Theoretical Informatics and Applications Will be set by the publisher
Informatique Théorique et Applications

PALINDROMES IN INFINITE TERNARY WORDS

L'ubomíra Balková[1], Edita Pelantová[2] and Štěpán Starosta[3]

Abstract. We study infinite words $\mathbf{u}$ over an alphabet $\mathcal{A}$ satisfying the property

$$\mathcal{P}: \qquad \mathcal{P}(n) + \mathcal{P}(n+1) = 1 + \#\mathcal{A} \quad \text{for any } n \in \mathbb{N},$$

where $\mathcal{P}(n)$ denotes the number of palindromic factors of length n occurring in the language of $\mathbf{u}$. We study also infinite words satisfying a stronger property

$$\mathcal{PE}: \quad \text{every palindrome of } \mathbf{u} \text{ has exactly one palindromic extension in } \mathbf{u}\,.$$

For binary words, the properties $\mathcal{P}$ and $\mathcal{PE}$ coincide and these properties characterize Sturmian words, i.e., words with the complexity $\mathcal{C}(n) = n + 1$ for any $n \in \mathbb{N}$. In this paper, we focus on ternary infinite words with the language closed under reversal. For such words $\mathbf{u}$, we prove that if $\mathcal{C}(n) = 2n + 1$ for any $n \in \mathbb{N}$, then $\mathbf{u}$ satisfies the property $\mathcal{P}$ and moreover $\mathbf{u}$ is rich in palindromes. Also a sufficient condition for the property $\mathcal{PE}$ is given. We construct a word demonstrating that $\mathcal{P}$ on a ternary alphabet does not imply $\mathcal{PE}$.

2000 Mathematics Subject Classification. 68R15.

1. Introduction

Sturmian words are the most intensively studied infinite words since their appearance in 1940. They were introduced by Morse and Hedlund [7] as aperiodic words with the minimal possible complexity, i.e., with the complexity $\mathcal{C}(n) = n + 1$ for any $n \in \mathbb{N}$. The complexity is the function $\mathcal{C} : \mathbb{N} \mapsto \mathbb{N}$ defined by

$$\mathcal{C}(n) = \text{number of factors of length } n \text{ occurring in } \mathbf{u}\,.$$

The set of all factors occurring in $\mathbf{u}$ is called the *language* of $\mathbf{u}$ and denoted throughout this paper by $\mathcal{L}(\mathbf{u})$. There exist many equivalent definitions of Sturmian words. Already in [7], Sturmian words are characterized by their balance property. In the center of our attention will be another characteristics of Sturmian words, recently proved in [5]. This characteristics uses the palindromic complexity of $\mathbf{u}$, which

Keywords and phrases: ternary infinite words, palindromes, generalized Sturmian words, rich words

[1] Doppler Institute & Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, Praha 2 120 00, Czech Republic; e-mail: l.balkova@centrum.cz

[2] Doppler Institute & Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, Praha 2 120 00, Czech Republic; e-mail: edita.pelantova@fjfi.cvut.cz

[3] Institut de Mathématiques de Luminy, Campus de Luminy, Case 907, 13288 MARSEILLE Cedex 9 & Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, Praha 2 120 00, Czech Republic; e-mail: starosta@iml.univ-mrs.fr

is the function $\mathcal{P} : \mathbb{N} \mapsto \mathbb{N}$ defined by

$$\mathcal{P}(n) = \text{number of palindromic factors of length } n \text{ occurring in } \mathbf{u}\,.$$

Droubay and Pirillo proved that an infinite word $\mathbf{u}$ is Sturmian if and only if its palindromic complexity is

$$\mathcal{P}(n) = \begin{cases} 1 & \text{if } n \text{ is even,} \\ 2 & \text{if } n \text{ is odd.} \end{cases}$$

Since the empty word is the only palindrome of length 0 and the letters of the alphabet $\mathcal{A}$ are the only palindromes of length 1 in $\mathbf{u}$, the previous property can be rewritten in a compact form for binary infinite words as

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

Being inspired by Sturmian words, we generalize the previous property for infinite words over any alphabet $\mathcal{A}$ as

$$\mathcal{P}: \qquad \mathcal{P}(n) + \mathcal{P}(n+1) = 1 + \#\mathcal{A} \quad \text{for any } n \in \mathbb{N}.$$

It is again readily seen that the property $\mathcal{P}$ is equivalent with the property

$$\mathcal{P}(n) = \begin{cases} 1 & \text{if } n \text{ is even,} \\ \#\mathcal{A} & \text{if } n \text{ is odd.} \end{cases}$$

Examples of infinite words over multilateral alphabets satisfying the property $\mathcal{P}$ are Arnoux-Rauzy words and nondegenerate words coding the r-interval exchange transformation with the permutation $\pi = (r, r-1, r-2, \ldots, 2, 1)$.

When studying in details the proof of Droubay and Pirillo, we learn that a binary word $\mathbf{u}$ is Sturmian if and only if $\mathbf{u}$ satisfies the following condition

$$\mathcal{PE}: \qquad \text{any palindromic factor of } \mathbf{u} \text{ has a unique palindromic extension in } \mathbf{u}\,.$$

In other words, for any palindrome $p \in \mathcal{L}(\mathbf{u})$ there exists a unique letter $a \in \mathcal{A}$ such that $apa \in \mathcal{L}(\mathbf{u})$. In fact, our two examples of words with the property $\mathcal{P}$ - namely Arnoux-Rauzy words and words coding interval exchange - have even the property $\mathcal{PE}$.

