CODES CIRCULAIRES

J. Berstel[+], D. Perrin[*]

L.I.T.P.

+ Université Paris 6

* Université de Rouen

A paraître dans "Progress in Combinatorics on Words"

(L.J. Cummings ed.) Addison Wesley

INTRODUCTION

Le texte qui suit est une synthèse sur les propriétés connues d'une classe de codes que nous avons baptisés circulaires. Il s'agit essentiellement des codes permettant le déchiffrage de manière unique des mots circulaires. On peut envisager ces mots circulaires de trois façons équivalentes :

(i) une suite de lettres écrites sur un cercle.

(ii) un mot doublement infini périodique.

(iii) la classe d'équivalence d'un mot par décalage circulaire de ses lettres.

Selon les cas, l'une ou l'autre de ces représentations se prête mieux à l'intuition ou aux démonstrations formelles. L'idée d'étudier les propriétés combinatoires des mots circulaires apparait déjà dans l'article d'Axel Thue [23] que l'on peut considérer comme la première contribution importante à la combinatoire des mots.

Les codes circulaires eux-mêmes ont été introduits sous une forme un peu différente par S. Golomb et B. Gordon : les codes à délai de synchronisation bornée. Ces derniers sont eux-mêmes une généralisation des codes "comma free" qui ont excité les imaginations à une époque où une coïncidence numérique avait fait penser que le code génétique était un code comma free : le nombre des acides aminés codés par des suites de trois bases pris parmi un ensemble de quatre bases possibles (A,C,G,U) est égal à 20. Et ce nombre, comme on le verra en section 3, est aussi le nombre maximum $l_3(4)$ d'éléments d'un code comma free formé de mots de longueur 3 que un alphabet à quatre lettres. Malheureusement, on sait maintenant que le code génétique n'est même pas un code, au sens que nous donnons ici à ce terme. Il reste que les codes circulaires forment une famille interessante de codes qui intervient dans de nombreux problèmes de combinatoire des mots.

Le texte est divisé en quatre sections. Dans la première, on donne des définitions et des propriétés générales des codes circulaires. Dans la seconde on étudie le lien entre diverses sous familles parametrées des cadres circulaires : codes limités, codes uniformément synchrones. Dans la troisième section, l'attention est portés sur les distributions par longueurs des codes circulaires. La quatrième section porte sur les factorisations des monoïdes libres qui font intervenir de façon essentielle les codes circulaires.

En général, les énoncés donnés dans le texte ne sont pas démontrés Les preuves peuvent être trouvés dans les articles originaux cités en références. Elles apparaitront aussi dans le livre sur les codes que nous préparons. Ce texte est en fait une première version abrégée du chapitre VII de ce livre. Nous donnons ici une seule preuve complète : celle du Théorème d'Eastman-Scholtz sur les codes comma-free (Théorème 3.3). La preuve que nous présentons, sans être très différente de celle de Scholtz, est basée sur l'idée de factorisation du monoïde libre. Elle nous semble de ce fait plus facile à suivre que celle de Scholtz.

Les notations générales que nous utilisons sont celles de Monsieur Lothaire.

PLAN

1. CODES CIRCULAIRES

Soit X une partie de A^*, c'est-à-dire un ensemble de mots sur l'alphabet A. Soit

$$\varphi : B \to X$$

une bijection d'un ensemble B sur X. On note encore φ le morphisme de B^* dans A^* qui prolonge φ. On dit que X est un <u>code</u> si le morphisme φ est injectif.

Si deux mots w, w' de B^* sont conjugués, c'est-à-dire si

$$w = uv, \quad w' = vu$$

alors $\varphi(w)$ et $\varphi(w')$ sont conjugués puisque

$$\varphi(w) = \varphi(u)\varphi(v), \quad \varphi(w') = \varphi(v)\varphi(u).$$

Ainsi φ définit une application

$$\Phi : \mathcal{D} \to \mathcal{C}$$

de l'ensemble $\mathcal{D}$ des classes de conjugaison de B^* dans l'ensemble $\mathcal{C}$ des classes de conjugaison de A^*.

On dit que X est un code circulaire si l'application Φ est injective. Cette terminologie est justifiée par le fait que la propriété d'être un code circulaire ne dépend que de X et pas de la bijection φ considérée. Clairement, tout code circulaire est un code.

En fait un ensemble $X \subset A^*$ est un code circulaire ssi pour tous $n,m \geq 1$ et $x_1,x_2,\ldots,x_n \in X$ et $y_1, y_2,\ldots, y_m \in X$ et $p \in A^*$ et $s \in A^+$ les égalités.

$$s\, x_2x_3\cdots x_n\, p = y_1y_2\cdots y_m \qquad (1.1)$$

$$x_1 = ps \qquad (1.2)$$

impliquent $n = m$, $p = 1$ et $x_i = y_i$ $(1 \leq i \leq n)$ (voir Figure 1.1). Pour que X soit un code, il faut et il suffit que la condition ci-dessus soit réalisée quand $p = 1$ (et donc $s = x_1$).

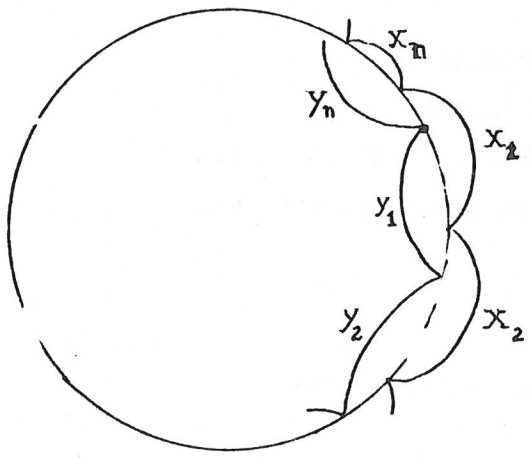

Figure 1.1. Deux factorisations d'un mot circulaire

Intuitivement, un code circulaire est donc un ensemble X de mots tel que tout "mot circulaire" se factorise d'au plus une façon en mots de X.

Toute partie d'un code circulaire est évidemment un code circulaire.

Nous allons maintenant caractériser de plusieurs façons les sous monoïdes engendrés par les codes circulaires.

On dit qu'un sous monoïde M de A^* est pur si pour tous $x \in A^*$ et $n \geq 1$, on a

$$x^n \in M \Rightarrow x \in M \qquad (1.3)$$

On dit que M est très pur si pour tous $u,v \in A^*$ on a

$$uv,\ vu \in M \Rightarrow u,v \in M \qquad (1.4)$$

Un sous monoïde très pur est pur. La réciproque est fausse (cf. Exemple 1.3).

On vérifie facilement la Proposition suivante :

PROPOSITION 1.1. Un sous monoïde M de A^* est très pur ssi il est engendré par un code circulaire.

On peut montrer qu'un sous-monoïde M de A^* est libre, c'est-à-dire engendré par un code ssi pour tous $u,v \in A^*$ on a

$$u,uv,vu \in M \Rightarrow v \in M \qquad (1.5)$$

Ceci fait apparaître directement qu'un sous monoïde très pur est libre.

