BULLETIN DE LA S. M. F.

BRIGITTE MOSSÉ

Reconnaissabilité des substitutions et complexité des suites automatiques

Bulletin de la S. M. F., tome 124, n° 2 (1996), p. 329-346

<http://www.numdam.org/item?id=BSMF_1996__124_2_329_0>

Bull. Soc. math. France,
124, 1996, p. 329–346.

RECONNAISSABILITÉ DES SUBSTITUTIONS ET COMPLEXITÉ DES SUITES AUTOMATIQUES

PAR

BRIGITTE MOSSÉ (*)

RÉSUMÉ. — Nous rappelons tout d'abord en les complétant quelques notions et un théorème de reconnaissabilité concernant les mots infinis points fixes de substitutions primitives. Dans le cas où la substitution considérée est de longueur constante q, nous étudions la fonction de complexité $p(n)$ du point fixe u, c'est-à-dire le nombre de facteurs de longueur n de u. Nous donnons une méthode pour calculer $p(n)$ par des formules de récurrence linéaire et nous montrons que la suite $(p(n+1)-p(n))_{n\in\mathbb{N}}$ est q-automatique.

ABSTRACT. — We first recall some notions and one theorem of recognizability about infinite words fixed points of primitive substitutions. When the length of the substitution is constant equal to q, we study the complexity function $p(n)$ of the fixed point u, which is the number of factors of u of length n. We give a method to compute $p(n)$ with linear recurrence formulas and we prove that the sequence $(p(n+1)-p(n))_{n\in\mathbb{N}}$ is q-automatic.

Introduction et notations

Soit $\mathfrak{a}$ un ensemble fini de cardinal $g \geq 2$, appelé *alphabet*; les éléments $a_1, \ldots, a_g$ de $\mathfrak{a}$ sont appelés lettres; soit $\mathfrak{a}^*$ l'ensemble des mots non vides de longueur finie sur $\mathfrak{a}$.

Si $v \in \mathfrak{a}^*$ et $x \in \mathfrak{a}^{\mathbb{N}}$, nous notons $|v|$ la longueur de v et $x[i,j]$ le mot $x_i x_{i+1} \cdots x_j$ (on suppose $j \geq i$).

Nous désignerons par σ une *substitution* sur $\mathfrak{a}$, c'est-à-dire une application de $\mathfrak{a}$ dans $\mathfrak{a}^*$ prolongée par concaténation à $\mathfrak{a}^*$ et à $\mathfrak{a}^{\mathbb{N}}$; nous supposerons que σ admet un point fixe u, qui lui sera désormais implicitement associé, et nous noterons L_u l'ensemble des facteurs du mot infini u.

La substitution σ est dite *primitive* s'il existe un entier naturel n tel que, pour tout couple (a,b) de lettres, b apparaît dans le mot $\sigma^n(a)$. Ceci

(*) Texte reçu le 13 septembre 1993, révisé le 9 février 1995.
B. MOSSÉ, Laboratoire de Mathématiques Discrètes UPR 9016 CNRS 163, Avenue de Luminy 13288 Marseille CEDEX 9.

BULLETIN DE LA SOCIÉTÉ MATHÉMATIQUE DE FRANCE 0037–9484/1996/329/$ 5.00

revient à dire que la matrice associée à σ, définie par $M = (m_{ij})_{1\leq i,j\leq g}$, où m_{ij} est le nombre d'apparitions de a_i dans $\sigma(a_j)$, admet une puissance strictement positive.

Soit enfin T le *décalage* défini sur $\mathfrak{a}^{\mathbb{N}}$ par $T\big((x_n)_{n\in\mathbb{N}}\big) = (x_{n+1})_{n\in\mathbb{N}}$.

Nous notons K_u la fermeture de l'orbite de u sous l'action de T pour la topologie produit sur $\mathfrak{a}^{\mathbb{N}}$.

Rappelons que le système dynamique (K_u, T) est strictement ergodique. En particulier, ce système est minimal : tout élément de $\mathfrak{a}^*$ apparaissant dans u y apparaît une infinité de fois, avec des lacunes bornées (*cf.* [4], [9]).

Dans tout ce qui suit nous ne considérons plus que des substitutions primitives.

Une famille particulière de substitutions retiendra notre attention par la suite :

• Une substitution σ est dite de *longueur constante* lorsqu'il existe un entier q tel que $|\sigma(a)| = q$ pour toute lettre a de $\mathfrak{a}$.

Une notion plus large que celle de point fixe d'une substitution de longueur constante est celle de suite q-automatique.

• Une suite v est dite *q-automatique* si la famille $((v_{a+nq^r})_n; a, r \in \mathbb{N})$ est finie.

Cette propriété équivaut entre autres à l'existence d'une substitution de longueur constante sur un alphabet $\mathfrak{a}$ admettant un point fixe u , d'un alphabet $\mathfrak{b}$ et d'une application f de $\mathfrak{a}$ vers $\mathfrak{b}$ tels que $v = f \circ u$.

Si $v(n)$ est à valeurs dans un ensemble fini d'entiers, elle équivaut également à l'existence d'une suite

$$\big(V(n)\big) = \big(v_0(n), \ldots, v_{r-1}(n)\big)$$

de vecteurs à valeurs entières, dont $v(n)$ est la première coordonnée, et pour laquelle il existe q matrices $r \times r$ à coefficients entiers $A_0, \ldots, A_{q-1}$ vérifiant :

$$V(nq + i) = A_i V(n)$$

pour tout i inférieur à $(q - 1)$ (*cf.* [2], [3]).