Infinite words over a multilateral alphabet satisfying the property $\mathcal{P}$ or $\mathcal{PE}$ may be understood as one of the possible generalizations of Sturmian words. It is evident that $\mathcal{PE}$ implies $\mathcal{P}$. The inverse implication holds over a binary alphabet, but it need not hold in general. The validity of $\mathcal{P}$ or $\mathcal{PE}$ guarantees that the language $\mathcal{L}(\mathbf{u})$ contains infinitely many distinct palindromic factors. Such a language need not be closed under reversal. Nevertheless in the sequel, we concentrate on the study of ternary words whose language is closed under reversal. It is readily seen that such words are recurrent and their Rauzy graphs have a non-trivial automorphism that will serve as a powerful tool in our consideration.

We will prove the following two theorems:

Theorem 1.1. *An infinite ternary word whose language is closed under reversal has the property $\mathcal{P}$ if its complexity satisfies $\mathcal{C}(n) = 2n + 1$.*

For the description of $\mathcal{PE}$, an important role is played by the notion of a left special factor: a factor $w \in \mathcal{L}(\mathbf{u})$ is called *left special* if there exist at least two different letters a, b such that both $aw \in \mathcal{L}(\mathbf{u})$ and $bw \in \mathcal{L}(\mathbf{u})$. A left special factor w is called *maximal* if for any letter $c \in \mathcal{A}$, the factor wc is not left special.

Theorem 1.2. *An infinite ternary word $\mathbf{u}$ whose language is closed under reversal has the property $\mathcal{PE}$ if its complexity satisfies $\mathcal{C}(n) = 2n + 1$ and $\mathbf{u}$ has no maximal left special factor.*

It is interesting to mention two corollaries of the previous theorems. Vuillon [9] showed that a binary infinite word is Sturmian if and only if each of its factors has exactly two return words, i.e., Sturmian words are precisely binary words satisfying the property

$$\mathcal{R}: \quad \text{any factor of } \mathbf{u} \text{ has exactly } \#\mathcal{A} \text{ return words.}$$

In the paper [3], it is shown that a ternary infinite uniformly recurrent word $\mathbf{u}$ has the property $\mathcal{R}$ if and only if its complexity satisfies $\mathcal{C}(n) = 2n+1$ and $\mathbf{u}$ has no maximal left special factor. Consequently, for ternary infinite words with the language closed under reversal, $\mathcal{R}$ implies $\mathcal{PE}$.

Theorem 1.1 says that for infinite words whose language is closed under reversal and whose complexity satisfies $\mathcal{C}(n) = 2n+1$, the following equation holds

$$\mathcal{P}(n) + \mathcal{P}(n+1) = 2 + \mathcal{C}(n+1) - \mathcal{C}(n)\,. \tag{1.1}$$

Infinite words fulfilling the above equation are in a certain sense the richest in palindromes, since according to [2], any infinite word whose language is closed under reversal satisfies

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

Different descriptions of rich words defined by the equation (1.1) can be found in [6]. Infinite ternary words with the language closed under reversal and the complexity $\mathcal{C}(n) = 2n+1$ form a further class of rich words.

In Section 2, we recall basic notions from combinatorics on words. Section 3 contains the proofs of Theorems 1.1 and 1.2. Section 4 provides two examples of words: the first one shows that the properties $\mathcal{P}$ and $\mathcal{PE}$ are not equivalent and the second one proves that the implications in Theorems 1.1 and 1.2 cannot be reversed.

2. Preliminaries

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 $w = w_0w_1\dots w_{n-1}$ 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 assignment

$$w = w_0w_1\dots w_{n-1} \quad \mapsto \quad \overline{w} = w_{n-1}w_{n-2}\dots w_0$$

is a bijection on $\mathcal{A}^*$, the word $\overline{w}$ 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* $\mathbf{u}$ we understand an infinite string $\mathbf{u} = u_0u_1u_2\dots$ of letters from $\mathcal{A}$. A finite word w is a *factor* of a word v (finite or infinite) if there exist words $w^{(1)}$ and $w^{(2)}$ such that $v = w^{(1)}ww^{(2)}$. If $w^{(1)} = \varepsilon$, then w is said to be a *prefix* of v, if $w^{(2)} = \varepsilon$, then w is a *suffix* of v. We say that a prefix, 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 $\overline{w}$.

For any factor $w \in \mathcal{L}(\mathbf{u})$, there exists an index i such that w is a prefix of the infinite word $u_iu_{i+1}u_{i+2}\dots$. 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 is easy to see that if the language of $\mathbf{u}$ is closed under reversal, then $\mathbf{u}$ is recurrent.

The *complexity* of an infinite word $\mathbf{u}$ is a map $\mathcal{C} : \mathbb{N} \mapsto \mathbb{N}$, defined by $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$. To determine the increment of the complexity, one has to count the possible extensions of factors of length n. A *right extension* of $w \in \mathcal{L}(\mathbf{u})$ is any letter $a \in \mathcal{A}$ such that $wa \in \mathcal{L}(\mathbf{u})$. The set of all right extensions of a factor w will be denoted by $\mathrm{Rext}(w)$. Of course, any factor of $\mathbf{u}$ has at least one right extension. A factor w is called *right special* if w has at least two right extensions. Clearly, any suffix of a right special factor is right special as well. A right special factor w which is not a suffix of any longer right special factor is called a *maximal right special* factor. Similarly, one can define a *left extension*, a *left special* factor and $\mathrm{Lext}(w)$. We will deal only with recurrent infinite words $\mathbf{u}$. In this case, any factor of $\mathbf{u}$ has at least one left extension. If $a \in \mathcal{A}$ and p is a palindrome and $apa \in \mathcal{L}(\mathbf{u})$, then apa is said to be a *palindromic* *extension* of p. We say that w is a *bispecial* factor if it is right and left special. The role of bispecial factors for the computation of the complexity can be nicely illustrated on Rauzy graphs.

