Séminaire de Théorie des Nombres
Année 1978-1979 - exposé n° 18

SUR LA CONSTRUCTION DE MOTS SANS CARRÉ

par

Jean BERSTEL

-:-:-:-

Résumé. - On montre que les quatre constructions de Thue, Morse-Hedlund, Braunholtz et Istrail définissent le même mot infini sans carré. On étudie alors leur lien avec les tag-systèmes. Enfin, on prouve que l'ensemble des mots infinis sans carré sur un alphabet à trois lettres n'est pas dénombrable.

0. - Introduction

La construction de mots sans carré, c'est-à-dire de mots ne contenant pas deux facteurs consécutifs égaux, a fait l'objet de nombreux travaux depuis le premier article consacré à ce sujet par Thue [19] en 1906. L'existence de mots arbitrairement longs sans carré est en effet surprenante et constitue une propriété combinatoire remarquable. De plus, ce problème intervient dans la preuve de l'existence de groupes infirmant la conjecture de Burnside (voir Novikov, Adjan [17]).

Le double but de cet exposé est de présenter et de comparer plusieurs constructions de mots infinis sans carré. Nous montrons notamment que les procédés de Thue [20], Morse, Hedlund [16], Braunholtz [4] et de Istrail [12] définissent le même mot infini sans carré. Il est remarquable que certaines de ces constructions peuvent s'interpréter comme des "tag-systems" (Nous utilisons ce terme dans l'acceptation de Cobham [7] qui diffère de celle de Minsky [15]. Les deux définitions sont des cas particuliers des automates à file de Franchi, Vauquelin [10]). Ceci amène à se poser, dans le cas particulier des mots infinis sans carré, la question que l'on peut formuler en termes généraux comme suit : Comment définir par un procédé fini un objet infini ? Connaissant plusieurs procédés de cette nature, on va en comparer la puissance "générative" en montrant que des suites obtenues par un des mécanismes ne peuvent pas être construites au moyen d'un autre.

L'esquisse d'une hiérarchie ainsi obtenue comporte quatre classes :
Les mots sans carré définis par itération d'un morphisme uniforme (Dejean [8], Leech [11], Zech [21]) ; ceux qui, sans être dans la première classe, peuvent être construits par itération d'un morphisme (donc non uniforme) : le mot de Thue déjà cité plus haut est de cette nature ; le mot d'Aršon [1] ne peut être obtenu de cette façon, mais est engendré par un tag-système [2] ; enfin en observant que les mots infinis sans carré sur un alphabet à trois lettres ne sont pas dénombrables, on constate que tous ces mots ne peuvent être produits de cette manière.

1. - Notations et définitions

Soit M un monoïde libre de type fini et X sa base ; alors M est isomorphe au monoïde, noté X^*, des suites finies d'éléments de X muni de la concaténation. La terminologie suivante est usuelle : X est appelé alphabet, ses éléments des lettres, les éléments de X^* des mots. Tout mot w s'écrit de manière unique comme produit de lettres :

$$w = x_1 x_2 \dots x_n \qquad n \geq 0 \ , \ x_i \in X \ .$$

L'entier n est la longueur de w, aussi notée $|w|$. Le seul mot de longueur 0 est l'élément neutre ε de X^*, nommé le mot vide. On pose $X^+ = X^* - \{\varepsilon\}$. Si a, u, b, w sont des mots, et $aub = w$, alors u est appelé un facteur de w. Si de plus $a = \varepsilon$ (resp. $b = \varepsilon$) alors u est facteur gauche (resp. droit) de w. Un mot est <u>sans carré</u> s'il ne possède pas de facteur de la forme uu, avec $u \neq \varepsilon$.

Un mot infini sur X est une application $\underline{\underline{x}} : \mathbb{N} \to X$. On l'écrit

$$\underline{\underline{x}} = \underline{\underline{x}}(0)\ \underline{\underline{x}}(1) \ldots \underline{\underline{x}}(n) \ldots$$

ou encore

$$\underline{\underline{x}} = x_o\, x_1 \ldots x_n \ldots \qquad (x_i = \underline{\underline{x}}(i) \in X)$$

et on note

$$\underline{\underline{x}}^{[k]} = x_o x_1 \ldots x_{k-1}$$

le facteur gauche de longueur k de $\underline{\underline{x}}$.

On note X^ω l'ensemble des mots infinis sur X et on pose $X^\infty = X^\omega \cup X^*$. Le terme "mot" sera toujours réservé aux mots finis.

