



**Murdoch**  
UNIVERSITY

**MURDOCH RESEARCH REPOSITORY**

<http://researchrepository.murdoch.edu.au>

*This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination.*

**Glen, A. and Justin, J. (2009) Episturmian words: a survey. RAIRO - Theoretical Informatics and Applications, 43 (3). pp. 403-442.**

<http://researchrepository.murdoch.edu.au/3803>

Copyright © 2009 EDP Sciences  
It is posted here for your personal use. No further distribution is permitted.

# EPISTURMIAN WORDS: A SURVEY\*

Amy Glen<sup>†</sup>

Jacques Justin<sup>‡</sup>

Submitted: December 11, 2007; Updated: February 3, 2009

---

## Abstract

In this paper, we survey the rich theory of infinite *episturmian words* which generalize to any finite alphabet, in a rather resembling way, the well-known family of *Sturmian words* on two letters. After recalling definitions and basic properties, we consider *episturmian morphisms* that allow for a deeper study of these words. Some properties of factors are described, including factor complexity, palindromes, fractional powers, frequencies, and return words. We also consider lexicographical properties of episturmian words, as well as their connection to the balance property, and related notions such as finite episturmian words, Arnoux-Rauzy sequences, and “episkew words” that generalize the skew words of Morse and Hedlund.

**Keywords:** combinatorics on words; episturmian words; Arnoux-Rauzy sequences; Sturmian words; episturmian morphisms.

MSC (2000): 68R15.

---

## 1 Introduction

### 1.1 From Sturmian to episturmian

Most renowned amongst the branches of combinatorics on words is the theory of infinite binary sequences called *Sturmian words*, which are fascinating in many respects, having been studied from combinatorial, algebraic, and geometric points of view. Their beautiful properties are related to many fields such as Number Theory, Geometry, Symbolic Dynamical Systems, Theoretical Physics, and Theoretical Computer Science (see [7, 83, 96] for recent surveys).

Since the seminal works of Morse and Hedlund [91], Sturmian words have been shown to admit numerous equivalent definitions and characterizations. For instance, it is well known that an infinite word  $\mathbf{w}$  over  $\{a, b\}$  is Sturmian if and only if  $\mathbf{w}$  is aperiodic and *balanced*: for any two factors  $u, v$  of  $\mathbf{w}$  of the same length, the number of  $a$ 's in each of  $u$  and  $v$  differs by at most 1. Sturmian words are also characterized by their *factor complexity* function (which counts the number of distinct factors of each length): they have exactly  $n + 1$  distinct factors of length  $n$  for each  $n$ . In this sense, Sturmian words are precisely the aperiodic infinite words of minimal factor complexity since, as is well known, an infinite word is ultimately periodic if and only if it has less than  $n + 1$  factors of length  $n$  for some  $n$  (see [37]). Many interesting properties of

---

\*This paper grew out of an invited lecture given by the second author at the Sixth International Conference on Words, Marseille, France, September 17–21, 2007.

<sup>†</sup>Corresponding author: LaCIM, Université du Québec à Montréal, C.P. 8888, succursale Centre-ville, Montréal, Québec, H3C 3P8, CANADA \ The Mathematics Institute, Reykjavik University, Kringlan 1, IS-103 Reykjavik, ICELAND (amy.glen@gmail.com).

<sup>‡</sup>LIAFA, Université Paris Diderot - Paris 7, Case 7014, 75205 Paris Cedex 13, FRANCE (jacjustin@free.fr).

Sturmian words can be attributed to their low complexity, which induces certain regularities in such words without, however, making them periodic. Sturmian words can also be geometrically realized as *cutting sequences* by considering the sequence of ‘cuts’ in an integer grid made by a line of irrational slope (see for instance [38, 13]). They also provide a symbolic coding of the orbit of a point on a circle with respect to a rotation by an irrational number (see [91, 4]).

All of the above characteristic properties of Sturmian words lead to natural generalizations on arbitrary finite alphabets. In one direction, the balance property naturally extends to an alphabet with more than two letters (e.g., see [68, 110, 115]) as does the following *generalized balance property* that also characterizes Sturmian words (see [49, 1]): the difference between the number of occurrences of a word  $u$  in any pair of factors of the same length is at most 1. In another direction, we could consider relaxing the minimality condition for the factor complexity  $p(n)$ . For example, *quasi-Sturmian words* are infinite words for which there exist two positive integers  $N$  and  $c$  such that  $n+1 \leq p(n) \leq n+c$  for all  $n \geq N$ . This generalization was introduced in [5] when studying the transcendence of certain continued fraction expansions. See also [31, 36, 66, 105] for similar extensions of Sturmian words with respect to factor complexity. From the geometric point of view, cutting sequences naturally generalize to trajectories in the hypercube billiard (e.g., see [25]), and codings of rotational orbits carry over to codings of interval exchange transformations (e.g., see [18]).

Two other very interesting natural generalizations of Sturmian words are *Arnoux-Rauzy sequences* [12, 97] and *episturmian words* [43, 73], which we will now define.

From the factor complexity of Sturmian words, it immediately follows that any Sturmian word is over a 2-letter alphabet and has exactly one *left special* factor of each length. A factor  $u$  of a finite or infinite word  $w$  is said to be *left special* (resp. *right special*) in  $w$  if there exists at least two distinct letters  $a, b$  such that  $au$  and  $bu$  (resp.  $ua, ub$ ) are factors of  $w$ . Extending the left special property of Sturmian words, a recurrent infinite word  $\mathbf{w}$  over a finite alphabet  $\mathcal{A}$  is said to be an *Arnoux-Rauzy sequence* (or a *strict episturmian word*) if it has exactly one left special factor and one right special factor of each length, and for every left (resp. right) special factor  $u$  of  $\mathbf{w}$ ,  $xu$  (resp.  $ux$ ) is a factor of  $\mathbf{w}$  for all letters  $x \in \mathcal{A}$ . A notable property that is shared by Sturmian words and Arnoux-Rauzy sequences is their closure under reversal, i.e., if  $u$  is a factor of such a word, then its reversal is also a factor. This nice property inspired Droubay, Justin, and Pirillo’s generalization of Sturmian words in [43]: an infinite word is *episturmian* if it is closed under reversal and has at most one left special factor of each length. Sturmian, Arnoux-Rauzy, and episturmian words all have *standard* (or *characteristic*) elements, which are those having all of their left special factors as prefixes. Within these families of words, standard words are good representatives in the sense that an infinite word belongs to one such family if and only if it has the same set of factors as some standard word in that family.

From the definitions, it is clear that the family of Arnoux-Rauzy sequences is a particular subclass of the family of episturmian words. More precisely, episturmian words are composed of the Arnoux-Rauzy sequences, images of the Arnoux-Rauzy sequences by *episturmian morphisms*, and certain periodic infinite words (see Section 5). In the 2-letter case, Arnoux-Rauzy sequences are exactly the Sturmian words whereas episturmian words include all recurrent balanced words, i.e., periodic balanced words and Sturmian words.

The study of episturmian words and Arnoux-Rauzy sequences has enjoyed a great deal of popularity in recent times, owing mostly to the many properties that they share with Sturmian words. In this paper we survey the purely combinatorial work on episturmian words, beginning with their definition and basic properties in Section 2. Then, in Section 3, we recall episturmian morphisms which allow for a deeper study of episturmian words. In particular, any episturmian

word is the image of another episturmian word by some so-called *pure episturmian morphism*. Even more, any episturmian word can be infinitely decomposed over the set of pure episturmian morphisms. This last property allows an episturmian word to be defined by one of its morphic decompositions or, equivalently, by a certain *directive word*, which is an infinite sequence of rules for decomposing the given episturmian word by morphisms. In Section 4 we consider notions such as *shifts*, *spins*, and *block-equivalence* in connection with directive words, which allow us to study when two different *spinned infinite words* direct the same episturmian word. We also consider periodic and purely morphic episturmian words. In Section 5, our discussion briefly turns to Arnoux-Rauzy sequences and *finite episturmian words*. Following this, we study in Section 6 some properties of factors of episturmian words (and Arnoux-Rauzy sequences), including factor complexity, palindromes, fractional powers, frequencies, and return words. Lastly, we consider more recent work involving *lexicographic order* and the balance property (including *Fraenkel's conjecture*).

## 1.2 Notation & terminology

We assume the reader is familiar with combinatorics on words and morphisms (e.g., see [82, 83]). In this section, we recall some basic definitions and properties relating to episturmian words which are needed throughout the paper. For the most part, we follow the notation and terminology of [43, 73, 75, 62].

Let  $\mathcal{A}$  denote a finite *alphabet*, i.e., a non-empty finite set of symbols called *letters*. A finite *word* over  $\mathcal{A}$  is a finite sequence of letters from  $\mathcal{A}$ . The *empty word*  $\varepsilon$  is the empty sequence. Under the operation of concatenation, the set  $\mathcal{A}^*$  of all finite words over  $\mathcal{A}$  is a *free monoid* with identity element  $\varepsilon$  and set of generators  $\mathcal{A}$ . The set of *non-empty* words over  $\mathcal{A}$  is the *free semigroup*  $\mathcal{A}^+ := \mathcal{A}^* \setminus \{\varepsilon\}$ .

A *right-infinite* (resp. *left-infinite*, *bi-infinite*) word over  $\mathcal{A}$  is a sequence indexed by  $\mathbb{N}^+$  (resp.  $\mathbb{Z} \setminus \mathbb{N}^+$ ,  $\mathbb{Z}$ ) with values in  $\mathcal{A}$ . For instance, a left-infinite word is represented by  $\mathbf{u} = \cdots b_{-2}b_{-1}b_0$  and a right-infinite word by  $\mathbf{v} = b_1b_2b_3 \cdots$  where  $b_i \in \mathcal{A}$ . The concatenation of  $\mathbf{u}$  and  $\mathbf{v}$  gives the bi-infinite word  $\mathbf{u.v} = \cdots b_{-2}b_{-1}b_0.b_1b_2b_3 \cdots$  with a dot written between  $b_0$  and  $b_1$  to avoid ambiguity. For easier reading, infinite words are hereafter typically typed in boldface to distinguish them from finite words.

The *shift map*  $T$  is defined for bi-infinite words  $\mathbf{b} = (b_i)_{i \in \mathbb{Z}}$  by  $T(\mathbf{b}) = (b_{i+1})_{i \in \mathbb{Z}}$  and its  $k$ -th iteration is denoted by  $T^k$ . This extends to right-infinite words for  $k \geq 0$  and left-infinite words for  $k \leq 0$ . For finite words  $w \in \mathcal{A}^*$ , the shift map  $T$  acts circularly, i.e., if  $w = xv$  where  $x \in \mathcal{A}$ , then  $T(w) = vx$ .

The set of all right-infinite words over  $\mathcal{A}$  is denoted by  $\mathcal{A}^\omega$ , and we define  $\mathcal{A}^\infty := \mathcal{A}^* \cup \mathcal{A}^\omega$ . An *ultimately periodic* right-infinite word can be written as  $uv^\omega = uvvv \cdots$ , for some  $u, v \in \mathcal{A}^*$ ,  $v \neq \varepsilon$ . If  $u = \varepsilon$ , then such a word is *periodic*. A right-infinite word that is not ultimately periodic is said to be *aperiodic*.

Given a finite word  $w = x_1x_2 \cdots x_m \in \mathcal{A}^*$  with each  $x_i \in \mathcal{A}$ , the *length* of  $w$ , denoted by  $|w|$ , is equal to  $m$ . By convention, the empty word  $\varepsilon$  is the unique word of length 0. The number of occurrences of a letter  $a$  in  $w$  is denoted by  $|w|_a$ . If  $|w|_a = 0$ , then  $w$  is said to be *a-free*. The *reversal*  $\tilde{w}$  of  $w$  is its mirror image:  $\tilde{w} = x_mx_{m-1} \cdots x_1$ , and if  $w = \tilde{w}$ , then  $w$  is called a *palindrome*. The reversal operator naturally extends to bi-infinite words; that is, the reversal of the bi-infinite word  $\mathbf{b} = \mathbf{l.r}$ , with  $\mathbf{l}$  left-infinite and  $\mathbf{r}$  right-infinite, is given by  $\tilde{\mathbf{b}} = \tilde{\mathbf{r.l}}$ .

A finite word  $w$  is a *factor* of a finite or infinite word  $z$  if  $z = u w v$  for some words  $u, v$  (which are finite or infinite depending on  $z$ ). In the special case  $u = \varepsilon$  (resp.  $v = \varepsilon$ ), we call  $w$  a *prefix*

(resp. *suffix*) of  $z$ . We use the notation  $p^{-1}w$  (resp.  $ws^{-1}$ ) to indicate the removal of a prefix  $p$  (resp. suffix  $s$ ) of a finite word  $w$ . Note that a prefix or suffix  $u$  of a finite word  $w$  is said to be *proper* if  $u \neq w$ . A factor  $u$  of a finite or infinite word  $w$  is *right* (resp. *left*) *special* if  $ua$ ,  $ub$  (resp.  $au$ ,  $bu$ ) are factors of  $w$  for some letters  $a, b \in \mathcal{A}$ ,  $a \neq b$ .

For any finite or infinite word  $w$ ,  $F(w)$  denotes the set of all its factors. Moreover, the *alphabet* of  $w$  is  $\text{Alph}(w) := F(w) \cap \mathcal{A}$  and, if  $w$  is infinite, we denote by  $\text{Ult}(w)$  the set of all letters occurring infinitely often in  $w$ . Any two infinite words  $\mathbf{x}$ ,  $\mathbf{y}$  are said to be *factor-equivalent* if  $F(\mathbf{x}) = F(\mathbf{y})$ , i.e., if  $\mathbf{x}$  and  $\mathbf{y}$  have the same set of factors.

A factor of an infinite word  $\mathbf{x}$  is *recurrent* in  $\mathbf{x}$  if it occurs infinitely often in  $\mathbf{x}$ , and  $\mathbf{x}$  itself is said to be *recurrent* if all of its factors are recurrent in it. For a bi-infinite word to be recurrent, any factor must occur infinitely often to the left and to the right. An infinite word is said to be *uniformly recurrent* if any factor occurs infinitely many times in it with bounded gaps [37].

A *morphism*  $\varphi$  on  $\mathcal{A}$  is a map from  $\mathcal{A}^*$  to  $\mathcal{A}^*$  such that  $\varphi(uv) = \varphi(u)\varphi(v)$  for any words  $u, v$  over  $\mathcal{A}$ . A morphism on  $\mathcal{A}$  is entirely defined by the images of letters in  $\mathcal{A}$ . All morphisms considered in this paper will be *non-erasing*: the image of any non-empty word is never empty. Hence the action of a morphism  $\varphi$  on  $\mathcal{A}^*$  can be naturally extended to infinite words; that is, if  $\mathbf{x} = x_1x_2x_3 \cdots \in \mathcal{A}^\omega$ , then  $f(\mathbf{x}) = f(x_1)f(x_2)f(x_3) \cdots$ . An infinite word  $\mathbf{x}$  can therefore be a *fixed point* of a morphism  $\varphi$ , i.e.,  $\varphi(\mathbf{x}) = \mathbf{x}$ . If  $\varphi$  is a (non-erasing) morphism such that  $\varphi(a) = aw$  for some letter  $a \in \mathcal{A}$  and  $w \in \mathcal{A}^+$ , then  $\varphi^n(a)$  is a proper prefix of the word  $\varphi^{n+1}(a)$  for each  $n \in \mathbb{N}$ , and the limit of the sequence  $(\varphi^n(a))_{n \geq 0}$  is the unique infinite word:

$$\mathbf{w} = \lim_{n \rightarrow \infty} \varphi^n(a) = \varphi^\omega(a) (= aw\varphi(w)\varphi^2(w)\varphi^3(w) \cdots).$$

Clearly,  $\mathbf{w}$  is a fixed point of  $\varphi$  and we say that  $\mathbf{w}$  is *generated* by  $\varphi$ . Furthermore, an infinite word generated by a morphism is said to be *purely morphic*.

In what follows, we will denote the composition of morphisms by juxtaposition as for concatenation of words.

## 2 Definitions & basic properties

In the initiating paper [43], episturmian words were defined as an extension of *standard episturmian words*, which were first introduced as a generalization of *standard* (or *characteristic*) Sturmian words using iterated palindromic closure (a construction due to de Luca [41]). Here we choose instead to begin with the following definition for deriving the main basic properties of episturmian words.

**Definition 2.1.** [43] *An infinite word  $\mathbf{t} \in \mathcal{A}^\omega$  is episturmian if  $F(\mathbf{t})$  is closed under reversal and  $\mathbf{t}$  has at most one left special factor (or equivalently, right special factor) of each length. Moreover, an episturmian word is standard if all of its left special factors are prefixes of it.*

*Note.* We can equivalently consider left or right special factors in the first part of the above definition since, by closure under reversal, a factor is left (resp. right) special if and only if its reversal is right (resp. left) special.

**Remark 2.2.** When  $|\mathcal{A}| = 2$ , Definition 2.1 gives the (aperiodic) Sturmian words, as well as the periodic balanced infinite words (also known as the *periodic Sturmian words*). See for instance [62] or Section 7.1.

The following theorem collects together some useful characteristic properties of standard episturmian words. Before stating it, let us first recall the some definitions.

Given two palindromes  $p, q$ , we say that  $q$  is a *central factor* of  $p$  if  $p = wq\tilde{w}$  for some  $w \in \mathcal{A}^*$ . The *palindromic right-closure*  $w^{(+)}$  of a finite word  $w$  is the (unique) shortest palindrome having  $w$  as a prefix (see [41]). That is,  $w^{(+)} = wv^{-1}\tilde{w}$  where  $v$  is the longest palindromic suffix of  $w$ . For example,  $(race)^{(+)} = racecar$ . The *iterated palindromic closure* function [71], denoted by  $Pal$ , is defined recursively as follows. Set  $Pal(\varepsilon) = \varepsilon$  and, for any word  $w$  and letter  $x$ , define  $Pal(wx) = (Pal(w)x)^{(+)}$ . For instance,  $Pal(abc) = (Pal(ab)c)^{(+)} = (abac)^{(+)} = abacaba$ . (See Sections 4.1 and 6.2.1 for further insight about palindromic closure.)