Revenons à la notion générale de substitution. Une manière naturelle de découper un point fixe u d'une substitution σ consiste à faire intervenir les ensembles :

$$E_k = \big\{0\big\} \cup \big\{|\sigma^k(u[0, p-1])|\,;\ p > 0\big\}.$$

Considérons par exemple le mot de Morse m, point fixe de la substitution σ_m définie sur $\mathfrak{a} = \{a, b\}$ par $\sigma_m(a) = ab$ et $\sigma_m(b) = ba$; nous avons ici $E_k = 2^k\mathbb{N}$:

$$\begin{aligned} m &= |ab|ba|ba|ab|ba|ab|ab|ba|ba|ab|ab|ba|\ldots \quad \text{(niveau 1 de découpage)} \\ &= |abba|baab|baab|abba|baab|abba|abba|\ldots \quad \text{(niveau 2 de découpage)} \\ &= |abbabaab|baababba|baababba|abba\ldots \quad \text{(niveau 3 de découpage)} \\ &\textit{etc.} \end{aligned}$$

Choisissons maintenant pour u le point fixe de la substitution σ définie sur $\mathfrak{a} = \{a, b, c\}$ par $\sigma(a) = abc$, $\sigma(b) = ca$ et $\sigma(c) = bb$:

$$\begin{aligned} u &= |abc|ca|bb|bb|abc|ca|ca|ca|ca|abc|ca|bb|bb|abc|\ldots \quad \text{(niveau 1)} \\ &= |abccabb|bbabc|caca|caca|abccabb|bbabc|\ldots \quad \text{(niveau 2)} \\ &= |abccabbbbabccaca|cacaabccabb|bbabc\ldots \quad \text{(niveau 3)} \\ &\textit{etc.} \end{aligned}$$

Entre deux barres du découpage de niveau k, nous trouvons un facteur de u de la forme $\sigma^k(\alpha)$ où α est un élément de l'alphabet $\mathfrak{a}$. Nous dirons que ce mot *provient* de α.

Soit maintenant B un facteur de u; on peut se demander comment B est découpé de la sorte aux rangs i où il apparaît dans u, et de quels mots il provient selon la valeur de i (étant bien entendu que si i ou $i + |B|$ n'est pas un élément de E^k, B commence ou finit par un suffixe ou un préfixe strict d'un certain $\sigma^k(a)$).

Le premier à avoir étudié ces problèmes de découpage est J.-C. Martin [7]; il qualifie de *déterminée à l'ordre k* une substitution σ pour laquelle dès que B est assez long, le découpage de niveau k et les lettres dont B provient par σ^k ne dépendent pas des rangs où B apparaît.

C'est la situation la plus simple qu'on puisse espérer dès que u n'est pas ultimement périodique (ce qui équivaut à périodique quand σ est primitive).

On peut se convaincre très vite par exemple du fait que la substitution de Morse vérifie cette propriété.

Malheureusement la seule primitivité ne suffit pas, comme nous le verrons ultérieurement, à assurer la détermination à tous les ordres, et B. Host et M. Queffélec ont introduit la notion moins exigeante de *reconnaissabilité* (voir [6], [9]), dont l'hypothèse fut rencontrée systématiquement par la suite.

Diverses notions de découpage et de reconnaissabilité ont été introduites, dont nous donnons un aperçu dans la première partie de cet article. Nous énonçons un théorème nouveau de reconnaissabilité, valable pour n'importe quelle substitution primitive à point fixe non périodique, qui complète les résultats précédemment obtenus dans [8].

Dans la deuxième partie, nous étudions la fonction de complexité $p(n)$ associée à un point fixe d'une substitution de longueur constante. Rappelons de quoi il s'agit. Soit σ une substitution de longueur constante. Si u est point fixe de σ, la *fonction de complexité* $p(n)$ *associée à* u est définie par :

$$p(n) = \operatorname{card}\{\omega \in L_u\,;\ |\omega| = n\}.$$

On sait qu'il suffit que σ soit primitive pour que $p(n)$ soit sous-affine. Ainsi, les suites qui sont points fixes de telles substitutions ont la propriété d'être de basse complexité : nous renvoyons le lecteur à [1] pour un bilan sur ce sujet.

En appliquant les théorèmes énoncés dans la première partie, nous montrons comment calculer $p(n)$. Plus précisément, nous exhibons des formules de récurrence linéaire, utilisant un nombre fini de matrices explicites permettant le calcul de $p(n)$ pour tout entier n donné.

Enfin, nous étudions la suite $(p(n+1) - p(n))_{n\in\mathbb{N}}$, sur l'intérêt de laquelle Gérard Rauzy a attiré notre attention. Cette suite est 2-automatique dans le cas de la suite de Morse (*cf.* [1]), et le cas plus général de l'alphabet à deux lettres a été étudié par T. Tapsoba (*cf.* [10]).

Nous prouvons ici que cette suite est q-automatique. Nous donnons également des précisions sur le cas où $p(n)$ est ultimement affine dans quelques exemples.

Première partie : Diverses notions de reconnaissabilité pour les substitutions

1. Définitions.

Dans [8], nous avons défini de la façon suivante la notion de découpage :

- Soient B un élément de L_u et i et j deux entiers naturels tels que

$$B = u\big[i, i + |B| - 1\big] = u\big[j, j + |B| - 1\big]\,;$$

on dit que B *admet le même* 1*-découpage au rang* i *et au rang* j si $E_1 \cap \{i, \ldots, i + |B| - 1\}$ et $E_1 \cap \{j, \ldots, j + |B| - 1\}$ sont images l'un de l'autre par la translation de $(j - i)$.

- Soient $x_0, x_1, \ldots, x_{r+1}$ des éléments de $\mathfrak{a}$, S un suffixe de $\sigma(x_0)$ et P

un préfixe de $\sigma(x_{r+1})$; on dit que $[S, \sigma(x_1), \ldots, \sigma(x_r), P]$ est un 1-*découpage (naturel) du mot*

$$B = u[i, i + |B| - 1] = S\sigma(x_1) \cdots \sigma(x_r)P$$

au rang i s'il existe un entier naturel j tel que

$$x_0 x_1 \cdots x_{r+1} = u[j, j + r + 1]$$

et le bloc B admet le même 1-découpage au rang i et au rang

$$\left|\sigma(u[0, j])\right| - |S|.$$

Par exemple, dans le mot de Morse, le mot *bab* n'a pas le même 1-découpage au rang 2 (où il admet le 1-découpage naturel $[ba, b]$) et au rang 11 (où il admet le 1-découpage naturel $[b, ab]$).