Tout morphisme $\alpha : X^* \to Y^*$ étant entièrement déterminé par sa donnée sur X s'étend en une application, encore notée α, de X^∞ dans Y^∞ en posant, pour $\underline{\underline{x}} \in X^\omega$,

$$\alpha(\underline{\underline{x}}) = \alpha(\underline{\underline{x}}(0))\,\alpha(\underline{\underline{x}}(1)) \ldots \alpha(\underline{\underline{x}}(n)) \ldots$$

Etant donné $\underline{\underline{x}}$ et une suite $\underline{\underline{x}}_1, \underline{\underline{x}}_2, \ldots, \underline{\underline{x}}_n, \ldots$ d'éléments de X^∞, on pose

$$\underline{\underline{x}} = \lim \underline{\underline{x}}_n$$

lorsque, pour tout entier K, il existe N tel que

$$\underline{\underline{x}}^{[K]} = \underline{\underline{x}}_n^{[K]} \quad \text{pour tout } n \geq N .$$

La topologie ainsi définie s'obtient comme une topologie produit de la manière suivante (voir aussi Boasson, Nivat [3]) : Soit y une lettre ne figurant pas dans X, soit $Z = X \cup \{y\}$ muni de la topologie discrète, Z^ω muni de la topologie produit. Alors X^∞ s'identifie à la partie T de Z^ω définie par

$$\underline{\underline{z}} \in T \quad \text{si} \quad \underline{\underline{z}}(n) \neq y,\ n > 0 \Rightarrow \underline{\underline{z}}(n-1) \neq y .$$

Clairement T est une partie compacte de Z^ω puisque X est fini ; la topologie induite sur X^∞ par celle de Z^ω est précisément celle donnée ci-dessus.

Lorsque $w_1, w_2, \dots, w_n, \dots$ est une suite d'éléments de X^* de longueurs non bornées, la limite, si elle existe, est un mot infini. Ceci est réalisé notamment lorsque chaque w_n est facteur gauche de w_{n+1} et distinct de w_{n+1}, et plus particulièrement dans la situation suivante. Soit $\alpha : X^* \to X^*$ un morphisme, et supposons qu'il existe une lettre $x_o \in X$, telle que $\alpha(x_o) = x_o b$ pour un mot non vide b. Posons $w_n = \alpha^n(x_o)$; on a :

$$w_{n+1} = \alpha^n(\alpha(x_o)) = \alpha^n(x_o b) = w_n \alpha^n(b) .$$

Si $\alpha^n(b) \neq \varepsilon$, la limite $\lim \alpha^n(x_o)$ est donc un mot infini que nous notons $\alpha^\omega(x_o)$. On a : $\alpha(\alpha^\omega(x_o)) = \alpha^\omega(x_o)$.

L'usage de mots infinis est commode lorsque l'on a affaire à une propriété P sur les mots (finis) qui est préservée par passage aux facteurs, autrement dit telle que $P(aub) \Rightarrow P(u)$, ce qui estle cas pour la propriété "être sans carré". Comme l'alphabet X est fini, donc X^∞ est compact, l'existence d'une infinité de mots vérifiant P équivaut alors à l'existence d'un mot infini $\underline{\underline{x}}$ dont tous les facteurs ont la propriété P, ce que l'on exprime en disant que $\underline{\underline{x}}$ lui-même vérifie P. De manière duale, si l'ensemble des mots (finis) vérifiant P est un idéal bilatère I de X^*, le fait que tout mot infini ait un facteur dans I équivaut à dire que X^*-I est fini (Justin [13]).

2. - Quatre définitions équivalentes d'un même mot

Sur un alphabet à deux lettres x et y, l'on a vite fait le tour des mots sans carré. Les seuls mots de cette nature sont x, y, xy, yx, xyx et yxy. Sur un alphabet à trois lettres, la situation change totalement. Il en existe une infinité, donc des mots infinis sans carré. Le premier mot infini sans carré a été donné par Thue [19] en 1906. D'autres de ces mots ont depuis été trouvés (entre autres) par Arson [1], Dejean [8], Istrail [12], Leech [14], Morse, Hedlund [16], Pleasants [18], Thue [20], Zech [21]. Dans cette section, nous prouvons que les quatre mots infinis sans carré construits indépendamment, et de manière différente, par Thue [20], Morse et Hedlund [16], Braunholtz [4] et Istrail [12] sont en fait le même mot. Les trois premières constructions se font en deux étapes : la première consiste en la définition d'un mot infini particulier sur un alphabet à deux lettres qui, dans une deuxième phase, est transformé pour donner le mot cherché. La constuction d'Istrail est directe.

