RAIRO-Theor. Inf. Appl. **41** (2007) 425–446
DOI: 10.1051/ita:2007027

COMBINATOIRE DE MOTS RÉCURRENTS DE COMPLEXITÉ $n+2$

Idrissa Kaboré[1] et Théodore Tapsoba[2]

Abstract. ***Combinatorics on recurrent words with subword complexity n+2.*** We state some new properties on Sturmian words and classify words which have, for any nonnegative integer n, exactly $n+2$ subwords of length n. We also define the notion of k by k insertion on infinite words and we give a formula for the complexity function of words obtained by applying that notion to Sturmian words. Lastly we study balance property and palindrome complexity of a subclass of words with complexity $n+2$ called quasi-Sturmian words by insertion; we give a characterization of this subclass with Parikh vectors.

Résumé. Nous établissons quelques propriétés des mots sturmiens et classifions, ensuite, les mots infinis qui possèdent, pour tout entier naturel non nul n, exactement $n+2$ facteurs de longueur n. Nous définissons également la notion d'insertion k à k sur les mots infinis puis nous calculons la complexité des mots obtenus en appliquant cette notion aux mots sturmiens. Enfin nous étudions l'équilibre et la palindromie d'une classe particulière de mots de complexité $n+2$ que nous appelons mots quasi-sturmiens par insertion et que nous caractérisons à l'aide des vecteurs de Parikh.

Classification Mathématique. 68R15.

1. Introduction

La fonction de complexité p, qui calcule le nombre de facteurs de longueur donnée dans un mot, est souvent utilisée pour caractériser certaines familles de mots [3]. Ainsi on sait, depuis les résultats précurseurs de Morse et Hedlund [17,18],

Mots Clés. Mot sturmien, complexité, mot quasi-sturmien par insertion.
Sturmian word, complexity, quasi-Sturmian word by insertion.

[1] Institut Sciences Exactes et Appliquées, Univ. polytech. de Bobo-Dioulasso, 01 BP 1091 Bobo-Dioulasso 01, Burkina Faso ; `ikaborei@yahoo.fr`

[2] École Supérieure d'Informatique, Univ. polytech. de Bobo-Dioulasso, 01 BP 1091 Bobo-Dioulasso 01, Burkina Faso ; `theo_tapsoba@univ-ouaga.bf`

qu'un mot dont la fonction de complexité vérifie $p(n) \leq n$ pour un certain entier n est ultimement périodique, et que les mots sturmiens sont les mots de complexité minimale parmi les mots infinis non ultimement périodiques. Depuis ces travaux, de nombreuses études comme par exemple [5,8,9,11,12,16,21] ont été menées sur les mots sturmiens et ont conduit à de nombreuses généralisations parmi lesquelles les mots quasi-sturmiens. Un mot infini est dit quasi-sturmien s'il existe des entiers n_0 et k tels que $p(n) = n + k$ pour tout $n \geq n_0$. Dans les années 1975 on étudiait déjà ces mots [8, 9, 19], mais c'est au milieu des années 1990 que l'expression de « mot quasi-sturmien » a été utilisée pour les désigner du fait qu'ils ont des propriétés très proches des mots sturmiens. En effet, les mots quasi-sturmiens ont une combinatoire assez proche des mots sturmiens et leurs caractérisations font intervenir ces derniers [1,2,7,8,10,13,14].

Les palindromes et la fonction de complexité palindromique interviennent dans l'étude combinatoire des mots infinis. Des résultats intéressants ont été obtenus pour certaines classes de mots (voir par exemple [4]) et en particulier une caractérisation des mots sturmiens [11].

Dans ce travail, après avoir donné quelques notations et définitions utiles (sect. 3) nous établissons des propriétés combinatoires des mots sturmiens (sect. 4) puis nous classifions tous les mots quasi-sturmiens de complexité $n + 2$ (sect. 5). Ensuite nous introduisons, dans la section 6, la notion « d'insertion k à k » sur les mots infinis puis nous calculons la complexité des mots obtenus en appliquant l'insertion k à k aux mots sturmiens. Enfin, à la section 7, nous étudions l'équilibre et la palindromie d'une classe particulière des mots de complexité $n + 2$ que nous appelons mots quasi-sturmiens par insertion et que nous caractérisons à l'aide des vecteurs de Parikh.

2. Préliminaires

La plupart des notations utilisées ici peuvent être retrouvées dans le livre de Lothaire [15].

Soit A un alphabet fixé. A^*, l'ensemble des mots finis sur A, est le monoïde libre engendré par A, ε le mot vide étant l'élément neutre. A^+ est l'ensemble des mots finis non vides. Pour tout $u \in A^*$, $|u|$ désigne la longueur du mot u ($|\varepsilon| = 0$) et pour toute lettre x de A, $|u|_x$ est le nombre d'occurrences de x dans u. Un mot u de longueur n formé d'une seule lettre x est simplement noté $u = x^n$; par extension $x^0 = \varepsilon$. Soit $u = u_1 u_2 \cdots u_n$ un mot tel que $u_i \in A$ pour tout $i \in \{1, 2, \ldots, n\}$. Le mot miroir de u que l'on note $\overline{u}$ est le mot obtenu en lisant u de la droite vers la gauche, c'est-à-dire $\overline{u} = u_n u_{n-1} \cdots u_1$. Si u et v sont deux mots de A^* on a $\overline{uv} = \overline{v}\,\overline{u}$. On dit qu'un mot u est un palindrome s'il est identique à son image miroir, c'est-à-dire $\overline{u} = u$. Par exemple les mots « elle », « été », « ici » et « tôt » sont des palindromes en français.

Un mot infini est une suite de lettres de A indexée par $\mathbb{N}$. On désigne par A^ω l'ensemble des mots infinis sur A et on pose $A^\infty = A^* \cup A^\omega$. Pour tout mot $u \in A^\infty$, on note $alph(u)$ l'ensemble des lettres de A présentes dans u.

Un mot infini u est dit ultimement périodique s'il existe un entier non nul τ et un entier n_0 tels que $u_{i+\tau} = u_i$ pour tout $i \geq n_0$. Ainsi un mot infini u est ultimement périodique s'il existe deux mots $v \in A^*$ et $w \in A^+$ tels que $u = vw^\omega$ où $w^\omega = www\cdots$ est une concaténation infinie de w.

On définit la puissance n-ième d'un mot fini w comme étant la concaténation n fois de w ; on la note w^n.

Soient $u \in A^\infty$ et $v \in A^*$. Le mot v est un facteur de u s'il existe $u_1 \in A^*$ et $u_2 \in A^\infty$ tels que $u = u_1 v u_2$; on dit aussi que u contient v. Le facteur v est dit préfixe (resp. suffixe) si u_1 (resp. u_2) est le mot vide.

Soient $u \in A^\omega$, w un facteur de u et x une lettre de A. Le langage de longueur n de u, noté $L_n(u)$, est l'ensemble des facteurs de u de longueur n. L'ensemble de tous les facteurs de u est simplement noté $L(u)$. La lettre x est un prolongement à gauche (resp. à droite) de w si xw (resp. wx) appartient à $L(u)$. On note $\partial^- w$ (resp. $\partial^+ w$) le nombre de prolongements à gauche (resp. à droite) de w. On dira qu'un facteur w est biprolongeable (resp. triprolongeable) à droite si $\partial^+ w = 2$ (resp. $\partial^+ w = 3$). De la même manière on définit la notion de facteur biprolongeable ou triprolongeable à gauche. Un facteur est dit spécial à gauche (resp. à droite) s'il admet plusieurs prolongements à gauche (resp. à droite). Un facteur à la fois spécial à gauche et à droite est dit bispécial.

Un mot u est dit récurrent si tout facteur de u apparaît une infinité de fois dans u. Un mot est dit uniformément récurrent ou minimal si tout facteur apparaît avec des lacunes bornées, c'est à dire que pour tout $n \in \mathbb{N}$, il existe N tel que tout facteur de longueur N contient tous les facteurs de longueur n.

On appelle décalage l'application S de A^ω dans A^ω qui consiste à effacer la première lettre.

La fonction de complexité de u est l'application de $\mathbb{N}$ dans $\mathbb{N}^*$ définie par $p_u(n) = \#L_n(u)$, où $\#L_n(u)$ désigne le cardinal de $L_n(u)$. Dans tout ce qui suit la fonction de complexité d'un mot u sera simplement désignée par p. La fonction de complexité est liée aux facteurs spéciaux par la relation [6]

$$p(n+1) - p(n) = \sum_{n \in L_n(u)} \left(\partial^+(w) - 1\right).$$

Ainsi, le nombre de facteurs spéciaux à droite est au plus $p(n+1) - p(n)$. Il y a un unique facteur spécial de longueur n, qui est biprolongeable si et seulement si $p(n+1) - p(n) = 1$.

La fonction de complexité palindromique de u est l'application de $\mathbb{N}$ dans $\mathbb{N}$ qui compte le nombre de palindromes de longueur n contenus dans u, c'est-à-dire $pal_u(n) = \#\{v \in L_n(u) : \overline{v} = v\}$.