EXEMPLE 1.1. Soit $A = \{a,b\}$. Le code $X = a^*b$ est circulaire. En effet, on a $X^+ = A^*b$; ainsi, si $uv, vu \in X^*$ alors u et v se terminent par b (ou sont le mot vide) ; ainsi $u,v \in X^*$. □

A toute partie X de A^+ on associe son automate en pétales de la façon suivante :

Soit Q l'ensemble

$$Q = \{ (u,v) \in A^+ \times A^+ \mid uv \in X \} \cup 1.$$

On associe à chaque lettre $a \in A$ la QxQ matrice $\varphi_X(a)$ à coefficients entiers définie de la façon suivante :

$$[\varphi_X(a)]_{p,q} = 1$$

dans les cas suivants

(i) $p = (u,av)$, $q = (ua,v)$
(ii) $p = 1$, $q = (a,v)$
(3i) $p = (v,a)$, $q = 1$
(4i) $p = 1$, $q = 1$, $a \in X$.

$$[\varphi_x(a)]_{p,q} = 0$$

dans les autres cas.

EXEMPLE 1.2 Soit $A = \{a,b\}$ et $X = \{b, ab\}$ qui est un code circulaire puisque $X \subset a^*b$. Chacune des matrices $\varphi(a)$ et $\varphi(b)$ définit une QxQ relation qui peut être representée sur le graphe de la Figure 1.2.

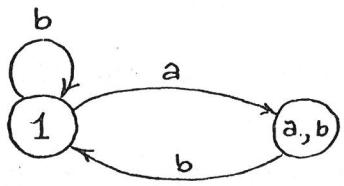

Figure 1.2. L'automate en pétales de $X = \{b, ab\}$.

On a

$$\varphi_X(a) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \varphi_X(b) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Le monoïde $\varphi_X(A^*)$ est constitué de 1, $\varphi_X(a)$, $\varphi_X(b)$ et des deux éléments

$$\varphi_X(ab) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \varphi_X(ba) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

□

On étend φ_X en un morphisme de A^* dans le monoïde des QxQ matrices à coefficients dans $\mathbb{N} \cup \infty$ (en fait toutes les matrices $\varphi(w)$ pour $w \in A^*$ sont à coefficients dans $\mathbb{N}$).

On démontre que pour tout mot w le coefficient

$$[\varphi_X(w)]_{1,1}$$

est égal au nombre de factorisations de w en mots de X.

PROPOSITION 1.2. Soit X une partie de A^+. Les conditions suivantes sont équivalentes :

(i) X est un code très pur.

(ii) Pour tout $w \in A^*$, la trace de la matrice $\varphi_X(w)$ est égale à 0 ou 1.

On peut déduire de la proposition précédente une caractérisation des codes circulaires finis dûe à A. Restivo [16], [17].

PROPOSITION 1.3 Soit $X \subset A^+$ un code fini. Les conditions suivantes sont équivalentes:

(i) X est un code circulaire

(ii) il existe des parties finies T,U,V,W de A telles que

$$X^* = T \cup (UA^* \cap A^*V) - A^*WA^*$$

Les parties (cf. [13], [5], [9], [15]) Y de A^+ satisfaisant la condition (ii) sont appelées strictement localement testables. Pour vérifier qu'un mot w est dans Y, il suffit de regarder s'il est dans T ou s'il a un début dans U, une fin dans V et qu'aucun de ses facteurs n'est dans W.

EXEMPLE 1.3. On a pour $A = \{a,b\}$ et $X = \{b,ab\}$

$$X^+ = A^*b - A^*a^2A^*$$

c'est-à-dire qu'avec les notations de la Proposition 1.3 on a $T = \{1\}$, $U = \{1\}$, $V = \{b\}$, $W = \{a^2\}$. □

Un code (resp. un code circulaire) X est maximal si pour tout code (resp. code circulaire) Y, l'inclusion $X \subset Y$ entraine l'égalité $X = Y$.

Une partie Y de A^* est dense dans A^* si tout mot de A^* est facteur d'un mot de Y. On a alors le résultat suivant :

THEOREME 1.4 Soit X un code (resp. un code circulaire). Alors X^* est dense ssi l'une des conditions suivantes est réalisée :

(i) X est maximal

(ii) X est dense.

L'énoncé concernant les codes est dû à Schützenberger (cf. [5]). Celui qui concerne les codes circulaires est dû à A. De Luca et A. Restivo [2].

On en déduit que si X est un code circulaire qui n'est pas dense, sa maximalité en tant que code équivaut à sa maximalité en tant que code circulaire.

Un code circulaire maximal X n'est jamais fini, sauf si $X = A$. En effet, si c'était le cas, X^* serait dense, Il existerait alors, du fait que X est fini, pour toute lettre $a \in A$ un entier $n \geq 1$ tel que $a^n \in X$. Comme X^* est pur cela impliquerait $a \in X$.

A l'opposé des codes circulaires maximaux, considérons les codes circulaires à deux éléments.

Soit $X = \{x,y\}$. Tout d'abord, X est un code ssi x et y ne sont pas puissance d'un même mot. Ensuite si $X = \{x,y\}$ est un code, soit $Z \subset x^*y \cup y^*x$ l'ensemble des mots $z \in x^*y \cup y^*x$ qui ne sont pas primitifs.

On peut démontrer que seuls les quatre cas suivants sont possibles (cf. [10], [11], [22]) :

(i) $Z = \{x,y\}$

(ii) $Z = \{x,z\}$ ou $Z = \{y,z\}$

(iii) $Z = \{z\}$

(iv) $Z = \emptyset$

On peut démontrer les assertions suivantes (cf. [10]) :

(1) X^* est pur ssi X satisfait la condition (iv)

(2) X est un code circulaire ssi X satisfait la condition (iv) et que les mots x,y ne sont pas conjugués.

Nous illustrons les différentes situations possibles sur trois exemples.

EXEMPLE 1.4. Le code $X = \{b,ab\}$ de l'Exemple 1.2 satisfait la condition (iv) puisque l'ensemble $b^*ab \cup (ab)^*b$ ne contient pas de mots imprimitifs. C'est un code circulaire. □

EXEMPLE 1.5. Le code $X = \{b,aba\}$ satisfait la condition (iii) avec

$$z = ba$$

puisque $z^2 = b(aba)$. Le sous monoïde X^* n'est pas pur. □

EXEMPLE 1.6. Le code $X = \{ab,ba\}$ satisfait la condition (iv). Le sous monoïde X^* est pur. En effet, supposons que $z^n \in X^*$. Supposons d'abord que z soit de la forme $z = uaav$.

Comme aa n'est pas facteur d'un mot de X on déduit de $uaavz^{n-1} \in X^*$ que $ua, avz^{n-1} \in X^*$. De la même façon, on déduit de $z^{n-1}uaav \in X^*$ que $z^{n-1}ua, av \in X^*$. Ainsi $ua, av \in X^*$ et donc $z = uaav \in X^*$.

De la même façon, si $z = ubbv$, on obtient $z \in X^*$. Dans le dernier cas où z est facteur de $(ab)^*$ on obtient $z \in X^*$ en raisonnant sur la parité de la longueur de z. □

On ne connait pas de classification analogue pour le cas d'un code

$$X = \{x,y,z\}$$

constitué de trois mots.