Dans ce paragraphe, X, Y, Z sont trois alphabets fixés : $X = \{0,1\}$, $Y = \{0, 1, 2\}$ et $Z = \{0, 1, 2, 3\}$.

a) Le mot de Thue-Morse. - Soit $\mu : X^* \to X^*$ le morphisme défini par

$$\mu(0) = 01 \ , \quad \mu(1) = 10 \ .$$

En vertu des remarques préliminaires, μ définit deux mots infinis

$$\underline{\underline{t}} = \mu^{\omega}(0) \ , \quad \bar{\underline{\underline{t}}} = \mu^{\omega}(1) \ .$$

La suite de Thue-Morse est le mot infini $\underline{\underline{t}}$:

$$\underline{\underline{t}} = 0110100110010110100101100110100110010110011010 01\ldots \ .$$

Cette suite a de nombreuses propriétés. Ainsi

$$\underline{\underline{t}}(n) \equiv d_2(n) \quad \text{mod. } 2$$

où $d_2(n)$ est le nombre de "1" figurant dans l'écriture binaire de n. Comme $\mu(\underline{\underline{t}}) = \underline{\underline{t}}$, on a notamment

$$\mu(t_n) = t_{2n}\, t_{2n+1} \qquad n \geq 0 \tag{1}$$

formule qui servira par la suite. La propriété fondamentale -pour notre propos- est la suivante :

PROPOSITION 1 (Thue [20], Morse, Hedlund [16]). - Le mot $\underline{\underline{t}}$ ne contient pas de facteur de la forme $xuxux$, avec $x \in X$, $u \in X^*$.

En particulier, $\underline{\underline{t}}$ ne contient pas de cube.

b) La méthode de Morse-Hedlund. - Soit $\underline{\underline{x}} = x_0 x_1 \ldots x_n \ldots \in X^{\omega}$. Chaque facteur $x_n x_{n+1}$ peut être interprété comme un nombre compris entre 0 et 3 et écrit en base 2. Si l'on réécrit ce nombre en base 4, on obtient à partir de $\underline{\underline{x}}$ un mot infini $\underline{\underline{x}}'$ sur Z. Formellement, $\underline{\underline{x}}'$ est défini par

$$x'_n = 2x_n + x_{n+1} \qquad n \geq 0 \ .$$

A partir de $\underline{\underline{x}}'$ on construit un mot infini $\underline{\underline{x}}''$ sur l'alphabet Y en identifiant 3 à 0. Ainsi

$$\underline{\underline{x}}'' = \beta(\underline{\underline{x}}')$$

où le morphisme $\beta : Z^* \to Y^*$ est donné par

$$\beta(0) = \beta(3) = 0 , \quad \beta(1) = 1 , \quad \beta(2) = 2 .$$

Partant de la suite de Thue-Morse $\underline{\underline{t}}$, on obtient successivement :

$$\underline{\underline{t}}' = 1321201320121321201213201321201320121320132120 1\ldots$$

et

$$\underline{\underline{s}} = \underline{\underline{t}}'' = 102120102012102120121020102120102012102010212 01\ldots .$$

PROPOSITION 2 (Morse-Hedlund [16]). - <u>Le mot infini</u> $\underline{\underline{s}}$ <u>est sans carré</u>.

La transformation qui à $\underline{\underline{x}}$ associe $\underline{\underline{x}}''$ est susceptible d'une définition directe intéressante. On vérifie en effet sans peine que

$$x''_n \equiv x_{n+1} - x_n \quad \text{mod. } 3 \tag{2}$$

Cette formule permet de déduire de la proposition 2 de la proposition 1. La suite $\underline{\underline{t}}'$ peut aussi être définie par itération d'un morphisme.

PROPOSITION 3. - <u>On a</u> $\underline{\underline{t}}' = \alpha^{\omega}(1)$, <u>où</u> $\alpha : Z^* \to Z^*$ <u>est le morphisme donné par</u> :

$$\alpha(0) = 12 , \quad \alpha(1) = 13 , \quad \alpha(2) = 20 , \quad \alpha(3) = 21 .$$

<u>Preuve</u>. - Nous étendons d'abord l'opération ' aux mots en posant, pour $w = x_o x_1 \ldots x_p \in X^*$, $w' = x'_o x'_1 \ldots x'_{p-1} \in Z^*$. Ensuite, on vérifie les formules

$$(\mu^n(0)0)' = \alpha^n(0) , \ (\mu^n(0)1)' = \alpha^n(1) , \ (\mu^n(1)0)' = \alpha^n(2) ,$$
$$(\mu^n(1)1)' = \alpha^n(3)$$

par récurrence sur n , le cas $n=0$ étant évident. Comme $\mu^n(0)$ (resp. $\mu^n(1)$) commence par un "0" (resp. un "1"), on a par exemple

$$(\mu^{n+1}(0)0)' = (\mu^n(0)\ \mu^n(1)0)' = (\mu^n(0)1)'(\mu^n(1)0)'$$
$$= \alpha^n(1)\ \alpha^n(2) = \alpha^{n+1}(0) .$$