Nous dirons de plus que $B = u[i, i + |B| - 1]$ *provient du mot* $x_0 x_1 \ldots x_{r+1}$ *(au rang* i *par* σ*)* s'il existe un entier naturel j' tel que

$$x_0 x_1 \ldots x_{r+1} = u[j', j' + r + 1] \quad \text{et} \quad i = \left|\sigma\big(u[0, j']\big)\right| - |S|.$$

DÉFINITION 1 (voir [7]). — On dit que σ est *déterminée à l'ordre* 1 s'il existe un entier $L > 0$ tel que si B est un élément de L_u de longueur $\geq L$ pour lequel il existe deux éléments $x_0 x_1 \cdots x_{r+1}$ et $x'_0 x'_1 \cdots x'_{r'+1}$ de L_u vérifiant :

$$B = S\sigma(x_1) \cdots \sigma(x_r)P = S'\sigma(x'_1) \cdots \sigma(x'_{r'})P',$$

où S est un suffixe non vide de $\sigma(x_0)$, P un préfixe non vide de $\sigma(x_{r+1})$, S' un suffixe non vide de $\sigma(x'_0)$ et P' un préfixe non vide de $\sigma(x'_{r'+1})$, alors $r = r'$ et $x_i = x'_i$ pour $0 \leq i \leq r + 1$, $S = S'$, et $P = P'$.

Autrement dit, si un mot B est suffisamment long, le 1-découpage et les lettres dont B provient ne dépendent pas de ses rangs d'apparition .

DÉFINITION 2 (voir [6], [9]). — On dit que σ est *reconnaissable* (ou *reconnaissable à droite*) s'il existe un entier $L > 0$ tel que si $u[i, i+L-1] = u[j, j + L - 1]$ et $i \in E_1$, alors $j \in E_1$.

Autrement dit, si un mot B est suffisamment long, le 1-découpage de B ne dépend pas de ses rangs d'apparition, sauf éventuellement sur son suffixe de longueur L. De plus cette propriété équivaut à dire que l'image $\sigma(K_u)$ par la substitution σ de la fermeture de l'orbite de u sous l'action du décalage est ouverte (*cf.* [6]).

DÉFINITION 3 (voir [8]). — On dit que σ est *bilatéralement reconnaissable* s'il existe un entier $L > 0$ tel que si $u[i-L, i+L] = u[j-L, j+L]$ et $i \in E_1$, alors $j \in E_1$.

Autrement dit si un mot B est suffisamment long, le 1-découpage de B ne dépend pas de ses rangs d'apparition, sauf éventuellement sur son préfixe de longueur L et son suffixe de longueur L.

2. Théorèmes de reconnaissabilité.

La reconnaissabilité bilatère est la « bonne » notion de reconnaissabilité, ainsi que le montre le théorème suivant (*cf.* [8]) :

THÉORÈME 1. — *Soit σ une substitution primitive admettant un point fixe non périodique u ; la substitution σ est bilatéralement reconnaissable.*

Il n'en est pas de même pour la reconnaissabilité qui peut ne pas avoir lieu, ou ne pas se conserver par itération (*cf.* [5], [8]).

Le théorème 1 n'aborde pas la question de l'éventuelle unicité des mots dont proviennent les facteurs B de u considérés. Il y a unicité de ces mots dans le cas où pour deux lettres distinctes α et β, $\sigma(\alpha)$ et $\sigma(\beta)$ commencent et finissent par deux lettres distinctes. Il y a seulement un problème éventuel au début et à la fin de B, lorsque σ est injective sur $\mathfrak{a}$.

Le théorème suivant montre cependant qu'on peut toujours « désubstituer » les mots considérés, sauf peut-être leurs bords (un préfixe et un suffixe de longueur L ne dependant que de σ).

THÉORÈME 2. — *Soit σ une substitution primitive admettant un point fixe non périodique u. Il existe un entier $L > 0$ tel que si*

$$u[i-L, j+L] = u[i'-L, j'+L],$$

alors $u[i,j]$ et $u[i',j']$ ont le même 1-découpage et proviennent du même mot aux rangs i et i'.

Démonstration. — Il existe en effet un entier $p > 0$ tel que pour toutes lettres a et b, si $\sigma^{p-1}(a) \neq \sigma^{p-1}(b)$, alors $\sigma^k(a) \neq \sigma^k(b)$ pour tout entier $k \geq 0$. Le mot dont provient un facteur $\sigma^p(a)$ de u apparaissant entre deux rangs de E_p est donc entièrement déterminé. Soit L' un entier convenant pour la propriété de reconnaissabilité bilatère de σ^p. Il suffit de poser $L = L' + \max\{|\sigma^p(a)| \,;\, a \in \mathfrak{a}\}$ pour conclure. □

Deuxième partie : Application à la complexité en longueur constante

Nous allons maintenant appliquer les résultats précédents à l'étude de la fonction de complexité $p(n)$ du point fixe d'une substitution primitive σ de longueur constante, égale à un entier $q \geq 2$. Il sera utile d'introduire les quantités et notations suivantes :

- le nombre $p(n, a)$ de facteurs de longueur n de u n'apparaissant qu'à des rangs congrus à a modulo q ;
- le nombre $p(n; gE, dF)$ de facteurs de longueur n de u prolongeables simultanément à gauche par au moins un élément de E et à droite par au moins un élément de F, E et F désignant deux parties de L_u ;
- les nombres

$$p(n, a; gE, dF), \quad p(n; gE), \quad p(n; dF), \quad p(n, a; gE), \quad p(n, a; dF)$$

dont les définitions sont calquées sur les précédentes.

Étant donnés deux nombres entiers a et b, $b \neq 0$, nous noterons $\mathrm{div}_b(a)$ et $\mathrm{mod}_b(a)$ le quotient et le reste de la division euclidienne de a par b.

