Partie 01 : Mots, Langages

1. Mots :

I. Définitions préliminaires :

1. Alphabet :

On appelle alphabet un ensemble fini quelconque. Les éléments d'un alphabet sont appelés lettres, caractères ou symboles.

Exemple 01:

X={a, b,…,z}

X={0,1,…,9}

2. Mot :

On appelle mot sur un alphabet X toute suite finie $a_1,a_2,\ldots,a_n$ d'éléments de X. On note usuellement un mot en écrivant en séquence, sans séparateur, les lettres qui la composent.

Exemple 02: X={a, b,…,z}

W_1=ab ; w_2=thp sont deux mots définis sur X.

Notations :

- Le mot vide est noté ε. $(|\varepsilon| = 0)$
- Les mots formés à partir d'un alphabet X est noté X^*. $(|X| = \propto)$.
- X^+ est l'ensemble des mots non vides formés à partir d'un alphabet X :

$$X^* = X^+ \cup \{\varepsilon\}$$

Exemple 03: X={a, b }

$X^* = \{\varepsilon, a, b, ab, ababa, aaa, \ldots\}$

3. Concaténation :

On peut définir la concaténation comme la juxtaposition de deux mots w_1 et w_2 et on note $w_1.w_2$

On vérifie facilement que la concaténation est une opération associative admettant le mot vide comme élément neutre. Soit :

- $\forall\, x, y, z \in X^*\quad x.(y.z) = (x.y).z$
- $\forall\, x \in X^*\quad x.\varepsilon=\varepsilon.x$

4. Longueur :

On appelle longueur d'une chaîne le nombre d'éléments de la suite la définissant. La longueur d'un mot w sera notée $|w|$.
Formellement on a :

- $|\varepsilon|=0$
- $|a|=1$; avec $a \in X$
- $|a.w|=1+|w|$; $\forall a \in X$ et ; $\forall w \in X^*$

Exercice 01: Montrer que $\forall w_1, w_2 \in X^*$, on a $|w_1.w_2|=|w_1|+|w_2|$
Indications : La démonstration se fait par récurrence sur le nombre de lettres de w_1 (ou bien w_2).

5. Miroir :

Le miroir d'un mot w noté w^R est le mot w 'lue à l'envers'. Par exemple : w=thp, w^R=pht. La définition récursive du miroir d'un mot :

$$w^R = \begin{cases} w & \text{Si } w=\varepsilon \\ v^R a & \text{Si } w=av \text{ ; } a \in X, v \in X^* \end{cases}$$

Exercice 02: **Montrer que $\forall u,v \in X^*$, $(uv)^R = v^R u^R$**

Solution :

Soit $u,v \in X^*$. On démontre $(uv)^R = v^R u^R$ par récurrence sur $|u|$:
Si $|u|=0$ alors $u=\varepsilon=u^R$ et $(uv)^R = (\varepsilon v)^R = v^R \varepsilon = v^R u^R$
On suppose maintenant que la formule est vraie $\forall u,v \in X^*$ tel que : $|u| \leq n$
Et vérifions qu'elle restera vraie pour l'ordre n+1.
Soit $u=a.w$ tel que $|u|=n+1$ avec : $w \in X^*$ ($|w|=n$) et $a \in X$
On a : $(uv)^R=(a.wv)^R=(wv)^R a=v^R w^R a= v^R u^R$. ($u=a.w$ et donc $u^R=w^R a$)

Par hypothèse de récurrence tous les mots w de taille <=n vérifiant la propriété, c'est-à-dire : $(wv)^R = v^R w^R$

Conclusion :
$\forall u,v \in X^*$, $(uv)^R = v^R u^R$

6. Puissance d'un mot :

La puissance d'un mot w est défini par récurrence comme suit :

$w^0=\varepsilon$

$w^{n+1} = w^n.w$

7. Factorisation :

Etant donné un mot w sur un alphabet X, un mot u est un sous-mot de w s'il existe $x,y \in X^*$ tels que $w = xuy$. Le sous mot u est un facteur gauche (préfixe) de w si $x = \varepsilon$; un facteur droit (suffixe) de w si $y = \varepsilon$.