De même pour les autres cas. Ceci montre que $\alpha^n(1)$ est facteur gauche de $\underline{\underline{t}}'$, pour tout n ; l'assertion en découle. $\square$

COROLLAIRE 4. - On a :

$$\underline{\underline{s}} = \beta(\alpha^{\omega}(1)) . \qquad \square$$

c) La méthode de Braunholtz. - Une autre méthode pour transformer $\underline{\underline{t}}$ en un mot sans carré a été donnée par Braunholtz [4]. On verra qu'elle s'apparente à la méthode de Thue. En vue de la proposition 1, deux occurences consécutives de la lettre "0" dans $\underline{\underline{t}}$ sont séparées par 0 , 1 ou 2 occurences de la lettre "1" . A partir de $\underline{\underline{t}}$, on peut donc construire un mot infini

$$\underline{\underline{b}} = b_o b_1 \dots b_n \dots$$

sur Y , où b_n est le nombre de "1" entre la n-ième et la n+1-ième occurence de la lettre "0" dans $\underline{\underline{t}}$ (on commence la numérotation à zéro). Procédant ainsi, on obtient le mot

$$\underline{\underline{b}} = 210201210120210201202 10\dots$$

ce qui suggère la

PROPOSITION 5. - On a

$$b_n \equiv 1+s_n \quad \underline{\text{mod.}}\ 3 \quad \underline{\text{pour}}\ n\geq 0 .$$

Preuve. - Notons p_n l'indice de la n-ième lettre de $\underline{\underline{t}}$ qui est égale à "0" . On a $p_o = 0$, $p_1 = 3$, $p_2 = 5$, $p_3 = 6, \dots$. Comme $t_{2n} t_{2n+1} = 01$ ou 10 pour tout n , on a $p_n = 2n$ ou $p_n = 2n+1$. Plus précisément, la formule (1) donne :

$$p_n = \begin{cases} 2n & \text{si } t_n = 0 \\ 2n+1 & \text{si } t_n = 1 \end{cases}$$

soit

$$p_n = 2n + t_n \quad n\geq 0 .$$

Par définition :

$$b_n = p_{n+1} - p_n - 1 ,$$

d'où, par la formule (2)

$$b_n = 2(n+1) + t_{n+1} - 2n - t_n - 1 = t_{n+1} - t_n + 1 \equiv s_n + 1 \quad \text{mod. } 3 . \qquad \square$$

d) La méthode de Thue. - A. Thue [20] considère le morphisme $\delta : Y^* \to X^*$ défini par

$$\delta(0) = 0 \ , \quad \delta(1) = 01 \ , \quad \delta(2) = 011 \ .$$

Comme $\delta(Y)$ est un code (suffixe), le mot $\underline{\underline{t}}$ se factorise de manière unique en mots de $\delta(Y)$. Par conséquent, $\delta^{-1}(\underline{\underline{t}})$ est un mot infini sur Y qui est bien défini. Il est clair que l'on a $\delta^{-1}(\underline{\underline{t}}) = \underline{\underline{b}}$.

e) La méthode d'Istrail. - La construction d'Istrail [12] est intéressante parce qu'elle est indépendante de la suite $\underline{\underline{t}}$. Par sa simplicité, elle fournit aussi la preuve la plus courte de l'existence d'un mot infini sans carré.

On considère le morphisme $\sigma : Y^* \to Y^*$ défini par

$$\sigma(0) = 12 \ , \quad \sigma(1) = 102 \ , \quad \sigma(2) = 0 \ .$$

PROPOSITION 6. - On a :

$$\sigma^{\omega}(1) = \underline{\underline{s}} \ .$$

Preuve. - En utilisant les morphismes α et β définis en b), on va montrer que

$$\sigma^n(1) = \beta(\alpha^{n-1}(132)), \quad \sigma^n(0) = \beta(\alpha^{n-1}(12)), \quad \sigma^n(2) = \beta(\alpha^{n-1}(0))$$

d'où il résulte que $\sigma^n(1)$ est facteur gauche de $\underline{\underline{s}}$ et par conséquent que $\sigma^{\omega}(1) = \underline{\underline{s}}$. Les formules sont évidentes pour $n = 1$. Ensuite, on a

$$\begin{aligned}\sigma^{n+1}(1) &= \sigma^n(102) = \beta(\alpha^{n-1}(132))\ \beta(\alpha^{n-1}(12))\ \beta(\alpha^{n-1}(0)) \\ &= \beta(\alpha^{n-1}(132120)) = \beta(\alpha^n(132)) .\end{aligned}$$