La substitution σ étant reconnaissable, pour n assez grand, tout mot de longueur n apparaît à des rangs congrus modulo q et on a l'égalité :

$$p(n) = \sum_{0 \leq a < q} p(n, a).$$

Nous traiterons tout d'abord le cas où σ est déterminée à l'ordre un, qui par sa simplicité éclairera le lecteur sur la méthode utilisée.

THÉORÈME 3. — *Soient u un point fixe d'une substitution primitive σ de longueur constante égale à q, déterminée à l'ordre un, et L un entier attaché à cette propriété.*

(i) *Pour $n \geq L$, la suite $p(n)$ est donnée par les relations :*

$$p(n) = \begin{cases} p(\mathrm{div}_q(n)) + (q-1)p(\mathrm{div}_q(n)+1) & \text{si } \mathrm{mod}_q(n) = 0, \\ (q - \mathrm{mod}_q(n) + 1)p(\mathrm{div}_q(n)+1) & \\ \qquad +(\mathrm{mod}_q(n) - 1)p(\mathrm{div}_q(n)+2) & \text{sinon.} \end{cases}$$

(ii) *Pour $n \geq L$, la suite $(p(n+1) - p(n))_{n \in \mathbb{N}}$ est une suite q-automatique donnée par les relations :*

$$p(n+1) - p(n) = \begin{cases} p(\mathrm{div}_q(n)+1) - p(\mathrm{div}_q(n)) & \text{si } \mathrm{mod}_q(n) = 0, \\ p(\mathrm{div}_q(n)+2) - p(\mathrm{div}_q(n)+1) & \text{sinon.} \end{cases}$$

En particulier cette suite est à valeurs dans un ensemble fini.

Démonstration. — Pour $n \geq L$, on a l'égalité :

$$(*) \qquad p(n) = \sum_{0 \leq a < q} p(n, a)$$

Comme σ est déterminée à l'ordre 1, nous allons évaluer le nombre de facteurs v de longueur n de u apparaissant à des rangs congrus à a modulo q en considérant les mots V dont ils proviennent. Ceci revient simplement à calculer la longueur de V, qui dépend du 1-découpage de v, donc de a et de $\mathrm{mod}_q(n)$. On obtient ainsi :

$$p(n, 0) = \begin{cases} p\big(\mathrm{div}_q(n)\big) & \text{si } \mathrm{mod}_q(n) = 0, \\ p(\mathrm{div}_q(n) + 1) & \text{sinon,} \end{cases}$$

et si $0 < a < q - 1$,

$$p(n, a) = \begin{cases} p(\mathrm{div}_q(n) + 1) & \text{si } \mathrm{mod}_q(n) \leq q - a, \\ p(\mathrm{div}_q(n) + 2) & \text{sinon.} \end{cases}$$

On déduit alors l'expression de $p(n)$, puis celle de $p(n+1) - p(n)$ de l'égalité (*).

En particulier, $p(n+1) - p(n)$ ne peut prendre qu'un nombre fini de valeurs.

Notons maintenant $d(n) = p(n+1) - p(n)$, et soit $D(n)$ le vecteur colonne dont les coordonnées sont $d(n)$ et $d(n+1)$. Pour tous les entiers naturels k et r tels que $r < q$, on a :

$$D(kq + r) = A_r D(k),$$

où A_0 est la matrice unité 2×2, et où pour $r \neq 0$ on a $A_r = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$.

On en déduit que la suite d est q-automatique comme première coordonnée de D. □

Remarque. — Supposons de plus que $p(n)$ est de la forme $an + b$ pour n assez grand. Ceci équivaut à dire que $p(n+1) - p(n)$ est constant sur l'intervalle $[\mathrm{div}_q(L), L - 1]$ en supposant par exemple que q ne divise pas L. De l'expression de $p(n)$ on déduit que pour k assez grand,

$$p(kq) = akq + b = (ak + b) + (q - 1)\big(a(k+1) + b\big),$$

d'où $a = -b$ et $p(n) = a(n-1)$ à partir d'un certain rang.

Exemples.

1) Dans le cas de la suite de Morse, p est donnée par :

$$p(0) = 1,\ p(1) = 2,\ p(2) = 4,\ p(3) = 6,\ p(4) = 10$$

et pour $n \geq 4$,

$$p(n+1) - p(n) = \begin{cases} p(k+1) - p(k) & \text{si } n = 2k, \\ p(k+2) - p(k+1) & \text{si } n = 2k+1. \end{cases}$$

2) La substitution de Rudin-Shapiro est définie par

$$\sigma(a) = ab,\ \sigma(b) = ac,\ \sigma(c) = db,\ \sigma(d) = dc.$$

On a

$$p(0) = 1,\ p(1) = 4,\ p(2) = 8,\ p(3) = 16$$

et pour $n \geq 3$,

$$p(n+1) - p(n) = \begin{cases} p(k+1) - p(k) & \text{si } n = 2k, \\ p(k+2) - p(k+1) & \text{si } n = 2k+1, \end{cases}$$

ce qui donne finalement $p(n) = 8(n-1)$ pour $n \geq 2$.

THÉORÈME 4. — *Soient u un point fixe d'une substitution primitive σ de longueur constante égale à q et $p(n)$ la fonction de complexité de la suite u. Il existe un entier naturel r, des suites $p_0, p_1, \ldots, p_{r-1}$ à valeurs dans $\mathbb{N}$, $p_0(n)$ étant précisément égal à $p(n)$, et q matrices $r \times r$ $A_0, A_1, \ldots, A_{q-1}$ à coefficients dans $\mathbb{N}$ tels que si*

$$V(n) = \begin{pmatrix} p_0(n) \\ \vdots \\ p_{r-1}(n) \end{pmatrix},$$

on a :

$$V(nq+i) = A_i V(n) \quad \textit{pour tout } i \textit{ inférieur à } (q-1).$$

Autrement dit, p est q-régulière au sens de J.-P. Allouche et J. Shallit (*cf.* [2]).

Démonstration. — Nous écartons le cas où u est périodique, qui ne présente pas de difficulté. D'après le théorème 2, il existe un entier $L > 0$ tel que le 1-découpage et la provenance des facteurs de longueur $> 2L$ sont entièrement déterminés, sauf peut-être sur leurs suffixes et leurs préfixes de longueur L. De plus, σ étant de longueur constante, les facteurs de u de longueur $> 2L$ admettent un unique 1-découpage d'un bout à l'autre.