2. CODES LIMITES

Nous avons vu (Proposition 1.3) que les codes circulaires finis sont caractérisés par une condition concernant leur déchiffrage. Cette condition a été introduite sous diverses formes légèrement différentes : "codes à délai de synchronisation bornée", "codes localement déchiffrables", etc... Dans cette section nous étudions cette question de façon systèmatique. Pour cela nous introduisons des familles particulières de codes circulaires définies par des restrictions de plus en plus fortes sur les chevauchements entre mots du code. La famille la plus particulière est celle des codes comma free, sur laquelle nous reviendrons dans la section suivante.

Soient $p,q \geqslant 0$ deux entiers. On dit qu'un sous monoïde M de A^* vérifie la condition $C(p,q)$ si pour toute suite

$$u_0, u_1, \ldots, u_{p+q}$$

de mots de A^*, l'hypothèse

$$u_{i-1}\, u_i \in M \qquad (1 \leqslant i \leqslant p+q) \qquad (2.1)$$

implique

$$u_p \in M$$

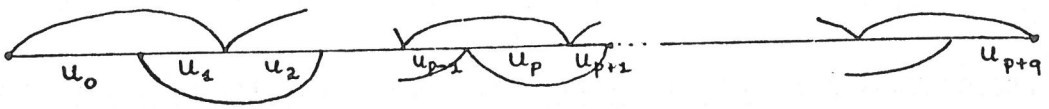

Figure 2.1. La condition $C(p,q)$ (pour p impair et q pair)

Par exemple, la condition $C(1,0)$ s'écrit

$$uv \in M \Rightarrow v \in M$$

et la condition $C(1,1)$ s'écrit

$$uv,\ vw \in M \Rightarrow v \in M$$

Les conditions $C(p,q)$ ont été introduites par Schützenberger [21] sous la forme de conditions $\mathcal{U}_s(p,q)$ pour $p \leqslant 0 \leqslant q$ qui équivalent à $C(-p,q)$.

On vérifie facilement que si M satisfait $C(p,q)$, il satisfait $C(p',q')$ pour $p' \geq p$, $q' \geq q$.

On vérifie facilement la propriété suivante :

PROPOSITION 2.1. Soient $p,q \geq 0$ et M un sous-monoïde de A^*. Si M satisfait $C(p,q)$, il est très pur.

Soit M un sous monoïde qui vérifie une condition $C(p,q)$. D'après la Proposition précédente, M est engendré par un code circulaire X. On dira que X est un code (p,q)-limité, ou simplement limité.

EXEMPLE 2.1. Le seul code $(0,0)$ limité sur A est $X = A$. □

EXEMPLE 2.2. Le code $X = a^*b$ de l'Exemple 1.1. est $(1,0)$-limité. □

EXEMPLE 2.3. Soit $A = \{a_i \mid i \geq 0\}$ et $X = \{a_i\, a_{i+1} \mid i \geq 0\}$.
Le code X est circulaire mais il n'est pas limité. □

D'après la Proposition 2.1, tout code limité est circulaire. L'Exemple 2.3 montre que la réciproque est fausse. Nous allons cependant montrer que, dans le cas des codes reconnaissables, elle est vraie.

Rappelons qu'une partie Y de A^+ est reconnaissable s'il existe un morphisme $\varphi : A^* \to M$ de A^* sur un monoïde fini M tel que $\varphi^{-1}\varphi(Y) = Y$.

PROPOSITION 2.2. Un code reconnaissable X est limité ssi il est circulaire.

L'implication limité $\Rightarrow$ circulaire résulte de la Proposition 2.1. Pour établir l'autre implication, on utilise le fait que si X est un code reconnaissable, il existe un ensemble fini Q et une représentation de A^* par des $Q \times Q$ matrices à éléments 0 ou 1 telle que

$$X^* = \{x \in A^* \mid [\varphi(x)]_{1,1} = 1\}.$$

Pour p+q assez grand on aura pour toute suite u_i $(0 \leqslant i \leqslant p+q)$ satisfaisant (2.1) un $s \in Q$ et des indices k, l avec $k < l$ tels que

$$[\varphi(u_k u_{k+1} \ldots u_1)]_{s,s} = 1$$

$$[\varphi(u_k u_{k+1} \ldots u_1)]_{1,1} = 1$$

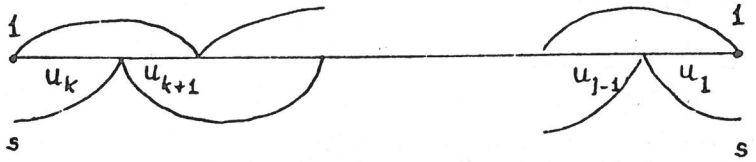

Figure 2.2

En posant $u = u_k$, $v = u_{k+1} \ldots u_1$, on a uv, $vu \in X^*$ d'où $u, v \in X^*$. Ceci montre que $u_k \in X^*$. On montrerait de même que $u_1 \in X^*$

Ceci permet de montrer que X est limité.

Soit $X \subset A^+$ un code. On dit que X est <u>uniformément synchrone</u> s'il existe un entier $s \geqslant 0$ tel que

$$x \in X^s, \quad u, v \in A^*, \quad uxv \in X^* \Rightarrow ux, xv \in X^*. \qquad (2.2)$$

Si l'implication (2.2) est vraie pour un entier s, elle est vraie aussi pour $s' \geqslant s$. Le plus petit entier s tel que (2.1) soit vraie s'appelle le <u>délai de synchronisation</u> de X. On le note $\sigma(X)$.

La notion de code uniformément synchrone apparaît dans [7], [17], [13] ("codes with bounded synchronization delay" ou "locally parsable codes").

EXEMPLE 2.4. Soit $A = \{a_1, a_2, \ldots, a_{2k}\}$ et $X = \{a_i a_j \mid i < j\}$. On peut vérifier que X est uniformément synchrone avec $\sigma(X) = k$. □

PROPOSITION 2.3. Un <u>code uniformément synchrone est limité.</u>

On montre en fait que si $\sigma(X) = s$ alors X est $(2s, 2s)$ limité. L'exemple suivant montre que la réciproque de la Proposition 2.3 est fausse.

EXEMPLE 2.5. Soit $X = ab^*c \cup b$. Le code X est (1,1) limité. Par contre il n'est pas uniformément synchrone : on a pour tout $s \geqslant 0$, $b^s \in X^s$, $ab^sc \in X$ mais ab^s, $b^sc \notin X$. □

Cependant dans le cas des codes finis, on a le résultat suivant :

THEOREME 2.4. Soit X un code fini. Les conditions suivantes sont équivalentes :

(i) X est circulaire.

(ii) X est limité.

(3i) X est uniformément synchrone.

L'équivalence (i) $\Leftrightarrow$ (ii) est un cas particulier de la Proposition 2.2. puisque toute partie finie est reconnaissable

Pour établir (i) $\Rightarrow$ (3i) on peut utiliser la proposition 1.2. L'équivalence (ii) $\Leftrightarrow$ (3i) est due à Restivo [17].

