INFORMATIQUE THÉORIQUE ET APPLICATIONS

J. BERSTEL
P. SÉÉBOLD

A remark on morphic sturmian words

Informatique théorique et applications, tome 28, nº 3-4 (1994), p. 255-263
<http://www.numdam.org/item?id=ITA_1994__28_3-4_255_0>

© AFCET, 1994, tous droits réservés.

L'accès aux archives de la revue « Informatique théorique et applications » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

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

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 28, n° 3-4, 1994, p. 255 à 263)

A REMARK ON MORPHIC STURMIAN WORDS (1)

by J. BERSTEL (2) and P. SÉÉBOLD (3)

Abstract. – *This Note deals with binary Sturmian words that are morphic,* i.e. *generated by iterating a morphism. Among these, characteristic words are a well-known subclass. We prove that for every characteristic morphic word* $\mathbf{x}$*, the four words* $a\mathbf{x}$*,* $b\mathbf{x}$*,* $ab\mathbf{x}$ *and* $ba\mathbf{x}$ *are morphic.*

1. INTRODUCTION

Combinatorial properties of finite and infinite words are of increasing importance in various fields of physics, biology, mathematics and computer science. Infinite words generated by various devices have been considered [11]. We are interested here in a special family of infinite words, namely Sturmian words. Sturmian words represent the simplest family of quasi-crystals (*see e.g.* [2]). They have numerous other properties, related to continued fraction expansion (*see* [4, 6] for recent results, in relation with iterating morphisms). There are numerous relations with fractals and data compression [7, 8, 9, 12], with molecular biology [10].

In this Note, we prove a combinatorial property of a special class of morphic words, namely morphic Sturmian words. The property is best explained by an example. Consider the infinite Fibonacci word

$$\mathbf{f} = abaababaabaab\ldots$$

(1) Partially supported by PRC "Mathématiques et Informatique" and by ESPRIT BRA working group 6317 – ASMICS 2.

(2) LITP Institut Blaise Pascal, Université Pierre-et-Marie-Curie, 4, place Jussieu, 75252 Paris Cedex 05.
(3) LAMIFA Faculté de Mathématiques et Informatique, 33, rue Saint-Leu, 80039 Amiens Cedex.

generated by the morphism

$$a \mapsto ab$$
$$b \mapsto a$$

In contrast to words generated by substitutions (*i. e.* codings of morphic words), morphic words do not well behave with respect to the shift. However, the Fibonacci word is an exception. Indeed, both words

$$a\,\mathbf{f} = aabaababaabaab\ldots$$

and

$$b\,\mathbf{f} = babaababaabaab\ldots$$

are morphic, with generators

$$\begin{array}{ll} a \mapsto aab & a \mapsto baa \\ b \mapsto ab & b \mapsto ba \end{array}$$

respectively. Moreover, the two words

$$ba\,\mathbf{f} = baabaababaabaab\ldots \qquad ab\,\mathbf{f} = ababaababaabaab\cdots$$

are both morphic, with the same generator

$$a \mapsto aba$$
$$b \mapsto ba$$

taking as starting letters either b or a.

This property will be shown to hold for a wide class of words, namely for all morphic words which are characteristic Sturmian words.

2. PRELIMINARIES

An infinite word is a mapping

$$\mathrm{x}: \quad \mathbb{N}_+ \to A$$

where $\mathbb{N}_+ = \{1, 2\ldots\}$ is the set of positive integers and A is an alphabet. In the sequel, we consider binary words, that is words over a two letter alphabet $A = \{a, b\}$. A^ω is the set of *infinite words* on A and $A^\infty = A^* \cup A^\omega$.

