

# Cours: Théorie des langages & Compilation

## Chapitre 1: Notions fondamentales en théorie des langages

Réalisé par:

Dr. Sakka Rouis Taoufik

<https://github.com/srtaoufik/Cours-TLA-et-Compilation>

1

### Ch1: Notions fondamentales en théorie des langages

#### I. Objectif

Ce cours présente une Notions fondamentales en théorie des langages en traitant les trois aspects suivants :

- **L'aspect reconnaissance** par les automates finis, les automates à pile et les machines de Turing,
- **L'aspect génération** par les grammaires régulières, non contextuelles et contextuelles,
- **L'aspect représentation** par propriétés mesurables, définitions récursives et expressions régulières.

L'objectif est d'introduire des connaissances en théorie des langages et des automates afin de pouvoir les étendre à la description des langages de programmation et leur analyse syntaxique en vue de leur compilation.

2

## Ch1: Notions fondamentales en théorie des langages

### I. Objectif

	Langages		
Processus	Réguliers	Hors contexte	Contextuel
Représentation	Expressions régulières	Propriété mesurable + Définition récursive	Propriété mesurable + Définition récursive
Reconnaissance	Automates Finis	Automates à Pile	Machine de Turing
Génération	Grammaires Réguliers	Grammaires Non Contextuelles	Grammaires Contextuelles

3

## Ch1: Notions fondamentales en théorie des langages

### II. Vocabulaire

#### Définition 1.

**Alphabet:** Un alphabet (ou vocabulaire) est un ensemble fini, non vide de symboles. On le note généralement  $X$ .

#### Définition 2.

**Fermeture:** Soit  $X$  un alphabet. On note par  $X^*$ , l'ensemble de toutes les séquences finies de symboles de  $X$ .

**Rq le symbole  $*$**  est une fonction qui, appliquée à un ensemble non vide  $X$ , donne un autre ensemble infini  $X^*$ .

On dit que  $X^*$  est la fermeture de  $X$ .

4

## Ch1: Notions fondamentales en théorie des langages

## II. Vocabulaire

**Définition 3.**

**Chaine vide:** Une chaine vide est une chaine qui ne contient aucun symbole du vocabulaire (appelée aussi **mot vide**). Une chaine vide est un élément de  $X^*$ . On la note :  $\varepsilon$

**Définition 4.**

**Longueur d'une chaine:** La longueur d'une chaine finie  $\omega$  est le **nombre de symboles** qu'elle contient.  
On la note  $|\omega|$ .

5

## Ch1: Notions fondamentales en théorie des langages

## III. Mot

**Définition 5.**

**Mot :** Un mot  $\omega$  est une application d'un segment initial de longueur  $n$  vers le vocabulaire:  $\omega : [n] \rightarrow X$

Où  $\omega_i$  est l'image de  $i$  dans  $X$ ;  $i$  est le rang de  $\omega_i$  dans  $\omega$ ;

si  $\omega_i = a$  et  $a \in X$  alors

$\omega_i$  est une occurrence de  $a$  dans le mot  $\omega$ .

**Exercice :**

Donner les  $\omega_i$  associés aux mots suivants :

1) abba sur le vocabulaire  $\{a, b\}$

2)  $(x_1^*(x_2+x_1))$  sur le vocabulaire  $\{x_1, x_2, +, *, (, )\}$

6

## Ch1: Notions fondamentales en théorie des langages

## III. Mot

**Corrigé:**

1) Si le vocabulaire  $X = \{a, b\}$  alors dans le mot abba,

$$\omega_1 = a$$

$$\omega_2 = b$$

$$\omega_3 = b$$

$$\omega_4 = a$$

2) Si le vocabulaire est  $X = \{x_1, x_2, +, *, (, )\}$  alors dans le mot  $(x_1^*(x_2+x_1))$

$\omega_1 = ($	$\omega_2 = x_1$	$\omega_3 = *$
$\omega_4 = ($	$\omega_5 = x_2$	$\omega_6 = +$
$\omega_7 = x_1$	$\omega_8 = )$	$\omega_9 = )$

7

## Ch1: Notions fondamentales en théorie des langages

## III. Mot

**Définition 6.**

**Longueur d'un mot :** La longueur d'un mot  $\omega$  est le nombre de symboles qu'il contient, on le note  $|\omega|$ .

