

# Théorie des langages : THL

## CM 9

Uli Fahrenberg

EPITA Rennes

S5 2024

# Aperçu

# Programme du cours

- ① Langages rationnels, automates finis
  - TP 1 : flex
- ③ Langages algébriques, grammaires hors-contexte, automates à pile
- ④ Parsage LL
  - TP 2 : LL
- ⑥ Parsage LR
- ⑨ Conclusion
  - TP 3 : bison
  - TP 4 : flex & bison

# Résumé du cours

# Hiérarchie de Chomsky

type	langages	grammaires	automates
4	finis ⋈	à choix finis ⋈	finis acycliques ⋈
3	rationnels ⋈	régulières ⋈	finis ⋈
2	algébriques ⋈	hors-contexte ⋈	à pile
1	contextuels ⋈	contextuelles ⋈	linéairement bornés ⋈
0	rékursivement énumérables	syntagmatiques	de Turing

# Hiérarchie de Chomsky

type	langages	grammaires	automates
4	finis ⋈	à choix finis ↓	finis acycliques ↓
3	rationnels ⋈	régulières ↓	finis ↓
2	algébriques ⋈	hors-contexte ↓	à pile
1	contextuels ⋈	contextuelles ↓	linéairement bornés ↓
0	récurivement énumérables	syntagmatiques	de Turing

# Dans le poly

- ① Langages rationnels, automates finis
  - chapitre 2 **sauf** 2.3.2, 2.3.3, 2.3.4, 2.3.5, 2.4.4
  - chapitre 3 **sauf** 3.1.3
  - chapitre 4 **sauf** 4.1.3, 4.2.1, 4.3, 4.4
- ③ Langages algébriques, grammaires hors-contexte, automates à pile
  - chapitre 6 **sauf** 6.3.1
  - 9.2.2
  - Sipser 2.2
- ④ Parsage LL
  - chapitre 7
  - section 8.1
- ⑥ Parsage LR
  - section 8.2

# Rationnel vs. algébrique

## langages rationnels

**grammaires** régulières :

- linéaire à droite :  $N \rightarrow \Sigma N \mid \Sigma \mid \varepsilon$
- linéaire à gauche :  $N \rightarrow N \Sigma \mid \Sigma \mid \varepsilon$

**automates** finis

- déterministes / non-déterministes

**décidabilité** :

- appartenance ( $w \in L$ ) ✓
- vacuité ( $L = \emptyset$ ) ✓
- universalité ( $L = \Sigma^*$ ) ✓

## langages algébriques

**grammaires** hors contexte :

- $N \rightarrow (N \cup \Sigma)^*$
- Greibach :  $N \rightarrow \Sigma N N \mid \Sigma N \mid \Sigma \mid \varepsilon$

**automates** à pile

- pas de déterminisation

**décidabilité** :

- appartenance ( $w \in L$ ) ✓
- vacuité ( $L = \emptyset$ ) ✓
- universalité ( $L = \Sigma^*$ ) ✗



# Zoom sur type 3 (régulier / rationnel)

## syntaxe

automates finis dét. complets

$\cap$

automates finis déterministes

$\cap$

automates finis

$\cap$

aut. finis à transitions spontanées

---

expressions rationnelles

---

grammaires régulières

$L(\cdot)$   
→

## sémantique

langages reconnaissables

$\parallel$

langages reconnaissables

$\parallel$

langages reconnaissables

$\parallel$

langages reconnaissables

$\parallel$

langages rationnelles

$\parallel$

langages réguliers

# Zoom sur type 2 (hors contexte / algébrique)

## syntaxe

## sémantique

grammaires hc forme Greibach

 $\cap$ 

grammaires hors-contexte

 $\cup$ 

grammaires hc forme Chomsky

 $\xrightarrow{L(\cdot)}$ 

langages algébriques

 $\parallel$ 

langages algébriques

 $\parallel$ 

langages algébriques

 $\parallel$ 

langages algébriques

 $\parallel$ 

langages algébriques

 $\cup$ 

automates à pile

 $\cup$ 

automates à pile sans transitions spont.