Exemple 04:

Le mot abba admet les sous mots ε, a, b, ab, bb, ba, abb, bba, abba.

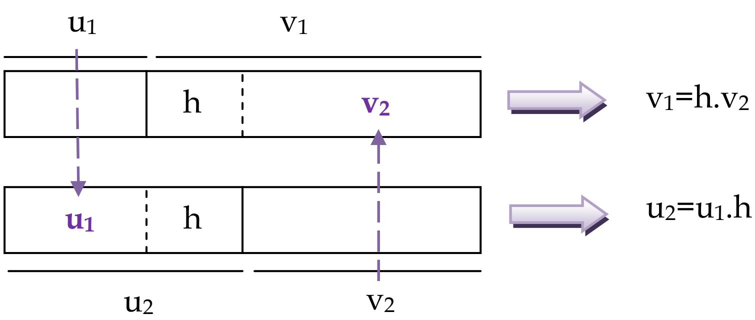
- Les facteurs gauches de abba sont ε, a, ab, abb et abba;
- Ses facteurs droits sont ε, a, ba, bba et abba.

II. Lemme de Levi :

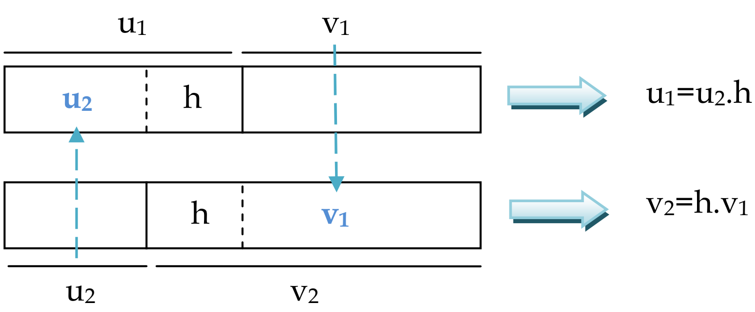
Soient $w=u_1.v_1=u_2.v_2$ on a alors 03 cas possibles : (w,u_1,v_1,u_2 et $v_2 \in X^*$)

1. Si $|u_1| = |u_2|$ Alors $u_1=u_2$ et $v_1=v_2$

2. Si $|u_1|<|u_2|$ Alors $u_2=u_1.h$ et $v_1=h.v_2$



3. Si $|u_1|>|u_2|$ Alors $u_1=u_2.h$ et $v_2=h.v_1$



III. Exercice corrigé sur les mots :

Enoncé :

1. Montrer que Si $xy = yz$, avec $x \neq \varepsilon$ alors $\exists u, v \in X^*$ et un entier $k \geq 0$ tels que :
$x = uv$, $y = (uv)^k u = u(vu)^k$, $z = vu$.

2. Montrer que Si $xy = yx$, avec $x \neq \varepsilon$, $y \neq \varepsilon$ alors $\exists u \in X^*$ et deux indices i et j tels que :
$x = u^i$ et $y = u^j$.

1. Preuve :

- Si $|x| \geq |y|$, alors le résultat précédent nous permet d'écrire directement $x = yt$, ce qui, en identifiant u et y, et v à t, nous permet de dériver directement les égalités voulues pour $k = 0$.
- Le cas où $|y| > |x|$ se traite par induction sur la longueur de y. Le cas où $|y|$ vaut 1 étant immédiat,
- Supposons la relation vraie pour tout y de longueur au moins n, et considérons y avec $|y| = n+1$. Il existe alors t tel que $y = xt$, d'où l'on dérive $xtz = xxt$, soit encore $tz = xt$, avec $|t| \leq n$. L'hypothèse de récurrence garantit l'existence de u et v tels que $x = uv$ et $t = (uv)^k u$, d'où $y = uv(uv)^k u = (uv)^{k+1} u$.

