



**HAL**  
open science

## A local balance property of episturmian words

Gwenaël Richomme

► **To cite this version:**

| Gwenaël Richomme. A local balance property of episturmian words. 2007. hal-00130229v1

**HAL Id: hal-00130229**

**<https://hal.science/hal-00130229v1>**

Preprint submitted on 9 Feb 2007 (v1), last revised 6 Feb 2008 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LaRIA : Laboratoire de Recherche en Informatique d'Amiens  
Université de Picardie Jules Verne – CNRS FRE 2733  
33, rue Saint Leu, 80039 Amiens cedex 01, France  
Tel : (+33)[0]3 22 82 88 77  
Fax : (+33)[0]3 22 82 54 12  
<http://www.laria.u-picardie.fr>

**A local balance property of episturmian words**

**G. Richomme<sup>a</sup>**

**LaRIA RESEARCH REPORT : LRR 2007-02**

(February 2007)

---

<sup>a</sup> LaRIA, Université de Picardie Jules Verne, [gwenael.richomme@u-picardie.fr](mailto:gwenael.richomme@u-picardie.fr)

# A local balance property of episturmian words

Gwénaél Richomme

UPJV, LaRIA,  
33, Rue du Moulin Neuf  
80039 Amiens cedex 01  
gwénael.richomme@u-picardie.fr,  
<http://www.laria.u-picardie.fr/~richomme/>

**Abstract.** We prove that episturmian words and Arnoux-Rauzy sequences can be characterized using a local balance property. We also give a new characterization of epistandard words and show that the set of finite words that are not factors of an episturmian word is not context-free.

**Keywords:** Arnoux-Rauzy sequences, episturmian words, balance property.

## 1 Introduction

M. Morse and G.A. Hedlund [19] were the firsts to study in depth a family of words called Sturmian words. Now a large literature exists on these words for which have been proved numerous characterizations more fascinating the ones than the others (see for instance [1,3,20]).

Sturmian words are defined over a binary alphabet. From their various characteristic properties, some generalizations of Sturmian words have emerged over larger alphabets. One of them, the so-called Arnoux-Rauzy sequences, is based on the notion of complexity of a word and is interesting by its geometrical, arithmetic, ergodic and combinatorial aspects (see for instance [20]).

One of the first properties of Sturmian words stated by M. Morse and G.A. Hedlund [19] is the balance property: any infinite word  $w$  over the alphabet  $\{a, b\}$  is Sturmian if and only if it is non-ultimately periodic and balanced, that is the number of occurrences of the letter  $a$  differs in two factors of same length of  $w$  by at most one. Generalizations of these words were studied for instance by P. Hubert [13] (see also [23] for a survey of this property). J. Justin and L. Vuillon have stated a non-characteristic kind of balance property [15] for Arnoux-Rauzy sequences. Although it was first conjectured that Arnoux-Rauzy sequences are balanced [9], J. Cassaigne, S. Ferenczi and L.Q. Zamboni have proved that this does not necessarily hold [6].

In 1973, E.M. Coven and G.A. Hedlund [7] stated that a word  $w$  over  $\{a, b\}$  is not balanced if and only if there exists a palindrome  $t$  such that  $ata$  and  $btb$  are both factors of  $w$ . This could be seen as a local balance property of Sturmian words since to check the balance property we do not have to compare all factors

of the same length but only factors on the sets  $AtA$  for  $t$  factors of  $w$ . The previous property can be rephrased in: an infinite word  $w$  over the alphabet  $A = \{a, b\}$  is Sturmian if and only if it is non-ultimately periodic and for any factor  $t$  of  $w$ , the set of factors belonging to  $AtA$  is a subset of  $atA \cup Ata$  or a subset of  $btA \cup Atb$ . In Section 3, we show that this result can be generalized to Arnoux-Rauzy sequences contrarily to the balance property.

Actually our result concerns a larger family of infinite words presented in Section 2. Based on ideas of A. de Luca [8], Episturmian words were proposed by X. Droubay, J. Justin and G. Pirillo [9] as a generalization of Sturmian words. They have observed that Arnoux-Rauzy words are special episturmian words they called strict episturmian words. In the binary case episturmian words are the Sturmian words and the balanced periodic infinite words. Let us note that the case of remaining balanced words, namely the skew ones, have recently been generalized [11,12].

In [9], episturmian words are defined as an extension to standard episturmian words (Here we will call *epistandard* these standard episturmian words) previously introduced as a generalization of standard Sturmian words. In Section 4, we generalize to epistandard words a characterization of standard words proving a converse of a theorem in [14] and stating that an infinite word  $w$  is epistandard if and only if there exists at least two letters such that  $aw$  and  $bw$  are both episturmian. The interested reader can also consult [12] and its references for other characterizations of episturmian words using left extension in the context of an ordered alphabet.

In [10], a precision is made on the result of E.M. Coven and G.A. Hedlund, and it is proved that the set of non-balanced words is context-free. In Section 5, we show that this is no longer true for the set of words that are not factors of episturmian words.

Our last section comes back to the generalization of the local balance property introduced by E.M. Coven and G.A. Hedlund. One another way to rephrase it is: an infinite word  $w$  over the alphabet  $A = \{a, b\}$  is Sturmian if and only if it is non-ultimately periodic and for any factor  $t$  of  $w$ , the set of factors belonging to  $AtA$  is balanced. This yields a new family of words on which we give partial results.

