ELSEVIER
Theoretical Computer Science 215 (1999) 31–49
Theoretical Computer Science

Caractérisation des N-écritures et application à l'étude des suites de complexité ultimement $n + c^{ste}$

Gilles Didier*

Institut de Mathématiques de Luminy, CNRS-UPR 9016, Case 930, 163 avenue de Luminy, 13288 Marseille Cedex 9, France

Received June 96; revised April 97
Communicated by D. Perrin

Résumé

Nous appelons N-écriture la transformation sur les mots qui consiste à faire glisser une fenêtre de taille N le long d'un mot et à le recoder en associant à chaque nouveau facteur une lettre différente.

Nous donnons une caractérisation complète des mots obtenus par cette transformation et proposons une méthode permettant de retrouver un antécédent particulier.

Enfin nous appliquons les résultats obtenus pour obtenir une nouvelle description des suites dont le nombre de facteurs de longueur n est de la forme $n + c^{ste}$ pour n assez grand. En particulier, nous explicitons les liens existant entre ces suites et les suites sturmiennes.

Abstract

We call N-écriture the transformation on words which consists on shifting a window of length N along a word and recoding by associating at each factor of length N a new letter.

We give a complete characterisation of the words obtained by this transformation and describe a method to get a particular antecedent.

Finally we apply previous results to give a new description of sequences which contain $n + c^{st}$ different subwords of length n for n great enough. In particular, we explain relations between these sequences and sturmian sequences.

1. Introduction

Etant donné un mot fini ou infini, il est toujours possible de faire glisser une fenêtre de longueur constante le long de ce mot et de recoder selon la configuration observée. Un recodage naturel consiste à associer une lettre distincte à chaque nouveau facteur.

* E-mail: didier@iml.univ-mrs.fr.

0304-3975/99/$ – see front matter
PII: S0304-3975(97)00122-9

On obtient ainsi un nouveau mot que nous appellerons par la suite N-écriture du précédent, où N est la longueur de la fenêtre.

La N-écriture d'un mot donné conserve certaines de ses propriétés et peut être alors étudiée pour déterminer certaines de ses caractéristiques. Par exemple Martine Queféllec utilise cette transformation pour calculer les fréquences des facteurs de points fixes de substitutions [11].

Dans la première partie de cet article nous donnons une caractérisation complète des mots sur un alphabet fini qui sont des N-écritures en distinguant tout d'abord le cas où $N = 2$. Dans tous les cas, le critère de décision ne tient compte que des facteurs de longueur 2 apparaissant dans le mot. Nous proposons également, dans le cas où le critère est vérifié, une construction qui permet d'associer à une N-écriture un antécédent particulier.

Dans la seconde partie, nous utilisons les résultats obtenus pour caractériser les mots infinis (ou suites) dont le nombre de facteurs différents de longueur n, ou complexité est égal à $n + c^{ste}$ pour n assez grand.

Comme son nom l'indique, la fonction de complexité d'une suite traduit assez bien l'idée que l'on se fait intuitivement de sa "complexité". En particulier une suite est ultimement périodique – donc "simple" – si et seulement si sa fonction de complexité est constante à partir d'un certain rang. Comme cette fonction est croissante, une suite non ultimement périodique est de complexité au moins $n + 1$, borne effectivement atteinte par certaines suites.

De nombreux exemples de complexité de suites et diverses références sur ce sujet peuvent être trouvés dans [2]. Parmi les développements postérieurs à ce survol, nous citerons les travaux de J. Cassaigne qui a, entre autres, montré que si une fonction de complexité est sous-affine, la suite de ses différences premières est bornée [5].

Le cas des suites de complexité $n + 1$ ou *suites sturmiennes* a été longuement et efficacement étudié. Pour en savoir plus, nous conseillons la lecture de la synthèse de J. Berstel consacrée à ce sujet; celle-ci contient également les principales références sur ce thème [4]. On connaît aujourd'hui de multiples caractérisations de ces suites et diverses façons de les générer.

En particulier, M. Morse et G. A. Hedlund ont montré que ces suites peuvent s'interpréter comme le codage de l'orbite d'un point sur le cercle unité sous l'action d'une rotation d'angle α irrationnel lorsqu'on partitionne le cercle en deux intervalles de longueurs respectives α et $1 - \alpha$ [8, 9].

Une autre méthode de construction a été proposé par G. Rauzy et P. Arnoux qui ont démontré que l'on peut également générer ces suites en itérant infiniment deux morphismes particuliers [3]. Les suites de complexité (ultimement) $n + c^{ste}$ ont une combinatoire proche des suites sturmiennes et leurs caractérisations font naturellement intervenir ces dernières. Citons le travail de P. Alessandri qui a caractérisé les suites de complexité $n + 2$ en s'attachant à en donner une interprétation géométrique [1].

Plus récemment, S. Ferenczi et C. Mauduit ont proposé une caractérisation des suites de complexité $n + c^{ste}$ [7]. En utilisant des méthodes analogues à [3], ils montrent qu'une suite de complexité $n+c^{ste}$ récurrente est l'image par un morphisme σ d'une suite sturmienne.

Notre résultat est différent : il fait intervenir d'autres morphismes dont tous les itérés des lettres sont de longueurs constantes égales à un entier q (des morphismes q-uniformes) et les N-écritures.

L'idée d'utiliser les N-écritures pour étudier les suites de complexité $n+c^{ste}$ vient naturellement lorsqu'on remarque que la N-écriture d'une suite sturmienne est de complexité $p(n)=n+N$ pour tout entier naturel non nul n et s'interprète comme le codage de la trajectoire d'un point x sous l'action d'une rotation d'angle α irrationnel, le cercle étant découpé en $N+1$ intervalles déduits du découpage $(\alpha, 1-\alpha)$.

Dans la dernière partie de ce travail, nous verrons que les suites de complexité $n+c^{ste}$, même dans le cas récurrent, ne sont pas toujours des N-écritures de suites sturmiennes : il faut parfois faire intervenir des morphismes très particuliers que nous appelons I-morphismes (le "I" pour "insérant" car leur application revient à insérer entre chaque lettre de la séquence initiale un nombre constant de lettres toutes distinctes n'appartenant pas à l'alphabet de départ).

2. Notations et définitions

Etant donné un ensemble E, on note $\#E$ le cardinal de E. On appelle *alphabet* tout ensemble fini non vide d'éléments appelés lettres ou symboles.

Un *mot* sur un alphabet X est une suite de lettres de X indexée sur $\{0,1,\ldots,n\}$ pour un entier naturel n, si le mot est fini, et sur $\mathbb{N}$ si le mot est infini. La *longueur* d'un mot w, notée $|w|$, est le nombre de lettres le composant. On désignera par:
- X^r: l'ensemble des mots de longueur r sur X,
- X^*: l'ensemble des mots finis sur X,
- $X^{\mathbb{N}}$: l'ensemble des mots infinis ou suites sur X.

Le *concaténé* de deux mots non vides $u=u_0\ldots u_{|u|-1}$ et $v=v_0\ldots v_{|v|-1}$, noté uv, est le mot $u_0\ldots u_{|u|-1}v_0\ldots v_{|v|-1}$ le mot vide étant neutre pour la concaténation.

Soient A et B deux alphabets. Toute application de A vers B^* se prolonge par concaténation en morphisme de A^* vers B^* et en application de $A^{\mathbb{N}}$ vers $B^{\mathbb{N}}$. En particulier, on appellera *projection* les morphismes "lettre à lettre" (prolongeant une application de A vers B).