On procède de même pour les autres cas. □

3. - Tag-systèmes et automates finis

Nous présentons maintenant un cadre suffisamment général pour donner une définition unifiée de la plupart des mots infinis sans carré explicitement connus. Il s'agit de ce que A. Cobham appelle tag-systèmes [7]. Le terme de tag-système est employé dans une acception plus large par Minsky [15]. Appelons k-uniforme un morphisme $\psi : X^* \to Y^*$ tel que $|\psi(x)| = k$ pour toute lettre x , et uniforme un morphisme qui est k-uniforme pour un entier k .

a) Définition. - Un tag-système T est composé de deux alphabets A, B, d'une lettre $y \in B$, et de deux morphismes $\theta : B^* \to B^*$ et $\gamma : B^* \to A^*$, le tout soumis aux conditions :

$$\theta(y) \in yB^+ \; ; \; \theta(z) \neq \varepsilon \quad (z \in B) \; ; \; \gamma(B) = A \; . \tag{3}$$

Si le morphisme θ est k-uniforme, le système T est lui-même dit uniforme de module k. On a alors nécessairement $k \geq 2$. Les deux premières des conditions (3) assurent l'existence du mot infini

$$\theta^{\omega}(y)$$

appelé la séquence interne de T ; la dernière des conditions (3) montre que

$$\gamma(\theta^{\omega}(y))$$

est un mot infini, appelé la séquence produite par T.

b) Lien avec les automates finis. - Considérons un tag-système uniforme T de module k. Suivant Cobham [7] on lui associe un automate fini M de la manière que voici : l'alphabet d'entrée de M est $\underline{k} = \{0, 1, \dots, k-1\}$, l'ensemble d'états est B, l'état initial y. La fonction de transition est

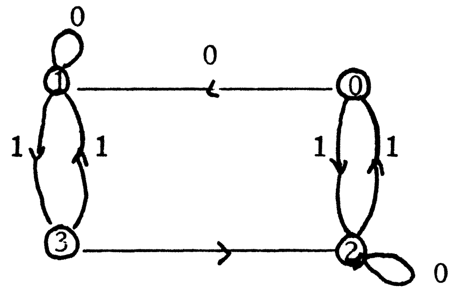
$$z.i = z_i \quad \text{si} \quad \theta(z) = z_0 z_1 \dots z_{k-1} \quad (i \in \underline{k}, \; z_i \in B).$$

On vérifie alors facilement que pour un mot $w \in k^*$, on obtient

$$y.w = y_n \; ,$$

où n est l'entier représenté par w en base k et où y_n est la n-ième lettre de la séquence interne $\theta^{\omega}(y)$. Ainsi, le mot infini $\theta^{\omega}(y)$ décrit la séquence des états de M atteints lorsque les mots $w \in \underline{k}^*$ décrivent la suite des entiers. L'application γ est une fonction de sortie qui à chaque état de M associe un symbole. Lorsque $A = \{0, 1\}$, on peut interpréter $\gamma^{-1}(1) \subset B$ comme l'ensemble des états terminaux de M, et $\{n \mid y_n = 1\}$ est un ensemble k-reconnaissable de nombres au sens de Eilenberg [9]. Notons qu'il existe un lien très remarquable entre ces constructions et les séries algébriques sur un corps fini (Christol; Kamae, Mendes France, Rauzy [5]).

c) <u>Deux exemples</u>.-Du couple de morphismes α , β défini en 2.b résulte un tag-système dont voici le graphe de l'automate.



L'état initial est 1 . Ici $k = 2$, donc l'ensemble $\{n \mid \underline{\underline{s}}(n) = 1\}$ est un ensemble 2-reconnaissable de nombres.

