e-mail: lubomira.balkova@fjfi.cvut.cz, edita.pelantova@fjfi.cvut.cz, staroste@fjfi.cvut.cz

STURMIAN JUNGLE (OR GARDEN?) ON MULTILITERAL ALPHABETS

Ľubomíra Balková, Edita Pelantová and Štěpán Starosta[1]

Abstract. The properties characterizing Sturmian words are considered for words on multiliteral alphabets. We summarize various generalizations of Sturmian words to multiliteral alphabets and enlarge the list of known relationships among these generalizations. We provide a new equivalent definition of rich words and make use of it in the study of generalizations of Sturmian words based on palindromes. We also collect many examples of infinite words to illustrate differences in the generalized definitions of Sturmian words.

1991 Mathematics Subject Classification. 68R15.

1. Introduction

Sturmian words, i.e., aperiodic words with the lowest factor complexity, appeared first in the paper of Hedlund and Morse in 1940. Since then Sturmian words have been in the center of interest of many mathematicians and the number of discoveries of new properties and connections keeps growing. The charm of Sturmian words consists in their natural appearance while studying diverse problems. Many equivalent definitions have been found that way. Sturmian words are binary and every property characterizing Sturmian words asks for a fruitful extension to an analogy on a larger alphabet. Well-known examples of such efforts are Arnoux-Rauzy words, words coding interval exchange transformations, or billiard words. All these words belong to well established classes and their descriptions and properties can be found in many works [5, 6, 11, 27, 35, 40, 46]. An overview of some generalizations of Sturmian words is provided in [12] and [50].

The aim of this paper is to attract attention to other generalizations of Sturmian words. Our motivation stems from recent results on palindromes in infinite words that have ended in the definition of words rich in palindromes, the definition of defect, the description of a relation between factor and palindromic complexity, etc. [3, 7, 15]. Impulses for such an intensive research of palindromes come concededly from the article [22] which characterizes Sturmian words by palindromes, the article [23] which investigates the number of palindromes in prefixes of infinite words and last, but not least, the discovery of the role of palindromes in description of the spectrum of Schrödinger operators with aperiodic potentials [31]. While generalizing Sturmian words we have taken into consideration the characterization of Sturmian words by return words from [49] and a recent definition of Abelian complexity [42, 43], which is closely connected with balance properties.

We consider the following properties (k denotes the cardinality of alphabet $\mathcal{A}$):

(1) Property $\mathcal{C}$:
    the factor complexity of $\mathbf{u}$ satisfies $\mathcal{C}(n) = (k-1)n + 1$ for all $n \in \mathbb{N}$.
(2) Property $\mathcal{LR}$:
    $\mathbf{u}$ contains one left special and one right special factor of every length.

---

(3) Property $\mathcal{BO}$:
   all bispecial factors of **u** are ordinary.
(4) Property $\mathcal{R}$:
   any factor of **u** has exactly k return words.
(5) Property $\mathcal{P}$:
   the palindromic complexity of **u** satisfies $\mathcal{P}(n) + \mathcal{P}(n+1) = k+1$ for all $n \in \mathbb{N}$.
(6) Property $\mathcal{PE}$:
   every palindrome has a unique palindromic extension in **u**.
(7) Balance properties:
   (a) Property $\mathcal{B}_\forall$:
      u is aperiodic and for all $a \in \mathcal{A}$ and for all factors $w, v \in \mathcal{L}(\mathbf{u})$ with $|w| = |v|$ it holds

$$||w|_a - |v|_a| \leq k - 1.$$

   (b) Property $\mathcal{B}_\exists$:
      u is aperiodic and there exists $a \in \mathcal{A}$ such that for all factors $w, v \in \mathcal{L}(\mathbf{u})$ with $|w| = |v|$ it holds
$$||w|_a - |v|_a| \leq k - 1.$$
   (c) Property $\mathcal{AC}$:
      u is aperiodic and the abelian complexity of **u** satisfies $\mathcal{AC}(n) = k$ for all $n \in \mathbb{N}$, $n \geq 1$.

All properties are equivalent on a binary alphabet and they characterize Sturmian words. No two of them are equivalent on the set of infinite words over a multiliteral alphabet. The non-equivalence is shown by counterexamples. However some properties imply others, or it can be shown that a couple of properties are equivalent on a certain class of infinite words. For instance, on the class of uniformly recurrent ternary words Properties $\mathcal{R}$ and $\mathcal{BO}$ are equivalent.

There exist more equivalent definitions of Sturmian words, for instance the definition based on balance properties of subfactors of factors [25], on the index of an infinite word [37], or Richomme's characteristics of Sturmian words [41]. We do not pay attention to these definitions in our survey.

The paper is organized as follows. In section 2 we recall the notions playing an important role in the definitions of Properties 1 through 7. We recall the notion of substitution which is irrelevant for the generalizations of Sturmian words but is used to construct most of examples of infinite words. Section 3 is focused on the study of palindromes in infinite words: we summarize older and new results concerning palindromes, we define palindromic branches. A new result in this section is Theorem 10 providing a new characterization of rich words by means of bilateral orders. Section 4 shortly summarizes essential results on Sturmian words. Section 5 is devoted to an overview of known relations among different generalizations of Sturmian words, mostly from articles [7,9,16,30,42,43]. New results are in Theorems 21 and 25, and Corollaries 23 and 24. The last section is a brief summary of selected relations and examples illustrating the studied Properties.

2. NOTATION AND DEFINITIONS

By $\mathcal{A}$ we denote a finite set of symbols, usually called *letters*; the set $\mathcal{A}$ is therefore called an *alphabet*. A finite string $w = w_0 w_1 \ldots w_{n-1}$ of letters of $\mathcal{A}$ is said to be a *finite word*, its length is denoted by $|w| = n$. Finite words over $\mathcal{A}$ together with the operation of concatenation and the empty word ε as the neutral element form a free monoid $\mathcal{A}^*$. The map

$$w = w_0 w_1 \ldots w_{n-1} \quad \mapsto \quad \tilde{w} = w_{n-1} w_{n-2} \ldots w_0$$

is a bijection on $\mathcal{A}^*$, the word $\tilde{w}$ is called the *reversal* or the *mirror image* of w. A word w which coincides with its mirror image is a *palindrome*.

Under an *infinite word* **u** over the alphabet $\mathcal{A}$ we understand an infinite string $\mathbf{u} = u_0 u_1 u_2 \ldots$ of letters from $\mathcal{A}$ such that every letter of $\mathcal{A}$ occurs in **u**. We call an infinite word **u** *eventually periodic* if there

exist finite words w, v such that $\mathbf{u} = wv^\omega$, where ω means 'repeated infinitely many times'. If $w = \varepsilon$, then $\mathbf{u}$ is said to be *(purely) periodic*. If $\mathbf{u}$ is not eventually periodic, then we call $\mathbf{u}$ *aperiodic*.

A finite word w is a *factor* of a word v (finite or infinite) if there exist words p and s such that $v = pws$. If $p = \varepsilon$, then w is said to be a *prefix* of v, if $s = \varepsilon$, then w is a *suffix* of v. We say that a prefix or a suffix is *proper* if it is not equal to the word itself.

The *language* $\mathcal{L}(\mathbf{u})$ of an infinite word $\mathbf{u}$ is the set of all its factors. The factors of $\mathbf{u}$ of length n form the set denoted by $\mathcal{L}_n(\mathbf{u})$. Using this notation, we may write $\mathcal{L}(\mathbf{u}) = \cup_{n \in \mathbb{N}} \mathcal{L}_n(\mathbf{u})$.

We say that the language $\mathcal{L}(\mathbf{u})$ is *closed under reversal* if $\mathcal{L}(\mathbf{u})$ contains with every factor w also its reversal $\tilde{w}$.

An infinite word $\mathbf{u}$ over $\mathcal{A}$ is called *c-balanced* if for every $a \in \mathcal{A}$ and for every pair of factors w, v of $\mathbf{u}$ of the same length $|w| = |v|$, we have $||w|_a - |v|_a| \le c$, where $|w|_a$ means the number of letters a contained in w. Note that in the case of a binary alphabet, say $\mathcal{A} = \{0, 1\}$, this condition may be rewritten in a simpler way: an infinite word $\mathbf{u}$ is *c-balanced*, if for every pair of factors w, v of $\mathbf{u}$ with $|w| = |v|$, we have $||w|_0 - |v|_0| \le c$. We call 1-balanced words simply *balanced*.

We say that two words $w, v \in \mathcal{A}^*$ are *abelian equivalent* if for each letter $a \in \mathcal{A}$, it holds $|w|_a = |v|_a$. It is easy to see that the abelian equivalence defines indeed an equivalence relation on $\mathcal{A}^*$. If $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$, then the *Parikh vector* associated with the word $w \in \mathcal{A}^*$ is defined as

$$\Psi(w) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_k}).$$

We call *abelian complexity* (as defined in [42]) of an infinite word $\mathbf{u}$ the function $\mathcal{AC} : \mathbb{N} \to \mathbb{N}$ given by

$$\mathcal{AC}(n) = \#\{\Psi(w) \mid w \in \mathcal{L}_n(\mathbf{u})\}.$$

For any factor $w \in \mathcal{L}(\mathbf{u})$, there exists an index i such that w is a prefix of the infinite word $u_i u_{i+1} u_{i+2} \ldots$. Such an index i is called an *occurrence* of w in $\mathbf{u}$. If each factor of $\mathbf{u}$ has at least two occurrences in $\mathbf{u}$, the infinite word $\mathbf{u}$ is said to be *recurrent*. It can be easily shown that each factor of a recurrent word occurs infinitely many times. It is readily seen that if the language of $\mathbf{u}$ is closed under reversal, then $\mathbf{u}$ is recurrent. The infinite word $\mathbf{u}$ is said to be *uniformly recurrent* if for any factor w of $\mathbf{u}$ the distances between successive occurrences of w form a bounded sequence.

Let $j, k, j < k$, be two successive occurrences of a factor w in $\mathbf{u}$. Then $u_j u_{j+1} \ldots u_{k-1}$ is called a *return word* of w. Return words were first studied in [24] and [32]. The set of all return words of w is denoted by $R(w)$,

$$R(w) = \{u_j u_{j+1} \ldots u_{k-1} \mid j, k \text{ being successive occurrences of } w \text{ in } \mathbf{u}\}.$$

If v is a return word of w, then the word vw is called a *complete return word* of w. It is obvious that an infinite recurrent word is uniformly recurrent if and only if the set of return words of any of its factors is finite.

The *(factor) complexity* of an infinite word $\mathbf{u}$ is the map $\mathcal{C} : \mathbb{N} \mapsto \mathbb{N}$, defined by $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$. To determine the increment of complexity, one has to count the possible extensions of factors of length n. A *left extension* of $w \in \mathcal{L}(\mathbf{u})$ is any letter $a \in \mathcal{A}$ such that $aw \in \mathcal{L}(\mathbf{u})$. The set of all left extensions of a factor w will be denoted by $\text{Lext}(w)$. We will mostly deal with recurrent infinite words $\mathbf{u}$. In this case, any factor of $\mathbf{u}$ has at least one left extension. A factor w is called *left special* (or LS for short) if w has at least two left extensions. Clearly, any prefix of a LS factor is LS as well. It makes therefore sense to define an *infinite LS branch* which is an infinite word whose all prefixes are LS factors of $\mathbf{u}$. Similarly, one can define a *right extension*, a *right special* (or RS) factor, $\text{Rext}(w)$, and an *infinite RS branch* which is a left-sided infinite word whose all suffixes are RS factors of $\mathbf{u}$.