Soit u un mot sur un alphabet A. On note $L_n(u)$ l'ensemble des facteurs de longueur n de u, c'est-à-dire des mots de la forme $u_iu_{i+1}\ldots u_{i+n-1}$. Dans la suite, dire que u est un mot sur un alphabet A sous-entendra que toute lettre de A admet au moins une occurrence dans u (i.e. $A=L_1(u)$).

On appellera *préfixe* de longueur i (inférieure ou égale à $|u|$) de u le mot $u_0u_1\ldots u_{i-1}$ si $i>0$ et le mot vide sinon. De même, si u est fini, on appellera *suffixe* de longueur i de u le mot $u_{|u|-i}u_{|u|-i+1}\ldots u_{|u|-1}$ si $i>0$ et le mot vide sinon. Un préfixe (resp. un suffixe) *strict* de u est de longueur strictement inférieure à $|u|$.

Une suite u est dite *récurrente* si tout facteur fini de u y apparaît une infinité de fois.

Définition 1. On appelle fonction de complexité de u et on note $p(u,n)$ la fonction qui à tout entier n non nul fait correspondre le nombre de facteurs différents de longueur n de $u: p(u,n) = \#L_n(u)$. On note alors $s(u,n)$ la différence première de la fonction de complexité de u:
$s(u,n) = p(u,n+1) - p(u,n)$.

Définition 2. Soient A et B deux alphabets et m un entier naturel. Une fonction de bloc Θ de rayon m est une application de A^m vers B. Cette application se prolonge en applications de l'ensemble des mots sur A de longueur supérieure à m vers B^* et de $A^{\mathbb{N}}$ vers $B^{\mathbb{N}}$ en associant à tout mot w le concaténé des images des facteurs de longueur m apparaissant successivement dans w:

$$\Theta(w) = \Theta(w_0 w_1 \dots w_{m-1})\Theta(w_1 w_2 \dots w_m)\dots .$$

On continue de noter Θ les applications ainsi obtenues à valeurs dans B^* et $B^{\mathbb{N}}$ respectivement.

On remarque que l'image d'un mot w ($|w| \geqslant m$) par une fonction de bloc de rayon m est de longueur $|w| - m + 1$.

Définition 3. Soit u un mot et N un entier naturel. On appelle fonction de N-écriture de u toute fonction de bloc de rayon N bijective de $L_N(u)$ sur un alphabet B. L'image du mot u par une telle fonction est appelée N-écriture de u.
Par abus, on note $u(N)$ toute N-écriture de u.

Remarques. (1) Pour tout entier n, les fonctions de complexité de u et de $u(N)$ sont liées par la relation

$$p(u(N), n) = p(u, n+N-1).$$

(2) Si N est supérieur ou égal à 2, la N-écriture d'un mot u est la 2-écriture de sa $(N-1)$-écriture. On a:

$$u(N) = (u(N-1))(2).$$

(3) A toute lettre de B correspond un unique facteur de longueur N de u. Pour tout entier i inférieur à N, on notera φ_i l'application de B vers A qui à toute lettre de B associe la lettre de rang i du facteur de u lui correspondant. On a alors: $\varphi_i(u(N)) = u_i u_{i+1} \dots$

3. Caractérisation des mots qui sont des N-écritures

Il est toujours possible, étant donné un mot u, de déterminer sa N-écriture. Notre propos est inversement, étant donné u, de décider s'il est la N-écriture d'un mot v. Le critère que nous donnons fait intervenir des relations d'équivalence sur les lettres du mot considéré.

3.1. Cas où $N = 2$

Soit u un mot (fini ou infini) sur l'alphabet A. On note $r_u^{(0)}$ la relation sur A définie par:

(i) tout symbole de u est en relation avec lui même,

(ii) deux symboles a et b sont en relation s'il existe un symbole c tel que:

- soit les facteurs ca et cb apparaissent dans u,
- soit $c\, r_u^{(0)} a$ et $c\, r_u^{(0)} b$.

La règle (i) ne sert qu'à assurer la réflexivité de $r_u^{(0)}$. En effet si la première lettre de u n'admet qu'une occurrence, elle ne peut pas apparaître au second rang d'un facteur de longueur deux de u et n'est pas mise en relation avec elle-même par la règle (ii).

La règle (ii) est récursive. Elle exprime que $r_u^{(0)}$ est obtenue en saturant par transitivité la relation qui met en liaison deux lettres de A si elles peuvent être précédées par la même lettre dans le langage de u.

Ainsi définie, $r_u^{(0)}$ est une relation d'équivalence et on note $\mathscr{R}_u^{(0)}$ la partition de l'alphabet A par $r_u^{(0)}$.

On définit de la même façon $r_u^{(1)}$:

(i) tout symbole de u est en relation avec lui-même,

(ii) deux symboles a et b sont en relation s'il existe un symbole c tel que

- soit les facteurs ac et bc apparaissent dans u,
- soit $c\, r_u^{(1)} a$ et $c\, r_u^{(1)} b$.

On note $\mathscr{R}_u^{(1)}$ la partition de l'alphabet A par $r_u^{(1)}$.

Soient u un mot sur un alphabet A et C un sous-ensemble de A. On note :

- $P_u(C)$ l'ensemble des lettres de A préfixes d'un facteur de longueur deux de u dont la deuxième lettre appartient à C:

$$P_u(C) = \{a \in A \,/\, \exists b \in C, ab \in L_2(u)\}.$$

- $S_u(C)$ l'ensemble des lettres de A suffixes d'un facteur de longueur deux de u dont la première lettre appartient à C:

$$S_u(C) = \{a \in A \,/\, \exists b \in C, ba \in L_2(u)\}.$$

On remarque que pour tout sous-ensemble C de A on a:

$$C \subset P_u(S_u(C)) \text{ et } C \subset S_u(P_u(C)).$$

Le lemme suivant montre qu'à toute classe de $\mathscr{R}_u^{(0)}$ est naturellement associée une unique classe de $\mathscr{R}_u^{(1)}$, et inversement, sauf dans un cas bien particulier que nous préciserons.

Lemme 1. *Soit u mot sur un alphabet A. Pour toute classe p de* $\mathscr{R}_u^{(0)}$ *et pour toute classe q de* $\mathscr{R}_u^{(1)}$:
(i) *Si* $P_u(p)$ *est non vide alors* $P_u(p)$ *est une classe de* $\mathscr{R}_u^{(1)}$.
(ii) *Si* $S_u(q)$ *est non vide alors* $S_u(q)$ *est une classe de* $\mathscr{R}_u^{(0)}$.

Preuve. Montrons tout d'abord que si p est une classe de $\mathscr{R}_u^{(0)}$ telle que $P_u(p)$ soit non vide alors il existe q une classe de $\mathscr{R}_u^{(1)}$ telle que $P_u(p) \subset q$. Soient p une classe de $\mathscr{R}_u^{(0)}$, a et b deux éléments de p suffixes de facteurs de longueur deux de u. Si e et f sont deux lettres de A telles que ea et fb soient facteurs de u alors $e\, r_u^{(1)} f$. Autrement dit, e et f appartiennent à une même classe de $\mathscr{R}_u^{(1)}$. En effet, si a et b appartiennent à une même classe de $\mathscr{R}_u^{(0)}$ alors, par définition, il existe une lettre c telle que:
– soit les facteurs ca et cb apparaissent dans u, on a $c\, r_u^{(1)} e$, $c\, r_u^{(1)} f$ et donc $e\, r_u^{(1)} f$.
– soit $c\, r_u^{(0)} a$ et $c\, r_u^{(0)} b$ et par induction il existe deux lettres h et h' telles que $h\, r_u^{(1)} e$, $h'\, r_u^{(1)} f$ et $h'\, r_u^{(1)} h$; donc $e\, r_u^{(1)} f$.

