

Constructions of words rich in palindromes and pseudopalindromes

Edita Pelantová¹

Štěpán Starosta²

¹ Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic

² Faculty of Information Technology, Czech Technical University in Prague, Czech Republic

received 2015-6-12, revised 2016-6-20, accepted 2016-10-18.

A narrow connection between infinite binary words rich in classical palindromes and infinite binary words rich simultaneously in palindromes and pseudopalindromes (the so-called *H*-rich words) is demonstrated. The correspondence between rich and *H*-rich words is based on the operation S acting over words over the alphabet $\{0, 1\}$ and defined by $S(u_0u_1u_2\dots) = v_1v_2v_3\dots$, where $v_i = u_{i-1} + u_i \pmod{2}$. The operation S enables us to construct a new class of rich words and a new class of *H*-rich words. Finally, the operation S is considered on the multiliteral alphabet \mathbb{Z}_m as well and applied to the generalized Thue–Morse words. As a byproduct, new binary rich and *H*-rich words are obtained by application of S on the generalized Thue–Morse words over the alphabet \mathbb{Z}_4 .

Keywords: palindrome, palindromic defect, rich words, full words, Rote words

1 Introduction

In the present paper we concentrate on construction of infinite words which are filled with palindromes or pseudopalindromes to the highest possible level, the so-called *rich* words. Before we explain the expression “the highest possible level” we recall the basic notions we work with. We understand by an *infinite word* over a finite alphabet \mathcal{A} a sequence $\mathbf{u} = (u_n)_{n \in \mathbb{N}} = u_0u_1u_2\dots$, where $u_n \in \mathcal{A}$ for each $n \in \mathbb{N}$. A *factor* of \mathbf{u} is a finite sequence $w = w_0w_1\dots w_{n-1}$ of letters from \mathcal{A} such that $w = u_iu_{i+1}\dots u_{i+n-1}$ for some $i, n \in \mathbb{N}$. The set of all factors of \mathbf{u} is the *language* of \mathbf{u} , usually denoted $\mathcal{L}(\mathbf{u})$. A finite word $w = w_0w_1\dots w_{n-1}$ is called *palindrome* if w coincides with its reversal $R(w) = w_{n-1}w_{n-2}\dots w_1w_0$.

Infinite words whose language contains infinitely many palindromes are being studied by many authors. Apart from the impulses from outside mathematics (such as Hof et al. (1995) where these words are used in a model of solid materials with finite local complexity) the main reason of the interest of mathematicians is the variety of characterizations of rich words. To specify the expression “the highest possible level” one can adopt two distinct points of view: local and global.

From the local point of view, one looks at a finite piece of the infinite word, i.e., at a factor of \mathbf{u} , and counts the number of distinct palindromes occurring in this factor. A motivation for rich word definition was an inequality due to Droubay and Pirillo, see Droubay and Pirillo (1999), which states that a finite word of length n contains at most $n + 1$ distinct palindromes (the empty word is counted as a palindrome). An infinite word is *rich*, or *full*, if every its factor of length n contains $n + 1$ distinct palindromes.

From the global point of view, one counts the palindromes of length n in the set of all factors of \mathbf{u} , i.e., in the language $\mathcal{L}(\mathbf{u})$. Let $\mathcal{C}_{\mathbf{u}}(n)$ and $\mathcal{P}_{\mathbf{u}}(n)$ denote the number of factors of length n and the number of palindromic factors of length n , respectively. As shown in Baláži et al. (2007), if $\mathcal{L}(\mathbf{u})$ is closed under reversal, then the number of palindromes in \mathbf{u} is bounded from above by the relation

$$\mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 2 \geq \mathcal{P}_{\mathbf{u}}(n+1) + \mathcal{P}_{\mathbf{u}}(n) \quad \text{for every } n \in \mathbb{N}. \quad (1)$$

In Bucci et al. (2009), Bucci, De Luca, Glen and Zamboni show that for infinite words with language closed under reversal the local and global points of view coincide. More precisely, \mathbf{u} is rich if and only if the inequality in (1) can be written as an equality for every $n \in \mathbb{N}$.