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 at the vertex x and ends at the vertex y if x is a prefix and y is a suffix of e. If the word $\mathbf{u}$ is recurrent, the graph Γ_n is strongly connected for

$$x = w_0 w_1 \cdots w_{n-1} \;\bullet \xrightarrow{\quad e = w_0 w_1 \cdots w_{n-1} w_n \quad} \bullet\; y = w_1 \cdots w_{n-1} w_n$$

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

every $n \in \mathbb{N}$, i.e., there exists a directed path from every vertex x to every vertex y of the graph.

The *outdegree* (*indegree*) of a vertex $x \in \mathcal{L}_n(\mathbf{u})$ is the number of edges which start (end) in x. Obviously the outdegree of x is equal to $\#\mathrm{Rext}(x)$ and the indegree of x is $\#\mathrm{Lext}(x)$.

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 we have

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

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

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

Let us restrict our consideration to recurrent words, then a non-zero contribution to $\Delta\mathcal{C}(n)$ is given only by those factors $x \in \mathcal{L}_n(\mathbf{u})$, for which $\#\mathrm{Rext}(x) \geq 2$ or $\#\mathrm{Lext}(x) \geq 2$, i.e., for right or left special factors. The relation (2.1) can be rewritten as

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

If the language of the infinite word $\mathbf{u}$ is closed under reversal, then the operation that to every vertex x of the graph associates the vertex $\overline{x}$ and to every edge e associates $\overline{e}$ maps the Rauzy graph Γ_n onto itself. In this case, we will draw the Rauzy graph Γ_n axially symmetric in the plane: the positions of vertices x and $\overline{x}$ are symmetrical with respect to an axis. Thus, x is a palindrome if and only if the vertex x lies on the axis, and e is a palindrome of length $n+1$ if and only if the edge e crosses the axis.

3. Proof of Theorems 1.1 and 1.2

The proofs of Theorems 1.1 and 1.2 will be a consequence of the following three lemmas that determine the number of palindromic extensions of palindromic factors with respect to the number of their left extensions.

Lemma 3.1. *Let* $\mathbf{u}$ *be an infinite word over an alphabet* $\mathcal{A}$ *whose language is closed under reversal. If a palindrome* $p \in \mathcal{L}(\mathbf{u})$ *is not left special (and thus neither right special), then* p *has a unique palindromic extension.*

Proof. Since $p \in \mathcal{L}(\mathbf{u})$ is not left special, there exists a unique $x \in \mathcal{A}$ such that $xp \in \mathcal{L}(\mathbf{u})$. By reversal closeness, $\mathcal{L}(\mathbf{u})$ contains also px. As p has a unique left extension x, the factor px has x as its unique left extension, too. Thus xpx is the unique palindromic extension of p. $\square$

Lemma 3.2. *Let* $\mathbf{u}$ *be an infinite word over a ternary alphabet* $\mathcal{A}$ *with the complexity* $\mathcal{C}(n) = 2n + 1$ *for any* $n \in \mathbb{N}$ *and with the language* $\mathcal{L}(\mathbf{u})$ *closed under reversal. If a palindrome* $p \in \mathcal{L}(\mathbf{u})$ *has* $\#\mathrm{Lext}(p) = 3$, *then* p *has a unique palindromic extension.*

Proof. As $\Delta\mathcal{C}(n) = 2$, the palindrome p is the only left special factor of length $n = |p|$, and by reversal closeness, the only right special factor of length n, too.

(1) First, assume that there exists a letter x such that $\mathrm{Lext}(px) = \mathcal{A}$. It means that xpx is a factor of $\mathbf{u}$, hence xpx is a palindromic extension of p. If there exists another palindromic extension of p, i.e., $ypy \in \mathcal{L}(\mathbf{u})$ for $y \neq x$, then since $y \in \mathrm{Lext}(px)$, it follows that ypx and xpy belong to $\mathcal{L}(\mathbf{u})$. Therefore $x, y \in \mathrm{Lext}(py)$, which implies

$$\Delta\mathcal{C}(n+1) \ \geq\ \#\mathrm{Lext}(px) - 1 + \#\mathrm{Lext}(py) - 1 \ \geq\ 3$$

- a contradiction.

(2) Second, suppose that for every letter x, it holds that $\mathrm{Lext}(px) \neq \mathcal{A}$. Let us recall that if w is a left special factor of length $n+1$, then its prefix of length n is necessarily left special, too. As a consequence, together with the fact that $\Delta C(n+1) = 2$, there exist two left special factors px and py in $\mathcal{L}_{n+1}(\mathbf{u})$ for $x \neq y$ with $\#\mathrm{Lext}(px) = \#\mathrm{Lext}(py) = 2$. Denote $\mathrm{Lext}(px) = \{a, b\}$ and $\mathrm{Lext}(py) = \{A, B\}$. Since our alphabet is ternary, we may assume WLOG that $a = A$. By reversal closeness, it follows that xpa and ypa belong to the language, and therefore, the factor pa is left special as well. WLOG $a = x$ and $b = y$. Denote by c the third letter of $\mathcal{A}$. Since $c \in \mathrm{Rext}(p)$ and by recurrence of $\mathbf{u}$, there exists a letter C such that $Cpc \in \mathcal{L}(\mathbf{u})$. However, since pc is not left special, this C is unique.

– If $C = a$, then $\mathrm{Lext}(pa) = \mathcal{A}$ - a contradiction.

– If $C = b$, then necessarily $B = c$ and apa is the unique palindromic extension of p, as claimed.

– If $C = c$, then $B = b$. The Rauzy graph Γ_n has a unique vertex of indegree > 1 (see Figure 2, where the straight lines denote edges and the zig zag lines denote paths) – the bispecial factor p. Consequently, the vertex p is the unique common vertex of three cycles. Since $\mathrm{Lext}(cp) = \{c\}$, after coming to the vertex p using the edge cp, we cannot leave p but using the edge pc. Hence, we move eventually in a unique cycle and the word $\mathbf{u}$ is thus eventually periodic - a contradiction with the complexity.