**Exercice :**

Quelle est la longueur des mots abba et  $\epsilon$  sur le Vocabulaire  $\{a,b\}$

**Corrigé:**

Le mot abba est de longueur 4,  $|abba| = 4$

Le mot  $\epsilon$  est de longueur 0,  $|\epsilon| = 0$

8

## Ch1: Notions fondamentales en théorie des langages

## III. Mot

**Définition 7.****Nombre d'occurrences d'un symbole dans un mot :**

Le nombre d'occurrences d'un symbole  $x$  dans un mot  $\omega$  est le nombre de fois où ce symbole apparaît dans ce mot  $\omega$ . On le note  $|\omega|_x$ .

$$|\omega| = \sum_{x \in X} |\omega|_x$$

**Exercice :**

Quel est le nombre d'occurrences de  $b$  dans les mots  $abba$  et  $\varepsilon$

**Corrigé :**

$$|abba|_b = 2$$

$$|\varepsilon|_b = 0$$

9

## Ch1: Notions fondamentales en théorie des langages

## IV. Langage

**Définition 8.**

**Langage** : On définit un langage sur un alphabet  $X$  comme un sous ensemble de  $X^*$

**Exemple:**

Si le vocabulaire est  $X = \{0, 1, 2, 3, \dots, 9\}$

$L = \{\text{représentations décimales des nombres entiers naturels}\}$

Si le vocabulaire est  $X = \{x_1, x_2, +, *, (, )\}$   $x_1 + x_2 \in X^*$

$L = \{\text{expressions arithmétiques parenthèses}\}$

$x_1$ ,  $x_2$  et  $(x_1 + x_2)$  sont 4 mots du langage  $L$

10

## Ch1: Notions fondamentales en théorie des langages

## IV. Langage

Si  $X$  est la vocabulaire du langage de programmation  $C$

$X = \{\text{main}, (, ), \#, \text{Include}, <, >, ., ;, \text{id}, \text{nb}, \dots\}$

$L = \{\text{programmes } C \text{ corrects syntaxiquement}\}$

(main) element de  $X^*$

main(){ }

Si  $X$  est le vocabulaire de la logique des propositions

$X = \{p, (, ), \rightarrow, \wedge, \vee, \neg\}$  où  $p$  désigne une proposition

$L = \{\text{formules bien formées de la logique des propositions}\}$

$p \rightarrow p$

11

## Ch1: Notions fondamentales en théorie des langages

## IV. Langage

Si  $X$  est le vocabulaire de la logique des propositions

$X = \{p, (, ), \rightarrow, \wedge, \vee, \neg\}$  où  $p$  désigne une proposition

$L = \{\text{formules bien formées de la logique des propositions}\}$

L'ensemble des mots du langage  $L$  est défini récursivement comme suit :

- 1)  $P$  est un mot
- 2) Si  $F$  et  $G$  sont deux mots alors  $F \rightarrow G$ ,  $F \wedge G$ ,  $F \vee G$  le sont aussi
- 3) Si  $F$  est un mot alors  $(F)$  et  $\neg F$  le sont aussi

12

## Ch1: Notions fondamentales en théorie des langages

## V. Opérations sur les langages

**Définition 9.****Concaténation de deux mots:**

soient  $u$  et  $v$  deux mots /  $|u| = n$  et  $|v| = m$  alors:

$$u.v = \omega \quad / \quad \omega_i = u_i \quad \forall i \in [n] \text{ et } \\ \omega_{n+j} = v_j \quad \forall j \in [m]$$

Le mot vide  $\varepsilon$  est un élément neutre de la concaténation

$$u.\varepsilon = \varepsilon.u = u$$

13

## Ch1: Notions fondamentales en théorie des langages

## V. Opérations sur les langages

Soient  $A$  et  $B$  deux langages alors on a les opérations suivantes:

<b>Intersection</b>	$A \cap B = \{\omega / \omega \in A \text{ et } \omega \in B\}$
<b>Union</b>	$A \cup B = \{\omega / \omega \in A \text{ ou } \omega \in B\}$ (notée aussi $A+B$ )
<b>Complémentation</b>	$A \setminus B = \{\omega / \omega \in A \text{ et } \omega \notin B\}$ (notée aussi $A-B$ )
<b>Concaténation</b>	$A.B = \{\omega / \exists u \in A \text{ et } \exists v \in B \text{ et } \omega = u.v\}$

**Propriétés**

Soient  $A, B, C$  des langages, on a

$$A.(B \cup C) = A.B \cup A.C$$

$$(A \cup B).C = A.C \cup B.C$$

14

## Ch1: Notions fondamentales en théorie des langages

## V. Opérations sur les langages

**Notation:**

$a^k = \text{aaaaa} \dots \text{aaaa}$   
 $\langle \dots \dots \dots \rangle$   
 k fois

$$a^2 = aa$$

$$a^0 = \epsilon$$

$$a^* = \{a^i / i \geq 0\} = \{a^0, a^1, a^2, \dots, a^i, \dots\}$$

$$a^+ = \{a^i / i > 0\} = \{a^1, a^2, \dots, a^i, \dots\}$$

**L'opération \***

$$A^* = \bigcup A^i \quad \forall i \geq 0 \quad \text{avec}$$

$$A^0 = \{\epsilon\}$$

$$A^1 = A$$

$$A^{i+1} = A.A^i \quad \forall i \geq 0$$

**L'opération +**

$$A^+ = \bigcup A^i \quad \forall i \geq 1$$

15

## Ch1: Notions fondamentales en théorie des langages

## V. Opérations sur les langages

**Exercice :**

Calculer  $A^*$  pour chacun des ensembles  $A$  suivants:

1)  $A = \{a\}$

2)  $A = \{\omega \in X^* / |\omega| = 2k+1 / k \geq 0\}$

**Corrigé :**

1) Si  $A = \{a\}$  alors  $A^* = \{a\}^* = a^*$

Car  $A^* = A^0 \cup A^1 \cup A^2 \cup \dots \cup A^i \cup \dots$

$$A^0 = \{\epsilon\} = \{a^0\}$$

$$A^1 = AA^0 = \{a\} \{\epsilon\} = \{a\} = \{a^1\}$$

$$A^2 = AA^1 = \{a\} \{a\} = \{aa\} = \{a^2\}$$

...

$$A^i = \{a^i\}$$

$$A^{i+1} = AA^i = \{a\} \{a^i\} = \{a^{i+1}\}$$

...

$$A^* = \{\epsilon, a, aa, aaa, \dots\} = \{a^0, a^1, a^2, a^3, \dots\} = a^*$$

16

## Ch1: Notions fondamentales en théorie des langages

## V. Opérations sur les langages

2) si  $A = \{\omega \in X^* \mid |\omega| = 2k+1 \mid k \geq 0\}$

$A^* = X^*$

Car  $A^* = A^0 \cup A^1 \cup A^2 \cup \dots \cup A^i \cup \dots$

$A^0 = \{\epsilon\} = \{X^0\}$

$A^1 = AA^0 = A\{\epsilon\} = \{\omega \in X^* \mid |\omega| = 2k+1 \mid k \geq 0\}$

$A^2 = AA^1 = \{\omega \in X^* \mid |\omega| = 2k+1 \mid k \geq 0\} \cup \{\omega \in X^* \mid |\omega| = 2k+1 \mid k \geq 0\}$   
 $= \{\omega \in X^* \mid |\omega| = 2k+2 \mid k \geq 0\}$   
 $= \{\omega \in X^* \mid |\omega| = 2k \mid k > 0\}$

$A^0 \cup A^2 = \{\epsilon\} \cup \{\omega \in X^* \mid |\omega| = 2k \mid k > 0\}$   
 $= \{\omega \in X^* \mid |\omega| = 2k \mid k \geq 0\}$