Un morphisme f est une application de A^* dans lui-même telle que $f(uv) = f(u)f(v)$ pour tous u, $v \in A^*$.

On dit qu'un mot infini u est engendré par un morphisme f s'il existe une lettre a telle que les mots a, $f(a)$, $f^2(a)$, $\cdots$, $f^n(a)$, $\cdots$ sont des préfixes de plus en plus longs de u. On note alors $u = f^\omega(a)$.

3. Mots sturmiens

Dans cette section, l'alphabet est $A = \{a, b\}$.

Définition 3.1. Un mot infini u sur A est un mot sturmien si pour tout entier n, $p(n) = n + 1$.

Exemple. Le mot sturmien le plus connu est le célèbre mot de Fibonacci engendré par le morphisme Φ défini par $\Phi(a) = ab$ et $\Phi(b) = a$.

Remarque 3.2. (i) Les mots ab et ba sont facteurs de tout mot sturmien.
(ii) Tout mot sturmien a en facteur un et un seul des deux mots a^2 et b^2.

Définition 3.3. Soit u un mot sturmien. On dit que u est a-sturmien (resp. b-sturmien) lorsqu'il contient a^2 (resp. b^2).

Définition 3.4. Soit u un mot infini sur A. On dit que u est équilibré si pour tout entier n et tous $v, w \in L_n(u)$, $\Big| |v|_x - |w|_x \Big| \leq 1$ pour tout $x \in A$.

Théorème 3.5. [9] *Un mot infini u sur A est non équilibré si et seulement s'il existe un unique mot t de longueur minimale tel que u contient ata et btb. Le mot t est alors un palindrome.*

La caractérisation suivante est fort utile.

Théorème 3.6. [9] *Un mot infini u est sturmien si et seulement s'il est non ultimement périodique et équilibré.*

Lemme 3.7. *Soit u un mot a-sturmien. Alors il existe une suite sturmienne $(\epsilon_i)_{i\geq 1}$ sur l'alphabet $\{0, 1\}$ et un entier naturel non nul n tel que u s'écrit*

$$u = a^{n_0} b a^{n+\epsilon_1} b a^{n+\epsilon_2} b a^{n+\epsilon_3} b \cdots$$

avec $n_0 \leq n + 1$.

Preuve. Soit u un mot a-sturmien. Considérons alors n l'entier minimal tel que u contient $ba^n b$. Ainsi, u ne contient pas a^{n+2} par le théorème 3.5. Donc u est de la forme

$$u = a^{n_0} b a^{n+\epsilon_1} b a^{n+\epsilon_2} b a^{n+\epsilon_3} b \cdots$$

où $n_0 \leq n+1$ et $(\epsilon_i)_{i\geq 1}$ une suite sur l'alphabet $\{0, 1\}$ [16]. Il reste donc à prouver que la suite $(\epsilon_i)_{i\geq 1}$ est sturmienne.
Supposons $(\epsilon_i)_{i\geq 1}$ non sturmienne. Alors $(\epsilon_i)_{i\geq 1}$ est ultimement périodique ou non équilibrée.

Si $(\epsilon_i)_{i\geq 1}$ était ultimement périodique alors u le serait aussi, ce qui est impossible car u est sturmien. Par suite, $(\epsilon_i)_{i\geq 1}$ est non équilibrée, et possède de ce fait deux facteurs de la forme $0t0$ et $1t1$. Considérons les facteurs de $(\epsilon_i)_{i\geq 1}$; en les écrivant de telle sorte que chaque lettre précède et succède un b puis en remplaçant 0 par a^n et 1 par a^{n+1} on retrouve certains facteurs de u. Ainsi u admet deux facteurs de la forme $ba^n T a^n b$ et $ba^{n+1} T a^{n+1} b$. En posant $w = a^n T a^n$ on remarque que les mots awa et bwb sont dans u. Donc u est non équilibré. On obtient une contradiction, car u est un mot sturmien. Par conséquent $(\epsilon_i)_{i\geq 1}$ est une suite sturmienne. □

Soient $\alpha \in]0,\, 1[$ et R la rotation d'angle α du cercle unité (identifié à l'intervalle $[0,\, 1[$). On a $R(\rho) = (\rho+\alpha) \bmod 1 = \{\rho+\alpha\}$, partie fractionnaire de $\rho+\alpha$. L'orbite d'un point ρ du cercle unité par la rotation d'angle α est l'ensemble des points $\{\{\rho+n\alpha\}\,,\ n\geq 0\}$. Le codage de l'orbite de ρ sous la rotation R d'angle α sur le cercle unité partitionné en deux intervalles complémentaires $I_0 = [0,\, 1-\alpha[$ et $I_1 = [1-\alpha,\, 1[$ (ou $I_0 =]0,\, 1-\alpha]$ et $I_1 =]1-\alpha,\, 1[\cup\{0\}$) est le mot $u_0u_1\cdots u_n\cdots$ défini sur l'alphabet $\{0,\, 1\}$ par : $u_n = \begin{cases} 1 & \text{si } \{\rho+n\alpha\}\in I_1 \\ 0 & \text{sinon} \end{cases}$.

L'orbite d'un point ρ du cercle unité par une rotation d'angle irrationnel est dense sur le cercle unité. Pour plus de détails sur les codages de rotations voir par exemple [1,2] et chapitre 6 section 6.1.2 de [20]. Rappelons la caractérisation classique des mots sturmiens due à Morse et Hedlund [17,18].

Théorème 3.8. *Un mot u est sturmien si et seulement s'il existe un nombre irrationnel $\alpha \in]0,\, 1[$ et un nombre réel ρ tels que u est le codage de l'orbite de ρ sous la rotation R d'angle α.*

Soit u le mot codant l'orbite d'un point ρ sous la rotation d'angle irrationnel α. Un mot fini $w = w_1w_2\cdots w_n$ sur l'alphabet $\{0,\, 1\}$ est un facteur de u si et seulement s'il existe un entier naturel k tel que :

$$R^k(\rho)\in I_w = \bigcap_{j=0}^{n-1} R^{-j}\left(I_{w_{j+1}}\right) \quad \text{où} \quad I_{w_{j+1}} = \begin{cases} I_0 & \text{si } w_{j+1}=0 \\ I_1 & \text{sinon} \end{cases}$$

([2,15], chap. 2 sect. 2.1).

La caractérisation ci-dessus nous permet d'établir le résultat suivant.

Théorème 3.9. *Soient $u = u_0u_1u_2\cdots$ un mot sturmien et $k\geq 1$ un entier. Tout facteur w de u, de longueur n quelconque, apparaît dans u à toutes les positions modulo k, c'est-à-dire*

$$\forall i\in\{0,\,\cdots,\,k-1\}\ \exists l_i\in\mathbb{N}\ : w = u_{kl_i+i}u_{kl_i+i+1}\cdots u_{kl_i+i+n-1}.$$

Preuve. Soient u un mot sturmien et $k\geq 1$ un entier. D'après le théorème 3.8, il existe un nombre réel ρ et un nombre irrationnel $\alpha\in]0,\, 1[$ tel que u soit le codage de l'orbite de ρ par la rotation d'angle α sur le cercle unité muni de la partition $\mathcal{P} = ([0,\, 1-\alpha[,\, [1-\alpha,\, 1[)$ ou $\mathcal{P} = (]0,\, 1-\alpha],\,]1-\alpha,\, 1[\cup\{0\})$. Considérons w un facteur de u de longueur n non nulle. Alors I_w est un intervalle de longueur non nulle. Pour tout $i\in\{0,\ \cdots,\ k-1\}$ posons $\rho' = \rho+i\alpha$ et $\alpha' = k\alpha$. Il vient que α' est irrationnel comme α. Par suite ρ' est d'orbite dense dans $[0,\, 1[$ par la rotation d'angle irrationnel α'. Par conséquent, il existe $l_i\geq 0$ tel que $\rho'+l_i\alpha' \mod 1\in I_w$. Autrement dit, il existe $l_i\geq 0$ tel que $\rho+(kl_i+i)\,\alpha \mod 1\in I_w$ et donc w apparaît dans u à la position kl_i+i. □

En vertu de ce résultat on fait le constat suivant :

Remarque 3.10. Soient u un mot sturmien et $k > 0$ un entier. Pour tout $i \in \{0, 1, \cdots, k-1\}$ et tout entier $n \geq 1$, on a :

$$\#\{u_{kl+i}u_{kl+i+1}\cdots u_{kl+i+n-1} : l \geq 0\} = n+1.$$

4. Une sous-classe de mots de complexité $n+2$

Dans cette section l'alphabet est $A = \{a, b, c\}$.

