QUELQUES PROBLEMES COMBINATOIRES DE LA THEORIE DES AUTOMATES

par M.P. SCHUTZENBERGER

Cours professé à l'Institut de
Programmation en 1966/1967,
rédigé par J.F. PERROT, Assistant.

TABLE DES MATIERES

-o-o-o-o-o-

Chap. 0 : Rappel de définitions.

Chap. I : "Régularités inévitables" dans le monoïde libre.

A. Théorème de Ramsey I.2

B. Suites bi-idéales I.9

C. Théorème de Van der Waerden et "Cadences" I.11

Chap. II : Equations à deux inconnues dans le monoïde libre.

A. Notions fondamentales II.1

B. Les mots primitifs et l'équation uv = vu II.4

C. L'équation générale à deux inconnues II.8

Chap. III : Propriétés combinatoires de l'ordre lexicographique

A. L'ordre lexicographique III.2

B. Les mots de Lyndon III.4

C. Application à un problème de synchronisation III.9

Chap. IV : Automates finis et K-langages

A. Le théorème de Kleene IV.2

B. Automate minimal reconnaissant un K-langage IV.14

C. Automate "boustrophédon". IV.21

Chap. V : Automates finis et évènements récurrents

A. Codes préfixes V.2

B. Evènements récurrents. V.8

CHAPITRE 0 .

-o-o-o-o-

RAPPEL DE DEFINITIONS

I.- Notion de monoïde.

Un monoïde $(M,.,e)$ est le triplet formé d'un ensemble M, d'une application de $M \times M$ dans M, notée $(m,m') \longrightarrow m.m'$ et appelée multiplication, et d'un élément distingué e de M, satisfaisant les axiomes suivants :

- la multiplication est associative , i e. $m.(m'.m") = (m.m').m"$ quels que m, m' et $m"$ dans M ;
- e est élément neutre pour la multiplication, i.e. $e.m = m.e = m$ pour tout m dans M.

Par abus de langage, on désignera par le même symbole M le monoïde et l'ensemble M.

Remarques :

- Il est clair que si $f \in M$ est élément neutre pour la multiplication, $f=e$.
- Si on n'exige pas la présence d'un élément neutre, la définition précédente devient celle d'un demi-groupe. La théorie des monoïdes apparaît comme un sous-ensemble de celle des demi-groupes ; sur cette dernière, voir CLIFFORD & PRESTON [2]+ ou LJAPIN [4]. Sur la notion de monoïde, cf. CHEVALLEY [1] .
- Si M est un monoïde, l'ensemble $\underline{P}(M)$ des parties de M reçoit canonique-

+ Les numéros entre crochets renvoient à la bibliographie placée en fin de chapitre.

ment une structure de monoïde, avec pour multiplication $A.B = \{a.b;\ a \in A,\ b \in B\}$, et pour élément neutre $\{e\}$.

Pour les notions de sous-monoïde, homomorphisme, congruence, monoïde-quotient, etc. on se reportera aux traités classiques d'algèbre et particulièrement à CHEVALLEY [1]. Notons simplement ici que :

- Un sous-monoïde doit toujours contenir l'élément neutre.
- Dans un homomorphisme de monoïdes, l'élément neutre de l'objet est toujours envoyé sur l'élément neutre de l'image.

II.- Monoïdes libres.

Etant donné un système générateur X d'un monoïde M, on dit que M est libre sur X, ou que X est un système générateur libre de M, ou que X engendre M librement, si pour toute application f de X dans un monoïde quelconque A il existe un homomorphisme φ de M dans A prolongeant f.

Il est équivalent de dire que tout élément de M admet une décompositon et une seule comme produit d'éléments de X.

Il est clair que, si M et M' sont libres respectivement sur X et X', et si on a une bijection de X sur X', alors M et M' sont isomorphes.

Plus généralement, si M' est engendré par X', et si on a une application de X sur X', on en tire un homomorphisme de tout monoïde librement engendré par X sur M'. Tout monoïde est donc image homomorphe d'un monoïde libre.

Il reste à prouver qu'il existe effectivement un monoïde libre au moins engendré par un ensemble X quelconque : on en obtient un noté X^* en prenant comme ensemble M l'ensemble des suites finies d'éléments de X, y compris la suite vide qui n'a aucun élément et qu'on notera e, avec pour mul-

tiplication la "concaténation" ou juxtaposition bout à bout de deux suites ; il est clair que cette opération est associative et que la suite vide e est élément neutre.(On peut définir sur l'ensemble en question d'autres structures de monoïde ; voir par exemple FOATA [3].)

Par abus de langage, X^* sera désigné dans la suite comme le monoïde libre engendré par l'ensemble X, et le vocabulaire suivant sera employé :

- X sera appelé alphabet, et ses éléments seront des lettres ;
- les suites finies $\in X^*$ seront des mots, e sera le mot vide ; l'ensemble des mots non vides est $X^* \setminus \{e\} = XX^*$; lorsqu'une lettre $x \in X$ apparaîtra dans un mot $w \in X^*$, on parlera d'une occurrence de x dans w ;
- le nombre d'occurences de x dans w sera noté $l_x(w)$, et le nombre total d'occurences des différentes lettres de X dans w sera la longueur de w $l(w) = \sum_{x \in X} l_x(w)$; les lettres $x \in X$ sont identifiées aux mots de longueur 1 ; le mot vide e est le seul mot de longueur nulle ;
- On dira que w' est un facteur de w si $w = uw'v$, avec $u, v \in X^*$; w' est un facteur gauche (resp. droit) si $u = e$ (resp. $v = e$) ;
- pour $A \subset X^*$, le sous-monoïde de X^* engendré par A sera désigné par A^*; pour $A = \{w\}$, on le notera simplement w^*.

Exemple : Le monoïde libre à un seul générateur $\{x\}^*$ est isomorphe au monoïde additif N des entiers naturels : si X est quelconque, et $Y \subset X$, l'application $w \to l_Y(w) = \sum_{y \in Y} l_y(w)$ est un homomorphisme de X^* sur N.

REFERENCES

[1] C. CHEVALLEY — Fundamental concepts of algebra, Academic Press 1956, Chap. I.

[2] A.H. CLIFFORD & G.B. PRESTON — The algebraic theory of semigroups Vol. I, Mathematical Survey Nr 7, American Mathematical Society 1961.

[3] D. FOATA — Etude algébrique de certains problèmes d'analyse combinatoire et du calcul des probabilités (thèse), Publ. Inst. Stat. Univ. Paris 14 (1965) p. 81-241.

[4] E.S. LJAPIN — Semigroups, Translations of mathematical monographs Vol. 3, American Mathematical Society 1963.

C H A P I T R E I

-o-o-o-o-o-o-o-o-

" Régularités inévitables"(1) dans le monoïde libre.

A - Théorème de Ramsey

B - Suites bi-idéales

C - Théorème de Van der Waerden et "Cadences"

(1) Cette expression est empruntée à G.Th. GUILBAUD [5].

A. THEOREME DE RAMSEY

I. Le théorème suivant a rendu célèbre le nom du mathématicien anglais F.P. RAMSEY qui le démontra en 1928 à la Société mathématique de Londres :

Soient E un ensemble, $P_m(E)$ l'ensemble des parties de E ayant m éléments, et $\mathcal{Q}$ une partition de $P_m(E)$ en k classes : il existe, quelque soit n, un entier $R(n,m,k)$ tel que $\text{Card}(E) \geq R(n,m,k)$ entraîne l'existence d'un sous-ensemble à n éléments F de E tel que $P_m(F)$ soit contenu tout entier dans une classe modulo $\mathcal{Q}$.

(Exposé, démonstration et bibliographie dans RYSER [8] chap. 4 ; sur les rapports avec la théorie des graphes voir ORE [7] chap. 13, sec. 5 ; on trouvera une discussion approfondie dans la thèse de C. FRASNAY [2]).

Pour $m = 1$, on a $R(n,1,k) = 1 + k(n-1)$ en vertu du principe des tiroirs selon lequel:" si $1 + k(n-1)$ éléments sont répartis entre k tiroirs, alors l'un des tiroirs renferme au moins n éléments "[2].

Pour $m > 1$ la fonction $R(n,m,k)$ n'est pas connue.

Pour $m = 2$, nous en déduisons le résultat suivant :

THEOREME : Soient X un ensemble non vide, X^* le monoïde libre engendré par X, et une partition de l'ensemble XX^* des mots de longueur positive

de X^* en k classes $A_1, A_2, \ldots, A_k$: pour tout entier n, il existe un entier $r_k(n)$ tel que, pour tout mot $w \in X^*$ de longueur supérieure à $r_k(n)$ il existe une classe A_i telle que w admette n facteurs non vides consécutifs dans A_i, i.e. $w \in X^* A_i^n X^*$.

On a certainement $r_k(n) \leq R(n+1,2,k)$. Soit $E = \{1, 2, \ldots l(w)\}$; à chaque sous-ensemble à deux éléments $\{a,b\}$ de E $(a < b)$ nous associons bijectivement le facteur $w_{a,b}$ de w commençant au rang a et se terminant au rang b : la partition induite sur l'ensemble des facteurs de w par celle de XX^* permet de définir une partition de $P_2(E)$ en au plus k classes, dont l'une au moins contient tous les sous-ensembles à deux éléments de $F \subset E, F = \{p_0, p_1, \ldots, p_n\}$ avec $p_i < p_{i+1}$. On a donc $w_{p_0, p_1} w_{p_1, p_2} \ldots w_{p_{n-1}, p_n} \in A_i^n$ pour une certaine classe A_i.

II. Nous allons maintenant préciser ce résultat en montrant directement que $r_k(n) = n^k$.

a) Cas où $k = 2$, $A_1 = A$, $A_2 = B$.

Exemple :

On montre que $r_2(2) = 4$, en raisonnant par l'absurde ; supposons qu'il existe un mot w de longueur supérieure ou égale à 4 et qui ne soit ni dans $X^* A^2 X^*$ ni dans $X^* B^2 X^*$: il est clair que, si w a une longueur strictement supérieure à 4, tous ses facteurs de longueur 4 possèdent la même propriété ; on peut donc supposer que w est de longueur 4, soit

$w = xyzt$, et, sans restreindre la généralité, que x est dans A. Il s'ensuit nécessairement que y est dans B, z dans A et t dans B. Dans ces conditions, le mot yz ne peut être ni dans A ni dans B sans contredire l'hyptohèse faite sur w, hypothèse qui est par conséquent absurde. On a ainsi établi que $r_2(2)$ est inférieur ou égal à 4, et l'exemple trivial suivant montre qu'on a bien l'égalité : $X = x,y$ et $A = X^*x$, $B = X^*y$, $w = xyx$, ou $w = yxy$.

Démonstration que $r_2(n) \leq n^2$:

A tout mot $w \in X^*$ associons le couple d'entiers (a_w, b_w) défini par : $a_w = \max \{k \mid w \in X^* A^k\}$; $b_w = \max \{k \mid w \in X^* B^k\}$. Si on a $w = fg$, avec $f \neq e$, $g \neq e$, on a nécessairement $(a_w, b_w) \neq (a_f, b_f)$.

En effet, $(a_w, b_w) = (a_f, b_f)$ entraînerait $w \in X^* A^{a_w} g$ et $w \in X^* B^{b_w} g$; si $g \in A$ (resp. $g \in B$), on en tire $w \in X^* A^{a_w+1}$ (resp. $w \in X^* B^{b_w+1}$) ce qui est contradictoire avec la maximalité de a_w (resp. de b_w).

En particulier, cette application envoie deux facteurs gauches différents d'un même mot w sur deux points différents du treillis des points à coordonnées entières du plan. Si w a pour longueur n^2, cela implique que, pour au moins un facteur f de w, on ait soit a_f soit b_f supérieur ou égal à n, puisqu'il n'y a que n^2 points à coordonnées entières du premier quadrant dont les deux coordonnées sont strictement inférieures à n et qu'un mot de longueur 1 a exactement 1 facteurs

gauches différents non vides auxquels il faut adjoindre e dont l'image est l'origine. Supposons $a_f \geq n$: $w \in f X^*$ et $f \in X^* A^{a_f} \subset X^* A^n X^*$ entraînent $w \in X^* A^n X^*$, ce qu'il fallait démontrer.

Exemple montrant que $n_2(n)$ ne saurait en général être inférieur à n^2 :

On considère l'alphabet à deux lettres $X = \{x,y\}$ et la partition $A = xx^*$ (mots formés exclusivement à l'aide de la lettre x) et $B = X^* yX^*$ (mots contenant au moins une occurrence de la lettre y). On forme alors la famille infinie de mots f_n $(1 \leq n \leq \infty)$ de la forme $(x^{n-1}y)^{n-1}x^{n-1}$, qui sont de longueur n^2-1, et dans chacun desquels il est clair qu'il ne se trouve aucun facteur de la forme A^n ou B^n.

b) Cas général d'une partition de XX^* en un nombre fini quelconque k de classes A_i $(i = 1, 2, \ldots, k)$:

On associe à chaque mot w les k entiers $a_i(w)$ $a_i(w) = \max (t \mid w \in X^* A_i^t)$ et on montre de la même façon que précédemment que si $w = fg$ avec f et g de longueur non nulle les suites $(a_i(m) : i = 1, \ldots, k)$ et $(a_i(n) : i = 1, \ldots, k)$ ne peuvent être identiques. On en déduit alors le résultat en appliquant la suite du raisonnement à des mots de longueur n^k et au treillis des points à coordonnées entières à k dimensions.

Exercice : Montrer que si $l(w) \geq n_1 n_2 \ldots n_k$ il existe i,j, $1 \leq i, j \leq k$, tels que $w \in X^* A_j^{n_i} X^*$.

Afin d'établir l'optimalité de ce résultat, on peut considérer la famille d'exemples suivante : pour chaque valeur de k, on utilise la partition de $X_k X_k^*$, où X_k est l'alphabet à k lettres $x_1, x_2, \ldots, x_k$, définie par $A_1 = x_1(x_1)^*$ et $A_i = (x_1, \ldots, x_i)(x_1, \ldots, x_i)^* - (x_1, \ldots, x_{i-1})^*$ pour $1 < i \leq k$. On construit alors la famille des mots f_n^k $(1 < n \leq \infty)$ par récurrence sur k : la famille (f_n^2) est par définition la famille (f_n) définie au paragraphe précédent ; la famille (f_n^k) se déduit de la famille (f_n^{k-1}) de la façon suivante : remplacer les $n-1$ occurrences de x_{k-1} dans f_n^{k-1} par autant d'occurrences de x_k, puis substituer à chaque sous-mot situé soit avant la première occurrence de x_k, soit après la dernière, soit entre deux occurrences (tous ces sous-mots sont égaux à f_n^{k-2}) le mot f_n^{k-1} lui même. On a ainsi $f_2^2 = x_1\ x_2\ x_1$, $f_3^2 = x_1\ x_1\ x_2\ x_1\ x_1\ x_2\ x_1\ x_1$, et $f_2^3 = x_1\ x_2\ x_1\ x_3\ x_1\ x_2\ x_1$,

$f_3^3 = x_1\ x_1\ x_2\ x_1\ x_1\ x_2\ x_1\ x_1\ x_3\ x_1\ x_1\ x_2\ x_1\ x_1\ x_2\ x_1\ x_1\ x_3\ x_1\ x_1\ x_2\ x_1\ x_1\ x_2\ x_1 x_1$.

On vérifie immédiatement que f_n^k est de longueur n^k-1 et ne peut pas avoir n facteurs consécutifs situés dans la même classe.

Exercice : Montrer qu'un K-langage (cf. infra, chap. IV) infini sur X fini contient nécessairement un sous-ensemble de la forme $u\ w^*\ v$, $u, v \in X^*$, $w \in XX^*$.

III. Extension aux séquences infinies :

On considère maintenant l'ensemble X^∞ des suites infinies de lettres de l'alphabet X.

Etant donné une partition de XX^* en k classes A_i, on montre que

$X^\infty = X^* \left(\bigcup_{i=1}^{k} A_i^\infty \right)$, c'est-à-dire que toute séquence infinie de lettre dans X a une infinité de facteurs consécutifs dans l'une au moins des k classes A_i.

a) Démonstration dans le cas où $k = 2$, $A_1 = A$, $A_2 = B$.

Supposons que la suite $s \in X^\infty$ ne soit ni dans AB^∞ ni dans BA^∞ : cela implique que s soit dans un produit infini de la forme $A^{h_1} B^{k_1} A^{h_2} B^{k_2} \ldots A^{h_n} B^{k_n} A^{h_{n+1}} B^{k_{n+1}} \ldots$ où les h_i et les k_i sont maximaux, i.e. tous les facteurs commençant après A^{h_i} sont dans B et tous les facteurs commençant après B^{k_j} sont dans A. Or il est clair qu'une pareille situation est impossible, car le facteur de s situé dans $A^{h_i} B^{k_i}$ ne peut être ni dans A à cause de la maximalité de $A^{h_{i+1}}$ ni dans B à cause de la maximalité de $B^{k_{i-1}}$.

b) Cas général :

La démonstration se fait au moyen du résultat suivant que l'on établit par récurrence, et dont l'énoncé précédent se déduit immédiatement :

Si on a une suite infinie de facteurs consécutifs pris dans k classes différentes, telle que le produit d'un nombre fini de ces facteurs consécutifs soit encore dans une des k classes considérées, alors il y a une infinité de facteurs consécutifs dans l'une au moins de ces k classes.