Aršon [1] a donné en 1937 un mot infini sur un alphabet à trois lettres distinct du mot $\underline{\underline{s}}$ décrit plus haut. Ce mot est produit par le tag-système que voici (cf. [2])

$$B = \{0, 1, 2, \overline{0}, \overline{1}, \overline{2}\}, \quad A = \{0, 1, 2\}, \quad y = 0$$

$$\theta(0) = 0\overline{1}2 \quad , \quad \theta(0) = \overline{2}1\overline{0}$$

$$\theta(1) = 1\overline{2}0 \quad , \quad \theta(1) = \overline{0}2\overline{1}$$

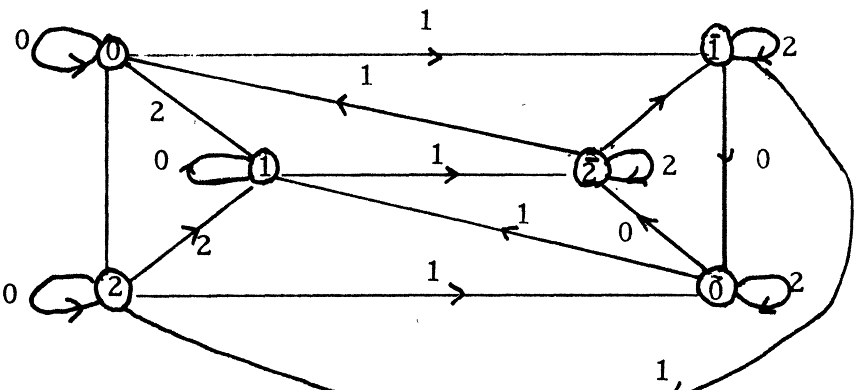
$$\theta(2) = 2\overline{0}1 \quad , \quad \theta(2) = \overline{1}0\overline{2}$$

$$\gamma(x) = \gamma(\overline{x}) = x \quad (x = 0, 1, 2).$$

La suite d'Aršon est

$$\underline{\underline{a}} = \gamma(\theta^{\omega}(0)) = 0120212012102010212012101201 02\ldots .$$

L'automate qui lui correspond est le suivant :



d) Deux conséquences. - Le lien étroit avec les tag-systèmes permet de prouver que certaines suites ne peuvent être engendrées au moyen d'un mécanisme moins puissant. L'outil de base, pour ce faire, est le théorème suivant :

THEOREME DE COBHAM [6,7]. - Si $\underline{\underline{x}}$ est le mot infini engendré par deux tag-systèmes T et T' uniformes de module k et ℓ respectivement, et si $\underline{\underline{x}}$ n'est pas ultimement périodique, alors il existe deux entiers $n, m > 0$ tels que $k^n = \ell^m$.

COROLLAIRE 7. - Il n'existe pas de morphisme uniforme $\psi : \{0,1,2\}^* \to \{0,1,2\}^*$ tel que le mot infini $\underline{\underline{s}}$ soit égal à $\psi^\omega(1)$.

Notons que $\underline{\underline{s}} = \sigma^\omega(1)$, où σ est le morphisme non uniforme d'Istrail. Le corollaire dit donc qu'il n'est pas possible de remplacer σ par un morphisme uniforme.

Preuve. - Supposons que $\underline{\underline{s}} = \psi^\omega(1)$, où ψ est un morphisme k-uniforme. En vertu du corollaire 4, $\underline{\underline{s}}$ est engendré par un tag-système de module 2. Comme $\underline{\underline{s}}$ n'est pas ultimement périodique puisque sans carré, l'entier k est, par le théorème de Cobham, une puissance de 2, soit $k = 2^q$.

Comme $\underline{\underline{s}} = \psi(\underline{\underline{s}})$, le mot $\psi(102120)$ est le facteur gauche de longueur $6k$ de $\underline{\underline{s}}$. En particulier, les lettres $\underline{\underline{s}}(k)$ et $\underline{\underline{s}}(5k)$ sont égales, puisqu'elles sont égales à la première lettre de $\psi(0)$.

D'autre part, et avec les notations 2.b), comme 132120 est facteur gauche de $t' = \alpha^\omega(1)$, le mot $\alpha^q(132120) = \alpha^q(1)\,\alpha^q(3)\,\alpha^q(212)\,\alpha^q(0)$ est le facteur gauche de longueur 6.2^q de t'. Par définition de α, la première lettre de $\alpha^q(3)$ (resp. $\alpha^q(0)$) est "2" (resp. "1"), donc $\underline{\underline{t}}'(k) = 2$, $\underline{\underline{t}}'(5k) = 1$, d'où $\underline{\underline{s}}(k) = 2$ et $\underline{\underline{s}}(5k) = 1$. Ceci donne la contradiction. □

Cette observation montre donc que la suite sans carré $\underline{\underline{s}}$ est plus "compliquée" que certaines autres suites sans carré, comme par exemple celles de Zech [21], Leech [14], ou Dejean [8] qui elles s'obtiennent par itération d'un morphisme uniforme. Néanmoins, il n'est pas encore clair dans quelle mesure ceci se reflète dans certaines propriétés des suites. Ainsi, la suite de Dejean possède des propriétés remarquables.