Théorème 4.1. *Soient $a^{n_0}ba^{n+\epsilon_1}ba^{n+\epsilon_2}ba^{n+\epsilon_3}b\cdots$ un mot a-sturmien sur $\{a, b\}$ et soit v le mot sur $\{a, b, c\}$ défini par $v = bx_1bx_2bx_3\cdots$ où pour tout entier non nul i, $x_i = a$ si $\epsilon_i = 0$ et $x_i = c$ si $\epsilon_i = 1$. Alors v est de complexité $n+2$ pour tout $n \geq 1$.*

Preuve. La suite $(\epsilon_i)_{i\geq 1}$ est sturmienne d'après le lemme 3.7. Par conséquent le mot $\boldsymbol{x} = x_1x_2x_3\cdots$ est sturmien sur $\{a, c\}$ car a et c s'identifient respectivement à 0 et 1 dans la suite $(\epsilon_i)_{i\geq 1}$.
Comme v ne contient pas de carré de lettre alors aucun facteur non vide de v n'est triprolongeable à droite. Il nous suffit donc de montrer que v admet un et un seul facteur biprolongeable à droite de longueur n, pour tout $n \geq 1$.
Pour $n = 1$, b est le seul facteur biprolongeable à droite. Supposons la propriété vérifiée jusqu'au rang n $(n \geq 1)$. Comme $\boldsymbol{x}$ est sturmien alors v est récurrent et non ultimement périodique. Par conséquent, v admet au moins un facteur biprolongeable à droite de chaque longueur.

Supposons que v admette deux facteurs biprolongeables à droite de longueur $n+1$. Soient D_1 et D_2 lesdits facteurs. Leurs suffixes sont aussi des facteurs biprolongeables à droite. D'après l'hypothèse de récurrence, v n'admet qu'un unique facteur biprolongeable à droite D de longueur n. Par suite, D se prolonge à gauche par deux lettres distinctes pour donner D_1 et D_2. Comme les mots aa, bb, cc, ac et ca ne sont pas facteurs de v, D commence par b et on peut écrire $D_1 = aD$ et $D_2 = cD$. De même D se termine par b et les mots

$$aDa,\ aDc,\ cDa \text{ et } cDc$$

sont dans $L_{n+2}(v)$. En effaçant les lettres b dans ces quatre facteurs de v, on obtient quatre facteurs de $\boldsymbol{x}$ de la forme

$$ata,\ atc,\ cta \text{ et } ctc.$$

Par suite $\boldsymbol{x}$ n'est pas équilibré puisqu'il contient deux facteurs de la forme ata et ctc. On obtient une contradiction, car $\boldsymbol{x}$ est sturmien. Ainsi, v admet un unique facteur biprolongeable à droite de longueur $n+1$.
Finalement, pour tout $n \geq 1$, v admet un unique facteur biprolongeable à droite de longueur n. D'où $p(n+1) = p(n)+1$ et comme $p(1) = 3$ il en résulte que $p(n) = n+2$ pour tout $n \geq 1$. $\square$

Corollaire 4.2. *Soit v un mot infini sur A de la forme*

$$v = x_0 z x_1 z x_2 z x_3 \cdots$$

où $x_0 \in \{\varepsilon, x, y\}$ et $x_i \in \{x, y\}$ pour $i = 1, 2, 3, \dots$ avec $\{x, y, z\} = \{a, b, c\}$. Le mot v est de complexité $n+2$ $(n \geq 1)$ si et seulement si le mot extrait

$$v' = x_0 x_1 x_2 x_3 \cdots$$

est un mot sturmien sur $\{x, y\}$.

Preuve. D'après la preuve du théorème 4.1 la condition est suffisante. Montrons qu'elle est nécessaire. Pour ce faire supposons $v' = x_0 x_1 x_2 x_3 \cdots$ non sturmien. Alors v' est ultimement périodique ou non équilibré. Si v' était ultimement périodique, alors v, qui est image morphique de v', le serait aussi, ce qui est exclu puisque sa complexité n'est pas bornée. Ainsi, v est non ultimement périodique. Donc v' est non équilibré et possède de ce fait deux facteurs spéciaux à droite de longueur n (pour un $n > 1$) de la forme

$$xt = xt_1 t_2 \cdots t_{n-1} \text{ et } yt = yt_1 t_2 \cdots t_{n-1}.$$

Par conséquent les mots

$$xtx,\ xty,\ ytx \text{ et } yty$$

sont des facteurs de v'. Par suite les mots

$$xrx,\ xry,\ yrx \text{ et } yry \text{ où } r = zt_1 zt_2 z \cdots zt_{n-1} z$$

sont des facteurs de même longueur de v. Ainsi xr et yr sont deux facteurs spéciaux à droite de v de même longueur. On obtient une contradiction, car v étant de complexité $n+2$ $(n \geq 1)$, il admet un et un seul facteur biprolongeable à droite pour toute longueur n non nulle donnée. Donc v' est sturmien. □

Définition 4.3. Soit u un mot récurrent sur A. On dira que u est un mot quasi-sturmien par insertion (la terminologie sera justifiée à la remarque 6.3) s'il est de complexité $n+2$ et de la forme

$$v = x_0 z x_1 z x_2 z x_3 \cdots$$

où $x_0 \in \{\varepsilon, x, y\}$ et $x_i \in \{x, y\}$ pour $i = 1, 2, 3, \dots$ avec $\{x, y, z\} = \{a, b, c\}$.

Exemple. Considérons le mot de Fibonacci

$$F = abaababaabaababaababaabaababa \cdots .$$

Le mot quasi-sturmien par insertion issu de F et commençant par a est

$$Q = acbcacacbcacbcacacbcacacbcacbcacacbcacbcacacbcacacbcacbca \cdots .$$

Remarque 4.4. Soit u un mot quasi-sturmien par insertion.
(i) u ne comporte aucun carré de lettre, c'est-à-dire toute lettre de $alph(u)$ est isolée dans u.
(ii) Une et une seule des lettres présentes dans u est biprolongeable. Elle est biprolongeable à la fois à gauche et à droite.
(iii) $\forall n \in \mathbb{N}^*,\ \exists!\,(D_n,\, G_n) \in L_n^2(u)\ : \partial^+ D_n = \partial^- G_n = 2$.

Dans toute la suite on dira qu'un facteur est bispécial lorsqu'il est biprolongeable à la fois à gauche et à droite dans le mot donné.

Définition 4.5. Un mot u, quasi-sturmien par insertion est dit a-spécial (resp. b-spécial, c-spécial) si la lettre a (resp. b, c) est biprolongeable dans u.

Par unicité de la lettre biprolongeable tout mot u quasi-sturmien par insertion est soit a-spécial, soit b-spécial, soit c-spécial.

Définition 4.6. Soit u un mot quasi-sturmien par insertion. On appelle mot extrait de u et on note u' le mot sturmien obtenu de u en effaçant la lettre biprolongeable.

5. Classification des mots de complexité $n+2$

La classification des mots de complexité $n+2$ a été établie par Alessandri [1]. Nous en donnons une formulation légèrement différente qui permet de construire de façon explicite tous les mots de complexité $n+2$ en nous appuyant essentiellement sur la forme des mots sturmiens décrite dans le lemme 3.7.

Lemme 5.1. *Soit u un mot de l'une des formes suivantes, à une permutation de lettres près :*
(1) cv avec v sturmien sur $\{a,\, b\}$;
(2) $S^{k_0}\big(a^{k+\epsilon_1}bca^{k+\epsilon_2}bca^{k+\epsilon_3}bc\cdots\big)$;
(3) $S^{k_0}\big((ab)^{k+\epsilon_1}\,c\,(ab)^{k+\epsilon_2}\,c\,(ab)^{k+\epsilon_3}\,c\cdots\big)$
où S est le décalage, $k_0 \geq 0$, $k > 0$ et $(\epsilon_i)_{i\geq 1}$, une suite sturmienne sur $\{0,\, 1\}$. Alors, u est de complexité $n+2$.

Preuve. (1) Pour tout $n \geq 1$, tout préfixe de longueur n n'apparaît qu'une seule fois dans u. Donc $p_u(n) = 1 + p_v(n) = 1 + (n+1)$ puisque v est sturmien. D'où, $p_u(u) = n+2$.

(2) u n'est pas ultimement périodique car il est un décalé de l'image du mot sturmien
$$v = a^{k+\epsilon_1}ba^{k+\epsilon_2}ba^{k+\epsilon_3}b\cdots$$
par le morphisme non périodique : $(a \longmapsto a,\ b \longmapsto bc)$. Donc u possède au moins un facteur spécial pour toute longueur donnée n. Nous observons que a est la seule lettre spéciale dans u : la lettre a se prolonge par a et c, à gauche et par a et b, à droite. Donc tout facteur spécial de u est biprolongeable. Supposons que u possède deux facteurs T_1 et T_2 spéciaux à droite pour une certaine longueur n. On peut prendre n minimal et nous aurons ($T_1 = aT$ et $T_2 = bT$) ou ($T_1 = bT$

et $T_2 = cT$) ou ($T_1 = aT$ et $T_2 = cT$) avec $T \in A^+$. Si bT est dans u, alors T commence nécessairement par c, puisque b ne précède que c dans u. Par suite, comme a ne précède pas c dans u, le premier cas ne peut se produire. Il en est de même pour le second cas puisque cc n'est pas dans u. Ainsi, le seul cas possible est $T_1 = aT$ et $T_2 = cT$. Il en résulte que aTa, $cTb \in L(u)$ et donc aTa, $bcTb \in L(u)$. En effaçant c dans aTa et $bcTb$, on retrouve deux facteurs ata et btb du mot v. Donc v n'est pas équilibré. On obtient une contradiction, car v est sturmien. Donc $p_u(n+1) - p_u(n) = 1$, pour tout $n \geq 1$ et comme $p_u(1) = 3$ on déduit que $p_u(n) = n+2$.