Le résultat est trivial pour $k = 1$; pour $k = 2$ la démonstration est exactement la même que celle qui a été faite précédemment dans le cas d'une partition de XX^* en deux classes. Le pas inductif se fait ainsi :

Supposons que les hypothèses soient satisfaites pour une famille de

$k + 1$ classes, et qu'il n'y ait point de suite infinie de facteurs dans A_{k+1} : en effet, dans le cas contraire il n'y a rien à démontrer. Si à partir d'un certain rang il n'apparaît aucun facteur qui soit dans A_{k+1}, alors la famille des k premières classes vérifie les hypothèses et l'hypothèse de récurrence donne le résultat ; sinon, chaque occurrence d'un mot de A_{k+1} est située à l'intérieur d'une suite finie maximale de facteurs dans A_{k+1} qui se termine en un point où ne commencent que des facteurs situés dans les k premières classes. L'ensemble des facteurs qui joignent de tels points satisfait aux hypothèses et grâce à l'induction le résultat est complètement démontré.

Exercice : Remarquer que la proposition pour les suites infinies ne découle pas du résultat établi auparavant pour les mots (suites finies). Toutefois, la méthode de démonstration ci-dessus peut être adaptée pour établir l'existence de $r_k(n)$: quelle évaluation de $r_k(n)$ obtient-on par ce procédé ?

B. SUITES BI-IDEALES

On appelle suite bi-idéale une suite infinie $\{k_n ; n \in \underline{N}\}$ de mots de X^* satisfaisant aux hypothèses suivantes :

- $k_o \neq e$
- $k_{n+1} \in k_n X^* k_n \quad \forall_n$.

Par exemple, la suite (f_n^k) définie précédemment est pour n fixé et k allant de 1 à l'infini une suite bi-idéale, alors que f_n^k pour k fixé et n variant n'en est pas une.

Il est facile de constater que le terme courant d'une suite bi-idéale a nécessairement une structure assez particulière et c'est ce qui fait l'intérêt du résultat suivant, démontré et utilisé dans [1] :

Si l'alphabet X est fini, pour tout entier n il existe un entier $k(n)$ tel que tout mot de X^* de longueur supérieure ou égale à $k(n)$ ait au moins un facteur qui est le n-ième terme d'une suite bi-idéale.

Remarquons d'abord que ce résultat cesse d'être vrai si X est infini : il suffit de considérer les mots où une même lettre n'apparaît pas deux fois.

Si X est fini, soit q nombres de lettres et posons $k(0) = 1$, $k(n+1) = (1+q^{k(n)})k(n)$. Cette suite de nombres est très rapidement croissante. On démontre qu'elle jouit de la propriété annoncée en raisonnant par récurrence sur n.

Le résultat est trivial pour $n = 1$.

Soit f dans $X^{k(n+1)}$: d'après la définition de $k(n+1)$ on peut écrire $f = f_1 f_2 \ldots f_s$ avec $s = 1 + q^{k(n)}$ et f_i dans $X^{k(n)}$.

Comme il n'y a que $q^{k(n)}$ mots de longueur $k(n)$ différents, il y a nécessairement deux des f_i au moins qui sont égaux, soit f_i et f_j : ces deux mots, en vertu de l'hypothèse de récurrence, ont un facteur k_n qui est n-ième terme d'une suite bi-idéale, $f_i = f_j = g\, k_n g'$, et f s'écrit $f = X^* gk_n\, g'\, X^*\, gk_n\, g'\, X^*$, d'où, avec le facteur $k_n\, X^*\, k_n$, le résultat.

C. LE THEOREME DE VAN DER WAERDEN ET LES "CADENCES"

Le théorème de Van der WAERDEN sur les progressions arithmétiques (cf. KHINTCHINE [6]) s'énonce comme suit :

Etant donnés deux entiers naturels quelconques k et l, il existe un entier $n(k,l)$ tel que, si $A = (a, a+1, a+2, \ldots, a+n(k,l)-1)$ est un segment quelconque de $\underline{N}$ de longueur $n(k,l)$, divisé d'une façon arbitraire en k classes, alors dans l'une au moins de ces k classes apparaît une progression arithmétique de l termes.

G. Th. GUILBAUD [5] en a donné l'interprétation suivante (voir aussi J. GARDELLE [3] et J. GARDELLE et G. Th. GUILBAUD [4]) :

Soit X un alphabet fini : dira que $m \in X^*$ contient une "cadence" d'ordre p s'il existe une lettre $x \in X$ telle que, parmi les occurrences de x dans m, il s'en trouve p dont les rangs forment une progression arithmétique.

Exemple : le mot aabaabba contient une cadence d'ordre 3 (aabaabba) mait le mot abbaabba n'en contient pas ([3]).

Comme il s'établit entre un mot de longueur n et un segment de même longueur une bijection naturelle qui induit un partage du segment en $k=$Card(X) classes (dont chacune est formée des rangs des occurrences d'une même lettre) et réciproquement, on peut donner au théorème de Van der Waerden la forme équivalente que voici ([4]) :

Etant donnés deux entiers k et l, il existe un entier $n(k,l)$ tel que tout mot de longueur $n(k,l)$ sur un alphabet de k lettres contienne au moins une cadence d'ordre l.

La démonstration suivante est la simple traduction en terme de cadences de celle qui est donnée par [6].

On procède par récurrence sur l : en vertu du "principe des tiroirs" on a $n(k,2) = k+1$. Supposons donc connu $n(k,l)$ quel que soit k pour un $l \geq 2$: nous allons construire $n(k,l+1)$, en définissant deux suites d'entiers (q_r) et (n_r) récursivement par : $q_0 = 1$, $n_0 = n(k,l)$ et $q_r = 2\, n_{r-1}\; q_{r-1}, n_r = n(k^{q_r}, l)$, et en montrant que l'on peut choisir $n(k,l+1) = q_k$.

Remarquons d'abord que tout mot sur X de longueur rp peut être considéré comme un mot de longueur p sur l'alphabet (à k^r lettres) X^r il s'ensuit que tout mot m de longueur q_r, considéré comme mot sur $X^{q_{r-1}}$, a longueur $2n_{r-1}$ et que par définition de n_{r-1} sa moitié gauche contient une cadence d'ordre l, c'est à dire qu'un même facteur de longueur q_{r-1} s'y répète l fois à intervalles réguliers ; soit d la longueur constante (comptée en lettres de X) qui sépare deux termes successifs de cette cadence. Nous adjoignons à cette cadence un $(l+1)^{ème}$ facteur de longueur q_{r-1}, celui qui se trouve à la distance d du dernier terme de la cadence : il est peut-être différent des l premiers termes, et peut-être empiète-t-il sur la moitié droite du mot m, mais il ne déborde certainement pas de m. Nous avons ainsi construit une suite de $l+1$ facteurs équidistants de m, dont les l premiers sont identiques et dont le dernier a même longueur que les autres, à savoir q_{r-1}.

Au mot m de longueur q_r nous associons ainsi une suite $D(m)$ de $l+1$ entiers, formée des $l+1$ rangs des lettres initiales des facteurs de longueur q_{r-1} de la cadence ci-dessus définie, complété par le $(l+1)^{ème}$ facteur. Il est clair

que la suite d'entiers $D(m)$ permet de définir sur un mot quelconque m' de longueur q_r une suite de $l+1$ facteurs de m', non nécessairement identiques, mais tous de longueur q_{r-1}. Nous pouvons donc considérer $D(m)$ comme un opérateur qui à un mot m' de longueur q_r fait correspondre une suite de $l+1$ facteurs de longueur q_{r-1} équidistants de m', que nous noterons $D(m)m'$. En particulier, $D(m)m$ est la suite de $l+1$ facteurs définie au paragraphe précédent.

Soit à présent m de longueur q_k : appelons m_i, $1 \leq i \leq l+1$ lesfacteurs successifs de $D(m)m$, m_{i_1,i_2}, $1 \leq i_2 \leq l+1$ ceux de $D(m_1)\, m_{i_1}$, m_{i_1,i_2,i_3}, $1 \leq i_3 \leq l+1$ ceux de $D(m_{1,1})m_{i_1,i_2}$ et ainsi de suite. Les facteurs m_i sont de longueur q_{k-1}, les $m_{i,j}$ de longueur q_{k-2}, etc ... Par conséquent les $m_{i_1,i_2,\ldots,i_k}$ seront de longueur $q_0 = 1$, c'est-à-dire des lettres de X. Comme par construction $m_{i_1,i_2,\ldots,i_r} = m_{j_1,j_2,\ldots,j_r}$ pour $1 \leq r \leq k$ et $1 \leq i_1,i_2,\ldots,i_r, j_1,\ldots,j_r \leq l$ (car $i_1 \leq l$ et $j_1 \leq l$ entraînent $m_{i_1} = m_{j_1}$, donc $D(m_1)\, m_{i_1} = D(m_1)\, m_{j_1}$ et ainsi de suite) nous avons :

(1) $m_{i_1,i_2,\ldots,i_r,i_{r+1},\ldots i_k} = m_{j_1,j_2,\ldots,j_r,i_{r+1},\ldots,i_k}$ pour $1 \leq r \leq k$ et $1 \leq i_1,i_2,\ldots,i_r,j_1,\ldots,j_r \leq l$, $1 \leq i_{r+1},i_{r+2},\ldots,i_k \leq l+1$.

De la même façon, du fait que $m_{i_1,i_2,\ldots i_{r-1},i_r}$ et $m_{i_1,i_2,\ldots i_{r-1},i_r+1}$ sont toujours des facteurs "voisins" dans $D(\underbrace{m_{1,1,\ldots,1}}_{r-1})m_{i_1,i_2,\ldots,i_{r-1}}$, dont la distance d_{r-1} ne dépend que de r, on déduit que les lettres

(2) $m_{i_1,\ldots,i_{r-1},i_r,i_{r+1},\ldots,i_k}$ et $m_{i_1,\ldots,i_{r-1},i_r+1,i_{r+1},\ldots,i_k}$ sont aussi distantes de d_{r-1}, pour $1 \leq r \leq k$ et $1 \leq i_1,\ldots,i_k \leq l+1$.

Parmi les k+1 lettres

$$m_{\underbrace{l+1,l+1,\ldots,l+1}_{k}}$$

$$m_{1,\underbrace{l+1,l+1,\ldots,l+1}_{k-1}}$$

$$m_{1,1,\underbrace{l+1,\ldots,l+1}_{k-2}}$$

$$m_{1,1,1,l+1,\ldots,l+1}$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$m_{\underbrace{1,1,\ldots,1}_{k-1},l+1}$$

$$m_{\underbrace{1,1,\ldots,1}_{k}},$$

au moins deux sont identiques, soit

(3) $$m_{\underbrace{1,1,\ldots,1}_{r},\underbrace{l+1,\ldots,l+1}_{k-r}} = m_{\underbrace{1,1,\ldots,1}_{s},\underbrace{l+1,\ldots,l+1}_{k-s}} \quad \text{avec} \quad r < s .$$

Il en résulte que les l+1 lettres $x_i, x_i = m_{\underbrace{1,\ldots,1}_{r},i,i,\ldots,i,\underbrace{l+1,l+1,\ldots,l+1}_{r-s}}$

sont toutes identiques : les l premières le sont en vertu de (1), et (3) signifie que $x_1 = x_{l+1}$. Il reste à prouver que ces l+1 lettres forment une cadence, c'est à dire qu'elles sont équidistantes. Soit d(i) la distance séparant x_i et x_{i+1} ; appelons d(i,p), pour $1 \leq p \leq s-r$, la distance séparant les lettres

$$m_{\underbrace{1,\ldots,1}_{r},\underbrace{i+1,i+1,\ldots,I+1}_{p-1},\underbrace{i,i,\ldots,i}_{s-r-p+1},\underbrace{l+1,\ldots l+1}_{k-s}} \quad \text{et}$$

$$m_{\underbrace{1,\ldots,1}_{r},\underbrace{i+1,\ldots,i+1}_{p},\underbrace{i,\ldots,i}_{s-r-p},\underbrace{l+1,\ldots,l+1}_{k-s}} .$$

Or, on a d'une part $d(i) = \sum_{p=1}^{p=s-r} d(i,p)$, d'autre part d'après (2)

$d(i,p) = d_{r+p-1}$, d'où il résulte que d(i) ne dépend pas de i, C.Q.F.D.

REFERENCES DU CHAPITRE I.

[1] M. COUDRAIN et M.P. SCHÜTZENBERGER, Une condition de finitude des monoïdes finiment engendrés, C.R. Acad. Sci. Paris, 262 (1966) p.1149-1151.

[2] C. FRASNAY ; Quelques problèmes combinatoires concernant les ordres totaux et les relations monomorphes, Ann.Inst. Fourier, Grenoble, 15 ,2, (1965).

[3] J. GARDELLE, A propos des "cadences", Mathématiques et Sciences humaines, n° 8 (1964) p. 36.

[4] J. GARDELLE et G. Th. GUILBAUD, Cadences, Mathématiques et Sciences humaines n° 9 (1964) p. 31-38.

[5] G. Th. GUILBAUD, Un exercice sur les permutations, Mathématiques et Sciences humaines n° 2 (1963) p. 37.

[6] A. Y. KHINCHIN, Three pearls of Number Theory, Graylock Press, Rochester, N.Y., 1952.

[7] O. ORE, Theory of Graphs, Am. Math. Soc. Coll. Publications Vol. 38 ; Providence 1962.

[8] H. J. RYSER ; Combinatorial Mathematics, Carus Mathematical Monographs N° 14 1963

C H A P I T R E II

-o-o-o-o-o-o-o-o-o-

Equations à deux inconnues dans le monoïde libre.

A - Notions fondamentales

B - Les mots primitifs et l'équation $uv = vu$

C - L'équation générale à deux inconnues

A. NOTIONS FONDAMENTALES

Le résultat principal de ce chapitre est que, si deux mots $u,v \in X^*$ ($X \neq \emptyset$, mais non nécessairement fini) vérifient une équation propre (en un sens qui sera précisé ultérieurement), ils sont nécessairement puissances d'un même troisième w, $u=w^h$, $v=w^k$. Nous utiliserons les outils combinatoires suivants :

I. Le lemme de F.W. LEVI :

Soient quatre mots $a,b,c,$ et de X^* tels que $ab=cd$. On a :

- si $l(a)>l(c)$ il existe dans X^* un mot f unique tel que $a=cf$ et $d=fb$;
- si $l(a)=l(c)$ $a=c$ et $b=d$;
- si $l(a)<l(c)$ il existe dans X^* un mot f unique tel que $af=c$ et $b=fd$.

Tout cela se vérifie immédiatement ; on peut remarquer que le lemme de LEVI reste valable dans des monoïdes plus généraux, pourvu qu'une notion de "longueur" y soit définie, comme par exemple celui qu'on obtient en mettant bout à bout des graphes de fonctions réelles définies sur des segments.

II. LA RELATION DE CONJUGAISON DANS X^*.

Soient a, b, et c dans X^* tels que ab=bc. Alors il existe u et v dans X^* et p entier tels que : a=uv, c=vu, $b=(uv)^p u=u(vu)^p$.

La démonstration se fait par récurrence sur l(b) :

- le résultat est vrai si l(b)=0, il suffit de choisir u=e=b, v=a=c, et p=0.
- Si l(a) > l(b), d'après le lemme de LEVI il existe f tel que a=bf et c=fb ; il suffit dans ce cas de choisir u=b et v=f, avec p=0.
- Si l(b)=l(a), il en va de même avec f=v=e.
- Si l(a) < l(b), il existe f tel que b=af et b=fc, où l(f) < l(b) ; en vertu de l'hypothèse d'induction, on a : a=uv, c=vu et $f=(uv)^p u$, d'où $b=uv\,(uv)^p u=(uv)^{p+1}u$.

Remarquons que ce résultat repose essentiellement sur le caractère discret de la longueur des mots dans X^* et cesserait d'être vrai, en l'absence d'hypothèses supplémentaires, dans l'exemple évoqué ci-dessus.

Réciproquement ; il est clair que si a=uv et c=vu, pour b=u on a ab=bc=uvu , par conséquent on peut énoncer :

Définition : deux mots a et c de X^* sont appelés conjugués s'ils satisfont à l'une quelconque des deux conditions équivalentes :

- il existe b dans X^* tel que $ab=bc$;
- il existe u et v dans X^* tels que $a=uv$ et $c=vu$.

Proposition : La relation de conjugaison est une équivalence.

Réflexivité et symétrie sont évidentes ; quant à la transitivité on a : $ab=bc$, $cb'=b'd$ entraîne $abb'=bcb'=bb'd$.

Remarque 1 : Si a est conjugué de b, a est obtenu par "permutation circulaire" des lettres de b.