**Theorem 2.3.** *For an infinite word  $\mathbf{s} \in \mathcal{A}^\omega$ , the following properties are equivalent.*

- i)  $\mathbf{s}$  is standard episturmian.*
- ii) Any first occurrence of a palindrome in  $\mathbf{s}$  is a central factor of some palindromic prefix of  $\mathbf{s}$  (property  $P_i$ ).*
- iii) If  $w$  is a prefix of  $\mathbf{s}$ , then  $w^{(+)}$  is also a prefix of  $\mathbf{s}$  (property  $Al$ ).*
- iv) There exists an infinite word  $\Delta = x_1x_2\cdots$  ( $x_i \in \mathcal{A}$ ), called the directive word of  $\mathbf{s}$ , such that  $\mathbf{s} = \lim_{n \rightarrow \infty} Pal(x_1 \cdots x_n)$ .*

**Remark 2.4.** The palindromes  $Pal(x_1 \cdots x_n)$  are very often denoted by  $u_{n+1}$  in the literature (and we will sometimes use the latter notation when convenient). By construction, these palindromes are exactly the palindromic prefixes of  $\mathbf{s}$ . Moreover,  $\mathbf{s}$  is uniquely determined by the directive word  $\Delta$ .

*Proof of Theorem 2.3.  $i) \Rightarrow ii)$ :* Let  $\mathbf{s} = u\mathbf{p}\mathbf{t}$ ,  $u \in \mathcal{A}^*$ ,  $\mathbf{t} \in \mathcal{A}^\omega$  showing the first occurrence of some palindrome  $p$  in  $\mathbf{s}$ . Suppose  $p$  is not the central factor of a palindromic prefix. Then we have  $\mathbf{s} = vxwp\tilde{w}y\mathbf{t}'$ ,  $x \neq y \in \mathcal{A}$ . By the reversal property,  $ywp\tilde{w}x \in F(\mathbf{s})$ , thus  $wp\tilde{w}$  is left special, hence is a prefix of  $\mathbf{s}$ . Thus  $p$  has another occurrence strictly on the left of the considered one, a contradiction.

*$i) \Rightarrow iii)$ :* If  $iii)$  is false, let  $w = ux$ , with  $u \in \mathcal{A}^*$  and  $x \in \mathcal{A}$ , be the shortest prefix of  $\mathbf{s}$  such that  $w^{(+)}$  is not a prefix of  $\mathbf{s}$ . Thus  $u^{(+)}$  is a prefix of  $\mathbf{s}$ . If  $u$  were not a palindrome then  $w$  would be a prefix of  $u^{(+)}$ ; whence  $w^{(+)} = u^{(+)}$ , a contradiction. Thus  $u$  is a palindrome. Now let  $q$  be the longest palindromic suffix of  $w$ . Then  $w^{(+)} = w_1q\tilde{w}_1 = w\tilde{w}_1$  where  $w = w_1q$ , and  $w^{(+)} = w_1qfyz$  and  $w_1qfz$  is a prefix of  $\mathbf{s}$  for some  $y \neq z \in \mathcal{A}$ ,  $f, g \in \mathcal{A}^*$ . Hence  $y\tilde{f}q \in F(\tilde{w}) \subset F(\mathbf{s})$  and  $z\tilde{f}q \in F(\mathbf{s})$ . Therefore  $\tilde{f}q$  is a left special prefix of  $\mathbf{s}$ . As  $qf$  is a prefix of  $\tilde{w} = xu$ ,  $x^{-1}qf$  is a prefix of  $u$ , hence  $x^{-1}qf\alpha$  is a prefix of  $u$  for some letter  $\alpha$ . So we have  $x^{-1}qf\alpha = \tilde{f}q$ , whence  $\alpha = x$  and  $qfx = x\tilde{f}q$ . This word is a palindrome and, as it is a suffix of  $w$ , this contradicts the minimality of  $|q|$ .

*$iii) \Rightarrow iv)$ :* Trivial.

At this stage, we have proved that standard episturmian words satisfy  $ii), iii), iv)$ . The equivalence of these three properties is proved in [43, Theorem 1]. Finally, if  $\mathbf{s}$  satisfies them, then  $F(\mathbf{s})$  is closed under reversal and by [43, Proposition 5] all of its left special factors are prefixes of it, thus  $\mathbf{s}$  is standard episturmian.  $\square$

**Remark 2.5.** Hereafter, we adopt “epistandard” as a shortcut for “standard episturmian”, as in [64, 99, 101]. Also, unless stated otherwise, the notation  $\Delta = x_1x_2x_3\cdots$  ( $x_i \in \mathcal{A}$ ) will remain for the directive word of an epistandard word  $\mathbf{s}$ .





(see [43, Theorem 7]); such words are exactly the  $k$ -letter *Arnoux-Rauzy sequences*, the study of which began in [12] (see also [74, 105] for example). In particular, the 2-strict episturmian words correspond to the (aperiodic) Sturmian words. Arnoux-Rauzy sequences will be discussed further in Section 5.

**Remark 2.7.** A noteworthy fact is that an episturmian word is periodic if and only if  $|\text{Ult}(\Delta)| = 1$  (see [73, Proposition 2.9]). The exact form of a periodic episturmian word is given by Theorem 4.15 in Section 4.4. We first need to consider *episturmian morphisms*.

### 3 Episturmian morphisms

From Lemma 4 in [43], if  $\mathbf{s}$  is epistandard with first letter  $a = x_1$ , then  $a$  is separating for  $\mathbf{s}$  and its factors, i.e., any factor of  $\mathbf{s}$  of length 2 contains the letter  $a$ . Any episturmian word  $\mathbf{t}$  that is factor-equivalent to  $\mathbf{s}$  also has separating letter  $a$ , and hence can be factorized with a code:

$$\begin{cases} \{a\} \cup a(\mathcal{A} \setminus \{a\}) & \text{if } \mathbf{t} \text{ begins with } a, \\ \{a\} \cup (\mathcal{A} \setminus \{a\})a & \text{otherwise.} \end{cases}$$

This leads to *episturmian morphisms*, which were introduced by Justin and Pirillo [73] in order to study deeper properties of episturmian words. As we shall see in Section 3.2, episturmian morphisms on an alphabet consisting of at least two letters are precisely the morphisms that preserve the set of aperiodic episturmian words (i.e., the morphisms that map aperiodic episturmian words onto aperiodic episturmian words). Such morphisms naturally generalize to any finite alphabet the *Sturmian morphisms* on two letters. A morphism  $\varphi$  is said to be *Sturmian* if  $\varphi(\mathbf{s})$  is Sturmian for any Sturmian word  $\mathbf{s}$ . The set of Sturmian morphisms over  $\{a, b\}$  is closed under composition, and consequently it is a submonoid of the endomorphisms of  $\{a, b\}^*$ . Moreover, it is well known that the monoid of Sturmian morphisms is generated by the three morphisms:  $(a \mapsto ab, b \mapsto a)$ ,  $(a \mapsto ba, b \mapsto a)$ ,  $(a \mapsto b, b \mapsto a)$  and that Sturmian morphisms are precisely the morphisms that map Sturmian words onto Sturmian words (see [16, 87]).

#### 3.1 Generators & monoids

By definition (see [43, 73]), the monoid of all *episturmian morphisms*  $\mathcal{E}$  is generated, under composition, by all the morphisms:

- $\psi_a$ :  $\psi_a(a) = a$ ,  $\psi_a(x) = ax$  for any letter  $x \neq a$ ;
- $\bar{\psi}_a$ :  $\bar{\psi}_a(a) = a$ ,  $\bar{\psi}_a(x) = xa$  for any letter  $x \neq a$ ;
- $\theta_{ab}$ : exchange of letters  $a$  and  $b$ .

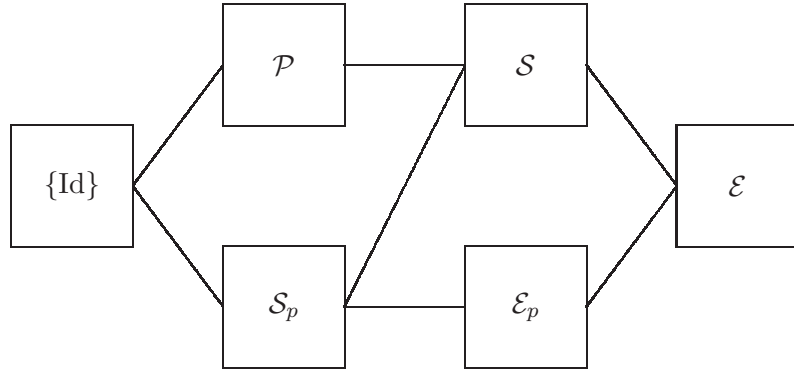
*Note.* This system of generators is far from minimal, e.g.,  $\psi_a = \theta_{ab}\psi_b\theta_{ab}$ , but gives simpler formulae.

Moreover, the monoid of so-called *epistandard morphisms*  $\mathcal{S}$  is generated by all the  $\psi_a$  and the  $\theta_{ab}$ , and the monoid of *pure episturmian morphisms*  $\mathcal{E}_p$  (resp. *pure epistandard morphisms*  $\mathcal{S}_p$ ) is generated by the  $\psi_a$  and  $\bar{\psi}_a$  only (resp. the  $\psi_a$  only). The monoid  $\mathcal{P}$  of the *permutation morphisms* (i.e., the morphisms  $\varphi$  such that  $\varphi(\mathcal{A}) = \mathcal{A}$ ) is generated by all the  $\theta_{ab}$ . The importance of the monoid of pure episturmian morphisms will become clearer in the next section where we shall see that such morphisms are strongly linked to *spinned directive words* of episturmian words, which



can be viewed as infinite sequences of rules for decomposing episturmian words by morphisms (see Theorems 3.1 and 3.3, to follow). In particular, any episturmian word is the image of another episturmian word by some pure episturmian morphism.

The following diagram illustrates the inclusions between the monoids defined above.



$$\text{Semidirect products: } \mathcal{S} = \mathcal{S}_p \times \mathcal{P}, \quad \mathcal{E} = \mathcal{E}_p \times \mathcal{P}$$

We note in particular that the monoid  $\mathcal{E}$  is a semidirect product of the submonoids of its pure morphisms and of its permutations. Consequently, any episturmian morphism  $\varphi \in \mathcal{E}$  can be expressed in a unique way as  $\varphi = \pi\mu = \mu'\pi$ , where  $\mu, \mu'$  are pure episturmian morphisms and  $\pi$  is a permutation.

*Note.* The episturmian morphisms are exactly the Sturmian morphisms when  $|\mathcal{A}| = 2$ .

Clearly, all episturmian morphisms on  $\mathcal{A}$  can be viewed as automorphisms of the free group generated by  $\mathcal{A}$  (e.g., see [57, 65, 99, 116]) and it follows that they are injective and that the monoids  $\mathcal{E}$  and  $\mathcal{S}$  are *left cancellative* (see [99, Lemma 7.2]) which means that for any episturmian morphisms  $f, g, h$ , if  $fg = fh$  then  $g = h$ . Other fundamental properties of episturmian morphisms will be discussed in the next section and in Section 4. For an in-depth study of some further properties of these morphisms, the interested reader is referred to Richomme's paper [99], in which he considers invertibility, presentation, cancellativity, unitarity, characterization by conjugacy, and so on. Most of the results in [99] naturally generalize those already known for Sturmian morphisms, but some new ones are also proved, such as a characterization of episturmian morphisms that preserve palindromes. In [100, 103], Richomme also characterized the episturmian morphisms that preserve finite and infinite *Lyndon words* and those that preserve a lexicographic order on words.

### 3.2 Relation with episturmian words

We now state two insightful characterizations of epistandard and episturmian words, which show that any episturmian word can be *infinitely decomposed* over the set of pure episturmian morphisms.

In the 'standard' case:

**Theorem 3.1.** [73, Corollary 2.7] *An infinite word  $\mathbf{s} \in \mathcal{A}^\omega$  is **epistandard** if and only if there exists an infinite word  $\Delta = x_1x_2\cdots$  over  $\mathcal{A}$  and a sequence  $(\mathbf{s}^{(i)})_{i \geq 0}$  of recurrent infinite words such that  $\mathbf{s}^{(0)} = \mathbf{s}$  and  $\mathbf{s}^{(i-1)} = \psi_{x_i}(\mathbf{s}^{(i)})$  for  $i > 0$ .  $\square$*

In [73], Justin and Pirillo showed that the infinite word  $\Delta$  appearing in the above theorem is exactly the directive word of  $\mathbf{s}$  that arises from the equivalent definition of epistandard words given in Theorem 2.3. In the binary case, the directive word  $\Delta$  is related to the continued fraction expansion of the slope of the straight line represented by a standard word (see Chapter 2 in [83]).

**Example 3.2.** Recall the Tribonacci word  $\mathbf{r}$ , which has directive word  $\Delta = (abc)^\omega$ . We have  $\mathbf{r} = \psi_a(\mathbf{r}^{(1)})$ , where  $\mathbf{r}^{(1)}$  is directed by  $T(\Delta) = (bca)^\omega$ . Notice that  $\mathbf{r}^{(1)} = \pi(\mathbf{r})$  with  $\pi = (abc)$ ; a very particular case.

More generally, the following result (Theorem 3.3) extends the notion of a directive word to all episturmian words. Before stating the theorem, we need to introduce some more notation. First we define a new alphabet,  $\bar{\mathcal{A}} := \{\bar{x} \mid x \in \mathcal{A}\}$ . A letter  $\bar{x} \in \bar{\mathcal{A}}$  is considered to be  $x$  with *spin* 1, whilst  $x$  itself has *spin* 0. The notion of a spin provides a convenient way to call upon the elementary pure episturmian morphisms  $\psi_x$  and  $\bar{\psi}_x$ . Moreover, as well shall see in Section 4, it allows us to derive many properties of episturmian words from episturmian morphisms (as a consequence of the next theorem). This approach is used for instance in [23, 60, 81, 101, 102, 105] and of course in the papers of Justin *et al.*

A finite or infinite word over  $\mathcal{A} \cup \bar{\mathcal{A}}$  is said to be a *spinned* word. Given a finite or infinite word  $w = x_1x_2\cdots$  over  $\mathcal{A}$ , we sometimes denote by  $\check{w} = \check{x}_1\check{x}_2\cdots$  any spinned word such that  $\check{x}_i = x_i$  if  $x_i$  has spin 0 and  $\check{x}_i = \bar{x}_i$  if  $x_i$  has spin 1. Such a word  $\check{w}$  is called a *spinned version* of  $w$ .

**Theorem 3.3.** [73, Theorem 3.10] *An infinite word  $\mathbf{t} \in \mathcal{A}^\omega$  is episturmian if and only if there exists a spinned infinite word  $\check{\Delta} = \check{x}_1\check{x}_2\check{x}_3\cdots$  over  $\mathcal{A} \cup \bar{\mathcal{A}}$  and an infinite sequence  $(\mathbf{t}^{(i)})_{i \geq 0}$  of recurrent infinite words such that*

$$\mathbf{t}^{(0)} = \mathbf{t} \quad \text{and} \quad \mathbf{t}^{(i-1)} = \psi_{x_i}(\mathbf{t}^{(i)}) \quad \text{or} \quad \mathbf{t}^{(i-1)} = \bar{\psi}_{x_i}(\mathbf{t}^{(i)}) \quad \text{for all } i > 0,$$

according to the spin 0 or 1 of  $\check{x}_i$ , respectively.

For any epistandard word (resp. episturmian word)  $\mathbf{t}$  and infinite word  $\Delta$  (resp. spinned infinite word  $\check{\Delta}$ ) satisfying the conditions of the Theorem 3.1 (resp. Theorem 3.3), we say that  $\Delta$  (resp.  $\check{\Delta}$ ) is a *directive word* (resp. a *spinned directive word*) for  $\mathbf{t}$  or  $\mathbf{t}$  is *directed by*  $\Delta$  (resp.  $\check{\Delta}$ ).

**Remark 3.4.** It follows immediately from Theorem 3.3 that if  $\mathbf{t}$  is an episturmian word directed by a spinned infinite word  $\check{\Delta}$ , then each  $\mathbf{t}^{(n)}$  (as defined in the theorem) is an episturmian word directed by  $T^n(\check{\Delta}) = \check{x}_{n+1}\check{x}_{n+2}\check{x}_{n+3}\cdots$ .

The following important fact links Theorems 3.1 and 3.3.

**Remark 3.5.** [73] If  $\mathbf{t}$  is an episturmian word directed by a spinned version  $\check{\Delta}$  of an infinite word  $\Delta$  over  $\mathcal{A}$ , then  $\mathbf{t}$  is factor-equivalent to the (unique) epistandard word  $\mathbf{s}$  directed by  $\Delta$ .

Moreover, with the same notation as in the above remark, the episturmian word  $\mathbf{t}$  is periodic if and only if the epistandard word  $\mathbf{s}$  is periodic, and this holds if and only if  $|\text{Ult}(\Delta)| = 1$  (see Remark 2.7 or Theorem 4.15 later).

**Example 3.6.** Consider the episturmian word  $\mathbf{m} = baabacabab\cdots$  directed by  $\check{\Delta} = \bar{a}\bar{b}\bar{c}(abc)^\omega$ . Observe that  $\mathbf{m}$  is factor-equivalent to the Tribonacci word  $\mathbf{r}$ , and we have

$$\mathbf{m} = \bar{\psi}_a(\mathbf{m}^{(1)}) = \bar{\psi}_a\psi_b(\mathbf{m}^{(2)}) = \bar{\psi}_a\psi_b\bar{\psi}_c(\mathbf{m}^{(3)}),$$

where  $\mathbf{m}^{(3)}$  is directed by  $T^3(\check{\Delta}) = (abc)^\omega$ , i.e.,  $\mathbf{m}^{(3)} = \mathbf{r}$ .

**Example 3.7.** We now consider an example where the condition that the  $\mathbf{t}^{(i)}$  in Theorem 3.3 are recurrent is not satisfied. Let  $\mathbf{t} = d\mathbf{r} = dabacabaabacaba \cdots$  where  $\mathbf{r}$  is the Tribonacci word and  $d$  is a letter different from  $a, b, c$ . Then  $\mathbf{t} = \bar{\psi}_a(\mathbf{t}^{(1)})$ ,  $\mathbf{t}^{(1)} = \bar{\psi}_b(\mathbf{t}^{(2)})$ ,  $\mathbf{t}^{(2)} = \bar{\psi}_c(\mathbf{t}^{(3)})$ , and so on; however, these  $\mathbf{t}^{(i)}$  are not recurrent (and  $\mathbf{t}$  is not episturmian). The infinite word  $\mathbf{t} = d\mathbf{r}$  is actually an example of an *episkew word*, i.e., a non-recurrent infinite word having episturmian factors. Such words are discussed in more detail in Section 7.2.