La forme (3) se montre de la même manière qu'en (2). □

Théorème 5.2. *Soit u un mot infini sur l'alphabet $\{a, b, c\}$. Le mot u est de complexité $n+2$ pour $n \geq 1$ si et seulement s'il a l'une des formes suivantes à une permutation de lettres près :*

(1) cv avec v sturmien sur $\{a, b\}$;

(2) $S^{k_0}\left(a^{k+\epsilon_1}bca^{k+\epsilon_2}bca^{k+\epsilon_3}bc\cdots\right)$;

(3) $S^{k_0}\left(\left(ab\right)^{k+\epsilon_1}ac\left(ab\right)^{k+\epsilon_2}ac\left(ab\right)^{k+\epsilon_3}ac\cdots\right)$;

(4) $S^{k_0}\left(\left(ab\right)^{k+\epsilon_1}c\left(ab\right)^{k+\epsilon_2}c\left(ab\right)^{k+\epsilon_3}c\cdots\right)$

où S est le décalage, $k_0 \geq 0$, $k > 0$ et $(\epsilon_i)_{i\geq 1}$, une suite sturmienne sur $\{0, 1\}$.

Preuve. Nécessité : Soit u un mot de complexité $n+2$ sur $\{a, b, c\}$. Alors, u est soit récurrent, soit non récurrent.

– Supposons u non récurrent. Il existe n_0 non nul tel que le préfixe de longueur n_0 n'apparaisse qu'une seule fois dans u. En prenant $v = T(u)$, nous avons $p_v(n) = p_u(n) - 1 = n+1$ pour tout $n \geq n_0$. S'il existe un entier non nul m tel que $p_v(m) \geq m+2$ alors pour tout n, $p_v(n+m) \geq n+m+2$ puisque la fonction de complexité est strictement croissante, contradiction. Donc, $p_v(n) = n+1$ pour tout n et v est sturmien. Ainsi, v est binaire et s'écrit par exemple avec a et b. Par suite, $u = cv$ puisque u est ternaire.

– Supposons u récurrent. Alors pour tout n, le mot u possède un unique facteur spécial de longueur n. Par suite, u possède soit une lettre bispéciale dont le carré est facteur de u, soit une lettre bispéciale dont le carré n'est pas facteur de u, soit enfin une lettre spéciale à gauche et une autre lettre spéciale à droite :

Cas 1. Supposons que a est la lettre bispéciale avec aa dans u. Comme u est récurrent nous avons :

$$L_2(u) = \{aa, ab, bc, ca\} \text{ ou } L_2(u) = \{aa, ac, ba, cb\}.$$

En effet, $aa \in L(u)$ et comme a est biprolongeable à droite (puisque a est bispéciale) il vient que soit $ab \in L(u)$, soit $ac \in L(u)$.

Supposons que $ab \in L(u)$. Donc $ac \notin L(u)$. $bb \notin L(u)$ car, b n'étant pas spéciale à droite, si son unique prolongement à droite était b, u serait ultimement périodique ($u = wb^\omega$). De même $cc \notin L(u)$. $ba \notin L(u)$ car sinon c serait absent de u puisque les deux prolongements à droite (resp. à gauche) de a seraient a et b et l'unique prolongement à droite (resp. à gauche) de b serait a. Donc, le prolongement à droite de b est c, c'est-à-dire $bc \in L(u)$. Ainsi $\{aa, ab, bc\} \subset L_2(u)$. $cb \notin L(u)$

car sinon b serait spéciale à gauche puisque ab est déjà dans u. Par suite, le seul prolongement à droite de c ne peut être que a puisque $cc \notin L(u)$ et $cb \notin L(u)$. Ainsi le quatrième élément de $L_2(u)$ est ca.
De même en supposant que $ac \in L(u)$, on obtient $L_2(u) = \{aa,\, ac,\, ba,\, cb\}$.

Quitte à échanger b et c on peut supposer pour continuer que $L_2(u) = \{aa,\, ab,\, ca,\, bc\}$. Définissons les morphismes suivants :

$$f: \begin{array}{rcl} \{a,\, b\} & \longrightarrow & \{a,\, b,\, c\} \\ a & \longmapsto & a \\ b & \longmapsto & bc \end{array} \quad \text{et} \quad g: \begin{array}{rcl} \{a,\, b,\, c\} & \longrightarrow & \{a,\, b\}. \\ a & \longmapsto & a \\ b & \longmapsto & b \\ c & \longmapsto & \varepsilon \end{array}$$

Remarquons que $g \circ f = Id_{\{a,\, b\}}$. Posons $v = g(u)$. Supposons que u ne commence pas par c. Nous avons $u = f(v)$. Montrons que v est un mot a-sturmien. bb n'est pas dans v puisque $f(bb) = bcbc$ contient cb qui n'est pas dans u. Le mot v est non ultimement périodique puisque $f(v)$ est non ultimement périodique. Donc v possède au moins un facteur spécial à droite pour toute longueur donnée. Considérons w un facteur spécial à droite de v. Nous avons $wa,\, wb \in L(v)$. Donc $f(wa),\, f(wb) \in L(u)$, c'est-à-dire $f(w)\,a,\, f(w)\,bc \in L(u)$. Ainsi $f(w)$ est spécial à droite dans u.

Soient w et w' deux facteurs spéciaux à droite de même longueur dans v. Alors $f(w)$ et $f(w')$ sont spéciaux à droite dans u. Soit N un entier tel que $N \geq \max(|f(w)|\,,\, |f(w')|)$. u possède un unique facteur spécial z de longueur N. Le suffixe de longueur $|f(w)|$ de z est spécial à droite donc égal à $f(w)$. De même le suffixe de longueur $|f(w')|$ de z est égal à $f(w')$. Par suite $f(w)$ et $f(w')$ sont suffixes de z. Donc les mots $g \circ f(w) = w$ et $g \circ f(w') = w'$ sont suffixes de $g(z)$. Par conséquent $w = w'$ puisqu'ils sont de même longueur. Finalement v possède un et un seul facteur spécial à droite pour toute longueur donnée. Ainsi, v est sturmien. Comme bb n'est pas dans v alors v est a-sturmien. Donc, il existe un entier naturel non nul k tel que v s'écrit

$$v = S^{k_0}\left(a^{k+\epsilon_1} b a^{k+\epsilon_2} b a^{k+\epsilon_3} b \cdots\right)$$

avec $k_0 \geq 0$ et $(\epsilon_i)_{i \geq 1}$ une suite sturmienne sur l'alphabet $\{0,\, 1\}$.
D'où $u = f\big(S^{k_0}\left(a^{k+\epsilon_1} b a^{k+\epsilon_2} b a^{k+\epsilon_3} b \cdots\right)\big) = T^{k'_0} f\big(\left(a^{k+\epsilon_1} b a^{k+\epsilon_2} b a^{k+\epsilon_3} b \cdots\right)\big) = S^{k_0}\left(a^{k+\epsilon_1} bca^{k+\epsilon_2} bca^{k+\epsilon_3} bc \cdots\right)$ avec $k'_0 \geq k_0$.

En définitive u vérifie (2).

Supposons maintenant que u commence par c. Soit alors $u' = bu$. Montrons que $L(u') = L(u)$. Si $w \in L(u')$ et $w \notin L(u)$ alors w est préfixe de u' et ainsi $w = bcw'$ où cw' est préfixe de u. Comme u est récurrent, il existe une autre occurrence de cw' dans u, et comme $L_2(u) = \{aa,\, ab,\, ca,\, bc\}$, cette occurrence est nécessairement précédée de b. Donc $w = bcw' \in L(u)$, contradiction. On peut donc appliquer ce qui précède à u', ce qui donne $u' = S^{k_0}\left(a^{k+\epsilon_1} bca^{k+\epsilon_2} bca^{k+\epsilon_3} bc \cdots\right)$ et donc $u = T^{k_0+1}\left(a^{k+\epsilon_1} bca^{k+\epsilon_2} bca^{k+\epsilon_3} bc \cdots\right)$.