De la même façon, si q est une classe de $\mathscr{R}_u^{(1)}$ telle que $S_u(q)$ soit non vide alors il existe p une classe de $\mathscr{R}_u^{(0)}$ telle que $S_u(q) \subset p$.

Considérons p une classe de $\mathscr{R}_u^{(0)}$ telle que $P_u(p)$ soit non vide et q la classe de $\mathscr{R}_u^{(1)}$ vérifiant $P_u(p) \subset q$. L'ensemble $S_u(q)$ est alors non vide. Il est donc inclus dans une classe de $\mathscr{R}_u^{(0)}$ qui ne peut être que p (i.e. on a aussi $S_u(q) \subset p$). On a les inclusions suivantes:

$$q \subset P_u(S_u(q)) \subset P_u(p) \subset q \text{ et } p \subset S_u(P_u(p)) \subset S_u(q) \subset p,$$

d'où l'on déduit que $P_u(p) = q$ et $S_u(q) = p$. □

Remarques. Le cas où $P_u(p)$ est vide arrive uniquement lorsque u_0 n'admet qu'une seule occurrence dans u. La classe $\{u_0\}$ est alors la seule classe de $\mathscr{R}_u^{(0)}$ telle que $P_u(\{u_0\})$ soit vide. De la même façon, si u est fini et si $u_{|u|-1}$ n'admet qu'une occurrence, la classe $\{u_{|u|-1}\}$ est alors la seule classe de $\mathscr{R}_u^{(1)}$ telle que $S_u(\{u_{|u|-1}\})$ soit vide. Hormis ces deux cas, on peut mettre en correspondance les éléments de $\mathscr{R}_u^{(0)}$ et de $\mathscr{R}_u^{(1)}$.

Proposition 1. *Soit u un mot sur un alphabet A. Il existe un mot v tel que u soit la 2-écriture de v si et seulement si l'intersection de deux classes quelconques de* $\mathscr{R}_u^{(0)}$ *et de* $\mathscr{R}_u^{(1)}$ *contient au plus un élément.*

Preuve. Supposons tout d'abord qu'il existe un mot v sur un alphabet B telle que $u = v(2)$. Par définition, il existe une fonction de 2-écriture Θ de $L_2(v)$ vers A telle que $u = \Theta(v)$. Considérons φ_0 (resp. φ_1) l'application de A vers B qui à tout élément de A fait correspondre la première (resp. la seconde) lettre du mot de longueur deux sur B qui lui est associé. Autrement dit toute lettre a de A est l'image par Θ de $\varphi_0(a)\varphi_1(a)$.

On note $\mathscr{F}_0$ (resp. ($\mathscr{F}_1$) la partition de A induite par φ_0 (resp. par φ_1). Par construction, si deux symboles a et b de A sont en relation par $r_u^{(0)}$ alors $\varphi_0(a)=\varphi_0(b)$. En effet si $a\, r_u^{(0)} b$ alors

- soit $a=b=u_0$ et $\varphi_0(a)=\varphi_0(b)$,
- soit il existe c tel que:
  - soit ca et cb apparaissent dans u et $\varphi_0(a)=\varphi_1(c)=\varphi_0(b)$,
  - soit $c\, r_u^{(0)} a$ et $c\, r_u^{(0)} b$ et par induction : $\varphi_0(a)=\varphi_0(c)=\varphi_0(b)$.

Tout élément de la partition $\mathscr{F}_0$ est donc une réunion d'éléments de $\mathscr{R}_u^{(0)}$.

On montre de la même façon que tout élément de la partition $\mathscr{F}_1$ est une réunion d'éléments de $\mathscr{R}_u^{(1)}$.

De plus, l'intersection d'un élément de $\mathscr{F}_0$ et d'un élément de $\mathscr{F}_1$ est de cardinal inférieur ou égal à 1. En effet, si deux lettres a et b de A sont telles que $\varphi_0(a)=\varphi_0(b)$ et $\varphi_1(a)=\varphi_1(b)$, alors elles sont égales (on a: $a=\Theta(\varphi_0(a)\varphi_1(a))=\Theta(\varphi_0(b)\varphi_1(b))=b$).

Ceci implique que l'intersection de deux classes de $\mathscr{R}_u^{(0)}$ et de $\mathscr{R}_u^{(1)}$ contient au plus un élément.

Inversement, supposons que quels que soient p un élément de $\mathscr{R}_u^{(0)}$ et q un élément de $\mathscr{R}_u^{(1)}$, le cardinal de l'intersection de p et de q est inférieur ou égal à 1. Nous allons construire un mot v et une fonction de 2-écriture Θ tels que $\Theta(v)=u$. Il faut avant tout que nous définissions l'alphabet X de v.

Notons X_1 l'ensemble des couples de la forme $(p,P_u(p))$ où p est une classe de $\mathscr{R}_u^{(0)}$ et X_2 l'ensemble des couples $(S_u(q),q)$ où q est une classe de $\mathscr{R}_u^{(1)}$. D'après le lemme 1, X_1 et X_2 coïncident pour tous les couples $(p,P_u(p))$ dont les composantes sont non vides. Il y a éventuellement deux couples n'appartenant pas à $X_1\cap X_2$: $(\{u_0\},\emptyset)$ si la première lettre de u n'admet qu'une occurrence et, dans le cas où u est fini, $(\emptyset,\{u_{|u|-1}\})$ si la dernière lettre de u n'admet qu'une occurrence.

Prenons comme alphabet X l'union de X_1 et X_2. Toute classe p de $\mathscr{R}_u^{(0)}$ apparaît dans un unique couple (p,q) de X. De même, à chaque classe q de $\mathscr{R}_u^{(1)}$ est associé un unique élément (p,q) de X.

On définit alors la projection π_0 (resp. π_1) de A vers X qui à chaque élément a de A fait correspondre l'élément (p,q) de X tel que a appartienne à p (resp. q). Soient a et b deux lettres de A ; si ab est un facteur de longueur deux de u, alors on a: $\pi_0(b)=\pi_1(a)$.

Les projections π_0 et π_1 nous permettent de construire un mot v qui convient. Il faut ici distinguer le cas fini du cas infini:

- si u est fini, on pose $v=\pi_0(u)\pi_1(u_{|u|-1})=\pi_0(u_0)\pi_1(u)$,
- si u est infini, on pose $v=\pi_0(u)=\pi_0(u_0)\pi_1(u)$.

Que v soit fini ou infini, pour tout facteur xy de v, il existe une lettre a de A telle que $x=\pi_0(a)$ et $y=\pi_1(a)$. De plus, par hypothèse, si pour deux lettres a et b de A on a $\pi_0(a)=\pi_0(b)$ et $\pi_1(a)=\pi_1(b)$, alors $a=b$ (sinon a et b appartiendraient à une même classe de $\mathscr{R}_u^{(0)}$ et à une même classe de $\mathscr{R}_u^{(1)}$). A tout facteur de longueur deux de v

correspond une unique lettre de A. On peut donc définir l'application Θ de $L_2(v)$ vers A par $\Theta(\pi_0(a)\pi_1(a)) = a$. On a alors:
– dans le cas fini:

$$\begin{aligned}\Theta(v) &= \Theta(\pi_0(u_0)\pi_0(u_1))\Theta(\pi_0(u_1)\pi_0(u_2))\dots\Theta(\pi_0(u_{|u|-1})\pi_1(u_{|u|-1}))\\ &= \Theta(\pi_0(u_0)\pi_1(u_0))\Theta(\pi_0(u_1)\pi_1(u_1))\dots\Theta(\pi_0(u_{|u|-1})\pi_1(u_{|u|-1}))\\ &= u_0u_1\dots u_{|u|-1}\\ &= u,\end{aligned}$$