**Remark 3.8.** Let us point out that the construction of epistandard words by palindromic closure (given in Theorem 2.3) extends to all episturmian words: when  $\check{x}_n = \bar{x}_n$  write  $x_n$  on the *left* and use *palindromic left-closure*. Here  $\mathbf{m}$  (from the above example) appears step by step on the right:

$$\begin{aligned} & \underline{a} \cdot \\ & a \cdot \underline{ba} \\ & ab\underline{aca} \cdot ba \\ & abaca \cdot ba\underline{abacaba} \end{aligned}$$

When an episturmian word is aperiodic, we have the following fundamental link between the words  $(\mathbf{t}^{(n)})_{n \geq 0}$  and the spinned infinite word  $\check{\Delta}$  occurring in Theorem 3.3: if  $a_n$  is the first letter of  $\mathbf{t}^{(n)}$ , then  $\mu_{\check{x}_1 \dots \check{x}_n}(a_n)$  is a prefix of  $\mathbf{t}$  and the sequence  $(\mu_{\check{x}_1 \dots \check{x}_n}(a_n))_{n \geq 1}$  is not ultimately constant (since  $\check{\Delta}$  is not ultimately constant), then  $\mathbf{t} = \lim_{n \rightarrow \infty} \mu_{\check{x}_1 \dots \check{x}_n}(a_n)$ . This fact is a slight generalization of a result of Risley and Zamboni [105, Prop. III.7] on *S-adic representations* for standard Arnoux-Rauzy sequences. See also the recent paper [23] for S-adic representations of Sturmian words. Note that *S-adic dynamical systems* were introduced by Ferenczi [50] as *minimal dynamical systems* (e.g., see [96]) generated by a finite number of substitutions. In the case of episturmian words, the notion itself is actually a reformulation of the well-known *Rauzy rules*, as studied in [98]. In fact, it is well known that the *subshift* of an aperiodic episturmian word  $\mathbf{t}$  (i.e., the topological closure of the shift orbit of  $\mathbf{t}$ ) is a minimal dynamical system, i.e., it consists of all the episturmian words with the same set of factors as  $\mathbf{t}$ .

It is not hard to see that when  $|\mathcal{A}| \geq 2$  a morphism on  $\mathcal{A}$  is episturmian (resp. epistandard) if and only if it preserves the set of aperiodic episturmian (resp. epistandard) words (see [73]). Even more:

**Theorem 3.9.** [73, Theorem 3.13] *When  $|\mathcal{A}| \geq 2$ , a morphism  $\varphi$  on  $\mathcal{A}$  is episturmian (resp. epistandard) if there exist strict episturmian (resp. epistandard) words  $\mathbf{m}, \mathbf{t}$  such that  $\mathbf{m} = \varphi(\mathbf{t})$ .  $\square$*

Purely morphic episturmian words (i.e., those generated by morphisms) are discussed further in Section 4, where we consider the relationship between spins and the shifts that they induce. These ideas were used in [75] to obtain a complete answer to the question: if an episturmian word is purely morphic, which shifts of it, if any, are also purely morphic? (See Theorem 4.19, to follow.) Such rigidity issues are discussed in more detail in Sections 4.4 and 8.

In [75], Justin and Pirillo also made use of bi-infinite words, which often allow for more natural formulations. Indeed, the characterization (Theorem 3.3) of right-infinite episturmian words by a sequence  $(\mathbf{t}^{(i)})_{i \geq 0}$  extends to bi-infinite episturmian words, with all the  $\mathbf{t}^{(i)}$  now bi-infinite episturmian words. That is, as for right-infinite episturmian words, we have bi-infinite words of the form  $\mathbf{l}^{(i)}.\mathbf{r}^{(i)}$  where  $\mathbf{l}^{(i)}$  is a left-infinite episturmian word and  $\mathbf{r}^{(i)}$  is a right-infinite episturmian word. Moreover, if the bi-infinite episturmian word  $\mathbf{b} = \mathbf{l}.\mathbf{r}$  is directed by  $\check{\Delta}$  with associated bi-infinite episturmian words  $\mathbf{b}^{(i)} = \mathbf{l}^{(i)}.\mathbf{r}^{(i)}$ , then  $\mathbf{r}$  is directed by  $\check{\Delta}$  with associated right-infinite episturmian words  $\mathbf{r}^{(i)}$ .

## 4 Spins, shifts, and directive words

In this section, we discuss in more detail the notion of spins, the *shifts* they induce, and the concept of *block-equivalence* in connection with directive words. These notions allow us to study in particular when two different spinned infinite words direct the same episturmian word. Indeed, as we shall see in Section 4.3, the correspondence between episturmian words and spinned directive words is not one-to-one.

### 4.1 Notation for pure episturmian morphisms

For  $a \in \mathcal{A}$ , let  $\mu_a = \psi_a$  and  $\mu_{\bar{a}} = \bar{\psi}_a$ . This operator  $\mu$  can be naturally extended (as done in [73]) to a pure episturmian morphism: for any spinned finite word  $\check{w} = \check{x}_1 \cdots \check{x}_n$  over  $\mathcal{A} \cup \bar{\mathcal{A}}$ , we define  $\mu_{\check{w}} := \mu_{\check{x}_1} \cdots \mu_{\check{x}_n}$  and set  $\mu_\varepsilon$  equal to the identity morphism Id.

Viewing  $w = x_1 x_2 \cdots x_n$  as a prefix of the directive word  $\Delta = x_1 x_2 x_3 \cdots \in \mathcal{A}^\omega$ , it is clear from Theorem 3.1 that the words

$$\mu_{x_1 \cdots x_{n-1}}(x_n), \quad n \geq 1,$$

are prefixes of the epistandard word  $\mathbf{s}$  directed by  $\Delta$ .

**Example 4.1.** We observe that any epistandard word  $\mathbf{s} \in \mathcal{A}^\omega$  has the form  $\mathbf{s} = \mu_w(\mathbf{s}')$  for some uniquely determined finite word  $w$  and strict epistandard word  $\mathbf{s}'$ . Indeed, if  $\Delta = x_1 x_2 x_3 \cdots \in \mathcal{A}^\omega$  is the directive word of  $\mathbf{s}$  and  $m$  is the smallest positive integer such that  $\text{Alph}(x_{m+1} x_{m+2} \cdots) = \text{Alph}(\Delta)$ , then  $x_1 \cdots x_m$  is the shortest prefix of  $\Delta$  that contains all the letters not appearing infinitely often in  $\Delta$ . Moreover, by Theorem 3.1,  $\mathbf{s} = \mu_{x_1 \cdots x_m}(\mathbf{s}^{(m)})$  where  $\mathbf{s}^{(m)}$  is the epistandard word directed by  $\text{T}^m(\Delta) = x_{m+1} x_{m+2} \cdots$ . Since  $\text{Ult}(\text{T}^m(\Delta)) = \text{Alph}(\text{T}^m(\Delta))$  by construction, the epistandard word  $\mathbf{s}^{(m)}$  is strict. For example, with  $\Delta = c(ab)^\omega$ , we have  $\mathbf{s} = \psi_c(\mathbf{s}^{(1)})$  where  $\mathbf{s}^{(1)}$  is directed by  $(ab)^\omega$ , i.e.,  $\mathbf{s}^{(1)}$  is the well-known *Fibonacci word* over  $\{a, b\}$ .

For  $n \geq 1$ , let  $u_{n+1} := \text{Pal}(x_1 \cdots x_n)$  and set  $u_1 = \varepsilon$ . Then by part *iv*) of Theorem 2.3, the epistandard word  $\mathbf{s}$  directed by  $\Delta$  is given by  $\mathbf{s} = \lim_{n \rightarrow \infty} u_n$ . We have the following useful formula from [73]:

$$u_{i+1} = \mu_{x_1 \cdots x_{i-1}}(x_i) u_i \quad \text{for } i > 0. \quad (4.1)$$

For letters  $(x_j)_{1 \leq j \leq i}$ , formula (4.1) inductively leads to:

$$u_{i+1} = \mu_{x_1 \cdots x_{i-1}}(x_i) \cdots \mu_{x_1}(x_2) x_1 = \prod_{1 \leq j \leq i} \mu_{x_1 \cdots x_{j-1}}(x_j). \quad (4.2)$$

(Note that by convention,  $x_1 \cdots x_0 = \varepsilon$  in the above product.) For example, with  $\Delta = abcb \cdots$ , we compute:

$$u_3 = \text{Pal}(abcb) = \mu_{abc}(b) \mu_{ab}(c) \mu_a(b) a = abacab \cdot abac \cdot ab \cdot a.$$

### 4.2 Shifts

Now let  $\check{w} = \check{x}_1 \check{x}_2 \cdots \check{x}_n$  be a spinned version of  $w = x_1 x_2 \cdots x_n$  (viewed as a prefix of a spinned version  $\check{\Delta}$  of  $\Delta$ ). Then, for any finite word  $v$ , we have

$$\mu_{\check{w}}(v) = S_{\check{w}}^{-1} \mu_w(v) S_{\check{w}} \quad \text{where } S_{\check{w}} = \prod_{\substack{i=n, \dots, 1 \\ |\check{x}_i = \bar{x}_i}} \mu_{x_1 \cdots x_{i-1}}(x_i). \quad (4.3)$$

Observe that  $S_{\check{w}}$  is a prefix of  $\text{Pal}(w)$ ; in particular  $S_{\check{w}} = \text{Pal}(w)$  by equation (4.2). Note also that  $\mu_{\check{w}}(v) = \text{T}^{|S_{\check{w}}|}(\mu_w(v))$ . The word  $S_{\check{w}}$  is called the *shifting factor* of  $\mu_{\check{w}}$  and its length  $|S_{\check{w}}|$  is called the *shift induced* by the prefix  $\check{w}$  of  $\check{\Delta}$  of length  $n$  [75].

**Example 4.2.** If we take  $\check{w} = \bar{a}\bar{b}\bar{c}\bar{a}$ , then

$$S_{\check{w}} = \mu_{abc}(a)\mu_a(b) = abacaba \cdot ab.$$

Thus since  $\mu_{abca}(ca) = abacabaab \cdot acabacaba$ , we have

$$\mu_{\bar{a}\bar{b}\bar{c}\bar{a}}(ca) = T^9(\mu_{abca}(ca)) = acabacaba \cdot abacabaab.$$

Likewise, for any infinite word  $\mathbf{y} \in \mathcal{A}^\omega$ ,  $\mu_{\check{w}}(\mathbf{y}) = S_{\check{w}}^{-1}\mu_w(\mathbf{y})$ . For example, if we take  $\check{w} = \bar{a}\bar{b}$ , then  $S_{\check{w}} = \text{Pal}(ab) = aba$ , and hence  $\mu_{\bar{a}\bar{b}}(\mathbf{y}) = (aba)^{-1}\mu_{ab}(\mathbf{y})$  for any infinite word  $\mathbf{y}$ .

*Note.* The morphisms  $\mu_w$  and  $\mu_{\check{w}}$  are *conjugate* morphisms [99].

### 4.3 Block-equivalence & directive words

By Theorem 2.3 (and also Theorem 3.1), any epistandard word  $\mathbf{s} \in \mathcal{A}^\omega$  has a unique directive word over  $\mathcal{A}$ , but  $\mathbf{s}$  also has infinitely many other spinned directive words (see [73, 75, 63]). For example, the Tribonacci word is directed by  $(abc)^\omega$  and also by  $(abc)^n\bar{a}\bar{b}\bar{c}(ab\bar{c})^\omega$  for each  $n \geq 0$ , as well as infinitely many other spinned words. The natural question: “does any spinned word direct a unique episturmian word?” was answered in [73].

**Proposition 4.3.** [73]

1. Any spinned infinite word  $\check{\Delta}$  having infinitely many letters with spin 0 directs a unique episturmian word beginning with the left-most letter having spin 0 in  $\check{\Delta}$ .
2. Any spinned infinite word  $\check{\Delta}$  with all spins ultimately 1 directs exactly  $|\text{Ult}(\Delta)|$  episturmian words.
3. Let  $\check{\Delta}$  be a spinned infinite word having all its letters with spin 1 and let  $a \in \text{Ult}(\Delta)$ . Then  $\check{\Delta}$  directs exactly one episturmian word starting with  $a$ .  $\square$

*Note.* The above statement corrects a small error in Proposition 3.11 of [73] where item 3 was stated in the more general case when  $\check{\Delta}$  has all spins ultimately 1. In this case,  $\check{\Delta}$  still directs exactly one episturmian word for each letter  $a$  in  $\text{Ult}(\Delta)$ , but contrary to what is written in [73], nothing can be said about its first letter.

*Block-equivalence* for spinned words was introduced in [75] as a way of studying when  $\check{\Delta}$  and  $\hat{\Delta}$  (two spinned versions of a directive word  $\Delta$ ) direct the same bi-infinite episturmian word. We do not recall the full details here, only a few notions relating to it.

**Notation.** If  $v \in \mathcal{A}^+$ , then  $\bar{v} \in \bar{\mathcal{A}}^+$  is  $v$  with all spins 1.

A word of the form  $xvx$ , where  $x \in \mathcal{A}$  and  $v \in (\mathcal{A} \setminus \{x\})^*$ , is called a ( $x$ -based) *block*. A ( $x$ -based) *block-transformation* is the replacement in a spinned word of an occurrence of  $xv\bar{x}$  (where  $xvx$  is a block) by  $\bar{x}\bar{v}x$  or vice-versa. Two finite spinned words  $w, w'$  are said to be *block-equivalent* if we can pass from one to the other by a (possibly empty) chain of block-transformations, in which case we write  $w \equiv w'$ . For example,  $\bar{b}\bar{a}\bar{b}\bar{c}\bar{b}\bar{a}\bar{c}$  and  $b\bar{a}\bar{b}\bar{c}\bar{b}\bar{a}\bar{c}$  are block-equivalent because  $\bar{b}\bar{a}\bar{b}\bar{c}\bar{b}\bar{a}\bar{c} \rightarrow \bar{b}\bar{a}\bar{b}\bar{c}\bar{b}\bar{a}\bar{c} \rightarrow b\bar{a}\bar{b}\bar{c}\bar{b}\bar{a}\bar{c}$  and vice-versa. Note that if  $w \equiv w'$  then  $w$  and  $w'$  are spinned versions of the same word over  $\mathcal{A}$ . Block-equivalence extends to (right-)infinite words as follows.

Let  $\Delta_1, \Delta_2$  be spinned versions of  $\Delta$ . We write  $\Delta_1 \rightsquigarrow \Delta_2$  if there exist infinitely many prefixes  $f_i$  of  $\Delta_1$  and  $g_i$  of  $\Delta_2$  with the  $g_i$  of strictly increasing lengths, and such that, for all  $i$ ,  $|g_i| \leq |f_i|$  and  $f_i \equiv g_i c_i$  for a suitable spinned word  $c_i$ . Infinite words  $\Delta_1$  and  $\Delta_2$  are said to be *block-equivalent* (denoted by  $\Delta_1 \equiv \Delta_2$ ) if  $\Delta_1 \rightsquigarrow \Delta_2$  and  $\Delta_2 \rightsquigarrow \Delta_1$ .

**Remark 4.4.** If  $x$  is a letter and  $v \in \mathcal{A}^*$  is  $x$ -free, then  $\bar{x}v\bar{x}$  and  $xv\bar{x}$  are block-equivalent and they induce the same shift, i.e.,  $\mu_{\bar{x}v\bar{x}} = \mu_{xv\bar{x}}$  [75, Theorem 2.2]. Thus the monoid of pure episturmian morphisms,  $\mathcal{E}_p$ , is isomorphic to the quotient of  $(\mathcal{A} \cup \bar{\mathcal{A}})^*$  by the block-equivalence generated by

$$\{\bar{x}v\bar{x} \equiv xv\bar{x} \mid x \in \mathcal{A}, v \text{ is } x\text{-free}\}.$$

Note that this has some relation to the study of conjugacy and episturmian morphisms carried out by Richomme [99].

From what we have already learned about bi-infinite episturmian words (in Sections 2.2 and 3.2), it is clear that Justin and Pirillo's results about spinned infinite words directing the same bi-infinite episturmian word are still valid for words directing the same (right-infinite) episturmian word. Roughly speaking, two spinned infinite words direct the same episturmian word if and only if they are block-equivalent. For instance, we have the following results for wavy spinned versions of  $\Delta \in \mathcal{A}^\omega$ . A spinned version  $\check{\Delta}$  of  $\Delta$  is said to be wavy if  $\check{\Delta}$  contains infinitely many letters of spin 0 and infinitely many letters of spin 1.

**Theorem 4.5.** [75, Theorem 3.4] *Suppose  $\check{\Delta}$  and  $\hat{\Delta}$  are wavy versions of  $\Delta \in \mathcal{A}^\omega$  with  $|\text{Ult}(\Delta)| > 1$ . Then  $\check{\Delta}$  and  $\hat{\Delta}$  direct the same episturmian word if and only if  $\check{\Delta} \equiv \hat{\Delta}$ .  $\square$*

For example,  $ba(\bar{b}c\bar{a})^\omega$  and  $\bar{b}a\bar{b}(c\bar{a}\bar{b})^\omega$  direct the same episturmian word, namely  $\mu_{ba\bar{b}c}(\mathbf{cr})$  ( $= \mu_{\bar{b}a\bar{b}c}(\mathbf{cr})$ ) where  $\mathbf{r}$  is the Tribonacci word.

**Theorem 4.6.** [75, Prop. 3.6] *Let  $\check{\Delta}$ ,  $\hat{\Delta}$  be two spinned versions of  $\Delta \in \mathcal{A}^\omega$  with  $|\text{Ult}(\Delta)| > 1$ ,  $\check{\Delta}$  wavy, and  $\hat{\Delta}$  having all spins ultimately 0 or 1. If  $\check{\Delta}$  and  $\hat{\Delta}$  direct the same episturmian word, then  $\check{\Delta} \rightsquigarrow \hat{\Delta}$ .  $\square$*

Similar results also hold when all spins are ultimately 0 or 1 and in the periodic case. See Propositions 3.7 and 3.10 in [75].

**Remark 4.7.** In [75], the study of block-equivalence for finite spinned words led to numeration systems that resemble the *Ostrowski systems* [20] associated with Sturmian words. A matrix formula for computing the number of representations of an integer in such a system was also given in [75, Section 2].