We say that a factor w of $\mathbf{u}$ is a *bispecial* (or BS) factor if it is both RS and LS. The role of BS factors for the computation of complexity can be nicely illustrated on Rauzy graphs (introduced in [6]).

Let $\mathbf{u}$ be an infinite word and $n \in \mathbb{N}$. The *Rauzy graph* Γ_n of $\mathbf{u}$ is a directed graph whose set of vertices is $\mathcal{L}_n(\mathbf{u})$ and set of edges is $\mathcal{L}_{n+1}(\mathbf{u})$. An edge $e \in \mathcal{L}_{n+1}(\mathbf{u})$ starts in the vertex w and ends in the vertex v if w is a prefix and v is a suffix of e, see Figure 1. If the word $\mathbf{u}$ is recurrent, the graph Γ_n is strongly connected for every $n \in \mathbb{N}$, i.e., there exists a directed path from every vertex w to every vertex v of the graph.

$$\bullet \xrightarrow{\quad e = w_0 w_1 \cdots w_{n-1} w_n \quad} \bullet$$
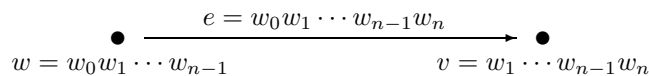$$w = w_0 w_1 \cdots w_{n-1} \qquad\qquad\qquad v = w_1 \cdots w_{n-1} w_n$$

FIGURE 1. Incidence relation between an edge and vertices in a Rauzy graph.

If the language $\mathcal{L}(\mathbf{u})$ of the infinite word $\mathbf{u}$ is closed under reversal, then the operation that to every vertex w of the graph associates its mirror image, the vertex $\tilde{w}$, and to every edge e associates $\tilde{e}$ maps the Rauzy graph Γ_n onto itself.

The *outdegree* (*indegree*) of a vertex $w \in \mathcal{L}_n(\mathbf{u})$ is the number of edges which start (end) in w. Obviously the outdegree of w is equal to $\#\mathrm{Rext}(w)$ and the indegree of w is $\#\mathrm{Lext}(w)$. The sum of outdegrees over all vertices is equal to the number of edges in every directed graph. Similarly, it holds for indegrees. In particular, for the Rauzy graph Γ_n we have

$$\sum_{w \in \mathcal{L}_n(\mathbf{u})} \#\mathrm{Rext}(w) \;=\; \mathcal{C}(n+1) \;=\; \sum_{w \in \mathcal{L}_n(\mathbf{u})} \#\mathrm{Lext}(w)\,.$$

The first difference of complexity $\Delta\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n)$ is thus given by

$$\Delta\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} \big(\#\mathrm{Rext}(w) - 1\big) \;=\; \sum_{w \in \mathcal{L}_n(\mathbf{u})} \big(\#\mathrm{Lext}(w) - 1\big)\,.$$

A non-zero contribution to $\Delta\mathcal{C}(n)$ in the left-hand sum is given only by those factors $w \in \mathcal{L}_n(\mathbf{u})$ for which $\#\mathrm{Rext}(w) \geq 2$, and for recurrent words, a non-zero contribution to $\Delta\mathcal{C}(n)$ in the right-hand sum is provided only by those factors $w \in \mathcal{L}_n(\mathbf{u})$ for which $\#\mathrm{Lext}(w) \geq 2$. The last relation can be thus rewritten for recurrent words $\mathbf{u}$ as

$$\Delta\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u}),\ w\ \mathrm{RS}} \big(\#\mathrm{Rext}(w) - 1\big) \;=\; \sum_{w \in \mathcal{L}_n(\mathbf{u}),\ w\ \mathrm{LS}} \big(\#\mathrm{Lext}(w) - 1\big)\,.$$

If we denote $\mathrm{Bext(w)} = \{awb \in \mathcal{L}(\mathbf{u}) \mid a, b \in \mathcal{A}\}$, then the second difference of complexity $\Delta^2\mathcal{C}(n) = \Delta\mathcal{C}(n+1) - \Delta\mathcal{C}(n) = \mathcal{C}(n+2) - 2\mathcal{C}(n+1) + \mathcal{C}(n)$ is given by

$$\Delta^2\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} \big(\#\mathrm{Bext}(w) - \#\mathrm{Rext}(w) - \#\mathrm{Lext}(w) + 1\big)\,. \tag{2.1}$$

Denote by $\mathrm{b}(w)$ the quantity

$$\mathrm{b}(w) := \#\mathrm{Bext}(w) - \#\mathrm{Rext}(w) - \#\mathrm{Lext}(w) + 1\,.$$

The number $\mathrm{b}(w)$ is called the *bilateral order* of the factor w and was introduced in [18]. It is readily seen that if w is not a BS factor, then $\mathrm{b}(w) = 0$. Bispecial factors are distinguished according to their bilateral order in the following way

- if $\mathrm{b}(w) > 0$, then w is a *strong* BS factor,
- if $\mathrm{b}(w) < 0$, then w is a *weak* BS factor,
- if $\mathrm{b}(w) = 0$ then w is an *ordinary* BS factor.

A *substitution* on $\mathcal{A}$ is a morphism $\varphi : \mathcal{A}^* \to \mathcal{A}^*$ such that there exists a letter $a \in \mathcal{A}$ and a non-empty word $w \in \mathcal{A}^*$ satisfying $\varphi(a) = aw$ and $\varphi(b) \neq \varepsilon$ for all $b \in \mathcal{A}$. Since a morphism satisfies $\varphi(vw) = \varphi(v)\varphi(w)$ for all $v, w \in \mathcal{A}^*$, any substitution is uniquely determined by the images of letters. Instead of classical $\varphi(a) = w$, we sometimes write $a \to w$. A substitution can be naturally extended to an infinite word $\mathbf{u} = u_0 u_1 u_2 \ldots$ by the prescription $\varphi(\mathbf{u}) = \varphi(u_0)\varphi(u_1)\varphi(u_2)\ldots$ An infinite word $\mathbf{u}$ is said to be a *fixed point* of the substitution φ if it fulfills $\mathbf{u} = \varphi(\mathbf{u})$. It is obvious that every substitution φ has at least one fixed point, namely $\lim_{n \to \infty} \varphi^n(a)$ (to be understood in the sense of product topology).

3. WORDS OPULENT IN PALINDROMES

In resemblance to the factor complexity $\mathcal{C}(n)$ of an infinite word $\mathbf{u}$, let us define the *palindromic complexity* of $\mathbf{u}$ as the map $\mathcal{P} : \mathbb{N} \to \mathbb{N}$ given by

$$\mathcal{P}(n) = \#\{w \in \mathcal{L}_n(\mathbf{u})|\ w = \tilde{w}\}.$$

If $a \in \mathcal{A}$ and w is a palindrome and $awa \in \mathcal{L}(\mathbf{u})$, then awa is said to be a *palindromic extension* of w. The set of all palindromic extensions of w is denoted by $\mathrm{Pext}(w)$.

Similarly as in the case of left special and right special branches, one can define a *palindromic branch* of $\mathbf{u}$.

Definition 1. *Let $\mathbf{u}$ be an infinite word. A two-sided infinite word $v = \ldots v_3 v_2 v_1 v_1 v_2 v_3 \ldots$ is a palindromic branch with center ε of the word $\mathbf{u}$ if for every $n \in \mathbb{N}$ the word $v_n v_{n-1} \ldots v_2 v_1 v_1 v_2 \ldots v_{n-1} v_n$ is a factor of $\mathbf{u}$. Let a be a letter. A two-sided infinite word $v = \ldots v_3 v_2 v_1 a v_1 v_2 v_3 \ldots$ is a palindromic branch with center a of the word $\mathbf{u}$ if for every $n \in \mathbb{N}$ the word $v_n v_{n-1} \ldots v_2 v_1 a v_1 v_2 \ldots v_{n-1} v_n$ is a factor of $\mathbf{u}$.*

It follows from the König's theorem that if $\mathbf{u}$ has infinitely many palindromes, then $\mathbf{u}$ has at least one palindromic branch. In any Sturmian word on $\{0, 1\}$ there exist exactly three palindromic branches with centers ε, 0 and 1. See also Section 5.1.

Uniformly recurrent words containing infinitely many distinct palindromes satisfy that for any factor w, every sufficiently large palindrome in $\mathbf{u}$ contains w, thus such a palindrome contains $\tilde{w}$ as well. As a consequence, we have the following theorem.

Theorem 2. *If $\mathbf{u}$ is a uniformly recurrent word that contains infinitely many distinct palindromes, then its language $\mathcal{L}(\mathbf{u})$ is closed under reversal.*

The opposite implication is not true as illustrated by the following example.

Example 1 (uniform recurrence + language closed under reversal $\not\Rightarrow$ infinitely many palindromes)**.** The infinite word $\mathbf{u}$ on $\{a, b\}$ (constructed in [13]) whose prefixes u_n are given by the following recurrent formula

$$u_0 = ab, \quad u_{n+1} = u_n a b \widetilde{u_n},$$

is uniformly recurrent and its language is closed under reversal. However, $\mathbf{u}$ contains only a finite number of palindromes.

When we relax the condition of uniform recurrence, the statement of Theorem 2 is not true any more.

Example 2 (infinitely many palindromes $\not\Rightarrow$ language closed under reversal)**.** The infinite word $\mathbf{u}$ on $\{a, b, c\}$ whose prefixes u_n are given by the following recurrent formula

$$u_0 = \varepsilon, \quad u_{n+1} = u_n a b c^{n+1} u_n$$

is clearly recurrent. Infinitely many palindromes are represented by the factors c^n for every n. As the factor ba does not occur, the set of factors is not closed under reversal. A similar example can be found in [16].

The word $\mathbf{u}$ may be recoded to a binary alphabet while preserving the mentioned properties. We may for instance recode $\mathbf{u}$ using the following mapping:

$$a \to 0110,\ b \to 1001,\ c \to 1.$$

An interesting relation between the palindromic and factor complexity has been revealed in [7].

Theorem 3. *Let $\mathbf{u}$ be an infinite word with the language closed under reversal. Then*

$$\mathcal{P}(n+1) + \mathcal{P}(n) \leq \Delta \mathcal{C}(n) + 2 \quad \text{for all } n \in \mathbb{N}. \tag{3.1}$$

In fact, the above relation is stated in [7] for uniformly recurrent words, however the proof requires only recurrent words. Theorem 3 implies that infinite words reaching the equality in (3.1) are in a certain sense opulent in palindromes. Another measure of opulence in palindromes has been provided in [23].