Cas 2. Supposons que a est la lettre bispéciale. Comme aa n'est pas dans u nous avons :

$$L_2(u) = \{ba, ca, ac, ab\}.$$

Dans ce cas en posant

$$\begin{array}{llll} f_2: & \{a, b\} \longrightarrow \{a, b, c\} & \text{et} \quad g_2: & \{a, b, c\} \longrightarrow \{a, b\} \\ & a \longmapsto ac & & a \longmapsto \varepsilon \\ & b \longmapsto ab & & b \longmapsto b \\ & & & c \longmapsto c \end{array}$$

puis en raisonnant comme au cas 1 on obtient que u a la forme (3) du théorème.
Cas 3. Supposons que a est la lettre spéciale à gauche et b la lettre spéciale à droite. Nous avons

$$L_2(u) = \{ba, bc, ca, ab\}.$$

En effet, $aa \notin L(u)$, car a n'est pas spécial à droite et si son unique prolongement était à droite était a, u serait ultimement périodique ($u = wa^\omega$). Donc les prolongements à gauche de a sont b et c. De même, $bb \notin L(u)$, donc les prolongements à droite de b sont a et c. Ainsi $\{ba, bc, ca,\} \subset L_2(u)$. Le quatrième élément de $L(u)$ ne peut être que ab (car a a un prolongement à droite, et b a un prolongement à gauche). Dans ce cas aussi en posant

$$\begin{array}{llll} f_3: & \{a, b\} \longrightarrow \{a, b, c\} & \text{et} \quad g_3: & \{a, b, c\} \longrightarrow \{a, b\} \\ & a \longmapsto ab & & a \longmapsto a \\ & b \longmapsto c & & b \longmapsto \varepsilon \\ & & & c \longmapsto b \end{array}$$

puis en suivant une démarche analogue au cas 1 on vérifie que u est de la forme (4).
Suffisance. On applique le lemme 5.1 pour les formes (1), (2) et (4). Pour la forme (3) qui correspond aux mots quasi-sturmiens par insertion, on applique le lemme 3.7 et le corollaire 4.1. □

6. Complexité des mots par insertion k à k des mots sturmiens

Considérons u un mot infini sur un alphabet fini A. Écrivons u à l'aide d'une factorisation par des mots de longueur constante k non nulle. On a :

$$u = m_0 m_1 m_2 \cdots m_i \cdots.$$

avec $m_i \in L_k(u),\ i \in \mathbb{N}$. Maintenant, insérons entre deux facteurs consécutifs m_i et m_{i+1} de u une lettre c étrangère à l'alphabet de u ($c \notin A$). On obtient le mot $v = m_0 c m_1 c m_2 c \cdots c m_i c \cdots$. Nous appelons le mot v “mot par insertion k à k” de u.

Proposition 6.1. *Soient u un mot infini, $k \geq 1$ un entier et v le mot par insertion k à k de u. Si p_u et p_v désignent respectivement les fonctions de complexité de u*

et de v, et $n \geq 0$, $0 \leq r \leq k$ des entiers, nous avons l'inégalité :

$$p_v(kn+n+1+r) \leq (r+1)\, p_u\,(kn+r) + (k-r)p_u\,(kn+r+1)\,.$$

Preuve. Remarquons que les facteurs de longueur $(k+1)\,n+1+r$ de v proviennent de certains facteurs de u de longueur $kn+r$ et $kn+r+1$. De plus, tout facteur de u de longueur $kn+r$ (resp. $kn+r+1$) produit par le biais de l'insertion au plus $r+1$ (resp. $k-r$) facteurs de longueur $(k+1)\,n+1+r$ de v. D'où l'inégalité :

$$p_v(kn+n+1+r) \leq (r+1)\, p_u\,(kn+r) + (k-r)p_u\,(kn+r+1)\,.$$

□

Nous allons montrer que l'inégalité ci-dessus devient une égalité lorsque u est un mot sturmien.

Théorème 6.2. *Soient u un mot sturmien et $k \geq 1$ un entier. La fonction de complexité p_v du mot v obtenu par insertion k à k de u est donnée par :*

$$p_v(n) = \begin{cases} n^2+n+1 & \forall n \leq k \\ kn+k+1 & \forall n > k \end{cases}.$$

Preuve. Soient $q \geq 1$ et $0 \leq r \leq k$ deux entiers. En vertu du théorème 3.9, de tout facteur de longueur supérieure à k de u, en insérant la lettre étrangère c à partir de ses préfixes de longueur inférieure ou égale à k, on obtient des facteurs de v. Par conséquent, on obtient, par insertion de tout facteur de u de longueur $kq+r$ (resp. $kq+r+1$), $r+1$ (resp. $k-r$) facteurs de v de longueur $(k+1)q+1+r$. Les facteurs de v de longueur $(k+1)q+1+r$ contiennent q ou $q+1$ occurrences de la lettre étrangère c à u. D'où les égalités :

$$\begin{aligned} p_v((k+1)q+1+r) &= (r+1)\, p_u\,(kq+r) + (k-r)p_u\,(kq+r+1) \\ &= (r+1)\,(kq+r+1) + (k-r)\,(kq+r+1+1) \\ &= k\,((k+1)\,q+r+1) + k + 1. \end{aligned}$$

Ainsi

$$\forall\, n \geq k+2, \quad p_v\,(n) = kn+k+1.$$

Soit maintenant $1 \leq n \leq k$. Tout facteur de longueur n de u est, en vertu du théorème 3.9, aussi facteur de v. Le mot v contient aussi des facteurs de longueur n qui contiennent une occurrence de la lettre c, obtenus par insertion de cette lettre dans un facteur de longueur $n-1$ de u. Chaque facteur de longueur $n-1$ de u donne ainsi n facteurs de v. On a donc :

$$p_v(n) = p_u(n) + np_u(n-1) = n+1+n^2.$$

Finalement, les facteurs de longueur $k+1$ de v sont tous obtenus en insérant une lettre c dans les facteurs de longueur k de u, d'où

$$p_v(k+1) = (k+1)p_u(k) = k^2 + 2k + 1 = kn + k + 1 \text{ si } n = k+1.$$

□

Remarque 6.3. Pour qu'un mot v obtenu par insertion k à k soit quasi-sturmien il faut et il suffit que $k = 1$. Il est alors de complexité $n+2$.

Cette remarque justifie la définition 4.3. Nous nous intéressons, dans le paragraphe qui suit, aux propriétés combinatoires de cette classe de mots.

7. Combinatoire des mots quasi-sturmiens par insertion

Dans ce paragraphe nous généralisons la notion d'équilibre vue à la section 3 puis nous étudions la palindromie.

7.1. Equilibre, vecteurs de Parikh

Définition 7.1. Un mot u est dit k-équilibré si k est le plus petit entier tel que pour tout couple (v, w) de facteurs de u de même longueur et toute lettre a de A on ait :

$$\Big| |v|_a - |w|_a \Big| \leq k.$$

Théorème 7.2. *Tout mot u quasi-sturmien par insertion est 2-équilibré.*

Preuve. Soit u un mot quasi-sturmien par insertion. On peut supposer sans perte de généralité que u est a-spécial.

1[ère] **étape :** Soient v et w deux facteurs de longueur égale et paire. De v et w on peut extraire deux mots v' et w' de longueur commune $\frac{|v|}{2}$ ne contenant pas a, la lettre spéciale de u. Les mots v' et w' sont en réalité deux facteurs du mot sturmien u' extrait de u. Par conséquent

$$\Big| |v'|_x - |w'|_x \Big| \leq 1$$

pour tout $x \in \{b, c\}$. De plus, remarquons que :

$$|v'|_x = |v|_x \text{ et } |w'|_x = |w|_x .$$

Par ailleurs nous avons aussi :

$$\Big| |v'|_a - |w'|_a \Big| = \left| \frac{|v|}{2} - \frac{|w|}{2} \right| = 0 \leq 1.$$

Il en résulte que :

$$\Big| |v|_x - |w|_x \Big| \leq 1$$

pour tout $x \in \{a, b, c\}$.

$2^{\text{ème}}$ **étape :** Soient r et s deux facteurs de longueur égale et impaire. On peut écrire $r = vx$ et $s = wy$ où (x, y) est un couple de lettres et (v, w) un couple de facteurs de longueur égale et paire de u. On a :

$$\begin{aligned}
\Big| |r|_z - |s|_z \Big| &= \Big| \Big(|v|_z + |x|_z \Big) - \Big(|w|_z + |y|_z \Big) \Big| \\
&= \Big| \Big(|v|_z - |w|_z \Big) + \Big(|x|_z - |y|_z \Big) \Big| \\
&\leq \Big| |v|_z - |w|_z \Big| + \Big| |x|_z - |y|_z \Big| \\
&\leq 1 + 1 = 2
\end{aligned}$$

(la dernière majoration est obtenue à partir du résultat de la $1^{\text{ère}}$ étape).

$3^{\text{ème}}$ **étape :** Le majorant 2 est atteint.
En effet, u contient l'un des mots bab et cac puisque u' contient bb ou cc. Supposons pour continuer que u contienne bab. On sait que aca est dans u. D'où

$$\Big| |bab|_b - |aca|_b \Big| = 2.$$

En définitive u est 2-équilibré. □

Soit $v \in A^+$, on appelle vecteur des occurrences des lettres dans v ou vecteur de Parikh de v et on note $\mathbf{c}(v)$ le triplet $(|v|_a, |v|_b, |v|_c)$.

Soient u un mot infini sur A et n un entier naturel non nul. On désigne par $V_n(u)$ l'ensemble $\{\mathbf{c}(v) : v \in L_n(u)\}$.