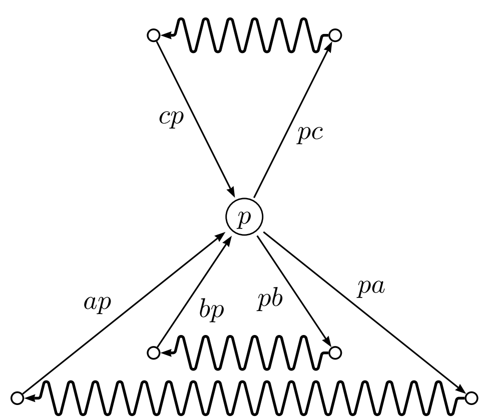

FIGURE 2

□

Lemma 3.3. *Let* $\mathbf{u}$ *be an infinite word over a ternary alphabet with the complexity* $\mathcal{C}(n) = 2n+1$ *for any* $n \in \mathbb{N}$ *and with the language* $\mathcal{L}(\mathbf{u})$ *closed under reversal. Let* $p \in \mathcal{L}(\mathbf{u})$ *be a palindrome with* $\#\mathrm{Lext}(p) = 2$ *and let* $\mathcal{P}(|p|) + \mathcal{P}(|p| + 1) = 4$.

(1) If p has no palindromic extension, then p is a maximal left special factor and there exists a palindrome q of the same length such that q has two palindromic extensions.
(2) If p has two palindromic extensions, then there exists a palindrome q of the same length such that q has no palindromic extension and q is a maximal left special factor.

Proof. Denote $\mathrm{Lext}(p) = \{a, b\}$ and $|p| = n$. Since $\Delta\mathcal{C}(n) = 2$, there exists a factor $q \neq p$ of the same length such that $\#\mathrm{Lext}(q) = 2$. Denote $\mathrm{Lext}(q) = \{A, B\}$.

(1) Assume that p has no palindromic extension. The only factors with length $n+2$ of the form $\ell_1 p \ell_2$ where $\ell_1, \ell_2 \in \mathcal{A}$ are apb and bpa. The factor p is thus a maximal left special factor. Let us recall that any prefix of a left special factor is again left special. This together with $\Delta\mathcal{C}(n+1) = 2$ implies that there exist two left special factors of length $n+1$: qx and qy for $x \neq y$ with

$$\mathrm{Lext}(qx) = \{A, B\} = \mathrm{Lext}(qy)\,. \tag{3.1}$$

By reversal closeness of $\mathcal{L}(\mathbf{u})$, we obtain that $\mathrm{Lext}(\overline{q}A) = \{x, y\}$ and $\mathrm{Lext}(\overline{q}B) = \{x, y\}$. Since there are no other left special factors besides qx and qy in $\mathcal{L}_{n+1}(\mathbf{u})$, necessarily $\overline{q} = q$ and $\{A, B\} = \{x, y\}$. Because of (3.1), we deduce that both xqx and yqy belong to the language $\mathcal{L}(\mathbf{u})$, i.e., the palindrome q has two palindromic extensions.

(2) Suppose that p has two palindromic extensions apa and bpb. In the Rauzy graph Γ_n, the bispecial factor p has the indegree and outdegree 2, the left special factor q has the indegree 2 and the right special factor $\overline{q}$ has the outdegree 2. Moreover, the palindromes of length n are exactly the vertices lying on the axis of symmetry and the palindromes of length $n+1$ are exactly the edges crossing the axis. These facts together with $\mathcal{P}(n) + \mathcal{P}(n+1) = 4$ imply that the Rauzy graph Γ_n can only look as depicted in Figure 3. Note that q and $\overline{q}$ may coincide. Let us first show, that

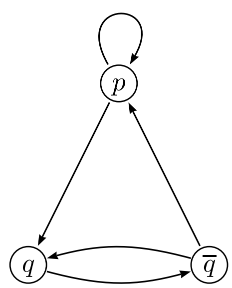

FIGURE 3

necessarily $apb \in \mathcal{L}(\mathbf{u})$. If not, then it is impossible in Γ_n to leave the cycles in which only the vertex p has the indegree or the outdegree bigger that 1. It means that the word $\mathbf{u}$ is eventually periodic - a contradiction with the complexity. Thus $apb, bpa \in \mathcal{L}(\mathbf{u})$. Consequently both pa and pb are left special factors of length $n+1$ with $\mathrm{Lext}(pa) = \{a, b\} = \mathrm{Lext}(pb)$. Since $\Delta\mathcal{C}(n+1) = 2$, no other left special factor of the same length exists. Thus q is the maximal left special and there exists a unique letter x and a unique letter y such that Aqx and Bqy belong to the language $\mathcal{L}(\mathbf{u})$ and $x \neq y$. It implies that $\mathrm{Lext}(\overline{q}) = \{x, y\}$, i.e., the factor $\overline{q}$ is a left special factor of length n, and therefore, $\overline{q} = q$ and $\{A, B\} = \{x, y\}$. The Rauzy graph Γ_n has two vertices with indegrees > 1 - the bispecial factors p and q, see Figure 4. Since q is a maximal left special factor, we have two disjoint possibilities: the first one is that xqx and yqy belong to the language, the second one is that xqy and yqx belong to the language. But the first possibility implies that in the Rauzy graph Γ_n, it is impossible to leave the cycles containing only one bispecial factor q - a contradiction. Therefore the second situation occurs and q has no palindromic extension.