Parmi toutes les conditions introduites dans cette section, voici la plus restrictive : on dit qu'un code X est comma free si pour tous $x \in X^+$ et $u,v \in A^*$ on a

$$uxv \in X^* \quad \Rightarrow \quad u,v \in X^*.$$

Ainsi un code comma free vérifie $\sigma(X) = 1$. On peut de plus vérifier qu'il est (p,q)-limité pour tous p,q tels que $p+q = 3$

3. DISTRIBUTION PAR LONGUEURS

Dans tout ce paragraphe, on suppose que l'alphabet A est fini. Soit X une partie de A^+. Posons pour $n \geqslant 1$

$$\alpha_n = \mathrm{Card}(X \cap A^n) \tag{3.1}$$

On dit que la suite $\alpha = (\alpha_n)_{n \geqslant 1}$ est la distribution par longueurs de X. Nous allons étudier dans ce paragraphe les distributions par longueurs des codes circulaires.

Si X est un code, on a avec $\mathrm{Card}(A) = k$ l'inégalité dite de Kraft-McMillan

$$\sum_{n \geqslant 1} \alpha_n k^{-n} \leqslant 1 \tag{3.2}$$

Réciproquement, si une suite $\alpha = (\alpha_n)_{n \geq 1}$ d'entiers vérifiant (3.2) il existe un code X dont α est la distribution par longueurs. On peut même choisir X préfixe.

Nous allons maintenant déterminer une inégalité plus forte que (3.1) pour les distributions par longueurs des codes circulaires.

Soit $X \subset A^+$ un code. On dit que deux mots $x,y \in X^*$ sont X-conjugués s'il sont de la forme $x = uv$, $y = vu$ avec $u,v \in X^*$. on dit que $x \in X^*$ est X-primitif si pour $y \in X^*$ et $n \geq 1$ l'égalité $x = y^n$ implique $n = 1$. Quand $X = A$, on retrouve les notions usuelles de conjugaison et de primitivité.

On note $l_n(X)$ la nombre de classes de X-conjugaison X-primitives dans $X^* \cap A^n$. Par définition, si X est un code circulaire, on a pour tout $n \geq 1$.

$$l_n(X) \leq l_n(A) \tag{3.3}$$

On a l'égalité dans (3.3) ssi chaque classe de conjugaison primitive dans A^n rencontre X^*.

Il est classique que

$$l_n(A) = \sum_{d \mid n} \mu(d) k^{n/d} \tag{3.4}$$

où μ désigne la fonction de Möbius (cf. [12] par exemple). Nous allons donner une expression de $l_n(X)$ en fonction de la distribution par longueur α de X.

Pour $i \geq 1$, notons

$$\alpha_n^{(i)} = \mathrm{Card}(X^i \cap A^n)$$

Le calcul des nombres $\alpha_n^{(i)}$ se fait facilement par la formule

$$\sum_{n \geq 1} \alpha_n^{(i)} t^n = \left(\sum_{n \geq 1} \alpha_n t^n\right)^i \tag{3.5}$$

On a par exemple

$$\alpha_1^{(1)} = \alpha_1$$

$$\alpha_2^{(1)} = \alpha_2, \quad \alpha_2^{(2)} = \alpha_1^2$$

$$\alpha_3^{(1)} = \alpha_3, \quad \alpha_3^{(2)} = 2\alpha_1\alpha_2, \quad \alpha_3^{(3)} = \alpha_1^3 .$$

Dans l'énoncé qui suit μ désigne la fonction de Möbius et (i,n) le pgcd, des nombres i et n.

PROPOSITION 3.1. Soit $X \subset A^+$ un code. Le nombre de classes de X-conjugaison X-primitives dans $X^i \cap A^n$ est :

$$1_n^{(i)}(X) = \frac{1}{i} \sum_{d \mid (i,n)} \mu(d) \, \alpha_{n/d}^{(i/d)} \qquad (3.6)$$

On a évidemment

$$1_n(X) = \sum_{i \geq 1} 1_n^{(i)}(X)$$

d'où la formule :

$$1_n(X) = \sum_{i \geq 1} \frac{1}{i} \sum_{d \mid (i,n)} \mu(d) \, \alpha_{n/d}^{(i/d)} \qquad (3.7)$$

Dans le cas où X = A, on retrouve évidemment la Formule (3.4) La Formule (3.7) montre que le nombre $1_n(X)$ ne dépend que de la suite α et nous noterons à partir de maintenant $1_n(\alpha)$ au lieu de $1_n(X)$. On notera aussi $1_n(k)$ au lieu de $1_n(A)$ pour Card(A) = k. La Formule (3.7) permet de donner une forme explicite aux inégalités (3.3). Pour n = 1,2,3 on obtient les inégalités suivantes.

$$\alpha_1^{(1)} \leq k$$

$$\alpha_2^{(1)} + \frac{1}{2}(\alpha_2^{(2)} - \alpha_1^{(1)}) \leq \frac{1}{2}(k^2-k)$$

$$\alpha_3^{(1)} + \frac{1}{2}\alpha_3^{(2)} + \frac{1}{3}(\alpha_3^{(3)} - \alpha_1^{(1)}) \leq \frac{1}{3}(k^3-k)$$

$$\alpha_4^{(1)} + \frac{1}{2}(\alpha_4^{(2)} - \alpha_2^{(1)}) + \frac{1}{3}\alpha_4^{(3)} + \frac{1}{4}(\alpha_4^{(4)} - \alpha_2^{(2)}) \leq \frac{1}{4}(k^4-k^2)$$

En termes de la suite (α_n) on obtient

$$\alpha_1 \leq k$$

$$\alpha_2 + \frac{1}{2}(\alpha_1^2 - \alpha_1) \leq \frac{1}{2}(k^2-k)$$

$$\alpha_3 + \alpha_1\alpha_2 + \frac{1}{3}(\alpha_1^3 - \alpha_1) \leq \frac{1}{3}(k^3-k)$$

$$\alpha_4 + \frac{1}{2}(\alpha_2^2 + 2\alpha_1\alpha_3 - \alpha_2) + \alpha_1\alpha_2 + \frac{1}{4}(\alpha_1^4 - \alpha_1^2) \leq \frac{1}{4}(k^4-k^2)$$

La table 3.1 donne pour $k = 2$ les suites $(\alpha_n)_{1 \leq n \leq 4}$ telles que les inégalités ci-dessus soient satisfaites. On n'a figuré que les suites qui sont maximales pour l'ordre usuel.

α_1	2	1	1	1	1	1	0
α_2	0	1	1	0	0	0	1
α_3	0	1	0	2	1	0	2
α_4	0	1	2	1	2	3	3

Table 3.1.

Les inégalités (3.3) avec la forme de $1_n(\alpha)$ donnée par (3.7) ont été données par Golomb et Gordon [7]. Ceux-ci ont émis la conjecture que pour toute suite finie α satisfaisant ces inégalités, il existe un code uniformément synchrone dont α est la distribution par longueurs. Cette conjecture a été démontrée par Schützenberger [21] qui a établi le résultat suivant (retrouvé par Scholtz [18]) :

THEOREME 3.2. *Soit A un alphabet à $k \geq 1$ lettres. Pour toute suite $\alpha = (\alpha_n)_{n \geq 1}$ d'entiers telle que*

$$1_n(\alpha) \geq 1_n(k) \qquad (n \geq 1) \qquad (3.8)$$

il existe un code circulaire X dont α est la distribution par longueurs.