Coven et Hedlund ont montré dans [9] qu'un mot infini u est sturmien si et seulement si $\#V_n(u) = 2$ pour tout $n \geq 1$. Dans [21], Rauzy pose la question de généralisation de cette propriété.

Théorème 7.3. *Un mot récurrent u est quasi-sturmien par insertion si et seulement si les conditions suivantes sont satisfaites :*
(i) $\#V_1(u) = 3$,
(ii) $\#V_n(u) = 2$ pour tout n pair non nul,
(iii) $\#V_n(u) = 4$ pour tout n impair et différent de 1.

Preuve. Soit u un mot récurrent.
Supposons u quasi-sturmien par insertion sur A. On peut supposer u a-spécial sans perte de généralité.
(i) Nous avons $V_1(u) = \{(1, 0, 0); (0, 1, 0); (0, 0, 1)\}$ et donc $\#V_1(u) = 3$.
(ii) Soit n un entier pair non nul et soit u' le mot sturmien extrait de u. Soit θ_n, l'application de $L_n(u)$ dans $L_{\frac{n}{2}}(u')$ telle que pour tout $v \in L_n(u)$, $\theta_n(v) = v'$ est le mot obtenu en effaçant la lettre spéciale a dans v. Cette application est trivialement surjective mais non injective. En effet, soit $v' = x_1 x_2 \cdots x_{\frac{n}{2}}$ un élément de $L_{\frac{n}{2}}(u')$.

Alors les mots

$$v_1 = ax_1ax_2\cdots ax_{\frac{n}{2}} \text{ et } v_2 = x_1ax_2a\cdots x_{\frac{n}{2}}a$$

sont deux éléments distincts de $L_n(u)$ tels que $\theta_n(v_1)=\theta_n(v_2)=v'$.
Pour tout élément v de $L_n(u)$ on a :

$$\mathbf{c}(v)=\Big(\frac{n}{2},\,|v'|_b\,,\,|v'|_c\Big) \text{ où } v'=\theta_n(v).$$

Comme θ est surjectif on en déduit :

$$\{\mathbf{c}(v) : v\in L_n(u)\}=\Big\{\Big(\frac{n}{2},\,|v'|_b\,,\,|v'|_c\Big) : v'\in L_{\frac{n}{2}}(u')\Big\}.$$

D'où

$$\#V_n(u)=\#\Big\{\Big(\frac{n}{2},\,|v'|_b\,,\,|v'|_c\Big) : v'\in L_{\frac{n}{2}}(u')\Big\}=\#V_{\frac{n}{2}}(u')=2$$

car u' est sturmien.
(iii) Soient n un entier impair différent de 1 et $v\in L_n(u)$. Désignons par F_1 (resp. F_2) l'ensemble des éléments de $L_n(u)$ commençant par a (resp. ne commençant pas par a). Nous avons :

$$L_n(u)=F_1\cup F_2 \text{ avec } F_1\cap F_2=\emptyset.$$

Soit φ l'application de F_1 dans $L_{\frac{n-1}{2}}(u')$ consistant à effacer la lettre spéciale a. On a l'égalité vectorielle

$$\mathbf{c}(v)=\Big(\frac{n+1}{2},\,|v'|_b\,,\,|v'|_c\Big) \text{ où } v'=\theta_n(v).$$

À tout élément $v'=x_1x_2\cdots x_{\frac{n-1}{2}}$ de $L_{\frac{n-1}{2}}(u')$ correspond l'unique élément $v=ax_1ax_2\cdots ax_{\frac{n-1}{2}}a$ de F_1. Il en résulte que φ est bijective. D'où

$$\{\mathbf{c}(v) : v\in F_1\}=\Big\{\Big(\frac{n+1}{2},|v'|_b\,,|v'|_c\Big) : v'\in L_{\frac{n-1}{2}}(u')\Big\}.$$

Ainsi,

$$\#\{\mathbf{c}(v) : v\in F_1\}=\#\Big\{\Big(\frac{n+1}{2},|v'|_b\,,|v'|_c\Big) : v'\in L_{\frac{n-1}{2}}(u')\Big\}=\#V_{\frac{n-1}{2}}(u')=2$$

car u' est sturmien. De même, en utilisant ψ l'application de F_2 dans $L_{\frac{n+1}{2}}(u')$ qui consiste à effacer la lettre spéciale a, on a

$$\#\{\mathbf{c}(v) : v\in F_2\}=2$$

car u' est sturmien. D'autre part, comme F_1 et F_2 forment une partition de $L_n(u)$, on a :

$$V_n(u)=\{\mathbf{c}(v) : v\in F_1\}\cup\{\mathbf{c}(v) : v\in F_2\}.$$

De plus la réunion est disjointe car

$$\left\{\left(\tfrac{n+1}{2}, |v'|_b\,, |v'|_c\right) : v' \in L_{\frac{n-1}{2}}(u')\right\} \cap \left\{\left(\tfrac{n-1}{2},\, |v'|_b\,,\, |v'|_c\right) : v' \in L_{\frac{n+1}{2}}(u')\right\} = \emptyset.$$

Par suite

$$\#V_n(u) = \#\left\{\mathbf{c}(v) : v \in F_1\right\} + \#\left\{\mathbf{c}(v) : v \in F_2\right\}.$$

Ainsi,

$$\#V_n(u) = 2 + 2 = 4.$$

Réciproquement supposons les conditions (i), (ii) et (iii) satisfaites par u. On déduit de (i) que les trois lettres sont présentes dans u.

De la condition (ii) on a $\#V_2(u) = 2$. Si (2, 0, 0) est dans $V_2(u)$ alors le deuxième vecteur est soit (1, 1, 0), soit (1, 0, 1). En prenant $V_2(u) = \Big\{(2,\,0,\,0);\ (1,\,1,\,0)\Big\}$, on déduit qu'aucun facteur de longueur 2 de u ne contient la lettre c. En d'autres termes, c n'est pas dans u ; contradiction. En prenant $V_2(u) = \Big\{(2,\,0,\,0);\ (1,\,0,\,1)\Big\}$ on aboutit à la même conclusion. Donc (2, 0, 0) n'est pas dans $V_2(u)$. De même on montre que (0, 0, 2) et (0, 2, 0) ne sont pas dans $V_2(u)$. Ainsi, les mots aa, bb et cc ne sont pas facteurs de u. Par conséquent $V_2(u)$ est l'une des trois paires :

$$\Big\{(1,\,1,\,0);\ (1,\,0,\,1)\Big\},\ \Big\{(1,\,1,\,0);\ (0,\,1,\,1)\Big\} \text{ et } \Big\{(1,\,0,\,1);\ (0,\,1,\,1)\Big\}.$$

On peut supposer pour continuer la preuve que

$$V_2(u) = \Big\{(1,\,1,\,0);\ (1,\,0,\,1)\Big\}.$$

Par suite, les facteurs de longueur 2 de u, formés chacun de lettres distinctes contiennent toujours la lettre a. Donc, les facteurs de longueur 2 de u sont ab, ac, ba et ca puisque u est récurrent. Plus précisément les lettres b et c n'admettent qu'un seul prolongement à droite et à gauche qui est a. Par conséquent u est de la forme

$$u = x_0ax_1ax_2ax_3\cdots$$

où $x_0 \in \{\varepsilon,\, b,\, c\}$ et $x_i \in \{b,\, c\}$ pour $i = 1,\, 2,\, 3,\, \ldots$.

Le mot u est non ultimement périodique. En effet, en supposant le contraire, u est alors périodique puisqu'il est récurrent. Par suite, $\#V_m(u) = 1$ où m est la période de u. Ce qui est impossible car selon les conditions (i), (ii) et (iii) $\#V_n(u) \neq 1$ pour tout entier naturel n non nul.