## 2 Episturmian and epistandard words

Even if we assume the reader is familiar with combinatorics on words (see, e.g., [18]), we precise our notations. Given an alphabet  $A$  (a finite non-empty set of letters),  $A^*$  is the set of finite words over  $A$  including the empty word  $\varepsilon$ . The length of a word  $w$  is denoted by  $|w|$  and the number of occurrences of a letter  $a$  in  $w$  is denoted by  $|w|_a$ . The *mirror image* of a finite word  $w = w_1 \dots w_n$  ( $w_i \in A$ , for  $i = 1, \dots, n$ ) is the word  $w_n \dots w_1$  (the mirror image of  $\varepsilon$  is  $\varepsilon$  itself). A word equals to its mirror image is a *palindrome*. A word  $u$  is a *factor* of  $w$  if there exist words  $p$  and  $s$  such that  $w = pus$ . If  $p = \varepsilon$  (resp.  $s = \varepsilon$ ),  $u$  is a *prefix* (resp. *suffix*) of  $w$ . A word  $u$  is a *left special* (resp. *right special*) factor of  $w$  if

there exist (at least) two different letters  $a$  and  $b$  such that  $au$  and  $bu$  (resp.  $ua$  and  $ub$ ) are factors of  $w$ . A *bispecial factor* is any word which is both a left and a right special factor (see, e.g., [5]) for more informations on special factors). The set of factors of a word  $w$  will be denoted  $Fact(w)$ .

The previous notions can be extended in a natural way to any infinite words. Moreover any *ultimately periodic* infinite word can be written  $uv^\omega$  for two finite words  $u, v$  ( $v \neq \varepsilon$ ): it is then the infinite word obtained concatenating infinitely often  $v$  to  $u$ . If  $u = \varepsilon$ , the word is said *periodic*.

A word  $w$  is *episturmian* if and only if its set of factors is closed by mirror image and  $w$  contains at most one left (or equivalently right) special factor of each length. A word  $w$  is *epistandard Sturmian* or *epistandard*, if  $w$  is episturmian and all its left special factors are prefixes of  $w$ . Let us note that, in [9], epistandard words were introduced by several equivalent ways, and then episturmian words were defined as words having same set of factors than an epistandard one.

The two theorems below recall a very useful property of episturmian words which is the possibility to decompose infinitely an episturmian word using some morphisms. This property already seen for Arnoux-Rauzy sequences in [2] is related to the notion of S-adic dynamical system (see, e.g. [20] for more details). This property could be useful to get information on the structure of episturmian words (see for instance [4,16,17,22] for some uses in the binary cases).

Given an alphabet  $A$ , a *morphism*  $f$  on  $A$  is an application from  $A^*$  to  $A^*$  such that  $f(uv) = f(u)f(v)$  for any words  $u, v$  over  $A$ . A morphism on  $A$  is entirely defined by the images of elements of  $A$ .

*Episturmian morphisms* studied in [14,21] are the morphisms defined by composition of the permutation morphisms and the morphisms  $L_a$  and  $R_a$  defined, for  $a$  a letter, by

$$L_a \begin{cases} a \mapsto a \\ b \mapsto ab, \text{ if } b \neq a, \end{cases} \quad R_a \begin{cases} a \mapsto a \\ b \mapsto ba, \text{ if } b \neq a. \end{cases}$$

**Theorem 2.1.** [14] *An infinite word  $w$  is epistandard if and only if there exist an infinite sequence of infinite words  $(w^{(n)})_{n \geq 0}$  and an infinite sequence of letters  $(x_n)_{n \geq 1}$  such that  $w^{(0)} = w$  and for all  $n \geq 1$ ,  $w^{(n-1)} = L_{x_n}(w^{(n)})$ .*

**Theorem 2.2.** [14] *An infinite word  $w$  is episturmian if and only if there exist an infinite sequence of recurrent infinite words  $(w^{(n)})_{n \geq 0}$  and an infinite sequence of letters  $(x_n)_{n \geq 1}$  such that  $w^{(0)} = w$  and for all  $n \geq 0$ ,  $w^{(n-1)} = L_{x_n}(w^{(n)})$  or  $w^{(n-1)} = R_{x_n}(w^{(n)})$ .*

*Moreover,  $w$  has the same set of factors than the epistandard word directed by  $(x_n)_{n \geq 1}$ .*

The infinite sequence  $(x_n)_{n \geq 1}$  which appears in the two previous theorem is called the *directive word* of  $w$  and is denoted  $\Delta(w)$ : Actually in terms of [14], it is the directive word of the epistandard word having the same set of factors than  $w$ . Each episturmian word has a unique directive word.

It is worth noting that any episturmian word is *recurrent*, that is, each factor of  $w$  occurs infinitely often. An infinite word  $w$  is recurrent if and only if each factor of  $w$  occurs at least twice. Equivalently each factor of  $w$  occurs at a non-prefix position. Thus an infinite word  $w$  over an alphabet  $A$  is recurrent if and only if for each of its factors  $u$  the set  $AuA$  (or simply  $Au$ ) is not empty.

We denote as in [14]  $Ult(w)$  the set of letters occurring infinitely often in  $\Delta(w)$ . For  $B$  a subset of the alphabet, we introduce a new definition: we call *ultimately  $B$ -strict episturmian* any episturmian word  $w$  for which  $Ult(\Delta(w)) = B$ . Of course this notion is related to the notion of  *$B$ -strict episturmian* word (see [14, def. 2.3]) which is a ultimately  $B$ -strict episturmian word whose alphabet (the letters occurring in  $w$ ) is exactly  $B$ , and which is nothing else than an Arnoux-Rauzy sequence over  $B$ .