Let $f : A^* \to A^*$ be a morphism. Assume that, for some letter a, the word $f(a)$ starts with a. Then $f^{n+1}(a)$ starts with $f^n(a)$ for all n. If the set $\{f^n(a) | n \geqq 0\}$ is infinite, then there exists a unique infinite word $\mathbf{x}$

such that every $f^n(a)$ is a prefix of $\mathbf{x}$. The word $\mathbf{x}$ is said to be *generated* by iterating f. For general results, *see* [13]. An infinite word $\mathbf{x}$ is *morphic* if it is generated by iterating a morphism. Any morphism that generates $\mathbf{x}$ is a *generator.*

The *complexity* function of an infinite word $\mathbf{x}$ is the function $P_{\mathbf{x}}$ where $P_{\mathbf{x}}(n)$ is the number of factors of length n of $\mathbf{x}$. It is well-known (e. g. [5]) that $\mathbf{x}$ is ultimately periodic as soon as $P_{\mathbf{x}}(n) \leqq n$ for some $n \geqq 0$. A word $\mathbf{x}$ is *Sturmian* if $P_{\mathbf{x}}(n) = n + 1$ for all n. For any $w \in A^{\infty}$, $\mathrm{Fact}(w)$ denotes the set of *finite factors* of w. Setting, for any $u, v \in A^*$ such that $|u| = |v|$, $\delta(u, v) = ||u|_a - |v|_a|$, we call *balanced* a word $w \in A^{\infty}$ such that $\delta(u, v) \leqq 1$ for any $u, v \in \mathrm{Fact}(w)$ with $|u| = |v|$. Sturmian words are intimately related to cutting sequences in the plane (also known as Beatty sequences. For a recent exposition, *see* [4]). Let α, ρ be real numbers with $0 \leqq \alpha \leqq 1$. Consider the infinite words

$$\mathbf{s}_{\alpha, \rho} = a_1 \dots a_n \dots, \qquad \mathbf{s}'_{\alpha, \rho} = b_1 \dots b_n \dots$$

defined by

$$a_n = \begin{cases} a & \text{if} \quad \lfloor \alpha(n+1) + \rho \rfloor = \lfloor \alpha n + \rho \rfloor \\ b & \text{otherwise} \end{cases}$$

and

$$b_n = \begin{cases} a & \text{if} \quad \lfloor \alpha(n+1) + \rho \rfloor = \lfloor \alpha n + \rho \rfloor \\ b & \text{otherwise} \end{cases}$$

The following theorem states well-known characterizations of Sturmian words.

THEOREM 2.1 [5, 15]: *Let* $\mathbf{x}$ *be an infinite binary word. The following conditions are equivalent:*

(i) $\mathbf{x}$ *is Sturmian;*

(ii) $\mathbf{x}$ *is balanced and not ultimately periodic;*

(iii) *there exist an irrational number* α $(0 < \alpha < 1)$ *and a real* ρ *such that* $\mathbf{x} = \mathbf{s}_{\alpha, \rho}$ *or* $\mathbf{x} = \mathbf{s}'_{\alpha, \rho}$.

A Sturmian word $\mathbf{x}$ is *characteristic* if $\mathbf{x} = \mathbf{s}_{\alpha, 0}$ for some irrational α $(0 < \alpha < 1)$. We write then $\mathbf{c}_{\alpha} = \mathbf{s}_{\alpha, 0}$. In this case, $\mathbf{s}_{\alpha, 0} = \mathbf{s}'_{\alpha, 0}$. In view of the preceding theorem, characteristic words are also described by

COROLLARY 2.2: *A Sturmian word* $\mathbf{x}$ *is characteristic iff both* $a\mathbf{x}$ *and* $b\mathbf{x}$ *are Sturmian.*

Proof: If $\mathbf{x} = \mathbf{s}_{\alpha,0} = \mathbf{s}'_{\alpha,0}$ for some irrational α $(0 < \alpha < 1)$, then a straightforward calculation shows that

$$a\mathbf{x} = \mathbf{s}_{\alpha,-\alpha}, \qquad b\mathbf{x} = \mathbf{s}'_{\alpha,-\alpha}$$

