

# Langages Formels

- **Anne Gazon**

- tel 02 23 23 **39 35**
- bureau D174 niveau orange

- **Gilles Lesventes**

- tel 02 23 23 **39 39**
- bureau de la direction ISTIC

- **Catherine Belleannée**

- tel 02 99 84 **73 20**
- bureau A106 niveau orange

- 10 séances de CM et 10 séances de TD
- Ce cours est un “prérequis” à
  - **Compilation** (licence L3 au S6)
  - Compilation (Master 1)
- Il est relativement lié à LOG (L3 au S6)
- Il est relativement lié à AGR1 (L3 en ce moment)

**Documents à disposition** (dont syllabus, diapos de cours)

- depuis une salle de TP de l'ISTIC : /share/l3info/lf
- de l'extérieur : <http://etudiant.istic.univ-rennes1.fr/current/l3info/lf>

# Évaluation

- **Contrôle continu (CC)**

sans document, a priori séances 4 et 8 de TD

- **Terminal (T)**

documents de cours et TD autorisés exclusivement,  
dont le poly n°94 (gratuit, pdf en ligne : <http://ndc.istic.univ-rennes1.fr/>)

- **Session 1**

$$(2*T + CC)/3$$

- **Session 2**

$$\text{MAX}(T, (2*T + CC)/3)$$

# Prérequis mathématiques

- ensembles
- produit cartésien
- union
- intersection
- relations (relations d'ordre et relations d'équivalence)
- raisonnement par récurrence (on fera un rappel)

# Lettres et mots

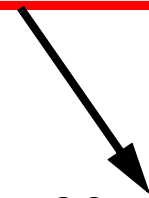
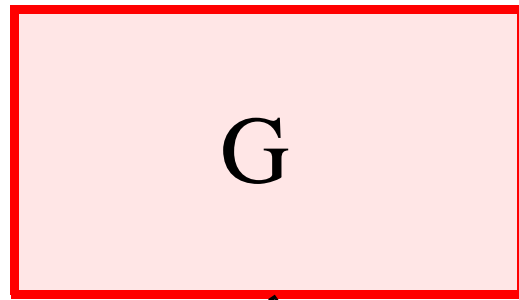
| des lettres, lexèmes                               | des mots                       |
|--|--------------------------------|
| le symbole '   '                                   | ,                              |
| 0, 1, 2, 3, 4, 5, 6, 7, 8, 9                       | 578 , 00002013                 |
| 0, 1, 2, 3, 4, 5, 6, 7, 8, 9<br>et le symbole ', ' | 3,14116                        |
| 0, ..., 9 et les symboles<br><, >, +, /, =, ... ,  | $3 < 5+5$ , $3 > 5+5$          |
| •, — et le symbole espace                          | • — — • — • • (anne)           |
| - , -, -   | -----                          |
| e, i, f, l, s, (, ), {, }                          | if ( ) { if ( ) { } else { } } |
| if, else, (, ), {, }                               | if ( ) { if ( ) { } else { } } |

# Mots et langages

| des mots                       | des langages  |
|--------------------------------|---|
|                                | l'ensemble des suites non vides de bâtons (B)                         |
| 578 ,      00002013            | l'ensemble des entiers naturels (du moins un codage des entiers) (IN) |
| 3,14116                        | l'ensemble des nombres décimaux (ID)                                  |
| $3 < 5+5$ , $3 > 5+5$          | l'ensemble des inégalités arithmétiques (I)                           |
| .   —     — .     — .     .    | le Morse (M)  |
| - - - - -<br>- - - - -         | "les drapeaux en couleurs" (C)  |
| if ( ) { if ( ) { } else { } } | l'ensemble des programmes <b>Java</b> syntaxiquement corrects (J)     |

## 2 mécanismes

**(a) engendrer un langage : grammaire**



32, 256, 00125, .....