$A^0 \cup A^1 \cup A^2 = \{\omega \in X^* \mid |\omega| = 2k+1 \mid k \geq 0\} \cup$   
 $\{\omega \in X^* \mid |\omega| = 2k \mid k \geq 0\} = \{\omega \in X^*\} = X^*$

Donc  $A^* = X^*$

17

## Ch1: Notions fondamentales en théorie des langages

## IV. Propriétés des langages

## Exercice :

Calculer  $A \cdot \emptyset$  et  $A \cdot \{\epsilon\}$

Montrer que  $A \cdot (B \cap C) \neq A \cdot B \cap A \cdot C$

Montrer que  $A^+ = A^* - \{\epsilon\}$  n'est vraie lorsque A ne contient pas  $\epsilon$

## Corrigé:

1-  $A \cdot \emptyset = \emptyset \cdot A = \emptyset$

2-  $A \cdot \{\epsilon\} = \{\epsilon\} \cdot A = A$

3- **Contre exemple** : soient  $A = \{\epsilon, x\}$ ,  $B = \{xy\}$ ,  $C = \{y\}$

$(B \cap C) = \emptyset$  donc  $A \cdot (B \cap C) = \emptyset$

$A \cdot B = \{xy, xxy\}$  et  $A \cdot C = \{y, xy\}$  donc  $A \cdot B \cap A \cdot C = \{xy\}$

→ On peut conclure que :  $A \cdot (B \cap C) \neq A \cdot B \cap A \cdot C$

18

## Ch1: Notions fondamentales en théorie des langages

## VI. Propriétés des langages

4- On n'a pas  $A^+ = A^* \setminus \{\epsilon\}$  par contre on  $A^* = A^+ \cup \{\epsilon\}$

Pour  $A = \{\epsilon, a\}$

$$A^+ = A^1 \cup A^2 \cup \dots = \{\epsilon, a\} \cup \{\epsilon, a, aa\} \cup \dots = \{\epsilon, a, aa, aaa, \dots\}$$

$$A^* = A^0 \cup A^1 \cup A^2 \cup \dots = \{a^i / i \geq 0\}$$

$$A^* \setminus \{\epsilon\} = \{a^i / i > 0\} \neq A^+ \text{ car } \{\epsilon\} \subseteq A$$

$A^+ = A^* \setminus \{\epsilon\}$  est vraie lorsque A ne contient pas  $\epsilon$

19

## Ch1: Notions fondamentales en théorie des langages

## IV. Définition des langages

## Définition par propriété mesurable

L1 est l'ensemble des mots sur  $\{a, b\}$  de longueur paire

$$L1 = \{\omega \in \{a, b\}^* / |\omega| = 2k, K \geq 0\}$$

L2 est l'ensemble des mots sur  $\{a, b\}$  ayant un nombre impaire de b

$$L2 = \{\omega \in \{a, b\}^* / |\omega|_b = 2k+1, K \geq 0\}$$

L3 est l'ensemble des mots sur  $\{a, b\}$  où tous les a précèdent les b et sont de même nombre

$$L3 = \{a^n b^n, n \geq 0\}$$

20

## Ch1: Notions fondamentales en théorie des langages

### VII. Définition des langages

#### Définition récursive

$L_3$  est l'ensemble des mots sur  $\{a, b\}$  où tous les  $a$  précèdent les  $b$  et sont de même nombre

**Définition par propriété mesurable est :**

$$L_3 = \{ a^n b^n, n \geq 0 \}$$

**Définition récursive du même langage est:**

$$L_3 = \{ \omega \in \{a, b\}^* \mid \omega = \varepsilon \text{ ou } \omega = a \omega_1 b \text{ et } \omega_1 \in L_3 \}$$

21

## Ch1: Notions fondamentales en théorie des langages

### VII. Définition des langages

#### Définition récursive

$L_4$  est l'ensemble des mots palindromes sur le vocabulaire  $\{a, b\}$ , de longueur paire.

$$L_4 = \{ \omega \in \{a, b\}^* \mid \omega = aa \text{ ou } \omega = bb \text{ ou } \omega = a \omega_1 a \text{ ou } \omega = b \omega_1 b \text{ et } \omega_1 \in L_4 \}$$