$\square$

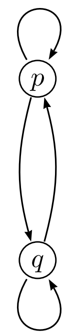

FIGURE 4

Proof of theorem 1.1. We will proceed by mathematical induction on n. Obviously, $\mathcal{P}(0) = 1$ and $\mathcal{P}(1) = 3$. Assume that $\mathcal{P}(n) + \mathcal{P}(n+1) = 4$ for some $n \geq 0$. Let $p \in \mathcal{L}_n(\mathbf{u})$ be a palindrome with zero or two palindromic extensions. According to Lemma 3.3 there exists a palindrome q of the same length, which is a left special factor as well. Since $\Delta\mathcal{C}(n) = 2$, all other factors including palindromes in $\mathcal{L}_n(\mathbf{u})$ have a unique left extension. According to Lemma 3.1, these palindromes have a unique palindromic extension. By Lemma 3.3, the palindromes p and q together have two palindromic extensions. Therefore, the number of palindromic extensions of all palindromes in $\mathcal{L}_n(\mathbf{u})$ together is equal to the number of palindromes of length n. Since every palindrome of length $n+2$ is a palindromic extension of a palindrome of length n, we obtain $\mathcal{P}(n+1) + \mathcal{P}(n+2) = 4$. □

Proof of theorem 1.2 is a direct consequence of the previous lemmas.

4. Counterexamples

In this last section, we will show that for ternary words, unlike binary words, the properties $\mathcal{P}$ and $\mathcal{PE}$ are not equivalent and we will provide counterexamples to reversed implications in Theorems 1.1 and 1.2.

We have seen that for the computation of the first difference of complexity $\Delta\mathcal{C}(n)$, an important role is played by left and right special factors. See Formula (2.1). In the sequel, it will be helpful to use a formula for the second difference of complexity $\Delta^2\mathcal{C}(n)$, introduced by Cassaigne [4]. Let us explain that for the computation of $\Delta^2\mathcal{C}(n)$, bispecial factors are crucial. Since every factor of length $n+2$ can be written as xwy, where $x, y \in \mathcal{A}$ and $w \in \mathcal{L}(\mathbf{u})$, it holds

$$\mathcal{C}(n+2) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} \#\{xwy \mid xwy \in \mathcal{L}(\mathbf{u})\},$$

and similarly,

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

The second difference of complexity $\Delta^2\mathcal{C}(n) = \Delta\mathcal{C}(n+1) - \Delta\mathcal{C}(n) = \mathcal{C}(n+2) - 2\mathcal{C}(n+1) + \mathcal{C}(n)$ may be obtained as follows

$$\Delta^2\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} \Big(\#\{xwy \mid xwy \in \mathcal{L}(\mathbf{u})\} - \#\mathrm{Lext}(w) - \#\mathrm{Rext}(w) + 1\Big). \tag{4.1}$$

Denote by

$$b(w) := \#\{xwy \mid xwy \in \mathcal{L}(\mathbf{u})\} - \#\mathrm{Lext}(w) - \#\mathrm{Rext}(w) + 1.$$

The number $b(w)$ is called the *bilateral order* of the factor w. It is readily seen that if w is not a bispecial factor, then $b(w) = 0$. Bispecial factors will be distinguished according to their bilateral order in the following way

- if $b(w) > 0$, then we call w a *strong* bispecial factor,
- if $b(w) < 0$, then we call w a *weak* bispecial factor,
- if $b(w) = 0$ and w is bispecial, then we call it *ordinary.*

Evidently, for the value of $\Delta^2\mathcal{C}(n)$, only strong and weak bispecial factors are of importance.

Remark 4.1. If p is a palindromic factor of a reversal closed language $\mathcal{L}(\mathbf{u})$, then $\#\{xpy \mid xpy \in \mathcal{L}(\mathbf{u})\}$ and the number of palindromic extensions of p in $\mathbf{u}$ have the same parity. Moreover, $\#\mathrm{Lext}(p) = \#\mathrm{Rext}(p)$. Therefore, the following simple observation holds

$$p \ \text{ has a unique palindromic extension in } \mathbf{u} \quad \Longrightarrow \quad b(p) \text{ is even.}$$

4.1. $\mathcal{P}$ and $\mathcal{PE}$ are not equivalent

The construction of a ternary infinite word $\mathbf{v}$ with the desired properties is inspired by Arnoux and Rauzy [1] and Rote [8]. Let $\mathbf{v}$ be the ternary infinite word defined by $\mathbf{v} = \Psi(\mathbf{u})$, where $\Psi : \{a,b\}^* \to \{0,1,2\}^*$ is the morphism given by

$$\Psi(a) = 12 \quad \text{and} \quad \Psi(b) = 100, \tag{4.2}$$

and $\mathbf{u}$ is the fixed point of the morphism $\varphi : \{a,b\}^* \to \{a,b\}^*$ defined by

$$\varphi(a) = abbabba, \quad \varphi(b) = aba. \tag{4.3}$$

In the sequel, we will show that $\mathbf{v}$ satisfies $\mathcal{P}$, but does not satisfy $\mathcal{PE}$. We will proceed in two steps. First, we will study several properties of the binary infinite word $\mathbf{u}$. Second, we will prove, using the properties of $\mathbf{u}$, that $\mathbf{v}$ satisfies $\mathcal{P}$, but does not satisfy $\mathcal{PE}$.

Step 1:

Let us show that the binary word $\mathbf{u}$ being the fixed point of the morphism φ given in (4.3) has the language $\mathcal{L}(\mathbf{u})$ closed under reversal and let us provide the list of all weak and strong bispecial factors of $\mathbf{u}$.

Let us start with an important observation.

Observation 4.2. *Every factor v of $\mathbf{u}$ can be decomposed as $v = v^{(0)}v^{(1)}\dots v^{(m)}$, $m \geq 1$, so that $v^{(i)} \in \{aba, abbabba\}$ for $i \in \{1,\dots,m-1\}$, $v^{(0)}$ is a proper suffix of abbabba and $v^{(m)}$ is a proper prefix of abbabba. Obviously, for every such decomposition, there exists $\tilde{v} \in \{a,b\}^*$ satisfying*

$$v = v^{(0)}\varphi(\tilde{v})v^{(m)}. \tag{4.4}$$

If the decomposition is unique, the corresponding $\tilde{v}$ is necessarily a factor of $\mathbf{u}$.