Theorem 4. *Every finite word w contains at most $|w| + 1$ palindromes (including the empty word).*

Definition 5. *An infinite word $\mathbf{u}$ satisfying that every factor w of $\mathbf{u}$ contains $|w| + 1$ palindromes is called rich in palindromes.*

The following equivalent definitions of richness have been proved in [30], [16], [17], respectively.

Theorem 6. *For any infinite word $\mathbf{u}$ the following conditions are equivalent:*
  (1) $\mathbf{u}$ is rich,
  (2) any complete return word of a palindromic factor of $\mathbf{u}$ is a palindrome,
  (3) for any factor w of $\mathbf{u}$, every factor of $\mathbf{u}$ that contains w only as its prefix and $\tilde{w}$ only as its suffix is a palindrome,
  (4) each factor of $\mathbf{u}$ is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.

We will need for our further purposes an implication that holds only for languages closed under reversal.

Corollary 7 ([16]). *Let $\mathbf{u}$ be a rich infinite word with the language closed under reversal. Then for any factor w of $\mathbf{u}$, the occurrences of w and $\tilde{w}$ alternate.*

A natural question is whether infinite words reaching the equality in (3.1) coincide with rich words. The following theorem proved in [16] provides an answer.

Theorem 8. *Let $\mathbf{u}$ be an infinite word with the language closed under reversal. Then $\mathbf{u}$ is rich if and only if $\mathcal{P}(n + 1) + \mathcal{P}(n) = \Delta \mathcal{C}(n) + 2$ for all $n \in \mathbb{N}$.*

Let us mention as an open problem the following question. "Does the equivalence of richness and the equality in (3.1) hold for a larger class than words with the language closed under reversal? For instance for all recurrent words?"

The following observations may serve as hints:
- It does not hold for non-recurrent infinite words in general. The infinite word ab^ω is given in [16] as an example of a rich non-recurrent infinite word (with the language of course not closed under reversal), which does not reach the equality in (3.1) for all $n \in \mathbb{N}$.
- Notice that both rich infinite words and infinite words reaching the equality in (3.1) contain infinitely many palindromes.
- If $\mathbf{u}$ is rich and recurrent, then $\mathcal{L}(\mathbf{u})$ is closed under reversal (proved in [30], Proposition 2.11).

The rest of this section is devoted to the relation between richness and bilateral orders of factors. The following proposition reveals some information on bilateral orders of palindromic bispecial factors in an infinite word with the language closed under reversal.

Proposition 9. *Let $\mathbf{u}$ be an infinite word whose language is closed under reversal. Then the bilateral order $\mathrm{b}(w)$ of a palindromic bispecial factor $w \in \mathcal{L}(\mathbf{u})$ has a different parity than the number of palindromic extensions of w.*

Proof. Let w be a palindromic BS factor of $\mathbf{u}$. On one hand, as the language is closed under reversal, we have $\#\mathrm{Lext}(w) = \#\mathrm{Rext}(w)$. Consequently, from the definition of bilateral order one can see that the parity of $\#\mathrm{Bext}(w)$ is different from the parity of $\mathrm{b}(w)$. On the other hand, the parity of the number of palindromic extensions of w equals the parity of $\#\mathrm{Bext}(w)$ since for any $a, b \in \mathcal{A}$, if $awb \in \mathcal{L}(\mathbf{u})$, then $bwa \in \mathcal{L}(\mathbf{u})$. □

In the sequel, we will state and prove a new equivalent definition of rich words by means of bilateral orders.

Theorem 10. *Let $\mathbf{u}$ be an infinite word with the language closed under reversal. Then $\mathbf{u}$ is rich if and only if any bispecial factor w of $\mathbf{u}$ satisfies:*

- *if w is non-palindromic, then*

$$\mathrm{b}(w) = 0,$$

- *if w is a palindrome, then*

$$\mathrm{b}(w) = \#\mathrm{Pext}(w) - 1.$$

The following lemma will provide the most important tool for the proof of Theorem 10.

Lemma 11. *Let* **u** *be a rich infinite word whose language is closed under reversal. Then it holds for any bispecial factor w:*

- *if w is non-palindromic, then*

$$\mathrm{b}(w) \geq 0,$$

- *if w is a palindrome, then*

$$\mathrm{b}(w) \geq \#\mathrm{Pext}(w) - 1.$$

Proof. Let w be a non-palindromic BS factor. By the definition of $\mathrm{b}(w)$, we want to prove

$$\#\mathrm{Bext}(w) \geq \#\mathrm{Rext}(w) + \#\mathrm{Lext}(w) - 1.$$

We will construct a bipartite oriented graph G having its set of vertices V defined as

$$V = \{wa | a \in \mathrm{Rext}(w)\} \cup \{\tilde{w}a | a \in \mathrm{Rext}(\tilde{w})\}.$$

There is an oriented edge from wa to $\tilde{w}b$ if there exists a factor $vb \in \mathcal{L}(\mathbf{u})$ such that wa is its prefix, $\tilde{w}b$ is its suffix and factors w and $\tilde{w}$ occur each exactly once in vb. Furthermore, there is an oriented edge from $\tilde{w}x$ to wy if there exists a factor $vy \in \mathcal{L}(\mathbf{u})$ such that $\tilde{w}x$ is its prefix, wy is its suffix and factors w and $\tilde{w}$ occur each exactly once in v.
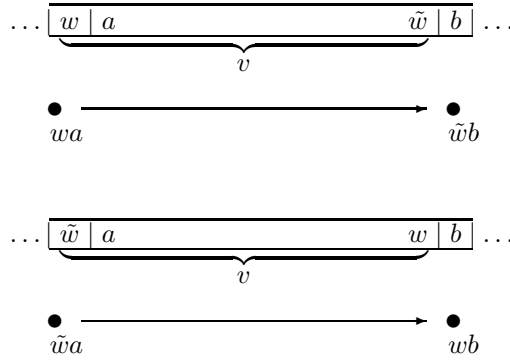

FIGURE 2. Incidence relation in the graph G.

Due to Theorem 6, such a factor v is a palindrome. Therefore the existence of an edge from wa to $\tilde{w}b$ implies $a\tilde{w}b \in \mathcal{L}(\mathbf{u})$, and so $bwa \in \mathcal{L}(\mathbf{u})$, too. Analogously, if there is an edge from $\tilde{w}x$ to wy, we have $xwy \in \mathcal{L}(\mathbf{u})$.

By Corollary 7, the occurrences of w and $\tilde{w}$ alternate. Thus, to any factor of **u** corresponds a path in G. As **u** is recurrent, the graph G is strongly connected.

As a consequence, the number of pairs of its vertices which are connected by an edge is greater than or equal to the number of its vertices minus 1. We have

$$\#\mathrm{Bext}(w) \geq \#\mathrm{Rext}(w) + \#\mathrm{Rext}(\tilde{w}) - 1.$$

Since $\mathrm{Rext}(\tilde{w}) = \mathrm{Lext}(w)$ the proof of the first part is finished. $\quad\square$

Let w be a palindromic BS factor. Let us consider this time a graph G whose set of factors V is defined as

$$V = \{wa | a \in \mathrm{Rext}(w)\} .$$

There is an edge from wa to wb if there exists a factor $vb \in \mathcal{L}(\mathbf{u})$ such that v is a complete return word to w that has wa as a prefix. As $\mathbf{u}$ is rich, v is a palindrome. Due to the recurrence of $\mathbf{u}$, for every $awb \in \mathcal{L}(\mathbf{u})$, $a \neq b$, there exists an edge in G going from wa to wb. As the language is closed under reversal, the edge going from wb to wa is in G, too. Therefore

$$\# \{awb \in \mathcal{L}(\mathbf{u}) | a \neq b\} = 2 \times \text{ the number of pairs of distinct vertices connected by an edge.}$$

Owing to the recurrence of $\mathbf{u}$, the graph G is strongly connected, thus the number of pairs of distinct vertices connected by an edge is greater than or equal to the number of vertices of G minus 1, which equals $\#\mathrm{Rext}(w) - 1$. We find

$$\#\mathrm{Bext}(w) = \# \{awb \in \mathcal{L}(\mathbf{u}) | a \neq b\} + \#\mathrm{Pext}(w) \geq 2 \left(\#\mathrm{Rext}(w) - 1\right) + \#\mathrm{Pext}(w).$$

As $\mathrm{Rext}(w) = \mathrm{Lext}(w)$, the statement is proved. $\quad\square$

Proof of Theorem 10. ($\Leftarrow$): Let us show by mathematical induction that

$$\Delta\mathcal{C}(n) + 2 = \mathcal{P}(n+1) + \mathcal{P}(n) \quad \text{for all } n \in \mathbb{N}.$$

Since $\mathcal{L}(\mathbf{u})$ is closed under reversal, this means by Theorem 8 that $\mathbf{u}$ is rich.

The assumption on bilateral orders and the fact that non-bispecial palindromic factors have a unique palindromic extension guarantee the following equality for all $n \in \mathbb{N}$:

$$\Delta^2\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} \mathrm{b}(w) = \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \tilde{w}}} (\#\mathrm{Pext}(w) - 1) = \mathcal{P}(n+2) - \mathcal{P}(n). \tag{3.2}$$

For $n = 0$, we can write $\Delta\mathcal{C}(0) + 2 = \mathcal{C}(1) - \mathcal{C}(0) + 2 = \#\mathcal{A} + 1$. On the other hand we have $\mathcal{P}(1) + \mathcal{P}(0) = \#\mathcal{A} + 1$.

Take $N \in \mathbb{N}$. Assume $\Delta\mathcal{C}(n) + 2 = \mathcal{P}(n+1) + \mathcal{P}(n)$ holds for all $n < N$. Using the induction assumption and (3.2), we obtain

$$\begin{aligned}
\Delta\mathcal{C}(N) + 2 &= (\Delta\mathcal{C}(N) - \Delta\mathcal{C}(N-1)) + (\Delta\mathcal{C}(N-1) + 2) \\
&= \Delta^2\mathcal{C}(N-1) + (\mathcal{P}(N-1) + \mathcal{P}(N)) \\
&= (\mathcal{P}(N+1) - \mathcal{P}(N-1)) + (\mathcal{P}(N-1) + \mathcal{P}(N)) \\
&= \mathcal{P}(N+1) + \mathcal{P}(N).
\end{aligned}$$

($\Rightarrow$): Take $n \in \mathbb{N}$ arbitrary. We will prove the statement of the theorem for all BS factors of length n.