Classic examples of rich words on binary alphabets include Sturmian words, i.e., infinite words over binary alphabet with the factor complexity $\mathcal{C}_{\mathbf{u}}(n) = n + 1$ for each $n \in \mathbb{N}$. Sturmian words can be generalized to multiliteral alphabets in many ways, see for example Balková et al. (2010). Two of these generalizations, namely k -ary Arnoux–Rauzy words and words coding k -interval exchange transformation with symmetric interval permutation, are rich as well. Both mentioned classes have their language closed under reversal.

Blondin Massé, Brlek, Garon and Labbé showed in Blondin Massé et al. (2011) that rich words include complementary-symmetric Rote words. They can be defined as binary words with factor complexity $\mathcal{C}_{\mathbf{u}}(n) = 2n$ for every nonzero integer n and with language closed under the exchange of letters, see Rote (1993). This implies that the language of a complementary-symmetric Rote word is closed under two mappings acting on the set $\{0, 1\}^*$ of all finite binary words: the first is R and the second is E defined by $E(w_0 \cdots w_n) = E(w_n) \cdots E(w_0)$ for letters w_i and $E(0) = 1$ and $E(1) = 0$. Thus, the language of a complementary-symmetric Rote word is closed under all elements of a group $H = \{R, E, ER, \text{Id}\}$. The same property has the language of the famous Thue–Morse word \mathbf{t} , nevertheless, it is well-known that \mathbf{t} is not rich.

For binary words having language closed under all elements of H , we show in Pelantová and Starosta (2013) that

$$\mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 4 \geq \mathcal{P}_{\mathbf{u}}(n+1) + \mathcal{P}_{\mathbf{u}}(n) + \mathcal{P}_{\mathbf{u}}^E(n+1) + \mathcal{P}_{\mathbf{u}}^E(n) \quad \text{for every } n \geq 1, \quad (2)$$

where $\mathcal{P}_{\mathbf{u}}^E$ is the function counting E -palindromes – words fixed by E – in the word \mathbf{u} . Analogously to the case of equality in (1), we say that an infinite word with language closed under all elements of H is H -rich if in (2) the equality holds for all $n \geq 1$. We also demonstrated that the Thue–Morse word \mathbf{t} is H -rich. In Starosta (2012) the second author proved that the binary generalization $\mathbf{t}_{b,2}$ of the Thue–Morse word is H -rich for all $b \geq 2$ (the definition of $\mathbf{t}_{b,2}$ is recalled in Preliminaries). In fact, the words $\mathbf{t}_{b,2}$ are the only H -rich words that have been found up to now.

One of the main aims of the present article is to describe a procedure which produces new H -rich words. We have found an inspiration in a connection between complementary-symmetric Rote words and Sturmian words due to Rote in Rote (1993). Given an infinite word $\mathbf{u} = u_0 u_1 \dots \in \{0, 1\}^{\mathbb{N}}$, we set $S(\mathbf{u}) = v_1 v_2 \dots \in \{0, 1\}^{\mathbb{N}}$ with $v_i = (u_{i-1} + u_i) \bmod 2$ for all positive integer i . The operator S defines the mentioned relation: a word \mathbf{u} is a complementary-symmetric Rote word if and only if $S(\mathbf{u})$ is a Sturmian word.

In Section 4, we investigate binary words which are simultaneously H -rich and also rich in the classical sense. In particular, we prove that every complementary-symmetric Rote word is H -rich, see Corollary 14. The main result concerning H -richness is presented in Theorem 24. On the one hand the theorem

says that the operator S applied to an H -rich word produces a rich word. Using the examples of H -rich words mentioned earlier, we get a new class of rich words, namely the words $S(\mathbf{t}_{b,2})$ for all $b \geq 2$. On the other hand, the theorem transforms the task to discover new H -rich words to the task to discover a new class of rich words with special structures of palindromes. One such class is described in Starosta (2016). Section 5 is devoted to the notion of G -richness on a multiliteral alphabet. In particular, the operation S is defined over the alphabet \mathbb{Z}_m . Theorem 31 illustrates that even on a multiliteral alphabet the operation S connects G -richness and G' -richness for, in general, distinct groups G and G' . In this sense Theorem 31 is a weaker version of Theorem 24.