As shown in [9], there is a close relation between the directive word of an episturmian word and its special words. Corollary 2.5 below will show it again for ultimately strict episturmian words.

Let  $w$  be an episturmian word and  $\Delta(w) = (x_n)_{n \geq 1}$  its directive word. With notations of Theorem 2.2, for  $n \geq 1$ , we denote  $u_{n,w}$  (or simply  $u_n$ ) the word :

$$u_{n,w} = L_{x_1}(L_{x_2}(\dots(L_{x_{n-1}}(\varepsilon)x_{n-1})\dots)x_2)x_1$$

When  $n = 1$ ,  $u_{n,w} = \varepsilon$ . These words play an important role in the initial definition of episturmian word by palindromic closure (see [14, Sec. 2]). In particular, each  $u_n$  is a palindrom (see for instance [14, Lem. 2.5]). One can also observe that, if  $Ult(\Delta(w))$  contains at least two letters, then each  $u_n$  is a bispecial factor of  $w$ . Indeed for  $n \geq 1$ ,  $u_n$  is a prefix of the epistandard word  $s$  directed by  $\Delta(w)$  and so, by definition of an epistandard word, it is a left special factor of  $s$  and so of  $w$  by Theorem 2.2. Since the set of factors of  $w$  is closed by mirror image and since  $u_n$  is a palindrom,  $u_n$  is a right special factor of  $w$ . Conversely let us observe that any bispecial factor of an episturmian word is a palindrom. Indeed if  $u$  is a bispecial factor, then  $u$  and its mirror image  $\tilde{u}$  are left special factors of an infinite word containing at most one left special word of length  $|u|$ . It follows the construction of an epistandard word  $w$  by palindromic closure [9], that the the words  $u_{n,w}$  are the only palindroms prefixes of  $w$ . From what precedes, we deduce the following fact that does not seem to have been already quoted in the literature:

*Remark 2.3.* For an episturmian word  $w$  with directive word  $(x_n)_{n \geq 1}$ , a factor  $u$  is bispecial if and only if  $u = u_{n,w}$  for an integer  $n \geq 1$ .

Another result involving the palindroms  $u_n$  is:

**Theorem 2.4.** [9, Th. 6] *Let  $s$  be an epistandard word over the alphabet  $A$  with directive word  $\Delta(s) = (x_n)_{n \geq 1}$ . For  $n \geq 1$  and  $x \in A$ ,  $u_{n,s}x$  (or equivalently  $xu_{n,s}$ ) is a factor of  $s$  if and only if  $x$  belongs to  $\{x_i \mid i \geq n\}$ .*

By Theorem 2.2, an episturmian word  $w$  with a directive word  $\Delta$  has the same set of factors than the epistandard word with directive word  $\Delta$ . Hence the previous theorem is still valid for any episturmian word, and we can deduce:

**Corollary 2.5.** *Let  $w$  be an episturmian word over an alphabet  $A$  and let  $B \subseteq A$  be a set containing at least two different letters. The word  $w$  is a ultimately  $B$ -strict episturmian word if and only if for an integer  $n_0$ , each left special factor with  $|u| \geq n_0$  verifies  $Au \cap \text{Fact}(w) = Bu$ .*

*Moreover for each left special factors with  $|u| < n_0$ ,  $Bu \subseteq \text{Fact}(w)$ .*

The restriction on the cardinality of  $B$  ( $\geq 2$ ) will be used in all the rest of the paper. It is needed to have special factors of arbitrary length.

### 3 A new characterization of episturmian words

Now we give our first main result presented in the introduction as a kind of local characteristic balance property of episturmian words.

**Theorem 3.1.** *For a recurrent infinite word  $w$ , the following assertions are equivalent:*

1.  $w$  is episturmian;
2. for each factor  $u$  of  $w$ , a letter  $a$  exists such that  $AuA \cap \text{Fact}(w) \subseteq auA \cup Aua$ ;
3. for each palindromic factor  $u$  of  $w$ , a letter  $a$  exists such that  $AuA \cap \text{Fact}(w) \subseteq auA \cup Aua$ .

In the previous theorem, the letter  $a$  and the cardinality of the set  $AuA$  depends on  $u$ . This is shown for instance by the Fibonacci word  $(abaababaabaab\dots)$ , the epistandard word having  $(ab)^\omega$  as director word, for which  $A\varepsilon A \cap \text{Fact}(w) = \{aa, ab, ba\}$ ,  $AaA \cap \text{Fact}(w) = \{aab, baa\}$ ,  $AbA \cap \text{Fact}(w) = \{aba\}$ ,  $AaaA \cap \text{Fact}(w) = \{baab\}, \dots$