Remarque 2 : Si $a=w^n$ alors tout conjugué a' de a est de la forme $a'=w'^n$, où w' est conjugué de w.

Soit en effet $a=uv=w^n$, on en tire $u=w^{n_1}w_1$ et $v=w_2w^{n_2}$, avec $w_1 w_2=w$ et $n=n_1+n_2+1$. De $a'=vu$, avec $w'=w_2 w_1$, on tire $a'=w'^n$.

B. LES MOTS PRIMITIFS ET L'EQUATION uv=vu.

I. Définitions : On dira que deux mots g et f sont proprement conjugués s'il existe dans XX^* deux mots u et v tels que g=uv et f=vu. Ceci équivaut à dire que fh=hg pour h qui n'est puissance ni de f ni de g : en effet, sous ces hypothèses on a f=uv et g=vu, $h=(uv)^n u$ et ni u ni v ne peuvent se réduire au mot vide ; la réciproque est évidente. La relation de propre conjugaison n'est évidemment plus réflexive : un mot qui n'est pas proprement conjugué de lui-même est appelé primitif.

Nous démontrerons que :

(1) Si un mot f de XX^* est proprement conjugué de lui-même, il est puissance d'un mot h de XX^*, i.e. $f=h^n, n>1$, et réciproquement.

Il en résulte qu'un mot est primitif si, et seulement s'il n'est puissance d'aucun autre mot que lui-même, et on peut raffiner l'énoncé ci-dessus :

(1') Tout mot proprement conjugué de lui-même est puissance (>1) d'un mot primitif. (On verra ultérieurement que ce mot primitif est uniquement déterminé). Remarquons également que tout conjugué d'un mot primitif est aussi primitif, et que si $a=w^n$, w primitif, alors tout a' conjugué de a s'écrit $a'=w'^n$, avec w' primitif, conjugué de w.

II. Afin d'établir (1), montrons que :

(2) Si u et v sont permutables, i.e. s'ils satisfont uv=vu, alors u et v sont puissances d'un même mot w. ([3], lemme 3 ; ce résultat a été ensuite retrouvé par différents auteurs, p. ex. GINSBURG & SPANIER [2], lemme 5.1).

Il signifie que les seules solutions de l'équation uv=vu sont de la forme $u=w^h$ $v=w^k$.

Vu la définition de la propre conjugaison, il est clair que (2) implique (1). (2) se démontre par récurrence sur les longueurs de u et v : si u=e, le résultat est trivial. Si u n'est pas puissance de v, uv=vu exprime que v est proprement conjugué de lui-même et donc v=u'v'=v'u', v' et u' étant chacun de longueur strictement moindre que celle de v , donc $v'=w^m$, et $u'=w^n$; de plus on a $u=(u'v')^iu'$ qui est donc puissance de w ainsi que v.

Pour montrer que le mot primitif dont l'existence est assurée par (1') est uniquement déterminé, nous ferons appel au résultat plus général suivant, dû à FINE & WILF ([1], théorème 1) qui précise le lemme 4 de [3].

III. Lemme : Soient $a, b \in X^*$, $\alpha = l(a)$, $\beta = l(b)$ et $\delta = PGCD(\alpha,\beta)$. S'il existe deux puissance a^p et b^q de a et b qui se mettent sous la forme $a^p=vfw$ et $b^q=v'fw'$, avec $l(f)=\alpha+\beta-\delta$ et $l(v)\equiv l(v')$ (mod. δ), alors a et b sont tous deux puissances d'un même mot c, $a=c^m$ et $b=c^n$.

La démonstration que voici, inédite, est due à M. A. LENTIN.

Supposons d'abord $\delta=1$ et $\alpha < \beta$, et montrons que dans ces conditions a et b sont nécessairement puissances d'une même lettre de X. On sait que les résidus mod. β des puissances successives de α forment exactement l'ensemble $\{1, 2, \ldots, \beta-1\}$: il en résulte que les β premières lettres de f sont identiques, d'où le résultat. En effet la première et la $(\alpha+1)$-ième le sont à cause de la périodicité en a ; si $h,k < \beta$, $h \equiv r\,\alpha \pmod{\beta}$ et $k\equiv(r+1)\alpha \pmod{\beta}$, alors la h-ième et la k-ième lettres de f sont identiques, car la h-ième et la $(h+\alpha)$-ième le sont par périodicité en a , et parce que $h+\alpha \leq l(f)$, et soit $k=h+\alpha$, soit $k=h+\alpha-\beta$ et alors la périodicité en b joue.

Si à présent $\delta \neq 1$, nous pouvons considérer a, b et f comme des mots sur l'alphabet X^{δ}. Le raisonnement précédent montre alors que $f=g^{k}$, avec $l(g)=\delta$, d'où $a=g'^{m}$ et $b=g''^{n}$, g'' et g' étant conjugués à g. Or la condition $l(v)\equiv l(v') \pmod{\delta}$ signifie précisément que g' et g'' sont le même conjugué de g.

Exercice : trouver des exemples montrant que les deux conditions de l'énoncé sont effectivement nécessaires.

Il résulte en particulier de ce lemme que si on avait $f=h^{n}=h'^{n'}$ il serait impossible que h et h' fussent primitifs et différents. Ceci achève de prouver que :

Soit $a \in XX^*$: il existe un mot primitif b et un entier $k \geq 1$ uniques tels que $a=b^k$. ([3], corollaire 4.2)

Exercices: Donner le nombre de conjugués différents d'un mot primitif.

Démontrer que $u^2 v^2=w^2$ entraîne $uv = vu$.

C. L'EQUATION GENERALE A DEUX INCONNUES.

I. NOTION D'EQUATION A N INCONNUES.

Nous formaliserons l'idée intuitive de "relation liant n mots dans X^*" de la façon suivante :

Soit $\Xi = \{\xi_1, \xi_2, \ldots, \xi_n\}$ un alphabet de n lettres, disjoint de X. Une <u>équation</u> à n inconnues dans X^* est définie par la donnée de deux mots μ et υ de Ξ^*. L'équation sera dite <u>propre</u> si μ et υ sont non vides et si $\mu \neq \upsilon$.

Résoudre cette équation, signifie trouver tous les homomorphismes φ de Ξ^* dans X^* - donc tous les n-uples de mots de X^* $\varphi\xi_1, \varphi\xi_2, \ldots, \varphi\xi_n$ - tels que $\varphi\mu=\varphi\upsilon$. La solution $\varphi\xi_i=e$, $i=1\ldots n$ sera dite triviale.

Ainsi, l'équation à deux variables $uv=vu$ est formellement définie par la donnée des deux mots $\xi\eta$ et $\eta\xi$ de Ξ^* ($\Xi=\{\xi,\eta\}$) c'est donc bien une équation propre, et nous avons montré que les seuls homomorphismes φ convenables sont obtenus par $\varphi\xi=w^m$, $\varphi\eta=w^n$, $w \in X^*$, $m, n \in \underline{N}$. Nous allons voir que c'est aussi, quand il en existe d'autres que la solution triviale, l'ensemble des solutions d'une équation propre quelconque à deux variables. Dans le cas de trois variables, le problème est ouvert.

II. SOLUTION DE L'EQUATION PROPRE A DEUX VARIABLES.

La démonstration par récurrence utilisée pour obtenir (2) s'étend au cas général comme suit :

Il est clair qu'on peut toujours se ramener par simplification, à l'un des deux cas :

$$\mu \in \xi\Xi^* \xi \qquad \text{et} \qquad \upsilon \in \eta\Xi^* \eta$$

ou

$$\mu \in \xi\Xi^* \eta \qquad \text{et} \qquad \upsilon \in \eta\Xi^* \xi.$$

Remarquons aussi que dans ces conditions, si φ est une solution $l(\varphi\xi)=l(\varphi\eta)$ entraîne $\varphi\xi=\varphi\eta$, donc le résultat.

Pour prouver que la solution non triviale φ (dont nous supposons l'existence) est nécessairement de la forme prescrite, nous pouvons donc supposer $l(\varphi\xi) \geqslant l(\varphi\eta)$, et raisonner par récurrence sur $l(\varphi\xi)$: pour $l(\varphi\xi)=1$ le résultat est trivial.

D'après le lemme de LEVI, appliqué à $\varphi\mu=\varphi\upsilon$, nous avons $\varphi\xi=(\varphi\eta)c$, pour un certain $c \in XX^*$. Considérons donc l'alphabet $\Xi'=\{\eta,\zeta\}$, et l'homomorphisme Ψ de Ξ'^* dans X^* défini par : $\Psi\eta=\varphi\eta$ et $\Psi\zeta=c$; en remplaçant $\varphi\xi$ par $(\varphi\eta)c$ dans $\varphi\mu$ et dans $\varphi\upsilon$ et en simplifiant par le $\varphi\eta$ initial commun, on obtient deux mots égaux de X^* qui sont les images par Ψ de deux mots μ' et υ' de Ξ'^* l'un et l'autre non vides et <u>différents</u> : en effet, la suppression de $\varphi\eta$ dans $\varphi\mu$ (qui commence par $\varphi\xi$) laisse en début de mot un facteur c, donc μ' commence par ζ ; il en résulte que la suppression de $\varphi\eta$

dans $\varphi\upsilon$ (qui commence par $\varphi\eta$) laisse un facteur non vide, lequel commence par $\varphi\xi$ ou par $\varphi\eta$, et en tous cas υ' commence par η. μ' et υ' définissent donc une équation propre, qui possède une solution Ψ telle que $\max(l(\Psi\eta), l(\Psi\zeta)) < l(\varphi\xi)$: l'hypothèse de récurrence entraîne $\Psi\eta = w^m$, $\Psi\zeta = w^n$, d'où nous tirons $\varphi\eta = \Psi\eta = w^m$ et $\varphi\xi = (\varphi\eta)c = (\Psi\eta)(\Psi\zeta) = w^{m+n}$.

Exercice : Résoudre l'équation à trois inconnues

$$\xi\ \zeta\ \xi = \eta\ \eta$$

REFERENCES

[1] N.J. FINE & H.S. WILF Uniqueness theorems for periodic functions, Proc. American Math. Soc. 16(1965) p. 109-114.

[2] S. GINSBURG & E.H. SPANIER, Bounded ALGOL-like languages, Trans. American Math. Soc. 113(1964) 333-368.

[3] R.C. LYNDON & M.P. SCHÜTZENBERGER The equation $a^m=b^nc^p$ in a free group, Michigan Math. J. 9(1962) 289-298.

C H A P I T R E III

-o-o-o-o-o-o-o-o-o-o-

Propriétés combinatoires de l'ordre lexicographique

A - L'ordre lexicographique

B - Les mots de Lyndon

C - Application à un problème de synchronisation

A. L'ORDRE LEXICOGRAPHIQUE

Nous donnons dans ce chapitre une propriété classique de l'ordre lexicographique. Le lecteur désireux d'approfondir la question pourra se reporter à la thèse de D. FOATA ([3], chap. 5 et bibliographie) ; sur les rapports de cette théorie avec les algèbres de Lie libres, voir P.M. COHN ([2], p. 289 sq.)

I. DEFINITION DE L'ORDRE LEXICOGRAPHIQUE.

Soit X un alphabet (non nécessairement dénombrable) totalement ordonné : on peut étendre à X^* l'ordre de X par le moyen suivant :

- si $f,g \in X^*$ et $f \in g\,X^*$, alors $g \leq f$;
- si $f=hxh'$ et $g=hx'h''$, où x et x' sont dans X, si $x > x'$, alors $f > g$.

La relation ainsi définie sur X^* est appelée ordre lexicographique induit sur X^* par l'ordre de X.

Exercices :

- Vérifier que les axiomes d'une relation d'ordre sont effectivement satisfaits et que deux mots de X^* sont toujours comparables (ordre total).
- On suppose que X est fini : soit n le nombre de ses lettres. On pose pour chaque f dans X^*,
  $rf = \sum_{1}^{l(f)} (\text{n° dans } X \text{ de } x_i)\, n^{-i}$ où x_i désigne la i-ième lettre de f. Montrer que $rf \leq rg \iff f \leq g$.
- L'ordre lexicographique peut-il être un bon ordre ?

II. PROPRIETE FONDAMENTALE DE L'ORDRE LEXICOGRAPHIQUE.

On déduit immédiatement des deux axiomes que si on a $f > g$ l'une et l'une seulement des deux éventualités suivantes se produit :

- soit $f \in g X^*$;
- soit quels que soient h et k dans X^* on a $fh > gk$.

Il en résulte en particulier que la concaténation dans X^* n'est pas monotone à droite, c'est-à-dire qu'on peut avoir $f < g$ et $gh < fh$, comme c'est le cas par exemple pour $f=a, g=ab$ et $h=c$ (l'alphabet étant $\{a, b, c\}$). En revanche, il est clair que la concaténation est monotone à gauche, i.e. que $f < g$ entraîne $hf < hg$ pour tout h dans X^*.

B. LES MOTS DE LYNDON.

Dans tout ce qui suit, la relation d'ordre sur X^* est l'ordre lexicographique.

I. Définition : Un mot de LYNDON est un mot h possédant l'une des 3 propriétés suivantes, qui sont équivalentes :

1- h est primitif, et pour tout conjugué h' de h, on a $h \leq h'$;

2- pour tout conjugué propre h' de h, on a $h < h'$;

3- pour tout facteur droit propre k de h, on a $h < k$.

Il est clair que 1- est équivalent à 2-.

2- implique 3- car supposons que $h=lk$, avec $l \neq e$ et $k < h$, (i.e. que h ne satisfait pas 3-) ; nous avons : soit $h \notin k X^*$, d'où, en vertu de la propriété fondamentale de l'ordre lexicographique, $kl < h$ et h ne vérifie pas 2-, soit $h \in k X^*$, donc $h=kl'$, alors si h vérifie 2- $kl' < kl$, ce qui signifie $l' < l$ et, comme l et l' sont de même longueur, $l \notin l' X^*$, donc $l'k < lk=h$ ce qui est contradictoire avec l'hypothèse que h satisfait 2-.

3- implique 2- car si on a $h'=kl < lk=h$ avec $l \neq e$, ceci entraîne $k < h$ puisque $k < kl$.

Exercice : Donner une autre démonstration de 2- $\Rightarrow$ 3- en choisissant pour k le facteur droit minimum (dans l'ordre lexicographique) de h,

en remarquant que 1 et 1' sont conjugués et enfin qu'un mot de la forme uvu, où $u \neq e$ et $v \neq e$, ne peut jamais satisfaire 2-.

Il est clair que tout mot de X^* possède une factorisation au moins en mots de LYNDON, puisque les mots de longueur 1 en sont. Nous allons voir que tout mot possède une décomposition remarquable en mots de LYNDON, laquelle est unique.

II. THEOREME DE CHEN, FOX et LYNDON.[1] : Tout mot de X^* admet une factorisation unique en mots de LYNDON, les facteurs pris dans l'ordre où ils apparaissent formant une suite non croissante, c'est-à-dire $f=h_1 h_2 \ldots h_m$ où les h_i sont des mots de LYNDON et où $h_1 \geq h_2 \geq \ldots \geq h_m$.

Soit f un mot de X^* et g son facteur droit minimum dans l'ordre lexicographique : d'après 3- g est un mot de LYNDON ; si $f \neq g$, posons $f=f'g$ et opérons de même sur f', obtenant ainsi un second mot de LYNDON g' ; montrons que $g \leq g'$. En effet g'g est un facteur droit de f et d'après la définition de g, $g'g > g$; or si on avait $g' < g$, on aurait soit $g \notin g' X^*$ et donc $g'g < g$, ce qui est n'est pas soit $g=g'k$, auquel cas on doit avoir $k > g$ puisque g est un mot de LYNDON, ce qui entraîne $g=g'k > g'g$ comme précédemment. Si $f' \neq g'$, nous pouvons renouveler l'opération en posant $f'=f''g'$, on obtiendra un troisième mot de LYNDON g" et on aura $g \leq g' \leq g''$. Il est clair qu'au bout d'un nombre fini de pas on arrivera à un ultime facteur gauche de f qui sera un mot de LYNDON : à ce moment, on aura obtenu une factorisation de f en mots de LYNDON qui peut s'écrire, au prix d'une renumérotation, $f=h_1 h_2 \ldots h_m$ avec

$h_i \geq h_{i+1}$. Montrons maintenant que cette décomposition est unique.

Soit $h_1 \geq h_2 \geq .. \geq h_m$ une suite finie non croissante de mots de LYNDON et $f = h_1 h_2 ... h_m$: si on effectue sur f la décomposition précédente on retrouvera la même suite de mots de LYNDON, il suffit pour le prouver de montrer que le facteur droit minimum de f soit g, est nécessairement h_m. En effet, g ne peut pas être plus court que h_m puisque h_m, en tant que mot de LYNDON, est inférieur à tous ses facteurs droits, et g ne peut pas non plus être plus long car en ce cas on aurait $g = k h_i h_{i+1} ... h_m$ pour un certain entier $i \leq m$, donc $g > k$ et comme k est facteur droit de h_{i-1}, $k \geq h_{i-1} \geq h_m$, et $g > h_m$ ce qui est contradictoire avec la définition de g comme facteur droit minimum de f.