La construction donnée en [21] est la suivante : on définit une suite $(X_n)_{n \geq 1}$ de codes circulaires en posant $X_1 = A$ puis

$$X_{n+1} = x_n^*(X_n - x_n)$$

avec $x_n \in X_n$ un mot de longueur

$$k = \min\{ i \geq 1 \mid \mathrm{Card}(X_n \cap A^i) > \alpha_i \}$$

Si un tel entier k n'existe pas on pose $X_{n+1} = X_n$.

On pose enfin

$$X = \bigcap_{n \geq 1} X_n .$$

EXEMPLE 3.1. Prenons pour suite α la cinquième colonne de la Table 3.1 :

$$\alpha_1 = 1, \quad \alpha_2 = 0, \quad \alpha_3 = 1, \quad \alpha_4 = 2$$

La suite des codes X_n est la suivante (on ne conserve que les mots de longueur au plus 4) :

$$X_1 = \{ a,b \}$$

$$X_2 = \{ b,ab,a^2b,a^3b,\ldots \}$$

$$X_3 = \{ b,a^2b,ab^2,a^3b,\ldots \}$$

$$X_4 = \{ b,ab^2,a^3b,a^2b^2,\ldots \}$$

□

Le théorème précédent appelle quelques remarques :

1. D'après le Théorème 3.2, toute suite α de nombres entiers vérifiant les inégalités (3.8) vérifie aussi l'inégalité (3.2). Ce fait ne semble pas simple à démontrer directement, c'est-à-dire sans un argument combinatoire sur les mots.

2. Ensuite, si une suite α vérifie, en plus des inégalités (3.8), l'inégalité

$$\sum_{n \geq 1} (1_n(\alpha) - 1_n(k)) < +\infty \qquad (3.9)$$

alors α est la distribution par longueurs d'un code circulaire maximal. En effet chacun des codes X_n de la construction précédente est maximal et si α vérifie (3.9) la suite des X_n est stationnaire à partir d'un certain rang. La suite α vérifie alors

$$\sum_{n \geq 1} \alpha_n k^{-n} = 1. \qquad (3.10)$$

Le problème se pose de savoir sous quelles hypothèses (3.9) et (3.10) sont équivalentes pour la distribution par longueurs d'un code circulaire. Notamment, est-il vrai qu'un code circulaire reconnaissable soit maximal ssi sa distribution par longueurs vérifie (3.9), c'est-à-dire ssi, pour tout entier n assez grand, toute classe de conjugaison primitive dans A^n rencontre X^* ?

.3. Les codes X_n de la construction précédente sont préfixes. Cela montre que pour tout code circulaire X, il existe un code circulaire préfixe ayant la même distribution par longueurs. Ceci est lié à l'énoncé suivant [14] : pour tout code circulaire X il existe un réarrangement des lettres des mots de X qui transforme X en un code préfixe.

4. La suite α définie par $\alpha_m = 1_m(k)$ pour un entier $m \geq 1$ et $\alpha_n = 0$ sinon vérifie les inégalités (3.8). On déduit donc du Théorème 3.2. que pour tout entier $m \geq 1$, il existe un code circulaire $X \subset A^m$ tel que

$$\mathrm{Card}(X) = 1_m(k) \qquad (3.11)$$

C'est-à-dire que X est un système de représentants des classes de conjugaison primitives dans A^m. La question se pose de déterminer le délai de synchronisation minimum que l'on peut obtenir pour ces codes.

Golomb, Gordon et Welsh ont conjecturé en [6] que pour tout entier m impair il existe un code comma-free $X \subset A^m$ vérifiant (3.11). Pour m pair cette propriété est fausse et on ne connait pas de borne exacte pour le cardinal d'un code comma free $X \subset A^{2m}$(cf. [7], [1]). Cette conjecture a été résolue par Eastman [4] qui a donné une construction pour tout entier m impair d'un code comma-free $X \subset A^m$ vérifiant (3.11). Quelques années plus tard, SCHOLTZ [19] a donné une autre construction dont le principe est le suivant :

On pose $X_1 = A$ puis récursivement pour $n \geq 1$.

$$X_{n+1} = x_n^*(X_n - x_n)$$

où x_n est un mot de X_n choisi parmi les mots de longueur impaire minimale. Soit

$$Z = \{ x_n \mid n \geq 1 \}$$

THEOREME 3.3. Pour tout entier m impair, le code $X = Z \cap A^m$ est comma free et vérifie (3.11).

La preuve du Théorème 3.3 que nous présentons repose sur une série de lemmes.

LEMME 3.4. Pour tout entier m impair, on a

$$\mathrm{Card}(Z \cap A^m) = 1_m(k).$$

Démonstration : Montrons d'abord par récurrence sur $n \geqslant 0$ que chaque classe de conjugaison dans A^* rencontre un et un seul des sous monoïdes

$$x_1^*, x_2^*, \ldots, x_n^*, X_{n+1}^*$$

Cela est vrai pour $n = 0$. Supposons ensuite la propriété vraie pour $n-1$. Soit $w \in A^n$; alors, soit w a un conjugué dans l'un des sous monoïdes $x_1^*, \ldots, x_{n-1}^*$ soit il a un conjugué w' dans X_n^*. Dans le dernier cas, ou bien $w' \in x_n^*$ ou bien $w' \in X_n^*(X_n - x_n)X_n^*$ et alors w' a un conjugué dans X_{n+1}^*.

Soit maintenant m un entier impair et $w \in A^m$ un mot primitif. Soit n un entier tel que X_{n+1} ne contient aucun mot de longueur impaire inférieure ou égale à m. Alors w ne peut avoir de conjugué dans X_{n+1}^* et donc il existe un unique entier i tel que w ait un conjugué w' dans x_i^*. Comme w est primitif cela implique $w' = x_i$.

Ceci montre que chaque classe de conjugaison primitive de mots de longueur m contient un unique élément de $Z \cap A^m$, d'où la formule cherchée. □

On pose :

$$U = \bigcup_{i \geqslant 1} X_i \,, \quad Y = U \cap (A^2)^* \,, \quad Z = U \cap A(A^2)^* \,.$$

Ainsi Y est l'ensemble des mots de U de longueur paire et Z l'ensemble des mots de U de longueur impaire.

Pour $u \in U$, posons

$$\nu(u) = \min\{ i \in \mathbb{N} \mid u \in X_i \} - 1$$

$$\delta(u) = \sup\{ i \in \mathbb{N} \mid u \in X_i \}$$

On a alors

$$Y = \{ u \in U \mid \delta(u) = +\infty \} \,, \quad Z = \{ x_i \mid i \geqslant 1 \}$$

Notons que

$$\delta(x_i) = i,$$

et que si pour $u \in U-A$, on a $\nu(u) = q$ alors

$$u = x_q v$$

avec $v \in X_{q+1}$. De plus pour tout $u \in U$, et $n \geq 1$

$$\nu(u) \leqslant n < \delta(u) \Rightarrow x_n u \in U \qquad (3.12)$$

Le lemme suivant établit une propriété qui est un exemple des factorisations de monoïdes libres que nous évoquerons à la Section 4.