*Proof of Theorem 3.1.*

*Proof of 1  $\Rightarrow$  2.* Assume  $w$  is episturmian. Since the result deals only with factors of  $w$ , and since by Theorem 2.2 an episturmian word have the same set of factors than an epistandard word, without loss of generality we can assume that  $w$  is epistandard. Let  $u$  be a factor of  $w$ . Property 2 is immediate if  $u$  is not a bispecial factor of  $w$ . If  $u$  is bispecial in  $w$ , by Remark 2.3, an integer  $n \geq 1$  exists such that  $u = u_{n,w}$ . Let  $\Delta = (x_i)_{i \geq 1}$  be the directive word of  $w$ , let  $s$  (resp.  $t$ ) be the epistandard word with  $(x_i)_{i \geq n}$  (resp.  $(x_i)_{i \geq n+1}$ ) as directive word and let  $a = x_n$ . Letters occurring in  $t$  are exactly the letters of the set  $B = \{x_i \mid i \geq n+1\}$ . Since  $s = L_{x_n}(t)$ , the factors of length 2 in  $s$  are the words  $ab$  and  $ba$  with  $b \in B$ . By definition of  $\Delta$  and  $u_{n,w}$ ,  $w = L_{x_1}(L_{x_2}(\dots L_{x_{n-1}}(s)\dots))$  and  $u_{n,w} = L_{x_1}(L_{x_2}(\dots (L_{x_{n-1}}(\varepsilon)x_{n-1})\dots)x_2)x_1$ . Hence by an easy induction on  $n$ , we deduce  $AuA \cap \text{Fact}(w) = auB \cup Bua \subseteq auA \cup Aua$ .

*Proof of 2  $\Rightarrow$  1.* Assume that, for any factor  $u$  of  $w$ , a letter  $a$  exists such that  $AuA \cap \text{Fact}(w) \subseteq auA \cup Aua$ . In particular, considering the empty word, we

deduce that  $AA \cap \text{Fact}(w) \subseteq aA \cup Aa$  for a letter  $a$ . Hence, for an infinite word  $x$ ,  $w = L_a(y)$  if  $w$  starts with  $a$  and  $w = R_a(y)$  otherwise.

Let us prove that for each factor  $v$  of  $y$ ,  $AvA \cap \text{Fact}(w) \subseteq bvA \cup Avb$  for a letter  $b$ . We consider  $w = L_a(y)$  (resp.  $w = R_a(y)$ ). Let  $v$  be a factor of  $y$  and let  $u = L_a(v)a$  (resp.  $u = aR_a(v)$ ). We observe that for letters  $c, d$ , the words  $cud$  is a factor of  $w$  if and only if  $cvd$  is a factor of  $y$ . By hypothesis there exists a letter  $b$  such that  $AuA \cap \text{Fact}(w) \subseteq buA \cup Aub$ . Hence  $AvA \cap \text{Fact}(w) \subseteq bvA \cup Avb$ .

Letting  $x_1 = a$  and iterating infinitely the previous step, we get an infinite sequence of letters  $(x_i)_{i \geq 1}$  and an infinite sequence of words  $(w^{(i)})_{i \geq 0}$  such that  $w^{(0)} = w$  and for all  $i \geq 1$ ,  $w^{(i-1)} = L_{x_i}(w^{(i)})$  or  $w^{(i-1)} = R_{x_i}(w^{(i)})$ . Due to the fact that  $w$  is recurrent, each word  $w^{(i)}$  is also recurrent. By Theorem 2.2, the word  $w$  is episturmian.

The proof of  $1 \Leftrightarrow 3$  is similar to the proof of  $1 \Leftrightarrow 2$ . Actually,  $1 \Rightarrow 3$  is a particular case of  $1 \Rightarrow 2$ . When proving  $3 \Rightarrow 1$ , we need to prove in the inductive step that  $u$  is a palindrome if and only if  $v$  is a palindrome. This is stated by Lemma 2.5 in [14] : *a word  $u$  is a palindrome if and only if the word  $L_a(u)a = aR_a(u)$  is a palindrome.*

□

We end this section with few remarks concerning results that can be proved similarly.

*Remark 3.2.* Since an infinite word  $w$  over an alphabet  $A$  is recurrent if and only if for each factor of  $w$  the set  $AuA$  is not empty, we have: an infinite word is episturmian if and only if for each (resp. *palindromic*) factor  $u$  of  $w$ ,  $AuA$  is not empty and a letter  $a$  exists such that  $AuA \cap \text{Fact}(w) \subseteq auA \cup Aua$ .

*Remark 3.3.* We have already said that Arnoux-Rauzy sequences over an alphabet  $A$  are exactly the (ultimately)  $A$ -strict episturmian word. One can ask for a characterization of these words in a quite similar way than Theorem 3.1. Corollary 2.5 can fulfill this purpose. But the proof of Theorem 3.1 can also be easily reworked to state : *an episturmian word  $w$  over an alphabet  $A$  is a ultimately  $B$ -strict episturmian word with  $B \subseteq A$  if and only if for all  $n \geq 0$ , there exists a (resp. palindromic) word  $u$  of length at least  $n$  and a letter  $a$  such that  $AuA \cap \text{Fact}(w) = auB \cup Bua$ .*

*Remark 3.4.* Another adaptation of the proof of Theorem 3.1 concerns finite words: *a finite word  $w$  is a factor of an infinite episturmian word if and only if for each factor  $u$  of  $w$ , a letter  $a$  exists such that  $AuA \cap \text{Fact}(w) \subseteq auA \cup Aua$ .* We let the reader verify this result. The main difficulty of the proof is that in the “if part”, we do not have necessarily  $w = L_a(y)$  or  $w = R_a(y)$ . But we have one of the four following cases depending on the fact that  $w$  ends or not with  $a$ :  $w = L_a(y)$ ,  $w = aL_a(y)$ , or  $wa = L_a(y)$  or  $wa = aL_a(y)$ . Except in small cases, we have  $|y| < |w|$  and the technique of the proof of Theorem 3.1 can be applied.