An essential role for the description of bispecial factors and palindromes in $\mathbf{u}$ is played by the map $T : \{a,b\}^* \to \{a,b\}^*$ defined by

$$T(w) = ba\varphi(w)ab \quad \text{for every } w \in \{a,b\}^*. \tag{4.5}$$

Let us summarize the properties of T in the following lemma.

Lemma 4.3. *Let T be the map defined in (4.5). Then, for every $w \in \{a,b\}^*$ and for all $c,d \in \{a,b\}$, it holds*

a) if w is a palindrome, then $T(w)$ is a palindrome, *b) cwd is a factor of* $\mathbf{u}$ *if and only if $cT(w)d$ is a factor of* $\mathbf{u}$*, in particular, if w is a factor of* $\mathbf{u}$*, then $T(w)$ is a factor of* $\mathbf{u}$*.*

Proof. $a)$ Since $\varphi(a) = aba$ and $\varphi(b) = abbabba$ are palindromes, it implies that $\varphi(w)$ is a palindrome, thus $T(w) = ba\varphi(w)ab$ is a palindrome, too.

$b)$ ($\Rightarrow$): If awb is a factor of $\mathbf{u}$, then $\varphi(awb)$ is in $\mathcal{L}(\mathbf{u})$. As $aT(w)b$ is a factor of $\varphi(awb) = aba\varphi(w)abbabba$, it follows that $aT(w)b$ is also a factor of $\mathbf{u}$. The proofs for the other cases awa, bwa, bwb are similar. ($\Leftarrow$): Let $aT(w)b$ be a factor of $\mathbf{u}$. It is readily seen that the unique decomposition of the form (4.4) of $aT(w)b$ is $aT(w)b = \varphi(aw)abb$. Since abb is a prefix of $\varphi(b)$, but not of $\varphi(a)$, it follows that $awb \in \mathcal{L}(\mathbf{u})$. The proofs for the other cases $aT(w)a, bT(w)a, bT(w)b$ are analogous.

□

Remark 4.4. Lemma 4.3 has several useful consequences.

(1) According to Lemma 4.3, the language $\mathcal{L}(\mathbf{u})$ contains infinitely many palindromes. Together with the primitivity of the substitution φ, thus the uniform recurrence of $\mathbf{u}$, it implies that the language $\mathcal{L}(\mathbf{u})$ is closed under reversal.
(2) For any factor $w \in \mathcal{L}(\mathbf{u})$, its bilateral order $b(w) = b(T(w))$ by Item $b)$ of Lemma 4.3.
(3) If w is a palindrome in $\mathcal{L}(\mathbf{u})$, then $T(w)$ is a palindrome with the same number of palindromic extensions by Lemma 4.3.

Since the word $\mathbf{u}$ is built from the factors $abbabba$ and aba, it is clear that the words

$$aaa, bbb, abab, baba, aabbaa, babbab$$

are not in $\mathcal{L}(\mathbf{u})$. Observing then the prefix of $\mathbf{u}$

$$\mathbf{u} = abbabbaabaabaabbabbaabaaba\ldots,$$

it follows that the only left special factors of length ≤ 4 are: $\varepsilon, a, b, ab, ba, abb, baa, abba, baab$; among them, only ε and $baab$ are strong bispecial factors and only $abba$ is a weak bispecial factor.

Lemma 4.5. *For every bispecial factor $v \in \mathcal{L}(\mathbf{u})$ of length at least 5, there exists a factor $w \in \mathcal{L}(\mathbf{u})$ such that $v = T(w)$. Moreover, $b(w) = b(T(w))$.*

Proof. Every prefix of a left special factor is left special, too. Since $abba$ and $baab$ are the only left special factors of length 4 and $abba$ is a weak bispecial factor, thus cannot be extended to the right staying left special, we learn that every bispecial (thus left special) factor v of length ≥ 5 has to start in $baab$. Since the language $\mathcal{L}(\mathbf{u})$ is closed under reversal, the bispecial (thus right special) factor v has to end in $baab$. Then, it is clear from the form of the substitution φ that $v = ba\varphi(w)ab$ is a unique decomposition of the form (4.4) of v. Thus, by Observation 4.2, w is a factor of $\mathbf{u}$ such that $v = T(w)$. The last statement is a consequence of Item $b)$ of Lemma 4.3. □

As a consequence of Lemma 4.5, we obtain the set of all strong bispecial factors

$$\{V^{(n)} \mid n \in \mathbb{N}\}, \quad \text{where } V^{(0)} = \varepsilon \text{ and } V^{(n)} = T(V^{(n-1)}) \text{ for } n \geq 1, \tag{4.6}$$

and the set of all weak bispecial factors

$$\{U^{(n)} \mid n \in \mathbb{N},\ n \geq 1\}, \quad \text{where } U^{(1)} = abba \text{ and } U^{(n)} = T(U^{(n-1)}) \text{ for } n \geq 2. \tag{4.7}$$

It is easy to see that $b(V^{(0)}) = b(\varepsilon) = 1$ and $b(U^{(1)}) = b(abba) = -1$. Item $b)$ of Lemma 4.3 implies that $b(V^{(n)}) = 1$ and $b(U^{(n)}) = -1$ for all n. Moreover, by Item $a)$ of Lemma 4.3, they are palindromes.

Step 2:

We may now study the ternary word $\mathbf{v} = \Psi(\mathbf{u})$ defined in (4.2). In the sequel, it will be shown that

(1) the language $\mathcal{L}(\mathbf{v})$ is closed under reversal,
(2) the complexity of $\mathbf{v}$ is $\mathcal{C}(n) = 2n + 1$ for all $n \in \mathbb{N}$,
(3) the language $\mathcal{L}(\mathbf{v})$ contains infinitely many distinct palindromes that do not have a unique palindromic extension.