More recently, Glen, Levé, and Richomme [63] established the following complete characterization of pairs of spinned infinite words directing the same unique episturmian word. Not only does the following theorem provide the relative forms of two spinned infinite words directing the same episturmian word, but it also fully solves the periodic case, which was only partially solved in [75].

**Theorem 4.8.** [63] *Given two spinned infinite words  $\Delta_1$  and  $\Delta_2$ , the following assertions are equivalent.*

*i)  $\Delta_1$  and  $\Delta_2$  direct the same right-infinite episturmian word.*

*ii) One of the following cases holds for some  $i, j$  such that  $\{i, j\} = \{1, 2\}$ :*

- 1.  $\Delta_i = \prod_{n \geq 1} v_n$ ,  $\Delta_j = \prod_{n \geq 1} z_n$  where  $(v_n)_{n \geq 1}, (z_n)_{n \geq 1}$  are spinned words such that  $\mu_{v_n} = \mu_{z_n}$  for all  $n \geq 1$ ;*



2.  $\Delta_i = wx \prod_{n \geq 1} v_n \check{x}_n$ ,  $\Delta_j = w' \bar{x} \prod_{n \geq 1} \bar{v}_n \hat{x}_n$  where  $w, w'$  are spinned words such that  $\mu_w = \mu_{w'}$ ,  $x$  is a letter,  $(v_n)_{n \geq 1}$  is a sequence of non-empty  $x$ -free words, and  $(\check{x}_n)_{n \geq 1}$ ,  $(\hat{x}_n)_{n \geq 1}$  are sequences of non-empty spinned words over  $\{x, \bar{x}\}$  such that, for all  $n \geq 1$ ,  $|\check{x}_n| = |\hat{x}_n|$  and  $|\check{x}_n|_x = |\hat{x}_n|_x$ ;
3.  $\Delta_1 = w\mathbf{x}$  and  $\Delta_2 = w'\mathbf{y}$  where  $w, w'$  are spinned words,  $x$  and  $y$  are letters, and  $\mathbf{x} \in \{x, \bar{x}\}^\omega$ ,  $\mathbf{y} \in \{y, \bar{y}\}^\omega$  are spinned infinite words such that  $\mu_w(x) = \mu_{w'}(y)$ .

□

In items 1 and 2 of Theorem 4.8, the two considered directive words are spinned versions of the same infinite word. This does not hold in item 3, which concerns only periodic episturmian words. In particular, we observe the following:

**Remark 4.9.** If an *aperiodic* episturmian word is directed by two different spinned infinite words  $\Delta_1$  and  $\Delta_2$ , then  $\Delta_1$  and  $\Delta_2$  are spinned versions of the same word  $\Delta$ .

As an example of item 3, one can consider the periodic episturmian word  $(bcba)^\omega$  which is directed by both  $bca^\omega$  and  $b\bar{a}c^\omega$  (since  $\mu_{bc}(a) = \mu_{b\bar{a}}(c)$ ). Note also that  $(bcba)^\omega$  is epistandard and has the same set of factors as the epistandard word  $(babc)^\omega$  directed by  $bac^\omega$ . Actually, in view of Remark 3.5, we observe the following:

**Remark 4.10.** The subshift of any aperiodic episturmian word contains a unique (aperiodic) epistandard word, whereas the subshift of a periodic episturmian word contains exactly two (periodic) epistandard words, except if this word is  $a^\omega$  with  $a$  a letter.

We also observe that  $x$  and  $y$  can be equal in item 3 of Theorem 4.8; for example  $(ab)^\omega$  is directed by  $a\bar{b}b^\omega$  and by  $ab^\omega$ .

**Example 4.11.** [63] For  $a, b, c$  three different letters in  $\mathcal{A}$ , the spinned infinite words  $\Delta_1 = a(bc\bar{a})^\omega$  and  $\Delta_2 = \bar{a}(\bar{b}\bar{c}\bar{a})^\omega$  direct the same episturmian word that starts with the letter  $a$ . Indeed, these two directive words fulfill item 2 of Theorem 4.8 with  $w = w' = \varepsilon$ ,  $x = a$ , and for all  $n$ ,  $v_n = bc$  and  $\check{x}_n = \hat{x}_n = \bar{a}$ . Moreover the fact that  $\Delta_1$  starts with the letter  $a$  shows that the word it directs starts with  $a$ . Similarly  $\Delta'_1 = \bar{a}b(cab)^\omega$  and  $\Delta'_2 = \bar{a}\bar{b}(\bar{c}\bar{a}\bar{b})^\omega$  direct the same episturmian word starting with the letter  $b$ . Since  $\Delta_2 = \Delta'_2$ , this shows that the relation “direct the same episturmian word” over spinned infinite words is not an equivalence relation.

Items 2 and 3 of Theorem 4.8 show that any episturmian word is directed by a spinned infinite word having infinitely many letters of spin 0, but also by a spinned word having both infinitely many letters of spin 0 and infinitely many letters of spin 1 (i.e., a wavy word). To emphasize the importance of these facts, let us recall from Proposition 4.3 that if  $\check{\Delta}$  is a spinned infinite word over  $\mathcal{A} \cup \bar{\mathcal{A}}$  with infinitely many letters of spin 0, then there exists a unique episturmian word  $\mathbf{t}$  directed by  $\check{\Delta}$ . Uniqueness comes from the fact that the first letter of  $\mathbf{t}$  is fixed by the first letter of spin 0 in  $\check{\Delta}$ . We also note that if an episturmian word  $\mathbf{t}$  has two directive words satisfying items 2 or 3, then  $\mathbf{t}$  has infinitely many directive words (this was shown in [63]).

When studying repetitions in Sturmian words, Berthé, Holton, and Zamboni [23] proved that any Sturmian word has a unique directive word over  $\{a, b, \bar{a}, \bar{b}\}$  containing infinitely many letters of spin 0, but no factor of the form  $\bar{a}\bar{b}^n a$  or  $\bar{b}\bar{a}^n b$  with  $n$  an integer. Levé and Richomme [80] recently generalized this result to episturmian words by introducing a way to ‘normalize’ the directive word(s) of an episturmian word so that any episturmian word can be defined uniquely by its so-called *normalized directive word*, defined by some factor avoidance, as follows. This idea has since proved useful in the study of *quasiperiodic* episturmian words (see Section 8); in particular, it provides an effective way to decide whether or not a given episturmian word is quasiperiodic.

**Theorem 4.12.** [63, 80] *Any episturmian word  $\mathbf{t} \in \mathcal{A}^\omega$  has a spinned directive word  $\check{\Delta}$  containing infinitely many letters of spin 0, but no factor in  $\bigcup_{a \in \mathcal{A}} \bar{a}\bar{\mathcal{A}}^*a$ . Such a directive word is unique if  $\mathbf{t}$  is aperiodic, in which case  $\check{\Delta}$  is called the normalized directive word for  $\mathbf{t}$ .  $\square$*

*Note.* Uniqueness does not necessarily hold for periodic episturmian words. For example, the periodic episturmian word  $(ab)^\omega = \psi_a(b^\omega) = \bar{\psi}_b(a^\omega)$  is directed by  $ab^\omega$  and by  $\bar{b}a^\omega$  (since  $\psi_a(b) = ab = \bar{\psi}_b(a)$ ).

The following result tells us precisely which episturmian words have a unique directive word.

**Theorem 4.13.** [63] *An episturmian word has a unique directive word if and only if its (normalized) directive word contains:*

- 1) *infinitely many letters of spin 0,*
- 2) *infinitely many letters of spin 1,*
- 3) *no factor in  $\bigcup_{a \in \mathcal{A}} \bar{a}\bar{\mathcal{A}}^*a$ ,*
- 4) *and no factor in  $\bigcup_{a \in \mathcal{A}} a\mathcal{A}^*\bar{a}$ .*

*Such an episturmian word is necessarily aperiodic.*

For instance, a particular family of episturmian words having unique directive words consists of those directed by *regular wavy words* [58, 64], i.e., spinned infinite words having both infinitely many letters of spin 0 and infinitely many letters of spin 1 such that each letter occurs with the same spin everywhere in the directive word. More formally, a spinned version  $\check{w}$  of a finite or infinite word  $w$  is said to be *regular* if, for each letter  $x \in \text{Alph}(w)$ , all occurrences of  $\check{x}$  in  $\check{w}$  have the same spin (0 or 1). For example, the regular wavy word  $(\bar{a}\bar{b}\bar{c})^\omega$  is the unique directive word for the episturmian word  $\mathbf{ar} = aabacabaabacab \cdots$  where  $\mathbf{r}$  is the Tribonacci word.

In the Sturmian case, we have:

**Proposition 4.14.** [63] *Any Sturmian word has either a unique spinned directive word or infinitely many spinned directive words. Moreover, a Sturmian word has a unique directive word if and only if its (normalized) directive word is regular wavy.  $\square$*

As pointed out in [63], Proposition 4.14 shows a great difference between Sturmian words and episturmian words constructed over alphabets with at least three letters. Indeed, when considering words over a ternary alphabet, one can find episturmian words having exactly  $m$  directive words for any  $m \geq 1$ . For instance, the episturmian word  $\mathbf{t}$  directed by  $\check{\Delta} = a(\bar{b}\bar{a})^{m-1}\bar{b}\bar{c}(ab\bar{c})^\omega$  has exactly  $m$  directive words, namely  $(\bar{a}\bar{b})^i a(\bar{b}\bar{a})^j \bar{b}\bar{c}(ab\bar{c})^\omega$  with  $i + j = m - 1$ . Notice that the suffix  $\bar{b}\bar{c}(ab\bar{c})^\omega$  of  $\check{\Delta}$  is regular wavy, and the other  $m - 1$  spinned versions of  $\Delta$  that also direct  $\mathbf{t}$  arise from the  $m - 1$  words that are block-equivalent to the prefix  $a(\bar{b}\bar{a})^{m-1}$ .

#### 4.4 Periodic and purely morphic episturmian words

We are now ready to describe periodic and purely morphic episturmian words.

Recall from Remark 2.7 that the periodic episturmian words correspond to  $|\text{Ult}(\Delta)| = 1$ . The following theorem gives the form of such words in terms of pure episturmian morphisms.

**Theorem 4.15.** [73] *An episturmian word is periodic if and only if it is  $(\mu_{\check{w}}(x))^\omega$  for some spinned finite word  $\check{w}$  and letter  $x$ .  $\square$*

For example,  $(\mu_{\bar{a}\bar{b}}(c))^\omega = (acab)^\omega$  is the periodic episturmian word directed by  $\bar{a}\bar{b}c^\omega$  (in fact, it is epistandard as it is also directed by  $acb^\omega$ ).

The next theorem characterizes purely morphic episturmian words with respect to their directive words.

**Theorem 4.16.** [73, Theorem 3.14] *An aperiodic episturmian word is purely morphic (i.e., generated by a morphism) if and only if it is directed by a periodic spinned infinite word  $\Delta = (\check{f})^\omega$  for some spinned word  $f$ . Moreover it can be generated by  $\mu_{\check{f}}$ .*  $\square$

We observe from Theorem 4.16 that any purely morphic episturmian word is strict (i.e., an Arnoux-Rauzy sequence) as  $\text{Ult}(\Delta) = \text{Alph}(f) = \text{Alph}(\Delta)$ . The proof of this theorem makes use of Proposition 4.3 and Theorem 3.9.

**Example 4.17.** The Tribonacci word is generated by  $\mu_{abc}$ . Notice that  $\mu_{abc} = \sigma^3$  where  $\sigma$  is the Tribonacci morphism defined by  $\sigma : (a, b, c) \mapsto (ab, ac, a)$ .

**Remark 4.18.** Purely morphic standard Sturmian words were previously characterized independently in the following papers: [16, 38, 77]. Yasutomi [118] has since established a characterization of all purely morphic Sturmian words with respect to their slopes and intercepts (when viewed as cutting sequences). An alternative geometric proof of Yasutomi's result was recently given by Berthé *et al.* in [21].

Using the notion of block-equivalence, Justin and Pirillo [75] explicitly determined which shifts, if any, of a purely morphic episturmian word are also purely morphic.

**Theorem 4.19.** [75] *If an episturmian word  $\mathbf{t}$  is purely morphic, then its shift  $T^i(\mathbf{t})$  is also purely morphic if and only if  $i$  belongs to some particular interval.*  $\square$

See Section 4 of [75] for specific (and very technical) details.

**Example 4.20.** For the Tribonacci word  $\mathbf{r}$ , only itself and  $T^{-1}(\mathbf{r})$  are purely morphic. Note that  $T^{-1}(\mathbf{r})$  corresponds to three episturmian words:  $\mathbf{ar}$ ,  $\mathbf{br}$ ,  $\mathbf{cr}$ , directed by  $(\bar{a}\bar{b}\bar{c})^\omega$ ,  $(\bar{a}\bar{b}\bar{c})^\omega$ ,  $(\bar{a}\bar{b}\bar{c})^\omega$ , respectively.

**Remark 4.21.** Theorem 4.19 corrects an error in [73, Section 5.1] where it was mistakenly said that if an episturmian word is purely morphic then any shift of it is also purely morphic. Indeed, this is false even in the Sturmian case as Fagnot [48] has shown that if  $\mathbf{s}$  is a purely morphic standard Sturmian word on  $\{a, b\}$ , then  $\mathbf{as}$ ,  $\mathbf{bs}$ ,  $\mathbf{abs}$ ,  $\mathbf{bas}$  (which are purely morphic [17]) are the only purely morphic Sturmian words related to  $\mathbf{s}$  by a shift.

## 5 Arnoux-Rauzy sequences

We now briefly turn our attention to Arnoux-Rauzy sequences since their combinatorial properties are also considered in the sections that follow.

*Arnoux-Rauzy sequences* are uniformly recurrent infinite words over a finite alphabet  $\mathcal{A}$  with factor complexity  $(|\mathcal{A}| - 1)n + 1$  for each  $n \in \mathbb{N}$ , and exactly one right and one left special factor of each length. They were introduced by Arnoux and Rauzy [97, 12], who studied them using *Rauzy graphs*, with particular emphasis on the case  $|\mathcal{A}| = 3$ . (Note that the foregoing definition is equivalent to the one given in the introduction.)

As mentioned previously (in Section 2.3), Arnoux-Rauzy sequences are exactly the strict episturmian words; in particular, any episturmian word has the form  $\varphi(\mathbf{t})$  with  $\varphi$  an episturmian

morphism and  $\mathbf{t}$  an Arnoux-Rauzy sequence. In this sense, episturmian words are only a slight generalization of Arnoux-Rauzy sequences. For example, the family of episturmian words on three letters  $\{a, b, c\}$  consists of the Arnoux-Rauzy sequences over  $\{a, b, c\}$ , the Sturmian words over  $\{a, b\}$ ,  $\{b, c\}$ ,  $\{a, c\}$  and their images under episturmian morphisms on  $\{a, b, c\}$ , and periodic infinite words of the form  $\varphi(x)^\omega$  where  $\varphi$  is an episturmian morphism on  $\{a, b, c\}$  and  $x \in \{a, b, c\}$ .

Arnoux-Rauzy sequences have deep properties studied in the framework of dynamical systems, with connections to geometrical realizations such as *Rauzy fractals* [11] and interval exchanges. When  $|\mathcal{A}| = 3$ , the condition on the special factors distinguishes Arnoux-Rauzy sequences from other infinite words of complexity  $2n+1$ , such as those obtained by coding trajectories of 3-interval exchange transformations (e.g., see [51]). In [12], it was shown how Arnoux-Rauzy sequences of complexity  $2n + 1$  (i.e., the 3-strict episturmian words) can be geometrically realized by an exchange of six intervals on the unit circle, which generalizes the representation of Sturmian sequences by rotations.

An alternative way of introducing and studying Arnoux-Rauzy sequences is in the context of *S-adic dynamical systems*, as done in [105] for instance (see our remarks following Theorem 3.3 in Section 3.2). In [40], Damanik and Zamboni give a kind of survey on this approach by considering *Arnoux-Rauzy subshifts* and answering various combinatorial questions concerning linear recurrence, maximal powers of factors, and the number of palindromes of a given length. They also present some applications of their results to the spectral theory of discrete one-dimensional Schrödinger operators with potentials given by Arnoux-Rauzy sequences.

Arnoux-Rauzy sequences also have interesting arithmetical properties. For instance, if one considers the *frequencies* of letters (as discussed later in Section 6.4), they are well-defined, and renormalization by an episturmian morphism leads to a generalization of the continued fraction algorithm that associates to each  $k$ -letter Arnoux-Rauzy sequence an infinite array of  $k \times k$  rational numbers. In the special case  $k = 2$ , these fractions are consecutive Farey numbers arising from the continued fraction expansion of the frequencies of the two letters. More generally, given an Arnoux-Rauzy sequence on  $k$ -letters, its directive word is determined by the ‘multi-dimensional’ continued fraction expansion of the frequencies of the first  $k - 1$  letters. Unfortunately, this generalized algorithm (except for the case  $k = 2$  when it is exactly the usual continued fraction algorithm) is only defined on a set of measure zero in  $\mathbb{R}^{k-1}$ . This reduces its interest and explains why it has not been appropriately studied since its inception (see Sections 6.2.1 and 6.4 for further details). Nonetheless, a nice arithmetical characterization of 3-letter Arnoux-Rauzy sequences can be given, as follows. We say that a triple  $(a, b, c)$  does not satisfy the triangular inequality if one of the coordinates is larger than the sum of the other two (e.g.,  $a > b + c$ ). In that case, we can renormalize in a unique way to obtain the triple  $(a - b - c, b, c)$ . The set of allowable frequencies for 3-letter Arnoux-Rauzy sequences is exactly the set of triples  $(a, b, c)$  that can be infinitely renormalized, each time to a triple that does *not* satisfy the triangular inequality (see [12]). The resulting picture exhibits a kind of *Sierpinski carpet*.

For further details on Arnoux-Rauzy sequences, we refer the reader to the interesting survey [22] in which Berthé, Ferenczi, and Zamboni discuss connections between Arnoux-Rauzy sequences and rotations of the 2-torus; coding of two-dimensional actions and two-dimensional Sturmian words; and interval exchanges and sequences of low complexity. See also [35], Section 12.2.3 in [96], and J. Berstel’s nice survey paper [15] in which he compares some combinatorial properties of Arnoux-Rauzy sequences (as well as episturmian words) to those of Sturmian words.