 $\cup$ 

automates à pile déterministes

langages algébriques déterministes

## Zoom sur LR

syntaxesémantique

grammaires hors-contexte  
 $\cup$   
 grammaires hc non-ambiguës  
 $\cup$   
 grammaires hc déterministes  
 $\cup$   
 grammaires LR( $k$ )  
 $\cup$   
 $\vdots$   
 $\cup$   
 grammaires LR(1)  
 $\cup$   
 grammaires LALR(1)  
 $\cup$   
 grammaires SLR(1)  
 $\cup$   
 grammaires LR(0)

 $L(\cdot)$   
 $\longrightarrow$ 

langages algébriques  
 $\cup$   
 lang. algébriques non-ambiguës  
 $\cup$   
 lang. algébriques déterministes  
 $\parallel$   
 lang. algébriques déterministes  
 $\parallel$   
 $\vdots$   
 $\parallel$   
 lang. algébriques déterministes  
 $\cup$   
 langages LALR(1)  
 $\cup$   
 langages SLR(1)  
 $\cup$   
 langages LR(0)

# Zoom sur LL

## syntaxe

grammaires hors-contexte

$\cup$

grammaires hc non-ambiguës

$\cup$

grammaires hc déterministes

$\cup$

grammaires  $LL(k)$

$\cup$

$\dots$

$\cup$

grammaires  $LL(2)$

$\cup$

grammaires  $LL(1)$

$L(\cdot)$   
 $\longrightarrow$

## sémantique

langages algébriques

$\cup$

lang. algébriques non-ambiguës

$\cup$

lang. algébriques déterministes

$\cup$

langages  $LL(k)$

$\cup$

$\dots$

$\cup$

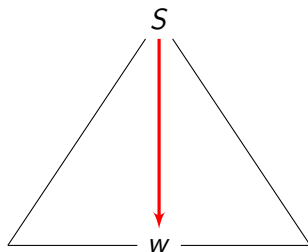
langages  $LL(2)$

$\cup$

langages  $LL(1)$

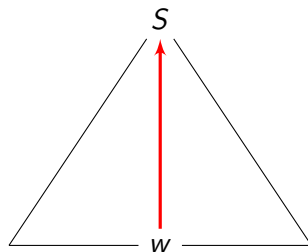
# Parsage

# Passage



descendante

↑  
LL



ascendante

↑  
LR

# Parcours LL(1) : approche descendante

- ① entrée : une grammaire hors contexte  $G = (N, \Sigma, P, S)$ 
  - si-dessous,  $V = N \cup \Sigma$
  - éliminer récursion à gauche dans  $G$  ; factoriser  $G$  à gauche
- ② calculer NULL
  - $\text{NULL} = \{A \in N \mid A \Rightarrow^* \varepsilon\}$
- ③ construire la table FIRST
  - $\text{FIRST}(A) = \{a \in \Sigma \mid \exists w \in V^* : A \Rightarrow^* aw\}$
- ④ construire la table FOLLOW
  - $\text{FOLLOW}(A) = \{a \in \Sigma \mid \exists B \in N, \alpha, \beta \in V^* : B \Rightarrow^* \alpha A a \beta\}$
- ⑤ construire la TABLE de parcours :
  - ① pour chaque production  $X \rightarrow w$  ( $n$ ) :
    - ① pour chaque  $a \in \text{FIRST}(w)$  :  $\text{TABLE}(X, a) += \{n\}$
    - ② si  $w \in \text{NULL}$  ou  $w = \varepsilon$  :
      - pour chaque  $a \in \text{FOLLOW}(X)$  :  $\text{TABLE}(X, a) += \{n\}$

# Re : passage ascendant : the basics

```
function BULRP( $\alpha$ )  
  if  $\alpha = S$  then  
    return True  
  for  $i \leftarrow 1$  to  $|\alpha|$  do  
    for  $j \leftarrow i$  to  $|\alpha|$  do                                ▷ décalage / SHIFT  
      for  $A \in N$  do  
        if  $A \rightarrow \alpha_i \dots \alpha_j$  then                            ▷ réduction / REDUCE  
          if BULRP( $\alpha_1 \dots \alpha_{i-1} A \alpha_{j+1} \dots \alpha_n$ ) then  
            return True  
  return False
```