Conversely, assume that $a\mathbf{x}$ and $b\mathbf{x}$ are Sturmian, and that $\mathbf{x} = \mathbf{s}_{\alpha,\rho}$ for some ρ. Since $\mathbf{s}_{\alpha,\rho} = \mathbf{s}_{\alpha',\rho'}$ or $\mathbf{s}_{\alpha,\rho} = \mathbf{s}'_{\alpha',\rho'}$ or $\mathbf{s}'_{\alpha,\rho} = \mathbf{s}'_{\alpha',\rho'}$ implies $\alpha = \alpha'$ and $\rho = \rho'$, it follows that $a\mathbf{x} = \mathbf{s}_{\alpha,\rho-\alpha}$ and $b\mathbf{x} = \mathbf{s}'_{\alpha,\rho-\alpha}$. But then, considering the first letters of these two words, we get the conditions

$$\lfloor \alpha + \rho \rfloor = \lfloor \rho \rfloor, \qquad \text{and} \qquad \lceil \alpha + \rho \rceil = \lceil \rho \rceil$$

which imply that $\rho = 0$. ■

It is easily seen that neither $a\mathbf{x}$ nor $b\mathbf{x}$ are characteristic.

A morphism $f : A^* \to A^*$ is a *Sturmian morphism* if $f(\mathbf{x})$ is Sturmian for all Sturmian words. The following are known for Sturmian morphisms.

THEOREM 2.3 [16]: *Every Sturmian morphism is a composition of the three morphisms*

$$E: \begin{array}{l} a \mapsto b \\ b \mapsto a \end{array} \qquad D: \begin{array}{l} a \mapsto ab \\ b \mapsto a \end{array} \qquad G: \begin{array}{l} a \mapsto ba \\ b \mapsto a \end{array}$$

in any order and number.

THEOREM 2.4 [1]: *A morphism f is Sturmian if $f(\mathbf{x})$ is Sturmian for one Sturmian word* $\mathbf{x}$.

THEOREM 2.5 [6]: *Let $\mathbf{c}_\alpha$ and $\mathbf{c}_\beta$ be characteristic words. If $\mathbf{c}_\alpha = f(\mathbf{c}_\beta)$, then the morphism f is a composition of E and D.*

We call a Sturmian morphism *positive* if it is a composition of E and D. An explicit description of positive Sturmian morphisms can be given in terms of standard pairs. For this, we consider the family $\mathcal{R}$ of (unordered) pairs of words of A^* defined as the smallest set of pairs of words such that

$$(1) \quad \{a, b\} \in \mathcal{R};$$
$$(2) \quad \{u, v\} \in \mathcal{R} \Rightarrow \{uv, u\} \in \mathcal{R}$$

Pairs in $\mathcal{R}$ are *standard pairs*, the components of standard pairs are called *standard words*. Observe that the two components of a standard pair always end with different letters. The relation between Sturmian morphisms and standard pairs is the following:

PROPOSITION 2.6: *A morphism $f : A^* \to A^*$ is a positive Sturmian morphism iff the set $\{f(a), f(b)\}$ is a standard pair.*

Proof: Assume first that f is a positive Sturmian morphism. Arguing by induction on $|f(a)| + |f(b)|$, assume that $\{f(a), f(b)\}$ is a standard pair. If $g = f \circ E$, then $\{g(a), g(b)\} = \{f(a), f(b)\}$ is standard. If $g = f \circ D$, then $\{g(a), g(b)\} = \{f(a) f(b), f(a)\}$ is again standard.

Conversely, assume that $\{f(a), f(b)\}$ is a standard pair, and that $|f(a)| > |f(b)|$. Then $f(a) = f(b) v$ for some word v, and $\{v, f(b)\}$ is a standard pair. By induction, there is a Sturmian positive morphism g such that $\{g(a), g(b)\} = \{v, f(b)\}$. If $g(a) = f(b)$ and $g(b) = v$, then $f = g \circ D$, in the other case, $f = g \circ E \circ D$. ■

We need the following property of standard words:

PROPOSITION 2.7 [14]: *Every standard word w is either a letter or of the form $w = pxy$, with p a palindrom word, and x, y distinct letters.*

3. RESULTS

Let C be the family of morphic Sturmian characteristic words. In view of the preceding results, we get:

THEOREM 3.1: *For any $\mathbf{c} \in C$, the infinite words $a\mathbf{c}$ and $b\mathbf{c}$ are morphic.*

Proof: Let $\mathbf{c}$ be a morphic Sturmian characteristic word, and let f be one of its generators. In view of theorem 2.5, the morphism f is positive and, by proposition 2.6, the pair $\{f(a), f(b)\}$ is standard.