When proven, the statements (1) and (2) imply that **the property $\mathcal{P}$ holds** (by Theorem 1.1) and the statement (3) has as a consequence that **the property $\mathcal{PE}$ does not hold**.

Proof of Step 2:
Let us start with a similar observation as Observation 4.2.

Observation 4.6. *Every factor v of $\mathbf{v}$ can be decomposed as $v = v^{(0)}v^{(1)} \dots v^{(m)}$, $m \geq 1$, so that $v^{(i)} \in \{12, 100\}$ for $i \in \{1, \dots, m-1\}$, $v^{(0)}$ is a proper suffix either of 12 or of 100 and $v^{(m)}$ is a proper prefix of 100. Obviously, for every such decomposition, there exists $\tilde{v} \in \{a, b\}^*$ such that*

$$v = v^{(0)}\Psi(\tilde{v})v^{(m)}. \tag{4.8}$$

If the decomposition is unique, the corresponding $\tilde{v}$ is necessarily a factor of $\mathbf{u}$.

The crucial tool for the proof of (1) – (3) is the map $H : \{a, b\}^* \to \{0, 1, 2\}^*$ defined by

$$H(w) = \Psi(w)1 \quad \text{for every } w \in \{a, b\}^*. \tag{4.9}$$

Its properties are stated in the following lemma.

Lemma 4.7. *Let H be the map defined in (4.9). Then it holds for every $w \in \{a, b\}^*$*

a) if w is a factor of $\mathbf{u}$, then $H(w)$ is a factor of $\mathbf{v}$,
b) if w is a palindrome, then $H(w)$ is a palindrome,
c) if w is a factor of $\mathbf{u}$, then $b(w) = b(H(w))$.

Proof. $a)$ There exists a letter $x \in \{a, b\}$ such that $wx \in \mathcal{L}(\mathbf{u})$. Then $\Psi(wx)$ is a factor of $\mathbf{v} = \Psi(\mathbf{u})$ and $\Psi(wx)$ contains $H(w) = \Psi(w)1$.

$b)$ It suffices to notice that $1^{-1}\Psi(a)1 = \overline{\Psi(a)}$ and $1^{-1}\Psi(b)1 = \overline{\Psi(b)}$, where $1^{-1}\Psi(a)1$ is the word obtained when the prefix 1 is cut from $\Psi(a)1$.

$c)$ The statement will be proven if we show that the relation between the extensions of w and $H(w)$ is as follows

$$\begin{aligned} awa \in \mathcal{L}(\mathbf{u}) \quad &\Leftrightarrow \quad 2H(w)2 \in \mathcal{L}(\mathbf{v}) \\ awb \in \mathcal{L}(\mathbf{u}) \quad &\Leftrightarrow \quad 2H(w)0 \in \mathcal{L}(\mathbf{v}) \\ bwa \in \mathcal{L}(\mathbf{u}) \quad &\Leftrightarrow \quad 0H(w)2 \in \mathcal{L}(\mathbf{v}) \\ bwb \in \mathcal{L}(\mathbf{u}) \quad &\Leftrightarrow \quad 0H(w)0 \in \mathcal{L}(\mathbf{v}) \end{aligned}$$

($\Rightarrow$): If $awb \in \mathcal{L}(\mathbf{u})$, then $\Psi(awb) = 12\Psi(w)100$ is a factor of $\mathbf{v}$ and $\Psi(awb)$ contains $2H(w)0$. The proofs for the other cases awa, bwa, bwb are similar. ($\Leftarrow$): It is easy to see that $2H(w)0 = 2\Psi(w)10$ is a unique decomposition of $2H(w)0$ of the form (4.8). Moreover, since 2 is a suffix of $\Psi(a)$, but not of $\Psi(b)$, and 10 is a prefix of $\Psi(b)$, but not of $\Psi(a)$, it follows that awb is a factor of $\mathbf{u}$. The proofs for the other cases $2H(w)2, 0H(w)2, 0H(w)0$ are analogous.

□

(1) According to its construction, the word $\mathbf{v}$ is uniformly recurrent. Using Items $a)$ and $b)$ of Lemma 4.7, it is clear that $\mathcal{L}(\mathbf{v})$ contains infinitely many distinct palindromes. Relating these two facts, $\mathcal{L}(\mathbf{v})$ is closed under reversal.

(2) In order to describe all strong and weak bispecial factors, the following lemma is helpful. However, it is useful to notice first that the only left special factors of length ≤ 2 are: $\varepsilon, 0, 1, 10, 12$. Among them, the only strong bispecial factor is 1 and the only weak bispecial factor is 0.

Lemma 4.8. *Let v be a bispecial factor of $\mathbf{v}$ of length ≥ 3. There exists a factor w of $\mathbf{u}$ such that $v = H(w)$. Moreover, $b(w) = b(H(w))$.*

Proof. Since every prefix of a bispecial factor is left special, v has to start in 1. Since the language $\mathcal{L}(\mathbf{v})$ is closed under reversal and v is right special, v has to end in 1. Then, observing the morphism Ψ, $v = \Psi(w)1$ is a unique decomposition of v the form (4.8). Thus, by Observation 4.6, w is a factor of $\mathbf{u}$ satisfying $v = H(w)$. The last statement follows by Item c) of Lemma 4.7. $\square$