# Une grammaire

<phrase> → <groupe-nominal> <verbe> <groupe-nominal>

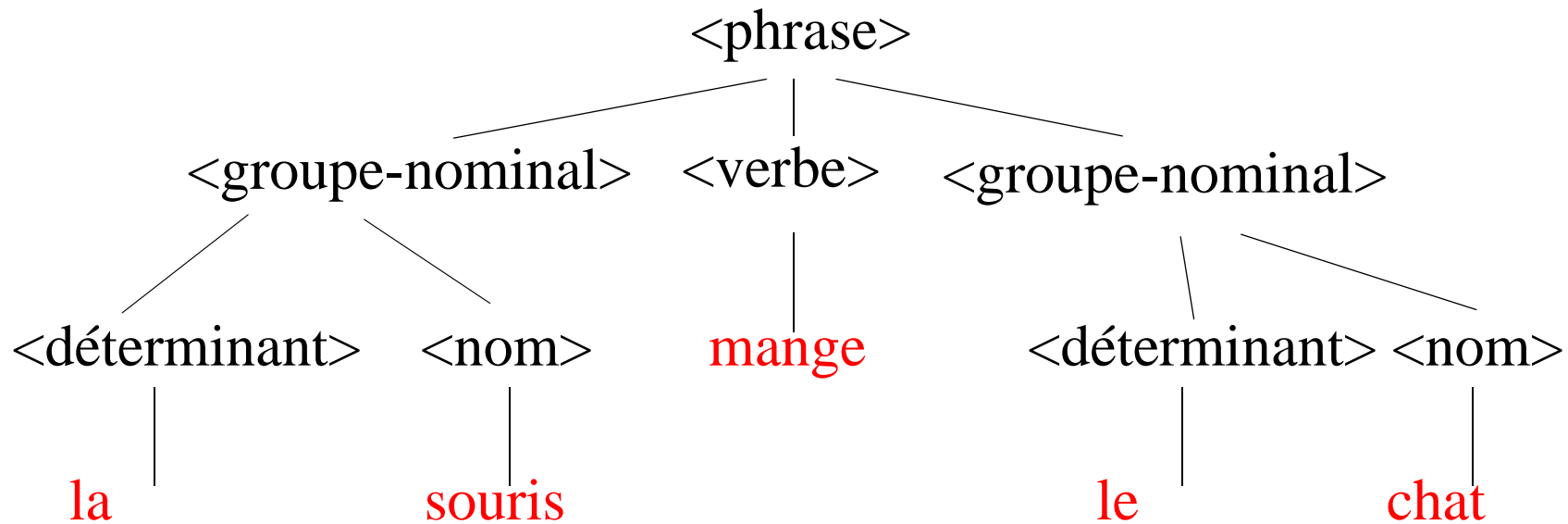
<phrase> → <groupe-nominal> <verbe> <groupe-nominal> qui <verbe>  
<groupe-nominal>