– dans le cas infini:

$$\begin{aligned}\Theta(v) &= \Theta(\pi_0(u_0)\pi_0(u_1))\Theta(\pi_0(u_1)\pi_0(u_2))\dots\\ &= \Theta(\pi_0(u_0)\pi_1(u_0))\Theta(\pi_0(u_1)\pi_1(u_1))\dots\\ &= u_0u_1\dots\\ &= u.\end{aligned}$$

La fonction Θ est une bijection de $L_2(v)$ vers A: elle est surjective car toute lettre de A apparaît dans u, et injective car si $a = b$ alors $\pi_0(a) = \pi_0(b)$ et $\pi_1(a) = \pi_1(b)$. Nous avons construit un mot v et une fonction Θ de 2-écriture de v tels que $u = \Theta(v)$, ce qui termine la démonstration. □

Corollaire 1. *Soit u une 2-écriture d'un mot w. Si v est le mot vérifiant $v(2) = u$ obtenu par la méthode décrite dans la seconde partie de la démonstration précédente, alors il existe une projection ϕ de $L_1(v)$ à valeurs dans $L_1(w)$ telle que $w = \phi(v)$. En particulier, l'alphabet de v est de cardinal supérieur ou égal au cardinal de l'alphabet de w.*

Preuve. C'est une conséquence directe de la première partie de la démonstration précédente.

Exemple. Considérons le mot fini v suivant:

$$v = a\,b\,a\,c\,d\,e\,a\,b\,d\,a\,c\,f\,a\,b\,g.$$

Construisons u, une de ses 2-écritures en numérotant ses facteurs de longueur 2 dans l'ordre de leur première apparition. On a:

$$u = 0\,1\,2\,3\,4\,5\,0\,6\,7\,2\,8\,9\,0\,10.$$

L'alphabet de u est alors:

$$A = L_1(u) = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\},$$

et l'ensemble des facteurs de longueur 2 apparaissant dans ce mot est:

$$L_2(u) = \{0\,1, 1\,2, 2\,3, 3\,4, 4\,5, 5\,0, 0\,6, 6\,7, 7\,2, 2\,8, 8\,9, 9\,0, 0\,10\}.$$

On peut ainsi calculer les partitions $\mathscr{R}_u^{(0)}$ et $\mathscr{R}_u^{(1)}$. On obtient:

$$\mathscr{R}_u^{(0)} = \{\{0\},\ \{1,6,10\},\ \{2\},\ \{3,8\},\ \{4\},\ \{5\},\ \{9\},\ \{7\}\} \text{ et}$$

$$\mathscr{R}_u^{(1)} = \{\{0\},\ \{1,7\},\ \{2\},\ \{3\},\ \{4\},\ \{5,9\},\ \{8\},\ \{10\}\}.$$

On vérifie que u est bien une 2-écriture car toute intersection d'un élément de $\mathscr{R}_u^{(0)}$ et d'un élément de $\mathscr{R}_u^{(1)}$ contient au plus un élément. En utilisant la méthode décrite dans la seconde partie de la démonstration, on construit tout d'abord le nouvel alphabet X en associant les classes de $\mathscr{R}_u^{(0)}$ et de $\mathscr{R}_u^{(1)}$:

$$\begin{array}{lll}
X & \mathscr{R}_u^{(0)} & \mathscr{R}_u^{(1)} \\
a & =(\{0\}, & \{5,9\}) \\
b & =(\{1,6,10\}, & \{0\}) \\
a' & =(\{2\}, & \{1,7\}) \\
c & =(\{3,8\}, & \{2\}) \\
d & =(\{4\}, & \{3\}) \\
e & =(\{5\}, & \{4\}) \\
d' & =(\{7\}, & \{6\}) \\
f & =(\{9\}, & \{8\}) \\
g & =(\emptyset, & \{10\})
\end{array}$$

Les projections π_0 et π_1 de l'alphabet A vers X se déduisent de ces associations:

π_0	π_1
$0 \to a$	$0 \to b$
$1 \to b$	$1 \to a'$
$2 \to a'$	$2 \to c$
$3 \to c$	$3 \to d$
$4 \to d$	$4 \to e$
$5 \to e$	$5 \to a$
$6 \to b$	$6 \to d'$
$7 \to d'$	$7 \to a'$
$8 \to c$	$8 \to f$
$9 \to f$	$9 \to a$
$10 \to b$	$10 \to g$

On peut alors construire un mot v':

$$\begin{aligned}
v' &= \pi_0(u)\pi_1(u_{|u|-1}) \\
&= \pi_0(0\,1\,2\,3\,4\,5\,0\,6\,7\,2\,8\,9\,0\,10)\pi_1(10) \\
&= a\,b\,a'\,c\,d\,e\,a\,b\,d'\,a'\,c\,f\,a\,b\,g,
\end{aligned}$$

et vérifier que la fonction de bloc suivante est telle que $\Theta(v') = u$:

$$\begin{array}{lcl} & \Theta & \\ ab & \to & 0 \\ ba' & \to & 1 \\ a'c & \to & 2 \\ cd & \to & 3 \\ de & \to & 4 \\ ea & \to & 5 \\ bd' & \to & 6 \\ d'a' & \to & 7 \\ cf & \to & 8 \\ fa & \to & 9 \\ bg & \to & 10 \end{array}$$

Le mot v' obtenu est différent du mot v de départ. Si l'on note ϕ la projection de $L_1(v')$ vers $L_1(v)$ qui identifie les lettres a et a' et d et d', on a: $v = \phi(v')$.

3.2. Cas général

Pour aborder le cas général, nous avons à définir de nouvelles relations d'équivalence. Pour tout entier $k \in \{0, 1, \ldots, N-1\}$, on note $r_u^{(k,N)}$ la relation sur A définie par:

Soient deux lettres a et b de A. Alors $a\ r_u^{(k,N)} b$ si au moins l'une des assertions suivantes est vérifiée:

(i) $a = b$,

(ii) il existe un entier $i \in \{1, \ldots, N-k-1\}$ et deux éléments $(a_1, a_2, \ldots, a_i)$ et $(b_1, b_2, \ldots, b_i)$ de A^i tels que:
- $a_i = b_i$,
- $\{a_i a_{i-1}, a_{i-1} a_{i-2}, \ldots, a_1 a, b_i b_{i-1}, b_{i-1} b_{i-2}, \ldots, b_1 b\} \subset L_2(u)$,

(iii) il existe un entier $i \in \{1, \ldots, k\}$ et deux éléments $(a_1, a_2, \ldots, a_i)$ et $(b_1, b_2, \ldots, b_i)$ de A^i tels que:
- $a_i = b_i$,
- $\{a a_1, a_1 a_2, \ldots, a_{i-1} a_i, b b_1, b_1 b_2, \ldots, b_{i-1} b_i\} \subset L_2(u)$,

(iv) il existe un symbole $c \in A$ tel que $c\ r_u^{(k,N)}\ a$ et $c\ r_u^{(k,N)}\ b$.

Pour toutes les valeurs de k permises, les relations $r_u^{(k,N)}$ sont des relations d'équivalence et on note $\mathscr{R}_u^{(k,N)}$ la partition de A par $r_u^{(k,N)}$.

Comme dans le cas précédent, on peut mettre en relation les éléments de deux partitions successives, c'est-à-dire correspondant à deux valeurs successives de k.