## 5.1 Finite Arnoux-Rauzy words

A finite word  $w$  is said to be *finite episturmian* if  $w$  is a factor of some infinite episturmian word. When considering factors of (infinite) episturmian words, it suffices to consider only the strict standard ones (i.e., the standard Arnoux-Rauzy sequences). Indeed, for any prefix  $u$  of an epistandard word, there exists a strict epistandard word also having  $u$  as a prefix. In particular, the words  $\mu_w(x)$ , with  $w \in \mathcal{A}^*$  and  $x \in \mathcal{A}$ , are the *standard* ones (cf. standard words, e.g., [83, Chapter 2]). They can be obtained by the *Rauzy rules* [98] (see also [43, Theorem 8]), and this has a strong connection with the set of periods of the palindromes  $u_{n+1} = \text{Pal}(x_1 \cdots x_n)$  (given in Theorem 2.3) and the Euclidean algorithm. This relation was studied by Castelli, Mignosi, and Restivo [34], who extended the well-known *Fine and Wilf Theorem* [82] to words having three periods. Justin [70] generalized this result even further to words having an arbitrary number of periods, which led to a characterization of finite episturmian words.

Finite episturmian words are exactly the *finite Arnoux-Rauzy words*. Such words were enumerated by Mignosi and Zamboni [88], who described a multi-dimensional generalization of the *Euler phi-function* that counts the number of finite Arnoux-Rauzy words of each length. Finite episturmian words have also been characterized with respect to lexicographic orderings in [62] (see Theorem 7.5 later).

## 6 Some properties of factors

### 6.1 Factor complexity

As mentioned previously, any  $k$ -strict episturmian word has complexity  $(k-1)n+1$  for all  $n \in \mathbb{N}$ . More generally:

**Theorem 6.1.** [43, Theorem 7] *Suppose  $\mathbf{t}$  is an episturmian word directed by  $\check{\Delta}$  with  $|\text{Ult}(\Delta)| > 1$ . Then, for  $n$  large enough,  $\mathbf{t}$  has complexity  $(k-1)n+q$  for some  $q \in \mathbb{N}^+$ , where  $k = |\text{Ult}(\Delta)|$ .  $\square$*

This theorem can be easily deduced from the fact that for sufficiently large  $n$ , any left special factor of  $\mathbf{t}$  of length at least  $n$  has exactly  $k = |\text{Ult}(\Delta)|$  different left extensions in  $\mathbf{t}$  (by Theorem 6 in [43]).

### 6.2 Palindromic factors

The *palindromic complexity* of episturmian words was established in [73] by carrying out a similar study to the one for Sturmian words in [44].

**Theorem 6.2.** [73, Theorem 4.4] *If  $\mathbf{t}$  is an  $\mathcal{A}$ -strict episturmian word, then there exists exactly*

- *one palindrome of length  $n$  for all even  $n$ ,*
- *one palindrome of length  $n$  and centre  $x$  for all odd  $n$  and  $x \in \mathcal{A}$ .  $\square$*

As shown in [44], the above property is characteristic in the Sturmian case, but not when  $\mathcal{A}$  contains more than two letters because it also holds for *billiard words*, which are not episturmian (see Borel and Reutenauer [25]).

**Theorem 6.3.** [73, Section 4.2] *If  $\mathbf{t}$  is episturmian, then there exist  $|\text{Ult}(\Delta)| + 1$  bi-infinite episturmian words of the form  $\tilde{\mathbf{m}}.\mathbf{m}$  and  $\tilde{\mathbf{m}}x\mathbf{m}$  giving all but a finite number of palindromic factors of  $\mathbf{t}$ . The spinned versions of  $\Delta$  directing these bi-infinite episturmian words can be easily constructed via a simple algorithm.  $\square$*

For more precise technical details, see Section 4.2 in [73].

**Example 6.4.** For the Tribonacci word,  $\tilde{\mathbf{r}}.\mathbf{r}$  is directed by  $(abc\bar{a}bc\bar{a}bc\bar{a}bc\bar{a})^\omega$ .

### 6.2.1 Iterated palindromic closure

In [105], Risley and Zamboni gave an alternative construction of the sequence  $(u_n)_{n \geq 1}$  of palindromic prefixes of an epistandard word (where  $u_1 = \varepsilon$  and  $u_{i+1} = \text{Pal}(x_1 \cdots x_i)$  for all  $i \geq 1$ ), using a ‘hat operation’ as opposed to palindromic closure. The so-called hat operation is defined as follows. We construct a new alphabet  $\mathcal{A}' := \mathcal{A} \cup \hat{\mathcal{A}}$  where  $\hat{\mathcal{A}} = \{\hat{x} \mid x \in \mathcal{A}\}$  and denote by  $\phi$  the morphism  $\phi : \mathcal{A}' \rightarrow \mathcal{A}$  defined by  $\phi(x) = \phi(\hat{x}) = x$  for all letters  $x \in \mathcal{A}$ . The morphism  $\phi$  extends to a morphism (also denoted by  $\phi$ ) from words over  $\mathcal{A}'$  to words over  $\mathcal{A}$ . Now, from a given directive word  $\Delta = x_1 x_2 x_3 \cdots \in \mathcal{A}^\omega$ , we construct a sequence of words  $(p_i)_{i \geq 1}$  as follows. We begin with  $p_1 = \varepsilon$  and  $p_2 = \hat{x}_1$ . Then, for  $n \geq 2$ ,  $p_{n+1}$  is obtained from  $p_n$  according to the rule: if  $\hat{x}_n$  does not occur in  $p_n$ , then  $p_{n+1} = p_n \hat{x}_n \phi(p_n)$ ; otherwise  $p_{n+1} = p_n \hat{x}_n \phi(s_n)$ , where  $s_n$  is the longest palindromic suffix of  $p_n$  containing no occurrence of  $\hat{x}_n$ .

**Example 6.5.** Let  $\Delta = (abc)^\omega$ . Then using the hat operation, we obtain:

$$\begin{aligned} p_1 &= \varepsilon \\ p_2 &= \hat{a} \\ p_3 &= \hat{a}\hat{b}\hat{a} \\ p_4 &= \hat{a}\hat{b}\hat{a}\hat{c}\hat{a}\hat{b}\hat{a} \\ p_5 &= \hat{a}\hat{b}\hat{a}\hat{c}\hat{a}\hat{b}\hat{a}\hat{b}\hat{a}\hat{c}\hat{a}\hat{b}\hat{a} \\ p_6 &= \hat{a}\hat{b}\hat{a}\hat{c}\hat{a}\hat{b}\hat{a}\hat{b}\hat{a}\hat{c}\hat{a}\hat{b}\hat{a}\hat{b}\hat{a}\hat{c}\hat{a}\hat{b}\hat{a} \\ &\vdots \end{aligned}$$

Now removing all hats (by applying  $\phi$ ), we see that the  $p_i$ ’s are precisely the palindromic prefixes of the Tribonacci word:  $abacabaabacababacabaabacaba \cdots$ .

As demonstrated by the above example, the hat operation is clearly the same as iterated palindromic closure; in fact, the relationship between these two constructions is evident by formula (4.1), which we now rewrite as:

$$\text{Pal}(x_1 \cdots x_n) = \mu_{x_1 \cdots x_{n-1}}(x_n) \text{Pal}(x_1 \cdots x_{n-1}) \quad \text{for } n > 0.$$

The above formula is actually a special case of formula (3) from [71], which also happens to be formula (3) in [73], namely:

$$\text{Pal}(vw) = \mu_v(\text{Pal}(w))\text{Pal}(v) \quad \text{for any words } w, v. \quad (6.1)$$

This formula is commonly referred to as *Justin’s Formula*, from which we deduce the following two special cases:

$$\text{Pal}(xw) = \psi_x(\text{Pal}(w))x \quad \text{and} \quad \text{Pal}(wx) = \mu_w(x)\text{Pal}(w) \quad \text{for any word } v \text{ and letter } x. \quad (6.2)$$

The first formula given in (6.2) tells us that  $\text{Pal}(xw)$  is obtained from  $\text{Pal}(w)$  simply by inserting the letter  $x$  before each letter different from  $x$  and then appending  $x$  to the resulting word. For example,  $\text{Pal}(bc) = bcb$  and  $\text{Pal}(abc) = abacaba$ . The second formula given in (6.2) provides



another way to compute the palindromic right-closure of  $wx$  by placing the finite epistandard word  $\mu_w(x)$  in front of  $Pal(w)$ . For example, to compute  $Pal(abcb)$  we need only compute the words  $\mu_{abc}(b) = abacab$  and  $Pal(abc) = abacaba$ , and then we have:

$$Pal(abcb) = \mu_{abc}(a)Pal(abc) = abacab \cdot abacaba.$$

In [71], Justin established some relations between the words  $Pal(w)$ ,  $\mu_w$ ,  $Pal(\tilde{w})$ , and  $\mu_{\tilde{w}}$  where  $w$  is any finite word. Moreover, he showed that his results can be explained by the similarity of the *incidence matrices* of  $\mu_w$  and  $\mu_{\tilde{w}}$ . One curious result is that  $|Pal(w)| = |Pal(\tilde{w})|$ . For example, with  $w = abac$ ,  $Pal(w) = abaabacabaaba$  and  $Pal(\tilde{w}) = cacbcacacacbcac$ , both of length 15.

Applying his results to a 2-letter alphabet, Justin [71] gave a new proof of a Galois theorem on continued fractions, by considering the epistandard words that are fixed points of  $\mu_w$  and  $\mu_{\tilde{w}}$  for any finite word  $w$ . From this point of view, Justin's result highlights the relevance of the previously mentioned 'multi-dimensional' continued fraction algorithm, proposed by Zamboni [119, 117] (see also [96, Section 12.2]). However, there still remains much work to be done in this direction, especially concerning the *generalized intercept* (coherent with the Sturmian case) introduced in [73, Section 5.4] and the generalized Ostrowski numeration systems [20, 75] (recall Remark 4.7).

*Note.* The aforementioned Galois theorem was used in the theory of Sturmian words to characterize so-called *Sturm numbers* (see [83, Theorem 2.3.26]).

### 6.2.2 Palindromic richness

In [43], Droubay, Justin, and Pirillo observed that any finite word  $w$  contains at most  $|w| + 1$  distinct palindromes (including the empty word). Even further, they proved that a word  $w$  contains exactly  $|w| + 1$  distinct palindromes if and only if the longest palindromic suffix of any prefix  $p$  of  $w$  occurs exactly once in  $p$  (i.e., every prefix of  $w$  has *Property Ju* [43]). Such words are 'rich' in palindromes in the sense that they contain the maximum number of different palindromic factors. Accordingly, we say that a finite word  $w$  is *rich* if it contains exactly  $|w| + 1$  distinct palindromes (or equivalently, if every prefix of  $w$  has *Property Ju*). For example,  $abac$  is rich since it is of length 4 and contains the following five palindromes:  $\varepsilon$ ,  $a$ ,  $b$ ,  $c$ ,  $aba$ . Naturally, an infinite word is rich if all of its factors are rich. For example, the periodic infinite words  $a^\omega = aaa \cdots$  and  $(ab)^\omega = ababab \cdots$  are clearly rich, whereas  $(abc)^\omega = abcabacabc \cdots$  is not rich since it contains the non-rich word  $abca$ .

Droubay *et al.* [43] showed that all finite and infinite episturmian words are rich. Specifically, they proved that if an infinite word has property *Pi* (and hence is epistandard – see Theorem 2.3), then all of its prefixes have property *Ju*. Consequently, any factor  $u$  of an epistandard word (and hence, of an episturmian word) contains exactly  $|u| + 1$  distinct palindromes, and is therefore rich (see Corollary 2 in [43]).

Another special class of rich words that encompasses the episturmian words consists of Fischler's sequences with "abundant palindromic prefixes". These words were introduced and studied in [54, 55] in the context of Diophantine approximation. See also papers by Adamczewski and Bugeaud [2, 3] concerning the transcendence of certain real numbers whose sequences of partial quotients contain arbitrarily long palindromes.

The theory of rich words has recently been further developed in a series of papers [61, 29, 42, 30]. In independent work, Ambrož, Frougny, Masáková, and Pelantová [8] have considered the same class of words which they call *full words*, following the earlier work of Brlek, Hamel, Nivat, and Reutenauer in [26].

### 6.3 Fractional powers & critical exponent

The study of *fractional powers* occurring in Sturmian words has been a topic of growing interest in recent times. See for instance [14, 23, 39, 72, 86, 111], as well as [73, 105, 59] for similar results concerning episturmian words and Arnoux-Rauzy sequences.

The following theorem extends the results in [72] on fractional powers in Sturmian words. Throughout this section, we let  $\mathbf{s}$  denote an epistandard word with directive word  $\Delta = x_1x_2\cdots \in \mathcal{A}^\omega$  (as usual), and for all  $n \geq 1$ , we denote by  $u_{n+1}$  the palindromic prefix  $Pal(x_1\cdots x_n)$  of  $\mathbf{s}$  given in Theorem 2.3. As in [72], we denote by  $L(m)$  the length of the longest factor  $v \in F(\mathbf{s})$  having period  $m \in \mathbb{N}$ , and write  $L(m) = em + r$ ,  $e \in \mathbb{N}^+$ ,  $0 \leq r < m$ . Given a finite or infinite word  $w$ , we denote by  $w(i)$  (resp.  $w(i, j)$ ) the letter in position  $i$  of  $w$  (resp. the factor of  $w$  beginning at position  $i$  and ending at position  $j$ ).

When  $L(m) \geq 2m$ , all factors of  $\mathbf{s}$  having period  $m$  and length  $L(m)$  are equal to a palindrome  $v$ , and for  $0 \leq i < e$ , the word  $v_i := v(1, im + r)$  is a palindromic prefix of  $\mathbf{s}$  by Lemma 4.1 in [73]. Moreover, with the preceding notation, we have:

**Theorem 6.6.** [73, Theorem 4.2] *Let  $m, n \in \mathbb{N}$  be such that  $|u_n| < m \leq |u_{n+1}|$  and  $\mathbf{s}(1, m) = w$  is primitive with  $\mathbf{s}(m) = x$  occurring in  $\mathbf{s}(1, m - 1)$ . Then the following properties hold.*

- i)  $L(m) \geq 2m$  if and only if  $w = \mu_{x_1\cdots x_n}(x)$  and  $x \in \text{Alph}(x_{n+1}x_{n+2}\cdots)$ .
- ii) Suppose  $L(m) \geq 2m$  and define  $p = \max\{i \leq n \mid x_i = x\}$  and  $t = \min\{j \in \mathbb{N}^+ \mid x_{n+j} \neq x\}$ . Then  $u_{n+t} = w^t u_p$  is the longest prefix of  $\mathbf{s}$  having period  $m$ . Moreover, if  $x \in \text{Alph}(x_{n+t+1}x_{n+t+2}\cdots)$ , then  $e = t + 1$ ; that is,  $v = w^{t+1}u_p$ , otherwise  $e = t$  and  $v = w^t u_p$ .  $\square$

**Remark 6.7.** Let us mention a few noteworthy facts.

- Exponents of powers in  $\mathbf{s}$  are bounded if and only if exponents of letters in  $\Delta$  are bounded [105, 73].
- Any Sturmian word has square prefixes and so do epistandard words [5, 105].
- Any episturmian word has infinitely many prefixes of the form  $uv^2$  with  $|u|/|v|$  bounded above.

The latter fact is readily deduced from the following result of Risley and Zamboni [105].

**Theorem 6.8.** [105, Prop. I.3] *If  $\mathbf{t}$  is an Arnoux-Rauzy sequence, then there exists a positive number  $\epsilon$  such that  $\mathbf{t}$  begins with infinitely many blocks of the form  $UVVV'$ , where  $V'$  is a prefix of  $V$  and  $\min\{|V'|/|V|, |V|/|U|\} > \epsilon$ .*  $\square$

*Note.* Such a result is motivated by transcendence issues; see for instance [52].

When  $\mathbf{s}$  is purely morphic, it is possible to give a rather explicit formula for the *critical exponent*:  $\gamma = \limsup_{n \rightarrow \infty} L(m)/m$ , as follows.

**Notation.** Let  $P$  be the function defined by  $P(n) = \sup\{i < n \mid x_i = x_n\}$  if this integer exists, undefined otherwise. That is, if  $x_n = a$ , then  $P(n)$  is the position of the right-most occurrence of the letter  $a$  in the prefix  $x_1x_2\cdots x_{n-1}$  of the directive word  $\Delta = x_1x_2x_3\cdots \in \mathcal{A}^\omega$ .

**Theorem 6.9.** [73, Theorem 5.2] *Let  $\mathbf{s}$  be an  $\mathcal{A}$ -strict epistandard word generated by a morphism with directive word  $\Delta$  having period  $q$ . Further, let  $l \in \mathbb{N}$  be maximal such that  $y^l \in F(\Delta)$  for some letter  $y$ , and define  $L = \{r, 0 \leq r < q \mid x_{r+1} = x_{r+2} = \dots = x_{r+l}\}$  and  $d(r) = r+q+1 - P(r+q-1)$  for  $0 \leq r < q$ . Then the critical exponent for  $\mathbf{s}$  is given by*

$$\gamma = l + 2 + \sup_{r \in L} \left\{ \lim_{i \rightarrow \infty} |u_{r+iq+1-d(r)}| / |h_{r+iq}| \right\}.$$

Moreover, for any letter  $x$  in  $\mathbf{s}$  the limit above can be obtained as a rational function with rational coefficients of the frequency  $\alpha_x$  of this letter.  $\square$

See also [86, 107, 111] for results on the critical exponent for the Fibonacci word, Tribonacci word, and Sturmian words, respectively.

**Example 6.10.** For the ever-so popular Fibonacci word  $\mathbf{f}$ , directed by  $(ab)^\omega$ , we have  $q = 2$ ,  $l = 1$ ,  $d(0) = d(1) = 2$ . Hence, since  $|u_{n-1}|/|h_n|$  has limit  $1/\varphi$  where  $\varphi = (1 + \sqrt{5})/2$  is the golden ratio, we obtain the well-known value  $2 + \varphi$  for the critical exponent, originally proved by Mignosi and Pirillo [86].

More generally, the  $k$ -bonacci word, directed by  $(a_1 a_2 \dots a_k)^\omega$ , has critical exponent  $2 + 1/(\varphi_k - 1)$ , where the  $k$ -bonacci constant  $\varphi_k$  is the (unique) positive real root of the  $k$ -th degree monic polynomial  $x^k - x^{k-1} - \dots - x - 1$ .