## 4 A characterization of epistandard words

Let us note that for any episturmian word  $w$ , there exists at least one letter  $a$  such that  $aw$  is also episturmian. Indeed, since any episturmian word is recurrent, for any prefix  $p$  of  $w$ , there exists a letter  $a_p$  such that  $a_pp$  is a factor of  $w$ . We work with a finite alphabet hence an infinity of letters  $a_p$  are mutually equal: there exists a letter  $a$  such that  $ap$  is a factor of  $p$  for an infinity of prefixes (and so for all prefixes) of  $w$ . The word  $aw$  has the same set of factors than  $w$ : it is episturmian.

In restriction to epistandard words, a more precise result is already known:

**Theorem 4.1.** [14, Th. 3.17] *If a word  $s$  is epistandard, then for each letter  $a$  in  $Ult(\Delta(s))$ ,  $as$  is episturmian.*

Up to now we know the converse of this result has already been stated only in the Sturmian case (see [3, Prop. 2.1.22]): *For every Sturmian word  $w$  over  $\{a, b\}$ ,  $w$  is standard episturmian if and only if  $aw$  and  $bw$  are both Sturmian.* We generalize here this result, proving a converse to Theorem 4.1.

**Proposition 4.2.** *A non-periodic word  $w$  is epistandard if and only if, for (at least) two different letters  $a$  and  $b$ ,  $aw$  and  $bw$  are episturmian.*

*Proof.* Let  $w$  be a non-periodic epistandard word. By [9, Th. 3], we know that  $Ult(\Delta(w))$  contains at least two different letters, say  $a$  and  $b$ . By Theorem 4.1,  $aw$  and  $bw$  are episturmian.

Assume now that for two different letters  $a$  and  $b$ ,  $aw$  and  $bw$  are episturmian. Since  $aw$  (and also  $bw$ ) is recurrent,  $w$  has the same set of factors than  $aw$  and so  $w$  is episturmian. Moreover each prefix  $p$  is left special (since  $ap$  and  $bp$  are factors of  $w$ ). Since any episturmian word has at most one left special factor for each length, the left special factors of  $w$  are its prefixes:  $w$  is epistandard.  $\square$

Let us give a more precise result:

**Theorem 4.3.** *Let  $w$  be an infinite word over the alphabet  $A$  and assume  $B \subseteq A$  contains at least two different letters. The two following assertions are equivalent:*

1. *The word  $w$  is ultimately  $B$ -strict epistandard;*
2. *For each letter  $a$  in  $A$ ,  $aw$  is episturmian if and only if  $a$  belongs to  $B$ .*

*Proof.* Assume first that  $w$  is  $B$ -strict epistandard, that is,  $Ult(\Delta(w)) = B$ . By Theorem 4.1, for each letter  $a$  in  $B$ ,  $aw$  is episturmian. For any integer  $n \geq 0$ , the word  $u_{n,w}$  is a prefix of  $w$ . If  $a$  does not belong to  $B$ , by Theorem 2.4, for at least one integer  $n \geq 0$ ,  $au_{n,w}$  is not a factor of  $w$ . Thus the word  $aw$  is not recurrent and so it is not episturmian. Hence if  $w$  is  $B$ -strict epistandard, for each letter  $a$  in  $A$ ,  $aw$  is episturmian if and only if  $a$  belongs to  $B$ .

Assume now that for each letter  $a$  in  $A$ ,  $aw$  is episturmian if and only if  $a$  belongs to  $B$ . Since  $B$  contains at least two letters, by Proposition 4.2,  $w$  is epistandard. As a consequence of Theorem 2.4, we can deduce  $Ult(\Delta(w)) = B$ .  $\square$

## 5 About the set of non-episturmian finite words

Now we come to a language theoretical question about the set of words that are not factors of episturmian words. In [10] (see also [3, Pb. 2.1.4, p. 102]), it is proved: *a word  $w$  is not balanced (and so not a factor of Sturmian word) if and only if it can be written  $w = xauayb\tilde{u}bz$  for some words  $u, x, y, z$  and distinct letters  $a$  and  $b$  ( $\tilde{u}$  is the mirror image of  $u$ ).* This result was then used to state that the set of non-balanced word is context-free. We prove that this is no longer true when considering words that are not factors of episturmian words over a (at least) ternary alphabet.

Given an alphabet  $A$ , we denote  $F_A$  the set of all possible factors of episturmian words over  $A$ .

**Proposition 5.1.** *If  $A$  contains at least three letters, the set  $A^* \setminus F_A$  is not context-free.*

This result can be deduced from the following fact using the usual pumping lemma for context-free languages.

**Fact 5.2** *If  $a, b, c$  are three distinct letters in  $A$ ,*

$$F_A \cap a^*ba^*ca^* = \{a^pba^qca^r \mid p \geq q+2 \text{ or } r \geq q+2 \text{ or } p = r = q+1\}$$

*Proof.* Let us first observe that a word  $a^pba^qca^r$  with  $p \geq q+2$  or  $r \geq q+2$  or  $p = r = q+1$  contains simultaneously the factors  $a^{q+2}$  and  $ba^qc$ , or simultaneously the factors  $aa^qb$ ,  $ba^qc$  and  $ca^qa$ . In each case, there does not exist a letter  $\alpha$  such that  $Aa^nA \cap \text{Fact}(w) \subseteq \alpha a^q A \cup Aa^q\alpha$ . By Theorem 3.1, the word is not episturmian.