Lemme 2. *Soit u mot sur un alphabet A. Soit un entier $k \in \{0, 1, \ldots, N-2\}$. Si p est une classe de $\mathscr{R}_u^{(k,N)}$ telle que $P_u(p)$ est non vide alors il existe une unique classe q de $\mathscr{R}_u^{(k+1,N)}$ telle que $P_u(p) \subset q$. Inversement, soit un entier $k \in \{1, 2, \ldots, N-1\}$. Si q*

est une classe de $\mathscr{R}_u^{(k,N)}$ telle que $S_u(q)$ soit non vide alors il existe une unique classe p de $\mathscr{R}_u^{(k-1,N)}$ telle que $P_u(q) \subset q$.

Preuve. Soient k un entier naturel strictement inférieur à $N-1$, a et b deux symboles de A tels que $a\, r_u^{(k,N)}\, b$.

Nous allons montrer que si e et f sont deux lettres de A telles que ea et fb sont des facteurs de u alors $e\, r_u^{(k+1,N)}\, f$, ce qui prouvera la première assertion.

Comme $a\, r_u^{(k,N)}\, b$, quatre cas se présentent:

cas (i): $a=b$. Les facteurs ea et fa apparaissent donc dans u et on a $e\, r_u^{(k+1,N)}\, f$ (l'assertion (iii) est vérifiée car $k+1 \geqslant 1$).

cas (ii): il existe un entier $i \in \{1,\dots,N-k-1\}$ et deux éléments $(a_1,a_2,\dots,a_i)$ et $(b_1,b_2,\dots,b_i)$ de A^i tels que:

$$a_i = b_i,$$

$$\{a_i a_{i-1}, a_{i-1}a_{i-2}, \dots, a_1 a, b_i b_{i-1}, b_{i-1}b_{i-2}, \dots, b_1 b\} \subset L_2(u),$$

d'où l'on déduit que $a_1\, r_u^{(k+1,N)}\, b_1$.

De plus ea et fb sont des facteurs de u. On en déduit que $a_1\, r_u^{(k+1,N)}\, e$ et $b_1\, r_u^{(k+1,N)}\, f$ (l'assertion (ii) est vérifiée), et par transitivité on a $e\, r_u^{(k+1,N)}\, f$.

cas (iii): il existe un entier $i \in \{1,\dots,k\}$ et deux éléments $(a_1,a_2,\dots,a_i)$ et $(b_1,b_2,\dots, b_i)$ de A^i tels que:

$$a_i = b_i,$$

$$\{aa_1, a_1a_2, \dots, a_{i-1}a_i, bb_1, b_1b_2, \dots, b_{i-1}b_i\} \subset L_2(u).$$

On a donc:

$$a_i = b_i,$$

$$\{ea, aa_1, a_1a_2, \dots, a_{i-1}a_i, fb, bb_1, b_1b_2, \dots, b_{i-1}b_i\} \subset L_2(u),$$

et par définition $e\, r_u^{(k+1,N)} f$ (l'assertion (iii) est vérifiée).

cas (iv): il existe une lettre c de A telle que $c\, r_u^{(k,N)}\, a$ et $c\, r_u^{(k,N)}\, b$, et par induction sur les cas précédents, on obtient également $e\, r_u^{(k+1,N)}\, f$.

On démontre de la même façon la seconde affirmation du lemme. □

Proposition 2. *Soit u un mot sur un alphabet A. Il existe un mot v tel que u soit la N-écriture de v si et seulement si pour tout couple (a,b) de lettres de A avec $a \neq b$, il existe un entier naturel k strictement inférieur à N tel que a et b ne sont pas en relation par $r_u^{(k,N)}$.*

Preuve. On suppose qu'il existe un mot v sur un alphabet B et une fonction de N-écriture Θ tels que $u=\Theta(v)$. On considère les projections φ_i pour tout entier $i \in \{0,1,\dots,N-1\}$. Tout symbole a de A est alors l'image par Θ du facteur $\varphi_0(a)\varphi_1(a)$

$\dots\varphi_{N-1}(a)$ de v. Ceci implique que si ab est un facteur de u, on a pour tout entier $k\in\{1,2,\dots,N-1\}$, $\varphi_k(a)=\varphi_{k-1}(b)$.

On montre alors que si deux lettres a et b de A sont telles que $a\ r_u^{(k,N)}\ b$ alors $\varphi_k(a)=\varphi_k(b)$.

En effet si $a\ r_u^{(k,N)}\ b$, on distingue les cas suivants:

cas (i): $a=b$. On a bien $\varphi_k(a)=\varphi_k(b)$.

cas (ii): il existe un entier $i\in\{1,\dots,N-k\}$ et deux éléments $(a_1,a_2,\dots,a_i)$ et $(b_1,b_2,\dots,b_i)$ de A^i tels que:

$$a_i=b_i,$$

$$\{a_ia_{i-1},a_{i-1}a_{i-2},\dots,a_1a,b_ib_{i-1},b_{i-1}b_{i-2},\dots,b_1b\}\subset L_2(u).$$

On a:

$\varphi_k(a)=\varphi_{k+1}(a_1)=\cdots=\varphi_{k+i}(a_i)$,
$\varphi_k(b)=\varphi_{k+1}(b_1)=\cdots=\varphi_{k+i}(b_i)$,
Et donc: $\varphi_k(a)=\varphi_k(b)$.

cas (iii): il existe un entier $i\in\{1,\dots,k\}$ et deux éléments de A^i $(a_1,a_2,\dots,a_i)$ et $(b_1,b_2,\dots,b_i)$ de A^i tels que:

$$a_i=b_i,$$

$$\{aa_1,a_1a_2,\dots,a_{i-1}a_i,bb_1,b_1b_2,\dots,b_{i-1}b_i\}\subset L_2(u)$$

On a:

$\varphi_k(a)=\varphi_{k-1}(a_1)=\cdots=\varphi_{k-i}(a_i)$,
$\varphi_k(b)=\varphi_{k-1}(b_1)=\cdots=\varphi_{k-i}(b_i)$,
Et donc: $\varphi_k(a)=\varphi_k(b)$.

cas (iv): il existe une lettre c de A telle que $c\ r_u^{(k,N)}\ a$ et $c\ r_u^{(k,N)}\ b$ et par induction sur les cas précédents, on obtient également $\varphi_k(a)=\varphi_k(b)$.

Comme Θ est une bijection, il ne peut y avoir deux symboles différents a et b de A, tels que pour tout entier $k\in\{0,1,\dots,N-1\}$, $\varphi_k(a)=\varphi_k(b)$. Il n'y donc pas non plus deux symboles a et b tels que $a\ r_u^{(k,N)}\ b$ pour tout $k\in\{0,1,\dots,N-1\}$.

Inversement supposons que si a et b sont deux symboles de A tels que $a\neq b$, alors a et b ne sont pas en relation par $r_u^{(k,N)}$, pour au moins un entier $k\in\{0,1,\dots,N-1\}$.

Désignons par X_1 l'ensemble des N-uplets de la forme $(q_0,\dots,q_{N-1})$ où q_0 est un élément de $\mathscr{R}_u^{(0,N)}$ et pour tout $i\in\{1,\dots,N-1\}$, q_i est la classe de $\mathscr{R}_u^{(i,N)}$ telle que $P_u(q_{i-1})\subset q_i$.

De façon symétrique, X_2 est l'ensemble des N-uplets de la forme $(q_0,\dots,q_{N-1})$ où q_{N-1} est un élément de $\mathscr{R}_u^{(N-1,N)}$, et pour tout $i\in\{0,\dots,N-2\}$, q_i est la classe de $\mathscr{R}_u^{(i,N)}$ telle que $S_u(q_{i+1})\subset q_i$.