LEMME 3.5. Tout mot $w \in A^*$ se factorise de façon unique en

$$w = y\, z_1 z_1 \cdots z_n \qquad (3.13)$$

avec $y \in Y^*$, $z_i \in Z$, $n \geq 0$ et $\delta(z_1) \geqslant \delta(z_2) \geqslant \ldots \geqslant \delta(z_n)$.

Démonstration : Montrons d'abord que pour $n \geq 1$ on a

$$\underline{X}_n^* = \underline{X}_{n+1}^* \underline{x}_n^* \qquad (3.14)$$

En effet, on a par définition $X_{n+1} = x_n^*(X_n - x_n)$. Le produit de x_n^* par $X_n - x_n$ est inambigu parce X_n est un code. Donc on a en séries formelles.

$$\underline{X}_{n+1} = \underline{x}_n^*(\underline{X}_n - \underline{x}_n)$$

Il en résulte que $\underline{X}_{n+1} = \underline{x}_n^* \underline{X}_n - \underline{x}_n^+ = \underline{x}_n^* \underline{X}_n - \underline{x}_n^* + 1$ ou encore $\underline{X}_{n+1} - 1 = \underline{x}_n^*(\underline{X}_n - 1)$ d'où (3.14).

On déduit de (3.14) par substitutions successives à partir de $X_1 = A$ que pour tout $n \geq 1$ on a

$$\underline{A}^* = \underline{X}_{n+1}^* \underline{x}_n^* \underline{x}_{n-1}^* \cdots \underline{x}_1^* \qquad (3.15)$$

Soit $w \in A^*$ et $p = |w|$. Soit n un entier tel que X_{n+1} ne contient aucun mot de longueur impaire $\leq p$. Par (3.15) on a une factorisation de w sous la forme

$$w = y\, z_1 z_2 \cdots z_k$$

avec $\delta(z_1) \geqslant \delta(z_2) \geqslant \ldots \geqslant \delta(z_k)$, $z_i \in Z$. De plus y est un mot de X_{n+1}^* de longueur $\leqslant p$; par le choix de n,y est produit de mots de X_{n+1} de longueur paire et on a donc $y \in Y^*$.

Ceci établit l'existence d'une factorisation (3.13). Supposons que w ait une deuxième factorisation

$$w = y' z_1' z_2' \ldots z_{k'}'$$

du même type. Soit m un entier supérieur ou égal à $\delta(z_1)$ et $\delta(z_1')$ et assez grand pour que $y, y' \in X_{m+1}^*$. Un tel choix est possible parce que tous les mots de longueur paire d'un code X_1 appartiennent aussi au code X_m pour $m \geqslant 1$. Les deux factorisations de w mises en évidence sont alors les mêmes par (3.15). □

Nous allons maintenant successivement caractériser la forme de la factorisation (3.13) pour les facteurs gauches puis pour les facteurs droits de mots de U .

LEMME 3.6 Tout facteur gauche propre w d'un mot de U admet une factorisation (3.13) avec $y = 1$.

Preuve : Chacun des X_n est un code préfixe maximal. Il en résulte que pour chaque $n \geqslant 0$ on a

$$\underline{A}^* = \underline{X}_{n+1}^* \underline{P}_{n+1}$$

où P_{n+1} est l'ensemble des facteurs gauches propres de mots de X_n. Il résulte de cette équation et de (3.15) que

$$\underline{P}_{n+1} = \underline{x}_n^* \underline{x}_{n-1}^* \cdots \underline{x}_1^*.$$

Soit w un facteur gauche propre d'un mot u de U . Le mot u appartient à un X_{n+1} et on a alors $w \in P_{n+1}$.

Ceci montre que w admet une factorisation (3.13) avec $y = 1$. □

LEMME 3.7. Pour tous $n, p \geqslant 1$ on a

$$x_n x_{n+p} \in Y^*.$$

Pour tous $z \in Z$ et $y \in Y$ on a

$$zy \in Y^* Z.$$

Preuve : Démontrons la première formule par récurrence sur p.

Pour $p = 1$ on a $x_n x_{n+1} \in U$ d'après (3.12) puisque $\nu(x_{n+1}) < n$.

Comme $x_n x_{n+1}$ est de longueur paire on a donc $x_n x_{n+1} \in Y$. Supposons la propriété vérifiée jusqu'à l'ordre $p-1$ et soit $q = \nu(x_{n+p})$.

Distinguons deux cas: si $q \leqslant n$, alors d'après (3.12) on a $x_n x_{n+p} \in U$. Ensuite, si $n < q$ on a $x_{n+p} \in U - A$ et donc $x_{n+p} = x_q u$ avec $u \in U$.

Comme $q < n+p$ on a $x_n x_q \in Y^*$ par hypothèse de récurrence. Comme u est de longueur paire, il est dans Y et donc $x_n x_{n+p} \in Y^*$ dans ce cas aussi.

Montrons le deuxième formule : posons $n = \delta(z)$ et $q = \nu(y)$. On a $z = x_n$ et $y = x_q x_t$. Si $n < q$ alors $x_n x_q \in Y^*$ d'après ce qui précède, donc $zy \in Y^*Z$. Si $q \leqslant n$ alors $x_n x_q x_t \in U$ d'après (3.12). Comme $x_n x_q x_t$ est de longueur impaire on a donc $x_n x_q x_t \in Z$. □

LEMME 3.8. Tout facteur droit w d'un mot de U admet une factorisation (3.13) avec $n = 0$ ou $n = 1$

Démonstration : il s'agit de montrer que tout facteur droit propre d'un mot $u \in U$ est dans $Y^* Z \cup Y$. Nous établissons cette propriété par récurrence sur $|u|$. Si $|u| = 1$, la propriété est évidente. Supposons $|u| \geqslant 2$.

Soit $n = \nu(u)$. Par définition de ν, on a $u \in X_{n+1}, u \notin X_n$ et donc $u = x_n u'$ avec $u' \in X_{n+1}$.

Soit w un facteur droit de u. Si w est facteur droit de u'. On a $w \in Y^* Z \cup Y^*$ par hypothèse de récurrence. Sinon on a $w = w'u'$, avec w' facteur droit de x_n. Par hypothèse de récurrence, on a $w' \in Y^*Z \cup Y^*$.

Si $w' \in Y^*$ alors $w'u' \in Y^*(Y \cup Z)$ et l'assertion est vérifiée. Il reste donc à examiner le cas où $w' \in Y^*Z$. Dans ce cas, posons $w' = yx_k$ avec $y \in Y^*$, $k \geqslant 1$. On a $k \leqslant n$ car $|x_k| \leqslant |w'| \leqslant |x_n|$.

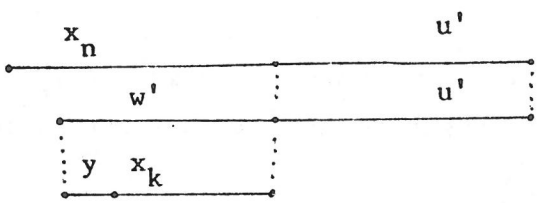

Figure 3.1.

Distinguons maintenant deux cas :

1° cas : $u' \in Y$. D'après le lemme 3.7 on a $x_k u' \in Y^* Z$ donc $w = yx_k u' \in Y^* Z$.