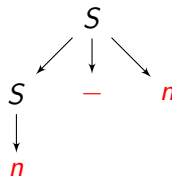


## Re : passage LR(0)

 $Z \rightarrow S\$$  (0) $S \rightarrow S-n$  (1) $| n$  (2)parser  $n - n\$$  :

entrée	pile	action
$n - n\$$	$\perp 0$	décaler
$-n\$$	$\perp 01$	réduire 2
$-n\$$	$\perp 02$	décaler
$n\$$	$\perp 024$	décaler
$\$$	$\perp 0245$	réduire 1
$\$$	$\perp 02$	décaler
	$\perp 023$	✓

état	action	$n$	$-$	$\$$	$S$
0	décaler	1			2
1	réduire 2				
2	décaler		4	3	
3	accepter				
4	décaler	5			
5	réduire 1				

 $S \rightarrow n$  $S \rightarrow S-n$ 

# Re : parsage SLR(1)

- ① calculer la table LR(0)
- ② si conflits : **conditionner** l'action par le FOLLOW

Exemple :  $Z \rightarrow S\$$  (0)  
 $S \rightarrow n-S$  (1)  
 $\quad \quad | n$  (2)

état	action	$n$	$-$	$\$$	$S$		état	$n$	$-$	$\$$	$S$
0	décaler	2			1		0	d.2			d.1
1	décaler			4			1			d.4	
2	réd. 2, déc.		3			$\Rightarrow$	2		d.3	r.2	
3	décaler	2			5		3	d.2			d.5
4	accepter						4		— accepter —		
5	réduire 1						5			r.1	

# Re : passage LR(1)

- conditionner l'action par le **contexte** : les symboles qui peuvent suivre

Exemple :

$Z \rightarrow S\$$  (0)

$S \rightarrow L=E$  (1)

$| E$  (2)

$L \rightarrow x$  (3)

$| *E$  (4)

$E \rightarrow L$  (5)

état	productions pointées élargies
0	$Z \rightarrow \bullet S\$ [\varepsilon]$ $S \rightarrow \bullet L=E [\$], S \rightarrow \bullet E [\$]$ $L \rightarrow \bullet x [=], L \rightarrow \bullet *E [=]$ $E \rightarrow \bullet L [\$]$
1	$Z \rightarrow S \bullet \$ [\varepsilon]$ $S \rightarrow L \bullet =E [\$], E \rightarrow L \bullet [\$ \checkmark]$
2	

## Exemple

	état	x	*	=	\$	S	L	E
	0	d.4	d.5			d.1	d.2	d.3
	1				d.6			
	2			d.7	r.5			
	3				r.2			
$Z \rightarrow S\$$ (0)	4			r.3	r.3			
$S \rightarrow L=E$ (1)	5	d.4	d.5				d.9	d.8
$  E$ (2)	6			— accepter —				
$L \rightarrow x$ (3)	7	d.12	d.13				d.11	d.10
$  *E$ (4)	8			r.4				
	9			r.5				
$E \rightarrow L$ (5)	10				r.1			
	11				r.5			
	12				r.3			
	13	d.12	d.13				d.11	d.14
	14				r.4			

# Parse LALR(1)

## Définition

Deux productions pointées élargies  $A \rightarrow \alpha \bullet \beta [a]$  et  $A \rightarrow \alpha' \bullet \beta' [b]$  sont **équivalent LALR(1)** si  $\alpha = \alpha'$  et  $\beta = \beta'$ .

- les **items** sont identiques, mais les contextes peuvent être différents

## Définition

L'**automate LALR(1)** d'une grammaire hors-contexte  $G$  est le quotient de l'automate LR(1) de  $G$  sous équivalence LALR(1).