Notons alors X l'union de X_1 et X_2. Cet ensemble va constituer le nouvel alphabet sur lequel nous allons construire un mot v tel que $u=v(N)$.

Pour tout $k\in\{0,\dots,N-1\}$, on note π_k la projection de A vers X qui associe à toute lettre a de A le N-uplet $(q_0,\dots,q_{N-1})$ tel que $a\in q_k$.

Les projections π_k nous permettent de construire un mot v à partir de u. Il faut ici encore distinguer le cas fini du cas infini:
- si u est fini, on pose $v=\pi_0(u)\pi_1(u_{|u|-1})\dots\pi_{N-1}(u_{|u|-1})$,
- si u est infini, on pose $v=\pi_0(u)$.

Par hypothèse, à tout facteur de longueur n de v correspond une unique lettre a de A telle que $v=\pi_0(a)\pi_1(a)\dots\pi_{N-1}(a)$ (sinon on aurait deux lettres a et b, avec $a\neq b$, telles que $a\, r_u^{(k,N)}\, b$ pour tout $k\in\{0,1,\dots,N-1\}$). On peut donc définir l'application Θ de $L_N(v)$ vers A par $\Theta(\pi_0(a)\pi_1(a)\dots\pi_{N-1}(a))=a$. On a alors:
- dans le cas fini:

$$\begin{aligned}\Theta(v) &= \Theta(\pi_0(u_0)\dots\pi_0(u_{N-1}))\Theta(\pi_0(u_1)\dots\pi_0(u_N))\dots\\ &\quad\ \Theta(\pi_0(u_{|u|-1})\dots\pi_{N-1}(u_{|u|-1}))\\ &= \Theta(\pi_0(u_0)\dots\pi_{N-1}(u_0))\Theta(\pi_0(u_1)\dots\pi_{N-1}(u_1))\dots\\ &\quad\ \Theta(\pi_0(u_{|u|-1})\dots\pi_{N-1}(u_{|u|-1}))\\ &= u_0u_1\dots u_{|u|-1}\\ &= u,\end{aligned}$$

- dans le cas infini:

$$\begin{aligned}\Theta(v) &= \Theta(\pi_0(u_0)\pi_0(u_1)\dots\pi_0(u_{n-1}))\Theta(\pi_0(u_1)\pi_0(u_2)\dots\pi_0(u_N))\dots\\ &= \Theta(\pi_0(u_0)\pi_1(u_0)\dots\pi_{N-1}(u_0))\Theta(\pi_0(u_1)\pi_1(u_1)\dots\pi_{N-1}(u_1))\dots\\ &= u_0u_1\dots\\ &= u.\end{aligned}$$

La fonction Θ est une bijection de $L_N(v)$ vers A: elle est surjective car toute lettre de A apparaît dans u, et injective car si $a=b$ alors $\pi_k(a)=\pi_k(b)$ pour tout entier naturel k strictement inférieur à N. Nous avons construit un mot v et une fonction Θ de N-écriture de v tels que $u=\Theta(v)$, ce qui termine la démonstration. □

On remarque que dans le cas où $N=2$ dire que l'intersection de deux classes quelconques de $\mathscr{R}_u^{(0)}$ et de $\mathscr{R}_u^{(1)}$ contient au plus un élément équivaut à dire que pour tout couple (a,b) de lettres de A avec $a\neq b$, a et b ne sont pas en relation à la fois par $r_u^{(0)}$ et par $r_u^{(1)}$.

On déduit de cette preuve un corollaire tout à fait analogue à celui énoncé dans le cas $N=2$.

Corollaire 2. *Soient u une N-écriture d'un mot w. Si v est le mot vérifiant $v(N)=u$ obtenu par la méthode décrite dans la seconde partie de la démonstration précédente, alors il existe un projection ϕ de $L_1(v)$ à valeurs dans $L_1(w)$ telle que $w=\phi(v)$. En particulier, l'alphabet de v est de cardinal supérieur ou égal au cardinal de l'alphabet de w.*

4. Application à l'étude des suites de complexité ultimement $n + c^{ste}$

Dorénavant nous ne nous intéresserons plus qu'à des mots infinis sur un alphabet que nous appellerons "suites".

Pour éviter d'alourdir inutilement nos notations, nous entendrons par suite de complexité $n + c^{ste}$: "suite dont la fonction de complexité est égale, pour tout entier $n \geqslant 1$, à $n + k$ où k est un entier naturel non nul fixé". Nous parlerons de suite de complexité ultimement $n + c^{ste}$ lorsque cette égalité n'est valable qu'à partir d'un certain rang.

Proposition 3. *Soit u une suite de complexité ultimement $n + c^{ste}$; alors il existe une suite v de complexité $n + c^{ste}$ et une projection ϕ de $L_1(v)$ vers $L_1(u)$ telles que $u = \phi(v)$. Réciproquement, soient v une suite de complexité $n + c^{ste}$ et ϕ une projection de $L_1(v)$ à valeurs dans un alphabet B. Si la suite $\phi(v)$ n'est pas ultimement périodique, elle est de complexité ultimement $n + c^{ste}$.*

Preuve. Soit u une suite et k et N deux entiers naturels non nuls tels que $p(u,n) = n+k$ pour tout entier $n \geqslant N$.

Considérons la N-écriture $u(N)$ de u. Il existe une projection ϕ telle que $u = \phi(u(N))$ et pour tout entier n non nul:

$$p(u(N),n) = p(u,n+N-1) = n+N-1+k.$$

Autrement dit, en posant $k' = N - 1 + k$, on a $p(u(N),n) = n + k'$ pour tout entier $n \geqslant 1$.

Réciproquement, soient v une suite et k' un entier supérieur ou égal à 1 tels que $p(v,n) = n + k'$ pour tout entier $n \geqslant 1$. Soient ϕ une projection de $L_1(v)$ vers un alphabet B et $\phi(v)$ l'image de v par ϕ.

On a $p(\phi(v),n) \leqslant p(v,n)$ pour tout entier $n \geqslant 1$. Comme $s(v,n) = 1$ pour tout entier $n \geqslant 1$, la différence première de la fonction de complexité de $\phi(v)$ ne peut prendre qu'un nombre fini de valeurs strictement supérieures à 1. La suite $\phi(v)$ est donc soit ultimement périodique, soit de complexité ultimement $n + c^{ste}$. □

Ainsi pour donner une description des suites de complexité ultimement $n + c^{ste}$, il nous suffira de nous intéresser aux suites de complexité $n + c^{ste}$ c'est-à-dire aux suites dont la différence première de la fonction de complexité est constante et égale à 1.

Pour commencer, nous introduisons trois transformations sur les suites conservant cette propriété. Nous montrerons ensuite que toute suite de complexité $n + c^{ste}$ peut s'obtenir à partir d'une suite sturmienne par combinaison de ces trois transformations.

4.1. La N-écriture

Nous avons vu précédemment que la fonction de complexité d'une suite u et celle de sa N-écriture $u(N)$ (où N est un entier naturel non nul) sont liées par la relation:

$$p(u(N),n) = p(u,n+N-1).$$

Il est clair que si u est tel que $s(u,n)=1$ pour tout entier $n\geqslant 1$, il en est de même pour $s(u(N),n)$.