<groupe-nominal> → <déterminant> <nom>

<déterminant> → le | la | un | une

<nom> → chat | souris | balle

<verbe> → mange | regarde | frappe

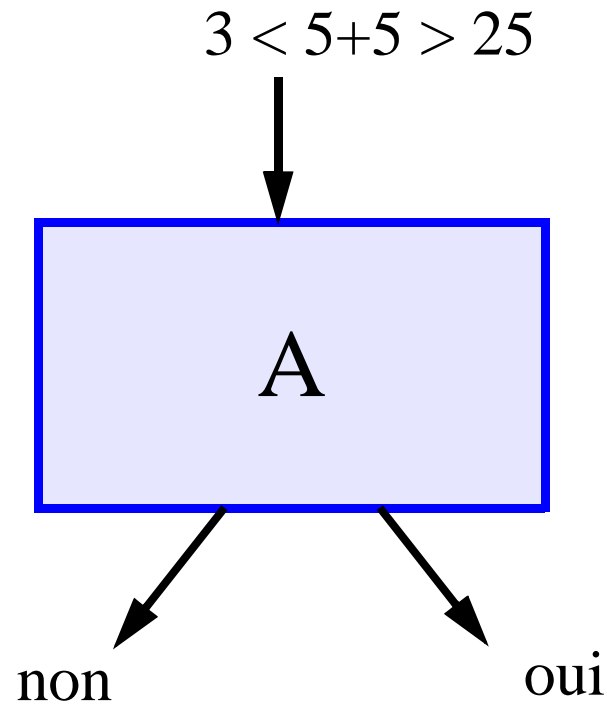


Le chat mange la souris qui regarde une balle

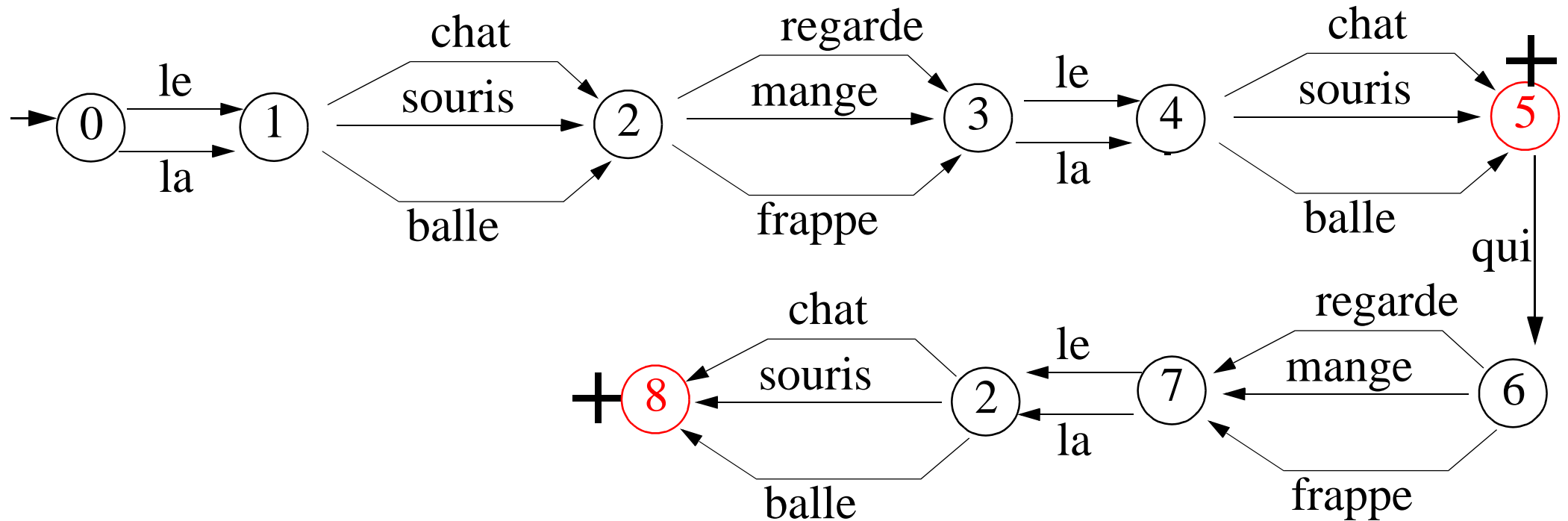


## 2 mécanismes

### (b) reconnaître un langage : automate



# Un automate



le chat regarde la souris : correct

la balle mange le souris qui frappe la chat : correct

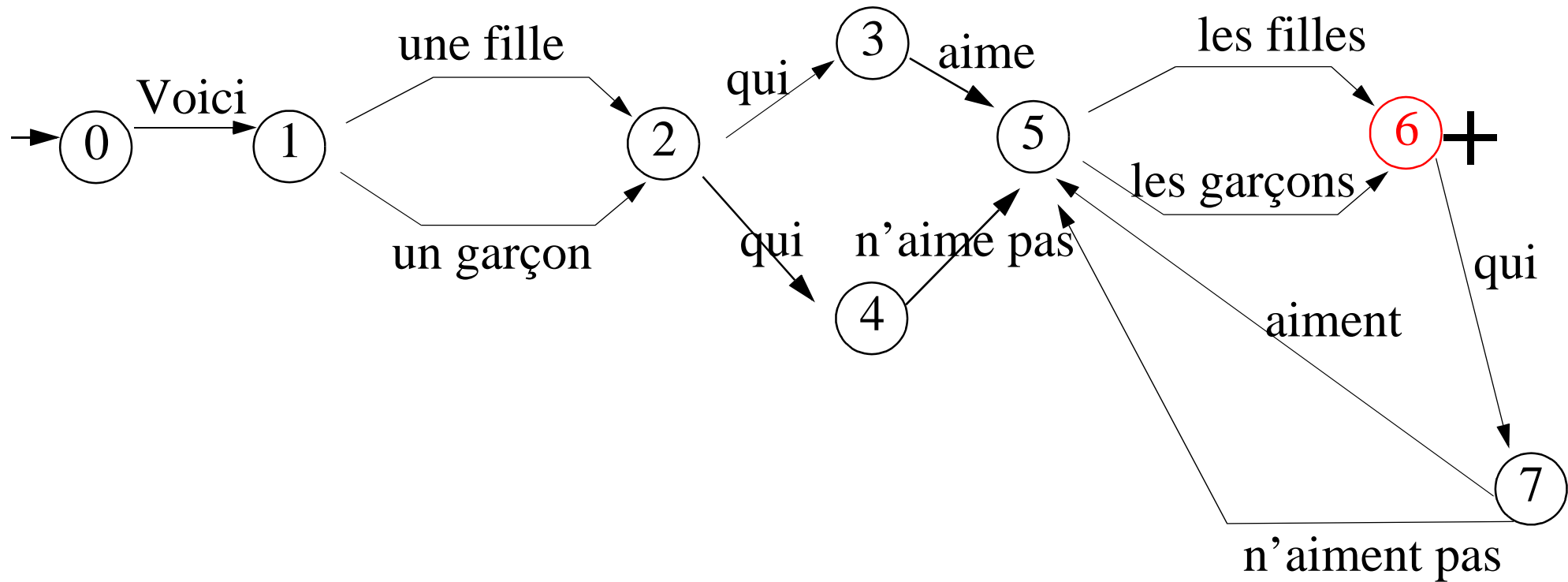
chat regarde souris : incorrect

le chat mange : incorrect

le chat regarde le chat qui regarde le chat qui regarde le chat : incorrect