Remarque 1 : Le théorème ci-dessus généralise la caractérisation 3- des mots de LYNDON.

Remarque 2 : Si $g_1 \geq g_2 \geq ... \geq g_p$ est une sous-suite de la suite de mots de LYNDON $h_1 \geq h_2 \geq ... \geq h_m$ $(p \leq m)$, alors $g_1 g_2 ... g_p \leq h_1 h_2 ... h_m$, ainsi qu'on le vérifie facilement par récurrence sur p.

III. Toutes les considérations précédentes devenant triviales dans le cas où le mot f considéré est lui-même un mot de LYNDON, on est amené à s'intéresser à l'ensemble des facteurs droits propres de f : le facteur droit propre minimum de f est encore un mot de LYNDON, soit g et $f = kg$ avec $k \neq e$. Soit maintenant $f = h$ un mot de LYNDON, alors k est également un mot de LYNDON et on a $k < g$. Comme h est

un mot de LYNDON et que g en est un facteur droit, on a $g > h > k$. Si k est de longueur 1, k est certainement un mot de LYNDON ; sinon, soit l un facteur droit propre de k, montrons que $l > k$. En effet lg est un facteur droit propre de h et par conséquent $lg > g$; or ceci implique $l \geq g$ car si on avait $g > l$, on aurait soit $g \notin l X^*$ et alors $lg < g$, soit $g = ll'$, et comme g est un mot de LYNDON $l' > g$ d'où $lg > ll' = g$; donc $l \geq g > k$ et k est un mot de LYNDON.

Réciproquement, si k et g sont des mots de LYNDON et si $g > k$, alors kg est un mot de LYNDON : montrons que si l est un facteur droit de kg, alors $l > kg$. Si l est plus long que g , $l = k'g$ où k' est un facteur droit de k , donc $k' > k$ et comme k' est moins long que k, $k' \notin k X^*$, et $k'g > kg$. Si l est un facteur droit de g , $l \geq g$ et il suffit de montrer que $g > kg$; comme $g > k$, soit $g \notin k X^*$ et $g > kg$, soit $g = kg'$ où g' est nécessairement $> g$ d'où $g = kg' > kg$.

Ainsi <u>le produit kg de deux mots de LYNDON k et g est un mot de LYNDON si (et seulement si) $k < g$ et réciproquement tout mot de LYNDON de longueur supérieure à 1 peut se mettre sous cette forme</u>. Toutefois, cette décomposition n'est pas unique : $aababb = a(ababb) = (aab)(abb) = (aabab)b$.

<u>Exercices</u> :

Dans toute la théorie précédente, remplacer l'expression "facteur droit minimum dans l'ordre lexicographique" par "facteur droit le plus long qui soit un mot de LYNDON".

- III.8 -

Montrer que si un mot de LYNDON h admet une décomposition $h=kg$, où k, g, sont des mots de LYNDON satisfaisant $k < g$, et $g < l$ pour tout facteur droit propre l de k , alors cette décomposition est unique.

C. APPLICATION A UN PROBLEME DE SYNCHRONISATION.

I. On envisage le problème suivant ([4], Remark 5) :

Etant donné un alphabet X à q lettres, et un entier positif quelconque n, on cherche un sous-ensemble K de X^n de taille minimale tel qu'un nombre fini seulement de mots de X^* n'aient aucun facteur dans K, (ensemble "coupant" Si on considère le graphe dont les sommets sont les mots de X^n et où un arc relie f à g si et seulement si il existe dans X des lettres x et y telles que $fx=yg$, cela revient à rechercher K de taille minimale tel que le sous-graphe ayant pour ensemble de sommets X^n K soit sans circuits.

Appelons $M_q(n)$ la taille minimale de K : on vérifie sans peine que

$$M_2(2) = 3 \quad \text{et que} \quad M_2(3) = 4.$$

Remarquons que si $q > 1$, $M_q(n+1) > M_q(n)$; en effet l'ensemble des facteurs gauches de longueur n des mots de K_{n+1} forment un ensemble K'_n qui, sa taille mise à part, satisfait à l'ordre n aux conditions imposées et d'autre part il existe nécessairement dans K_{n+1} deux mots différents qui ont même facteur gauche de longueur n, car sinon une famille infinie de mots n'ayant aucun facteur dans K_{n+1} pourrait être construite de la manière suivante : étant donné un mot f de longueur n quelconque, il existe dans X une lettre x telle que fx n'est pas dans K_{n+1}, cela est vrai aussi du facteur droit de longueur n de fx, d'où fxy qui n'a aucun facteur dans K_{n+1}, et on itère la construction en trouvant à chaque fois un mot plus long que le précédent et n'ayant aucun facteur dans K_{n+1}.

On voit aussi facilement que $M_{q+1}(n) > M_q(n)$, le monoïde libre à q

générateurs pouvant être plongé dans le monoïde libre à $q+1$ générateurs ; par conséquent, $M_q(n)$ tend vers l'infini avec $\max(q,n)$; d'autre part, il résulte de la définition que $q^{-n}M_q(n)$ reste inférieur à 1. On se propose de montrer qu'en fait $nq^{-n}M_q(n)$ tend vers 1 quand $\max(n,q)$ tend vers l'infini.

Montrons d'abord que $nq^{-n}M_q(n) \geq 1$. Pour cela remarquons que tout mot de longueur n, soit f, est puissance d'un mot primitif g de longueur divisant n, et que par suite les différents facteurs de longueur n des puissances successives f^m de f, qui sont tous les conjugués de f, sont puissances des différents conjugués du mot primitif g. Il en résulte que pour remplir sa mission K doit contenir au moins une puissance d'un représentant de chaque classe de mots primitifs conjugués de longueur divisant n. Soit $L_q(d)$ le nombre de classes de mots primitifs de longueur d (c'est aussi le nombre de mots de LYNDON de longueur d), comme une telle classe contient exactement d éléments (la conjugaison est équivalente à une permutation circulaire) et que le nombre de mots de longueur n est q^n, on en tire

$$q^n = \sum_{d|n} dL_q(d), \quad \text{d'où} \quad n^{-1}q^n = \sum_{d|n} n^{-1}dL_q(d) \leq \sum_{d|n} L_q(d) \leq M_q(n).$$

Nous allons maintenant donner, à l'aide de deux ensembles "coupants", deux majorations de $nq^{-n}M_q(n)$, l'une (II) tendant vers 1 quand q tend vers l'infini, l'autre (III) tendant vers 1 quand n tend vers l'infini.

II.a) Pour cela, montrons que si K_n est l'ensemble des facteurs gauches de longueur n des puissances des mots de LYNDON de longueur inférieure ou égale à n (l'ordre lexicographique étant défini à partir d'un ordre total quelconque sur X, qui est fini), tout mot suffisamment long a un facteur au moins dans K_n.

Plus précisément, on montrera que tout mot f de longueur supérieure ou égale à $(n-1)\,q^n+1$ a un facteur dans K_n, en raisonnant par récurrence sur le numéro du facteur gauche de longueur n de f dans l'ordre induit sur X^n par l'ordre lexicographique, lequel est au plus q^n, on montrera que si le facteur gauche de longueur n de f a pour numéro k, alors, si la longueur de f est supérieure ou égale à $k(n-1)+1$, f a un facteur dans K_n.

Si ce numéro est 1, alors nécessairement f commence par a^n, où a est la première lettre de l'alphabet, et a^n étant le facteur gauche de longueur n des puissances de a, qui est un mot de LYNDON de longueur 1, est dans K_n. Dans ces conditions il suffit que f ait une longueur supérieure ou égale à n $(n=1(n-1)+1)$ pour avoir un facteur dans K_n, et l'hypothèse de récurrence est vérifiée au rang 1.

Soit f' le facteur gauche de longueur n de f, et soit $k>1$ le numéro de f' dans X^n ; supposons que l'hypothèse de récurrence est vérifiée pour tous les rangs inférieurs à k.

Si f' n'est pas primitif, $f' = w^s$ où w est primitif ; si w est un mot de LYNDON, alors f' est dans K ; sinon, soit w' le mot de LYNDON conjugué de w(c'est l'élément minimum de la classe des conjugués de w) : $w' < w$ et comme w et w' ont même longueur on a pour tous x r, s dans X^* $wr > w's$.

En particulier, si f'' est le mot de longueur n commençant à la première occurence de w' dans f', le numéro de f'' dans X^n Soit k' est inférieur à k. Or le facteur droit de f qui commence avec f'' a une longueur certainement supérieure à $k'(n-1)+1$ si f a une longueur supérieure ou égale à

$k(n-1)+1$ et en vertu de l'hypothèse de récurrence f'' (donc aussi f) a un facteur dans K.

Si f' est primitif, s'il est un mot de LYNDON, f a un facteur dans K_n puisque les mots de LYNDON de longueur n sont dans K ; sinon f' admet suivant le théorème de CHEN, FOX et LYNDON une décomposition non triviale $f' = g' g''$ où g' est un mot de LYNDON et g'' un produit de mots de LYNDON avec $g'' < f'$ d'après la remarque 2 (on a $f'' = h_1 h_2 \ldots h_m$, $m > 1$ d'après le théorème, et on pose $g'=h_1$, $g''=h_2 h_3 \ldots h_m$). On considère alors le mot f'' de longueur n commençant avec g'' : si $f'' < f'$, l'hypothèse de récurrence entraîne que f'' a un facteur dans K, donc f aussi, dès que f est de longueur supérieure ou égale à $k(n-1)+1$ car alors, comme précédemment, le facteur gauche de f qui commence avec $\overline{f''V}$ est certainement de longueur supérieur ou égale à $(k-1)(n-1)+1$; sinon, on a certainement $f' = g''l$ car s'il n'en était pas ainsi on ne pourrait concilier $g'' < f'$ et $f'' = g'' w * \geq f'$; dans ces conditons $f'=g'g''=g''l$ montre que g' et l sont conjugués, donc il existe u et v dans X^* tels que $g' = uv$, $l=vu$ et $g''=(uv)^p u$, et $f'=(uv)^{p+1}u$, ce qui montre que f' est le facteur gauche de g'^{p+2} de longueur n, donc que f' est dans K_n.

b) Nous pouvons donc conclure que $M_q(n)$ est inférieur à $|K_n|$ qu'il s'agit à présent d'évaluer.

Remarquons d'abord que $|K_n| = |H_n|$ où H_n est l'ensemble des mots de LYNDON de longueur inférieure ou égale à n : étant donné un mot $g \in K_n$, il existe $h \in H_n$ tel que $g=h^s h'$, avec $h=h'h''$, par la définition de K_n ; on a $h > h'$ et par suite le facteur droit minimum de g n'est pas plus long

que h' car pour tout mot de la forme $kh^r h'$ où k est facteur droit de h, donc $k > h$, on a $kh^r h' > h > h'$; il en résulte que la décomposition de g par le théorème de CHEN, FOX et LYNDON commence par h, ce qui entraîne que h est unique. On a ainsi une bijection entre K_n et H_n, et il faut maintenant évaluer $|H_n|$. Or on a vu précédemment que

$q^n = \sum_{d|n} dL_q(n)$, d'où $L_q(n) \leq n^{-1}q^n$ et par suite

$$|H_n| = \sum_{i=1}^{n} L_q(i) \leq n^{-1}q^n \left(1+\sum_{j=1}^{n-1} nj^{-1}q^{j-n}\right),$$ le facteur entre parenthèses

peut s'écrire en posant $k=n-j$ et $x=q^{-1}$ $1 = \sum_{k=1}^{n-1} x^k n/n-k$. Or on a

$n/n-k \leq k+1$ comme on le vérifie facilement.

Donc on a $nq^{-n}M_q(n) \leq 1+\sum_{k=1}^{n-1} (k+1)x^k \leq 1+\sum_{k=1}^{\infty} (k+1)x^k$; et il reste

à montrer que la somme de la série (manifestement convergente) tend vers

0 avec x. Or cette série est la dérivée de la série $\sum_{h=2}^{\infty} x^h = x^2/1-x$, sa

somme est donc égale à $(2x-x^2)/(1-x)^2$, qui tend vers zéro comme x.

Ceci montre donc que <u>si q tend vers l'infini</u>, et quelque soit par ailleurs le comportement de n, <u>$nq^{-n}M_q(n)$ tend vers 1</u>. Il reste à obtenir le même résultat pour n tendant vers l'infini et q restant borné.

Nous allons pour cela construire un nouvel ensemble "coupant".

III. a) Il résulte de la démonstration précédente que l'ensemble K_{n-1} de mots de longueur $n-1$, possède la propriété que tout mot suffisamment long admet un facteur dans K_{n-1}. A partir de ce résultat, nous allons montrer que l'ensemble K'_n défini ainsi :

$$K'_n = \{x^n \; ; \; x \in X\} \cup \{xk \; ; \; k \in K_{n-1} \text{ et } k < x\}$$ possède la

même propriété. Il en résultera que le nombre d'éléments de K'_n sera supérieur à $M_q(n)$. Soit un mot f suffisamment long pour avoir un facteur dans K_{n-1}, soit k, et supposons que k n'est pas facteur gauche de f et que le facteur de longueur $n-1$ qui précède k n'est pas dans K_{n-1}, c'est-à-dire $k=gy$ et que, xk étant facteur de f, xg n'est pas dans K_{n-1} : alors nécessairement xk est dans K'_n, car nécessairement $k < x$. En effet s'il n'en était pas ainsi, xg serait dans K_{n-1} : $x \leq k$ entraîne $x \leq z$, où z est la première lettre de k. Examinons d'abord le cas où $x < z$: étant donné que g s'écrit, en vertu d'un raisonnement précédent $g=h^r h'$ où h est un mot de LYNDON, r un entier positif ou nul et h' un facteur gauche de h, avec en plus $r=0$ si et seulement si h est de longueur $n-1$, il est clair que puisque h, mot de LYNDON, commence par $z > x$, toutes les lettres de h, qui doivent être supérieures ou égales à z, sont strictement supérieures à x ; il en résulte que tous les facteurs de g dans sa décomposition de CHEN, FOX et LYNDON sont des mots de LYNDON supérieurs à x, donc à tout mot commençant par x en vertu des hypothèses, et que le mot xg est, comme on le voit par l'application répétée d'une proposition démontrée précédemment, un mot de LYNDON de longueur $n-1$, donc est dans K_{n-1}. Soit maintenant $x=z$:

si $g=x^{n-2}$, on a $xg=x^{n-1}\in K_{n-1}$; sinon, g a un facteur gauche qui s'écrit $x^u t$, où $t>x$, il résulte des mêmes considérations que précédemment que la décomposition de CHEN, FOX et LYNDON de g s'écrit $g=h_1 h_2 \ldots h_m x^v$ où $0 \leq v \leq u$ et où $h_i > x^{u+1} tw$ pour tout mot w , donc que $xh_1 \ldots h_m$ est un mot de LYNDON de longueur inférieure ou égale à $n-1$ commençant par x^{u+1}, et que par suite xg est dans K_{n-1}.

Nous avons ainsi montrer que si f admet un facteur xgy , si gy est dans K_{n-1} et si xg n'est pas dans K_{n-1} alors xgy est dans K'_n . Si cette situation n'est pas réalisée, c'est que tous les facteurs de longueur $n-1$ de f , à partir du premier sont dans K_{n-1}, car sinon on pourrait appliquer le raisonnement précédent au facteur droit de f dont le facteur gauche de longueur $n-1$ n'est pas dans K_{n-1}. Or en ce cas, soit f admet un facteur de la forme x^n, qui est dans K'_n soit il apparaît nécessairement dans f un digramme xz , avec $x>z$: le facteur de longueur $n-1$ commençant par z, soit k , étant par hypothèse dans K_{n-1} , on a $x>k$ et xk dans K'_n : ceci achève la démonstration que K'_n possède effectivement la propriété requise.

b) Une majoration du nombre d'éléments de K'_n se déduit de celle que nous avons déjà donnée pour le nombre d'éléments de K_n , à savoir $\sum^{n} k^{-1} q^k$, on a donc $M_q(n) \leq q+(q-1) \sum^{n-1} q^k k^{-1}$

$$= n^{-1}q^{n}(n/(n-1) + nq^{-n}\sum_{k=1}^{n-2} q^{k+1}/k(k+1)).$$

Dans le facteur entre parenthèses, le terme $n/(n-1)$ tend vers 1 quand n tend vers l'infini, il reste donc à montrer que le second terme tend alors vers 0. Il peut être écrit, en posant $h=n-k$ et $x=q^{-1}$ (le cas trivial $q=1$ étant éliminé, on peut supposer $x < 1$), $\sum_{h=2}^{n-1} x^{h-1}n/(n-h)(n-h+1)$. De même que précédemment, on a $n/(n-h) \leq h+1$, d'où $n/(n-h)(n-h+1) \leq (h+1)/(n-h+1) = n(h+1)/n(n-h+1) \leq (h+1)(h+2)/n$ par le même raisonnement. Par conséquent le terme en question est inférieur à $\sum_{h=2}^{n-1} x^{h-1}(h+1)(h+2)/n \leq n^{-1} \sum_{h=2}^{\infty} x^{h-1}(h+1)(h+2)$. La série est convergente et par conséquent ce terme tend vers zéro quand n tend vers l'infini, ce qui achève la démonstration de la propriété annoncée.