4.2. L'ajout de nouvelles lettres en préfixe

Lemme 3. *Soit $u=u_0u_1\dots u_{k-1}u_k\dots$ telle que u_0, $u_1,\dots$ et u_{k-1} n'admettent qu'une seule occurrence dans u, Si l'on note u' la suite $u_ku_{k+1}\dots$, on a pour tout entier naturel non nul n:*

$$p(u',n)=p(u,n)-k \quad et \quad s(u',n)=s(u,n).$$

Preuve. L'ensemble des mots de longueur n de u' contient tous les mots de longueur n de u sauf les mots $u_0u_1\dots u_{n-1}$, $u_1u_2\dots u_n,\dots,u_ku_{k+1}\dots u_{k+n-1}$ qui sont tous distincts. □

4.3. Les I-morphismes

Définition 4. Soient A et B deux alphabets et q un entier naturel non-nul. On appelle I-morphisme de longueur q tout morphisme q-uniforme τ de A^* vers B^*, tel qu'il existe un entier naturel $k<q$ vérifiant, pour tout couple (x,y) d'éléments de A:
(i) $\tau(x)_k=\tau(y)_k\Rightarrow x=y$,
(ii) pour tout entier $i\neq k$, $\tau(x)_i=\tau(y)_i$,
(iii) pour tout couple d'entier (i,j), $\tau(x)_i=\tau(y)_j\Rightarrow i=j$.

Exemple. Le morphisme τ suivant est un I-morphisme de longueur 4 à valeur dans $\{0,1,2,a,b,c\}^*$:

$$\begin{array}{l} \tau \\ 0\rightarrow ab0c \\ 1\rightarrow ab1c \\ 2\rightarrow ab2c \end{array}$$

Le lemme suivant relie la complexité d'une suite à celle de son image par un I-morphisme.

Pour calculer la fonction de complexité de la suite image, on la décompose en somme de fonctions intermédiaires à la manière de [10]: on note $p_i(u,n)$ le nombre de mots de longueur n apparaissant à un rang congru à i modulo q dans u.

Lemme 4. *Soient A et B deux alphabets, u une suite sur A récurrente et τ un I-morphisme de longueur q de A^* vers B^*. Les fonctions de complexité de u et de $\tau(u)$*

sont liées par la relation:

$$p(\tau(u), nq+r) = (q-r)p(u,n) + rp(u,n+1), \quad \textit{pour tout entier } n \geqslant 1.$$

Ceci entraîne de plus l'égalité : $s(\tau(u), nq+r) = s(u,n)$.

Preuve. D'après la définition des I-morphismes, tout symbole de B ne peut apparaître à deux rangs distincts dans des images de lettres par τ (propriété (iii)).

Tout facteur de $\tau(u)$ apparaît donc à un unique rang i modulo q et l'on a:

$$p(\tau(u), n) = \sum_{i=0}^{q-1} p_i(\tau(u), n).$$

Pour tout entier naturel n et tout entier naturel r strictement inférieur à q, tout facteur v de longueur $nq+r$ de $\tau(u)$ apparaît à un unique rang i modulo q et s'écrit donc de façon unique sous la forme $S\tau(w)P$ où:
- w est un facteur de u de longueur égale à
  - n si $i=0$ ou $r+i \geqslant q$,
  - $n-1$ sinon;
- S est un suffixe strict d'un $\tau(x)$, $(x \in A)$, de longueur égale à:
  - 0 si $i=0$,
  - $q-i$ sinon;
- P est un préfixe strict d'un $\tau(y)$, $(y \in A)$, de longueur égale à:
  - $r+i$ si $r+i<q$,
  - $r+i-q$ sinon.

De plus, la restriction de τ à A étant injective, à un facteur v de $\tau(u)$ donné correspond un unique mot w, mais éventuellement plusieurs lettres x et y.

Par définition, il existe un entier k inférieur à q tel que pour tout couple de symboles (x,y) de A on ait: $\tau(x)_k = \tau(y)_k \Rightarrow x=y$ et pour tout $i \neq k$, $\tau(x)_i = \tau(y)_i$.

Autrement dit, tous les suffixes des $\tau(x)$ coïncident jusqu'à la longueur $q-k-1$ et les préfixes des $\tau(x)$ coïncident jusqu'à la longueur k.

Comme u est récurrente, $\tau(u)$ l'est également et on ne modifie pas sa fonction de complexité en lui ôtant un préfixe fini. On peut donc par la suite supposer que $k=0$ quitte à ôter un préfixe de longueur $k-1$ à $\tau(u)$. La suite ainsi obtenue est l'image de u par un nouvel I-morphisme de même longueur que τ où $k=0$.

La suite u étant récurrente, pour tout facteur w de u et tout couple d'entiers naturels l et l', il existe un facteur de $\tau(u)$ de la forme $S\tau(w)P$ tel que $|S|=l$ et $|P|=l'$ (si la suite u n'était pas récurrente, il existerait alors un préfixe w de u n'admettant qu'une seule occurrence et le facteur $\tau(w)$ n'apparaîtrait qu'une seule fois, au rang 0 de $\tau(u)$).

Soient l et m deux entiers naturels strictement inférieurs à q et n un entier naturel. Le nombre de facteurs de la forme $S\tau(w)P$ de $\tau(u)$ pour lesquels $|S|=l$, $|P|=m$ et $|w|=n$ est égal au nombre de facteurs de u de longueur:
- $|w|$ si $|P|=0$,
- $|w|+1$ sinon.

On calcule les valeurs des $p_i(\tau(u), nq+r)$ selon r et i:

- Si $r=0$:

si $i=0$,	$\lvert w\rvert = n$,	$\lvert P\rvert = 0$,	$p_0(\tau(u), nq+r) = p(u,n)$
si $0<i<q$,	$\lvert w\rvert = n-1$,	$\lvert P\rvert > 0$,	$p_i(\tau(u), nq+r) = p(u,n)$

- Si $q>r>0$:

si $i=0$,	$\lvert w\rvert = n$,	$\lvert P\rvert > 0$,	$p_0(\tau(u), nq+r) = p(u,n+1)$
si $0<i<q-r$,	$\lvert w\rvert = n-1$,	$\lvert P\rvert > 0$,	$p_i(\tau(u), nq+r) = p(u,n)$
si $i=q-r$,	$\lvert w\rvert = n$,	$\lvert P\rvert = 0$,	$p_i(\tau(u), nq+r) = p(u,n)$
si $q-r<i<q$,	$\lvert w\rvert = n$,	$\lvert P\rvert > 0$,	$p_i(\tau(u), nq+r) = p(u,n+1)$

Dans les deux cas, on obtient $p(\tau(u), nq+r) = (q-r)p(u,n) + rp(u,n+1)$. D'où l'on déduit que $s(\tau(u), nq+r) = s(u,n)$. $\square$

Lemme 5. *Soit une suite u sur un alphabet A telle que toute lettre x de A soit suffixe d'un facteur de longueur deux de u. Si $s(u,1)=1$, alors:*
- *soit il existe une suite v telle que $v(2)=u$ et $p(v,1)=p(u,1)-1$,*
- *soit u est l'image par un I-morphisme τ d'une suite w sur deux lettres.*

Preuve. Les notations sont les mêmes que dans la première partie.

Par hypothèse, il n'existe qu'une seule lettre de A admettant deux prolongements à droite et qu'une seule lettre admettant deux prolongements à gauche. Le nombre d'éléments de $\mathscr{R}_u^{(0)}$ (resp. de $\mathscr{R}_u^{(1)}$) est donc égal à $\#A-1$: il n'y a qu'une seule classe de $\mathscr{R}_u^{(0)}$ (resp. de $\mathscr{R}_u^{(1)}$) qui contient deux symboles de A, toutes les autres sont des singletons.

