# 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

## 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 |



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.

## 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 : ε

**Définition 4.**
Longueur d'une chaine: La longueur d'une chaine finie ω est le nombre de symboles qu'elle contient.

On la note | ω |.





## III. Mot

**Définition 5.**
Mot : Un mot ω 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 ω;
si $\omega_i = a$ et $a \in X$ alors
$\omega_i$ est une occurrence de a dans le mot ω.

**Exercice** :
Donner les $\omega_i$ associés aux mots suivants :
1) abba sur le vocabulaire {a, b}

2) (x1*(x2+x1)) sur le vocabulaire {x1, x2, +, *, (, ) }

## 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 = {x1, x2, +, *, (, )} alors dans le mot (x1*(x2+x1))

| | | |
|---|---|---|
| $\omega_1 = ($ | $\omega_2 = x1$ | $\omega_3 = *$ |
| $\omega_4 = ($ | $\omega_5 = x2$ | $\omega_6 = +$ |
| $\omega_7 = x1$ | $\omega_8 = )$ | $\omega_9 = )$ |





## III. Mot

**Définition 6**.

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

**Exercice :**

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

**Corrigé:**

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

Le mot ε est de longueur 0, $| \varepsilon | = 0$

## III. Mot

**Définition 7.**
**Nombre d'occurrences d'un symbole dans un mot :**
Le nombre d'occurrences d'un symbole x dans un mot ω est le nombre de fois où ce symbole apparaît dans ce mot ω. On le note $| \omega |_x$.

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

**Exercice :**
Quel est le nombre d'occurrences de b dans les mots abba et ε

**Corrigé :**
$|abba|_b=2$
$| \varepsilon |_b=0$





## 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, …,9}
L = { représentations décimales des nombres entiers naturels}

Si le vocabulaire est X = {x1, x2, +, *, (, )}   x1+x2 ∈ X*
L = {expressions arithmétiques parenthèses}
x1, x2, + et *(x1+x2) sont 4 mots du langage L

## IV. Langage

Si X est la vocabulaire du langage de programmation C
X = {main, (, ), #, Include, <, >, ., ;, id, nb, …….}
L= {programmes C corrects syntaxiquement}
(main) element de X*
main(){ }

Si X est le vocabulaire de la logique des propositions
X ={p, (, ), →, ∧, ∨, ¬} où p désigne une proposition
L = {formules bien formées de la logique des propositions}
p ->p



Ch1: Notions fondamentales en théorie des langages

## IV. Langage

Si X est le vocabulaire de la logique des propositions
X ={p, (, ), →, ∧, ∨, ¬} où p désigne une proposition
L = {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→G, F ∧ G, F ∨ G le sont aussi
3) Si F est un mot alors (F) et ¬ F le sont aussi

## V. Opérations sur les langages

**Définition 9.**

**Concaténation de deux mots:**

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

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

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

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





## V. Opérations sur les langages

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

| | |
|---|---|
| **Intersection** | $A \cap B = \{\omega \,/\, \omega \in A$ et $\omega \in B\}$ |
| **Union** | $A \cup B = \{\omega \,/\, \omega \in A$ ou $\omega \in B\}$ (notée aussi A+B) |
| **Complémentation** | $A \backslash B = \{\omega \,/\, \omega \in A$ et $\omega \notin B\}$ (notée aussi A-B) |
| **Concaténation** | $A.B = \{\omega \,/\, \exists\, u \in A$ et $\exists\, v \in B$ et $\omega = u.v\}$ |

**Propriétés**

Soient A, B, C des langages, on a

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

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

Ch1: Notions fondamentales en théorie des langages

## V. Opérations sur les langages

**Notation**:

$a^k$ = aaaaa….aaaa    $a^2$ = aa

<………….. ….>

k fois

$a^0 = \varepsilon$

$a^* = \{a^i \mid i \geq 0\} = \{a^0, a^1, a^2, \ldots. a^i, \ldots\}$

$a^+ = \{a^i \mid i > 0\} = \{a^1, a^2, \ldots. a^i, \ldots\}$

**L'opération ***    $A^* = \cup A^i \quad \forall i \geq 0$ avec $A^0 = \{\varepsilon\}$, $A^1 = A$

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

**L'opération +**    $A^+ = \cup A^i \quad \forall i \geq 1$

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

**Corrigé :**

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

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

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

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

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

…

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

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

…

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

16

## V. Opérations sur les langages

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

$A^* = X^*$

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

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

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

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

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

$= \{\omega \in X^* \ / \ |\omega| = 2k \ / \ k > 0\}$

$A^0 \cup A^2 = \{\varepsilon\} \cup \{\omega \in X^* \ / \ |\omega| = 2k \ / \ k > 0\}$

