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>

© Bulletin de la S. M. F., 1996, tous droits réservés.

L'accès aux archives de la revue « Bulletin de la S. M. F. » (http://smf.emath.fr/Publications/Bulletin/Presentation.html) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/legal.php). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

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
© Société mathématique de France

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