Deux cas se présentent:
- Si l'on est dans les conditions de la proposition 1, c'est à dire si les intersections des classes de $\mathscr{R}_u^{(0)}$ et de $\mathscr{R}_u^{(1)}$ contiennent au plus un symbole, on utilise la construction décrite dans la seconde partie de la démonstration. On obtient ainsi une suite v telle que $v(2)=u$ sur un alphabet X de cardinal égal aux nombre de classes de $\mathscr{R}_u^{(0)}$ (et de $\mathscr{R}_u^{(1)}$), ici $\#A-1$. On a ainsi déterminé une suite v sur un alphabet contenant un élément de moins que l'alphabet A, et de complexité $p(v,n)=p(u,n-1)$ pour tout entier $n>1$.
- Sinon, la classe de $\mathscr{R}_u^{(0)}$ contenant deux éléments et la classe de $\mathscr{R}_u^{(1)}$ contenant deux éléments coïncident, et il existe quatre lettres de A notées x, y, c_r, c_0 telles que c_rx, c_ry, xc_0 et yc_0 apparaissent dans u.

  Si $c_r=x$ (resp. $c_r=y$), alors $c_0=x$ (resp. $c_0=y$) et il ne peut pas y avoir de lettre c différente de x et de y telle que cx ou cy apparaisse dans u: cela impliquerait soit qu'il y a une lettre qui n'apparaît pas comme suffixe d'un mot de longueur deux, soit que $s(u,1)$ est strictement supérieur à 1. Dans ce cas, l'alphabet de u ne contient que deux lettres et le morphisme τ est l'identité.

Si $c_r \neq x$ et $c_r \neq y$ et donc $c_0 \neq x$ et $c_0 \neq y$, c'est à dire $\{c_0, c_1 \ldots, c_r\} = A_{-\{x,y\}}$, alors les seuls mots de longueur deux apparaissant dans u sont (quitte à renuméroter $c_1, \ldots, c_{r-1}$):

xc_0, yc_0, c_0c_1, $\ldots c_{r-1}c_r$, c_rx et c_ry.

Soit a la première lettre de u, on définit le morphisme τ de $\{0,1\}$ vers A^* par:

- $\tau(0) = xc_0c_1 \ldots c_r$ et $\tau(1) = yc_0c_1 \ldots c_r$, si $a = x$ ou y,
- $\tau(0) = c_ic_{i+1} \ldots c_rxc_0 \ldots c_{i-1}$ et $\tau(1) = c_ic_{i+1} \ldots c_ryc_0 \ldots c_{i-1}$, si $a = c_i$.

Comme $c_i = c_j$ implique $i = j$, τ est bien un I-morphisme et la liste des mots de longueur deux de u montre que l'on peut construire une suite v sur $\{0,1\}$ telle que $u = \tau(v)$. □

Le lemme suivant montre que toute suite de complexité $n + c^{ste}$ est récurrente à un préfixe de longueur finie près.

Lemme 6. *Soit u une suite de complexité $n + c^{ste}$. Si la première lettre de u admet au moins deux occurrences, alors u est récurrente.*

Preuve. On se convaincra que u est récurrente si et seulement si tout préfixe de u admet au moins deux occurrences dans u.

Supposons que u ne soit pas récurrente. Soit alors N le plus petit entier tel que $u_0 \ldots u_N$ n'admette qu'une seule occurrence dans u. Par hypothèse N est supérieur ou égal à 1.

Notons w le mot $u_0 \ldots u_{N-1}$. Comme le facteur w admet au moins deux occurrences, il apparaît comme préfixe de deux mots de longueur $N + 1$ distincts: wu_N et wx où x est une lettre de $L_1(u)$ différente de u_N.

La suite u étant de complexité $n + c^{ste}$, w est le seul facteur de longueur N de u admettant deux continuations à droite, c'est à dire apparaissant comme préfixe d'exactement deux facteurs de longueur $N + 1$ distincts, tous les autres facteurs de longueur N de u n'admettant qu'une seule continuation.

La complexité de u implique également l'existence d'un unique facteur de longueur $N + 1$ admettant deux continuations à droite. Celui-ci est alors nécessairement de la forme yw où y est une lettre de $L_1(u)$ et les mots ywu_N et ywx sont facteurs de u. Ceci est en contradiction avec le fait que wu_N, préfixe de u, n'admet qu'une seule occurrence. □

Proposition 4. *Une suite u est de complexité $n + c^{ste}$ si et seulement si elle est de la forme $ov(N)$ où o est un préfixe de u éventuellement vide tel que tout symbole de o n'admette qu'une seule occurrence dans u, et $v(N)$ est la N-écriture d'une suite v image par un I-morphisme d'une suite sturmienne.*

Preuve. Supposons que u soit de complexité $p(u,n) = n + k$ pour tout entier $n \geqslant 1$.

Soit i le plus petit entier naturel tel que le symbole u_i admette au moins deux occurrences dans u. D'après les lemmes 6 et 3, la suite $w = u_iu_{i+1} \ldots$ est récurrente et de complexité $n+k-i$. De plus, tout symbole de $L_1(w)$ est suffixe d'un mot de longueur deux de w.

Notons $w^{(0)}$ la suite w. Elle satisfait bien les hypothèses du lemme 5 et deux cas se présentent:

- Soit $w^{(0)}$ s'écrit $w^{(1)}(2)$ avec $p(w^{(1)},n)=n+k-i-1$ pour tout entier $n \geqslant 1$. Tout symbole de $L_1(w(1))$ apparaît comme suffixe d'un mot de longueur deux de $w^{(1)}$. On applique à nouveau le lemme 5. Comme l'alphabet de départ est fini, on tombe au bout d'un nombre fini d'itérations dans le second cas.
- Soit $w^{(i)}$ s'écrit $\tau(v)$ avec $\#L_1(v)=2$ et τ est un I-morphisme; nécessairement $w^{(i)}$ est récurrente. Ceci implique d'après le lemme 4 que $s(v,n)=s(w^{(i)},nq+r)=1$ pour tout entier naturel non nul n. Autrement dit v est sturmienne.

Inversement, si w est sturmienne, son image par un I-morphisme τ est de complexité $n+c^{ste}$ (lemme 4). Pour tout entier naturel non nul N, la N-écriture de la suite $v=\tau(w)$ est de complexité $n+c^{ste}$. D'après le lemme 3, si o est un mot fini tel que toute lettre de o n'admette qu'une occurrence dans $ov(N)$ alors $ov(N)$ est de complexité $n+c^{ste}$. $\square$

Références

[1] P. Alessandri, Codages de rotations et basses complexités, Thèse de Doctorat, Université d'Aix-Marseille II, 1996.
[2] J.-P. Allouche, Sur la complexité des suites infinies, Bull. Belg. Math. Soc. 1 (1994) 133–143.
[3] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n+1$, Bull. Soc. Math. France 119 (1991) 199–215.
[4] J. Berstel, Recent results in sturmian words, DLT'95.
[5] J. Cassaigne, Complexité et facteurs spéciaux, Bull. Belg. Soc. Math. 4 (1997) 67–88.
[6] S. Ferenczi, Systems of finite rank, Colloquium Mathematicum 73 (1997) 35–65.
[7] S. Ferenczi, C. Mauduit, Transcendence of numbers with a low complexity expansion, preprint de l'I.M.L., 1995.
[8] G.A. Hedlund, M. Morse, Symbolic dynamics, Amer. J. Math. 60 (1938) 815–866.
[9] G.A. Hedlund, M. Morse, Symbolic dynamics II. Sturmian trajectories, Amer. J. Math. 62 (1940) 287–306.
[10] B. Mossé, Notions de reconnaissabilité pour les substitutions et complexité de suites automatiques, Bull. Soc. Math. France 124 (1996) 101–108.
[11] M. Queffélec, Substitution Dynamical System – Spectral Analysis, Lecture Notes in Mathematics, vol. 1294, Springer, Berlin, 1987.