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

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

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

Donc $A^* = X^*$





## IV. Propriétés des langages

**Exercice :**

Calculer A. ø et A.{ ε}

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

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

**Corrigé:**

1- A. ø = ø.A = ø

2- A. { ε} = { ε} .A= A

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

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

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

➔ On peut conclure que : $A.(B \cap C) \neq A.B \cap A.C$

## VI. Propriétés des langages

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

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

$A^+ = A^1 \cup A^2 \cup \ldots = \{\varepsilon, a\} \cup \{\varepsilon, a, aa\} \cup \ldots = \{\varepsilon, a, aa, aaa, \ldots.\}$

$A^* = A^0 \cup A^1 \cup A^2 \cup \ldots = \{a^i \mid i \geq 0\}$

$A^* \setminus \{\varepsilon\} = \{a^i \mid i > 0\} \quad \neq A^+$ car $\{\varepsilon\} \subseteq A$

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





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

## VII. Définition des langages

### Définition récursive

L3 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 :**

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

**Définition récursive du même langage est:**

$L3 = \{\omega \in \{a, b\}^* \ / \ \omega = \varepsilon \text{ ou } \omega = a\, \omega_1 b \text{ et } \omega_1 \in L3 \}$



Ch1: Notions fondamentales en théorie des langages

VII. Définition des langages

### Définition récursive

L4 est l'ensemble des mots palindromes sur le vocabulaire {a, b}, de longueur paire.

$L4 = \{\omega \in \{a, b\}^* / \ \omega = aa \text{ ou } \omega = bb \text{ ou } \omega = a\, \omega_1 a \text{ ou } \omega = b\, \omega_1 b \text{ et } \omega_1 \in L4\}$

L5 est l'ensemble des mots palindromes sur le vocabulaire {a, b}, de longueur impaire.

$L5 = \{\omega \in \{a, b\}^* / \ \omega = a \text{ ou } \omega = b \text{ ou } \omega = a\, \omega_1 a \text{ ou } \omega = b\, \omega_1 b \text{ et } \omega_1 \in L5\}$

L6 est l'ensemble des mots palindromes sur le vocabulaire {a, b}.

$L6 = \{\omega \in \{a, b\}^* / \ \omega = a \text{ ou } \omega = b \text{ ou } \omega = aa \text{ ou } \omega = bb \text{ ou}$

$\omega = a\, \omega_1 a \text{ ou } \omega = b\, \omega_1 b \text{ et } \omega_1 \in L5\}$

## VIII. Le lemme d'Arden

Pour deux langages A et B d'un vocabulaire X*,
Les équations L= AL ∪ B et L = LA ∪ B admettent respectivement comme solution minimale A*B et BA*. Cette solution est unique si ε ∉ A.





## IX. Exercices d'application

**Exercice 1:**
a) Soit E= {a, b, c}. Trouvez les ensembles suivants:
- $E^2$= E x E
- P(E)= l'ensemble de toutes les parties de E

b) Démentrez par l'absurde que si A ∩ B = ø et C ⊆ B alors A ∩ C = ø

**Exercice 2:**
a) Donner une définition formelle pour chacun des langages suivantes:
**L1=** {xyy, xxyyyy, xxxyyyyyy, ....}.
**L2=** {xy, xyxy, xyxyxy, ....}.
**L3=** {xx, xyx, xyyx,xyyyx,....}.
b) Soit L = {ω ∈ {x, y}* / ω =$x^{2n} y^{n} x^{m}$ , n, m ≥ 0}
Les mots suivants appartiennent t'ils à L?
xxyxxx, xx, xxxxyyzx,xxxyx,xxy

## IX. Exercices d'application

**Exercice 3:**
Soit l'alphabet X= {a, b, c, d}.
Un mot w est dit parfait si et seulement si ω =d ou ω =aubvc où u et v étant parfaits

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



Ch1: Notions fondamentales en théorie des langages

## IX. Exercices d'application

**Exercice 4:**
a) Montrez que pour tout A, B $\subseteq X^*$
$(A^*B^*)^* = (A \cup B)^* = (A^* \cup B^*)^* = (A^*B)^*A^*$
b) Montrez que pour tout L $\subseteq X^*$
Si $\varepsilon \in L$ et $L^2 = L$ alors $L^* = L$

**Exercice 5:**
a) Calculez $L^2$, $L^+$ et $L^*$ pour:

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

## IX. Exercices d'application

**Correction Exercice 5:**

- Pour L= {00,11}.

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

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

$L^+ = L^* \setminus \{\varepsilon\}$ car L ne contient pas ε





## IX. Exercices d'application

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

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

$L^* = \{\varepsilon\} \cup L \cup L^2 \cup \cdots \cup L^i = L$ ( car L contient ε)

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

## IX. Exercices d'application

- Pour $L = \{\omega \in \{0, 1\}^* \,/\, \omega = 1^{2n} \quad n > 0\}$

$L^2 = \{\omega \in \{0, 1\}^* \,/\, \omega = 1^{2n} \quad n >= 2\} \subset L$

$L^* = \{\varepsilon\} \cup L \cup L^2 \cup \cdots \cup L^i = \{\varepsilon\} \cup L$

$L^+ = L^* \setminus \{\varepsilon\}$ car L ne contient pas ε

➔ $L^+ = L$