## Exemple, re

	état	$\times$	$*$	$=$	$\$$	$S$	$L$	$E$
	0	d.4	d.5			d.1	d.2	d.3
$Z \rightarrow S\$$ (0)	1				d.6			
$S \rightarrow L=E$ (1)	2			d.7	r.5			
$  E$ (2)	3				r.2			
	4			r.3	r.3			
$L \rightarrow \times$ (3)	5	d.4	d.5				d.9	d.8
$  *E$ (4)	6			— accepter —				
$E \rightarrow L$ (5)	7	d.12	d.13				d.11	d.10
	8			r.4				
	9			r.5				
	10				r.1			
	11				r.5			
	12				r.3			
	13	d.12	d.13				d.11	d.14
	14				r.4			

## Exemple, re

	état	$\times$	$*$	$=$	$\$$	$S$	$L$	$E$
	0	d.4	d.5			d.1	d.2	d.3
$Z \rightarrow S\$$ (0)	1				d.6			
$S \rightarrow L=E$ (1)	2			d.7	r.5			
$  E$ (2)	3				r.2			
	4			r.3	r.3			
$L \rightarrow \times$ (3)	5	d.4	d.5				d.9	d.8
$  *E$ (4)	6			— accepter —				
$E \rightarrow L$ (5)	7	d.12	d.13				d.11	d.10
	8			r.4	r.4			
	9			r.5	r.5			
	10				r.1			

# Résolution de conflits

Exemple :

$Z \rightarrow E\$$  (0)

$E \rightarrow E+E$  (1)

|  $E * E$  (2)

|  $n$  (3)

état	+	*	$n$	\$	E
0			d.2		g.1
1	d.4	d.5		d.3	
2	r.3	r.3		r.3	
3			— accepter —		
4			d.2		g.6
5			d.2		g.7
6	d.4	d.5			
	r.1	r.1		r.1	
7	d.4	d.5			
	r.2	r.2		r.2	

- une grammaire **ambiguë**
- donc pas LR( $k$ ) pour n'importe quel  $k$



# Résolution de conflits

Exemple :

$$Z \rightarrow E\$ \quad (0)$$

$$E \rightarrow E+E \quad (1)$$

$$| E * E \quad (2)$$

$$| n \quad (3)$$

état	+	*	n	\$	E
0			d.2		g.1
1	d.4	d.5		d.3	
2	r.3	r.3		r.3	
3			— accepter —		
4			d.2		g.6
5			d.2		g.7
6	d.4	d.5			
	r.1	r.1		r.1	
7	d.4	d.5			
	r.2	r.2		r.2	

- une grammaire **ambiguë**
- donc pas LR( $k$ ) pour n'importe quel  $k$
- **associativité** : d.4  $\Rightarrow n + (n + n)$ ; r.1  $\Rightarrow (n + n) + n$
- **priorité** : d.5  $\Rightarrow n * (n + n)$ ; r.1  $\Rightarrow (n * n) + n$

# Résolution de conflits

Exemple :

$$Z \rightarrow E\$ \quad (0)$$

$$E \rightarrow E+E \quad (1)$$

$$| E * E \quad (2)$$

$$| n \quad (3)$$

état	+	*	n	\$	E
0			d.2		g.1
1	d.4	d.5		d.3	
2	r.3	r.3		r.3	
3			— accepter —		
4			d.2		g.6
5			d.2		g.7
6	d.4	d.5			
	r.1	r.1		r.1	
7	d.4	d.5			
	r.2	r.2		r.2	

- une grammaire **ambiguë**
- donc pas LR( $k$ ) pour n'importe quel  $k$
- **associativité** : d.4  $\Rightarrow n + (n + n)$ ; r.1  $\Rightarrow (n + n) + n$
- **priorité** : d.5  $\Rightarrow n * (n + n)$ ; r.1  $\Rightarrow (n * n) + n$
- solution : règles de **priorité**
- ici : **r.1 > d.4**, **r.2 > d.5**, **r.2 > d.4**, **d.5 > r.1**  $\Leftarrow$  !