REFERENCES

[1] K.T. CHEN, R.H. FOX, R.C. LYNDON, Free differential calculus IV, Ann. Math. 68(1958) p. 81-95.

[2] P.M. COHN, Universal Algebra, Harper & Row 1965.

[3] D. FOATA Etude algébrique de certains problèmes d'Analyse combinatoire et du Calcul des Probabilités, Publ. Inst. Statist. Univ. Paris 14(1965) p. 81-241.

[4] M.P. SCHÜTZENBERGER On the synchronizing properties of certain prefix codes, Information and Control 7(1964) p. 23-36.

C H A P I T R E IV

-o-o-o-o-o-o-o-o-

Automates finis et K-langages

A - Le théorème de Kleene

B - Automate minimal reconnaissant un K-langage

C - Automate "boustrophédon".

A. LE THEOREME DE KLEENE

N.B. La bibliographie des automates finis est abondante et recouvre des sujets variés, de l'algèbre à l'électronique. Pour situer les automates finis par rapport aux autres classes d'automates, on pourra consulter le livre de M. GROSS et A. LENTIN [4] qui donne, de plus une bibliographie critique ; sur le point de vue américain, voir M. HARRISON [3] chap. 9-15 ; l'ouvrage de référence des spécialistes de langue allemande est le petit livre de V.M. GLUSCHKOW [2], qui donne un exposé très clair et de précieux renseignements de détail ; on trouvera un traitement algébrique direct du théorème de KLEENE dans l'article de Mc KNIGHT [6].

I. Notion d'automate fini.

a) Définition : Un automate fini $\mathcal{A}$ est un quintuplet

$$\mathcal{A} = \langle X, S, S', s^\circ, f \rangle \quad \text{où}$$

X, appelé alphabet d'entrée de $\mathcal{A}$, est un ensemble non vide quelconque ;

S, ensemble d'états de $\mathcal{A}$, est un ensemble non vide fini disjoint de X ;

S', ensemble des états finaux de $\mathcal{A}$, est un sous-ensemble distingué de S ;

s°, état initial de $\mathcal{A}$, est un élément distingué de S ;

f, la fonction de transition de l'automate, est une application de SxX dans S.

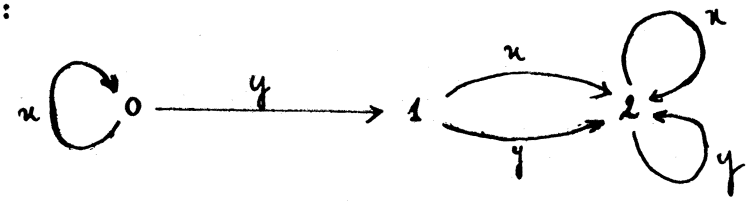
Lorsque X est fini, on peut représenter l'automate fini $\mathcal{A}$ par son graphe $G_{\mathcal{A}}$ (en anglais "state-graph") ainsi défini :

$G_{\mathcal{A}}$ est le multi-graphe (orienté) dont les sommets sont les états de $\mathcal{A}$ ($\in S$) ; de chaque sommet partant Card (X) arcs, chacun d'eux étant étiqueté par une lettre de X ; l'arc ayant $s \in S$ pour extrémité initiale, étiqueté $x \in X$ a pour extrémité terminale $f(s,x) \in S$.

Exemples : $\mathcal{A}_1$: $X = \{x,y\}$ $S = \{0,1,2\}$ $S' = \{1\}$ $s^\circ = 0$

$f(0,x)= 0$, $f(0,y)= 1$, $f(1,x)=f(1,y)= 2$, $f(2,x)=f(2,y)= 2$.

$G_{\mathcal{A}_1}$ prend la forme :

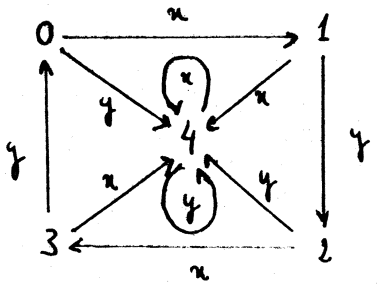


$\mathcal{A}_2$: $X = \{x,y\}$ $S = \{0,1,2,3,4\}$ $S' = \{0,2\}$ $s^\circ = 0$

$f(0,y)=f(1,x)=f(2,y)=f(3,x)=f(4,x)=f(4,y)=4$,

$f(0,x)=1$, $f(1,y)= 2$, $f(2,x)= 3$, $f(3,y)= 0$.

$G_{\mathcal{A}_2}$ prend la forme :



b) Rôle de X,f et s° : structure algébrique de l'automate.

Supposons X fixé : la donnée de f et de s° définit sur S une "structure algébrique" (cf. par exemple SZÀSZ [10] p. 37 et 188 sq. ; COHN [1] p. 47 sq.) ou "algèbre universelle" (KUROSH [5] p. 91 sq.) avec

- une opération zéroaire, qui distingue s° ;

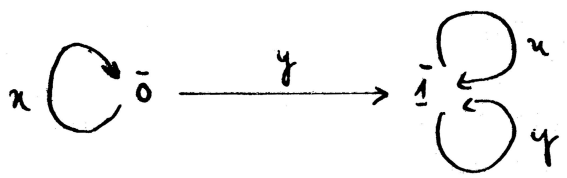
- X comme ensemble d'opérations unaires ; d'ailleurs, comme les applications de S dans lui-même sont en nombre fini, on peut se borner à prendre pour ensemble d'opérations unaires un quotient fini de X.

Nous avons donc les notions habituelles d'homomorphisme, de congruence, et d'automate quotient, de sous-automate (contenant nécessairement s°), de produit direct etc...

Rappelons quelques définitions :

- une <u>congruence</u> β de l'automate $\mathcal{A} = \langle X, S, S^X, s^\circ, f\rangle$ est une équivalence β sur S telle que, pour tout $x \in X$, $s \equiv t$ (mod. β) entraîne $f(s,x) \equiv f(t,x)$ (mod. β).
- dans ces conditions, <u>l'automate quotient</u> est $\mathcal{A}/\beta = \langle X, S/\beta , S'/\beta , \beta(s^\circ) , \bar{f}\rangle$ avec $\bar{f}\,(\beta(s) , x) = \beta(f(s,x))$
- un <u>homomorphisme</u> φ de l'automate $\mathcal{A} = \langle X, S, S', s^\circ, f\rangle$ dans l'automate $\mathcal{B} = \langle X, T, T', t^\circ, g\rangle$ est une application φ de S dans T telle que : $\varphi s^\circ = t^\circ$ et pour tous $x \in X$, $s \in S$, $\varphi f(s,x) = g(\varphi s, x)$.

<u>Exemple</u> : Dans l'automate $\mathcal{A}_1$, l'équivalence β qui a pour classes $\{0\}$ et $\{1,2\}$ est une congruence. Le graphe de $\mathcal{A}_1/\beta$ est



L'homomorphisme canonique de $\mathcal{A}_1$ sur $\mathcal{A}_1/\beta$ est φ : $\varphi\, 0 = \bar{0}$, $\varphi 1 = \varphi 2 = \bar{1}$.

Exercice : Déterminer toutes les congruences et toutes les images homomorphes de l'automate $\mathcal{A}_2$.

f peut être considérée comme une application de X dans le monoïde S^S des applications de S dans lui-même : on peut donc l'étendre canoniquement à un homomorphisme f^* de X^* dans S^S, ce qui se fait par récurrence :

- $f^*(s,e)=s$ pour tout $s \in S$;
- $f^*(s,wx)=f(f^*(s,w),x)$ pour tout $s \in S$, tout $w \in X^*$ et tout $x \in X$.

L'image de X^* par f^* est appelée le monoïde associé à l'automate noté $M_{\mathcal{A}}$: en tant que sous-monoïde de S^S il est fini. Réciproquement un monoïde fini quelconque M, image homomorphe de X^*, par φ permet de définir un automate fini ayant X comme alphabet d'entrée, les éléments de M pour états, l'état initial étant l'élément neutre φe, et pour fonction de transition $f(m,x)=m\varphi x$, obtenu en plongeant M dans M^M par sa représentation régulière droite.

Remarquons qu'un homomorphisme (d'automates) de $\mathcal{A}$ sur $\mathcal{B}$ induit un homomorphisme (de monoïdes) de $M_{\mathcal{A}}$ sur $M_{\mathcal{B}}$.

Se donner un automate fini d'alphabet d'entrée X n'est donc qu'une manière (particulièrement bien adaptée à certains problèmes, comme nous le verrons) de se donner un homomorphisme de X^* dans un monoïde fini.

Remarque : le même monoïde peut être associé à plusieurs automates différents.

Exemple : Le monoïde associé à l'automate $\mathcal{A}_1$ ci-dessus est $M_{\mathcal{A}_1}$:

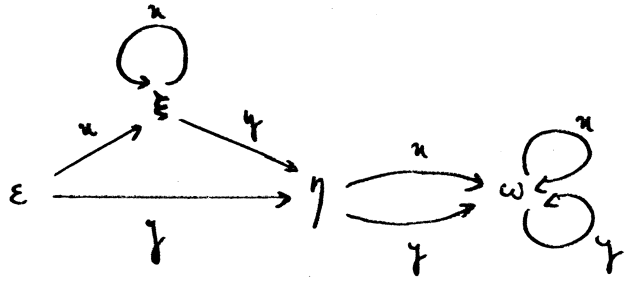
$M_{\mathcal{A}_1} = \{\omega, \varepsilon, \xi, \eta\}$, avec $\varepsilon = \varphi e$, $\xi = \varphi x$, $\eta = \varphi y$ et

$$\varepsilon\begin{cases}0 \to 0\\1 \to 1\\2 \to 2\end{cases} \quad \xi\begin{cases}0 \to 0\\1 \to 2\\2 \to 2\end{cases} \quad \eta\begin{cases}0 \to 1\\1 \to 2\\2 \to 2\end{cases} \quad \omega\begin{cases}0 \to 2\\1 \to 2\\2 \to 2\end{cases} \qquad \text{(zéro de M)}$$

Table de $M_{\mathcal{A}_1}$: :

	ω	ε	ξ	η
ω	ω	ω	ω	ω
ε	ω	ε	ξ	η
ξ	ω	ξ	ξ	η
η	ω	η	ω	ω

Graphe de l'automate associé à $M_{\mathcal{A}_1}$:



Exercices :

1.- Calculer le monoïde $M_{\mathcal{A}_2}$ associé à l'automate $\mathcal{A}_2$ précédent et l'automate déduit de $M_{\mathcal{A}_2}$.

2.- Soient $\mathcal{A}$ un automate fini ayant X pour alphabet d'entrée , $M_{\mathcal{A}} = \varphi X^*$ son monoïde associé et $\mathcal{A}'$ l'automate déduit de $M_{\mathcal{A}}$: montrer que

l'application $\varphi w \rightarrow f^*(s^\circ,w)$ où f est la fonction de transition de $\mathcal{A}$ et s° son état initial, est un homomorphisme d'automates.

c) Rôle de S' : le langage accepté par l'automate.

Définitions : Un mot $w \in X^*$ est accepté par l'automate $\mathcal{A} = \langle X,S,S',s^\circ,f\rangle$ si, et seulement si $f^*(s^\circ,w) \in S'$.

L'ensemble $L(\mathcal{A})$ de tous les mots de X^* qui sont acceptés par l'automate fini $\mathcal{A}$ est appelé le langage accepté par $\mathcal{A}$.

Une partie L de X^* pour laquelle il existe un automate fini tel que $L = L(\mathcal{A})$ est appelé un langage de KLEENE ou K-langage (sur cette terminologie, voir [4]). Nous noterons $\underline{K}(X)$ l'ensemble des K-langages contenus dans X^*.

Exemples : $L_1 = x^*y$ et $L_2 = (xy)^*$ sont des K-langages, car ils sont acceptés respectivement par les automates $\mathcal{A}_1$ et $\mathcal{A}_2$ décrits en I a).

Le langage $L = \{x^n y^n ; n \in \underline{N}\}$ n'est pas un K-langage : supposons qu'il soit accepté par un automate fini $\mathcal{A} = \langle\{x,y\}, S,S',s^\circ,f\rangle$ et soit $p = \text{Card}(S)$. Pour $n > p$ il existe certainement $n' \leq p$ tel que $f(s^\circ,x^n)=f(s^\circ,x^{n'})$ et dans ces conditions si $x^n y^n \in L(\mathcal{A})$ on a aussi $x^{n'} y^n \in L(\mathcal{A})$ donc $L \neq L(\mathcal{A})$.

Exercices :

1.- Déterminer les K-langages reconnus par les trois automates décrits jusqu'ici en prenant pour chacun tous les ensembles d'états terminaux possibles.

2.- Automates finis "non déterministes" : on appelle ainsi un automate fini dans lequel la fonction de transition est multivoque, i.e. $f(s,x) \subset S$ et non plus nécessairement $f(s,x) \in S$; les autres définitions sont les mêmes que dans le cas "déterministe". Montrer que si $L \subset X^*$ est accepté par un automate fini "non déterministe" il existe un automate fini "déterministe" qui accepte L. (Suggestion : passer aux parties de S). Ce résultat cesse d'être vrai pour d'autres classes d'automates (cf. [4]).

Si au lieu de considérer l'automate nous envisageons son monoïde associé, M , nous constatons que le mot $w \in X^*$ est accepté par l'automate si, et seulement si, son image φw dans M envoie s° dans S', $\varphi w(s^\circ) \in S'$; il en résulte que $L(\mathcal{A})$ vérifie $\varphi^{-1} \varphi L(\mathcal{A}) = L(\mathcal{A})$, i.e. que $L(\mathcal{A})$ est une partie saturée modulo la congruence nucléaire de l'homomorphisme φ . Réciproquement, soit $L \subset X^*$ une partie saturée modulo une congruence d'indèx fini θ : alors L est un K-langage, car il est accepté par l'automate associé au monoïde quotient $X^*/_\theta$ pour lequel on prend comme ensemble d'états finaux les classes mod. θ contenues dans L.

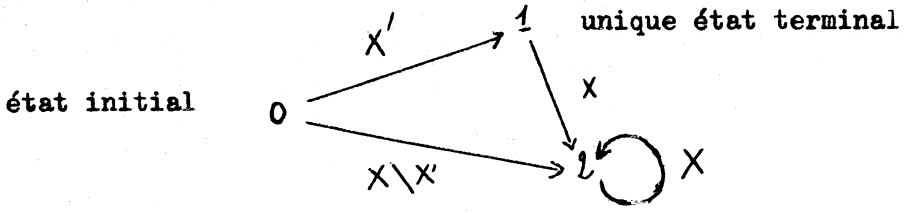
<u>Proposition</u> : $L \subset X^*$ est un K-langage si, et seulement s'il existe un homomorphisme φ de X^* dans un monoïde fini tel que $\varphi^{-1} \varphi L = L$.

<u>Corollaire</u> : L'image inverse d'un K-langage dans un homomorphisme de monoïdes libres est un K-langage.

<u>Exercice</u> : Montrer que si $u,v \in X^*$, $uv \neq vu$, alors $\{u^n v^n ; n \in \underline{N}\} \notin \underline{K}(X)$.

II - Le théorème de KLEENE : $\underline{K}(X)$ est la plus petite famille de parties de X^* contenant les parties de X et fermée par union, produit et passage au sous-monoïde engendré ("étoile").

Remarquons d'abord que $\underline{P}(X) \subset \underline{K}(X)$: $X' \subset X$ est accepté par l'automate ci-dessous :



a) Propriétés de fermeture de $\underline{K}(X)$.

Proposition : $\underline{K}(X)$ est fermé par union, complémentation (donc par intersection), produit et étoile.

- Complémentation : On passe de L à $X^* \setminus L$ en passant de S' à $S \setminus S'$.

- Union : On fait le produit direct des automates.

Si $L = L(\mathcal{A})$, $\mathcal{A} = \langle X, S, S', s^\circ, f \rangle$ et $L' = L(\mathcal{B})$, $\mathcal{B} = \langle X, T, T', t^\circ, g \rangle$ alors $L \cup L' = L(\mathcal{A} \times \mathcal{B})$, avec $\mathcal{A} \times \mathcal{B} = \langle X, S \times T, S' \times T \cup S \times T', (s^\circ, t^\circ), f \times g \rangle$ où $f \times g$ est définie par $f \times g((s,t),x) = (f(s,x), g(t,x))$.

Exercices :

1.- Démontrer directement la fermeture par intersection.

2.- Vérifier que le monoïde associé au produit direct de deux automates

est le produit direct des monoïdes associés aux deux automates composants, et en tirer une preuve de la fermeture de $\underline{K}(X)$ par union et intersection utilisant la proposition du § I c/ .