We consider the case where $|f(a)| < |f(b)|$, the other case is symmetric. By the definition of standard pairs, $f(a)$ is a prefix of $f(b)$. Let x be the last letter of $f(a)$. Then there exist two words r and s such that

$$f(a) = rx, \qquad f(b) = rxs$$

Define two morphisms f_a and f_b by

$$f_a : \begin{array}{l} a \mapsto xr \\ b \mapsto xsr \end{array} \qquad f_b : \begin{array}{l} a \mapsto \tilde{r}x \\ b \mapsto \tilde{r}\tilde{s}x \end{array}$$

where $\tilde{w}$ denotes the reversal of w (Observe that $f_b = \tilde{f_a}$).

LEMMA 3.2: *The following relations hold: $f_a(a\mathbf{c}) = x\mathbf{c}$, $f_b(b\mathbf{c}) = \bar{x}\mathbf{c}$, where $\bar{a} = b$ and $\bar{b} = a$.*

Consequently, a generator of $a\,\mathbf{c}$ will be f_a if $x = a$ or $f_b \circ f_a$ if $x = b$, and similarly for $b\,\mathbf{c}$. This proves the theorem. ■

Proof of lemma 3.2: With the notation above, we prove that for any word w,

$$x f(wa) = f_a(aw)\, x$$

and

$$\bar{x} f(wb) = f_b(bw)\, \bar{x} \qquad (\star)$$

The first relation is easily proved by induction on the number of a in w. For the second, we consider first the case where $f(a)$ is a letter, say x. Then $f(b) = x^m \bar{x}$ for some positive m, and the verification is straightforward. Otherwise by proposition 2.7,

$$f(a) = rx = p\bar{x}x, \qquad f(b) = rxs = p\bar{x}xqx\bar{x}$$

where p, q and $p\bar{x}xq$ are palindroms. In particular, for all $n \geqq 0$

$$(p\bar{x}x)^n\, q = q\,(x\bar{x}p)^n \qquad (\star\star)$$

We prove $(\star)$ by induction on the number of b in w. If $w = a^n$ for some $n \geqq 0$, then

$$\begin{aligned}\bar{x} f(a^n b) &= \bar{x}\,(p\bar{x}x)^n\, p\bar{x}xqx\bar{x}\\ &= \bar{x}qx\bar{x}p\,(x\bar{x}p)^n\, x\bar{x}\\ &= \bar{x}qx\bar{x}px\,(\bar{x}px)^n\, \bar{x}\\ &= f_b(ba^n)\,\bar{x}\end{aligned}$$

Next, if $w = uba^n$, then

$$\begin{aligned}\bar{x} f(wb) &= \bar{x} f(ub)\, f(a^n b) = f_b(bu)\, \bar{x} f(a^n b)\\ &= f_b(buba^n)\,\bar{x} = f_b(bw)\,\bar{x}\end{aligned}$$

This proves $(\star)$.

Let now wa be a prefix of $\mathbf{c}$. Then $f(wa)$ is a prefix of $f(\mathbf{c}) = \mathbf{c}$, and consequently $x f(wa)$ is a prefix of $x\,\mathbf{c}$. Since $x f(wa) = f_a(aw)\, x$, the word $f_a(aw)$ is a prefix of $x\,\mathbf{c}$. The word $\mathbf{c}$ being Sturmian, there are infinitely many such w, and consequently $f_a(a\,\mathbf{c}) = x\,\mathbf{c}$. The same proof holds for $f_b(b\,\mathbf{c}) = \bar{x}\,\mathbf{c}$. ■

The fact that the morphisms f_a and f_b are Sturmian are also an immediate consequence of more general results of [19]. We need here a more precise statement.