2. Démonstration :

- Ce résultat s'obtient de nouveau par induction sur la longueur de xy. Pour une longueur égale à 2 le résultat vaut trivialement. Supposons le valable jusqu'à la longueur n, et considérons xy de longueur n + 1. En utilisant le résultant précédent, il existe u et v tels que $x = uv$, $y = (uv)^k u$, d'où on déduit : $uv(uv)^k u = (uv)^k uuv$, soit encore $uv = vu$. En utilisant l'hypothèse de récurrence il vient alors : $u = t^i$, $v = t^j$, puis encore $x = t^{i+j}$ et $y = t^{i+k(i+j)}$, qui est le résultat recherché.

2. Langages :

Définitions :

Soit X un alphabet, on appelle langage formel défini sur X, tout sous ensemble de X*.

Exemple 05:

L_1=L'ensemble des mots $\{a, b\}^*$ qui commence par a et se termine par b. Donc :
$L_1=\{ab, aab, aaaab, abbbab,\ldots\}$.

L_2=L'ensemble des mots $\{a, b\}^*$ de taille inférieure strictement à 3.
$L_2=\{\varepsilon, a, aa, b, bb, ab, ba\}$.

- Un langage **fini** est un langage qui **contient un nombre fini de mots**. Un langage fini peut être décrit par l'énumération des mots qui le composent.
- Un langage **vide** est un langage qui ne contient aucun mot et il est noté $\varnothing$**.**
- Un langage est dit propre s'il ne contient pas le mot vide.
- Le langage $\varnothing$ est **différent** du langage $\{\varepsilon\}$.

Opérations sur les langages :

L'union :	$L_1 \cup L_2 = \{w / w\in L_1$ ou $w\in L_2\}$.
L'intersection :	$L_1 \cap L_2 = \{w / w\in L_1$ et $w\in L_2\}$.
L'inclusion :	$L_1 \subseteq L_2$ Si et seulement si $\forall w$ (Si $w\in L_1 \Rightarrow w\in L_2$)
La différence :	$L_1 - L_2 = \{w / w\in L_1$ et $w\notin L_2\}$.
La concaténation :	$L_1 . L_2 = \{w_1.w_2 / w_1\in L_1$ et $w_2 \in L_2\}$.
Langage miroir	$L^R=\{w / w^R\in L\}$.
Complément :	$L' = \{w / w\in X^*$ et $w \notin L\}$.
Puissance concaténative :	$L^0 =\{\varepsilon\}$ et $L^{n+1}=L^n.L$
Fermeture itérative ou étoile :	$L^*=L^0\cup L^1 \cup\ldots\ldots\cup L^k\cup\ldots\ldots= \cup_{i\geq 0} L^i$ $L^+=\cup_{i\geq 1} L^i$
$\varnothing.L_1=L_1.\varnothing= \varnothing$ $\{\varepsilon\}.L_1=L_1.\{\varepsilon\}=L$ $L^+=L.L^*= L^*.L$ $L^*=(L^*)^*$ $L^*.L^*=L^*$	

Propriétés sur la concaténation des langages :

- La concaténation des langages n'est pas idempotente, c'est-à-dire : $L.L \neq L$.
- La concaténation est associative.
- La concaténation des langages est distributive par rapport à l'union des langages.
- La concaténation des langages n'est pas distributive par rapport à l'intersection

des langages : $L_1.(L_2 \cap L_3) \neq L_1.L_2 \cap L_2.L_3$

- Soient L et M deux langages :

$(L^* . M^*)^* = (L \cup M)^*$

$(L.M)^*L = L.(M.L)^*$

$(L.M \cup L)^* .L = L.(M.L \cup L)^*$