Moreover each word  $a^pba^qca^r$  with  $p \leq q+1, r \leq q+1$  and  $(p, r) \neq (q+1, q+1)$  is a factor of  $a^{q+1}ba^qca^qb = L_a^q R_b R_c(a)$  or of  $ca^qba^qca^{q+1} = R_a^q L_c L_b(a)$ , and so is a factor of an episturmian word.  $\square$

*Remark 5.3.* Actually the proof above is similar to the one to prove that the set of balanced words (and so the set  $F_A$  for any alphabet  $A$  containing at least two letters) is not context-free.

The great difference between the binary case and larger cases is that for any letters  $a, b, c$  and any integer  $n \geq 0$ , the word  $a^{n+1}ba^nba^{n+1}$  is balanced (and so factor of a Sturmian word), whereas the still balanced word  $a^{n+1}ba^nca^{n+1}$  is not a factor of any episturmian word.

## 6 A new family of words

In this section, we consider recurrent infinite words  $w$  over an alphabet  $A$  having the following property:

Property  $\mathcal{P}$ : for any word  $u$  over  $A$ , the set of factors of  $w$  belonging to  $AuA$  is balanced,  
that is, for any word  $u$  and for any letters  $a, b, c, d$ , if  $aub$  and  $cud$  are factors of  $w$  then  $\{a, b\} \cap \{c, d\} \neq \emptyset$ .

Any word verifying Assertion 2 in Theorem 3.1 also verifies Property  $\mathcal{P}$ . As shown by the word  $(abc)^\omega$ , the converse does not hold. In other words, any episturmian word verifies Property  $\mathcal{P}$ , but this is not a characteristic property (except in the binary case for which it is immediate that a word  $w$  verifies Property  $\mathcal{P}$  if and only if for all words  $u$ ,  $aua$  or  $bub$  is not a factor of  $w$ ).

We prove:

**Proposition 6.1.** *A recurrent word  $w$  over an alphabet  $A$  verifies property  $\mathcal{P}$  if and only if one of the two following assertion holds:*

1.  $w$  is episturmian;
2. there exist an episturmian morphism  $f$ , three different letters  $a, b, c$  in  $A$  and a word  $w'$  over  $\{a, b, c\}$  such that  $w = f(w')$ ,  $w'$  verifies Property  $\mathcal{P}$  and the three words  $ab$ ,  $bc$  and  $ca$  are factors of  $w'$ .

This proposition is a consequence of the next two lemmas.

**Lemma 6.2.** *If a recurrent infinite word  $w$  verifies property  $\mathcal{P}$ , then one of the two following assertion holds:*

1.  $w = L_\alpha(w')$  or  $w = R_\alpha(w')$  for a letter  $\alpha$  and a recurrent infinite word  $w'$ ;
2. there exist three different letters  $a, b, c$  such that  $w \in \{a, b, c\}^\omega$  and the three words  $ab$ ,  $bc$  and  $ca$  are factors of  $w$ .

*Proof.* We first observe that if  $AA \cap \text{Fact}(w) \subseteq \alpha A \cup A\alpha$  then (as in the proof of Theorem 4.3)  $w = L_\alpha(w')$  or  $w = R_\alpha(w')$ , for a letter  $\alpha$  and a recurrent infinite word  $w'$ .

We assume from now on that  $AA \cap \text{Fact}(w) \not\subseteq \alpha A \cup A\alpha$ .

For any letter  $\alpha$  in  $A$ ,  $\alpha\alpha$  is not a factor of  $w$ . Indeed if such a word is a factor of  $w$ , then, for any factor  $\beta\gamma$  with  $\beta$  and  $\gamma$  letters, by Property  $\mathcal{P}$ ,  $\beta = \alpha$  or  $\gamma = \alpha$ , that is  $AA \cap \text{Fact}(w) \subseteq \alpha A \cup A\alpha$ .

The alphabet  $A$  contains at least three letters. Indeed if  $A$  contains at most two letters  $a$  and  $b$ , then Property  $\mathcal{P}$  implies that  $aa$  and  $bb$  are not simultaneously factors of  $w$ , and so we have  $AA \cap \text{Fact}(w) \subseteq aA \cup Aa$  or  $AA \cap \text{Fact}(w) \subseteq bA \cup Ab$ .

Let us prove that  $A$  contains exactly three letters. Assume by contradiction that  $A$  contains at least four letters. Let  $a$  (resp.  $b$ ) be the first (resp. the second) letter of  $w$ . Since  $aa$  is not a factor of  $w$ ,  $a \neq b$ . At least two other letters  $c$  and  $d$  occur in  $w$  ( $c, d \notin \{a, b\}$ ,  $c \neq d$ ). By Property  $\mathcal{P}$ , each occurrence of  $c$  is preceded by  $a$  or by  $b$ . Assume that  $ac$  occurs in  $w$ . Since  $ab$  also occurs, for any letter  $\alpha$  not in  $\{a, b, c\}$ , each occurrence of  $\alpha$  is preceded and followed by the letter  $a$ . But  $AA \cap \text{Fact}(w) \not\subseteq aA \cup Aa$ . Hence  $bc$  or  $cb$  occurs in  $w$ . But then the factor  $ad$  contradicts Property  $\mathcal{P}$ . Assume now that  $bc$  occurs in  $w$ . Since  $ab$  also occurs, for any letter  $\alpha$  not in  $\{a, b, c\}$ , each occurrence of  $\alpha$  is preceded and followed by  $b$ . But  $AA \cap \text{Fact}(w) \not\subseteq bA \cup Ab$ . Hence  $ac$  or  $ca$  occurs in  $w$ . But then the factor  $db$  contradicts Property  $\mathcal{P}$ .

