Chapitre 4

# Langage

## 1 Généralités sur les langages

### 1.1 Un peu de vocabulaire.

**Définition 1 :**

- Un **alphabet** est un ensemble fini ; ces éléments sont des **lettres**.
- Un **mot** sur l'alphabet $A$ est une suite finie de lettres de $A$ : $w = a_1a_2...a_n$ .
- La **longueur du mot** précédent, notée $|w|$ est $n$ (nombre de lettres).
- On note $|w|_a$ le nombre d'occurrences de la lettre $a$ dans $w$. ( on a donc $|w| = \sum\limits_{a\in A} |w|_a$ )
- On définit un **mot vide** noté généralement $\epsilon$ , constitué d'aucune lettre. Sa longueur est nulle.
- L'ensemble des mots de longueur $n$ est noté $A^n$ ; l'ensemble des mots est noté $A^*$ . L'ensemble des mots non vides est noté $A^+$.
- On définit la concaténation de deux mots $w_1 = a_1a_2...a_n$ et $w_2 = b_1b_2...b_m$ par $w_1.w_2 = a_1a_2...a_nb_1b_2...b_m$ de sorte que $|w_1.w_2| = |w_1| + |w_2|$ .
- $a^n$ désigne la concaténation de $n$ lettres $a$ (si $n = 0$, alors $a^0 = \epsilon$ ).
- Par convention, $\epsilon.w = w.\epsilon = w$ , de sorte que . définit sur $A$ une loi de composition interne associative, non commutative (sauf si $A$ est un singleton), de neutre $\epsilon$ . Ainsi, $(A^*, .)$ constitue un monoïde. Une autre convention consiste à confondre une lettre et le mot constitué uniquement de cette lettre.
- Un langage sur $A$ est une partie de $A^*$ .
- Si $w = a_1a_2...a_n$, les **préfixes**(resp. **suffixes**) de $w$ sont les mots de la forme$a_1a_2...a_k$ (resp. $a_ka_2...a_n$), avec $1 \leq k \leq n$ . De plus, par convention, $\epsilon$ est préfixe et suffixe de tout mot. Un préfixe (resp. suffixe)propre de $w$ est un préfixe (resp. suffixe) distinct de $\epsilon$.
- Un **facteur** de $w = a_1a_2...a_n$ est un mot de la forme $a_i...a_j$ , avec $1 \leq i \leq j \leq n$. Un **sous-mot** est un mot de la forme $a_{i_1}...a_{i_k}$ , avec $1 \leq i_1 < ... < i_k \leq n$. Ici encore, $\epsilon$ est un facteur et un sous-mot de tout mot.
- Le **miroir** de $w = a_1a_2...a_n$ est $w' = a_n...a_1$ . Si $w = w'$ , on dit que $w$ est un palindrome.

### 1.2 Propriétés.

**Propriété 1 :**
$\forall u, v, x, y \in A^*$ , $uv = xy \Rightarrow \exists t \in A^*$ tel que soit $u = xt$ et $tv = y$ , soit $x = ut$ et $v = ty$

**Preuve :**

**Théorème 1 :**
Si $xy = yz$ avec $x \neq \epsilon$ , alors $\exists u, v \in A^*$ et un entier $k \geq 0$ tel que :

$$x = uv \quad , \quad y = (uv)^k u = u(vu)^k \quad , \quad z = vu$$

**Preuve :**

**Théorème 2 :**
Si $xy = yx$ , avec $x \neq \epsilon$ , et $y \neq \epsilon$ alors $\exists u \in A^*$ et deux entiers $i, j$ tels que $x = u^i$ , $y = u^j$ .