As $\mathbf{u}$ is rich and the language $\mathcal{L}(\mathbf{u})$ is closed under reversal, we have by Theorem 8

$$\Delta\mathcal{C}(k) + 2 = \mathcal{P}(k+1) + \mathcal{P}(k) \quad \text{for all } k \in \mathbb{N}.$$

Applying this equality, we will deduce the form of $\Delta^2\mathcal{C}(n)$.

$$\Delta^2\mathcal{C}(n) = (\Delta\mathcal{C}(n+1) + 2) - (\Delta\mathcal{C}(n) + 2) = (\mathcal{P}(n+2) + \mathcal{P}(n+1)) - (\mathcal{P}(n+1) + \mathcal{P}(n)) = \mathcal{P}(n+2) - \mathcal{P}(n).$$

Consequently, we obtain

$$\sum_{w \in \mathcal{L}_n(\mathbf{u})} \mathrm{b}(w) = \Delta^2\mathcal{C}(n) = \mathcal{P}(n+2) - \mathcal{P}(n) = \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \tilde{w}}} (\#\mathrm{Pext}(w) - 1) .$$

Palindromic factors that are not BS have obviously exactly one palindromic extension. Thus, we can rewrite the previous equality

$$\sum_{w\in\mathcal{L}_n(\mathbf{u})} \mathrm{b}(w) = \sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w=\tilde{w},\,w\ \mathrm{BS}}} \left(\#\mathrm{Pext}(w)-1\right). \qquad (3.3)$$

Let us split the sum of bilateral orders into two parts and use Lemma 11

$$\sum_{w\in\mathcal{L}_n(\mathbf{u})} \mathrm{b}(w) = \sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w\neq\tilde{w},\,w\ \mathrm{BS}}} \mathrm{b}(w) + \sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w=\tilde{w},\,w\ \mathrm{BS}}} \mathrm{b}(w) \geq \sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w\neq\tilde{w},\,w\ \mathrm{BS}}} \mathrm{b}(w) + \sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w=\tilde{w},\,w\ \mathrm{BS}}} \left(\#\mathrm{Pext}(w)-1\right). \quad (3.4)$$

This in combination with (3.3) gives $\sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w\neq\tilde{w},\,w\ \mathrm{BS}}} \mathrm{b}(w) = 0.$ By Lemma 11, bilateral orders of such factors are non-negative, which implies $\mathrm{b}(w)=0$ for all non-palindromic BS factors. Since the equality is reached in (3.4), we obtain $\sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w=\tilde{w},\,w\ \mathrm{BS}}} \mathrm{b}(w) = \sum_{\substack{w\in\mathcal{L}_n(\mathbf{u})\\ w=\tilde{w},\,w\ \mathrm{BS}}} \left(\#\mathrm{Pext}(w)-1\right).$ Together with Lemma 11, this results in $\mathrm{b}(w)=\#\mathrm{Pext}(w)-1$ for all palindromic BS factors. $\qquad\square$

4. Equivalent definitions of Sturmian words

Let us stress a close link between periodicity and complexity (revealed by Hedlund and Morse [38]). On one hand, the complexity of eventually periodic words is bounded. On the other hand, if there exists $n \in \mathbb{N}$ such that $\mathcal{C}(n) \leq n$, then the complexity is bounded and the infinite word $\mathbf{u}$ is eventually periodic. In consequence, the complexity of aperiodic words satisfies $\mathcal{C}(n) \geq n+1$ for all $n \in \mathbb{N}$. *Sturmian words* are defined as infinite words with the complexity $\mathcal{C}(n) = n+1$ for all $n \in \mathbb{N}$. This condition on complexity implies many properties. Let us list some of them. If $\mathbf{u}$ is a Sturmian word, then $\mathbf{u}$ has the following properties:

- $\mathbf{u}$ is a binary word,
- $\mathbf{u}$ is aperiodic,
- the language $\mathcal{L}(\mathbf{u})$ is closed under reversal,
- the language $\mathcal{L}(\mathbf{u})$ contains infinitely many palindromes,
- the word $\mathbf{u}$ is uniformly recurrent,
- the language $\mathcal{L}(\mathbf{u})$ contains no weak bispecial factors,
- $\mathbf{u}$ is rich.

There exist many equivalent definitions of Sturmian words. The following theorem summarizes several of their well-known combinatorial characterizations.

Theorem 12. *Let $\mathbf{u}$ be an infinite word over the alphabet $\mathcal{A}$. The properties listed below are equivalent:*
  (i) *$\mathbf{u}$ is Sturmian, i.e., $\mathcal{C}(n) = n+1$ for all n,*
 (ii) *$\mathbf{u}$ is binary and contains a unique left special factor of every length,*
(iii) *$\mathbf{u}$ is binary, aperiodic and every bispecial factor is ordinary,*
 (iv) *any factor of $\mathbf{u}$ has exactly two return words,*
  (v) *$\mathbf{u}$ contains one palindrome of every even length and two palindromes of every odd length,*
 (vi) *$\mathbf{u}$ is binary and every palindrome has a unique palindromic extension,*
(vii) *$\mathbf{u}$ is aperiodic and balanced,*
(viii) *$\mathbf{u}$ is aperiodic and $\mathcal{AC}(n) = 2$ for all $n \in \mathbb{N}$, $n \geq 1$.*

The characterization by return words is due to Vuillon [49] and the one by the abelian complexity is a consequence of the works by Coven and Hedlund [20]. The equivalent definition based on the

balance property comes already from Hedlund and Morse [39]. The two equivalent properties concerning palindromes have been proved by Droubay and Pirillo [22]. Notice that the sixth property can be equivalently rewritten as

$$\mathcal{P}(n) + \mathcal{P}(n+1) = 3 \quad \text{for all } n \in \mathbb{N},$$

and also as

$$\mathcal{P}(n+2) = \mathcal{P}(n) \quad \text{for all } n \in \mathbb{N}.$$

Let us recall that $\mathcal{P}(0) = 1$ since the empty word is considered to be a palindrome.

5. Generalizations of Sturmian words

We have seen that Sturmian words can be defined in many equivalent ways. As a matter of course, various generalizations to multiliteral alphabets have been suggested and studied.

5.1. Two well-known generalizations

The most studied generalizations are Arnoux-Rauzy words and words coding k-interval exchange transformation.

Arnoux-Rauzy words (or *AR words* for simplicity) are infinite words with the language closed under reversal and containing exactly one LS factor w of every length, and such that every LS factor has the same number k of left extensions, i.e., #Lext(w) $= k$. Their alphabet $\mathcal{A}$ has k letters since the empty word has exactly k left extensions. AR words are aperiodic and satisfy $\mathcal{C}(n) = (k-1)n + 1$ for all $n \in \mathbb{N}$. They have been defined and studied in [23], the following properties have been proved ibidem. The language of AR words contains infinitely many palindromes, they are uniformly recurrent, rich, and have only ordinary BS factors. AR words form a subclass of extensively studied *episturmian words* (see for instance [29]), defined as infinite words that have the language closed under reversal and contain at most one LS factor of every length.

Another well-known generalization of Sturmian words is provided by *words coding k-interval exchange transformation*. Let us state their definition and then explain why such words generalize Sturmian words to k-letter alphabets. Take positive numbers $\alpha_1, \dots, \alpha_k$ such that $\sum_{i=1}^{k} \alpha_i = 1$. They define a partition of the interval $I = [0,1)$ into k subintervals

$$I_j = \Big[\sum_{i=1}^{j-1} \alpha_i, \sum_{i=1}^{j} \alpha_i\Big), \ j \in \{1, 2, \dots, k\}.$$

The *interval exchange transformation* is a bijection $T : I \to I$ given by the prescription

$$T(x) = x + c_j \quad \text{for all } x \in I_j, \ j \in \{1, 2, \dots, k\},$$

where c_j are suitably chosen constants. Since T is a bijection, the intervals $T(I_1), T(I_2), \dots, T(I_k)$ form a partition of I. The orders of $T(I_j)$ in the partition define a permutation $\pi : \{1, 2, \dots, k\} \to \{1, 2, \dots, k\}$ and this permutation π determines uniquely the constants c_j. For instance, if the permutation π is symmetric, i.e., $\pi = \big(\begin{smallmatrix} 1 & 2 & \dots & k-1 & k \\ k & k-1 & \dots & 2 & 1 \end{smallmatrix}\big)$, then the transformation T is of the following form

$$T(x) = x + \sum_{i>j} \alpha_i - \sum_{i<j} \alpha_i \quad \text{for} \quad x \in I_j.$$

The infinite word $\mathbf{u} = u_0 u_1 u_2 \dots$ over $\mathcal{A} = \{a_1, \dots, a_k\}$ associated with T is defined as

$$u_n := a_j \quad \text{if} \quad T^n(x) \in I_j$$

and is called a *word coding k-interval exchange transformation* (*k-iet word* for short).

From the point of view of combinatorics on words, an important role is played by those transformations whose orbit for an arbitrary $x \in I$ is dense in I, i.e., the closure of $\{T^n(x) \mid n \in \mathbb{N}\}$ is the whole interval I. A sufficient condition for this property represents the so-called i.d.o.c. (consult [35]) and the irreducibility of the permutation π. In the sequel, let us assume that T satisfies both of these properties. The k-iet word is then uniformly recurrent, its language does not depend on the position of the starting point x, but only on the transformation T, its complexity satisfies $\mathcal{C}(n) = (k-1)n + 1$ for all $n \in \mathbb{N}$ and no BS factor is weak.

The language of the k-iet word **u** is closed under reversal if and only if the permutation π is symmetric. In such a case, the language $\mathcal{L}(\mathbf{u})$ contains infinitely many palindromes and, as shown in [7], the equality in (3.1) is attained. Hence, according to Theorem 8, the k-iet words are rich. It is easy to describe the infinite palindromic branches for such k-iet words. The one with the empty word as its center is obtained as the coding of the orbit $\{T^n(x)|n \in \mathbb{Z}\}$ with the starting point $x = 1/2$ and the branch with the center $a_j \in \mathcal{A}$ as the coding of the orbit with the starting point $x = \sum_{i<j} \alpha_i + \alpha_j/2$.

The k-iet words provide a generalization of Sturmian words due to the well-known connection between Sturmian and mechanical words [36].

Theorem 13. *Let* **u** *be an infinite word. Then* **u** *is Sturmian if and only if* **u** *is a 2-iet word with an irrational partition of the unit interval.*

Recently, in [45], a different generalization of Sturmian sequences is considered. It in fact corresponds to a special subclass of k-iet words given by coding a trajectory in a regular $2n$-gon.

5.2. Combinatorial generalizations

Let us write down and baptize the generalizations of properties from Theorem 12. We will then refer to them and study their relations. Let **u** be an infinite word over the alphabet $\mathcal{A}$. Denote $k = \#\mathcal{A}$.

(1) Property $\mathcal{C}$:
    the factor complexity of **u** satisfies $\mathcal{C}(n) = (k-1)n + 1$ for all $n \in \mathbb{N}$.
(2) Property $\mathcal{LR}$:
    u contains one left special and one right special factor of every length.
(3) Property $\mathcal{BO}$:
    all bispecial factors of **u** are ordinary.
(4) Property $\mathcal{R}$:
    any factor of **u** has exactly k return words.
(5) Property $\mathcal{P}$:
    the palindromic complexity of **u** satisfies $\mathcal{P}(n) + \mathcal{P}(n+1) = k+1$ for all $n \in \mathbb{N}$.
(6) Property $\mathcal{PE}$:
    every palindrome has a unique palindromic extension in **u**.
(7) Balance properties:
   (a) Property $\mathcal{B}_\forall$:
       u is aperiodic and for all $a \in \mathcal{A}$ and for all factors $w, v \in \mathcal{L}(\mathbf{u})$ with $|w| = |v|$ it holds

$$||w|_a - |v|_a| \leq k - 1.$$

   (b) Property $\mathcal{B}_\exists$:
       u is aperiodic and there exists $a \in \mathcal{A}$ such that for all factors $w, v \in \mathcal{L}(\mathbf{u})$ with $|w| = |v|$ it holds
$$||w|_a - |v|_a| \leq k - 1.$$
   (c) Property $\mathcal{AC}$:
       u is aperiodic and the abelian complexity of **u** satisfies $\mathcal{AC}(n) = k$ for all $n \in \mathbb{N}$, $n \geq 1$.