## Exemple, complet

		état	productions pointées élargies
$Z \rightarrow E\$$ (0)	$E \rightarrow E * E$ (2)	0	$Z \rightarrow \bullet E\$ [\varepsilon], E \rightarrow \bullet E + E [\$+*]$
$E \rightarrow E + E$ (1)	$E \rightarrow n$ (3)		$E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$

état	+	*	n	\$	E
0			d.2		g.1

## Exemple, complet

					état	productions pointées élargies
$Z \rightarrow E\$$ (0) $E \rightarrow E * E$ (2)					0	$Z \rightarrow \bullet E\$ [\varepsilon], E \rightarrow \bullet E + E [\$+*]$ $E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$
$E \rightarrow E + E$ (1) $E \rightarrow n$ (3)					1	$Z \rightarrow E \bullet \$ [\varepsilon], E \rightarrow E \bullet + E [\$+*]$ $E \rightarrow E \bullet * E [\$+*]$
état	+	*	n	\$	E	
0			d.2		g.1	2
						$E \rightarrow n \bullet [\$+* \checkmark]$

## Exemple, complet

					état	productions pointées élargies
$Z \rightarrow E\$$ (0) $E \rightarrow E * E$ (2)					0	$Z \rightarrow \bullet E\$ [\varepsilon], E \rightarrow \bullet E + E [\$+*]$
$E \rightarrow E + E$ (1) $E \rightarrow n$ (3)					1	$E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$
					2	$Z \rightarrow E \bullet \$ [\varepsilon], E \rightarrow E \bullet + E [\$+*]$
					2	$E \rightarrow E \bullet * E [\$+*]$
					2	$E \rightarrow n \bullet [\$+* \checkmark]$
état	+	*	n	\$	E	
0			d.2		g.1	
1	d.4	d.5		d.3		

## Exemple, complet

					état	productions pointées élargies
$Z \rightarrow E\$$ (0) $E \rightarrow E * E$ (2)					0	$Z \rightarrow \bullet E\$ [\varepsilon], E \rightarrow \bullet E + E [\$ + *]$ $E \rightarrow \bullet E * E [\$ + *], E \rightarrow \bullet n [\$ + *]$
$E \rightarrow E + E$ (1) $E \rightarrow n$ (3)					1	$Z \rightarrow E \bullet \$ [\varepsilon], E \rightarrow E \bullet + E [\$ + *]$ $E \rightarrow E \bullet * E [\$ + *]$
état	+	*	n	\$	E	
0			d.2		g.1	2 $E \rightarrow n \bullet [\$ + * \checkmark]$
1	d.4	d.5		d.3		3 $Z \rightarrow E \$ \bullet [\$ \checkmark]$
					4	$E \rightarrow E + \bullet E [\$ + *], E \rightarrow \bullet E + E [\$ + *]$ $E \rightarrow \bullet E * E [\$ + *], E \rightarrow \bullet n [\$ + *]$
					5	$E \rightarrow E * \bullet E [\$ + *], E \rightarrow \bullet E + E [\$ + *]$ $E \rightarrow \bullet E * E [\$ + *], E \rightarrow \bullet n [\$ + *]$

## Exemple, complet

					état	productions pointées élargies
$Z \rightarrow E\$$ (0) $E \rightarrow E * E$ (2)					0	$Z \rightarrow \bullet E\$ [\varepsilon], E \rightarrow \bullet E + E [\$+*]$ $E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$
$E \rightarrow E + E$ (1) $E \rightarrow n$ (3)					1	$Z \rightarrow E \bullet \$ [\varepsilon], E \rightarrow E \bullet + E [\$+*]$ $E \rightarrow E \bullet * E [\$+*]$
état	+	*	n	\$	E	
0			d.2		g.1	2
1	d.4	d.5		d.3		3
2	r.3	r.3		r.3		4
3			— accepter —			5
						$E \rightarrow n \bullet [\$+* \checkmark]$ $Z \rightarrow E \$ \bullet [\$ \checkmark]$ $E \rightarrow E + \bullet E [\$+*], E \rightarrow \bullet E + E [\$+*]$ $E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$ $E \rightarrow E * \bullet E [\$+*], E \rightarrow \bullet E + E [\$+*]$ $E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$