THEOREM 3.3: *For any* $\mathbf{c} \in C$, *the infinite words* $ab\,\mathbf{c}$ *and* $ba\,\mathbf{c}$ *are morphic, and have the same generator.*

Proof: As in the proof of theorem 3.1, let $\mathbf{c} = f(\mathbf{c})$ and assume $|f(a)| < |f(b)|$, the other case is symmetric. Again, we do not consider the case where $f(a)$ is a letter. Then $f(a) = p\bar{x}x$, $f(b) = p\bar{x}xqx\bar{x}$ where p, q and $p\bar{x}xq$ are palindroms. Define a morphism g by

$$g: \begin{array}{l} a \mapsto x\bar{x}p \\ b \mapsto \bar{x}xqx\bar{x}p \,(= \bar{x}xp\bar{x}xq) \end{array}$$

Oberve that $g = \bar{f}$.

LEMMA 3.4: *The following relation holds:* $f(wab) = pg(bw)\,x\bar{x}$.

From this, it follows that $x\bar{x}\,\mathbf{c} = g(ab\,\mathbf{c})$ and $\bar{x}x\,\mathbf{c} = g(ba\,\mathbf{c})$. Indeed, for any prefix wab of $\mathbf{c}$ (and there are infinitely many of this kind), one has

$$x\bar{x}f(wab) = x\bar{x}pg(bw)\,x\bar{x} = g(abw)\,x\bar{x}$$

and similarly $\bar{x}xf(wab) = g(baw)\,x\bar{x}$. Thus the theorem holds for the generator g in the case $x = a$, and for the generator g^2 for $x = b$. ■

Proof of lemma 3.4: By induction on the number of factors ab in w. If this number is null, then $w = b^m a^n$, and using $(\star\star)$

$$\begin{aligned} f(wab) &= f(b^m a^n ab) = (p\bar{x}xqx\bar{x})^m (p\bar{x}x)^n p\bar{x}xp\bar{x}xqx\bar{x} \\ &= (p\bar{x}xqx\bar{x})^m p\bar{x}xqx\bar{x}p\,(x\bar{x}p)^n x\bar{x} \\ &= p\,(\bar{x}xqx\bar{x}p)^{m+1} (x\bar{x}p)^n x\bar{x} \\ &= pg(b^{m+1} a^n)\,x\bar{x} = pg(bw)\,x\bar{x} \end{aligned}$$

Next, if $wab = uabvab$, then

$$\begin{aligned} f(wab) &= f(uab)\,f(vab) = pg(bu)\,x\bar{x}pg(bv)\,x\bar{x} \\ &= pg(buabv)\,x\bar{x} = pg(bw)\,x\bar{x} \end{aligned}$$

This proves the lemma. ■

4. EXAMPLES

1. – Consider the Fibonacci morphism of the introduction. The morphisms f_a and f_b of the proof of theorem 3.1 are (with a and b interchanged):

$$f_a : \begin{array}{l} a \mapsto ab \\ b \mapsto a \end{array} \qquad f_b : \begin{array}{l} a \mapsto ba \\ b \mapsto a \end{array}$$

Thus, the morphism $f_b \circ f_a$ generates $b\mathbf{f}$ and $f_a \circ f_b$ generates $a\mathbf{f}$. The morphism g of the proof of theorem 3.3 is

$$g : \begin{array}{l} a \mapsto ab \\ b \mapsto a \end{array}$$

and the morphism g^2 generates both $ab\mathbf{f}$ and $ba\mathbf{f}$.

2. – In [18] (p. 69), A. Salomaa introduces the morphism $h : \quad a \mapsto aab$, $b \mapsto a$ in connection with equality sets. This is a positive Sturmian morphism, and $h = D \circ E \circ D$.