Until now we have proved that  $w$  is written on a three-letter alphabet and contains no word  $\alpha\alpha$  with  $\alpha$  a letter. Assume that, for two letters  $a$  and  $b$ ,  $ab$  is

a factor of  $w$  but not  $ba$ . Then for an integer  $n \geq 1$ ,  $a(bc)^n a$  (let recall that  $aa$ ,  $bb$ ,  $cc$  and  $ba$  are not factors of  $w$ ), and so  $ab$ ,  $bc$  and  $ca$  are factors of  $w$ . Now if, for all letters  $\alpha$  and  $\beta$ ,  $\alpha\beta$  and  $\beta\alpha$  are factors of  $w$  then denoting  $a$ ,  $b$  and  $c$  the letters occurring in  $w$ , once again  $ab$ ,  $bc$  and  $ca$  are factors of  $w$ .  $\square$

**Lemma 6.3.** *Let  $\alpha$  be a letter,  $w$  and  $w'$  be recurrent words such that  $w = L_\alpha(w')$  or  $w = R_\alpha(w')$ . The word  $w$  verifies Property  $\mathcal{P}$  if and only if  $w'$  verifies Property  $\mathcal{P}$ .*

*Proof.* We first assume  $w = L_\alpha(w')$ .

Assume that  $w$  does not verify Property  $\mathcal{P}$ :  $aub$  and  $cud$  are factors of  $w$  for some letters  $a, b, c, d$  and a word  $u$  such that  $\{a, b\} \cap \{c, d\} = \emptyset$ . At least one of the two letters  $a$  and  $b$  is different from  $\alpha$  and at least one of the two letters  $c$  and  $d$  is different from  $\alpha$ . Since  $w = L_\alpha(w')$ , we deduce that  $u \neq \varepsilon$ , and that  $u$  begins and ends with  $\alpha$ :  $u = L_\alpha(v)\alpha$  for a word  $v$ . Thus  $aub = aL_\alpha(v)\alpha b$  and  $cud = cL_\alpha(v)\alpha d$ . We observe that if  $a \neq \alpha$  (resp.  $c \neq \alpha$ ),  $\alpha aL_\alpha(v)\alpha b$  (resp.  $\alpha cL_\alpha(v)\alpha d$ ) is a factor of  $w$ . Thus we can deduce that  $avb$  and  $cvd$  are factors of  $w'$  (even if one of the letters  $a, b, c, d$  is  $\alpha$ ): the word  $w'$  does not verify Property  $\mathcal{P}$ .

Assume conversely that the word  $w'$  does not verify Property  $\mathcal{P}$ :  $aub$  and  $cud$  are factors of  $w'$  for some letters  $a, b, c, d$  and a word  $u$  such that  $\{a, b\} \cap \{c, d\} = \emptyset$ . The word  $aL_\alpha(u)\alpha b$  is a factor of  $w$  (if  $b = \alpha$ , this is still true since we work with infinite words and so in this case  $au\alpha b'$  is a factor of  $w$  for a letter  $b'$ ). Similarly  $cL_\alpha(u)\alpha d$  is a factor of  $w$ : the word  $w$  does not verify Property  $\mathcal{P}$ .

The proof when  $w = R_\alpha(w')$  is similar. Note that the fact that  $w'$  is recurrent is needed for the last part of the proof to know when  $a = \alpha$ , that  $a'\alpha ub$  is a factor of  $w'$  for a letter  $a'$ .  $\square$

*Proof of Proposition 6.1.* Assume  $w$  is a recurrent word that verifies Property  $\mathcal{P}$  but that does not verifies Assertion 2 of Lemma 6.2. Then  $w = L_\alpha(w')$  or  $w = R_\alpha(w')$ , with  $w'$  a recurrent word. By Lemma 6.3,  $w'$  verifies Property  $\mathcal{P}$ .

Thus using Lemmas 6.2 and 6.3, we can prove by induction that, for any integer  $n \geq 0$ , one of the two following assertions holds :

- there exist recurrent infinite word  $w^{(0)} = w, w^{(1)}, \dots, w^{(n)}$ , and letters  $a_1, \dots, a_n$  such that for each  $1 \leq p \leq n$ ,  $w^{(p-1)} = L_{a_p}(w^{(p)})$  or  $w^{(p-1)} = R_{a_p}(w^{(p)})$ , and  $w^n$  verifies property  $\mathcal{P}$ ;
- for an integer  $m \leq n$ , there exist recurrent infinite word  $w^{(0)} = w, w^{(1)}, \dots, w^{(m)}$ , and letters  $a_1, \dots, a_m$  such that for each  $1 \leq p \leq m$ ,  $w^{(p-1)} = L_{a_p}(w^{(p)})$  or  $w^{(p-1)} = R_{a_p}(w^{(p)})$ , and  $w^{(m)}$  verifies Assertion 2 of Lemma 6.2.