**Preuve :**

## 2 Opérations sur les langages .

Les langages sur $A$ étant des parties de $A^*$ , on dispose naturellement des notions de réunion,d'intersection , de différence et du complémentaire de langages.

$$L_1 \cup L_2 = \{w \in A^* | w \in L_1 \text{ ou } w \in L_2\}$$

$$L_1 \cap L_2 = \{w \in A^* | w \in L_1 \text{ et } w \in L_2\}$$

$$\overline{L_1} = \{w \in A^* | w \notin L_1\}$$

On peut de plus définir la concaténation ( ou le produit ) de deux langages $L_1$ et $L_2$ par :

$$L_1 L_2 = \{w_1 w_2 \in A^* | w_1 \in L_1 \text{ et } w_2 \in L_2\}$$

On définit alors de façon naturelle $L^n$ comme l'ensemble des mots constitués de $n$ mots de $L$ concaténés :

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

$$\forall n \in \mathbb{N} \quad , \quad L^{n+1} = L^n . L$$

**Définition 2 :**

**La fermeture de Kleene** ( ou **itération** ou encore **étoile** ) d'un langage $L$ est la réunion des $L_n$ , pour $n \in \mathbb{N}$. Elle est notée $L^*$ :

$$L^* = \bigcup_{n \in \mathbb{N}} L^n$$

elle contient toujours le mot vide.
$L^+$ désigne la réunion des $L^n$ pour $n \geq 1$ :

$$L^+ = \bigcup_{n \in \mathbb{N}^*} L^n$$

Ce langage contient le mot vide si et seulement si $\epsilon \in L$ . On a $L^+ = L.L^*$.

Exercice 1 : **Lemme d'Arden.**
Soient $A$ et $B$ des langages ( avec $\epsilon \notin A$) sur l'alphabet $\Sigma$ .
Montrer que l'équation de langage suivante :

$$L = A.L \cup B$$

admet pour unique solution le langage $L = A^* B$.

# 3 Langages et expressions rationnelles .

## 3.1 Langage rationnel :

**Définition 3 :**

Soit l'alphabet $A$. Les langages rationnels sur $A$ sont définis inductivement par :

- $\{\epsilon\}$ et $\varnothing$ sont des langages rationnels.
- $\forall a \in A$, $\{a\}$ est un langage rationnel.
- Si $L_1$ et $L_2$ sont des langages rationnels, alors $L_1 \cup L_2$ et $L_1.L_2$ également.
- Si $L$ est rationnel, alors $L^*$ l'est également.

**Remarques :**

- Tous les langages finis sont rationnels, puisqu'ils se déduisent des singletons par un nombre fini d'application des opérations d'union et de concaténation.
- Par définition, l'ensemble des langages rationnels est clos pour les trois opérations rationnelles. ( on dit qu'il est rationnellement clos.)
- La famille des langages rationnels correspond précisément au plus petit ensemble de langages qui contient tous les langages finis, et est rationnellement clos.

## 3.2 Expression rationnelle ( ou régulière ) :

**Définition 4 :**

L'ensemble $E$ des expressions rationnelles sur un alphabet $A$ est le plus petit ensemble tel que :

- $E$ contient les éléments de $A$, ainsi que deux éléments distincts des précédents $\epsilon$ et $\varnothing$.
- Si $\mathsf{e}_1$ et $\mathsf{e}_2$ sont deux éléments de $E$, alors $\mathsf{e}_1 + \mathsf{e}_2$ et$\mathsf{e}_1.\mathsf{e}_2$ également.
- Si $\mathsf{e}$ est un élément de $E$, alors $\mathsf{e}^*$ est un élément de $E$.

**Remarques :**

- On note parfois $\mathsf{e}_1|\mathsf{e}_2$ à la place de $\mathsf{e}_1 + \mathsf{e}_2$.
- Cette définition permet les preuves par induction : pour montrer une propriété $P(\mathsf{e})$ pour toute expression $\mathsf{e}$, il suffit d'établir $P(\varnothing)$, $P(\epsilon)$, $P(a)$ pour tout $a \in A$ , puis de montrer que si $P(\mathsf{e}_1)$ et $P(\mathsf{e}_2)$ sont vérifiées, alors $P(\mathsf{e}_1 + \mathsf{e}_2)$ , $P(\mathsf{e}_1.\mathsf{e}_2)$ et $P(\mathsf{e}_1^*)$ également.