Nous imposons d'autre part à L d'être un multiple de q pour simplifier l'expression des calculs : $L = \ell q$. Pour $n > 2L$, on a l'égalité :

$$p(n) = \sum_{0 \le a < q} p(n, a).$$

Soit donc v un facteur de u de longueur $n > 2L$ apparaissant à des rangs congrus à a modulo q ; $v = \omega_1 \omega \omega_2$, où $|\omega_1| = |\omega_2| = L$; le mot ω provient d'un mot $x_0 x_1 \cdots x_{r-1}$ uniquement déterminé dont la longueur r vaut $E((n - 2L)/q) + C$, où C est égal à 0, 1 ou 2 selon la valeur de a et de $\mathrm{mod}_q(n)$. Nous notons ainsi $C = C(\mathrm{mod}_q(n), a)$.

Découpons à nouveau le mot v en $v = \omega_1' \omega' \omega_2'$, où $|\omega_1'| = L - a$ et $\omega' = \sigma(x_0 x_1 \cdots x_{r-1})$.

Le mot ω_1' (resp. ω_2') provient d'un mot de longueur ℓ qui peut varier selon les rangs d'apparition de v ; le 1-découpage de ω_1' dépend de a et celui de ω_2' dépend de a et de $\mathrm{mod}_q(n)$. Pour des valeurs fixées de a et de n, considérons la partition de l'ensemble des facteurs de u de longueur ℓ en classes de mots dont peut provenir un même préfixe ω_1' : cette partition dépend du 1-découpage des préfixes, donc de a et nous notons $P(a)$ l'ensemble de ses classes. De façon analogue, $S(\mathrm{mod}_q(n), a)$ désigne l'ensemble des classes de mots de longueur ℓ dont peut provenir un même suffixe ω_2'.

Pour n assez grand on a l'égalité :

$$p(n, a) = \sum_{\substack{E \in P(a) \\ F \in S(\mathrm{mod}_q(n), a)}} p\Big(E\Big[\frac{n - 2L}{q}\Big] + C\big(\mathrm{mod}_q(n), a\big); gE, dF\Big).$$

Nous obtenons ainsi :

$$\begin{aligned} p(n) &= \sum_{0 \le a < q} p(n, a) \\ &= \sum_{0 \le a < q} \sum_{\substack{E \in P(a) \\ F \in S(\mathrm{mod}_q(n), a)}} p\Big(E\Big[\frac{n - 2L}{q}\Big] + C\big(\mathrm{mod}_q(n), a\big); gE, dF\Big). \end{aligned}$$

Pour chaque E et F intervenant dans l'expression précédente, on définit de même des ensembles $P(a; gE, dF)$ et $S(\mathrm{mod}_q(n), a; gE, dF)$ qui conduisent à l'expression suivante :

$$p(n; gE, dF) = \sum_{0 \le a < q} \sum_{E' \in P(a; gE, dF)} \sum_{F' \in S(\mathrm{mod}_q(n), a; gE, dF)} p\Big(E\Big[\frac{n - 2L}{q}\Big] + C\big(\mathrm{mod}_q(n), a; E, F\big); gE', dF'\Big).$$

On obtient de la même façon une expression de $p(n; gE', dF')$ pour tous les couples (E', F') rencontrés, et ainsi de suite.

Comme les longueurs des éléments des ensembles de suffixes et de préfixes obtenus restent uniformément bornées, le nombre de couples à considérer est fini. Les suites $p(n; gE, dF)$ correspondantes que nous notons $q_0 = p, \ldots, q_{s-1}$ vérifient pour tout entier i compris entre 0 et $(q-1)$ des relations de la forme :

$$(**) \qquad q_i(nq+i) = \varepsilon_{0,i} q_0(n+k_0) + \cdots + \varepsilon_{s-1,i} q_{s-1}(n+k_{s-1}),$$

où les $\varepsilon_{j,i}$ valent 0 ou 1.

Les suites q_i, parmi lesquelles figure ainsi la fonction de complexité p que nous calculons, sont donc liées par les relations de récurrence linéaires (**). Si ces relations ne se traduisent pas immédiatement par des égalités matricielles de la forme :

$$W(nq+i) = B_i W(n) \quad \text{pour tout } i \text{ inférieur à } q-1,$$

où

$$W(n) = \begin{pmatrix} q_0(n) \\ \vdots \\ q_{s-1}(n) \end{pmatrix},$$

et les B_i sont q matrices $s \times s$ à coefficients dans $\mathbb{N}$, il suffit de considérer un entier m suffisamment grand pour que les suites

$$q_0(n), q_0(n+1), \ldots, q_0(n+m), \ldots, \\ q_{s-1}(n), q_{s-1}(n+1), \ldots, q_{s-1}(n+m),$$

que nous rebaptisons $p_0, p_1, \ldots, p_{r-1}$, vérifient cette fois les conditions de l'énoncé. □

COROLLAIRE. — *Soit u un point fixe d'une substitution primitive σ de longueur constante égale à q. Si la suite $(p(n+1) - p(n))_{n\in\mathbb{N}}$ est bornée, elle est q-automatique.*

Démonstration. — Soit $V(n)$ le vecteur colonne considéré précédemment, et posons $D(n) = V(n+1) - V(n)$. Des relations

$$D(nq+i) = (A_{i+1} - A_i)V(n)$$

si $0 \leq i < q-1$ et

$$D(nq+q-1) = A_0 V(n+1) - A_{q-1} V(n),$$

on déduit que la suite $(p(n+1) - p(n))_{n\in\mathbb{N}}$ est q-régulière et bornée donc q-automatique. □

Nous avons vu précédemment que la suite $(p(n+1)-p(n))_{n\in\mathbb{N}}$ est bornée dans le cas où σ est déterminée à l'ordre un. Nous abordons maintenant la situation générale.