Hence the proposition is a consequence of Theorem 2.2.  $\square$

## 7 Conclusion

The reader has certainly noticed that words verifying Property  $\mathcal{P}$  are not completely characterized. For this, one should have to better know ternary recurrent words verifying Property  $\mathcal{P}$  and containing the words  $ab$ ,  $bc$  and  $ca$  as factors.

Let us give examples of such words. One can immediately verify that if  $ab$ ,  $bc$  and  $ca$  are the only words of length 2 that are factors of a word  $w$ , then  $w$  is  $(abc)^\omega$ ,  $(bca)^\omega$  or  $(cab)^\omega$ . When a recurrent word  $w$  verifying property  $\mathcal{P}$  has exactly the words  $ab$ ,  $bc$ ,  $ca$  and  $ba$  as factors of length 2, one can see that  $w$  is a suffix of a word  $f(w')$  where  $w'$  is a Sturmian word over  $\{a, b\}$  and  $f$  is the morphism defined by  $f(a) = (ab)^n c$  and  $f(b) = (ab)^{n+1} c$  for an integer  $n \geq 1$ . When  $f$  is replaced by one of the following morphisms  $g_1$  or  $g_2$ , we can get other examples of ternary words verifying Property  $\mathcal{P}$  (and containing exactly 5 factors of length 2 with amongst them  $ab$ ,  $bc$  and  $ca$ ):  $g_1(a) = (ab)^n c$ ,  $g_1(b) = (ab)^n cb$ ,  $g_2(a) = (ab)^n c$ ,  $g_2(b) = (ab)^{n+1} cb$ . Our final example is the periodic word  $(abcabacbabc)^\omega$  which verifies Property  $\mathcal{P}$  and contains as factors all words of length 2 except  $aa$ ,  $bb$ ,  $cc$ : this word could be seen as the morphic image of  $a^\omega$  by the morphism that maps  $a$  onto  $abcabacbabc$ .

All these examples lead to the question: Are all ternary recurrent words verifying Property  $\mathcal{P}$  and containing  $ab$ ,  $bc$  and  $ca$  as factors are suffix of a word  $f(w')$  with  $w'$  a recurrent balanced word (that is a Sturmian word or a periodic balanced word) and with  $f$  a morphism? If it is true, which are the possible values for  $f$ ?

**Acknowledgements.** The author would like to thanks J.-P. Allouche for his questions that have initiated the present work.

## References

1. J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, 2003.
2. P. Arnoux and G. Rauzy. Représentation géométrique de suites de complexités  $2n + 1$ . *Bull. Soc. Math. France*, 119:199–215, 1991.
3. J. Berstel and P. Séébold. *Algebraic Combinatorics on Words* (M. Lothaire, ed.), volume 90, chapter 2. Sturmian words. Cambridge Mathematical Library, 2002.
4. V. Berthé, C. Holton, and L. Q. Zamboni. Initial powers of sturmian sequences. *Acta arithmetica*, 122:315–347, 2006.
5. J. Cassaigne. Complexité et facteurs spéciaux. *Belg. Bull. Math. Soc.*, 4:67–88, 1997.
6. J. Cassaigne, S. Ferenczi, and L.Q. Zamboni. Imbalances in Arnoux-Rauzy sequences. *Ann. Inst. Fourier, Grenoble*, 50(4):1265–1276, 2000.
7. E. M. Coven and G.A. Hedlund. Sequences with minimal block growth. *Math. Syst. Th.*, 7:138–153, 1973.
8. A. de Luca. On standard Sturmian morphisms. *Theoretical Computer Science*, 178:205–224, 1997.
9. X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoretical Computer Science*, 255:539–553, 2001.
10. S. Dulucq and D. Gouyou-Beauchamps. Sur les facteurs des suites de Sturm. *Theoretical Computer Science*, 71:381–400, 1990.
11. A. Glen. A characterization of fine words over a finite alphabet. In *International School and Conference on Combinatorics, Automata and Number Theory (Cant'06)*. Université de Liège, Belgium, 2006. 9 pp.

12. A. Glen, J. Justin, and G. Pirillo. Characterizations of finite and infinite episturmian words via lexicographic orderings. *submitted*, 2006.
13. P. Hubert. Suites équilibrées. *Theoretical Computer Science*, 242:91–108, 2000.
14. J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoretical Computer Science*, 276(1-2):281–313, 2002.
15. J. Justin and L. Vuillon. Return words in Sturmian and episturmian words. *RAIRO Theoret. Infor. Appl.*, 34:343–356, 2000.
16. F. Levé and G. Richomme. Quasiperiodic infinite words: some answers. *Bull. Europ. Assoc. Theoret. Comput. Sci.*, 84:128–238, 2004.
17. F. Levé and G. Richomme. Quasiperiodic Sturmian words and morphisms. *Theoretical Computer Science*, To appear.
18. M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, 1983. Reprinted in the Cambridge Mathematical Library, Cambridge University Press, UK, 1997.
19. M. Morse and G.A. Hedlund. Symbolic Dynamics II: Sturmian trajectories. *Amer. J. Math.*, 61:1–42, 1940.
20. N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer, 2002. (V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel, editors).
21. G. Richomme. Conjugacy and episturmian morphisms. *Theoretical Computer Science*, 302:1-34, 2003.
22. G. Richomme. Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words. *Theoretical Computer Science*, (Words'05 special number), to appear.
23. L. Vuillon. Balanced words. *Bull. Belg. Math.Soc.*, 10(5):787–805, 2003.