$L_5$  est l'ensemble des mots palindromes sur le vocabulaire  $\{a, b\}$ , de longueur impaire.

$$L_5 = \{ \omega \in \{a, b\}^* \mid \omega = a \text{ ou } \omega = b \text{ ou } \omega = a \omega_1 a \text{ ou } \omega = b \omega_1 b \text{ et } \omega_1 \in L_5 \}$$

$L_6$  est l'ensemble des mots palindromes sur le vocabulaire  $\{a, b\}$ .

$$L_6 = \{ \omega \in \{a, b\}^* \mid \omega = a \text{ ou } \omega = b \text{ ou } \omega = aa \text{ ou } \omega = bb \text{ ou } \omega = a \omega_1 a \text{ ou } \omega = b \omega_1 b \text{ et } \omega_1 \in L_5 \}$$

22

## Ch1: Notions fondamentales en théorie des langages

### VIII. Le lemme d'Arden

Pour deux langages A et B d'un vocabulaire  $X^*$ ,  
Les équations  $L = AL \cup B$  et  $L = LA \cup B$  admettent respectivement comme solution minimale  $A^*B$  et  $BA^*$ . Cette solution est unique si  $\epsilon \notin A$ .

23

## Ch1: Notions fondamentales en théorie des langages

### IX. Exercices d'application

#### Exercice 1:

a) Soit  $E = \{a, b, c\}$ . Trouvez les ensembles suivants:

➤  $E^2 = E \times E$

➤  $P(E)$  = l'ensemble de toutes les parties de E

b) Démontrez par l'absurde que si  $A \cap B = \emptyset$  et  $C \subseteq B$  alors  $A \cap C = \emptyset$

#### Exercice 2:

a) Donner une définition formelle pour chacun des langages suivantes:

**L1** =  $\{xyy, xxyyyy, xxxyyyyy, \dots\}$ .

**L2** =  $\{xy, xyxy, xyxyxy, \dots\}$ .

**L3** =  $\{xx, xyx, xyyx, xyxyx, \dots\}$ .

b) Soit  $L = \{\omega \in \{x, y\}^* / \omega = x^{2n} y^n x^m, n, m \geq 0\}$

Les mots suivants appartiennent t'ils à L?

xyxxxx, xx, xxxxyyzx, xxxyx, xxy

24

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

**Exercice 3:**

Soit l'alphabet  $X = \{a, b, c, d\}$ .

Un mot  $\omega$  est dit parfait si et seulement si  $\omega = d$  ou  $\omega = aubvc$  où  $u$  et  $v$  étant parfaits

- Montrez que si  $w$  est parfait alors  $|\omega|_a = |\omega|_b = |\omega|_c = |\omega|_d - 1$
- Montrez qu'aucun facteur gauche (préfixe) propre d'un mot parfait n'est parfait

25

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

**Exercice 4:**

- Montrez que pour tout  $A, B \subseteq X^*$

$$(A^*B^*)^* = (A \cup B)^* = (A^* \cup B^*)^* = (A^*B)^*A^*$$

- Montrez que pour tout  $L \subseteq X^*$

Si  $\varepsilon \in L$  et  $L^2 = L$  alors  $L^* = L$

**Exercice 5:**

- Calculez  $L^2$ ,  $L^+$  et  $L^*$  pour:

- $L = \{00, 11\}$ .
- $L = \{\omega \in \{0, 1\}^* \mid |\omega|_0 = 2 \mid |\omega|_1\}$
- $L = \{\omega \in \{0, 1\}^* \mid \omega = 1^{2n} \ n > 0\}$