2 Preliminaries

The set \mathcal{A}^* is the set of all finite words over the *alphabet* \mathcal{A} which is a finite set of *letters*. The *length* of the word $w = w_0w_1 \cdots w_{n-1} \in \mathcal{A}^*$ with $w_i \in \mathcal{A}$ for all i is denoted $|w|$ and equals n . The *empty word* – the unique word of length 0 – is denoted ε . The set \mathcal{A}^* together with concatenation forms a free monoid with the neutral element ε . A word $v \in \mathcal{A}^*$ is a *factor* of $w \in \mathcal{A}^*$ if $w = uvz$ for some word $u, z \in \mathcal{A}^*$. If, moreover, $u = \varepsilon$, then we say that v is a *prefix* of w , if $z = \varepsilon$, the word v is a *suffix* of w . If w has the form $w = vz$, then z is denoted $z = v^{-1}w$ and the word $v^{-1}wv$ is a *conjugate* of the word w .

The *infinite word* over \mathcal{A} is a sequence $\mathbf{u} = (u_n)_{n \in \mathbb{N}} = u_0u_1u_2 \dots$. The symbol $\mathcal{A}^{\mathbb{N}}$ denotes the set of all infinite words over \mathcal{A} . A finite word $w \in \mathcal{A}^*$ of length $n = |w|$ is a *factor* of \mathbf{u} if there exists an index i such that $w = u_iu_{i+1} \cdots u_{i+n-1}$; the index i is an *occurrence* of the factor w . The symbol $\mathcal{L}_n(\mathbf{u})$ stands for the set of all factors of length n occurring in \mathbf{u} . The set of all factors of \mathbf{u} is the *language* of \mathbf{u} and is denoted by $\mathcal{L}(\mathbf{u})$.

An infinite word \mathbf{u} is *recurrent* if any factor of \mathbf{u} has at least two occurrences in \mathbf{u} . Equivalently, a word is recurrent if any factor has infinitely many occurrences. If moreover for any factor w the gaps between consecutive occurrences of w are bounded, then the word \mathbf{u} is *uniformly recurrent*. Let w and vw be factors of $\mathcal{L}(\mathbf{u})$ such that vw has a prefix w and w occurs in vw exactly twice. The word v is a *return word* of w and vw is a *complete return word* of w . One can say equivalently: a recurrent word \mathbf{u} is uniformly recurrent if any factor $w \in \mathcal{L}(\mathbf{u})$ has finite number of return words of w .

The *factor complexity* of \mathbf{u} is the mapping $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \rightarrow \mathbb{N}$, defined by $\mathcal{C}_{\mathbf{u}}(n) = \#\mathcal{L}_n(\mathbf{u})$. Given $a \in \mathcal{A}$ and $w \in \mathcal{A}^*$, a factor $wa \in \mathcal{L}(\mathbf{u})$ is a *right extension* of the factor w . Any factor of \mathbf{u} has at least one right extension, the set of all right extensions of w is denoted $\text{Rext}(w)$. If w has at least two right extensions we call it *right special*. Analogously one can define *left extension* and *left special* and $\text{Lext}(w)$. In a recurrent word \mathbf{u} any factor has at least one left extension. A factor w which is left and right special is *bispecial*. Special factors can be used to determine the factor complexity, in particular

$$\Delta \mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} (\#\text{Rext}(w) - 1).$$

If \mathcal{A} is a binary alphabet, we get

$$\Delta \mathcal{C}_{\mathbf{u}}(n) = \#\{w \in \mathcal{L}_n(\mathbf{u}) : w \text{ is right special}\}. \quad (3)$$

A mapping $\mu : \mathcal{A}^* \rightarrow \mathcal{B}^*$ is a *morphism* if $\mu(wv) = \mu(w)\mu(v)$ for all $w, v \in \mathcal{A}^*$. It is an *antimorphism* if $\mu(wv) = \mu(v)\mu(w)$ for all $w, v \in \mathcal{A}^*$. An infinite word \mathbf{u} is *closed under the mapping* μ if $\mu(w) \in \mathcal{L}(\mathbf{u})$ for any factor $w \in \mathcal{L}(\mathbf{u})$. Domain of a morphism $\varphi : \mathcal{A}^* \rightarrow \mathcal{A}^*$ can be naturally extended to $\mathcal{A}^{\mathbb{N}}$

by the prescription $\varphi(\mathbf{u}) = \varphi(u_0u_1u_2\dots) = \varphi(u_0)\varphi(u_1)\varphi(u_2)\dots$. An infinite word $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ is called *fixed point* of a morphism φ if $\varphi(\mathbf{u}) = \mathbf{u}$.