## 6.4 Frequencies

Let  $w$  be a non-empty finite word. For any  $v \in F(w)$ , the *frequency* of  $v$  in  $w$  is  $|w|_v/|w|$  where  $|w|_v$  denotes the number of distinct occurrences of  $v$  in  $w$ . The notion of frequency can be extended to infinite words in two ways, as follows.

**Definition 6.11.** *Suppose  $v$  is a non-empty factor of an infinite word  $\mathbf{x}$ . Then:*

- i) *the frequency of  $v$  in  $\mathbf{x}$  in the weak sense is  $\lim_{n \rightarrow \infty} |w(1, n)|_v/n$  if this limit exists;*
- ii)  *$v$  has frequency  $\alpha_v$  in  $\mathbf{x}$  in the strong sense if, for any sequence  $(w_n)_{n \geq 0}$  of factors of  $\mathbf{x}$  with increasing lengths, we have  $\alpha_v = \lim_{n \rightarrow \infty} |w_n|_v/|w_n|$ .*

In a purely combinatorial way, Justin and Pirillo [73, Section 6] proved that any factor occurring in an episturmian word has frequency in the strong sense.

Wozny and Zamboni [117] also studied frequencies (in the weak sense) for Arnoux-Rauzy sequences. Using a reformulation of a vectorial division algorithm, originally introduced in [105], they computed each allowable frequency of factors of the same length, as well as the number of factors with a given frequency. In particular, the authors of [117] gave simultaneous rational approximations of the frequencies by unreduced fractions having a common denominator. From this work, one recovers the results of Berthé [19] for Sturmian words in terms of *Farey* approximations arising from the continued fraction expansions of the frequencies of the letters. For instance, the frequencies of factors of the same length in a Sturmian word assume at most three values, which were explicitly given by Berthé [19], who also discovered that this result is in strong connection with the *three distance theorem* in Diophantine analysis.

## 6.5 Return words

Let us now recall the notion of a *return word*, which was introduced independently by Durand [45], and Holton and Zamboni [67] when studying primitive substitutive sequences.

**Definition 6.12.** *Let  $v$  be a recurrent factor of  $\mathbf{y} \in \mathcal{A}^\omega$ , starting at positions  $n_1 < n_2 < n_3 \dots$ . Then each word  $r_i = y_{n_i}y_{n_i+1} \dots y_{n_{i+1}-1}$  is called a return to  $v$  in  $\mathbf{y}$ . Moreover,  $\mathbf{y}$  can be factorized in a unique way as  $\mathbf{y} = y_1 \dots y_{n_1-1}r_1r_2r_3 \dots$  where  $r_1r_2r_3 \dots$ , viewed as a word on the  $r_i$ , is called the derived word of  $\mathbf{y}$  with respect to  $v$ .*

That is, a return to  $v$  in  $\mathbf{y}$  is a non-empty factor of  $\mathbf{y}$  beginning at an occurrence of  $v$  and ending exactly before the next occurrence of  $v$  in  $\mathbf{y}$ . Thus, if  $r$  is a return to  $v$  in  $\mathbf{y}$ , then  $rv$  is a factor of  $\mathbf{y}$  that contains exactly two occurrences of  $v$ , one as a prefix and one as a suffix. We call  $rv$  a *complete return* to  $v$  [76].

Return words play an important role in the study of *minimal subshifts* in symbolic dynamics; see for instance [45, 46, 47, 53, 106]. In the context of episturmian words, such words have recently proven to be a useful tool in the study of quasiperiodicity (see Section 8 for further details). This latest work made use of the following result of Justin and Vuillon [76] which completely describes the returns to any factor of an epistandard word. In fact, their result actually characterizes return words in episturmian words (not just epistandard words) since, by uniform recurrence, the returns to any factor  $v$  in an epistandard word  $\mathbf{s}$  are the same as the returns to  $v$  as a factor of any episturmian word  $\mathbf{t}$  having the same set of factors as  $\mathbf{s}$ .

**Theorem 6.13.** [76] *Let  $\mathbf{s}$  be an epistandard word directed by  $\Delta = x_1x_2x_3 \dots \in \mathcal{A}^\omega$  and consider any  $v \in F(\mathbf{s})$ . If  $u_{n+1}$  is the shortest palindromic prefix of  $\mathbf{s}$  containing  $v$  with  $u_{n+1} = fvg$ , then the returns to  $v$  in  $\mathbf{s}$  are the words  $f^{-1}\mu_{x_1 \dots x_n}(x)f$  where  $x \in \text{Alph}(x_{n+1}x_{n+2} \dots)$ . Moreover, the corresponding complete returns to  $v$  are the words  $f^{-1}(u_{n+1}x)^{(+)}g^{-1}$  and the derived word of  $\mathbf{s}$  with respect to  $v$  is given by  $\mathbf{s}^{(n)} = \mu_{x_1 \dots x_n}^{-1}(\mathbf{s})$ .  $\square$*

*Note.* It follows immediately that any factor of an  $\mathcal{A}$ -strict episturmian word has exactly  $|\mathcal{A}|$  return words.

Theorem 6.13 extends earlier work of Vuillon on return words in Sturmian words (see [114]). In particular, Vuillon proved that Sturmian words are characterized by the property that any non-empty factor has exactly 2 different return words in the given Sturmian word. However, contrary to what one might expect, such a property with 2 replaced by a positive integer  $k \geq 3$  does not characterize  $k$ -strict episturmian words. For instance, infinite words coding 3-interval exchange transformations, which constitute a different generalization of Sturmian words to 3-letter alphabets, are known to have the property that every factor has 3 different return words (see the work by Ferenczi, Holton, and Zamboni in [51]).

## 7 Balance & lexicographic order

### 7.1 $q$ -Balance

**Definition 7.1.** *A finite or infinite word is  $q$ -balanced if, for any two of its factors  $u, v$  with  $|u| = |v|$ , we have*

$$||u|_x - |v|_x| \leq q \quad \text{for any letter } x,$$

*i.e., the number of  $x$ 's in each of  $u$  and  $v$  differs by at most  $q$ .*

*Note.* A 1-balanced word is simply said to be *balanced*.

The term ‘balanced’ is relatively new; it appeared in [16, 17] (also see [83, Chapter 2]), and the notion itself dates back to [91, 37]. In the pioneering work of Morse and Hedlund [91], balanced infinite words over a 2-letter alphabet were called ‘Sturmian trajectories’ and belong to three classes: aperiodic Sturmian; periodic Sturmian; and infinite words that are ultimately periodic (but not periodic), called *skew words*. That is, the family of balanced infinite words consists of the (recurrent) Sturmian words and the (non-recurrent) skew infinite words, the factors of which are balanced. Skew words are ultimately periodic (but not periodic) suffixes of words of the form  $\mu(a^pba^\omega)$ , where  $\mu$  is a *pure standard Sturmian morphism* and  $p \in \mathbb{N}$ . For example,  $aba^\omega$  and  $\psi_b(aba^\omega) = bab(ba)^\omega$  are skew. See also [108, 109, 66, 95] for further work on skew words.

**Remark 7.2.** Nowadays, for most authors, only the aperiodic Sturmian words are considered to be ‘Sturmian’. However, from now on, we will use the term ‘Sturmian’ to refer to both aperiodic and periodic Sturmian words. In the context of cutting sequences, the aperiodic (resp. periodic) Sturmian words are precisely those with irrational slope (resp. rational slope).

It is important to note that a finite word is *finite Sturmian* (i.e., a factor of some Sturmian word) if and only if it is balanced [83, Chapter 2]. Accordingly, the balanced infinite words are precisely the infinite words whose factors are finite Sturmian. This concept was recently generalized in [62] by showing that the set of all infinite words whose factors are finite episturmian consists of the (recurrent) episturmian words and the (non-recurrent) *episkew* infinite words, as defined in the next section.

## 7.2 Episkew words

Inspired by the skew words of Morse and Hedlund [91], *episkew words* were recently defined in [62] as non-recurrent infinite words, all of whose factors are (finite) episturmian. The following theorem gives a number of equivalent definitions of such words, similar to those for (recurrent) episturmian words.

**Theorem 7.3.** [62] *An infinite word  $\mathbf{t}$  with  $\text{Alph}(\mathbf{t}) = \mathcal{A}$  is episkew if equivalently:*

- i)  $\mathbf{t}$  is non-recurrent and all of its factors are (finite) episturmian;
- ii) there exists an infinite sequence  $(\mathbf{t}^{(i)})_{i \geq 0}$  of non-recurrent infinite words and a directive word  $x_1x_2x_3 \cdots$  ( $x_i \in \mathcal{A}$ ) such that  $\mathbf{t}^{(0)} = \mathbf{t}, \dots, \mathbf{t}^{(i-1)} = \psi_{x_i}(\mathbf{t}^{(i)})$ , where  $\mathbf{t}^{(i-1)} = \mathbf{t}^{(i-1)}$  if  $\mathbf{t}^{(i-1)}$  begins with  $x_i$  and  $\mathbf{t}^{(i-1)} = x_i\mathbf{t}^{(i-1)}$  otherwise;
- iii) there exists a letter  $x \in \mathcal{A}$  and an epistandard word  $\mathbf{s}$  on  $\mathcal{A} \setminus \{x\}$  such that  $\mathbf{t} = v\mu(\mathbf{s})$ , where  $\mu$  is a pure epistandard morphism on  $\mathcal{A}$  and  $v$  is a non-empty suffix of  $\mu(\tilde{\mathbf{s}}_p x)$  for some  $p \in \mathbb{N}$ .

Moreover,  $\mathbf{t}$  is said to be *strict episkew* if  $\mathbf{s}$  is strict on  $\mathcal{A} \setminus \{x\}$ , i.e., if each letter in  $\mathcal{A} \setminus \{x\}$  occurs infinitely often in the directive word  $x_1x_2x_3 \cdots$ .  $\square$

A simple example of an episkew word on more than two letters is the infinite word  $c\mathbf{f} = cabaababa \cdots$  where  $\mathbf{f}$  is the Fibonacci word and  $c$  is a letter (see also Example 3.7).

Note that the episkew words on a 2-letter alphabet are precisely the skew words. Certainly, in the Sturmian case, the word  $\tilde{\mathbf{s}}_p x \mathbf{s}$  reduces to a word of the form  $a^pba^\omega$ .

**Remark 7.4.** Thanks to Richomme [104], we recently determined that episkew words actually have the following simpler characterization: *an infinite word  $\mathbf{t}$  is episkew if and only if  $\mathbf{t}$  is non-recurrent and  $\mathbf{t} = \varphi(xs)$  where  $\mathbf{s}$  is an epistandard word,  $x$  is a letter not occurring in  $\mathbf{s}$ , and  $\varphi$  is a pure episturmian morphism.* Note that the given condition on the form of  $\mathbf{t}$  is not sufficient for  $\mathbf{t}$  to be episkew, i.e., we must also assume that  $\mathbf{t}$  is non-recurrent. Indeed, if we consider for instance the infinite word  $\mathbf{t} = \psi_x(x\mathbf{f})$  where  $\mathbf{f}$  is the Fibonacci word and  $x$  is a letter different from  $a$  and  $b$ , then  $\mathbf{t} = xaxbxaaxbxa \cdots = \bar{\psi}_x(\mathbf{f})$  – a recurrent episturmian word (not episkew).

Episkew words were first alluded to (but not explicated) in the recent paper [60]. Following that paper, these words showed up again in the study of inequalities characterizing finite and infinite episturmian words with respect to lexicographic orderings [62]. In fact, as detailed in the next section, episturmian words have similar extremal properties to Sturmian words. See also [74, 69, 93, 94, 95, 60, 62] for other work in this direction.

### 7.3 Extremal properties

Suppose the alphabet  $\mathcal{A}$  is totally ordered by the relation  $<$ . Then we can totally order  $\mathcal{A}^*$  by the *lexicographic order*  $\leq$  defined as follows. Given two words  $u, v \in \mathcal{A}^+$ , we have  $u \leq v$  if and only if either  $u$  is a prefix of  $v$  or  $u = xau'$  and  $v = xbv'$ , for some  $x, u', v' \in \mathcal{A}^*$  and letters  $a, b$  with  $a < b$ . This is the usual alphabetic ordering in a dictionary. We write  $u < v$  when  $u \leq v$  and  $u \neq v$ , in which case we say that  $u$  is (strictly) *lexicographically smaller* than  $v$ . The notion of lexicographic order naturally extends to infinite words in  $\mathcal{A}^\omega$ . We denote by  $\min(\mathcal{A})$  the smallest letter in  $\mathcal{A}$  with respect to the given lexicographic order.

Let  $w$  be a finite or infinite word over  $\mathcal{A}$  and let  $k$  be a positive integer. We denote by  $\min(w|k)$  (resp.  $\max(w|k)$ ) the lexicographically smallest (resp. greatest) factor of  $w$  of length  $k$  for the given order (where  $|w| \geq k$  if  $w$  is finite). If  $w$  is infinite, then it is clear that  $\min(w|k)$  and  $\max(w|k)$  are prefixes of the respective words  $\min(w|k+1)$  and  $\max(w|k+1)$ . So we can define, by taking limits, the following two infinite words (see [94]):

$$\min(w) = \lim_{k \rightarrow \infty} \min(w|k) \quad \text{and} \quad \max(w) = \lim_{k \rightarrow \infty} \max(w|k).$$

That is, to any infinite word  $\mathbf{t}$  we can associate two infinite words  $\min(\mathbf{t})$  and  $\max(\mathbf{t})$  such that any prefix of  $\min(\mathbf{t})$  (resp.  $\max(\mathbf{t})$ ) is the lexicographically smallest (resp. greatest) amongst the factors of  $\mathbf{t}$  of the same length.

For a finite word  $w$  over  $\mathcal{A}$  and a given order on  $\mathcal{A}$ ,  $\min(w)$  denotes  $\min(w|k)$  where  $k$  is maximal such that all  $\min(w|j)$ ,  $j = 1, 2, \dots, k$ , are prefixes of  $\min(w|k)$ . In the case  $\mathcal{A} = \{a, b\}$ ,  $\max(w)$  is defined similarly (see [62]).

In 2003, Pirillo [93] (see also [94]) proved that, for infinite words  $\mathbf{s}$  on a 2-letter alphabet  $\{a, b\}$  with  $a < b$ , the inequality

$$a\mathbf{s} \leq \min(\mathbf{s}) \leq \max(\mathbf{s}) \leq b\mathbf{s} \tag{7.1}$$

characterizes *standard Sturmian words* (aperiodic and periodic). Actually, this result was known much earlier, dating back to the work of P. Veerman [112, 113] in the mid 80's. Since that time, these ‘Sturmian inequalities’ have been rediscovered numerous times under different guises, as discussed in the forthcoming survey paper [6].

Continuing his work in relation to inequality (7.1), Pirillo [94] proved further that, in the case of an arbitrary finite alphabet  $\mathcal{A}$ , an infinite word  $\mathbf{s} \in \mathcal{A}^\omega$  is epistandard if and only if, for any lexicographic order, we have

$$a\mathbf{s} \leq \min(\mathbf{s}) \quad \text{where } a = \min(\mathcal{A}). \tag{7.2}$$



Moreover,  $\mathbf{s}$  is a strict epistandard word if and only if (7.2) holds with strict equality for any order [74].

In a similar spirit, Glen, Justin, and Pirillo [62] recently established new characterizations of finite Sturmian and episturmian words via lexicographic orderings. As a consequence, they were able to characterize by lexicographic order all episturmian and episkew words. Similarly, they characterized by lexicographic order all balanced infinite words on a 2-letter alphabet; in other words, all Sturmian and skew infinite words, the factors of which are (finite) Sturmian. In the finite case:

**Theorem 7.5.** [62] *A finite word  $w$  on  $\mathcal{A}$  is episturmian if and only if there exists a finite word  $u$  such that, for any lexicographic order,*

$$au_{|m|-1} \leq m \quad (7.3)$$

where  $m = \min(w)$  and  $a = \min(\mathcal{A})$  for the considered order.  $\square$

**Example 7.6.** Consider the finite word  $w = baabacababac$ . For the different orders on  $\{a, b, c\}$ , we have

- $a < b < c$  or  $a < c < b$ :  $\min(w) = aabacababac$ ,
- $b < a < c$  or  $b < c < a$ :  $\min(w) = babac$ ,
- $c < a < b$  or  $c < b < a$ :  $\min(w) = cababac$ .

It can be verified that a finite word  $u$  satisfying (7.3) must begin with  $aba$  and one possibility is  $u = abacaaaaa$ ; thus  $w$  is a finite episturmian word.

*Note.* In the above example, any two orders with the same minimum letter give the same  $\min(w)$ , which is not true in general.

A corollary of Theorem 7.5 is the following new characterization of finite Sturmian words (i.e., finite balanced words).

**Corollary 7.7.** [62] *A finite word  $w$  on  $\mathcal{A} = \{a, b\}$ ,  $a < b$ , is not Sturmian (in other words, not balanced) if and only if there exists a finite word  $u$  such that  $aua$  is a prefix of  $\min(w)$  and  $bub$  is a prefix of  $\max(w)$ .  $\square$*

In the infinite case, the following characterization of all infinite words whose factors are finite episturmian follows almost immediately from Theorem 7.5.

**Corollary 7.8.** [62] *An infinite word  $\mathbf{t}$  on  $\mathcal{A}$  is episturmian or episkew if and only if there exists an infinite word  $\mathbf{u}$  such that, for any lexicographic order,*

$$a\mathbf{u} \leq \min(\mathbf{t}) \quad \text{where } a = \min(\mathcal{A}).$$

$\square$

Consequently, an infinite word  $\mathbf{s}$  on  $\{a, b\}$  ( $a < b$ ) is balanced (i.e., Sturmian or skew) if and only if there exists an infinite word  $\mathbf{u}$  such that

$$a\mathbf{u} \leq \min(\mathbf{s}) \leq \max(\mathbf{s}) \leq b\mathbf{u}.$$

Corollary 7.8 was recently refined in [58] where it was shown that, for any aperiodic episturmian word  $\mathbf{t}$ , the infinite word  $\mathbf{u}$  (as given in the corollary) is the unique epistandard word with the

same set of factors as  $\mathbf{t}$ . As an easy consequence, we obtain the following characterization of strict episturmian words that are *infinite Lyndon words* (Theorem 7.9). Recall that a non-empty finite word  $w$  over  $\mathcal{A}$  is a *Lyndon word* if it is lexicographically smaller than all of its proper suffixes for the given order  $<$  on  $\mathcal{A}$ . Equivalently,  $w$  is the lexicographically smallest primitive word in its conjugacy class; that is,  $w < vu$  for all non-empty words  $u, v$  such that  $w = uv$ . The first of these definitions extends to infinite words: an infinite word over  $\mathcal{A}$  is an *infinite Lyndon word* if and only if it is (strictly) lexicographically smaller than all of its proper suffixes for the given order on  $\mathcal{A}$ . That is, a finite or infinite word  $w$  is a Lyndon word if and only if  $w < T^i(w)$  for all  $i > 0$ .