At first, let us mention which properties are satisfied by the two generalizations of Sturmian words from Section 5.1. AR words fulfill Properties: $\mathcal{C}, \mathcal{LR}, \mathcal{BO}, \mathcal{R}, \mathcal{P}, \mathcal{PE}$ and k-iet words satisfy Properties:

$\mathcal{C}, \mathcal{BO}, \mathcal{R}$. If moreover the permutation defining the k-iet word is symmetric, then these words have Properties $\mathcal{P}$ and $\mathcal{PE}$. Property $\mathcal{LR}$ does not hold for k-iet words.

It follows directly from the definition that some Properties imply others. For instance, by (2.1) $\mathcal{BO}$ implies $\mathcal{C}$. They are not equivalent as shown by the following example taken from [26].

Example 3 ($\mathcal{C} \not\Rightarrow \mathcal{BO}$)**.** The infinite ternary word $\lim_{n \to \infty} \varphi^n(a)$, where $\varphi(a) = ab$, $\varphi(b) = cab$, $\varphi(c) = ccab$ – a recoding of the Chacon substitution – has the complexity $2n + 1$ for every $n \in \mathbb{N}$, but contains infinitely many strong and weak BS factors.

In the sequel, we will show that no two of these properties are equivalent on a multiliteral alphabet.

Concerning Properties $\mathcal{B}_\forall, \mathcal{B}_\exists$ and $\mathcal{AC}$, we will not treat them but in the last section since they are very restrictive, and consequently, satisfied only by a small class of infinite words.

5.3. Property $\mathcal{LR}$

Property $\mathcal{LR}$ does not characterize AR words since it is satisfied by a larger class of words. Infinite words with the language closed under reversal and satisfying Property $\mathcal{LR}$ coincide with extensively studied aperiodic episturmian words. Nevertheless, Property $\mathcal{LR}$ may be satisfied by words whose language is not closed under reversal, as illustrated in [23] by the following example. It shows also that Property $\mathcal{LR}$ does not guarantee Properties $\mathcal{C}, \mathcal{BO}, \mathcal{R}, \mathcal{P}, \mathcal{PE}$.

Example 4 ($\mathcal{LR} \not\Rightarrow$ language closed under reversal, $\mathcal{C}, \mathcal{BO}, \mathcal{R}, \mathcal{P}, \mathcal{PE}$)**.** If we construct an infinite word $\mathbf{u}$ so that we replace b with bc in the Fibonacci word $abaababaabaabab\ldots$, the fixed point of $\varphi : a \to ab$, $b \to a$, then bc is a factor of $\mathcal{L}(\mathbf{u})$, however cb not. It is easy to see that such a word has still a unique infinite RS and a unique LS branch (the infinite word $\mathbf{u}$ itself). Consequently, Property $\mathcal{LR}$ is preserved. However, both of these infinite special branches have only two extensions, hence Property $\mathcal{C}$ (and $\mathcal{BO}$ as well) fails. The factor c has only two return words $caab$ and cab, hence Property $\mathcal{R}$ does not hold. Moreover, as $\mathbf{u}$ is uniformly recurrent and its language is not closed under reversal, it contains by Theorem 2 only a finite number of palindromes. Therefore, Properties $\mathcal{P}$ and $\mathcal{PE}$ are not satisfied.

On the other hand, observing k-iet words, we learn that none of Properties $\mathcal{C}, \mathcal{BO}, \mathcal{R}, \mathcal{P}, \mathcal{PE}$ imply $\mathcal{LR}$. The problem to describe the class of infinite words with Property $\mathcal{LR}$ whose language is not closed under reversal requires a further study.

5.4. Property $\mathcal{R}$

Let us recall that infinite words with Property $\mathcal{R}$ are necessarily uniformly recurrent. If their language is not closed under reversal, then it cannot contain infinitely many palindromes by Theorem 2. Such words exist, as illustrated by the following example, therefore, Property $\mathcal{R}$ does not imply $\mathcal{P}$.

Example 5 ($\mathcal{R} \not\Rightarrow \mathcal{P}$)**.** The fixed point $\mathbf{u}$ of φ, where $\varphi(a) = aab$, $\varphi(b) = ac$, $\varphi(c) = a$, contains bac, but cab is not its factor. The fact that every factor of $\mathbf{u}$ has three return words is explained in [9] for a whole class of infinite words coding β-integers.

We have seen that AR words and k-iet words have both Property $\mathcal{R}$ and $\mathcal{C}$, however, as shown in [26] by the following example, Property $\mathcal{C}$ does not imply Property $\mathcal{R}$ on multiliteral alphabets.

Example 6 ($\mathcal{C} \not\Rightarrow \mathcal{R}$)**.** The fixed point of $\varphi : a \to ab$, $b \to cab, c \to ccab$ – the above mentioned recoding of the Chacon substitution – has the complexity $2n + 1$ for every $n \in \mathbb{N}$, but contains more than three return words of certain factors (for example the factor bc has 4 return words: bca, $bcca$, $bcaba$ and $bccaba$.

The following theorems come from the paper [9] that is devoted to the study of Property $\mathcal{R}$ for infinite words on multiliteral alphabets. Let us observe once more AR words and k-iet words, these classes satisfy not only Property $\mathcal{C}$, but also Property $\mathcal{BO}$. It is thus natural to ask whether Property $\mathcal{BO}$ guarantees $\mathcal{R}$. The corollary of the following theorem will provide an answer.

Theorem 14. *If $\mathbf{u}$ is an infinite word with no weak BS factors, then $\mathbf{u}$ has Property $\mathcal{R}$ if and only if $\mathbf{u}$ is uniformly recurrent and satisfies $\mathcal{C}$.*

Let us underline, an infinite word **u** has Property $\mathcal{BO}$ if and only if it has Property $\mathcal{C}$ and contains no weak BS factors. It results in the advertised corollary.

Corollary 15. *Let* **u** *be a uniformly recurrent infinite word. Then*

$$\mathcal{BO} \Rightarrow \mathcal{R}.$$

If we restrict our consideration to the ternary alphabet, the implication can be reversed.

Theorem 16. *Let* **u** *be a ternary uniformly recurrent infinite word. Then*

$$\mathcal{BO} \Leftrightarrow \mathcal{R}.$$

As soon as the alphabet has more than three letters, Property $\mathcal{R}$ does not imply Property $\mathcal{BO}$ any more.

Example 7 ($\mathcal{R} \not\Rightarrow \mathcal{BO}$)**.** The uniformly recurrent infinite word $\mathbf{u} = \lim_{n \to \infty} \varphi^n(a)$, where

$$\varphi(a) = acbca, \ \varphi(b) = acbcadbdaca, \ \varphi(c) = dbcbdacadbd, \ \varphi(d) = dbcbd,$$

satisfies $\mathcal{R}$, but not $\mathcal{C}$ (since $\mathcal{C}(n)$ is even for all $n \in \mathbb{N}$) and **u** contains, of course, weak BS factors. For details consult [9].

The question whether there exists a nice characterization of words with Property $\mathcal{R}$ on alphabets with more than three letters remains open.

5.5. Property $\mathcal{P}$ and $\mathcal{PE}$

The paper [8] is focused on the study of Properties $\mathcal{P}$ and $\mathcal{PE}$. As soon as an infinite word **u** has Property $\mathcal{PE}$, then **u** has exactly one infinite palindromic branch with center a for every letter $a \in \mathcal{A}$ and one infinite palindromic branch with center ε. Therefore, **u** contains exactly $\#\mathcal{A}$ palindromes for every odd length (central factors of palindromic branches with centers $a \in \mathcal{A}$) and one palindrome for every even length (central factor of the infinite palindromic branch with center ε). Consequently, Property $\mathcal{P}$ is also satisfied by **u**.

Let us recall that Property $\mathcal{P}$ may be reformulated in the following way

$$\mathcal{P}(n+2) = \mathcal{P}(n) \quad \text{for all } n \in \mathbb{N}, \tag{5.1}$$

where $\mathcal{P}(0) = 1$. We will equally use both of the forms of Property $\mathcal{P}$.

Let **u** be an infinite word satisfying $\mathcal{PE}$. The language $\mathcal{L}(\mathbf{u})$ contains infinitely many palindromes, but it need not be closed under reversal, neither recurrent nor rich as illustrated by the following example.

Example 8 ($\mathcal{PE} \not\Rightarrow$ language closed under reversal, $\mathcal{PE} \not\Rightarrow$ richness)**.** The infinite word **u** on the alphabet $\{a, b, c\}$ defined in the following way:

$$\mathbf{u} = caccb \underbrace{ccc}_{3\times} a \underbrace{cccc}_{4\times} b \underbrace{ccccc}_{5\times} a \underbrace{cccccc}_{6\times} b \underbrace{ccccccc}_{7\times} a \ldots$$

has three infinite palindromic branches with centers a, b and c

$$\ldots cccaccc \ldots, \quad \ldots cccbccc \ldots, \quad \ldots ccccccc \ldots$$

and one infinite palindromic branch with central factors of even length of the form $\ldots cccccccc \ldots$ The factor $accb$ occurs only once in **u**, thus **u** is not recurrent and hence $\mathcal{L}(\mathbf{u})$ is not closed under reversal. Moreover, **u** is not rich since the prefix $caccbccca$ of length 9 contains only 9 palindromes:
$\varepsilon, a, b, c, cc, cac, cbc, ccc$ and $ccbcc$.

However, if the language $\mathcal{L}(\mathbf{u})$ is closed under reversal, then it is possible to say more about the relation of Properties $\mathcal{P}$ and $\mathcal{C}$ and the richness of $\mathbf{u}$. When both $\mathcal{P}$ and $\mathcal{C}$ are satisfied, the equality in (3.1) is reached. Application of Theorem 8 provides us with the following corollary.

Corollary 17. *Let $\mathbf{u}$ be an infinite word whose language is closed under reversal. Then*

$$\mathcal{P} + \mathcal{C} \Rightarrow richness\ of\ \mathbf{u}.$$

The first example shows that Property $\mathcal{P}$ itself does not guarantee richness even if the language is closed under reversal. The second one illustrates that the implication in Corollary 17 cannot be reversed.