Supposons pour l'instant que σ est injective sur $\mathfrak{a}$, et admet un point fixe non périodique u; soit L un entier tel que tout facteur de u de longueur L admet un unique 1-découpage. Soit B un facteur de u de longueur $|B| \geq L + 2(q-1)$. Nous allons associer à B un mot B' obtenu en supprimant :

- d'une part les q-a premières lettres de B, si B commence à des rangs congrus à a modulo q avec $1 \leq a \leq q-1$;
- d'autre part les b dernières lettres de B, si B finit à des rangs congrus à b modulo q avec $1 \leq b \leq q-1$.

Ainsi $|B'| \geq L$ et B' provient d'un facteur V de u parfaitement déterminé; soit ϕ l'application définie par

$$\phi(B) = V$$

sur l'ensemble E des facteurs B de u de longueur $\geq L + 2(q-1)$, .

Étant donné un facteur B de u, nous noterons $\mathfrak{d}(B)$ (resp. $\mathfrak{g}(B)$) l'ensemble des lettres par lesquelles B est prolongeable à droite (resp. à gauche), et $D(B)$ (resp. $G(B)$) le nombre d'éléments de $\mathfrak{d}(B)$ (resp. $\mathfrak{g}(B)$).

Nous faisons tout d'abord deux remarques concernant les éléments B de E tels que $D(B) > 1$ et $D(B) = D(\phi(B))$:

1) Soit B un élément de E tel que $\mathfrak{d}(B) = \{x_1, \ldots, x_r\}$ (avec $r > 1$) et $V = \phi(B)$; si on a $D(B) = D(V) = r$, avec $\mathfrak{d}(V) = \{y_1, \ldots, y_r\}$, et si B finit à un rang congru à b modulo q, il existe un mot ω (vide éventuellement) et une permutation τ de $\{1, \ldots, r\}$ tels que :

— ω est le suffixe de longueur b de B,

— $\omega x_{\tau(i)}$ est un préfixe de $\sigma(y_i)$ pour tout $i \in \{1, \ldots, r\}$.

2) Inversement, si V est un mot tel que $\mathfrak{d}(V) = \{y_1, \ldots, y_r\}$ (avec $D(V) = r > 1$) et s'il existe un mot ω de longueur b et r lettres distinctes deux à deux $x_1, \ldots, x_r$ tels que ωx_i est un préfixe de $\sigma(y_i)$ pour tout $i \in \{1, \ldots, r\}$, alors tout élément B de E tel que $D(B) > 1$ et $\phi(B) = V$ vérifie $\mathfrak{d}(B) = \{x_1, \ldots, x_r\}$ et finit à un rang congru à b modulo q.

LEMME 1. — *Soit u un point fixe non périodique d'une substitution primitive σ de longueur constante égale à q, injective sur l'alphabet $\mathfrak{a}$. Avec les notatios précédentes, il existe un entier naturel k tel que pour tout entier $r \geq k$, si B est un facteur de u tel que $D(B) > 1$ et*

$$V_0 = B, V_1 = \phi(V_0), \ldots, V_r = \phi(V_{r-1})$$

sont tous des éléments de E, alors $D(B) = D(V_1) = \ldots = D(V_{r-k})$.

Démonstration. — Notons P_r le nombre de parties à r éléments de $\mathfrak{a}$ et posons

$$k = (g-1) \max_{j \geq 2} P_j.$$

Soient $r \geq k$ et B un facteur de u suffisamment long pour que $V_0 = B, V_1 = \phi(V_0), \ldots, V_r = \phi(V_{r-1})$ soient tous des éléments de E. On a la suite d'inégalités

$$D(V_0) \leq D(V_1) \leq \ldots \leq D(V_r).$$

Comme $r \geq k$, il existe $i \in \{0, \ldots, r\}$ et $j \in \{2, \ldots, g\}$ tels que

$$D(V_i) = \ldots = D(V_{i+P_j}) = j.$$

Deux termes au moins, V_{k_1} et V_{k_2} (avec $k_1 < k_2$), de la suite finie $V_i, \ldots, V_{i+P_j}$ vérifient donc $\mathfrak{d}(V_{k_1}) = \mathfrak{d}(V_{k_2})$. D'après les remarques précédentes, ceci entraîne que

$$D(V_0) = D(V_1) = \ldots = D(V_{k_2}) = j.$$

Soit alors k_3 le plus grand élément de $\{0, \ldots, r\}$ tel que $D(V_{k_3}) = j$. Si on avait $r - k_3 > k$, il existerait $i' \in \{k_3 + 1, \ldots, r\}$ et $j' \in \{j+1, \ldots, g\}$ tels que $D(V_{i'}) = \ldots = D(V_{i'+P_j}) = j'$, ce qui par le même argument que précédemment conduirait à

$$D(V_0) = D(V_1) = \ldots = D(V_{k_3+1}) = j'.$$

Donc $D(V_0) = D(V_1) = \ldots = D(V_{k_3})$ et $r - k_3 \leq k$. □

LEMME 2. — *Soit u un point fixe non périodique d'une substitution primitive σ de longueur constante égale à q, injective sur l'alphabet $\mathfrak{a}$. Il existe un entier naturel K_1 tel que si B est un élément de E tel que $D(B) > 1$ et de longueur $\geq K_1$, on a $D(B) = D(\phi(B))$.*

Démonstration. — Soit k un entier naturel satisfaisant les conditions du lemme 1. Soit B un facteur de u tel que $V_0 = B, V_1 = \phi(V_0), \ldots, V_{k+1} = \phi(V_k)$ soient tous des éléments de E. Il suffit pour cela que la longueur de B soit supérieure à un certain entier K_1. On a alors $D(B) = D(\phi(B))$ d'après le lemme 1. □

LEMME 3. — *Soit u un point fixe non périodique d'une substitution primitive σ de longueur constante égale à q, injective sur l'alphabet $\mathfrak{a}$. Il existe un entier naturel K_2 tel que $D(B) = D(\phi(B))$ si B est un élément de E tel que $D(B) > 1$ et de longueur $\geq K_2$, et $\mathfrak{d}(B)$ détermine $\mathfrak{d}(\phi(B))$ et le rang auquel finit B modulo q.*