3. – Consider the morphism

$$f = D \circ h \circ E : \begin{array}{l} a \mapsto ab \\ b \mapsto ababa \end{array}$$

similar to a morphism used in [16]. It generates the infinite word

$$\mathbf{x} = abababaabababaabababaababababa \ldots$$

By the constructions given above, $b\mathbf{x} = f_a(a\mathbf{x})$, $a\mathbf{x} = f_b(b\mathbf{x})$, $ab\mathbf{x} = g(ba\mathbf{x})$, $ba\mathbf{x} = g(ab\mathbf{x})$ with

$$f_a : \begin{array}{l} a \mapsto ba \\ b \mapsto babaa \end{array} \qquad f_b : \begin{array}{l} a \mapsto ab \\ b \mapsto aabab \end{array} \qquad g : \begin{array}{l} a \mapsto ba \\ b \mapsto ababa \end{array}$$

Words $a\mathbf{x}$, $b\mathbf{x}$ and $ab\mathbf{x}$, $ba\mathbf{x}$ are generated by $f_b \circ f_a$, $f_a \circ f_b$ and g^2 respectively.

REFERENCES

1. J. BERSTEL and P. SÉÉBOLD, A Characterization of Sturmian Morphisms, *in:* A. BORZYSKOWSKI, S. SOKOLOWSKI (eds.) MFCS'93, *Lect. Notes Comp. Sci.*, 1993, *711*, pp. 281-290.
2. E. BOMBIERI and J. E. TAYLOR, Which Distributions of Matter Diffract? An Initial Investigation, *J. Phys.*, 1986, *47*, Colloque C3, pp. 19-28.

3. J.-P. Borel and F. Laubie, Quelques mots sur la droite projective réelle, *J. théorie des nombres de Bordeaux*, 1993, *5*, pp. 23-52.
4. T. C. Brown, Descriptions of the Characteristic Sequence of an Irrational, *Canad. Math. Bull.*, 1993, *36*, 1, pp. 15-21.
5. E. Coven and G. Hedlund, Sequences with Minimal Block Growth, *Math. Systems Theory*, 1973, *7*, pp. 138-153.
6. D. Crisp, W. Moran, A. Pollington, P. Shiue, Subsitution Invariant Cutting Sequences, *J. théorie des nombres de Bordeaux*, 1993, *5*, pp. 123-138.
7. K. Culik II and S. Dube, Rational and Affine Expressions for Image Descriptions, *Discrete Appl. Math.*, 1993, *41*, pp. 85-120.
8. K. Culik II and S. Dube, L-Systems and Mutually Recursive Function Systems, *Acta Inform.*, 1993, *30*, pp. 279-302.
9. K. Culik II and S. Dube, Encoding Images as Words and Languages, *Intern. J. Algebra Comput.*, 1993, *3*, pp. 221-236.
10. K. Culik II and T. Harju, Dominoes, Slicing Semigroups and DNA, *Discrete Appl. Math.*, 1991, *31*, pp. 261-277.
11. K. Culik II and J. Karhumäki, Iterative Devices Generating Infinite Words, *Intern. J. Algebra Comput.* (to appear).
12. K. Culik II and J. Kari, Image Compression Using Weighted Automata, *Computer and Graphics*, 1993, *17*, pp. 305-313.
13. K. Culik II and A. Salomaa, On Infinite Words Obtained by Iterating Morphisms, *Theoret. Comput. Sci.*, 1982, *19*, pp. 29-38.
14. A. De Luca and F. Mignosi, Some Combinatorial Properties of Sturmian Words, *Theoret.. Comput. Sci.* (to appear). Also Available as Technical Report LITP 93-53, october 1993.
15. G. Hedglund and M. Morse, Symbolic Dynamics II: Sturmian Sequences, *Amer. J. Math.*, 1940, *61*, pp. 1-42.
16. F. Mignosi, P. Séébold, Morphismes sturmiens et règles de Rauzy, *J. théorie des nombres de Bordeaux*, 1993, *5*, pp. 221-233.
17. A. Salomaa, Morphisms on Free Monoids and Language Theory, in *Formal Language Theory: Perspectives and Open Problems*, 1980, pp. 141-166, Academic Press.
18. A. Salomaa, *Jewels of Formal Language Theory*, Computer Science Press, 1981.
19. Z.-X. Wen and Z.-Y. Wen, Local Isomorphisms of Invertible Substitutions, *C. R. Acad. Sci. Paris* (to appear).