Example 9 ($\mathcal{PE} \not\Rightarrow$ richness, $\mathcal{PE} \not\Rightarrow \mathcal{C}$)**.** A known example of an infinite word with the language closed under reversal and with a higher factor complexity is the billiard sequence on three letters, for which $\mathcal{C}(n) = n^2 + n + 1$. As shown in [14], such words satisfy Property $\mathcal{PE}$, hence $\mathcal{P}$ as well. Consequently, billiard sequences do not reach the upper bound in (3.1) and by Theorem 8 cannot be rich.

Example 10 (richness $\not\Rightarrow \mathcal{P}$, richness $\not\Rightarrow \mathcal{C}$)**.** Let φ be defined on an m-letter alphabet as follows:

$$\varphi(0) = 0^t 1, \quad \varphi(1) = 0^t 2, \quad \ldots, \varphi(m-2) = 0^t(m-1), \quad \varphi(m-1) = 0^s,$$

where $s, t \in \mathbb{N}$ and $t \geq s \geq 2$. The fixed point $\mathbf{u}$ of φ satisfies the equality $\mathcal{P}(n+1) + \mathcal{P}(n) = \Delta\mathcal{C}(n) + 2$ for all n. As the language is closed under reversal, by Theorem 8 $\mathbf{u}$ is rich. Property $\mathcal{P}$ is not satisfied since the sum $\mathcal{P}(n+1) + \mathcal{P}(n)$ is not constant. Further properties of palindromes in $\mathbf{u}$ can be found in [4].

Let us examine in the sequel the connection between Properties $\mathcal{C}$ and $\mathcal{P}$, resp. $\mathcal{C}$ and $\mathcal{PE}$.

5.5.1. *Ternary alphabet*

Let us limit our considerations to the ternary alphabet. The following theorem and examples come from [8].

Theorem 18. *Let $\mathbf{u}$ be an infinite ternary word with the language closed under reversal. Then*
  (1) $\mathcal{C} \Rightarrow \mathcal{P}$,
  (2) $\mathcal{BO} \Rightarrow \mathcal{PE}$.

The implication in Theorem 18 cannot be reversed. We have already illustrated in Example 9 that even the stronger property $\mathcal{PE}$ does not ensure $\mathcal{C}$. Let us provide one more counterexample - a fixed point of a substitution.

Example 11 ($\mathcal{PE} \not\Rightarrow \mathcal{C}$)**.** Denote by $\mathbf{u}$ the infinite ternary word being the fixed point of the substitution Φ defined by

$$\Phi(a) = aba, \quad \Phi(b) = cac, \quad \Phi(c) = aca. \tag{5.2}$$

Then the language of $\mathbf{u}$ is closed under reversal. On one hand, $\mathbf{u}$ has Property $\mathcal{PE}$, consequently, $\mathbf{u}$ has Property $\mathcal{P}$, too. On the other hand, Property $\mathcal{C}$ fails and $\mathcal{L}(\mathbf{u})$ contains infinitely many weak BS factors.

Properties $\mathcal{P}$ and $\mathcal{PE}$ are equivalent for binary words. However already for ternary words, the implication $\mathcal{P} \Rightarrow \mathcal{PE}$ does not hold any more.

Example 12 ($\mathcal{P} \not\Rightarrow \mathcal{PE}$)**.** Let $\mathbf{v}$ be the ternary infinite word defined by $\mathbf{v} = \Psi(\mathbf{u})$, where $\Psi : \{A, B\}^* \rightarrow \{a, b, c\}^*$ is the morphism given by

$$\Psi(A) = bc \quad \text{and} \quad \Psi(B) = baa,$$

and $\mathbf{u}$ is the fixed point of the substitution φ defined by

$$\varphi(A) = ABBABBA, \quad \varphi(B) = ABA.$$

Then $\mathbf{v}$ satisfies $\mathcal{P}$, but does not satisfy $\mathcal{PE}$.

The relation between $\mathcal{R}$ and $\mathcal{P}$ follows from Theorem 18 and Theorem 16.

Corollary 19. *Let* **u** *be an infinite ternary word with the language closed under reversal. Then*

$$\mathcal{R} \Rightarrow \mathcal{PE}.$$

The implication cannot be reversed.

Example 13 ($\mathcal{PE} \not\Rightarrow \mathcal{R}$)**.** Consider the fixed point **u** of the substitution in (5.2). As mentioned above, **u** contains weak BS factors. Then by Theorem 16, **u** does not satisfy $\mathcal{R}$.

Putting together Theorems 18 and Corollary 17, we obtain one more corollary.

Corollary 20. *Let* **u** *be an infinite ternary word with the language closed under reversal. Then*

$$\mathcal{C} \Rightarrow \text{richness of } \mathbf{u}.$$

In contrast with Corollary 17, we see that on a ternary alphabet already Property $\mathcal{C}$ itself ensures richness.

Neither in this case, the reversed implication holds. Consult Example 10 or the following example with a periodic word.

Example 14 (richness $\not\Rightarrow \mathcal{C}$)**.** The periodic infinite word $(abcba)^\omega$ is rich (since complete return words of palindromic factors are palindromes) and has a bounded complexity.

5.5.2. *Multiliteral alphabet*

In this section, two new theorems concerning Properties $\mathcal{P}$ and $\mathcal{PE}$ for multiliteral infinite words will be proved.

Theorem 21. *Let* **u** *be an infinite word with the language closed under reversal.*

$$\text{Assume } \mathcal{C}: \quad \mathcal{PE} \Leftrightarrow \mathcal{BO}.$$

Proof. ($\Leftarrow$): Let us prove the statement by contradiction. Assume that Property $\mathcal{BO}$ holds and Property $\mathcal{PE}$ does not. It is clear that the property $\mathcal{PE}$ can only be violated on a palindromic BS factor. By Property $\mathcal{BO}$, all palindromic factors have their bilateral order equal to zero. By Proposition 9, they have an odd number of palindromic extensions, particularly at least one.

Since the language is closed under reversal, Theorem 3 implies the inequality (3.1) for all $n \in \mathbb{N}$

$$\mathcal{P}(n) + \mathcal{P}(n+1) \leq 2 + \Delta\mathcal{C}(n).$$

Let w denote the shortest palindromic BS factor that does not have exactly one palindromic extension. Denote $N = |w|$. Then we have for all $n \leq N$,

$$\mathcal{P}(n) + \mathcal{P}(n+1) = \#\mathcal{A} + 1.$$

Since Property $\mathcal{BO}$ implies Property $\mathcal{C}$, we have $2 + \Delta\mathcal{C}(n) = 2 + (\#\mathcal{A} - 1)$, hence the equality in (3.1) is attained for all $n \leq N$.

Since w has to have at least 3 palindromic extensions, one can see that $\mathcal{P}(N+2) \geq \mathcal{P}(N)+2$. Thus, we obtain $\mathcal{P}(N+1)+\mathcal{P}(N+2) \geq \mathcal{P}(N+1)+\mathcal{P}(N)+2 = \#\mathcal{A}+3 = \Delta\mathcal{C}(N+1)+4$, which is a contradiction with (3.1). We conclude that Property $\mathcal{PE}$ holds.

($\Rightarrow$): Assume Property $\mathcal{PE}$ holds. Then Property $\mathcal{P}$ holds as well. By Corollary 17 **u** is rich. Consequently, we can apply Theorem 10 and we obtain $\mathrm{b}(w) = 0$ for all non-palindromic BS factors and $\mathrm{b}(w) = \#\mathrm{Pext}(w) - 1$ for all palindromic BS factors. By Property $\mathcal{PE}$ every palindromic BS factor has a unique palindromic extension, thus $\mathrm{b}(w) = 0$ for palindromic BS factors, too. $\qquad\square$

Let us deduce several corollaries of Theorem 21. The most straightforward concerns richness and Property $\mathcal{BO}$. It follows combining Theorems 21 and 8.

Corollary 22. *Let* **u** *be an infinite word with the language closed under reversal. Then*

$$\mathcal{BO} \Rightarrow richness\ of\ \mathbf{u}.$$

Putting together Theorems 2, 14 and 21, we obtain the following corollaries.

Corollary 23. *Let* **u** *be a uniformly recurrent infinite word.*

$$Assume\ \mathcal{C}: \quad \mathcal{PE} \Rightarrow \mathcal{R}.$$

The reversed implication does not hold. Property $\mathcal{R}$ does not even guarantee the weaker property $\mathcal{P}$.

Example 15 ($\mathcal{R} + \mathcal{C} \not\Rightarrow \mathcal{P}$)**.** Consider again the infinite word from the previous section: the fixed point **u** of φ, where $\varphi(a) = aab$, $\varphi(b) = ac$, $\varphi(c) = a$. Properties $\mathcal{C}$ and $\mathcal{R}$ are satisfied (as explained in [9]), **u** is uniformly recurrent and the language $\mathcal{L}(\mathbf{u})$ is not closed under reversal. By Theorem 2, **u** contains only a finite number of palindromes.

Notice that the assumptions in Corollary 23 imply that the language $\mathcal{L}(\mathbf{u})$ is closed under reversal. It is natural to ask whether the implication $\mathcal{R} \Rightarrow \mathcal{PE}$ holds for infinite words with the language closed under reversal. The answer is however negative. Property $\mathcal{R}$ does not imply even the weaker property $\mathcal{P}$.

Example 16. ($\mathcal{R} +$ language closed under reversal $\not\Rightarrow \mathcal{P}$) Consider again the uniformly recurrent infinite word from [9] given by $\mathbf{u} = \lim_{n \to \infty} \varphi^n(a)$, where

$$\varphi(a) = acbca,\ \varphi(b) = acbcadbdaca,\ \varphi(c) = dbcbdacadbd,\ \varphi(d) = dbcbd.$$

It satisfies $\mathcal{R}$, but $\mathcal{C}$ and $\mathcal{BO}$ are violated. It is not difficult to find infinitely many palindromes among weak BS factors. Thus, the language $\mathcal{L}(\mathbf{u})$ is closed under reversal. However $\mathcal{PE}$ is not satisfied because $cbc, dbd \in \mathcal{L}(\mathbf{u})$. Nor $\mathcal{P}$ holds since $\mathcal{P}(1) + \mathcal{P}(2) = 4 \neq 5$.

We notice in the previous examples that to demand either only the language closed under reversal or only Property $\mathcal{C}$ in order to reverse the implication in Corollary 23 is not sufficient. It is however not solved whether any infinite word with the language closed under reversal and having Properties $\mathcal{C}$ and $\mathcal{R}$ satisfies Property $\mathcal{PE}$ or at least $\mathcal{P}$ as well.

Corollary 24. *Let* **u** *be a uniformly recurrent infinite word.*

$$Assume\ \mathcal{PE}: \quad richness\ of\ \mathbf{u} \Leftrightarrow \mathcal{R}.$$

Proof. Recall that by Theorem 2, the language is closed under reversal.

($\Rightarrow$): Suppose **u** is rich. Then Property $\mathcal{PE}$ guarantees that Property $\mathcal{P}$ holds as well. Since the language is closed under reversal, Property $\mathcal{P}$ together with Theorem 8 implies $\mathcal{C}$ is also satisfied. The statement follows then by Corollary 23.