26

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

**Correction Exercice 4: Mq.  $(A^*B^*)^* = (A \cup B)^* = (A^* \cup B^*)^* = (A^*B)^*A^*$**

$$\begin{aligned}
 1) \quad & (A \cup B)^* \subseteq (A^*B^*)^* \\
 & \bullet A \subseteq A^* \subseteq A^*B^* \\
 & \bullet B \subseteq B^* \subseteq A^*B^* \\
 & \bullet \text{Donc } A \cup B \subseteq A^*B^* \quad \rightarrow (A \cup B)^* \subseteq (A^*B^*)^*
 \end{aligned}$$

$$\begin{aligned}
 2) \quad & (A^*B^*)^* \subseteq (A \cup B)^* \\
 & \bullet A^* \subseteq (A \cup B)^* \\
 & \bullet B^* \subseteq (A \cup B)^* \\
 & \bullet \text{Donc } A^*B^* \subseteq (A \cup B)^* \quad \rightarrow (A^*B^*)^* \subseteq (A \cup B)^*
 \end{aligned}$$

$$1) \text{ et } 2) \rightarrow (A^*B^*)^* = (A \cup B)^*$$

27

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

**Correction Exercice 4: Mq.  $(A^*B^*)^* = (A \cup B)^* = (A^* \cup B^*)^* = (A^*B)^*A^*$**

$$\begin{aligned}
 1) \quad & A \subseteq A^* \text{ \&\& } B \subseteq B^* \Rightarrow A \cup B \subseteq A^* \cup B^* \\
 & \quad \rightarrow (A \cup B)^* \subseteq (A^* \cup B^*)^*
 \end{aligned}$$

$$\begin{aligned}
 2) \quad & A^* \subseteq (A \cup B)^* \text{ \&\& } B^* \subseteq (A \cup B)^* \\
 & \quad \rightarrow A^* \cup B^* \subseteq (A \cup B)^*
 \end{aligned}$$

$$1) \text{ et } 2) \rightarrow (A^* \cup B^*)^* = (A \cup B)^*$$

28

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

**Correction Exercice 4:**  $(A^*B^*)^* = (A \cup B)^* = (A^* \cup B^*)^* = (A^*B)^*A^*$

$$1) A \subseteq A^*B \quad \&\& \quad B \subseteq A^*B \Rightarrow A \cup B \subseteq A^*B \\ \Rightarrow (A \cup B)^* \subseteq (A^*B)^* \subseteq (A^*B)^*A^*$$

$$2) A \subseteq (A \cup B) \quad \Rightarrow A^* \subseteq (A \cup B)^* \\ B \subseteq (A \cup B) \Rightarrow B \subseteq (A \cup B)^* \quad \Rightarrow A^*B \subseteq (A \cup B)^* \\ \Rightarrow (A^*B)^*A^* \subseteq (A \cup B)^*$$

$$1) \text{ et } 2) \Rightarrow (A^*B)^*A^* = (A \cup B)^*$$

29

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

**Correction Exercice 5:**

➤ Pour  $L = \{00, 11\}$ .

$$L^2 = \{0000, 0011, 1100, 1111\}$$

$$L^* = \{00, 11\}^*$$

$$L^+ = L^* \setminus \{\epsilon\} \text{ car } L \text{ ne contient pas } \epsilon$$

30

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

➤ Pour  $L = \{\omega \in \{0, 1\}^* \mid |\omega|_0 = 2 \mid |\omega|_1\}$

$$L^2 = \{\omega_1 \omega_2 \in \{0, 1\}^* \mid |\omega_1|_0 = \dots = 2 \mid |\omega_1|_1\} = L$$

$$L^* = \{\epsilon\} \cup L \cup L^2 \cup \dots \cup L^i = L \quad (\text{car } L \text{ contient } \epsilon)$$

$$L^+ = L \cup L^2 \cup \dots \cup L^i = L$$

31

## Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

➤ Pour  $L = \{\omega \in \{0, 1\}^* \mid |\omega| = 1^{2n} \quad n > 0\}$

$$L^2 = \{\omega \in \{0, 1\}^* \mid |\omega| = 1^{2n} \quad n \geq 2\} \subset L$$

$$L^* = \{\epsilon\} \cup L \cup L^2 \cup \dots \cup L^i = \{\epsilon\} \cup L$$

$$L^+ = L^* \setminus \{\epsilon\} \quad \text{car } L \text{ ne contient pas } \epsilon$$

$$\rightarrow L^+ = L$$

32