2° cas : $u' = x_m \in Z$. On a $x_m \in X_{n+1}$ donc $m > n$. Comme on a aussi $k \leqslant n$, il vient $k < m$ et $x_k x_m \in Y^*$ d'après le Lemme 3.7. Ainsi $w = y x_k x_m \in Y^*$ et l'assertion est encore vérifiée.

□

Démonstration du Théorème 3.3 : Soit m un entier impair et $X = Z \cap A^m$. D'après le lemme 3.4, on a $\mathrm{Card}(Z \cap A^m) = l_k(m)$. Montrons ensuite que Z est comma free.

Soient $x,x',x'' \in X$. Supposons que

$$xx' = ux''v \tag{3.16}$$

avec $u,v \in A^+$.

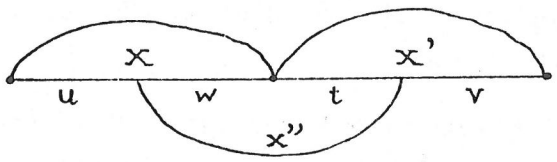

Figure 3.2

On a $x'' = wt$ avec $x = uw$, $tv = x'$. Comme x'' est de longueur impaire, l'un au moins des deux mots w,t doit être de longueur paire. Supposons par exemple que w soit de longueur paire. Comme w est facteur gauche propre de $x'' \in Z$, on a d'après le Lemme 3.6. $w = z_1 z_2 \ldots z_n$ avec $\delta(z_1) \geqslant \delta(z_2) \geqslant \ldots \geqslant \delta(z_n)$. D'autre part, comme w est facteur droit de $x \in Z$, on a d'après le Lemme 3.8, $w \in Y^*Z \cup Y^*$. Et comme w est de longueur paire, on a donc $w \in Y^*$. On obtient ainsi une contradiction avec l'unicité de la factorisation (3.13) du Lemme 3.5. Ainsi on ne peut avoir (3.16) et X est donc comma free.

□

EXEMPLE 3.2. Soit $A = \{a,b\}$. Une suite $(X_n)_{n \geqslant 1}$ satisfaisant les conditions de la construction ci-dessus est la suivante :

$$X_1 = \{ a,b \}$$

$$X_2 = \{ b, ab, a^2b, a^3b, a^4b, \ldots \}$$

$$X_3 = \{ ab, a^2b, a^3b, a^4b, \ldots \quad bab, ba^2b, ba^3b, b^2ab, b^2a^2b, b^3ab, \}$$

$$X_4 = \{ ab, bab, a^3b, a^4b, \dots \}$$
$$ba^2b, ba^3b$$
$$b^2ab, b^2a^2b,$$
$$b^3ab,$$
$$a^2bab,$$

$$X_5 = \{ ab, a^3b, a^4b, \dots \}$$
$$ba^2b, ba^3b,$$
$$b^2ab, b^2a^2b,$$
$$b^3ab,$$
$$a^2bab,$$
$$babab,$$

On n'a pris en compte que les mots de longueur au plus égale à 5 et on fait figurer verticalement la liste des mots de même longueur.

On obtient en prenant les mots de longueur 5 de X_5 tous les mots de longueur 5 de Z. On obtient donc un code comma-free $X \subset A^5$:

$$X = \{ a^4b, ba^3b, b^2a^2b, b^3ab, a^2bab, babab \}$$

tel que $\mathrm{Card}(X) = 1_2(5) = 6$. □

4. FACTORISATION DES MONOÏDES LIBRES

Les constructions données dans la section précédente pour le Théorème 3.2 et pour le Théorème 3.3 sont des cas particuliers de la notion de factorisation des monoïdes libres dont voici la définition :

Soit I un ensemble totalement ordonné et $(X_i)_{i \in I}$ une famille de parties de A^+ indicée par I. Une <u>factorisation ordonnée</u> d'un mot $w \in A^*$ est une factorisation

$$w = x_1 x_2 \dots x_n$$

avec $n \geq 0$, $x_i \in X_{j_i}$ et $j_1 \geqslant j_2 \geqslant \dots \geqslant j_n$.

On dit que la famille $(X_i)_{i \in I}$ est une factorisation de A^* si tout mot $w \in A^*$ a exactement une factorisation ordonnée.

On note $\underline{X}$ la série caractéristique (en variable non commutatives) d'une partie X de A^*. Par définition, la famille $(X_i)_{i \in I}$ est une factorisation ssi on a l'égalité en série formelles:

$$\underline{A}^* = \prod_{i \in I} \underline{X}_i^*$$

EXEMPLE 4.1 Le Lemme 3.5 exprime que la famille

$$(x_1, x_2, \ldots, Y)$$

est une factorisation de A^*. □

Le résultat de base sur les factorisations est le théorème suivant, dû à Schützenberger [20].

THEOREME 4.1. *Soit* $(X_i)_{i \in I}$ *une famille de parties de* A^+ *indicée par un ensemble totalement ordonné* I. *Deux des trois conditions suivantes impliquent la troisième* :

(i) *tout mot* $w \in A^*$ *a au moins une factorisation ordonnée.*

(ii) *tout mot* $w \in A^*$ *a au plus une factorisation ordonnée.*

(3i) *chacun des* $X_i (i \in I)$ *est un code circulaire et chaque classe de conjugaison dans* A^+ *rencontre un et un seul des sous monoïdes* X_i^*.

Le fait que (i) + (ii) $\Rightarrow$ (3i) implique par exemple (via le Lemme 3.4) que le code $X = Z \cap A^m$ du Théorème 3.3 vérifie (3.11).

La preuve du Théorème 4.1 repose sur une technique d'énumération qui donne peu d'information sur les liens entre les parties X_i. On peut espérer en dire plus dans le cas des *factorisations finies*, c'est-à-dire quand l'ensemble I est fini. On peut alors poser $I = \{1,2,..,n\}$ et une factorisation finie est donc une famille

$$X_n, X_{n-1}, \ldots, X_1$$

de codes tels que tout mot $w \in A^*$ s'écrive de façon unique

$$w = x_n x_{n-1} \cdots x_1$$

avec $x_i \in X_i^*$ $(1 \leqslant i \leqslant n)$. On peut encore écrire en séries formelles

$$\underline{A}^* = \underline{X}_n^* \underline{X}_{n-1}^* \cdots \underline{X}_1^*$$

La cas $n = 2$ est celui des *bisections*. On a le résultat suivant (cf. [20], [25] et [12]) :

THEOREME 4.2 Soit (P,Q) une partition de A^+. Il existe une unique bisection (X,Y) de A^* telle que $X \subset P$, $Y \subset Q$.

Cette bisection est obtenue ainsi : posons

$$X_1 = P \cap A, \quad Y_1 = Q \cap A$$

et pour $n \geq 2$,

$$Z_n = \bigcup_{i \geq 1} Y_i X_{n-i}$$

$$X_n = Z_n \cap P, \quad Y_n = Z_n \cap Q$$

Alors

$$X = \bigcup_{n \geq 1} X_n, \quad Y = \bigcup_{n \geq 1} Y_n.$$

Ce théorème donne une construction de toutes les bisections de la façon suivante :

a) on partitionne l'alphabet A en deux parties X_1 et Y_1

b) pour $n \geq 2$, on partitionne l'ensemble

$$Z_n = \bigcup_{i=1}^{n-1} Y_i X_{n-i}$$

en deux parties X_n, Y_n.