An antimorphism Ψ is *involutory* if $\Psi^2 = \text{Id}$. The most frequent involutory antimorphism is the reversal mapping R . If the word \mathbf{u} is closed under an involutory antimorphism, then \mathbf{u} is necessarily recurrent.

If $p = \Psi(p)$, the word p is a Ψ -*palindrome* or *pseudopalindrome*, if specification of the mapping Ψ is not needed. In the case $\Psi = R$, we say only *palindrome* instead of R -palindrome. The set of all Ψ -palindromes occurring as factors of a finite word w is denoted $\text{Pal}^{\Psi}(w)$. The Ψ -*palindromic complexity* of an infinite word \mathbf{u} is the mapping $P_{\mathbf{u}}^{\Psi} : \mathbb{N} \rightarrow \mathbb{N}$, defined by $P_{\mathbf{u}}^{\Psi}(n) = \#\{p \in \mathcal{L}_n(\mathbf{u}) : p = \Psi(p)\}$.

A Ψ -palindrome w is *centered at* $x \in \mathcal{A} \cup \{\varepsilon\}$ if $w = vx\Psi(v)$ for some word v . If a Ψ -palindrome is centered at ε , then it is of even length.

3 G -defect and G -richness

First, we recall the definition of palindromic defect as it was introduced by Brlek, Hamel, Nivat and Reutenauer in Brlek et al. (2004). This classical definition is based on the inequality

$$\#\text{Pal}^R(w) \leq |w| + 1 \quad \text{for all } w \in \mathcal{A}^*, \quad (4)$$

where $\text{Pal}^R(w)$ is the set of all R -palindromic factors of w including the empty word.

The R -*defect* of a finite word w is

$$D^R(w) = |w| + 1 - \#\text{Pal}^R(w),$$

and R -defect of an infinite word \mathbf{u} is

$$D^R(\mathbf{u}) = \sup\{D^R(w) : w \in \mathcal{L}(\mathbf{u})\}.$$

We prefer to use the name R -defect instead of the originally used “defect” because we will introduce an analogous notion for a general antimorphism Ψ as well. An infinite word \mathbf{u} with $D^R(\mathbf{u}) = 0$ is called R -*full* or R -*rich*. If $D^R(\mathbf{u})$ is finite, we say that \mathbf{u} is *almost R -rich*. In Brlek and Reutenauer (2011), the inequality (1) is used to introduce the value

$$T_{\mathbf{u}}(n) = \Delta C_{\mathbf{u}}(n) + 2 - \mathcal{P}_{\mathbf{u}}^R(n+1) - \mathcal{P}_{\mathbf{u}}^R(n) \quad \text{for every } n \in \mathbb{N}$$

and they conjectured that if \mathbf{u} is closed under reversal, then

$$2D^R(\mathbf{u}) = \sum_{n=1}^{\infty} T_{\mathbf{u}}(n). \quad (5)$$

Their conjecture was proven in Balková et al. (2013). In particular, it means that $D^R(\mathbf{u})$ is finite if and only if there exists $N \in \mathbb{N}$ such that $T_{\mathbf{u}}(n) = 0$ for all $n \geq N$, or in other words in (1) the equality holds for all $n \geq N$.

To prove R -richness we will use the characterization of R -rich words given in Balková et al. (2009). It exploits the notion of the bilateral order $\text{b}(w)$ of a factor w and the palindromic extension of a palindrome. The bilateral order was introduced in Cassaigne (1997) as

$$\text{b}(w) = \#\{awb \in \mathcal{L}(\mathbf{u}) : a, b \in \mathcal{A}\} - \#\text{Rext}(w) - \#\text{Lext}(w) + 1. \quad (6)$$

The set of all palindromic extensions of a palindrome $w \in \mathcal{L}(\mathbf{u})$ is defined by

$$\text{Pext}(w) = \{awa : awa \in \mathcal{L}(\mathbf{u}), a \in \mathcal{A}\}.$$

Theorem 1 (Balková et al. (2010)). *Let \mathbf{u} be an infinite word closed under reversal.*

1. *The word \mathbf{u} is R -rich if and only if any bispecial factor w of \mathbf{u} satisfies:*

$$b(w) = \begin{cases} \#\text{Pext}(w) - 1 & \text{if } w \text{ is a palindrome;} \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

2. *If the word \mathbf{u} is almost R -rich, then (7) is satisfied for all bispecial factors w up to finitely many exceptions.*