- Produit : soient $L = L(\mathcal{A})$ et $L' = L(\mathcal{B})$ comme précédemment ; alors $LL' = L(\mathcal{L})$ avec $\mathcal{L} = \langle X, S \times \underline{P}(T)$, $S \times \{A \subset T ; A \cap T' \neq \emptyset\}$, $(s^\circ, \emptyset)$, $h\rangle$ où h se définit par : $h((s,A),x) = (f(s,x), \{g(t,x); t \in A\})$ si $f(s,x) \notin S'$ et $h((s,A),x) = (f(s,x), \{g(t,x); t \in A\} \cup \{t^\circ\})$ si $f(s,x) \in S'$.

- Etoile : Si $L = L(\mathcal{A})$ comme ci-dessus, et si $e \in L$, alors $L^* = L(\mathcal{A}')$ avec $\mathcal{A}' = \langle \underline{P}(S)$, $\{A \subset S ; A \cap S' \neq \emptyset\}$, $\{s^\circ\}$, $F\rangle$ où F est définie par $F(\emptyset,x) = \emptyset$,

  $F(A,x) = \bigcup_{s \in A} F(\{s\},x)$ avec

  $F(\{s\},x) = \{f(s,x)\}$ si $f(s,x) \notin S'$ et $F(\{s\},x) = \{f(s,x), s^\circ\}$ si $f(s,x) \in S'$.

Si $e \notin L$, comme $L^* = (L \cup \{e\})^*$, nous nous ramenons au cas précédent en remarquant $\{e\} \in \underline{K}(X)$ (accepté par exemple par l'automate ci-dessus avec $S' = \{0\}$) et en faisant appel à la fermeture de $\underline{K}(X)$ par réunion.

Remarque : On peut également, pour le produit et pour l'étoile, construire des monoïdes permettant de prouver la fermeture de $\underline{K}(X)$ sans faire appel aux notations des automates finis : pour le produit, voir [8] ; pour l'étoile, la construction est compliquée et ne présente (actuellement) pas d'intérêt particulier.

b) L'algorithme de Mc NAUGHTON & YAMADA [7].

Cet algorithme permet, un automate fini étant donné par son graphe, d'obtenir une expression régulière décrivant le langage qu'il accepte.

Définition des expressions régulières sur un alphabet X :

- Si $X' \subset X$, X' est une expression régulière ;
- Si A et B sont des expressions régulières, $(A)^*$, $(A)(B)$ et $A \cup B$ sont des expressions régulières ;
- Toutes les expressions régulières sont obtenues à l'aide des deux règles ci-dessus.

Il est clair que toute expression régulière décrit un sous-ensemble de X^*, et que l'ensemble des parties de X^* qui sont susceptibles d'une telle description est exactement l'ensemble visé dans le théorème de KLEENE; nous savons d'après les résultats précédents qu'il est inclus dans $\underline{K}(X)$, et l'algorithme de McNAUGHTON & YAMADA va nous donner la réciproque.

Il est aussi clair que des expressions régulières différentes peuvent représenter le même K-langage : par exemple , avec $X = \{x,y\}$, $(yx)^*y \cup (yx)^* (yy \cup x) \left[x \cup y (yx)^* (yy \cup x)\right]^* y(yx)^*y$

$= (yx \cup (x \cup yy)\, x^*y)^*\, y.$

Décrire le langage accepté par l'automate, revient à décrire l'ensemble des chemins de son graphe conduisant de s° aux différents sommets de S' : en fait l'algorithme permet de décrire tous les chemins du graphe allant d'un sommet s à un sommet t, en procédant par récurrence sur le nombre de sommets intermédiaires empruntés par ces chemins.

Nous noterons $V(s,t,P)$, où $s,t \in S$, $P \subset S$ et $s,t \notin P$, l'ensemble des chemins allant de s à t en ne passant entre s et t que par des sommets $\in P$, et pour $s,t \in Q$, $W(s,t,Q)$ l'ensemble des chemins allant de s à t en ne passant par aucun sommet extérieur à Q : ainsi pour Card $P = 0$, $V(s,t,P)$ se réduit à l'ensemble des arcs du graphe joignant s à t, qu'on peut identifier à l'ensemble de leurs étiquettes $X_{s,t} \subset X$; et $W(s,t,Q) = \emptyset$ si Card $Q = 0$.

Pour Card $P = 1$, soit $P = \{p\}$, on aura $V(s,t,P) = X_{s,p}(X_{p,p})^* X_{p,t}$ sous forme d'expression régulière ; et pour Card $Q = 1$, $Q = \{q\}$, $W(q,q,Q) = (X_{q,q})^*$. Le but à atteindre est $W(s,t,S)$: supposons que pour tout Card $P \leq k <$ Card S et pour tout Card $Q \leq k <$ Card S nous soyons en mesure de décrire par expressions régulières tous les ensembles $V(s,t,P)$ et $W(s,t,Q)$, $s,t \in S$: nous venons de voir que cette supposition est réalisée pour $k = 0$ et $k = 1$, nous passons donc à $k + 1$. Soit :

Card $Q = k+1$: on a $W(s,t,Q) = (V(s,s,Q \setminus \{s\}))^* \left[\bigcup_{q \in Q \setminus \{s\}} X_{s,q} W(q,t,Q \setminus \{s\}) \right]$

comme par hypothèse on a $s \in Q$, Card $(Q \setminus \{s\}) = k$ et l'hypothèse de récurrence s'applique.

Soit Card $P = k+1$: on a $V(s,t,P) = \bigcup_{p,p' \in P} X_{s,p} W(p,p',P) X_{p',t}$.

Ceci termine la démonstration du théorème de KLEENE.

c) Conséquence du théorème de KLEENE : l'image d'un $\underline{K}$-langage sur un alphabet fini dans un homomorphisme de monoïdes libres est encore un K-langage.

Soit φ un homomorphisme de X^* dans Y^* ; alors $\varphi\ \underline{K}(X) \subset \underline{K}(Y)$, car si $L \subset X^*$ est obtenu à partir de $\underline{P}(X)$ par un nombre fini d'unions, de produits et d'étoiles, $\varphi\ L$ s'obtient de la même façon à partir de $\underline{P}(\varphi\ X)$; or, X étant fini, pour $X' \subset X$ $\varphi\ X'$ est une partie finie de Y^*, et par suite se calcule à partir des parties finies de Y par union et produit ; $\varphi\ L$ est donc un K-langage.

B. AUTOMATE MINIMAL ACCEPTANT UN K-LANGAGE DONNE.

Soit L un K-langage sur un alphabet X : si $\mathcal{A}$ est un automate fini qui accepte L , nous pouvons supposer que "tous les états de l'automate servent à quelque chose", i.e. que pour tout état s il existe $w \in X^*$ tel que $f^*(s^\circ,w)=s$. L'automate est alors dit connexe : dans ce qui suit <u>tous les automates seront supposés connexes</u>.

I. <u>Automates finis, équivalences régulières à droite et congruences</u>.

A tout automate fini $\mathcal{A}$ (connexe) ayant X comme alphabet d'entrée correspond une équivalence régulière à droite sur X^* notée $\theta_{\mathcal{A}}$, définie par $w \equiv w'$ (mod. $\theta_{\mathcal{A}}$) si, et seulement si, $f^*(s^\circ,w)=f^*(s^\circ,w')$; réciproquement, cette équivalence permet de définir un automate fini noté $\mathcal{A}_\theta$, dont les états sont les classes d'équivalence, l'état initial la classe de e , et la fonction de transition $\bar{f}(A,x)=$ classe de Ax , et qui est isomorphe à l'automate initial $\mathcal{A}$. De la même façon, nous associons à toute équivalence régulière à droite θ d'index fini un automate fini $\mathcal{A}_\theta$.

La relation entre deux équivalences régulières à droite θ et θ' : $\theta \leq \theta'$ se traduit pour les automates associés par "$\mathcal{A}_\theta$, est image homomorphe de $\mathcal{A}_{\theta'}$.

Soit $M = \varphi X^*$ le monoïde associé à l'automate fini $\mathcal{A}$: la congruence nucléaire θ_φ de φ est la congruence de X^* plus fine

que $\theta_{\mathcal{A}}$ maximale. Il est clair que $\theta_\varphi \leq \theta_{\mathcal{A}}$. Soit θ une congruence quelconque plus fine que $\theta_{\mathcal{A}}$: elle est aussi plus fine que θ_φ car $w \equiv w'$ (θ) entraîne $uw \equiv uw'$ (θ) quelque soit $u \in X^*$, et par suite $f^*(s^\circ,uw)=f^*(s^\circ,uw')$ (car $\theta \leq \theta_{\mathcal{A}}$) soit, avec $f^*(s^\circ,u)=s$, $f^*(s,w)=f^*(s,w')$; cette égalité vaut pour tout s puisque $\mathcal{A}$ est connexe et elle signifie $\varphi w = \varphi w'$, i.e. $w \equiv w'$ (θ_φ).

Enfin, il est clair que le langage accepté par l'automate est une partie saturée par l'équivalence régulière à droite associée et que réciproquement, si une partie L de X^* est saturée par une équivalence régulière à droite d'index fini, c'est un K-langage accepté par l'automate que définit cette équivalence, en prenant pour états finaux les classes d'équivalence incluses dans L.

Nous avons ainsi complètement traduit le calcul sur les automates finis en termes d'équivalences régulières à droite et de congruences.

Proposition : Une partie L de X^* est un K-langage si, et seulement s'il existe une équivalence régulière à droite d'index fini pour laquelle L est saturée.

Exercices :

1.- Remplacer dans la proposition ci-dessus "à droite" par "à gauche".

2.- Montrer que la donnée d'une équivalence régulière à droite (resp. à gauche) d'index fini pour laquelle un K-langage L est saturé

permet d'écrire immédiatement une K-grammaire droite (resp. gauche) engendrant L. (cf. [4])

II. Automate minimal, équivalence principale à droite et congruence syntaxique.

Soit L une partie quelconque de X^* : l'équivalence régulière à droite θ_L (resp. la congruence γ_L) maximale pour laquelle L est saturée a pour expression : $w \equiv w' \pmod{\theta_L}$ si, et seulement si, pour tout $v \in X^*$, $wv \in L \Leftrightarrow w'v \in L$; (resp. $w \equiv w' \pmod{\gamma_L}$ si, et seulement si, pour tous $u,v \in X^*$, $uwv \in L \Leftrightarrow uw'v \in L$) L est saturée mod. θ_L et mod. γ_L à cause de la présence dans X^* de l'élément neutre e ; il est clair que θ_L est une équivalence régulière à droite (resp. γ_L une congruence) ; et si θ est une équivalence régulière à droite pour laquelle L est saturée, on a $w \equiv w' \pmod{\theta} \Rightarrow wv \equiv w'v \pmod{\theta}$ donc $wv \in L \Leftrightarrow w'v \in L$ pour tout $v \in X^*$, i.e. $w \equiv w' \pmod{\theta_L}$ et θ_L est bien maximale (même démonstration pour la congruence).

γ_L s'appelle la congruence syntaxique et θ_L l'équivalence principale à droite de L. Nous noterons $\mathcal{A}_L$ l'automate associé à θ_L et M_L le monoïde quotient X^*/γ_L.

Il résulte immédiatement de leurs définitions et des remarques du paragraphe précédent que :

- si $\mathcal{A}$ est un automate (connexe) acceptant L , $\mathcal{A}_L$ est image homomorphe de $\mathcal{A}$. Nous dirons donc que $\mathcal{A}_L$ est <u>l'automate minimal</u> acceptant L.
- si φ est un homomorphisme de X^* sur un monoïde fini tel que $\varphi^{-1}\varphi L = L$, M_L est image homomorphe de φX^* ; M_L est appelé le <u>monoïde syntaxique</u> de L.

Réciproquement, ces propriétés caractérisent l'automate minimal et le monoïde syntaxique à un isomorphisme près.

De plus, on voit facilement que <u>le monoïde syntaxique de L est isomorphe au monoïde associé à l'automate minimal</u> acceptant L : ils sont en effet, en vertu dé ce qui précède, image homomorphe l'un de l'autre.

<u>Exemple</u> : Soit $L = x^*y$. θ_L a trois classes :

- x^*. dont tous les mots sont envoyés dans L par des facteurs (droite),de la forme x^*y ;
- x^*y, dont les mots ne peuvent être envoyés dans L que par le mot vide e ;
- x^*yXX^*, dont les mots ne peuvent pas être envoyés dans L par un facteur droit.

L'automate minimal est isomorphe à $\mathcal{A}_1$.

La congruence syntaxique γ_L de L a quatre classes :

- $\{e\}$ qui peut être envoyé dans L par $\begin{cases} x^* \text{ à gauche et } x^*y \text{ à droite} \\ x^*y \text{ à gauche et } e \text{ à droite} \end{cases}$

- xx^* qui peut être envoyé dans L par x^* à gauche et x^*y à droite seulement ;

- x^*y qui peut être envoyé dans L par x^* à gauche et e à droite ;

- x^*yXX^* qui ne peut pas être envoyé dans L.

La table du monoïde syntaxique est celle de M_{a_1} .

Exercice : Calculer l'automate minimal et le monoïde syntaxique de $(xy)^*$ et de $((xy)^2)^*$.

III. Calcul de l'automate minimal à partir d'un automate acceptant L.

Soit $M = \varphi X^*$ un monoïde quelconque tel que $\varphi^{-1}\varphi L = L$; on peut définir comme en II l'équivalence principale à droite θ_L^M de φL et la congruence syntaxique γ_L^M de φL dans M : ces notions ne font pas appel à la liberté de X^*. D'après le (premier) théorème d'isomorphisme, a_L et Π_L sont isomorphes respectivement à M/θ_L^M et à M/γ_L^M.

Si M est fini, on peut (théoriquement) calculer θ_L^M et γ_L^M par exhaustion :

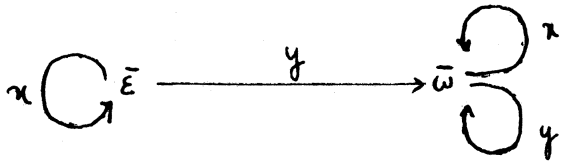
par exemple, soit $L = x^*yX^*$: en prenant $M = M_{a_1}$, il est clair que $L = \varphi^{-1}(\eta,\omega)$. On a alors :

$\xi \equiv \varepsilon$ (mod. θ_L^h) car $\xi\eta,\ \varepsilon\eta,\ \xi\omega,\ \varepsilon\omega \in \{\eta,\omega\}$ et $\xi\varepsilon,\ \xi^2,\ \varepsilon^2,\ \varepsilon \notin \{\eta,\omega\}$.

$\eta \equiv \omega$ (mod. θ_L^η) car $\eta^2,\ \omega\eta,\ \omega^2,\ \eta\omega,\ \eta\xi,\ \eta\varepsilon,\ \omega\xi,\ \omega\varepsilon \in \{\eta,\omega\}$

et $\varepsilon \neq \omega$.

D'où l'automate minimal acceptant L :



Toutefois, ce calcul peut être malaisé :

On peut simplifier ce calcul dans le cas où M est le monoïde associé à un automate fini $\mathcal{A}$ acceptant L : en effet, $\mathcal{A}$ définit dans M (comme dans tout monoïde dont M est image homomorphe) une équivalence régulière à droite (comme en I) qui est plus fine que θ_L^η : d'après le premier théorème d'isomorphisme, $\mathcal{A}_L$ est isomorphe à $\mathcal{A}/\beta$ où β est la congruence de $\mathcal{A}$ induite par θ_L^η .

Le calcul de β est facile : $s \equiv t$ (mod. β) si, et seulement si, pour tout $w \in X^*$, $f^*(s,w) \in S' \Leftrightarrow f^*(t,w) \in S'$, i.e. si, et seulement si, pour tout $m \in M$, $m(s) \in S' \Leftrightarrow m(t) \in S'$.

<u>Exemple</u> : $\mathcal{A}_1$ accepte L $x = x^*yx^*$ si on prend comme ensemble d'états finaux $S' = \{1,2\}$. On peut adopter pour le calcul la disposition suivante :

On écrit en colonne la liste des états de $\mathcal{A}$, et en ligne on énumère les mots de X^*, en supprimant ceux pour lesquels un représentant de la même classe (mod. φ) a déjà été écrit, i.e. pour lesquels on a déjà

rencontré un mot ayant même action sur S : M étant fini, au bout d'un certain temps on n'écrit plus aucun mot, on a alors développé le monoïde M tout entier, et il suffit de cocher dans le tableau les états de S' pour que les états équivalents soient dénoncés par l'identité des lignes de marques qui leur correspondent. Ici on aura :

	e	x	y	yx
0	0	0	1	2
1	1	2	2	2
2	2	2	2	2

car $x^2 \equiv x$, $xy \equiv y$, $y^2 \equiv yx$ et $y^3 \equiv y^2x \equiv yxy \equiv yx$, et on voit que $1 \equiv 2$ (mod. β). On retrouvera bien l'automate minimal précédent.

Exercice : Retrouver de cette manière les automates minimaux acceptant $(xy)^*$ et $((xy)^2)^*$.