Rq : c'est le même langage que dans l'exemple précédent

# Campagne de lutte contre l'homophobie par le ministère de l'enseignement supérieur



Attention : choix de l'alphabet à revoir



**VOICI UNE FILLE  
QUI AIME LES FILLES.  
MAIS CETTE FILLE  
QUI AIME LES FILLES  
N'AIME PAS  
LES FILLES  
QUI N'AIMENT PAS  
LES FILLES  
QUI AIMENT  
LES FILLES.**

**CETTE PHRASE EST COMPLIQUÉE MAIS  
MOINS QUE SA VIE D'ÉTUDIANTE HOMOSEXUELLE.**

© Junium - M3570pcom oct 08

**L'HOMOPHOBIE MÈNE À L'EXCLUSION ET AU REJET**

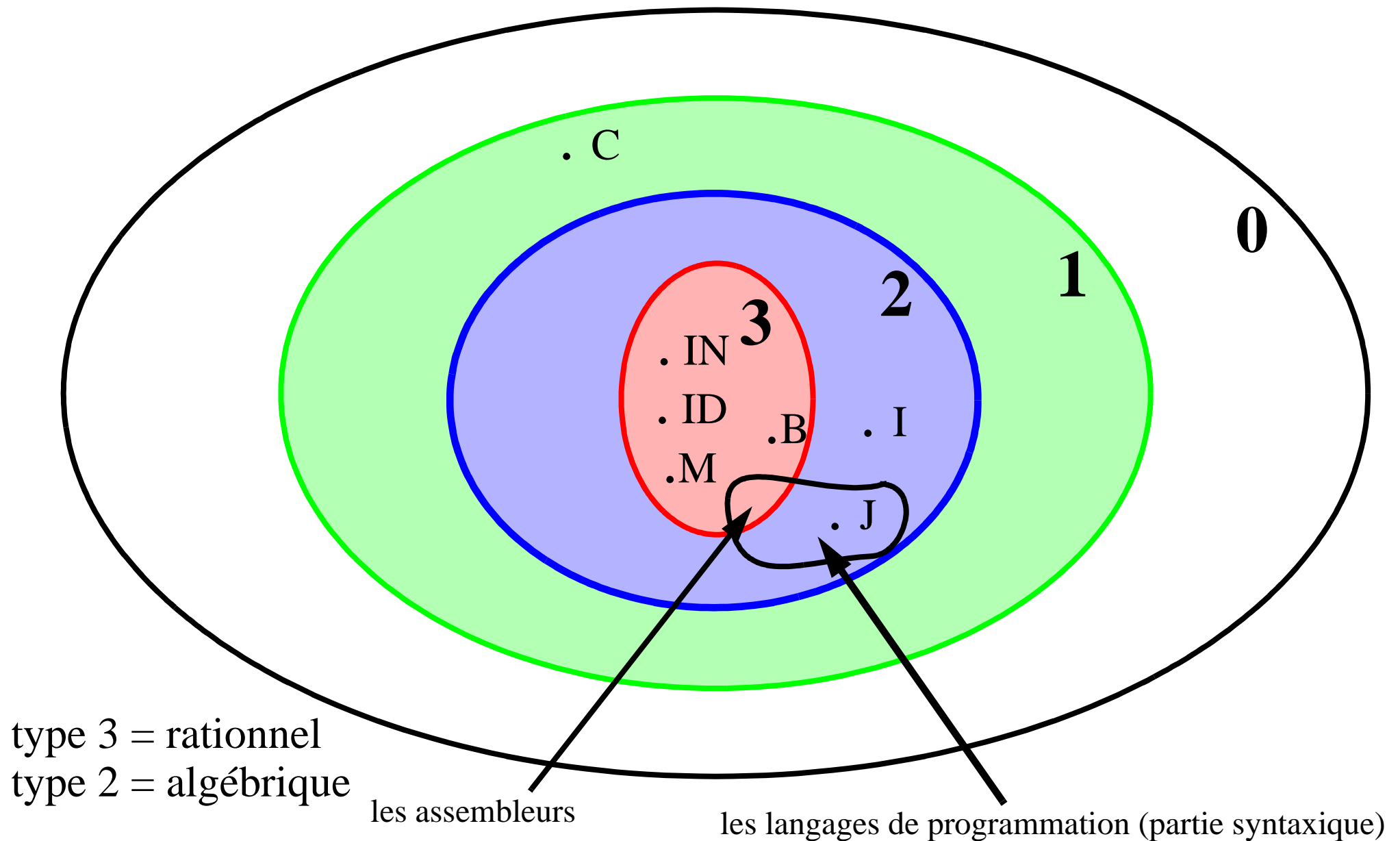
Les actes et les comportements homophobes peuvent avoir des conséquences désastreuses pour ceux qui en sont victimes.