Démonstration. — Soit k et K_1 deux entiers naturels comme dans les lemmes 1 et 2. Soit B un élément de E tel que

$$V_0 = B, V_1 = \phi(V_0), \ldots, V_k = \phi(V_{k-1})$$

soient tous, sauf peut-être V_k, de longueur $\geq K_1$. Il suffit pour cela que la longueur de B soit supérieure à un certain entier K_2. On a alors

$$D(V_0) = D(V_1) = \ldots = D(V_k).$$

Comme dans la démonstration du lemme 1, nous pouvons exhiber deux entiers naturels k_1 et k_2 tels que $0 \leq k_1 < k_2 \leq k$ et $\mathfrak{d}(V_{k_1}) = \mathfrak{d}(V_{k_2})$. Soit $F = \{\mathfrak{d}(V_{k_1}), \ldots, \mathfrak{d}(V_{k_2-1})\}$, $\mathfrak{d}(B) \in F$. Soit B' un autre facteur de u de longueur $\geq K_2$. Par la même construction, on obtient un ensemble $F' = \{\mathfrak{d}(V_{k'_1}), \ldots, \mathfrak{d}(V_{k'_2-1})\}$ tel que $\mathfrak{d}(B') \in F$.

D'après les remarques précédant le lemme 1, F et F' sont soit disjoints soit égaux, et si $\mathfrak{d}(B) = \mathfrak{d}(B')$, on a $F = F', \mathfrak{d}(\phi(B)) = \mathfrak{d}(\phi(B'))$ et B et B' finissent au même rang modulo q. □

Des résultats sur les prolongements à gauche s'obtiennent de manière analogue.

LEMME 4. — *Soit u point fixe d'une substitution primitive σ de longueur constante égale à q , injective sur l'alphabet $\mathfrak{a}$. La suite $(p(n+1)-p(n))_{n\in\mathbb{N}}$ est bornée.*

Démonstration. — Il suffit de considérer le cas où u est non périodique. Soit n un entier supérieur au nombre K_2 intervenant dans le lemme 3, et à son homologue à gauche. On suppose que pour tout entier $m < n$ et toute partie $\mathfrak{d}$ de l'alphabet de cardinal > 1, le nombre de facteurs de u de longueur m prolongeables à droite par les éléments de $\mathfrak{d}$ exactement est plus petit qu'une constante C. Fixons alors un ensemble $\mathfrak{d}$.

Soit B un facteur de u de longueur n tel que $\mathfrak{d}(B) = \mathfrak{d}$; d'après le lemme 3, le rang auquel finit B modulo q est parfaitement déterminé, et $\Phi(B)$ est prolongeable à droite par les éléments d'une partie $\mathfrak{d}'$ de $\mathfrak{a}$ également déterminée. Soit m la longueur de $\Phi(B)$.

Le mot B peut être prolongeable à gauche par les éléments d'une partie $\mathfrak{g}$ de $\mathfrak{a}$ de cardinal > 1, et $\Phi(B)$ est alors prolongeable à gauche par les éléments d'une partie $\mathfrak{g}'$ de $\mathfrak{a}$ de même cardinal que $\mathfrak{g}$.

Il peut aussi arriver que B soit prolongeable à gauche par une seule lettre, mais $\Phi(B)$ par les éléments d'une partie $\mathfrak{g}''$ de $\mathfrak{a}$ non réduite à un singleton. Enfin, B et $\Phi(B)$ peuvent n'admettre qu'un seul prolongement à gauche.

Remarquons que si B_1 est un facteur de u correspondant au premier cas, et B_2 au deuxième cas, les parties $\mathfrak{g}'$ et $\mathfrak{g}''$ correspondantes de $\mathfrak{a}$ sont disjointes, car les 1-découpages de B_1 et de B_2 sont les mêmes, fixés par le choix de n et de $\mathfrak{d}$. Ainsi de l'égalité

$$\begin{aligned}\operatorname{Card}\Big\{B \in L_u\,;\ |B| = n \text{ et } \mathfrak{d}(B) = \mathfrak{d}\Big\}\\ = \sum_{\mathfrak{g}} \operatorname{Card}\Big\{B \in L_u\,;\ |B| = n, \mathfrak{d}(B) = \mathfrak{d} \text{ et } \mathfrak{g}(B) = \mathfrak{g}\Big\}\end{aligned}$$

il découle que

$$\begin{aligned}\operatorname{Card}\Big\{B \in L_u\,;\ |B| = n \text{ et } \mathfrak{d}(B) = \mathfrak{d}\Big\}\\ \leq \operatorname{Card}\Big\{B \in L_u\,;\ |B| = m \text{ et } \mathfrak{d}(B) = \mathfrak{d}'\Big\} \leq C.\end{aligned}$$

On en déduit que $(p(n+1) - p(n))$ est bornée. □

LEMME 5. — *Soit u point fixe d'une substitution primitive σ de longueur constante égale à q. La suite $(p(n+1) - p(n))_{n\in\mathbb{N}}$ est bornée.*

Démonstration. — Il suffit de considérer le cas des points fixes non périodiques. Soit τ une substitution primitive sur un alphabet $\mathfrak{b}$, de point fixe v, et τ' la substitution obtenue à partir de τ en identifiant les éléments de $\mathfrak{b}$ ayant même image par τ. Soit π une application de $\mathfrak{b}$ dans une de ses parties $\mathfrak{b}'$ réalisant cette identification, que nous prolongeons à $\mathfrak{b}^*$ et à $\mathfrak{b}^{\mathbb{N}}$ par concaténation. La substitution τ' est donc définie par $\tau'(\beta) = \pi(\tau(\beta))$ pour tout $\beta \in \mathfrak{b}'$. Posons $v' = \pi(v)$; la suite v' est point fixe de τ' et on a $\tau(v') = v$.

Soit L un entier tel que tout facteur de v de longueur L admet un unique 1-découpage, et B un facteur de v de longueur $|B| \geq L + 2(q-1)$.