La suite d'Aršon s'insère dans ce cadre en vue de la

PROPOSITION 8 [2]. - Il n'existe pas de morphisme $\psi : \{0,1,2\}^* \to \{0,1,2\}^*$ tel que la suite d'Aršon $\underline{\underline{a}}$ soit égale à $\psi^{\omega}(0)$.

Ainsi, et contrairement à ce qui se passe pour $\underline{\underline{s}}$, la suite d'Aršon $\underline{\underline{a}}$ ne peut même pas être engendrée par un morphisme non uniforme.

4. - L'ensemble des mots infinis sans carré n'est pas dénombrable

Comme en section 2, on pose $Y = \{0,1,2\}$. Nous prouvons, en adaptant un argument de Kakutani (cité dans Gottschalk, Hedlund [11]) la

PROPOSITION 9. - L'ensemble des mots infinis sans carré sur Y n'est pas dénombrable.

L'ensemble des tag-systèmes étant par contre dénombrable, ceci montre que tout mot sans carré infini sur un alphabet à trois lettres ne peut être engendré par un tag-système.

Pour prouver la proposition, nous examinons d'un peu plus près la structure de la suite $\underline{\underline{s}}$ que l'on peut écrire sous la forme $\underline{\underline{s}} = \beta(\alpha^{\omega}(1))$ avec les notations de 2.b). Posons, pour $n>0$

$$a_n = \beta\alpha^n(0) \ , \quad b_n = \beta\alpha^n(1) \ , \quad c_n = \beta\alpha^n(2) \ , \quad d_n = \beta\alpha^n(3) \ .$$

Alors on a, par la définition même de α,

$$a_o = 0 \ , \quad b_o = 1 \ , \quad c_o = 2 \ , \quad d_o = 0 \ ,$$

$$a_1 = 12 \ , \quad b_1 = 10 \ , \quad c_1 = 20 \ , \quad d_1 = 21 \ ,$$

et

$$a_{n+1} = b_n c_n \ , \quad b_{n+1} = b_n d_n \ , \quad c_{n+1} = c_n a_n \ , \quad d_{n+1} = c_n b_n \quad n \geq 0 \ .$$

Un calcul immédiat montre que

$$\underline{\underline{s}}^{[2^n]} = b_n = 1 d_o d_1 d_2 \dots d_{n-1} \quad (n \geq 1) \ .$$

Pour simplifier l'exposition, nous écrirons $u \leq v$ lorsque u est facteur droit de v, i.e. lorsque $hu = v$ pour un mot h. Cette relation est un ordre, et $u \leq v \Rightarrow uw \leq vw$ pour tout mot w.

LEMME 10. - *On a* $b_n \leq b_{n+2}$; *de plus, si* $n \leq m$ *et* $m-n$ *est pair, alors* $b_n d_m \leq b_{m+1}$.

Preuve. - On a $b_{n+2} = b_{n+1} d_{n+1} = b_{n+1} c_n b_n$, montrant la première relation. La deuxième assertion est vraie par définition lorsque $m = n$; si $m > n$, alors on a $b_n \leq b_m$ en vertu de ce qui précède, d'où $b_n d_m \leq b_m d_m = b_{m+1}$. □

On considère maintenant l'ensemble F des applications strictement croissantes $f : \mathbb{N} \to \mathbb{N}$ vérifiant :

$$f(0) = 0$$
$$f(n+1) - f(n) \text{ impair } (n \geq 0).$$

On note $\underline{\underline{s}}(f)$ le mot infini suivant :

$$\underline{\underline{s}}(f) = 1\, d_{f(0)}\, d_{f(1)} \cdots d_{f(n)} \cdots .$$

Ainsi, si f est l'application identique, alors $\underline{\underline{s}}(f) = \underline{\underline{s}}$.

PROPOSITION 11. - *Pour tout* $f \in F$, *la suite* $\underline{\underline{s}}(f)$ *est sans carré. De plus, l'application* $f \mapsto \underline{\underline{s}}(f)$ *est injective.*

Comme F n'est pas dénombrable, la proposition 9 en découle aussitôt.

Preuve. - Soit $f \in F$. Nous montrons que pour tout $n \geq 0$,

$$u_n = 1\, d_{f(0)}\, d_{f(1)} \cdots d_{f(n)} \quad \text{est facteur droit de } b_{1+f(n)} . \tag{4}$$

Comme $b_{1+f(n)} \leq d_{2+f(n)}$, ceci montre que u_n est facteur de $\underline{\underline{s}}$, donc un mot sans carré. Par conséquent, $\underline{\underline{s}}(f)$ est aussi sans carré.