Montrons à présent que u est de complexité $n + 2$ ($n \geq 1$). Pour cela nous allons établir que pour tout entier naturel non nul n, u admet un unique facteur biprolongeable de longueur n. Comme $L_2\,(u) = \{ab,\, ac,\, ba,\, ca\}$ il vient que u n'a qu'un facteur biprolongeable, en l'occurrence la lettre a. En effet, a se prolonge à droite (resp. à gauche) par b et c. Par l'absurde, supposons que u admette deux facteurs biprolongeables à droite. Alors u contient deux facteurs biprolongeables à droite de la forme xr et yr où x, $y \in \{b, c\}$ (car seulement b et c ont un prolongement commun dans u) et $r \in L(u)$. Par suite, les mots :

$$xrx,\ xry,\ yrx \text{ et } yry$$

sont des facteurs de u.
Posons :

$$F = \{xrx, xry, yrx, yry\}.$$

On remarque aisément que :

$$\#\{\mathbf{c}(w) : w \in F\} = 3.$$

Par ailleurs

$$F' = \{axr, rxa, ayr, rya\}$$

est un ensemble de facteurs de u de même longueur que les éléments de F.
De plus

$$\#\{\mathbf{c}(w) : w \in F'\} = 2.$$

Les éléments de F' contiennent un a de plus que ceux de F. Donc

$$\{\mathbf{c}(w) : w \in F\} \cap \{\mathbf{c}(w) : w \in F'\} = \emptyset.$$

Considérons m la longueur commune des éléments de F et F'. Nous avons :

$$(F \cup F') \subset L_m(u).$$

Par conséquent

$$\begin{aligned} \#V_m(u) &\geq \#\{\mathbf{c}(w) : w \in F \cup F'\} \\ &= \#\{\mathbf{c}(w) : w \in F\} + \#\{\mathbf{c}(w) : w \in F'\} \\ &= 3 + 2 = 5. \end{aligned}$$

On obtient une contradiction, car selon (i), (ii) et (iii) on a :

$$\#V_m(u) \leq 4.$$

□

7.2. Palindromie

Lemme 7.4. *L'ensemble des facteurs d'un mot quasi-sturmien par insertion est stable par image miroir.*

Preuve. Soient u un mot quasi-sturmien par insertion et u' son mot sturmien extrait. Soient m un entier non nul et $v \in L_{2m}(u)$. On peut supposer sans perte de généralité que u est a-spécial. Par suite, on a :

$$v = ax_1ax_2 \cdots ax_m \text{ ou } v = x_1ax_2 \cdots ax_ma.$$

Ainsi le mot $v' = x_1x_2 \cdots x_m$ est un facteur de u', et puisque u' est un mot sturmien, alors $\overline{v'} = x_mx_{m-1} \cdots x_1$ est facteur de u' [5]. Pour tout facteur $y_1y_2 \dots y_n$ de u', les deux mots $ay_1ay_2a \dots ay_n$ et $y_1ay_2a \dots ay_na$ sont des facteurs de u car le mot $ay_1ay_2a \dots ay_na$ est un facteur de u. Par conséquent les mots

$$x_max_{m-1}a \cdots x_1a \text{ et } ax_max_{m-1} \cdots ax_1$$

sont dans u. D'où $\overline{v} \in L_{2m}(u)$ et l'application miroir préserve les facteurs de longueur paire de u. Par ailleurs, comme tout facteur de u de longueur impaire est préfixe d'un facteur de longueur paire plus grande, on déduit que les facteurs de u de longueur impaire sont aussi préservés par l'application miroir. □

Remarque 7.5. L'ensemble des facteurs d'un mot de complexité $n + 2$ de la forme (1), (2) ou (4) du théorème 5.2 n'est pas stable par l'application miroir.

En effet, l'ensemble des facteurs de longueur 2 d'un mot de complexité $n+2$ de la forme (1), (2) ou (4) du théorème 5.2 n'est pas préservé par l'application miroir.

Comme les facteurs de longueur paire des mots quasi-sturmiens par insertion commencent et finissent par des lettres différentes, il convient de faire la remarque suivante.

Remarque 7.6. Les facteurs bispéciaux des mots quasi-sturmiens par insertion sont de longueurs impaires.

Proposition 7.7. *L'image miroir d'un facteur biprolongeable à droite d'un mot quasi-sturmien par insertion est biprolongeable à gauche et inversement.*

Preuve. Soit u un mot quasi-sturmien par insertion que l'on peut supposer a-spécial sans perte de généralité. Soient n un entier non nul et D_n, $G_n \in L_n(u)$ tels que

$$\partial^- G_n = \partial^+ D_n = 2.$$

Alors nous avons :

$$D_nb,\ D_nc,\ bG_n,\ cG_n \in L_{n+1}(u).$$

Par suite, en vertu du lemme 7.4, nous avons :

$$b\overline{D_n},\ c\overline{D_n} \in L_{n+1}(u)$$

et donc $\partial^- \overline{D_n} = 2$. Ainsi $\overline{D_n}$ est biprolongeable à gauche.

De même on montre que $\overline{G_n}$ est biprolongeable à droite. □

Remarque 7.8. Soient u un mot quasi-sturmien par insertion et P un palindrome de u. Si P est biprolongeable à droite, alors P est biprolongeable à gauche, et inversement.

Lemme 7.9. *Soit u un mot quasi-sturmien par insertion. Alors, u possède, pour tout entier n impair, exactement trois palindromes de longueur n.*

Preuve. Soit u un mot quasi-sturmien par insertion.
Pour $n = 1$ la propriété est vraie car les trois lettres a, b et c présentes dans u sont des palindromes. Supposons la propriété vraie pour $n = 2m - 1$ supérieur à 1 : il existe exactement trois palindromes P_1, P_2 et P_3 de longueur $2m - 1$ dans u. Verifions que la propriété reste vraie pour la longueur $2m + 1$. Deux éventualités se présentent : soit l'un de ces trois palindromes est spécial, soit aucun d'eux ne l'est.

Cas 1. Les palindromes P_1, P_2 et P_3 ne sont pas spéciaux. Montrons qu'il existe des lettres x_1, x_2 et x_3 dans $\{a, b, c\}$ telles que $x_iP_ix_i \in L_{2m+1}(u)$ pour $i = 1, 2, 3$. Comme P_i $(i = 1, 2, 3)$ est non spécial, alors P_i admet un unique prolongement à droite (resp. à gauche) dans u. Soit x_i l'unique prolongement à droite de P_i. On a $P_ix_i \in L_{2m}(u)$ et d'après le lemme 7.4, $x_iP_i = \overline{P_ix_i} \in L_{2m}(u)$. Ainsi x_i est l'unique prolongement à gauche de P_i dans u. Par conséquent $x_iP_ix_i \in L_{2m+1}(u)$ pour $i = 1$, 2 ou 3. Ainsi il existe des lettres x_1, x_2 et x_3 dans $\{a, b, c\}$ telles que $x_iP_ix_i \in L_{2m+1}(u)$ pour $i = 1, 2, 3$. Ces trois mots

$$x_1P_1x_1,\ x_2P_2x_2 \text{ et } x_3P_3x_3$$

(qui sont distincts puisque P_1, P_2 et P_3 le sont) sont des palindromes sur u.

Cas 2. L'un des palindromes P_1, P_2 et P_3 est spécial. On peut supposer qu'il s'agit de P_1 pour continuer la preuve. Donc les palindromes P_2 et P_3 ne sont pas spéciaux. Ainsi on montre comme dans le cas 1 qu'il existe des lettres x_2 et x_3 dans $\{a, b, c\}$ telles que $x_iP_ix_i \in L_{2m+1}(u)$ avec $i = 2, 3$.

Considérons maintenant le palindrome biprolongeable P_1. On a $\partial^- P_1 = \partial^+ P_1 = 2$ et il existe donc deux lettres x et y telles que $P_1x, P_1y, xP_1, yP_1 \in L_{2m}(u)$. L'unique facteur biprolongeable à gauche de longueur $2m$ est soit P_1x, soit P_1y. On peut supposer sans perte de généralité que P_1x est ce facteur biprolongeable et donc P_1y et son image miroir yP_1 ne sont pas biprolongeables dans u. Par suite, $xP_1x, xP_1y, yP_1x \in L_{2m+1}(u)$ et $yP_1y \notin L_{2m+1}(u)$. Ce qui prouve en posant $\mathrm{x}_1 = x$ qu'il existe des lettres x_1, x_2 et x_3 dans $\{a, b, c\}$ telles que $x_iP_ix_i \in L_{2m+1}(u)$ pour $i = 1, 2, 3$.

Finalement u possède au moins trois palindromes de longueur $2m + 1$.

Soit R un palindrome de longueur $2m + 1$ sur u. Alors R est de la forme xPx avec $P \in L_{2m-1}(u)$ et $x \in A$. Comme R est un palindrome on a $R = \overline{R}$ c'est-à-dire $xPx = \overline{xPx} = x\overline{P}x$. D'où $P = \overline{P}$ et comme P est de longueur $2m - 1$ (impaire) alors par hypothèse de récurrence, $P \in \{P_1, P_2, P_3\}$. Par conséquent $R \in \{x_1P_1x_1, x_2P_2x_2, x_3P_3x_3\}$. Ainsi u admet exactement trois palindromes. $\square$

Remarque 7.10. Le résultat précédent est en fait une conséquence immédiate d'un résultat de Droubay et Pirillo (théorème 5 de [11]).

Proposition 7.11. *Pour tout mot infini périodique u, il existe un entier n_0 tel que la complexité palindromique vérifie $pal_u(n) \leq 2$ pour tout $n \geq n_0$.*

Preuve. Soient u un mot périodique de période τ et I_n, l'ensemble des palindromes de longueur n de u, avec $n \geq \tau$. Par la périodicité de u nous avons :

$$I_n = \{i < \tau : u_i u_{i+1} \cdots u_{i+n-1} = u_{i+n-1} \cdots u_{i+1} u_i\}.$$

Supposons I_n non vide et posons $i_0 = \min(I_n)$.