Nous allons associer comme précédemment à B un mot B' obtenu en supprimant :

- d'une part les q-a premières lettres de B, si B commence à des rangs congrus à a modulo q avec $1 \leq a \leq q-1$,
- d'autre part les b dernières lettres de B, si B finit à des rangs congrus à b modulo q avec $1 \leq b \leq q-1$.

Comme τ n'est pas nécessairement injective, B' peut provenir de plusieurs facteurs V de v, mais $W = \pi(V)$ est parfaitement déterminé. Supposons que la fonction de complexité p' de v' vérifie $p'(n+1) - p'(n) \leq c$.

Si W admet r prolongements à droite dans v', B en admet au plus r. Comme la différence $|W| - \operatorname{div}_q(n)$ vaut (-1) ou 0, on a $p(n+1) - p(n) \leq 2c$. Ainsi la suite $(p(n+1) - p(n))$ est bornée.

En itérant ce processus « d'injectivation » à partir de la substitution σ, on obtient en un nombre fini d'étapes une substitution injective σ', dont la fonction de complexité p' est telle que $(p'(n+1)-p'(n))$ est bornée, d'après le LEMME 4. On déduit alors de ce qui précède que $(p(n+1)-p(n))$ est bornée. □

Le corollaire du théorème 4 et le lemme 5 nous permettent d'énoncer le résultat suivant.

THÉORÈME 5. — *Soit u point fixe d'une substitution primitive σ de longueur constante égale à q. Alors la suite $(p(n+1)-p(n))_{n\in\mathbb{N}}$ est q-automatique.*

EXEMPLES.

1) Soit σ la substitution définie sur l'alphabet $\mathfrak{a}=\{a,b\}$ par

$$\sigma(a)=aab, \quad \sigma(b)=aba.$$

Pour $n\geq 6$, on a :

$$\left.\begin{aligned} &p(n,0)=p(k),\\ &p(n,1)=p(k;d\{a,b\})=p(k),\\ &p(n,2)=p(k+1) \end{aligned}\right\} \text{ si } n=3k,$$

$$\left.\begin{aligned} &p(n,0)=p(k;d\{a,b\})=p(k),\\ &p(n,1)=p(k+1),\\ &p(n,2)=p(k+1) \end{aligned}\right\} \text{ si } n=3k+1,$$

$$\left.\begin{aligned} &p(n,0)=p(k+1),\\ &p(n,1)=p(k+1;d\{a,b\})=p(k+1),\\ &p(n,2)=p(k+1), \end{aligned}\right\} \text{ si } n=3k+2,$$

ce qui conduit à

$$p(1)=2 \quad \text{et} \quad p(n)=2n-1 \ \text{ pour } \ n>1.$$

2) Soit σ la substitution définie sur $\mathfrak{a}=\{a,b\}$ par

$$\sigma(a)=a\omega, \quad \sigma(b)=b\omega,$$

où ω est un élément de $\mathfrak{a}^*$ de longueur ℓ. Posons $q=\ell+1$.

Pour n assez grand, on a :

$$p(n+1) - p(n) = p(k+1) - p(k) \quad \text{si} \quad k = \mathrm{mod}_q(n).$$

Si par exemple $\omega = aba$, la fonction de complexité de σ est donnée par

$$p(0) = 1,\ p(1) = 2,\ p(2) = 3,\ p(3) = 5,\ p(4) = 6,\ p(5) = 9$$

et pour $n \geq 5$,

$$p(n+1) - p(n) = p(k+1) - p(k) \quad \text{si } k = \mathrm{mod}_4(n).$$

En particulier, on a pour cette substitution $n < p(n) < 2n$ pour $n \geq 2$.

Remarquons de plus que dans le cas général on a $p(kq) = qp(k)$ pour k assez grand et donc que $b = 0$ et $p(n) = an$ si $p(n) = an + b$ à partir d'un certain rang.

3) De nombreux autres exemples sont à notre disposition grâce à un logiciel de calcul de $p(n)$ utilisant la méthode introduite dans la démonstration du théorème 4, logiciel réalisé par Gilles DIDIER, étudiant au Laboratoire de Mathématiques Discrètes à Marseille.

REMERCIEMENTS.

Ce travail a été effectué au sein de l'équipe *Arithmétique, automates et dynamique symbolique* du Laboratoire de Mathématiques Discrètes de Marseille dont je remercie les membres, et tout particulièrement Bernard HOST pour les conversations fructueuses que nous avons eues.

BIBLIOGRAPHIE

[1] ALLOUCHE (J.-P.). — *Sur la complexité des suites infinies*, prépublication.

[2] ALLOUCHE (J.-P.) et SHALLIT (J.). — *The ring of k-regular sequences*, Theoret. Comput. Sci., t. **98**, 1992, p. 163–187.

[3] CHRISTOL (G.), KAMAE (T.), MENDES-FRANCE (M.) et RAUZY (G.). — *Suites algébriques, automates et substitutions*, Bull. Soc. Math. France, t. **108**, 1980, p. 401–419.

[4] DEKKING (F.-M.). — *Combinatorial and statistical properties of sequences generated by substitutions*, Thèse, Nijmegen, 1980.

[5] GABRIEL (P.). — *Communication privée.*
[6] HOST (B.). — *Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable*, Ergodic Theory Dynamical Systems, t. **6**, 1986, p. 529–540.
[7] MARTIN (J.-C.). — *Minimal flows arising from substitutions of non constant length*, Math. Systems Theory, t. **7**, 1973, p. 73–82.
[8] MOSSÉ (B.). — *Puissances de mots et reconnaissabilité des points fixes de substitutions*, Theoret. Comput. Sci., t. **99**, 1992, p. 327–334.
[9] QUEFFÉLEC (M.). — *Substitution dynamical systems – Spectral analysis*, Lecture Notes in Math., t. **1294**, 1987.
[10] TAPSOBA (T.). — *Complexité de suites automatiques*, Thèse de 3e cycle, Aix–Marseille II, 1987.