By Lemma 4.8 and since 0 is the only weak and 1 the only strong bispecial factor of length ≤ 2, we obtain the set of all strong bispecial factors of $\mathbf{v}$ (recall that $V^{(n)}$ and $U^{(n)}$ are defined in (4.6) and in (4.7), respectively)

$$\{\hat{V}^{(n)} \mid n \in \mathbb{N}\}, \quad \text{where } \hat{V}^{(n)} = H(V^{(n)}),$$

and the set of all weak bispecial factors of $\mathbf{v}$

$$\{\hat{U}^{(n)} \mid n \in \mathbb{N}\}, \quad \text{where } \hat{U}^{(0)} = 0 \text{ and } \hat{U}^{(n)} = H(U^{(n)}) \text{ for } n \geq 1.$$

Since the factors $U^{(1)} = abba$ and $V^{(1)} = baab$ consist of the same "hand" of letters, it follows by the definition of $V^{(n)}$ and $U^{(n)}$ that $|V^{(n)}|_a = |U^{(n)}|_a$ and $|V^{(n)}|_b = |U^{(n)}|_b$, where $|w|_a$ denotes the number of letters a occurring in a word w. Therefore, we deduce that $|\hat{V}^{(n)}| = |\hat{U}^{(n)}|$ and by Lemma 4.8, it holds $b(\hat{V}^{(n)}) + b(\hat{U}^{(n)}) = b(V^{(n)}) + b(U^{(n)}) = 0$. By Formula (4.1), we have $\Delta^2\mathcal{C}(n) \equiv 0$, and since $\mathcal{C}(0) = 1$ and $\mathcal{C}(1) = 3$, it follows that $C(n) = 2n + 1$ for every $n \in \mathbb{N}$.

(3) The strong bispecial factors $\hat{V}^{(n)}$ are palindromes by Item b) of Lemma 4.7. Since $b(\hat{V}^{(1)}) = b(1) = 1$, we deduce using Item c) of Lemma 4.7 that $b(\hat{V}^{(n)}) = 1$ for all $n \in \mathbb{N}$. Applying Remark 4.1, the palindromes $\hat{V}^{(n)}$ do not have a unique palindromic extension. Using similar arguments, the palindromes $\hat{U}^{(n)}$ do not have a unique palindromic extension either.

4.2. Implications in Theorems 1.1 and 1.2 are irreversible

In order to show that the implications in Theorems 1.1 and 1.2 are irreversible, we will construct an infinite ternary word $\mathbf{U}$ whose language $\mathcal{L}(\mathbf{U})$ is closed under reversal and such that on one hand, $\mathbf{U}$ has the property $\mathcal{PE}$, consequently $\mathbf{U}$ has the property $\mathcal{P}$, too, on the other hand, the complexity $\mathcal{C}(n)$ of $\mathbf{U}$ does not satisfy $\mathcal{C}(n) = 2n + 1$ for all $n \in \mathbb{N}$.

Denote by $\mathbf{U}$ the infinite ternary word being the fixed point of the morphism $\Phi : \{A, B, C\}^* \to \{A, B, C\}^*$ defined by

$$\Phi(A) = ABA, \quad \Phi(B) = CAC, \quad \Phi(C) = ACA. \tag{4.10}$$

We will not provide a detailed proof of the announced properties, but only a helpful hint for the reader. Observing the substitution Φ, it is obvious that the image of a palindrome is again a palindrome. Therefore, $\mathcal{L}(\mathbf{U})$ contains infinitely many palindromes. Together with the uniform recurrence of $\mathbf{U}$, it implies that the language $\mathcal{L}(\mathbf{U})$ is closed under reversal. In addition, every palindrome p is a central factor of $\Phi^2(p)$, i.e., there exists $w \in \{A, B, C\}^*$ such that $\Phi^2(p) = wp\overline{w}$. In particular, $(\Phi^{2n}(A))$ is a sequence of palindromes with A as a central factor, $(\Phi^{2n}(B))$ is a sequence of palindromes with B as a central factor, $(\Phi^{2n}(C))$ is a sequence of palindromes with C as a central factor, and $(\Phi^{2n}(AA))$ is a sequence of palindromes of even length. It is easy to see that every palindrome is a central factor of one of the above families, thus the property $\mathcal{PE}$ holds.

Concerning the complexity, we have

$$\mathcal{L}_3(\mathbf{U}) = \{AAB, BAA, AAC, CAA, ABA, ACA, CAC, BAC, CAB\},$$

hence $\mathcal{C}(1) = 3$, $\mathcal{C}(2) = 5$, $\mathcal{C}(3) = 9$. Thus, it does not hold $\mathcal{C}(3) = 2 \cdot 3 + 1$. In fact, $\Delta\mathcal{C}(n) \neq 2$ for infinitely many $n \in \mathbb{N}$.

Acknowledgements

The authors acknowledge financial support by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic.

References

[1] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n+1$. Bull. Soc. Math. France **119** (1991) 199-215.
[2] P. Baláži, Z. Masáková, E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words. Theoret. Comput. Sci. **380** (2007) 266-275.
[3] L'. Balková, E. Pelantová, W. Steiner, Sequences with Constant Number of Return Words. Monatshefte für Mathematik, **155(3-4)** (2008) 251-263.
[4] J. Cassaigne, Complexity and special factors. Bull. Belg. Math. Soc. Simon Stevin 4 **1** (1997) 67-88.
[5] X. Droubay, G. Pirillo, Palindromes and Sturmian words. Theoret. Comput. Sci. **223** (1999) 73-85.
[6] A. Glen, J. Justin, S. Widmer, L. Q. Zamboni, Palindromic richness. Eur. J. Comb. **30** (2009) 510-531.
[7] G. A. Hedlund, M. Morse, Symbolic dynamics II - Sturmian trajectories. Amer. J. Math. **62** (1940), 1-42.
[8] G. Rote, Sequences with subword complexity $2n$. Journal of Number Theory **46** (1993) 196-213.
[9] L. Vuillon, A characterization of Sturmian words by return words. Eur. J. Comb. **22** (2001) 263-275.

Communicated by (The editor will be set by the publisher).
February 3, 2009.