Soit $j \in I_n$. Nous avons alors :

$$u_{i_0+h} = u_{i_0+n-1-h} \text{ et } u_{j+h} = u_{j+n-1-h}$$

pour tout $h \in \{0, 1, \ldots, n-1\}$. D'où

$$\begin{aligned} u_{i_0+h} &= u_{i_0+n-1-h} \\ &= u_{j+n-1-(h+j-i_0)} \\ &= u_{j+(h+j-i_0)} \text{ si } h \leq n-1-(j-i_0) \\ &= u_{i_0+h+2(j-i_0)}. \end{aligned}$$

Pour n suffisamment grand on a $n-1-(j-i_0) \geq \tau$ car $j - i_0 \leq \tau$. Soit n_0 le plus petit entier vérifiant cette condition. Alors pour $n \geq n_0$ on a $u_{i_0+h+2(j-i_0)} = u_{i_0+h}$ pour tout $h \in \{0, 1, \ldots, n-1-(j-i_0)\}$ et donc pour tout $h \in \{0, 1, \ldots, \tau-1\}$. Par conséquent la période τ divise $2(j-i_0)$. Ainsi $2(j-i_0) = 0$ ou τ puisque $j - i_0 < \tau$. Par suite, $j = i_0$ ou $j = i_0 + \frac{\tau}{2}$ et donc $\#I_n \leq 2$. □

Théorème 7.12. *Un mot récurrent u est quasi-sturmien par insertion si et seulement si les assertions suivantes sont vérifiées :*

(i) L'application miroir préserve les facteurs de u c'est-à-dire

$$\forall v \in L(u) : \ \overline{v} \in L(u).$$

(ii) $pal_u(2) = 0$.

(iii) $pal_u(n) = 3$ pour tout entier n impair.

(iv) Une et une seule des lettres présentes dans u est biprolongeable.

Preuve. Soit u un mot récurrent.

Supposons u quasi-sturmien par insertion. Les lemmes 7.4 et 7.9 nous assurent respectivement les assertions (i), (iii). Le mot u vérifie (ii) et (iv) comme tout mot quasi-sturmien par insertion.

Réciproquement supposons les assertions (i), (ii), (iii) et (iv) vraies pour u.

$1^{\text{ère}}$ **étape :** montrons que u est de la forme

$$u = x_0 z x_1 z x_2 z x_3 \cdots$$

où $x_0 \in \{\varepsilon, x, y\}$ et $x_i \in \{x, y\}$ pour $i = 1, 2, 3, \ldots$ avec $\{x, y, z\} = \{a, b, c\}$. Les trois palindromes de longueur 1 (conformément à l'assertion (*iii*)) de u sont les lettres a, b et c. En vertu de l'assertion (*ii*), les mots aa, bb et cc ne sont pas dans u. En d'autres termes, u ne contient pas de carré de lettre. Par conséquent

u ne possède pas de lettre triprolongeable et ainsi toute lettre spéciale dans u est biprolongeable à gauche ou à droite. De l'assertion (i) on déduit que toute lettre spéciale dans u est bispéciale. Donc, par (iv) on conclut que u ne possède qu'une seule lettre spéciale ; soit z cette lettre. Alors, les deux autres lettres désignées par x et y sont non spéciales. Par suite, les facteurs de longueur 2 de u sont nécessairement zx, zy, xz et yz. Par conséquent u est de la forme voulue.

$2^{\text{ème}}$ **étape :** montrons que u est de complexité $n+2$ ($n \geq 1$). Supposons que u est ultimement périodique. Alors u serait périodique puisqu'il est récurrent et contiendrait au plus 2 palindromes pour toute longueur donnée à partir d'un certain rang (cf. Proposition 7.11). On obtient une contradiction avec l'assertion (iii). Par suite, u possède au moins un facteur biprolongeable à droite de longueur n pour tout naturel non nul n. Ainsi il nous suffira de montrer que pour tout entier n non nul, u admet un et un seul facteur de longueur n biprolongeable à droite. Par l'absurde, supposons que u admette deux facteurs biprolongeables à droite de même longueur. Considérons R et R' de tels facteurs de longueur n minimale. Par minimalité de n nous aurons $R = xD$ et $R' = yD$ où D est l'unique facteur biprolongeable à droite de longueur $n-1$. Par suite xDx, xDy, yDx et yDy sont des facteurs de u. Par l'assertion (i), il vient que les mots $x\overline{D}x$, $y\overline{D}x$, $x\overline{D}y$ et $y\overline{D}y$ sont aussi des facteurs de u. Ainsi, les facteurs D et $\overline{D}$ sont bispéciaux et a fortiori biprolongeables à droite de longueur $n-1$ dans u. Ce qui prouve que $D = \overline{D}$ puisque le facteur biprolongeable de longueur $n-1$ est unique. En d'autres termes D est un palindrome. Par suite, $n-1$ est impair et selon (iii) on peut donc considérer P_1 et P_2 les deux autres palindromes de longueur $n-1$. Les palindromes P_1 et P_2 sont non biprolongeables car par minimalité de n, D est l'unique facteur biprolongeable de longueur $n-1$ de u. Par (i) on montre que P_k ($k = 1, 2$) se prolonge à gauche et à droite par la même lettre : il existe des lettres y_1 et y_2 telles que $y_1P_1y_1$ et $y_2P_2y_2$ soient deux facteurs de u. En somme, u contient au total quatre palindromes de longueur $n+1$, en l'occurence, xDx, yDy, $y_1P_1y_1$ et $y_2P_2y_2$ puisque D, P_1 et P_2 sont distincts, ce qui est contradictoire avec (iii). $\square$

8. Perspectives

Nous avons déterminé certaines propriétés des mots quasi-sturmiens par insertion (équilibre, vecteurs de Parikh, palindromie). Il serait intéressant d'étendre cette étude à tous les mots obtenus par insertion k à k des mots sturmiens. En particulier, existe-il une caractérisation à l'aide des vecteurs de Parikh de ces mots ou d'autres classes de mots ; cela répondrait à une question de Rauzy ([21], sect. 6.2).

Remerciements. Les auteurs sont très redevables à Julien Cassaigne et aux deux rapporteurs anonymes dont la lecture attentive et les suggestions ont beaucoup amélioré une première version de ce papier.

Références

[1] P. Alessandri, Classification et représentation des mots de complexité $n+2$. *Rapport technique*, Université Aix-Marseille II (1995).

[2] P. Alessandri, Codage de rotations et basses complexités. *Thèse*, Université Aix-Marseille II (1996).

[3] J.-P. Allouche, Sur la complexité des suites infinies. *Bull. Belg. Math. Soc. Simon Stevin* **1** (1994) 133–143.

[4] J.-P. Allouche, M. Baake, J. Cassaigne et D. Damanik, Palindrome complexity. *Theoret. Comput. Sci.* **292** (2003) 9–31.

[5] V. Berthée, Fréquences des facteurs des suites sturmiennes. *Theoret. Comput. Sci.* **165** (1996) 295–309.

[6] J. Cassaigne, Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997) 67–88.

[7] J. Cassaigne, Sequences with grouped factors, in *Developments in Language Theory III (DLT'97)*, pp. 211–222, Aristotle University af Thessaloniki (1998).

[8] E.M. Coven, Sequences with minimal block growth. *Math. Syst. Theory* **8** (1975) 376–382.

[9] E.M. Coven et G.A. Hedlund, Sequences with minimal block growth. *Math. Syst. Theory* **7** (1973) 138–153.

[10] G. Didier, Caractérisation des N-écritures et application à l'étude des suites de complexité ultimement $n + C^{\text{ste}}$. *Theoret. Comput. Sci.* **215** (1999) 31–49.

[11] X. Droubay et G. Pirillo, Palindromes and Sturmian words. *Theoret. Comput. Sci.* **223** (1999) 73–85.

[12] S. Dulucq et D. Gouyou-Beauchamps, Sur les facteurs des suites de Sturm. *Theoret. comput. Sci.* 71 (1990) 381–400.

[13] S. Ferenczi et C. Mauduit, Transcendency of numbers with a low complexity expansion. *J. Number Theory* **67** (1997) 146–161.

[14] A. Heinis, Arithmetics and combinatorics of words of low complexity. *P*h.D. Thesis, University of Leiden (2001).

[15] M. Lothaire, *Algebraic combinatorics on words*, vol. 90 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press (2002).

[16] F. Mignosi et P. Séébold, Morphismes sturmiens et règles de Rauzy. *J. Théor. Nombres Bordeaux* **5** (1993) 221–233.

[17] M. Morse et G.A. Hedlund, Symbolic dynamics. *Amer. J. Math.* **60** (1938) 815–866.

[18] M. Morse et G.A. Hedlund, Symbolic dynamics II : Sturmian trajectories. *Amer. J. Math.* 62 (1940) 1–42.

[19] M.E. Paul, Minimal symbolic flows having minimal block growth. *Math. Syst. Theory* **8** (1975) 309–315.

[20] N. Pytheas Fogg, Substitutions in Dynamics, Arithmetics and Combinatorics. *Lect. Notes Math.* **1794** (2002).

[21] G. Rauzy, Suites à termes dans un alphabet fini. *Sémin. Théor. Nombres Bordeaux* **25** (1982–1983) 1–16.

Communiqué par J. Berstel.
Received May 10, 2006. Accepted January 9, 2007.