_1b$  and

$$u = au'_1bu_2.$$

since  $\|u_1\| = 0$  and  $\|u\| = 0$  we get that  $\|u_1\| = \|u_2\| = 0$ .

Moreover  $\forall v \preceq u'_1, \|v\| \geq 0$ , and  $\forall v \preceq u_2, \|au'_1bv\| \geq 0$ , hence  $\|v\| \geq 0$ . By (IH), it follows that :

$$S_a \xrightarrow{*}_{H'} u'_1 \text{ and } S_a \xrightarrow{*}_{H'} u_2$$

hence  $S_a \rightarrow aS_abS_a \xrightarrow{*}_{H'} au'_1bu_2 = u$ .  $\square$



## Grammar equivalence

## Lemma

Let  $u \in A^*$ . If  $\|u\| = 0$  then  $S \xrightarrow{*}_{H'} u$ .

**Proof :**

By induction on  $|u|$ .

**Basis :**  $|u| = 0$ .

Then  $S \xrightarrow{1}_{H'} \varepsilon$ .

**Induction step :**  $|u| = n + 1$ .

Let

$$u_1 := \min_{\preceq} \{u' \preceq u \mid u' \neq \varepsilon \text{ and } \|u'\| = 0\}$$

By minimality, all prefixes of  $u_1$  must have a norm with the same sign.

## Grammar equivalence

**Case 1** :  $\forall v \preceq u_1, \|v\| \geq 0$ .

Then  $u_1 = au'_1b$  with  $\|u'_1\| = 0$  and  $\forall v \preceq u'_1, \|v\| \geq 0$ . Then  $u = au'_1bu_2$ . By the lemma above,

$$S_a \xrightarrow{*}_{H'} u'_1.$$

Moreover  $\|u_2\| = \|u\| - \|au'_1b\|$  showing that  $\|u_2\| = 0$ . By (IH)

$$S \xrightarrow{*}_{H'} u_2.$$

The two above derivations entail :

$$S \xrightarrow{1}_{H'} aS_a bS \xrightarrow{*}_{H'} au'_1bu_2 = u.$$

**Case 2** :  $\forall v \preceq u_1, \|v\| \leq 0$ .

By a similar reasoning

$$S \xrightarrow{1}_{H'} bS_b aS \xrightarrow{*}_{H'} u.$$



# Grammar equivalence

We have proved :

$H$  is **non-ambiguous**.

$$L(G_2, S) = L(H, S) = L(H', S) = L_{=}$$

# Grammar equivalence

Recall  $G_1 = \langle A, N, R, S \rangle$  with

$$A = \{a, b\}, \quad N = \{S\}, \quad S \longrightarrow aSS \mid b,$$

## Exercise

*Prove that*

$$L(G_1, S) = \{u \in \{a, b\}^* \mid \|u\| = -1 \text{ and } \forall v \prec u, \|v\| \geq 0\}$$

# Grammar equivalence

Recall  $G_2 = \langle A, N, R'_2, S \rangle$  with

$$A = \{a, b\}, \quad N = \{S\}, \quad S \longrightarrow aSbS \mid bSaS \mid \epsilon$$

## Exercice

Prove that  $G_2$  is *ambiguous*.

# Algorithms for non-ambiguity or grammar equivalence

We gave proof-methods. But are there **algorithms** solving the problems :

- INPUT : a c.f. grammar  $G = \langle A, N, R, S \rangle$   
QUESTION : is  $G$  **ambiguous**?
- INPUT : two c.f. grammars  
 $G_1 = \langle A, N_1, R_1, S_1 \rangle$ ,  $G_2 = \langle A, N_2, R_2, S_2 \rangle$   
QUESTION :  $L(G_1, S_1) = L(G_2, S_2)$ ?

Answer : **NO**, such algorithms **do not exist**.

# Algorithms for non-ambiguity or grammar equivalence

- There exist some **programs** that try to detect ambiguity and **succeed on many** grammars ([S. Schmitz, 2006] for example)
- There are subclasses of c.f. grammars : the simple grammars, the LL(k) grammars, the LR(k) grammars, that are **non-ambiguous**; the condition “to be simple”, “to be LL(k)” or “to be LR(k)” on grammars are **testable**.
- For the above classes of grammars, the equivalence-problem **can be tested** [G. Sénizergues 2002], **program** LALBLC on my web-page).