C. AUTOMATES FINIS "BOUSTROPHEDON"

Nous exposons ici un résultat de RABIN et SHEPHERDSON [9] : "Tout langage accepté par un automate fini "boustrophédon" (en anglais : "two-way finite automaton") est un K-langage".

I. Notion d'automate "boustrophédon".

Un automate fini "boustrophédon" est un quintuplet $\mathcal{A} = \langle X, S, S', s^\circ, F \rangle$ où X, S, S' et s° sont comme en A.I-a), mais où F est une fonction non plus de $S \times X$ dans S, mais de $S \times X$ dans $\{-1, 0, 1\} \times S$. Mis en présence d'un mot $w \in X^*$, l'automate "boustrophédon" se comporte comme suit :

- dans l'état s° , l'automate subit l'action de la première lettre de w , soit x , i.e. il passe dans l'état s tel que $F(s^\circ, x) = (i, s)$ et selon que $i = 1, 0$ ou -1 il va subir l'action de la deuxième lettre de w, subit à nouveau l'action de la première lettre de w ou s'arrête.
- d'une façon générale, si l'automate dans l'état s subit l'action de la k-ième lettre de w , soit x , il passe dans l'état t tel que $F(s, x) = (i, t)$ et selon que

i=1,0, ou -1 il subit ensuite l'action de la k+1-ème lettre de w , si elle existe, ou s'arrête dans le cas contraire, ou subit à nouveau l'action de la k-ième lettre, ou celle de la k-1-ème si elle existe ou s'arrête sinon.

On voit que l'automate a trois possibilités : s'arrêter en sortant "à gauche" de w , tourner indéfiniment sans sortir de w , ou s'arrêter en sortant "à droite" de w . Le mot w est <u>accepté</u> par l'automate si, et seulement si, ce dernier sort "à droite" du mot et s'arrête alors dans un état $s \in S'$. Comme précédemment, le langage accepté est l'ensemble de tous les mots acceptés par l'automate.

Il est clair que cette notion généralise celle d'automate fini : il suffit de prendre F telle que $F(s,x) \in \{1\} \times S$ pour retrouver la définition précédente.

II. <u>Démonstration du théorème.</u>

Soit $L = L(\mathcal{B})$ un langage accepté par un automate "boustrophédon" $\mathcal{B}$. Nous allons prouver que $L \in \underline{K}(X)$ en montrant que l'équivalence principale à droite θ_L est d'index fini.

Considérons l'application J de $X^* \setminus \{e\}$ dans l'ensemble des applications de $S \cup \{t^\circ\}$ dans $S \cup \{0\}$, où $0, t^\circ \notin S$, ainsi définie :

pour $w \in X^*$, $w \neq e$, $J(w)s = 0$ si l'automate, subissant dans l'état $s \in S$ l'action de la dernière lettre de w , et ensuite se comportant comme il a été dit ci-dessus, sort "a gauche" de w ou n'en sort jamais ;

$J(w)s = t$ si, dans les mêmes conditions que ci-dessus, l'automate sort "à droite" de w dans l'état $t \in S$;

$J(w)t^\circ = 0$ si l'automate, subissant l'action de la première lettre de w dans l'état s° , en sort ensuite à gauche ou n'en sort pas ;

$J(w)t^\circ = s$ si, dans les mêmes conditions, il sort de w "à droite" dans l'état s .

Il est clair que $J(w)=J(w')$ entraîne $w \equiv w' \pmod{\theta_L}$. L'équivalence sur X^* obtenue en laissant e seul dans sa classe et en divisant XX^* selon l'équivalence d'application de J est donc plus fine que θ_L ; mais, S étant fini, l'équivalence d'application de J est d'index fini, donc θ_L aussi, C.Q.F.D.

REFERENCES

[1] P.M. COHN, Universal algebra, Harper & Row 1965

[2] V.M. GLUSCHKOW, Abstrakte Theorie der Automaten, VEB Verlag der Wissen-schaften 1963

[3] M.A. HARRISON, Introduction to switching and automata theory, Mc Graw-Hill 1965

[4] M. GROSS et A. LENTIN, Notions sur les grammaires formelles, Gauthier-Villars 1967

[5] A.G. KUROSH, Lectures in general algebra, Pergamon press 1965

[6] J.D. Mc KNIGHT, Kleene quotient theorems, Pacific J. Math. 14 (1964) p. 1343-1352.

[7] R. Mc NAUGHTON and H. YAMADA, Regular expressions and state graphs for automata, IRE Trans. On Electronic Computers, EC-9 (1960) p. 39-47.

[8] M.P. SCHÜTZENBERGER, On finite monoïds having only trivial subgroups, Information and Control 8 (1965) p. 190-194.

[9] J.C. SHEPHERDSON The reduction of two-way automata to one-way automata, IBM Journal 3 (1959) p. 198-200.

[10] G. SZÀSZ, Einführung in die Verbandstheorie, Akadémiai Kiado, Budapest 1962 (trad. anglaise : Introduction to lattice theory, Academic press 1965).

C H A P I T R E V

-o-o-o-o-o-o-o-

Automates finis et évènements récurrents

A - Codes préfixes

B - Evènements récurrents.

A - CODES PREFIXES

-o-o-o-o-o-o-o-

I.- Circuits dans un automate.

Soient $\mathcal{A}$ un automate, X son alphabet d'entrée, S son ensemble d'états (non nécessairement fini) et f sa fonction de transitions : pour $s \in S$, l'ensemble des circuits du graphe $G_{\mathcal{A}}$ qui passent par s est décrit par

$$A_s^* = \{w \in X^* \ ; \ f^*(s,w) = s\} .$$

La notation A_s^* est justifiée car A_s^* est évidemment un sous-monoïde de X^* ; il vérifie de plus la condition :

$$\underline{U}_r : \ u \in A_s^* \quad \text{et} \quad uv \in A_s^* \quad \text{entraînent} \quad v \in A_s^* .$$

Par suite, nous pouvons choisir comme système générateur de A_s^* l'ensemble A_s des mots non vides de A_s^* qui n'ont aucun facteur gauche propre dans A_s^*, soit

$$A_s = \{w \in XX^* \ ; \ f^*(s,w) = s \quad \text{et} \quad w = uv, u, v \in XX^* \Rightarrow f(s,u) \neq s\}$$

ou encore

$$A_s = A_s^* \setminus (\{e\} \cup (A_s^* \setminus \{e\}) XX^*), \quad \text{et} \quad A_s \quad \text{vérifie la condition :}$$

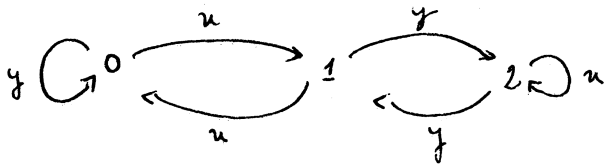
$$\underline{U}'_r : \ A_s \cap A_s XX^* = \emptyset .$$

De plus, on vérifie immédiatement que chaque mot de A_s^* admet une décomposition unique en produit de mots de A_s, i.e. que A_s est un système générateur libre de A_s^*. A_s est appelé un <u>code préfixe</u>.

Exemples : Soit $X = \{x,y\}$.

1.- $\mathcal{A} = \mathcal{A}_1$ du chapitre précédent ; pour $s \neq 4$, on a $A_s^* = ((xy)^2)^*$ et $A_s = (xy)^2$.

2.-Soit $\mathcal{A}$ fini donné par le graphe suivant :



$A_0 = \{y\} \cup x\,(yx^*y)^*x$.

3.-Soit $\mathcal{A}$ donné par : $S = \mathbb{Z}$ et $f(n,x) = n+1$, $f(n,y) = n-1$ pour tout $n \in \mathbb{Z}$:

$A_n^* = \{w \in X^* ; l_x(w) = l_y(w)\}$ pour tout $n \in \mathbb{Z}$, et A_n est l'ensemble des mots "irréductibles" qu'on peut définir par récurrence :

- xy et yx sont irréductibles ;
- si $l(w) > 2$, w est irréductible si et seulement si $w = x\,w_1 w_2 \ldots w_k y$, avec w_i irréductible et $w_i \in yX^*$ pour $i=1,2,\ldots k$ ou $w = y\,w_1 w_2 \ldots w_h\,x$, avec w_j irréductible et $w_j \in x\,X^*$ pour $j = 1,2,\ldots,h$.

<u>Exercice</u> : Trouver des sous-monoïdes de X^* qui ne satisfont pas $\underline{U}_r$ et vérifier que le raisonnement précédent ne donne pas un système générateur.

II.- <u>Codes préfixes</u>.

a) Codes : Un <u>code</u> est une partie A de X^* qui engendre librement A^*, i.e. telle que tout mot de A^* admette une décomposition unique en produit de mots de A.

Il est équivalent de dire que, si Y est un ensemble et φ une bijection de Y sur A, φ se prolonge en un monomorphisme de Y^* sur A^*. Ainsi, le résultat établi au chapitre II-C sur les équations à deux inconnues signifie qu'un ensemble de deux mots $A = \{u,v\}$ est un code si et seulement

si u et v ne sont pas puissances d'un même mot w. Pour une étude générale de ces objets, voir NIVAT [2].

Exercices : 1.- Soit $X = \{x,y\}$; vérifier que $A = \{xx, yxx, yyx, yx, yy\}$ est un code.

2.- Montrer que $A \subset X^*$ est un code si et seulement si A possède la propriété suivante :

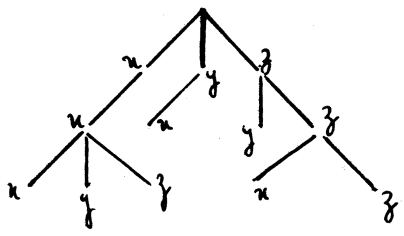
Quelque soit $f \in X^* \setminus A^*$, on a $fA^* \cap A^* \cap A^*f = \emptyset$.

b) Code préfixes : Un code préfixe est une partie A de X^* vérifiant la condition $\underset{-r}{U}'$: $A \cap AXX^* = \emptyset$.

On en tire immédiatment que A est un code et que A^* vérifie la condition $\underset{-r}{U}$.

Nous avons vu en I des exemples de codes préfixes finis et infinis. Remarquons que les codes préfixes finis sur un alphabet à n lettres correspondent bijectivement aux arbres de degré maximum ne dépassant pas n : par exemple, pour $X = \{x,y,z\}$

l'arbre



est associé au code

$\{ xxx, xxy, xxz, yx, zy, zzx, zzz \}$

Un code préfixe complet (ou maximal) sur X est un code préfixe A sur X tel que, quelque soit $B \subset X^*$, si $A \subsetneq B$, B n'est pas un code

préfixe sur X. On tire de cette définition la propriété caractéristique suivante :

$A \subset X^*$ est un code préfixe complet si et seulement si :

1.- Tout mot de X^* admet au plus un facteur gauche dans A ;

2.- Tout mot de X^* soit est facteur gauche d'un mot de A, soit admet un facteur gauche dans A .

On voit aussi que les arbres associés aux codes préfixes complets finis sont les arbres homogènes de degré Card (X).

Exercice : Soit A le code préfixe de l'exemple 3.- de I. Montrer que A est un code préfixe complet mais que $A \cup \{x\}$ est encore un code : A, maximal en tant que code préfixe, n'est pas maximal en tant que code.

Montrer que tout code préfixe complet fini est aussi maximal en tant que code.

c) Codes suffixes et codes bipréfixes : Un code suffixe (ou préfixe gauche) est une partie A de X^* vérifiant la condition

U_{-1} : $A \cap XX^* A = \emptyset$.

Il est clair que les codes suffixes ont des propriétés analogues à celles des codes préfixes, et que l'on passe des uns aux autres par "image miroir" (i.e. la transformation qui au mot $x_{i_1} x_{i_2} \ldots x_{i_k}$, avec $x_{i_j} \in X$ pour $j = 1,2,\ldots,k$, associe le mot $x_{i_k} x_{i_{k-1}} \ldots x_{i_2} x_{i_1}$).

Une partie A de X^* qui est à la fois un code préfixe et un code suffixe est appelée un code bipréfixe : c'est le cas de tous les exemples précédents, mais le lecteur vérifiera sans peine qu'un code préfixe peut ne pas être bipréfixe ! Nous avons vu en a) un code qui n'est ni préfixe ni suffixe.

III.- Codes préfixes et automates finis.

Parmi les exemples de codes préfixes que nous avons envisagés, jusqu'ici, tous, sauf le 3ème de I, sont des K-langages ; nous verrons dans la suite que les codes préfixes qui sont des K-langages ont des propriétés remarquables.

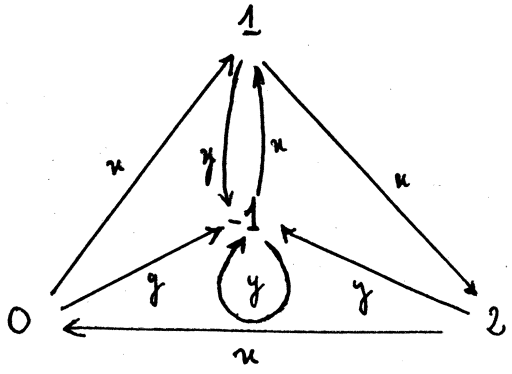
Notons que si un code préfixe A est un K-langage, alors $A = A_{s^o}$ (avec les notations de I) pour un automate fini $\mathcal{A}$.
En effet, il est clair que le code préfixe A est un K-langage si et seulement si A^* est un K-langage (car A^* vérifie $\underline{U}_r$ et on raisonne comme en I). Soit donc $A^* \in \underline{K}(X)$ accepté par un automate fini $\mathcal{A} = \langle S, X, S, S', s^o, f \rangle$. On a $s^o \in S'$ puisque $e \in A^*$; montrons que, si $\mathcal{A}$ est minimal (cf. chap. IV,B), $S' = \{s^o\}$: en effet on a en vertu de $\underline{U}_r$ $f^*(s^o, w) \in S' \iff f^*(s', w) \in S'$ quels que soient $s' \in S'$ et $w \in X^*$, donc, d'après l'algorithme de calcul de l'automate minimal acceptant A^*, s^o et s' sont confondus dans l'automate minimal, i.e. $S' = \{s^o\}$ et $A^* = A^*_{s^o}$.

Une classe importante de codes préfixes est définie ainsi : soit $B \subset XX^*$, on lui associe l'ensemble $A_B = \{w \in X^*,\ w \in X^*B,\ w = uv,\ u, v\ XX^* \Longrightarrow u \notin X^*B\}$ qui est un code préfixe complet, ainsi qu'on le vérifie facilement.

Exercice : Montrer que le code préfixe de l'exemple 2 de I- est complet et qu'il nest pas de la forme A_B.

Comme $A_B = X^*B \setminus X^*BXX^*$, A_B est un K-langage si B est un K-langage.

Par exemple, pour $B = \{x^3\}$, un automate acceptant A_B^* est :



Exercice : Décrire un procédé permettant de construire un automate acceptant A_B à partir d'un automate acceptant B.

B. EVENEMENTS RECURRENTS

I. Définitions.

Soit P une distribution de probabilités sur l'alphabet X, i.e. $P(x) > 0$ et $\sum_{x\in X} P(x) = 1$, et pour $w \in X^*$, $w = x_{i_1} x_{i_2} \ldots x_{i_k}$, soit $P(x) = P(x_{i_1})\, P(x_{i_2}) \ldots P(x_{i_k})$. On a, pour tout entier positif n, $\sum_{w\in X^n} P(w) = 1$; P ne définit donc pas une distribution de probabilités sur X^*, mais en définit une sur chaque X^n.

Pour $L \subset X^*$, notons $\varphi_L(z)$ la fonction génératrice des probabilités $P(X^n \cap L)$, i.e. $\varphi_L(z) = \sum_{n=1}^{\infty} z^n P(X^n \cap L)$. En vertu de la remarque précédente, $\varphi_L(z) \leq \frac{1}{1-z}$.

Considérons une propriété définie pour les mots de X^* : elle peut être symbolisée par sa fonction caractéristique F, $F(w) = 1 \Leftrightarrow w$ possède la propriété et $F(w) = 0 \Leftrightarrow w$ ne la possède pas.

On dit avec FELLER ([1] p. 282) que F définit un <u>évènement récurrent</u> si et seulement si :

Quelques soient u et v dans X^*, on a $F(u) = F(uv) = 1 \Leftrightarrow F(u) = F(v) = 1$;

Ceci signifie que $F^{-1}(1)$ (le <u>support</u> de l'évènement récurrent) est un sous-monoïde de X^* vérifiant la condition $\underline{U}_r$. Il en résulte que l'ensemble $A = \left\{w \in X^* \; ; \; F(w) = 1 \text{ et } w = uv,\ u, v \in XX^* \Rightarrow F(u) = 0\right\}$ est un code préfixe. Réciproquement, si A est un code préfixe, la propriété $F(w) = 1 \iff w \in A^*$, $F(w) = 0 \iff w \notin A^*$ définit un évènement récurrent dont A^* est le support.