**Définition 5 :**

On définit de façon inductive le langage associé à une expression rationnelle $\mathtt{e}$, noté $L(\mathtt{e})$ :

- $L(\epsilon) = \{\epsilon\}$ ; $L(\varnothing) = \varnothing$ ; $L(a) = \{a\}$ , $\forall a \in A$.
- $L(\mathtt{e}_1 + \mathtt{e}_2) = L(\mathtt{e}_1) \cup L(\mathtt{e}_2)$.
- $L(\mathtt{e}_1.\mathtt{e}_2) = L(\mathtt{e}_1).L(\mathtt{e}_2)$.
- $L(\mathtt{e}^*) = L(\mathtt{e})^*$

**Définition 6 :**

Un langage est dit rationnel si c'est le langage associé à une expression rationnelle. On note $\mathrm{Rat}(A)$ l'ensemble des langages rationnels : $\mathrm{Rat}(A) \subset \mathcal{P}(A^*)$.

## 3.3 Équivalence et réduction :

**Définition 7 :**

Deux expressions rationnelles sont équivalentes si elles dénotent le même langage.

**Exemple d'expressions équivalentes**

| | |
|---|---|
| $\varnothing a = a\varnothing = \varnothing$ | $\epsilon a = a\epsilon = a$ |
| $\varnothing^* = \epsilon$ | $\epsilon^* = \epsilon$ |
| $a + b = b + a$ | $a + \varnothing = a$ |
| $a + a = a$ | $a^* = (a^*)^*$ |
| $c(a + b) = ca + cb$ | $(a + b)c = ac + bc$ |
| $(ab)^*a = a(ba)^*$ | |
| $(a + b)^* = a^*(a + b)^*$ | $(a + b)^* = (a^* + b)^*$ |
| $(a + b)^* = (a^*b^*)^*$ | $(a + b)^* = (a^*b)^*a^*$ |

Exercice 2 : Soit $A = \{a, b\}$.

1. Montrer que le langage formé des mots comportant le facteur $ab$ est rationnel.
2. Montrer que le langage formé des mots ne comportant pas le facteur $ab$ est rationnel.
3. Montrer que le langage formé des mots comportant le facteur $ab$ mais pas le facteur $aa$ est rationnel.

<u>Exercice 3 :</u> Soit $A$ un alphabet, $L$ un langage et $u \in A^*$. On définit le quotient (ou résiduel) gauche de $L$ par $u$ par :

$$u^{-1}L = \{v \in A^* | uv \in L\}$$

Soit $a \in A$ une lettre et soient $K$ et $L$ deux langages.

1. Montrer que :
$$a^{-1}(K \cup L) = a^{-1}K \cup a^{-1}L$$

2. Montrer que :
$$a^{-1}(KL) = \begin{cases} (a^{-1}K)\,L & \text{si } \epsilon \notin K \\ (a^{-1}K)\,L \cup a^{-1}L & \text{si } \epsilon \in K \end{cases}$$

3. Montrer que :
$$a^{-1}L^* = \left(a^{-1}L\right) L^*$$

4. En déduire que si $L$ est un langage rationnel alors $a^{-1}L$ est aussi un langage rationnel.

5. Montrer que si $L$ est un langage rationnel et si $u \in A^*$ alors $u^{-1}L$ est aussi un langage rationnel.

<u>Exercice 4 :</u> On considère l'alphabet $\Sigma = a, b$. On note $L_1$ le langage des mots ayant un nombre pair de $b$ et $L_2$ le langage des mots ayant un nombre impair de $b$.