($\Leftarrow$): Let us prove the second implication by contradiction. Assume $\mathcal{R}$ is satisfied and **u** is not rich. Theorem 6 claims that there exists a palindrome w which has a complete return word that is not a palindrome itself. As $\mathcal{PE}$ holds, the language has $\#\mathcal{A} + 1$ biinfinite palindromic branches. As w is a palindrome, we can find it in the middle of one branch. Since **u** is uniformly recurrent, we can find w in a bounded distance from the center (on both sides) of the remaining $\#\mathcal{A}$ branches. Thus we have $\#\mathcal{A}$ distinct palindromic complete return words of w. As w was supposed to have a non-palindromic complete return word, we have a contradiction with $\mathcal{R}$. $\square$

In Theorem 21 for infinite words having Property $\mathcal{C}$, we have proved that Property $\mathcal{PE}$ coincides with Property $\mathcal{BO}$. Under the same assumption on the complexity, we are again able to characterize Property $\mathcal{P}$ imposing this time a weaker condition on bilateral orders of BS factors.

Theorem 25. *Let* **u** *be an infinite word with the language closed under reversal and satisfying Property* $\mathcal{C}$. *Then Property* $\mathcal{P}$ *holds if and only if any bispecial factor* w *of* **u** *satisfies:*

- *if* w *is non-palindromic, then*

$$\mathrm{b}(w) = 0,$$

- *if* w *is a palindrome, then*

$$\mathrm{b}(w) = \#\mathrm{Pext}(w) - 1.$$

Proof. ($\Leftarrow$): Theorem 10 implies that **u** is rich. Since the language is closed under reversal, we can use Theorem 8. By Property $\mathcal{C}$, we have $\mathcal{P}(n+1) + \mathcal{P}(n) = \Delta\mathcal{C}(n) + 2 = \#\mathcal{A} + 1$, thus Property $\mathcal{P}$ holds.

($\Rightarrow$): Corollary 17 states that **u** is rich. The statement about bilateral orders follows then by Theorem 10. □

This theorem may be immediately reformulated using Theorem 10.

Corollary 26. *Let* **u** *be an infinite word with the language closed under reversal.*

$$\textit{Assume } \mathcal{C}: \quad \mathcal{P} \Leftrightarrow \textit{richness of } \mathbf{u}.$$

Non-palindromic bispecial factors can really occur in infinite words with the language closed under reversal and satisfying Properties $\mathcal{C}$ and $\mathcal{PE}$, thus $\mathcal{P}$ as well. This means that there exist rich words with non-palindromic BS factors.

Example 17. A ternary word with such properties is $\mathbf{v} = \pi(\mathbf{u})$, where $\mathbf{u} = \varphi^2(\mathbf{u})$ and

$$\varphi : A \to CAC, \ B \to CACBD, \ C \to BDBCA, \ D \to BDB,$$

$$\pi : A \to ba, \ B \to b, \ C \to a, \ D \to abc.$$

The substitution φ satisfies for any letter $x \in \{A, B, C, D\}$, if we cut off the last two letters of $\varphi^{2n}(x)$, we get a palindrome. Together with the uniform recurrence of **u**, Theorem 2 implies that the language $\mathcal{L}(\mathbf{u})$ is closed under reversal. Every LS factor of **u** is a prefix of $\varphi^{2n}(B)$ or $\varphi^{2n}(C)$ for some $n \in \mathbb{N}$, consequently, $\Delta\mathcal{C}(n) = 2$ for all $n \in \mathbb{N}, \ n \geq 1$.

For every non-empty palindrome $w \in \mathcal{L}(\mathbf{u})$ (except for B and C), its morphic image $\pi(w)$ without first two letters is a palindrome. As **v** contains infinitely many distinct palindromes and is a morphic image of a uniformly recurrent word, thus uniformly recurrent, too, the language $\mathcal{L}(\mathbf{v})$ is closed under reversal. The word **v** has two infinite LS branches: every LS factor of **v** is either a prefix of $\pi(\varphi^{2n}(B))$ or of $\pi(\varphi^{2n}(C))$. Therefore, **v** satisfies Property $\mathcal{C}$. Moreover, **v** contains only ordinary BS factors. Applying Theorem 21, Property $\mathcal{PE}$ holds as well. Remark that the factor ba is a non-palindromic BS factor of **v**.

5.6. BALANCE PROPERTIES

It is a direct consequence of the definition that

$$\mathcal{AC} \ \Rightarrow \ \mathcal{B}_\forall \ \Rightarrow \ \mathcal{B}_\exists. \tag{5.3}$$

The first implication follows from the fact that if there are two factors v, w of the same length that contain a distinct number of letters a, say l and r, then there exist factors containing any number of letters a between l and r (they may be found in any factor having v as its prefix and w as its suffix, or vice versa).

Let us point out that our favorite generalizations of Sturmian words, namely AR words and k-iet words, violate the property $\mathcal{B}_\forall$. The paper [19] provides a construction of an AR word **u** that is not c-balanced for any c. The same property have also all 3-iet words given by the transformation T associated with the symmetric permutation and verifying the property i.d.o.c., which can be shown using methods from [1].

It is natural to ask whether infinite words on multiliteral alphabets with Property $\mathcal{AC}$ exist. A recent answer has been provided in [21]: there are no infinite words satisfying $\mathcal{AC}$ on alphabets containing more

than 3 letters. On the other hand, there exist ternary infinite words with Property $\mathcal{AC}$ as shown by the example taken from [42].

Example 18. Let $\mathbf{v}$ be any aperiodic infinite word on $\{A, B\}$ and put $\mathbf{u} = \pi(\mathbf{v})$, where π is the morphism defined by $\pi(A) = abc$, $\pi(B) = acb$. Then $\mathcal{AC}(n) = 3$ for all $n \in \mathbb{N}$, $n \geq 1$.

A more general theorem has been proved ibidem.

Theorem 27. *If an aperiodic uniformly recurrent infinite word $\mathbf{u}$ on a ternary alphabet is 1-balanced, then $\mathbf{u}$ has Property $\mathcal{AC}$.*

Let us underline in the following examples that none of the implications in (5.3) can be reversed. The first example comes from [43] and the second one is taken from [47].

Example 19 ($\mathcal{B}_\forall \not\Rightarrow \mathcal{AC}$)**.** The ternary Tribonacci word – the fixed point of the substitution $\varphi : a \rightarrow ab$, $b \rightarrow ac$, $c \rightarrow a$ – is 2-balanced, however its abelian complexity reaches five values: $3, 4, 5, 6, 7$. Notice that the Tribonacci word belongs to AR words, which satisfy Properties $\mathcal{C}, \mathcal{LR}, \mathcal{BO}, \mathcal{R}, \mathcal{P}, \mathcal{PE}$.

Example 20 ($\mathcal{B}_\exists \not\Rightarrow \mathcal{B}_\forall$)**.** The fixed point $\mathbf{u}$ of the substitution $\varphi : a \rightarrow aab$, $b \rightarrow c$, $c \rightarrow ab$ has the following properties (shown in [47]):

- for any factors $v, w \in \mathcal{L}(\mathbf{u})$ with $|v| = |w|$, it holds

$$||v|_x - |w|_x| \leq 2 \quad \text{if } x \in \{b, c\},$$

- there exist $v, w \in \mathcal{L}(\mathbf{u})$ with $|v| = |w|$ such that

$$||v|_a - |w|_a| = 3.$$

Thus, $\mathbf{u}$ has Property $\mathcal{B}_\exists$. The word $\mathbf{u}$ is a coding of distances between neighboring β-integers, where β is the largest root of the polynomial $x^3 - 2x^2 - x + 1$. The word $\mathbf{u}$ is moreover known (see [28]) to verify Property $\mathcal{BO}$, but not $\mathcal{LR}$. Theorem 14 implies that $\mathbf{u}$ has Property $\mathcal{R}$ as well. Its language is not closed under reversal, consequently, neither $\mathcal{PE}$ nor $\mathcal{P}$ holds.

Generally, it is difficult to decide whether an infinite word has Property $\mathcal{B}_\exists$ or $\mathcal{B}_\forall$. A slightly simpler problem is to study infinite words that are c-balanced for some c. The criterion for existence of such a constant c for fixed points of a primitive substitution has been provided in [2], observing the spectra of adjacence matrices of substitutions. In general, it is however impossible to determine the minimal value of c from the spectrum. To our knowledge, besides the ternary words considered in Examples 19 and 20, the only non-sturmian fixed points of primitive substitutions, for which the minimal value of c is known, have been examined in [10] and [48].

6. Overview of relations and examples

In this section we provide a brief overview of relations and examples presented in the paper. Most of the relations are depicted in Figure 3. Examples are listed in Table 1. The word is either a fixed point of the given substitution, the image by the morphism π of a fixed point of the substitution φ, the limit of the sequence (u_n) or otherwise specified.
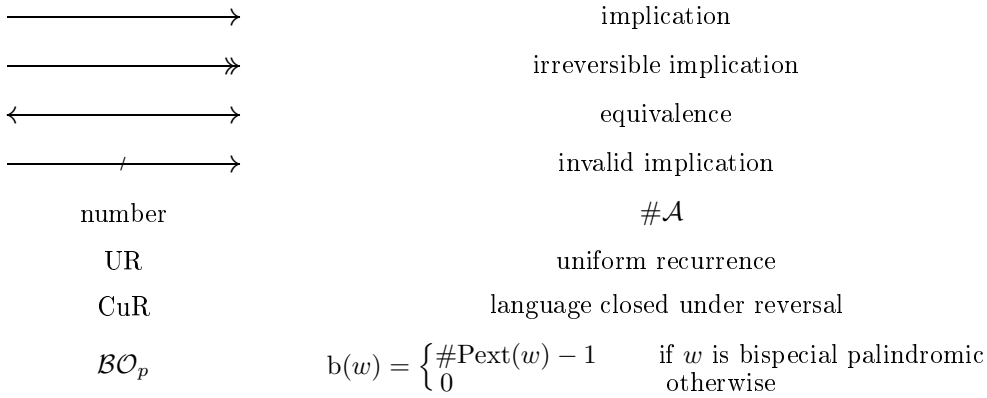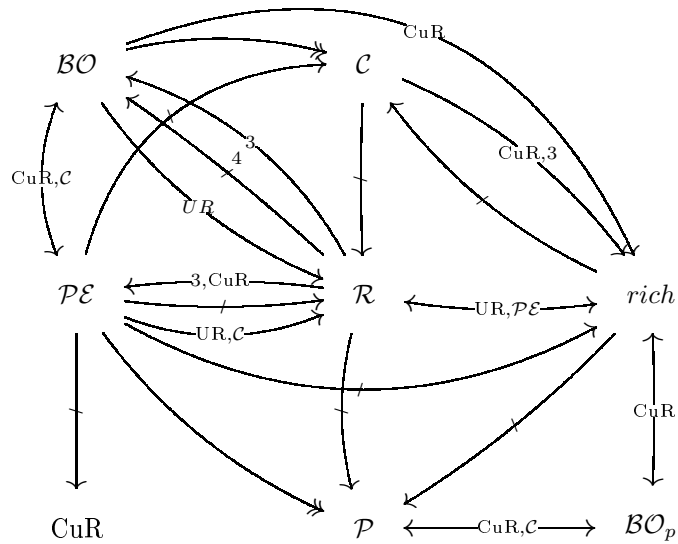
FIGURE 3. Diagram of known relations (assumptions are marked as labels of arrows)