## Exemple, complet

					état	productions pointées élargies
$Z \rightarrow E\$$ (0) $E \rightarrow E * E$ (2)					0	$Z \rightarrow \bullet E\$ [\varepsilon], E \rightarrow \bullet E + E [\$+*]$
$E \rightarrow E + E$ (1) $E \rightarrow n$ (3)					1	$E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$
					2	$Z \rightarrow E \bullet \$ [\varepsilon], E \rightarrow E \bullet + E [\$+*]$
					3	$E \rightarrow E \bullet * E [\$+*]$
					4	$E \rightarrow n \bullet [\$+* \checkmark]$
					5	$Z \rightarrow E \$ \bullet [\$ \checkmark]$
état	+	*	n	\$	E	
0			d.2		g.1	
1	d.4	d.5		d.3		
2	r.3	r.3		r.3		
3			— accepter —			
4			d.2		g.6	
5			d.2		g.7	



# Exemple, complet

					état	productions pointées élargies
$Z \rightarrow E\$$ (0) $E \rightarrow E * E$ (2)					0	$Z \rightarrow \bullet E\$ [\varepsilon], E \rightarrow \bullet E + E [\$+*]$
$E \rightarrow E + E$ (1) $E \rightarrow n$ (3)					1	$E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$
					2	$Z \rightarrow E \bullet \$ [\varepsilon], E \rightarrow E \bullet + E [\$+*]$
					3	$E \rightarrow E \bullet * E [\$+*]$
					4	$E \rightarrow n \bullet [\$+* \checkmark]$
					5	$Z \rightarrow E \$ \bullet [\$ \checkmark]$
					6	$E \rightarrow E + \bullet E [\$+*], E \rightarrow \bullet E + E [\$+*]$
					7	$E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$
					8	$E \rightarrow E * \bullet E [\$+*], E \rightarrow \bullet E + E [\$+*]$
					9	$E \rightarrow \bullet E * E [\$+*], E \rightarrow \bullet n [\$+*]$
					10	$E \rightarrow E + E \bullet [\$+*], E \rightarrow E \bullet + E [\$+*]$
					11	$E \rightarrow E \bullet * E [\$+*]$
					12	$E \rightarrow E * E \bullet [\$+*], E \rightarrow E \bullet + E [\$+*]$
					13	$E \rightarrow E \bullet * E [\$+*]$

état	+	*	n	\$	E
0			d.2		g.1
1	d.4	d.5		d.3	
2	r.3	r.3		r.3	
3			— accepter —		
4			d.2		g.6
5			d.2		g.7

## Exemple, complet

					état	productions pointées élargies
$Z \rightarrow E\$$ (0) $E \rightarrow E * E$ (2)					0	$Z \rightarrow \bullet E\$$ [ $\varepsilon$ ], $E \rightarrow \bullet E + E$ [ $\$ + *$ ]
$E \rightarrow E + E$ (1) $E \rightarrow n$ (3)					1	$E \rightarrow \bullet E * E$ [ $\$ + *$ ], $E \rightarrow \bullet n$ [ $\$ + *$ ]
					2	$Z \rightarrow E \bullet \$$ [ $\varepsilon$ ], $E \rightarrow E \bullet + E$ [ $\$ + *$ ]
					3	$E \rightarrow E \bullet * E$ [ $\$ + *$ ]
					4	$E \rightarrow n \bullet$ [ $\$ + * \checkmark$ ]
					5	$Z \rightarrow E \$ \bullet$ [ $\$ \checkmark$ ]
					6	$E \rightarrow E + \bullet E$ [ $\$ + *$ ], $E \rightarrow \bullet E + E$ [ $\$ + *$ ]
					7	$E \rightarrow \bullet E * E$ [ $\$ + *$ ], $E \rightarrow \bullet n$ [ $\$ + *$ ]
état	+	*	n	\$	E	
0			d.2		g.1	
1	d.4	d.5		d.3		
2	r.3	r.3		r.3		
3			— accepter —			
4			d.2		g.6	
5			d.2		g.7	
6	d.4	d.5				
	r.1	r.1		r.1		
7	d.4	d.5				
	r.2	r.2		r.2		

# Parsage LR généralisé

- *embrace non-determinism!*
- parsage GLR : en cas de conflit, suivre tous les chemins **en parallèle**
- « parsage parallèle », « parsage Tomita »
- implémenter l'automate (non-déterministe) de parsage sans détermination
- états : productions pointées, **pas de clôture**
- algorithme en temps **exponentiel**, pas linéaire
- optimisation : partager préfixes et suffixes de piles



*That's all Folks!*