1. Justifier que l'on a les relations :
$$L_1 = aL_1 + bL_2 + \epsilon$$
$$L_2 = aL_2 + bL_1$$

2. Résoudre le système à l'aide du lemme d'Arden.

# 4 Langages locaux .

**Définition 8 :**

On considère un langage $L$ sur un alphabet $A$.On définie :

- $P(L) = \{a \in A | a.A^* \cap L \neq \varnothing\} = \{a \in A | \exists u \in A^* \quad, a.u \in L\}$
  $P(L)$ correspond à l'ensemble des premières lettres des mots de $L$.
- $S(L) = \{a \in A | A^*.a \cap L \neq \varnothing\} = \{a \in A | \exists u \in A^* \quad, u.a \in L\}$
  $S(L)$ correspond à l'ensemble des dernières lettres des mots de $L$.
- $F(L) = \{(a,b) \in A^2 | A^*.ab.A^* \cap L \neq \varnothing\} = \{(a,b) \in A^2 | \exists (u,v) \in (A^*)^2, u.a.b.v \in L\}$
  $F(L)$ correspond à l'ensemble des facteurs de 2 lettres des mots de $L$.
- $N(L) = \{(a,b) \in A^2 | A^*.ab.A^* \cap L = \varnothing\} = \{(a,b) \in A^2 | \exists (u,v) \in (A^*)^2, u.a.b.v \notin L\}$
  $F(L)$ correspond à l'ensemble des mots de 2 lettres non facteurs des mots de $L$.

**Exemple 1 :**

Pour $L = L(a^*bb^*)$, sur l'alphabet $A = \{a,b\}$, on a :

- $P(L) = \{a,b\}$
- $S(L) = \{b\}$
- $F(L) = \{aa, ab, bb\}$
- $N(L) = \{ba\}$

**Définition 9 :**

Un langage $L$ sur un alphabet $A$ est local s'il existe deux parties $P$ et $S$ de $A$ et une partie $N$ de $A^2$ telles que $w$ est un mot de $L$ si et seulement si :

- la première lettre de $w$ est dans $P$ ;
- la dernière lettre de $w$ est dans $S$ ;
- Aucun facteur de longueur 2 de $w$ n'est dans $N$.

On a alors :

$$L \setminus \{\epsilon\} = (PA^* \cap A^*S) \setminus (A^*NA^*)$$

**Exemple 2 :**

Sur l'alphabet $A = \{a,b\}$, les langages suivants sont locaux :

- $L = L(ab)$ car $ab = aA^* \cap A^*b \setminus A^*\{aa, bb, ba\}A^*$.
- $L = L((ab)^*)$ car $(ab)^+ = aA^* \cap A^*b \setminus A^*\{aa, bb\}A^*$.
- $L = L((a)^*)$ car $a^+ = aA^* \cap A^*a \setminus A^*\{ab, ba, bb\}A^*$.

Exercice 5 : Montrer que, sur l'alphabet $A = \{a, b\}$, les langages suivants :

- $L_1 = L((a)^*) \cup L((ab)^*) = L(a^* + (ab)^*)$.
- $L_2 = L((a)^*).L((ab)^*) = L((a)^*.(ab)^*)$.

ne sont pas locaux.

**Théorème 3 :**

L'intersection de deux langages locaux est un langage local.

**Preuve :**

**Théorème 4 :**

Si $L_1$ et $L_2$ sont deux langages locaux définis sur des alphabets disjoints, alors $L_1 \cup L_2$ est encore un langage local.

**Preuve :**

**Théorème 5 :**

Si $L_1$ et $L_2$ sont deux langages locaux définis sur des alphabets disjoints, alors $L_1.L_2$ est encore un langage local.

**Preuve :**

**Théorème 6 :**

La fermeture de Kleene $L^*$ d'un langage local $L$ est aussi un langage local.

**Preuve :**

## 5 Expressions rationnelles linéaires.

**Définition 10 :**

Une expression rationnelle $e$ est dite linéaire lorsque tout caractère de $A$ apparaît au plus une fois dans $e$ :

$$\forall a \in A \quad |e|_a \leq 1$$

**Exemple 3 :**

1. $e_1 = ab$ est une expression rationnelle linéaire.
2. $e_2 = (ab)^*$ est une expression rationnelle linéaire.
3. $e_3 = (ab)^*a$ n'est pas une expression rationnelle linéaire.

**Théorème 7 :**

Toute expression rationnelle linéaire dénote un langage local.

**Preuve :**