**LUTTONS ENSEMBLE CONTRE L'HOMOPHOBIE À L'UNIVERSITÉ**



MINISTÈRE  
DE L'ENSEIGNEMENT SUPÉRIEUR  
ET DE LA RECHERCHE

# Chomsky : quatre types de langages...



# Quelques définitions

- alphabet
- mot
- longueur
- mot vide
- $X^*$
- concaténation, facteur
- monoïde

# Fait fondamental

(définit l'égalité dans le monoïde libre)

- Soient  $u = a_1 a_2 \dots a_p$  et  $v = b_1 b_2 \dots b_q$  deux mots de  $X^*$ , où les  $a_i$  et les  $b_j$  sont des **lettres** de  $X$ ,

on a  $\mathbf{u = v}$  si et seulement si

-  $\mathbf{p = q}$

et

- pour tout  $i \in \{1, 2, \dots, p\}$ ,  $\mathbf{a_i = b_i}$

- ce fait découle immédiatement de la définition d'un mot.

# Quelques définitions

- facteur
- facteur gauche, droit
- facteur propre
- factorisation
- occurrence



# Factorisations de abaab

- le mot abaab admet 6 factorisations de la forme  $u.v$  :
  - $\epsilon . abaab$
  - $a . baab$
  - $ab . aab$
  - $aba . ab$
  - $abaa . b$
  - $abaab . \epsilon$
- le mot abaab admet
  - 6 facteurs gauches:  $\epsilon , a , ab, \dots$
  - 12 facteurs:  $\epsilon , a , b , aa , ab, ba , aab , aba , baa , abaa , baab , abaab$
- $ba$  est facteur de abaab, mais ni facteur droit, ni facteur gauche.
- $|abaab|_a = 3$       et       $|abaab| = 5$

# Opérations sur les langages

- un **langage** est un ensemble fini ou infini de mots sur un alphabet  $X$
- union  $L \cup M$
- intersection  $L \cap M$
- complémentation  $X^* \setminus L$  ou  $\bar{L}$
- produit  $L . M$
- étoile  $L^*$  cas particulier  $L=X$
- étoile propre  $L^+$

# Propriétés des opérations sur les langages

- $\emptyset^* = \{ \epsilon \}$
- élément neutre :  $\{ \epsilon \} . L = L . \{ \epsilon \} = L$
- élément absorbant :  $\emptyset . L = L . \emptyset = \emptyset$
- idempotence de l'étoile :  $(L^*)^* = L^*$
- distributivité :  
$$L . (M \cup N) = (L . M) \cup (L . N)$$
$$(L \cup M) . N = (L . N) \cup (M . N)$$
- etc...