La formule (4) est vraie pour $n = 0$ car la condition $f(0) = 0$ implique $u_o = b_1$. Si $n \geq 1$, on obtient :

$$u_n = 1\, d_{f(0)}\, d_{f(1)} \cdots d_{f(n-1)}\, d_{f(n)} \leq b_{1+f(n-1)}\, d_{f(n)} \leq b_{1+f(n)} ,$$

en appliquant l'hypothèse de récurrence et le lemme 10.

Pour prouver l'injectivité, supposons que $\underline{\underline{s}}(f) = \underline{\underline{s}}(g)$ pour $f, g \in F$ et $f \neq g$; soit n le plus petit entier tel que $f(n) \neq g(n)$. Posons $p = f(n)$, $q = g(n)$; on peut supposer $p < q$, donc d_p est facteur gauche de d_q. D'autre part

$$d_q = c_p a_p a_{p+1} \dots a_{q-2} b_{q-1} ,$$

et comme $|d_p| = |c_p|$, on a $d_p = c_p$. Il en résulte, puisque

$$c_p = c_{p-1} a_{p-1} = c_{p-1} b_{p-2} c_{p-2} ,$$

$$d_p = c_{p-1} b_{p-1} = c_{p-1} b_{p-2} d_{p-2} ,$$

que $c_{p-2} = d_{p-2}$, et de proche en proche on aboutit à $c_o = d_o$ ou $c_1 = d_1$ ce qui n'est pas. □

-:-:-:-

REFERENCES

[1] S. ARSON, Démonstration de l'existence de suites asymétriques infinies, Mat. Sb. 44 (1937), p. 769-777.

[2] J. BERSTEL, Mots sans carré et morphismes itérés, rapport n° 78-42, Institut de Programmation, Paris, 1978.

[3] L. BOASSON, M. NIVAT, Adherences of languages, rapport n° 79-13, Laboratoire d'Informatique théorique et de programmation, Paris VII, 1979.

[4] C. BRAUNHOLTZ, An infinite sequence of three symbols with no adjacent repeats, American Math. Monthly 70 (1963), p. 675-676.

[5] G. CHRISTOL, T. KAMAE, M. MENDÈS FRANCE, G. RAUZY, Suites algébriques, automates et substitutions, à paraître.

[6] A. COBHAM, On the base-dependence of sets of numbers recognizable by finite automata, Math. Systems Theory 3 (1969), p. 186-192.

[7] A. COBHAM, Uniform tag sequences, Math. Systems Theory 6 (1972), p. 164-192.

[8] F. DEJEAN, Sur un théorème de Thue, J. Combinatorial Theory, Series A, 13 (1972), p. 90-99.

[9] S. EILENBERG, "Automata, Languages and Machines", vol. A, Academic Press 1974.

[10] P. FRANCHI ZANNETTACHI, B. VAUQUELIN, Automates à file, Theoret. Comput. Sci., à paraître.

[11] W. GOTTSCHALK, G. HEDLUND, Topological dynamics, Amer. Math. Soc. Colloq. Publ. vol. 36, 1955.

[12] S. ISTRAIL, On irreductible languages and nonrational numbers, Bull. Soc. Math. Roumanie 21 (1977), p. 301-308.

[13] J. JUSTIN, Généralisation du théorème de van der Waerden sur les semi-groupes répétitifs, J. Comb. Theory, Series A 12 (1972), p. 357-367.

[14] J. LEECH, Note 2726, A problem on strings of beads, Math. Gazette 41 (1957), p. 277-278.

[15] M.L. MINSKY, Computation, finite and infinite machines, Prentice Hall, 1967.

[16] M. MORSE, G. HEDLUND, Unending chess, symbolic dynamics and a problem in semi-groups, Duke Math. J. 11 (1944), p. 1-7.

[17] P.S. NOVIKOV, S.I. ADJAN, Infinite periodic groups, Math. USSR Izv. 2 (1968), pp. 209-236, 241-479, 665-685.

[18] P.A. PLEASANTS, Non-repetitive sequences, Proc. Cambridge Phi. Soc. 68 (1970), p. 267-274.

[19] A. THUE, Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat.-Nat. Kl., Christiania(1906), n° 7, p.1-22.

[20] A. THUE, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Vidensk. Skr. I. Mat.-Naturv. Kl. (1912), n°1, p. 1-67.

[21] T. ZECH, Wiederholungsfreie Folgen, Z. Angew. Math. Mech. 38 (1958), p. 206-209.

(texte reçu le 5 juin 1979)

-:-:-:-

Jean BERSTEL
Université de Paris VI
Institut de Programmation
Laboratoire d'Informatique Théorique
et Programmation - U.E.R. 50
75230 PARIS CEDEX 05