On pose enfin $X = \bigcup_{n \geq 1} X_n$, $Y = \bigcup_{n \geq 1} Y_n$.

Les codes X qui sont facteur gauche d'une bisection sont exactement les codes $(1,0)$-limités. Symétriquement, les codes Y qui sont les facteurs droits d'une bisection sont les codes $(0,1)$-limités. Les codes X et Y sont reliés par :

$$Y^* = A^* - XA^*, \quad X^* = A^* - A^*Y.$$

On peut se demander si toute factorisation finie est le résultat d'une suite de bisections. L'exemple suivant, dû à G. Viennot montre que cela n'est pas vrai.

EXEMPLE 4.2. Soit $A = \{a,b\}$. Le code

$$Z = \{b, ba, ba^2\}.$$

est $(0,1)$-limité. C'est donc le facteur droit d'une bisection (X',Z') de A^*. On a

$$X'^* = A^* - A^*Z'.$$

Soient

$$U = (ba)^* ba^3, \quad V = ba, \quad Z = \{ba, ba^2\} (ba)^*.$$

La paire (V,Z) est visiblement une bisection de Z'^*. Ensuite on peut vérifier que l'on a

$$U \subset X'.$$

Il en résulte que U est facteur droit d'une bisection (X,U) de X'^*. avec

$$X = U^*(X' - U).$$

On a de plus.

$$U^*V^* = \{ba, ba^3\}^*$$

Ainsi (U,V) est une bisection de Y^* avec $Y = \{ba, ba^3\}$.

On a donc en séries caractéristiques :

$$\underline{A}^* = \underline{X}'^* \; Z'^* = \underline{X}^*\underline{U}^* \; \underline{V}^*\underline{Z}^* = \underline{X}^*\underline{Y}^*\underline{Z}^*$$

et (X,Y,Z) est une trisection de A^*. Ni X^*Y^*, ni Y^*Z^* ne sont des sous-monoïdes. En effet, $ba \in Y$, $a \in X$ et $ba^2 \in Z$. Donc $ba^2 \notin X^*Y^*$. De même $b \in Z$, $ba^3 \in Y$ et $b^2a^3 \in X$ donc $b^2a^3 \notin Y^*Z^*$. Ceci montre que la trisection (X,Y,Z) ne peut pas être obtenue par deux bisections successives. □

Par contre, G. Viennot a démontré que toute trisection est obtenu par "recollement" de trois bisections [24], comme la trisection de l'exemple précédent :

THEOREME 4.3. Soit (X,Y,Z) une trisection de A^*. Il existe une bisection (U,V) de Y^* et une bisection (X',Z') de A^* telles que (X,U) est une bisection de X'^* et (V,Z) est une bisection de Z'^* :

$$\underline{A}^* = \underline{X}^* \underline{Y}^* \underline{Z}^* = (\underline{X}^*\underline{U}^*)(\underline{V}^*\underline{Z}^*) = \underline{X}'^* \underline{Z}'^*.$$

La preuve de ce résultat est assez délicate. L'un des lemmes préliminaires établit que si (X, Y, Z) est une trisection, alors X,Y et Z sont respectivement (2.0), (1,1) et (0,2)-limités.

Dans le cas d'une quelconque factorisation finie

$$A^* = X_n^* X_{n-1}^* \cdots X_1^*$$

On ne sait même pas prouver que tous les codes X_i sont limités. Le cas n=3 et le cas où la factorisation est obtenue par bisection successives pourraient suggérer la conjecture que le code X_i $(1 \leq i \leq n)$ est (i-1, n-i) - limité.

REFERENCES

[1] Cummings L.J., Comma free codes, unpublished notes, 1980

[2] De Luca, A. and A. Restivo, On some properties of very pure codes, Theoret. Comput. Sci., 10, 157-170, 1980.

[3] Devitt, J.S., Jackson, D.M., Comma free codes; an extension of certain enumerative techniques to recursively defined sequences, J. Comb. Theor., Ser. A, 30, 1-18, 1981.

[4] Eastman, W.L. On the construction of comma-free codes, IEEE Trans. Information Theory, vol. IT-11, 263-266, 1965

[5] Eilenberg, S. Automata, Languages and Machines, Academic Press, Vol. A(1974) Vol. B (1976).

[6] Golomb, S.W., B. Gordon and L.R. Welch, Comma free codes, Can. J. Math., 10, 202-209, 1958.

[7] Golomb, S.W. and B. Gordon, Codes with bounded synchronization delay, Inform. and Control, 8 , 355-376, 1965.

[8] Guibas, L. and A. Odlyzko, Maximal prefix synchronized codes, SIAM J. on Appl. Math.

[9] Hashiguchi, K. and N. Honda, Properties of code events and homomorphisms over regular events, J. Comput. Systems Sci., 12, 352-367(1976).

[10] Lentin, A. and M.P. Schützenberger, A combinatorial problem in the theory of free monoids, in Combinatorial Mathematics and its Applications (R.C. Base and T.A. Dowlings eds.) North Carolina Press, Chapell Hill, 1967, 128-144.

[11] Lerest, E. M, Sur les relations entre un nombre fini de mots, Thèse de 3^{e} cycle, Université de Rouen , 1979.

[12] Lothaire, M. Combinatorics on Words, Addison Wesley, 1982.

[13] Mc Naughton, R. and S. Papert, Counter-Free Automata, MIT Press, 1971.

[14] Perrin, D. et M.P. Schützenberger, Un problème élémentaire de la théorie de l'information, in Théorie de l'Information, Colloques internationaux du CNRS n°276, 249-260, 1977.

[15] Pin, J.E., Une caractérisation de trois variétés de langages bien connues, in Theoretical Computer Science (K. Weihrauch ed.), 4th GI Conférence, Lecture Notes in Comput. Sci 67, Springer Verlag, 1979, 233-243.

[16] Restivo, A., On a question of Mc Naughton and Papert, Inform. and Control, 25, 1, 1974.

[17] Restivo, A., A combinatorial property of codes having finite synchronisation delay, Theoret. Comput. Sci., 1, 95-101, 1975.

[18] Scholtz, R.A., Codes with synchronization capability, IEEE Trans. Information Theory, IT-12, 135-142, 1966.

[19] Scholtz, R.A., Maximal and variable word-length comma-free codes, IEEE Trans. Inform. Theory, IT-15, 1969.

[20] Schützenberger, M.P., On a factorisation of free monoids, Proc. Amer. Math. Soc., 16 , 21-24, 1965.

[21] Schützenberger, M.P., Sur une question concernant certains sous-monoïdes libres, C.R. Acad. Sci. Paris, 261, 2419-2420, 1965.

[22] Spehner, J.C., Quelques problème d'extension, de conjugaison et de présentation des sous-monoïdes d'un monoïde libre, Thèse, Université Paris 7, 1976.

[23] Thue, A., Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norske Vid. Selsk. I. Mat. Nat. Kl., Christiania N°7, 1-67, 1912.

[24] Viennot, G., Algèbres de Lie libres et Monoïdes libres, Thèse, Université Paris 7, 1974.

[25] Viennot, G., Algèbres de Lie libres et Monoides libres, Lecture Notes in Mathematics, 691, Springer Verlag, 1978.