# Grammaire algébrique (ou de type 2)

- Une grammaire **algébrique** est un quadruplet  $G = \langle X, V, S, P \rangle$  où :
  - $X$  est un alphabet, dit **terminal**
  - $V$  est un alphabet disjoint de  $X$ , dit **non-terminal**
  - $S \in V$  est **l'axiome** de  $G$
  - $P \subset V \times (X \cup V)^*$  est l'ensemble (toujours **fini**) des **règles** de production

# Exemple de grammaire algébrique

- $G_1 = \langle X, V, S, P \rangle$  avec :

$$- X = \{ a, b \}$$

$$- V = \{ S, T \}$$

$$- P = \{ (S, aSb), (S, aT), (T, b) \}$$

Autre notation pour P :

$$S \rightarrow aSb$$

$$S \rightarrow aT$$

$$T \rightarrow b$$

ou encore :

$$S \rightarrow aSb + aT$$

$$T \rightarrow b$$

# Grammaires linéaires

- Une grammaire **algébrique**  $\langle X, V, S, P \rangle$  est dite :

- **linéaire** si  $P \subset V \times (X^* V X^* \cup X^*)$

exemple :  $G_1$  avec  $P = \{ (S, aSb), (S, aT), (T, b) \}$

- linéaire **droite** si  $P \subset V \times (X^* V \cup X^*)$

- linéaire **gauche** si  $P \subset V \times (V X^* \cup X^*)$

# Dérivations dans une grammaire algébrique

- la relation “... se dérive directement en ...”
  - sur  $(X \cup V)^*$ , on définit  $f \rightarrow h$  ssi  $f = u g v$ ,  $h = u d v$  et  $(g, d) \in P$
- la relation “... se dérive en...”
  - $\rightarrow^*$  est la fermeture réflexive et transitive de  $\rightarrow$
- notation  $f \rightarrow^n h$  (n est l'ordre de dérivation)
- remarque :  $f \rightarrow^0 f$  et donc  $f \rightarrow^* f$

# Langage engendré par une grammaire algébrique

soit  $G = \langle X, V, S, P \rangle$

-  $L_G(S) = \{ f \in X^*, S \rightarrow^* f \}$

- langage élargi :  $\tilde{L}_G(S) = \{ f \in (X \cup V)^*, S \rightarrow^* f \}$

- langage algébrique

- famille des langages algébriques :  $\text{Alg}(X^*)$  et  $\text{Alg}$



# Lemme fondamental (v1)

- Soit  $G = \langle X, V, S, P \rangle$  une grammaire algébrique et  $f \in (X \cup V)^*$ .

Si l'on factorise  $f$  en

$$f = f_0 S_1 f_1 S_2 \dots f_{k-1} S_k f_k, \quad k \geq 1, \text{ avec } \forall i, f_i \in X^* \text{ et } S_i \in V,$$

alors pour tout mot  $g \in (X \cup V)^*$ ,

$f \rightarrow^* g$  si et seulement si il existe des mots  $h_1, h_2, \dots, h_k \in (X \cup V)^*$  tels que

$$g = f_0 h_1 f_1 h_2 \dots f_{k-1} h_k f_k, \text{ avec } \forall i, 1 \leq i \leq k, S_i \rightarrow^* h_i$$

# Lemme fondamental (v2)

- Soit  $G = \langle X, V, S, P \rangle$  une grammaire algébrique et  $f \in (X \cup V)^*$ .

Si l'on factorise  $f$  en

$$f = f_0 S_1 f_1 S_2 \dots f_{k-1} S_k f_k, \quad k \geq 1, \text{ avec } \forall i, f_i \in X^*, S_i \in V, \text{ alors}$$

**pour tout entier  $n \geq 0$**  et pour tout mot  $g \in (X \cup V)^*$ ,

$f \xrightarrow{n} g$  si et seulement si il existe des entiers  **$n_1, n_2, \dots, n_k \geq 0$**

et des mots  $h_1, h_2, \dots, h_k \in (X \cup V)^*$  tels que  $n = n_1 + n_2 + \dots + n_k$ ,

$g = f_0 h_1 f_1 h_2 \dots f_{k-1} h_k f_k$  et  $\forall i, 1 \leq i \leq k, S_i \xrightarrow{n_i} h_i$ .

- ce lemme se montre par récurrence sur  $n$