Assuming that  $|\mathcal{A}| > 1$  (since there are no Lyndon words on a 1-letter alphabet), we have:

**Theorem 7.9.** [58] *An  $\mathcal{A}$ -strict episturmian word  $\mathbf{t}$  is an infinite Lyndon word if and only if  $\mathbf{t} = as$  where  $a = \min(\mathcal{A})$  for the given order on  $\mathcal{A}$  and  $\mathbf{s}$  is an (aperiodic)  $\mathcal{A}$ -strict epistandard word. Moreover, if  $\Delta = x_1x_2\cdots \in \mathcal{A}^\omega$  is the directive word of  $\mathbf{s}$ , then  $\mathbf{t} = as$  is the unique episturmian word in the subshift of  $\mathbf{s}$  directed by the spinned version of  $\Delta$  having all spins 1, except when  $x_i = a$ .  $\square$*

The above theorem is actually a generalization of a result on (aperiodic) Sturmian words given by Borel and Laubie [24] (see also [102]).

Let  $\mathcal{A} = \{a_1, \dots, a_m\}$  be an alphabet ordered by  $a_1 < a_2 < \cdots < a_m$  with  $m \geq 2$ . Then Theorem 7.9 says that an  $\mathcal{A}$ -strict episturmian word  $\mathbf{t}$  is an infinite Lyndon word if and only if the (normalized) directive word of  $\mathbf{t}$  belongs to  $\{a_1, \bar{a}_2, \dots, \bar{a}_m\}^\omega$ . This can be reformulated as a generalization of Proposition 6.4 in [81]:

**Corollary 7.10.** [58] *An  $\mathcal{A}$ -strict episturmian word  $\mathbf{t}$  is an infinite Lyndon word if and only if it can be infinitely decomposed over the set of morphisms  $\{\psi_a, \bar{\psi}_x \mid x \in \mathcal{A} \setminus \{a\}\}$  where  $a = \min(\mathcal{A})$  for the given order on  $\mathcal{A}$ .  $\square$*

We observe that, contrary to the fact that there exists  $|\mathcal{A}|!$  possible orders of a finite alphabet  $\mathcal{A}$ , Theorem 7.9 shows that there exist exactly  $|\mathcal{A}|$  infinite Lyndon words in the subshift of a given  $\mathcal{A}$ -strict epistandard word  $\mathbf{s}$  (when  $|\mathcal{A}| > 1$ ). That is, for any order with  $\min(\mathcal{A}) = a$ , the subshift of  $\mathbf{s}$  contains a unique infinite Lyndon word beginning with  $a$ , namely  $as$ .

**Example 7.11.** With  $\Delta = (abc)^\omega$ , the spinned versions  $(a\bar{b}\bar{c})^\omega$ ,  $(\bar{a}b\bar{c})^\omega$ ,  $(\bar{a}\bar{b}c)^\omega$  and their ‘opposites’ (obtained by exchange of spins):  $(\bar{a}\bar{b}c)^\omega$ ,  $(a\bar{b}c)^\omega$ ,  $(ab\bar{c})^\omega$  direct episturmian words in the subshift of the Tribonacci word  $\mathbf{r}$ . Only the first three of these spinned infinite words direct episturmian Lyndon words:  $a\mathbf{r}$ ,  $b\mathbf{r}$ ,  $c\mathbf{r}$ , respectively.

The above results on strict episturmian Lyndon words have very recently been generalized to all episturmian words by Glen, Levé, and Richomme [64], as follows.

**Theorem 7.12.** [64] *Let  $\mathcal{A} = \{a_1, \dots, a_m\}$  be an alphabet ordered by  $a_1 < a_2 < \cdots < a_m$  and, for  $1 \leq i \leq m$ , let  $\mathcal{B}_i = \{a_i, \dots, a_m\}$ . An episturmian word  $\mathbf{t}$  is an infinite Lyndon word if and only if there exists an integer  $j$  such that  $1 \leq j < m$  and the (normalized) directive word of  $\mathbf{w}$  belongs to:*

$$(\bar{\mathcal{B}}_2^* a_1)^* \cdots (\bar{\mathcal{B}}_j^* a_{j-1})^* (\bar{\mathcal{B}}_{j+1}^* a_j)^* (\bar{\mathcal{B}}_{j+1}^+ \{a_j\}^+)^*.$$

$\square$

*Note.* In the above theorem, the word *normalized* appears between brackets since one can easily verify from Theorem 4.13 that a spinned infinite word of the given form is the unique directive word of exactly one episturmian word.

**Example 7.13.** [64] Let  $\mathcal{A} = \{a, b, c, d\}$ . Then the spinned infinite word  $(\bar{b}\bar{c}a)(\bar{d}\bar{c}b)^2(\bar{d}cc)^\omega$  directs a Lyndon episturmian word, and so does  $aa(\bar{d}c)^\omega$ , but  $\bar{c}a\bar{b}a\bar{d}c\bar{d}^\omega$  does not (since this spinned word directs a periodic word).

**Remark 7.14.** Theorems 4.13 and 7.12 show that any episturmian Lyndon word has a unique spinned directive word, but the converse is not true. For example, the regular wavy word  $(a\bar{b}c)^\omega$  is the unique directive word of the strict episturmian word:

$$\lim_{n \rightarrow \infty} \mu_{abc}^n(a) = acabaabacabacabaabaca \dots$$

which is clearly not an infinite Lyndon word by Theorem 7.12 and also by the fact that  $acabaaw$  is not a Lyndon word for any order on  $\{a, b, c\}$  and for any word  $w$ .

A key tool used in the proof of Theorem 7.12 was the following result of Richomme, which characterizes episturmian morphisms that preserve Lyndon words. A morphism  $f$  is said to preserve finite (resp. infinite) Lyndon words if for each finite (resp. infinite) Lyndon word  $w$ ,  $f(w)$  is a finite (resp. infinite) Lyndon word.

**Theorem 7.15.** [100, 103] *Let  $\mathcal{A} = \{a_1, \dots, a_m\}$  be an alphabet ordered by  $a_1 < a_2 < \dots < a_m$ . Then the following assertions are equivalent for an episturmian morphism:*

- $f$  preserves finite Lyndon words;
- $f$  preserves infinite Lyndon words;
- $f \in (\bar{\Psi}_{\{a_2, \dots, a_m\}}^* \psi_{a_1})^* \{\bar{\Psi}_{a_m}\}^*$  where  $\bar{\Psi}_{\mathcal{A}} = \{\bar{\psi}_x \mid x \in \mathcal{A}\}$ . □

## 7.4 Imbalance

We now return our attention to the notion of balance.

Episturmian words on three or more letters are generally unbalanced in the sense of 1-balance, except, of course, for those on a 2-letter alphabet, which are precisely the (periodic and aperiodic) Sturmian words. In fact, Cassaigne, Ferenczi, and Zamboni [33] have proved, by construction, that there exists an episturmian word that is not  $q$ -balanced for any  $q$ . Note, however, that the Tribonacci word is 2-balanced, for example. More generally, it can be shown by induction that the  $k$ -bonacci word, directed by  $(a_1 a_2 \dots a_k)^\omega$ , is  $(k - 1)$ -balanced. Even further, one can prove that any *linearly recurrent* strict episturmian word (or Arnoux-Rauzy sequence) is  $q$ -balanced for some  $q$ . Linearly recurrent Arnoux-Rauzy sequences were completely described in [105, 32]; they are the strict episturmian words for which each letter  $x$  occurs in  $\Delta$  with bounded gaps.

Using their main result on return words (Theorem 6.13), Justin and Vuillon [76] proved that episturmian words do in fact satisfy a kind of balance property. Specifically:

**Theorem 7.16.** [76, Theorem 5.2] *Let  $\mathbf{s} \in \mathcal{A}^\omega$  be an epistandard word and let  $\{d, e\}$  be a 2-letter subset of  $\mathcal{A}$ . Then, for any  $u, v \in F(\mathbf{s}) \cap \{d, e\}^*$  with  $|u| = |v|$ , we have  $||u|_d - |v|_d| \leq 1$ . □*

This property of episturmian words reduces to the balance property of Sturmian words when  $\mathcal{A}$  is a 2-letter alphabet (in which case it is characteristic); however, the property is far from being characteristic when  $\mathcal{A}$  consists of more than two letters.

More recently, Richomme [101] also proved that episturmian words and Arnoux-Rauzy sequences can be characterized via a nice ‘local balance property’. That is:

**Theorem 7.17.** [101] *For a recurrent infinite word  $\mathbf{t} \in \mathcal{A}^\omega$ , the following assertions are equivalent:*

- i)  $\mathbf{t}$  is episturmian;*
- ii) for each factor  $u$  of  $\mathbf{t}$ , there exists a letter  $a$  such that  $\mathcal{A}u\mathcal{A} \cap F(\mathbf{t}) \subseteq au\mathcal{A} \cup \mathcal{A}ua$ ;*
- iii) for each palindromic factor  $u$  of  $\mathbf{t}$ , there exists a letter  $a$  such that  $\mathcal{A}u\mathcal{A} \cap F(\mathbf{t}) \subseteq au\mathcal{A} \cup \mathcal{A}ua$ .* □

Roughly speaking, the above theorem says that for any factor  $u$  of a given episturmian word  $\mathbf{t}$ , there exists a unique letter  $a$  such that every occurrence of  $u$  in  $\mathbf{t}$  is immediately preceded or followed by  $a$  in  $\mathbf{t}$ . When  $|\mathcal{A}| = 2$ , property *ii)* of Theorem 7.17 is equivalent to the definition of balance. Indeed, Coven and Hedlund [37] stated that an infinite word  $\mathbf{w}$  over  $\{a, b\}$  is not balanced if and only if there exists a palindrome  $u$  such that  $aua$  and  $bub$  are both factors of  $\mathbf{w}$ . As pointed out in [101], this property can be rephrased as follows: an infinite word  $\mathbf{w}$  is Sturmian if and only if  $\mathbf{w}$  is aperiodic and, for any factor  $u$  of  $\mathbf{w}$ , the set of factors belonging to  $\mathcal{A}u\mathcal{A}$  is a subset of  $au\mathcal{A} \cup \mathcal{A}ua$  or a subset of  $bu\mathcal{A} \cup \mathcal{A}ub$ .

## 7.5 Fraenkel's conjecture

As discussed previously, the *recurrent* balanced infinite words on two letters are exactly the Sturmian words (aperiodic and periodic). A natural question to ask is then: “What are the balanced recurrent infinite words on more than two letters?” In this direction, Paquin and Vuillon [92] recently characterized the balanced episturmian words by classifying these words into three families, as follows.

**Theorem 7.18.** [92] *Any balanced standard episturmian sequence  $\mathbf{s}$  on a  $k$ -letter alphabet  $\mathcal{A}_k = \{1, 2, \dots, k\}$ ,  $k \geq 3$ , belongs to one of the following three families (up to letter permutation):*

- i)  $\mathbf{s} = p(k-1)p(kp(k-1)p)^\omega$ , with  $p = \text{Pal}(1^n 2 \cdots (k-2))$ ;*
- ii)  $\mathbf{s} = p(k-1)p(kp(k-1)p)^\omega$ , with*

$$p = \text{Pal}(123 \cdots (k-\ell-1)1(k-\ell) \cdots (k-2));$$
- iii)  $\mathbf{s} = [\text{Pal}(123 \cdots k)]^\omega$ .* □

The importance of the above result lies in the fact that it supports *Fraenkel's conjecture* [56]: a problem that arose in a number-theoretic context and has remained unsolved for over thirty years. Fraenkel conjectured that, for a fixed  $k \geq 3$ , there is only one covering of  $\mathbb{Z}$  by  $k$  *Beatty sequences* of the form  $(\lfloor \alpha n + \beta \rfloor)_{n \geq 1}$ , where  $\alpha, \beta$  are real numbers. A combinatorial interpretation of this conjecture may be stated as follows (taken from [92]). Over a  $k$ -letter alphabet with  $k \geq 3$ , there is only one recurrent balanced infinite word, up to letter permutation and shifts, that has mutually distinct letter frequencies. This supposedly unique infinite word is called *Fraenkel's sequence* and is given by  $(F_k)^\omega$  where the *Fraenkel words*  $(F_i)_{i \geq 1}$  are defined recursively by  $F_1 = 1$  and  $F_i = F_{i-1}iF_{i-1}$  for all  $i \geq 2$ . (Note that  $F_k = \text{Pal}(12 \cdots k)$ .) For further details, see for instance [92, 110] and references therein.

Amongst the classes of balanced episturmian words given in Theorem 7.18, only one class has mutually distinct letter frequencies and, up to letter permutation and shifts, corresponds to Fraenkel's sequence. That is:

**Theorem 7.19** (Paquin-Vuillon [92]). *Suppose  $\mathbf{t}$  is a balanced episturmian word with  $\text{Alph}(\mathbf{t}) = \{1, 2, \dots, k\}$ ,  $k \geq 3$ . If  $\mathbf{t}$  has mutually distinct letter frequencies, then up to letter permutation,  $\mathbf{t}$  is a shift of  $(F_k)^\omega$ .  $\square$*

More recently, it was proved in [61] that any recurrent balanced *rich* infinite word is necessarily episturmian, and hence such words obey Fraenkel’s conjecture (recall that rich words were defined Section 6.2.2).

**Remark 7.20.** An interesting known fact (e.g., see [68]) is that any balanced recurrent infinite word  $\mathbf{x}$  on  $k \geq 3$  letters having mutually distinct letter frequencies is necessarily periodic. Certainly, the image of  $\mathbf{x}$  under any morphism of the form:  $(a \mapsto a, \text{ other } x \mapsto b)$  is a Sturmian word. If, for one letter, the corresponding Sturmian word is aperiodic (i.e.,  $\mathbf{x}$  has irrational slope as a cutting sequence), then we meet impossibility; thus rather easily  $\mathbf{x}$  must be periodic.

## 8 Concluding remarks

In closing, we mention a number of very recent works involving episturmian words.

**Rigidity:** Krieger [78] has shown that any strict purely morphic epistandard word  $\mathbf{s}$  is *rigid*. That is, all of the morphisms that generate  $\mathbf{s}$  are powers of the same unique (epistandard) morphism. Krieger also showed that a certain class of ‘ultimately strict’ purely morphic epistandard words are not rigid, but it remains an open question as to whether or not all strict purely morphic episturmian words are rigid.

**Quasiperiodicity:** A finite or infinite word  $w$  is said to be *quasiperiodic* if there exists a word  $u$  (with  $u \neq w$  for finite  $w$ ) such that the occurrences of  $u$  in  $w$  entirely cover  $w$ , i.e., every position of  $w$  falls within some occurrence of  $u$  in  $w$ . Such a word  $u$  is called a *quasiperiod* of  $w$ . For example, the word  $w = abaababaabaababaaba$  has quasiperiods  $aba$ ,  $abaaba$ ,  $abaababaaba$ .

In the last fifteen years, quasiperiodicity and coverings of finite words has been extensively studied (see [9] for a brief survey on quasiperiodicity in ‘strings’). Quasiperiodic finite words were first introduced by Apostolico and Ehrenfeucht in [10]. The notion was later extended to infinite words by Marcus [85] who opened some questions, particularly concerning quasiperiodicity of Sturmian words. After a brief answer to some of these questions in [79], the Sturmian case was fully studied by Levé and Richomme [81] who proved that a Sturmian word is non-quasiperiodic if and only if it is an infinite Lyndon word. The study of quasiperiodicity in Sturmian words was very recently extended to episturmian words by Glen, Levé, and Richomme [58, 64, 80], who have completely described all of the quasiperiods of an episturmian word, yielding a characterization of quasiperiodic episturmian words in terms of their directive words. They have also characterized episturmian morphisms that map any word onto a quasiperiodic one. These results show that, unlike the Sturmian case, there exist non-quasiperiodic episturmian words that are not infinite Lyndon words. Key tools used in the study of quasiperiodicity in episturmian words were episturmian morphisms, normalized directive words (recall Theorem 4.12), and the following equivalent definition of quasiperiodicity in terms of return words introduced by Glen in [58]: a finite word  $v$  is a quasiperiod of an infinite word  $\mathbf{w}$  if and only if  $v$  is a recurrent prefix of  $\mathbf{w}$  such that all of the returns to  $v$  in  $\mathbf{w}$  have length at most  $|v|$ .

In [89], Monteil proved that any Sturmian subshift contains a *multi-scale quasiperiodic word*, i.e., an infinite word having infinitely many quasiperiods. A shorter proof of this fact was provided in [81] and this result has also been proven true for episturmian words in [64].

For more recent work on quasiperiodicity, see for instance [89, 90].

**$\theta$ -episturmian words:** Recall that an infinite word is episturmian if and only if its set of factors is closed under reversal and it has at most one left special factor of each length. With this definition in mind, Bucci, de Luca, De Luca, and Zamboni [27, 28] have recently introduced and studied a further extension of episturmian words in which the reversal operator is replaced by an arbitrary *involutory antimorphism* (i.e., a map  $\theta : \mathcal{A}^* \rightarrow \mathcal{A}^*$  such that  $\theta^2 = \text{Id}$  and  $\theta(uv) = \theta(v)\theta(u)$  for all  $u, v \in \mathcal{A}^*$ ). More precisely, an infinite word over  $\mathcal{A}$  is said to be  $\theta$ -episturmian if it has at most one left special factor of each length and its set of factors is closed under an involutory antimorphism  $\theta$  of the free monoid  $\mathcal{A}^*$ . Generalizing even further,  *$\theta$ -episturmian words with seed* are obtained by requiring the condition on special factors only for sufficiently large lengths (see [28]).

**Acknowledgements.** The authors would like to thank Jean Berstel and Pierre Arnoux for their helpful comments on a preliminary version of this paper. Many thanks also to the two anonymous referees whose thoughtful suggestions helped to improve the paper.