<u>Exemple</u> : Soit $X = x,y$ symbolisant le succès et l'échec dans des épreuves de BERNOULLI :

- L'évènement "exactement trois succès de suite viennent de se produire" est un évènement récurrent dont le support est le code préfixe défini en A-III.
- L'évènement "le nombre d'échecs est égal au nombre de succès" est aussi un évènement récurrent ; son support à été introduit en A-I (ex. 3).

On se propose, étant donnée P, de déterminer les fonctions génératrices associées aux ensembles A et A^*, A^* étant à la fois le support d'un évènement récurrent et un K-langage.

Nous dirons (toujours avec FELLER, [1] p. 283) que l'évènement récurrent de support A^* est <u>persistant</u> si $\varphi_A(1) = 1$ et <u>transitoire</u> si $\varphi_A(1) < 1$. Ce sont les deux seules éventualités possibles, car :

Pour tout code préfixe A et tout entier positif n on a :

$$\sum_{k=1}^{n} P(A \cap X^k) \leq 1.$$

La démonstration se fait par récurrence sur n : Soit m l'entier minimum tel que $A \cap X^m \neq \emptyset$; le résultat est trivial pour $n \leq m$. Supposons $n > m$: l'ensemble A' des mots de longueur m qui sont facteurs gauches propres de mots de A est inclus dans $X^m \setminus A$ puisque A est un code préfixe, donc $P(A') \leq 1 - P(A \cap X^m)$; de plus, pour chaque $w \in A'$, l'ensemble $A_w = \{v \in X X^* ; wv \in A\}$ est un code préfixe et on a :

$$\sum_{k=1}^{n} P(A \cap X^k) = \sum_{k=m}^{n} P(A \cap X^k) = P(A \cap X^m) + \sum_{w \in A'} P(w) \sum_{h=1}^{n-m} P(A_w \cap X^h).$$

D'après l'hypothèse de récurrence le second terme de la somme n'excède pas $P(A')$, d'où le résultat.

Remarque : Soit $A = \{a_1, \ldots, a_q\}$ un code fini sur un alphabet X fini à r lettres ; prenons $P(x) = 1/r$ pour tout $x \in X$. On déduit du résultat précédent pour $n \geq \max_i (l(a_i))$, l'inégalité de KRAFT-MACMILLAN :

$$1 \geq \sum_{i=1}^{q} r^{-l(a_i)}$$

(cf. ABRAMSON [3] chap. 3)

II. Calcul des fonctions génératrices.

a) Calcul de φ_{A^*} : Soit A^* accepté par un automate fini $= \langle X, S, \{s^\circ\}, s^\circ, f \rangle$ conformément à A-III. Pour $s \in S$, notons $W_s = \{w \in X^* ; f^*(s^\circ, w) = s\}$ et $\varphi_s(z) = \sum_{n=0}^{\infty} z^n P(X^n \cap W_s)$. Si $n \geq 1$, il est clair que l'ensemble $X^n \cap W_s$ est l'union des ensembles disjoints $(W_t \cap X^{n-1}) . x$ avec $f(t,x) = s$. On en tire $P(X^n \cap W_s) = \sum_{t \in S} P(W_t \cap X^{n-1}) \; P(\{x \in X ; f(t,x) = s\})$; d'autre part, $e \in W_s \Leftrightarrow s = s^\circ$.

D'où, en multipliant par z^n et en sommant, un système d'équations linéaires de la forme :

$$\varphi_{s^\circ}(z) = 1 + z \sum_{t \in S} \varphi_t(z) \; P(\{x \in X ; f(t,x) = s^\circ\})$$

$$\varphi_s(z) = z \sum_{t \in S} \varphi_t(z) \; P(\{x \in X ; f(t,x) = s\}) \quad \text{pour} \quad s \neq s^\circ.$$

Ce système ne fait que traduire la structure de l'automate .

Exemple : Soit $\mathcal{A}$ l'automate donné en A-III. Avec pour probabilités $P(x) = p$, $P(y) = q = 1-p$, on obtient le système :

$$\begin{cases} \varphi_0 &= 1 + zp\varphi_2 \\ \varphi_1 &= zp(\varphi_{-1} + \varphi_0) \\ \varphi_2 &= zp\varphi_1 \\ \varphi_{-1} &= zq(\varphi_0 + \varphi_1 + \varphi_2 + \varphi_{-1}). \end{cases}$$

D'où $\varphi_0 + \varphi_1 + \varphi_2 + \varphi_{-1} = \frac{1}{1-z}$ et $\varphi_0 = \frac{1-z+p^3qz^4}{(1-z)(1-p^3z^3)}$.

b) Calcul de φ_A : on obtient φ_A à partir de φ_{A^*} grâce au

Théorème fondamental : $\boxed{\varphi_{A^*} = (1-\varphi_A)^{-1} \quad , \quad \varphi_A = 1-(\varphi_{A^*})^{-1}}$ Comme on va le voir, ce théorème vaut pour tout code A préfixe ou non, qu'il soit ou non un K-langage.

En effet, comme $A^* = \{e\} \bigcup_1^\infty A^k$, nous pouvons écrire

$X^n \cap A^* = \bigcup_{k=1}^\infty X^n \cap A^k$ dès que $n \geq 1$, soit, comme les ensembles A^k sont deux à deux disjoints (car A est un code)

$P(X^n \cap A^*) = \bigcup_{k=1} P(X^n \cap A^k)$. Il vient

$$\varphi_{A^*}(z) - 1 = \sum_{n=1} \sum_{k=1} z^n P(X^n \cap A^k) = \sum_{k=1} \sum_{n=1} z^n P(X^n \cap A^k)$$ (car les séries sont absolument convergentes pour $z < 1$). Or $z^n P(X^n \cap A^k)$ s'écrit $\sum \left\{ z^n P(w) \; ; \; w = a_1 \ldots a_k \; , \; a_i \in A \; , \; \sum_1^k l(a_i) = n \right\}$ ou encore $\sum \left\{ z^{l(a_1)} P(a_1) z^{l(a_2)} P(a_2) \ldots z^{l(a_k)} P(a_k) \; ; \; a_1 a_2 \ldots a_k \in X^n \cap A^k \right\}$. Donc

$$\varphi_{A^*}(z) - 1 = \sum_{k=1}^\infty \left\{ z^{l(a_1)} P(a_1) \ldots z^{l(a_k)} P(a_k) \; ; \; a_1 \ldots a_k \in A^k \right\} =$$

$\sum_{k=1}^\infty (\varphi_A(z))^k$ d'où le résultat.

Conséquences :

- φ_A et $\varphi_A{}^*$ sont simultanément des fonctions rationnelles :

il résulte du calcul du paragraphe précédent que

Si le code préfixe A est un K-langage, φ_A et $\varphi_A{}^*$ sont des fonctions rationnelles.

FELLER ([1] p. 288) démontre que, si $X = \{x,y\}$ avec $P(x) = p$, $P(y) = q = 1-p$, et pour A le code préfixe de l'exemple 3.- de A-I, on a $\varphi_A{}^* = (1 - 4pqz^2)^{-1/2}$, qui est une fonction algébrique mais non rationnelle.

- Le calcul précédent n'est plus valable pour $z=1$; toutefois, puisque nécessairement $\varphi_A(1) \leq 1$ pour un code préfixe , on en déduit que

$\varphi_A{}^*(1)$ diverge $\Longleftrightarrow$ l'évènement récurrent de support A^* est persistant.

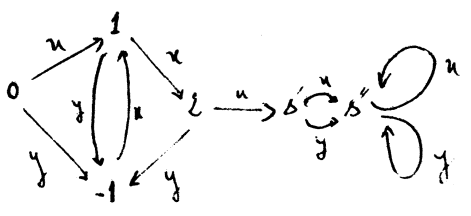
Exemple : L'événement récurrent dont la fonction génératrice a été calculée au paragraphe précédent est persistant.

c) Autre calcul de φ_A : On peut obtenir directement φ_A sans passer par le calcul de φ_{A^*} en utilisant un automate fini acceptant A. La condition $\underline{U'}_r$ satisfaite par A se traduit dans la structure de l'automate minimal acceptant A ainsi :

- il n'y a qu'un seul état final s', distinct de l'état initial s^o ;
- il existe un état s" unique, distinct de s' et de s^o, tel que, pour tout $x \in X$, $f(s',x) = f(s",x) = s"$.

En effet, $e \notin A$ entraîne que s^o n'est pas final ; $\underline{U'}_r$ signifie que pour $s' \in S'$, $f^*(s',w) \in S' \Longleftrightarrow w = e$, d'où $S' = \{s'\}$ dans l'automate minimal, et que, pour $s" = f^*(s',w)$, $w \in XX^*$, on a $f^*(s",v) \notin S'$ quelque soit $v \in X^*$, i.e. $f^*(s",X^*) = \{s"\}$, $s" \neq s^o$.
Soit donc $\mathcal{A}$ un automate fini acceptant A et satisfaisant aux deux conditions ci-dessus (un tel automate peut être construit, par exemple, a partir d'un automate acceptant A^* en lui adjoignant les deux états s' et s" et en remplaçant les transitions de la forme $f(s,x) = s^o$ par des transitions $f(s,x) = s'$). Avec les mêmes notations qu'en a), il est clair que $\varphi_A = \varphi_{s'}$ et que la structure de l'automate se traduit par un système linéaire de la même forme que celui de a), à ceci près que $\varphi_{so} = 1$, qui permet de calculer φ_A .

<u>Exemple</u> : Reprenons l'exemple précédent ; l'automate acceptant A, déduit de celui qui acceptait A^*, sera :



$$\begin{cases} \varphi_0 = 1 \\ \varphi_1 = zp\ (\varphi_0 + \varphi_1) \\ \varphi_2 = zp\ \varphi_1 \\ \varphi_{-1} = zp\ (\varphi_0 + \varphi_1 + \varphi_2 + \varphi_{-1}) \\ \varphi_{s'} = zp\ \varphi_2 \end{cases}$$

On en tire $\varphi_1 = \dfrac{zp}{1-zq-z^2pq-z^3p^2q}$, soit $\varphi_A = \varphi_{s'} = \dfrac{z^3p^3}{1-zq-z^2pq-z^3p^2q}$

L'élévement récurrent correspondant est bien persistant, car

$$1 - \varphi_A = \frac{1-zq-z^2pq-z^3p^2}{1-zq-z^2pq-z^3p^2q} = \frac{(1-z)\ (1+zp+z^2p^2}{1-zq-z^2pq-z^3p^2q}$$

Plus généralement :

d) On a : $\boxed{\varphi_A(z) = (1-z)\ \Psi(z).}$

Soit $\mathcal{A}$ l'automate fini acceptant A envisagé au paragraphe précédent :

notons $$\Psi(z) = \sum_{\substack{s\neq s' \\ s\neq s''}} \varphi_s(z).$$

On a $1 + z\,\Psi = \Psi + \varphi_{s'}$, d'où le résultat.

En effet, $\Psi(z) = \varphi_B(z)$ avec

$$B = \left\{ w \in X^* \ ;\ f^*(s^o,w) \neq s' \text{ et } f^*(s^o,w) \neq s'' \right\}$$

On a

$$\{e\} \cup BX = B \cup A$$

d'après la structure de l'automate $\mathcal{A}$, et d'autre part

$$P(B \cap X^n) = P\ (BX \cap X^{n+1}).$$

Il vient

$$z\ z^n\ P(B \cap X^n) = z^{\,n+1}\ P(B\ X \cap X^{n+1})$$

$$= z^{n+1}(P(B\cap X^{n+1}) + P(A\cap X^{n+1})$$

Car $B\cap A = \emptyset$. D'où, en sommant sur n, l'égalité annoncée.

d) Longueur moyenne des mots d'un code préfixe ;

C'est par définition $L_m(A) = \sum_{w\in A} l(w)P(w) = \sum_{n=0}^{\infty} \sum_{w\in A\cap X^n} n\, P(w)$

D'où $$L_m(A) = \frac{d\,\varphi_A(z)}{dz} \quad (1) .$$

Cette quantité n'est pas nécessairement finie : f. FELLER, [1] p. 288.

Exemple : Dans l'exemple précédent, on a

$$\frac{d\,\varphi_A}{dz} = \frac{3z^2p^3-2z^3p^3q-z^4p^4q}{(1-zq-z^2pq-z^3p^2q)^2} \qquad \text{d'où}$$

$$L_m(A) = \frac{1+p+p^2}{p^3}.$$

III.- Cas des codes préfixes complets qui sont des K-langages

Nous allons montrer qu'alors l'événement récurrent associé est persistant et que la longueur moyenne des mots du code est la somme des probabilités des mots qui sont facteurs gauches propres de mots du code, laquelle est finie.

Lemme : Un code préfixe A qui est un K-langage est complet si et seulement si il existe $a \in X^*$ tel que $X^* a X^* \subset AX^*$.

En effet, l'existence de a entraîne que, si $w \notin AX^*$, alors $wa \in AX^*$,

donc que tout mot de X^* soit admet un facteur gauche dans A, soit est facteur gauche d'un mot de A, i.e. que A est complet.

Réciproquement; si A est complet et accpeté par un automate fini satisfaisant les conditions du paragraphe précédent, alors pour tout état $s \neq s'$ et $s \neq s''$ il existe un mot a_s tel que $f^*(s,a_s) = s''$. On peut alors obtenir un mot a tel que pourtant s, $f^*(s,a) = s'''$, de la façon suivante :

Supposons donnée un ordre sur les états de $\mathcal{A}$, et désignons les par s_i , $i = 1,2,\ldots,r$. Constuisons la famille de mots a'_i , $i = 1,2,\ldots,r$, récursivement par $a'_1 = a_{s_1}$ et $a'_{k+1} = a'_k a_s$ où $s = f^*(s_{k+1},a'_k)$: il est clair que, pour tout $i \leq k$, $f^*(s_i,a'_k) = s''$ et on peut choisir $a = a'_r$.

Montrons maintenant que s'il existe $a \in X^*$ tel que $X^* aX^* \subset AX^*$, alors avec les notations du paragraphe précédent, on a $\Psi(1) < \infty$ et l'événement récurrent dont A^* est le support est persistant.

En effet, soit $\Psi(z) = \sum_{n=0}^{\infty} z^n f_n$, et $\Psi(1) = \sum_{n=0}^{\infty} f_n$: avec $\alpha = l(a)$, on peut écrire $\Psi(1) = \sum_{k=0}^{\alpha} \sum_{n=0}^{\infty} f_{k+n\alpha}$.

$f_{k+(n+1)\alpha} = P(B \cap X^{k+(n+1)\alpha}) \leq P(\{w ; w = uv, u \in B \cap X^{k+n\alpha}, v \neq a\})$ car pour tout $w \in B$ on a $wa \notin B$; d'où $f_{k+(n+1)\alpha} \leq f_{k+n\alpha}(1-P(a))$ et les k+1 séries $f_{k+n\alpha}$ sont toutes majorées par des séries géométriques de raison $(1-P(a))<1$, car $P(x) > 0$ pour tout x entraîne $P(a) > 0$, donc convergent toutes.

On en tire également que $\frac{d\Psi(z)}{dz}(1) < \infty$, donc que $\lim_{z \to 1} (1-z) \frac{d\Psi(z)}{dz} = 0$.

En reprenant l'équation $\varphi_A(z) = (1-z)\ \Psi(z)$ on en déduit

$$\frac{-d}{dz}\,\varphi_A(z) = (1-z)\,\frac{d\,\psi(z)}{dz} - \psi(z)$$

d'où pour $z=1$, en vertu de ce qui précède, $L_m(A) = \frac{d\varphi_A(z)}{dz}(1) = \psi(1) < \infty$.

Or $\psi(z) = \varphi_B(z)$ et B est, pour un code complet, l'ensemble des mots qui sont facteurs gauches propres de mots du code. Le résultat annoncé est donc complètement établi.

Exemple : dans l'exemple précédent, on peut prendre $a = x^3$,

on a $$\psi(z) = \frac{1+zp+z^2p^2}{1-zq-z^2pq-z^3p^2q} \quad \text{et}$$

$$L_m(A) = \frac{1+p+p^2}{p^3} \quad \text{comme précédemment.}$$

Exercice : Soit $X = \{x,y\}$, et $A = X^*yxx \setminus X^*yxxXX^*$

a) Donner la fonction génératrice de A en prenant 1/3 pour probabilité de x et 2/3 pour y. En déduire la longueur moyenne des mots de A.

b) Les probabilités de x et de y étant les mêmes que précédemment, calculer les longueurs moyennes des mots des codes préfixes

$$B = X^*xxy \setminus X^*xxyXX^* \quad \text{et} \quad C = X^*xyx \setminus X^*xyxXX^*$$

et comparer les résultats obtenus.

REFERENCES

[1] W. FELLER — An Introduction to Probability theory and its Applications, 2nd Edition, J. WILEY 1957

[2] M. NIVAT — Eléments de la théorie générale des codes, p. 278-294 in E.R. CAIANIELLO, Ed., Automata theory, Academic Press 1966.

[3] N. ABRAMSON — Information theory and Coding, Mc. Graw Hill 1963.