7. ACKNOWLEDGEMENTS

word	properties	reference
$u_0 = ab$, $u_{n+1} = u_n ab\widetilde{u_n}$	uniformly recurrent, closed under reversal, finite number of palindromes	ex. 1 on p. 5, [13]
$u_0 = \varepsilon$, $u_{n+1} = u_n abc^{n+1} u_n$	recurrent, ∞-many palindromes, not closed under reversal	ex. 2 on p. 5, [16]
$a \to ab$, $b \to cab$, $c \to ccab$	$\mathcal{C}$, not $\mathcal{BO}$, not $\mathcal{R}$	ex. 3 on p. 12, ex. 6 on p. 12, [26]
φ: $A \to AB$, $B \to A$; π: $A \to a$, $B \to bc$	$\mathcal{LR}$, not closed under reversal, finite number of palindromes, not $\mathcal{C}$, not $\mathcal{R}$	ex. 4 on p. 12
$a \to aab$, $b \to ac$, $c \to a$	$\mathcal{R}$, not closed under reversal	ex. 5 on p. 12, [9]
$a \to acbca$, $b \to acbcadbdaca$, $c \to dbcbdacadbd$, $d \to dbcbd$	$\mathcal{R}$, closed under reversal, not $\mathcal{C}$, not $\mathcal{P}$	ex. 7 on p. 13, ex. 16 on p. 16, [9]
$\mathbf{u} = ca\ \underbrace{cc}_{2\times}\ b\ \underbrace{ccc}_{3\times}\ a\ \underbrace{cccc}_{4\times}\ b\ \underbrace{ccccc}_{5\times}\ a\ldots$	∞-many palindromes, not closed under reversal, not rich	ex. 8 on p. 13
billiard sequence on three letters	closed under reversal, $\mathcal{PE}$, not $\mathcal{C}$, not rich	ex. 9 on p. 14, [14]
$a \to aab$, $b \to aac$, $c \to aa$	rich, not $\mathcal{C}$, not $\mathcal{P}$	ex. 10 on p. 14, [4]
$a \to aba$, $b \to cac$, $c \to aca$	closed under reversal, $\mathcal{PE}$, not $\mathcal{C}$, not $\mathcal{R}$	ex. 11 on p. 14, ex. 13 on p. 15, [8]
φ: $A \to ABBABBA$, $B \to ABA$; π: $A \to bc$, $B \to baa$	closed under reversal, $\mathcal{C},\mathcal{P}$, not $\mathcal{PE}$	ex. 12 on p. 14, [8]
$(abcba)^\omega$	rich, not $\mathcal{C}$	ex. 14 on p. 15
$a \to aab$, $b \to ac$, $c \to a$	$\mathcal{C}$, $\mathcal{R}$, not closed under reversal	ex. 15 on p. 16, [9]
φ: $A \to CAC$, $B \to CACBD$, $C \to BDBCA$, $D \to BDB$; π: $A \to ba$, $B \to b$, $C \to a$, $D \to abc$	$\mathcal{PE}$, $\mathcal{C}$, closed under reversal, rich, contains non-palindromic BS factors	ex. 17 on p. 17
$\mathbf{u} = \pi(\mathbf{v})$, π: $A \to abc$, $B \to acb$, $\mathbf{v}$ is an aperiodic word over $\{A, B\}$	$\mathcal{AC}$	ex. 18 on p. 18, [42]
$a \to ab$, $b \to ac$, $c \to a$	$\mathcal{LR}, \mathcal{BO}, \mathcal{R}, \mathcal{PE}, \mathcal{B}_\forall$, not $\mathcal{AC}$	ex. 19 on p. 18, [43]
$a \to aab$, $b \to c$, $c \to ab$	$\mathcal{B}_\exists$, not $\mathcal{B}_\forall$, not closed under reversal, $\mathcal{BO}$, not $\mathcal{LR}$, $\mathcal{R}$	ex. 20 on p. 18, [47]

TABLE 1. Example overview

References

[1] B. Adamczewski, Codages de rotations et phénomenes d'autosimilarité. J. Théor. Nombres Bordeaux **14** (2002) 351-386.

[2] B. Adamczewski, Balances for fixed points of primitive substitutions. Theoret. Comput. Sci. **307** (2003) 47-75.

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

[4] P. Ambrož, Ch. Frougny, Z. Masáková, E. Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers. Ann. Inst. Fourier **56** (2006) 2131-2160.

[5] P. Arnoux, C. Mauduit, I. Shiokawa, J.-I. Tamura, Complexity of sequences defined by billiards in the cube. Bull. Soc. Math. France **122** (1994) 1-12.

[6] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n + 1$. Bull. Soc. Math. France **119** (1991) 199-215.

[7] P. Baláži, Z. Masáková, E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words. Theoret. Comput. Sci. **380** (2007) 266-275.

[8] Ľ. Balková, E. Pelantová, Š. Starosta, Palindromes in infinite ternary words. *RAIRO*: Theoret. Informatics Appl. **43** (2009) 687-702.

[9] Ľ. Balková, E. Pelantová, W. Steiner, Sequences with constant number of return words. Monatsh. Math. **155(3-4)** (2008) 251-263.

[10] Ľ. Balková, E. Pelantová, O. Turek, Combinatorial and arithmetical properties of infinite words associated with quadratic non-simple Parry numbers. *RAIRO*: Theoret. Informatics Appl. **41** (2007) 307-328.

[11] Y. Baryshnikov, Complexity of trajectories in rectangular billiards. Comm. Math. Phys. **174** (1995) 43-56.

[12] J. Berstel, Recent results on extensions of Sturmian words. Int. J. Algebra Comput. **12** (2002), 371-385.

[13] J. Berstel, L. Boasson, O. Carton, I. Fagnot, Infinite words without palindromes. arXiv:0903.2382 (2009)

[14] J.P. Borel, Complexity and palindromic complexity of billiards words, in *Proceedings of WORDS 2005*, edited by S. Brlek, C. Reutenauer (2005) 175-183.

[15] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words. Int. J. Found. Comput. Sci. **2** (2004) 293-306.

[16] 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.

[17] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A new characteristic property of rich words. Theoret. Comput. Sci. **410** (2009) 2860-2863.

[18] J. Cassaigne, Complexity and special factors. Bull. Belg. Math. Soc. Simon Stevin 4 **1** (1997) 67-88.

[19] J. Cassaigne, S. Ferenczi, L.Q. Zamboni, Imbalances in Arnoux-Rauzy sequences. Ann. Inst. Fourier **50** (2000) 1265-1276.

[20] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth. Math. Systems Theory **7** (1973) 138-153.

[21] J. Currie, N. Rampersad, Recurrent words with constant Abelian complexity. Adv. in Appl. Math. (2010) DOI:10.1016/j.aam.2010.05.001

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

[23] X. Droubay , J. Justin , G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy. Theoret. Comput. Sci. **255** (2001) 539-553.

[24] F. Durand, A characterization of substitutive sequences using return words. Discrete Math. **179** (1998) 89-101.

[25] I. Fagnot, L. Vuillon, Generalized balances in Sturmian words. Discrete Applied Mathematics **121** (2002) 83-101.

[26] S. Ferenczi, Les transformations de Chacon: combinatoire, structure géométrique, lien avec les systèmes de complexité $2n+1$. Bull. Soc. Math. Fr. **123** (1995) 271-292.

[27] S. Ferenczi, L. Zamboni, Languages of k-interval exchange transformations. Bull. London Math. Soc. **40(4)** (2008) 705-714.

[28] C. Frougny, Z. Masáková, E. Pelantová, Complexity of infinite words associated with beta-expansions. *RAIRO*: Theoret. Informatics Appl. **38** (2004) 162-184.

[29] A. Glen, J. Justin, Episturmian words: a survey. *RAIRO*: Theoret. Informatics Appl. **43** (2009) 403-442.

[30] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness. Eur. J. Comb. **30** (2009) 510-531.

[31] A. Hof, O. Knill, B. Simon, Singular continuous spectrum for palindromic Schröodinger operators. Comm. Math. Phys. **174** (1995) 149-159.

[32] C. Holton, L.Q. Zamboni, Geometric realizations of substitutions. Bull. Soc. Math. France **126** (1998) 149–179.

[33] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms. Theoret. Comput. Sci. **276** (2002) 281-313.

[34] J. Justin, L. Vuillon, Return words in Sturmian and episturmian words. *RAIRO*: Theoret. Informatics Appl. **34** (2000) 343-356.

[35] M.S. Keane, Interval exchange transformations. Math. Z. **141** (1975) 25-31.

[36] M. Lothaire, *Algebraic combinatorics on words.* volume 90 of Encyclopedia of Mathematics and its Applications, Cambridge University Press (2002).

[37] Z. Masáková, E. Pelantová, Relation between powers of factors and the recurrence function characterizing Sturmian words. Theoret. Comput. Sci. **410 (38-40)** (2009) 3589-3596.

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

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

[40] G. Rauzy, Echanges d'intervalles et transformations induites. Acta Arith. **34** (1979) 315-328.

[41] G. Richomme, Another characterization of Sturmian words (one more). Bull. Eur. Assoc. Theor. Comput. Sci. EATCS **67** (1999) 173-175.

[42] G. Richomme, K. Saari, L.Q. Zamboni, Abelian complexity in minimal subshifts. arXiv:0904.2925v1 (2009)

[43] G. Richomme, K. Saari, L.Q. Zamboni, Balance and abelian complexity of the Tribonacci word. Adv. Appl. Math. **45** (2010) 212–231.

[44] G. Rote, Sequences with subword complexity $2n$. J. Number Th. **46** (1993) 196-213.

[45] J. Smillie, C. Ulcigrai, Beyond Sturmian sequences: coding linear trajectories in the regular octagon. Proc. London Math. Soc. (2010) DOI: 10.1112/plms/pdq018

[46] S. Tabachnikov, *Billiards.* Panoramas et synthèse, SMF, Numéro 1 (1995).

[47] O. Turek, Balances and Abelian complexity of a certain class of infinite ternary words. *RAIRO*: Theoret. Informatics Appl. DOI: 10.1051/ita/2010017

[48] O. Turek, Balance properties of the fixed point of the substitution associated to quadratic simple Pisot numbers. *RAIRO*: Theoret. Informatics Appl. **41** (2007) 123-135.

[49] L. Vuillon, A characterization of Sturmian words by return words. Eur. J. Comb. **22** (2001) 263-275.

[50] L. Vuillon, Balanced words. Bull. Belg. Math. Soc. Simon Stevin **10** (2003) 787-805.

[51] L. Vuillon, *On the number of return words in infinite words with complexity* $2n + 1$. LIAFA Research Report (2000).