## References

- [1] B. Adamczewski, Balances for fixed points of primitive substitutions, *Theoret. Comput. Sci.* 307 (2003) 47–75.
- [2] B. Adamczewski, Y. Bugeaud, Palindromic continued fractions, *Ann. Inst. Fourier (Grenoble)* 57 (2007) 1557–1574.
- [3] B. Adamczewski, Y. Bugeaud, Transcendence measure for continued fractions involving repetitive or symmetric patterns, *J. Eur. Math. Soc.*, to appear.
- [4] P. Alessandri, V. Berthé, Three distance theorems and combinatorics on words, *Enseign. Math.* 44 (1998) 103–132.
- [5] J.-P. Allouche, J.L. Davison, M. Queffélec, L.Q. Zamboni, Transcendence of Sturmian or morphic continued fractions, *J. Number Theory* 91 (2001) 39–66.
- [6] J.-P. Allouche, A. Glen, Extremal properties of (epi)sturmian sequences and distribution modulo 1, in preparation.
- [7] J.-P. Allouche, J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, UK, 2003.
- [8] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers, *Ann. Inst. Fourier (Grenoble)* 56 (2006) 2131–2160.
- [9] A. Apostolico, M. Crochemore, String pattern matching for a deluge survival kit, in: *Handbook of Massive Data Sets, Massive Comput.*, vol. 4, Kluwer Acad. Publ., Dordrecht, 2002.
- [10] A. Apostolico, A. Ehrenfeucht, Efficient detection of quasiperiodicities in strings, *Theoret. Comput. Sci.* 119 (1993) 247–265.
- [11] P. Arnoux, S. Ito, Pisot substitutions and Rauzy fractals, *Bull. Belg. Math. Soc. Simon Stevin* 8 (2001) 181–207.



- [12] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité  $2n + 1$ , *Bull. Soc. Math. France* 119 (1991) 199–215.
- [13] Yu. Baryshnikov, Complexity of trajectories in rectangular billiards, *Comm. Math. Phys.* 174 (1995) 43–56.
- [14] J. Berstel, On the index of Sturmian words, in: *Jewels Are Forever*, Springer-Verlag, Berlin, 1999, pp. 287–294.
- [15] J. Berstel, Recent results on extensions of Sturmian words, *Internat. J. Algebra Comput.* 12 (2002) 371–385.
- [16] J. Berstel, P. Séébold, A characterization of Sturmian morphisms, in: *Borzyszkowski, A.M. and Sokolowski, S. (Eds.), Mathematical Foundations of Computer Science 1993, Lecture Notes in Computer Science*, vol. 711, Springer-Verlag, Berlin, 1993, pp. 281–290.
- [17] J. Berstel, P. Séébold, A remark on morphic Sturmian words, *Theor. Inform. Appl.* 28 (1994) 255–263.
- [18] J. Berstel, L. Vuillon, Coding rotations on intervals, *Theoret. Comput. Sci.* 281 (2002) 99–107.
- [19] V. Berthé, Fréquences des facteurs des suites sturmiennes, *Theoret. Comput. Sci.* 165 (1996) 295–309.
- [20] V. Berthé, Autour du système de numération d’Ostrowski, *Bull. Belg. Math. Soc. Simon Stevin* 8 (2001) 209–239.
- [21] V. Berthé, H. Ei, S. Ito, and H. Rao, On substitution invariant Sturmian words: an application of Rauzy fractals, *Theoret. Inform. Appl.* 41 (2007) 329–349.
- [22] V. Berthé, S. Ferenczi, L.Q. Zamboni, Interactions between dynamics, arithmetics and combinatorics: the good, the bad, and the ugly, in: *Algebraic and topological dynamics, Contemp. Math.*, vol. 385, Amer. Math. Soc., Providence, RI, 2005, pp. 333–364.
- [23] V. Berthé, C. Holton, L.Q. Zamboni, Initial powers of Sturmian sequences, *Acta Arith.* 122 (2006) 315–347.
- [24] J.-P. Borel, F. Laubie, Quelques mots sur la droite projective réelle, *J. Théor. Nombres Bordeaux* 5 (1993) 23–51.
- [25] J.-P. Borel, C. Reutenauer, Palindromic factors of billiard words, *Theoret. Comput. Sci.* 340 (2005) 334–348.
- [26] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, *Internat. J. Found. Comput. Sci.* 15 (2004) 293–306.
- [27] M. Bucci, A. de Luca, A. De Luca, L.Q. Zamboni, On some problems related to palindrome closure, *Theor. Inform. Appl.* (in press), doi:10.1051/ita:2007064.
- [28] M. Bucci, A. de Luca, A. De Luca, L.Q. Zamboni, On different generalizations of episturmian words, *Theoret. Comput. Sci.* 393 (2008) 23–36.
- [29] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A connection between palindromic and factor complexity using return words, *Adv. in Appl. Math.* 42 (2009) 60–74.
- [30] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A new characteristic property of rich words, *Theoret. Comput. Sci.* (in press), doi:10.1016/j.tcs.2008.11.001.
- [31] J. Cassaigne, Sequences with grouped factors, in: *Developments in Language Theory III*, Aristotle University of Thessaloniki, 1998, pp. 211–222.
- [32] J. Cassaigne, N. Chekhova, Fonctions de récurrence des suites d’Arnoux-Rauzy et réponse à une question de Morse et Hedlund, *Ann. Inst. Fourier (Grenoble)* 56 (2006) 2249–2270.
- [33] J. Cassaigne, S. Ferenczi, L.Q. Zamboni, Imbalances in Arnoux-Rauzy sequences, *Ann. Inst. Fourier (Grenoble)* 50 (2000) 1265–1276.

- [34] M.G. Castelli, F. Mignosi, A. Restivo, Fine and Wilf's theorem for three periods and a generalization of Sturmian words, *Theoret. Comput. Sci.* 218 (1999) 83–94.
- [35] N. Chekhova, P. Hubert, A. Messaoudi, Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci, *J. Théor. Nombres Bordeaux* 13 (2001) 371–394.
- [36] E.M. Coven, Sequences with minimal block growth II, *Math. Systems Theory* 8 (1974) 376–382.
- [37] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory* 7 (1973) 138–153.
- [38] D. Crisp, W. Moran, A. Pollington, P. Shiue, Substitution invariant cutting sequences, *J. Théor. Nombres Bordeaux* 5 (1993) 123–137.
- [39] D. Damanik, D. Lenz, The index of Sturmian sequences, *European J. Combin.* 23 (2002) 23–29.
- [40] D. Damanik, L.Q. Zamboni, Combinatorial properties of Arnoux-Rauzy subshifts and applications to Schrödinger operators, *Rev. Math. Phys.* 15 (2003) 745–763.
- [41] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, *Theoret. Comput. Sci.* 183 (1997) 45–82.
- [42] A. de Luca, A. Glen, L.Q. Zamboni, Rich, Sturmian, and trapezoidal words, *Theoret. Comput. Sci.* 407 (2008) 569–573.
- [43] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* 255 (2001) 539–553.
- [44] X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* 223 (1999) 73–85.
- [45] F. Durand, A characterization of substitutive sequences using return words, *Discrete Math.* 179 (1998) 89–101.
- [46] F. Durand, A generalization of Cobham's theorem, *Theory Comput. Syst.* 31 (1998) 169–185.
- [47] F. Durand, Linearly recurrent subshifts have a finite number of non-periodic subshift factors, *Ergodic Theory Dynam. Systems* 19 (1999) 953–993.
- [48] I. Fagnot, A little more about morphic Sturmian words, *Theor. Inform. Appl.* 40 (2006) 511–518.
- [49] I. Fagnot, L. Vuillon, Generalized balances in Sturmian words, *Discrete Appl. Math.* 121 (2002) 83–101.
- [50] S. Ferenczi, Complexity of sequences and dynamical systems, *Discrete Math.* 206 (1999) 145–154.
- [51] S. Ferenczi, C. Holton, L.Q. Zamboni, Structure of three interval exchange transformations I. An arithmetic study, *Ann. Inst. Fourier (Grenoble)* 51 (2001) 861–901.
- [52] S. Ferenczi, C. Mauduit, Transcendence of numbers with a low complexity expansion, *J. Number Theory* 67 (1997) 146–161.
- [53] S. Ferenczi, C. Mauduit, A. Nogueira, Substitutional dynamical systems: algebraic characterization of eigenvalues, *Ann. Sci. École Norm. Sup.* 29 (1995) 519–533.
- [54] S. Fischler, Palindromic prefixes and episturmian words, *J. Combin. Theory Ser. A* 113 (2006) 1281–1304.
- [55] S. Fischler, Palindromic prefixes and diophantine approximation, *Monatsh. Math.* 151 (2007) 11–37.
- [56] A.S. Fraenkel, Complementing and exactly covering sequences. *J. Combinatorial Theory Ser. A* 14 (1973) 8–20.
- [57] A. Glen, On Sturmian and episturmian words, and related topics, Ph.D. Thesis, The University of Adelaide, Australia, April 2006.
- [58] A. Glen. Order and quasiperiodicity in episturmian words. in: *Proceedings of the 6th International Conference on Words*, Marseille, France, September 17-21, 2007, pp. 144–158.

- [59] A. Glen, Powers in a class of  $\mathcal{A}$ -strict standard episturmian words, *Theoret. Comput. Sci.* 380 (2007) 330–354.
- [60] A. Glen, A characterization of fine words over a finite alphabet, *Theoret. Comput. Sci.* 391 (2008) 51–60.
- [61] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, *European J. Combin.* 30 (2009) 510–531.
- [62] A. Glen, J. Justin, G. Pirillo, Characterizations of finite and infinite episturmian words via lexicographic orderings, *European J. Combin.* 29 (2008) 45–58.
- [63] A. Glen, F. Levé, G. Richomme, Directive words of episturmian words: equivalences and normalization, *Theor. Inform. Appl.* (in press), doi:10.1051/ita:2008029.
- [64] A. Glen, F. Levé, G. Richomme, Quasiperiodic and Lyndon episturmian words, *Theoret. Comput. Sci.* 409 (2008) 578–600.
- [65] E. Godelle, Représentation par des transvections des groupes dartin-tits, *Group, Geometry and Dynamics* 1 (2007) 111–133.
- [66] A. Heinis, R. Tijdeman, Characterisation of asymptotically Sturmian sequences, *Publ. Math. Debrecen* 56 (2000) 415–430.
- [67] C. Holton, L.Q. Zamboni, Descendants of primitive substitutions, *Theory Comput. Syst.* 32 (1999) 133–157.
- [68] P. Hubert, Suites équilibrés, *Theoret. Comput. Sci.* 242 (2000) 91–108.
- [69] O. Jenkinson, L.Q. Zamboni, Characterisations of balanced words via orderings, *Theoret. Comput. Sci.* 310 (2004) 247–271.
- [70] J. Justin, On a paper by Castelli, Mignosi, Restivo, *Theor. Inform. Appl.* 34 (2000) 373–377.
- [71] J. Justin, Episturmian morphisms and a Galois theorem on continued fractions, *Theor. Inform. Appl.* 39 (2005) 207–215.
- [72] J. Justin, G. Pirillo, Fractional powers in Sturmian words, *Theoret. Comput. Sci.* 255 (2001) 363–376.
- [73] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* 276 (2002) 281–313.
- [74] J. Justin, G. Pirillo, On a characteristic property of Arnoux-Rauzy sequences, *Theor. Inform. Appl.* 36 (2002) 385–388.
- [75] J. Justin, G. Pirillo, Episturmian words: shifts, morphisms and numeration systems, *Internat. J. Found. Comput. Sci.* 15 (2004) 329–348.
- [76] J. Justin, L. Vuillon, Return words in Sturmian and episturmian words, *Theor. Inform. Appl.* 34 (2000) 343–356.
- [77] T. Komatsu A.J. van der Poorten, Substitution invariant Beatty sequences, *Jpn. J. Math.* 22 (1996) 349–354.
- [78] D. Krieger, On stabilizers of infinite words, *Theoret. Comput. Sci.* 400 (2008) 169–181.
- [79] F. Levé, G. Richomme, Quasiperiodic infinite words: some answers, *Bull. Eur. Assoc. Theor. Comput. Sci. (EATCS)* 84 (2004) 128–138.
- [80] F. Levé, G. Richomme, Quasiperiodic episturmian words, in: *Proceedings of the 6th International Conference on Words*, Marseille, France, September 17–21, 2007, pp. 201–211.
- [81] F. Levé, G. Richomme, Quasiperiodic Sturmian words and morphisms, *Theoret. Comput. Sci.* 372 (2007) 15–25.

- [82] M. Lothaire, *Combinatorics on Words*, vol. 17 of *Encyclopedia of Mathematics and its Applications*, Addison-Wesley, Reading, Massachusetts, 1983.
- [83] M. Lothaire, *Algebraic Combinatorics on Words*, vol. 90 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, U.K., 2002.
- [84] M. Lothaire, *Applied Combinatorics on Words*, vol. 105 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, U.K., 2005.
- [85] S. Marcus, Quasiperiodic infinite words, *Bull. Eur. Assoc. Theor. Comput. Sci. (EATCS)* 82 (2004) 170–174.
- [86] F. Mignosi, G. Pirillo. Repetitions in the Fibonacci infinite word, *Theor. Inform. Appl.* 26 (1992) 199–204.
- [87] F. Mignosi, P. Séébold, Morphismes Sturmien et règles de Rauzy, *J. Théor. Nombres Bordeaux* 5 (1993) 221–233.
- [88] F. Mignosi, L.Q. Zamboni, On the number of Arnoux-Rauzy words, *Acta Arith.* 101 (2002) 121–129.
- [89] T. Monteil. Illumination dans les billards polygonaux et dynamique symbolique. PhD thesis, Université de la Méditerranée, Faculté des Sciences de Luminy, December 2005.
- [90] T. Monteil, S. Marcus, Quasiperiodic infinite words: multi-scale case and dynamical properties, *Theoret. Comput. Sci.*, to appear, arXiv:math/0603354v1.
- [91] M. Morse, G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories, *Amer. J. Math.* 62 (1940) 1–42.
- [92] G. Paquin, L. Vuillon, A characterization of balanced episturmian sequences, *Electron. J. Combin.* 14 (2007) #R33, pp. 12.
- [93] G. Pirillo, Inequalities characterizing standard Sturmian words, *Pure Math. Appl.* 14 (2003) 141–144.
- [94] G. Pirillo, Inequalities characterizing standard Sturmian and episturmian words, *Theoret. Comput. Sci.* 341 (2005) 276–292.
- [95] G. Pirillo, Morse and Hedlund’s skew Sturmian words revisited, *Ann. Comb.* 12 (2008) 115–121.
- [96] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, vol. 1794 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 2002.
- [97] G. Rauzy, Suites à termes dans un alphabet fini, in: *Sémin. Théorie des Nombres*, Exp. No. 25, pp. 16, Univ. Bordeaux I, Talence, 1982–1983.
- [98] G. Rauzy, Mots infinis en arithmétique, in: M. Nivat, D. Perrin (Eds.), *Automata On Infinite Words, Lecture Notes in Computer Science*, vol. 192, Springer-Verlag, Berlin, 1985, pp. 165–171.
- [99] G. Richomme, Conjugacy and episturmian morphisms, *Theoret. Comput. Sci.* 302 (2003) 1–34.
- [100] G. Richomme, Lyndon morphisms, *Bull. Belg. Math. Soc. Simon Stevin* 10 (2003) 761–785.
- [101] G. Richomme, A local balance property of episturmian words, in: *Proceedings of the 11th International Conference on Developments in Language Theory 2007 (DLT ’07)*, July 3–6, Turku, Finland, vol. 4588 of *Lecture Notes in Computer Science*, Springer, Berlin, 2007, pp. 371–381.
- [102] G. Richomme, Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words, *Theoret. Comput. Sci.* 380 (2007) 393–400.
- [103] G. Richomme, On morphisms preserving infinite Lyndon words, *Discrete Math. Theor. Comput. Sci.* 9 (2007) 89–108.
- [104] G. Richomme, *Private communication*, 2007.
- [105] R.N. Risley, L.Q. Zamboni, A generalization of Sturmian sequences: Combinatorial structure and transcendence, *Acta Arith.* 95 (2000) 167–184.

- [106] A. Siegel, Pure discrete spectrum dynamical systems and periodic tiling associated with a substitution, *Ann. Inst. Fourier (Grenoble)* 54 (2004) 341–381.
- [107] B. Tan, Z.-Y. Wen, Some properties of the Tribonacci sequence, *European J. Combin.* 28 (2007) 1703–1719.
- [108] R. Tijdeman, On complementary triples of Sturmian bisequences, *Indag. Math.* 7 (1996) 419–424.
- [109] R. Tijdeman, Intertwinings of Sturmian sequences, *Indag. Math.* 9 (1998) 113–122.
- [110] R. Tijdeman, Fraenkel’s conjecture for six sequences, *Discrete Math.* 222 (2000) 223–234.
- [111] D. Vandeth, Sturmian words and words with a critical exponent, *Theoret. Comput. Sci.* 242 (2000) 283–300.
- [112] P. Veerman, Symbolic dynamics and rotation numbers, *Physica A* 134 (1986) 543–576.
- [113] P. Veerman, Symbolic dynamics of order-preserving orbits, *Physica D* 29 (1987) 191–201.
- [114] L. Vuillon, A characterization of Sturmian words by return words, *European J. Combin.* 22 (2001) 263–275.
- [115] L. Vuillon, Balanced words, *Bull. Belg. Math. Soc. Simon Stevin* 10 (2003) 787–805.
- [116] Z.-X. Wen, Y. Zhang, Some remarks on invertible substitutions on three letter alphabet, *Chinese Sci. Bull.* 44 (19) (1999) 1755–1760.
- [117] N. Wozny, L.Q. Zamboni, Frequencies of factors in Arnoux-Rauzy sequences, *Acta Arith.* 96 (2001) 261–278.
- [118] S.-I. Yasutomi, On Sturmian sequences which are invariant under some substitutions, in: *Number theory and its applications* (Kyoto, 1997), Kluwer Acad. Publ., Dordrecht, 1999, pp. 347–373.
- [119] L.Q. Zamboni, Une généralisation du théorème de Lagrange sur le développement en fraction continue, *C.R. Acad. Sci. Paris Sér. I Math.* 327 (1998) 527–530.