The first attempt to study the number of Ψ -palindromes for an involutory antimorphism Ψ was made in Blondin Massé et al. (2008). Blondin Massé, Brlek, Garon and Labbé considered the binary alphabet $\{0, 1\}$ and the antimorphism E . They showed that

$$\#\text{Pal}^E(w) \leq |w| \quad \text{for all } w \in \mathcal{A}^* \setminus \{\varepsilon\}. \quad (8)$$

In Starosta (2011), this results is generalized for an arbitrary involutory antimorphism Ψ and arbitrary alphabet into the inequality

$$\#\text{Pal}^\Psi(w) \leq |w| + 1 - \gamma_\Psi(w) \quad \text{for all } w \in \mathcal{A}^*, \quad (9)$$

where $\gamma_\Psi(w) = \#\{\{a, \Psi(a)\} : a \in \mathcal{A}, a \text{ occurs in } w \text{ and } \Psi(a) \neq a\}$. Clearly, if $\Psi = E$ we have (8) as $\gamma_E(w) = 1$ for any $w \neq \varepsilon$, if $\Psi = R$ we have (4) as $\gamma_R(w) = 0$ for any w . Based on the inequality (9), the Ψ -defect of $w \in \mathcal{A}^*$ is defined by

$$D^\Psi(w) = |w| + 1 - \gamma_\Psi(w) - \#\text{Pal}^\Psi(w). \quad (10)$$

The Ψ -defect of an infinite word \mathbf{u} is defined analogously, i.e., $D^\Psi(\mathbf{u}) = \sup\{D^\Psi(w) : w \in \mathcal{L}(\mathbf{u})\}$.

Infinite words having finite Ψ -defect can be characterized by several properties, for more details about R -defect see Balková et al. (2011) and about Ψ -defect see Starosta (2011); Pelantová and Starosta (2012). In Pelantová and Starosta (2012) we showed that there exists a very narrow connection between words with finite defect and words with zero defect. We proved that if \mathbf{u} is closed under an involutory antimorphism Ψ and $D^\Psi(\mathbf{u})$ is finite, then \mathbf{u} is a morphic image of a word \mathbf{v} with $D^\Phi(\mathbf{v}) = 0$ for some involutory antimorphism Φ . If moreover \mathbf{u} is uniformly recurrent, then $\Phi = R$. In this sense, considering Ψ instead of R does not bring a broader variability into the concept of rich words.

The situation changes when we consider more antimorphisms. In Pelantová and Starosta (2013) we defined a generalization of the notion of defect. In what follows, the symbol G stands for a finite group consisting of morphisms and antimorphisms over \mathcal{A}^* and containing at least one antimorphism. The *orbit* of $w \in \mathcal{A}^*$ is the set

$$[w] = \{\mu(w) : \mu \in G\}. \quad (11)$$

We say that \mathbf{u} is closed under G if $[w] \subset \mathcal{L}(\mathbf{u})$ for any $w \in \mathcal{L}(\mathbf{u})$. Word $p \in \mathcal{A}^*$ is a G -palindrome if $p = \Psi(p)$ for some antimorphism $\Psi \in G$. The generalization of the set of all palindromic factors of a word is a set consisting of palindromic orbits, namely the set

$$\text{Pal}^G(w) = \{[p] : p \text{ occurs in } w \text{ and } p \text{ is a } G\text{-palindrome}\}.$$

Note that if $G = \{\text{Id}, \Psi\}$ where Ψ is an involutory antimorphism, then $\text{Pal}^\Psi(w)$ is in one-to-one correspondence with the set $\text{Pal}^G(w)$ (the only difference is that the latter is a set of orbits instead of factors). Let us stress that in $\text{Pal}^G(w)$ we count how many different orbits have a G -palindromic representative occurring in w .

Definition 2. Let w be a finite word. The G -defect of w is defined as

$$D^G(w) = |w| + 1 - \#\text{Pal}^G(w) - \gamma_G(w),$$

where

$$\gamma_G(w) = \#\{[a] : a \in \mathcal{A}, a \text{ occurs in } w, \text{ and } a \neq \Psi(a) \text{ for every antimorphism } \Psi \in G\}.$$

A finite word is G -rich if its G -defect is 0. An infinite word is G -rich if all its factors are G -rich. In Pelantová and Starosta (2013), a distinct and equivalent definition of G -richness is used: it is based on a specific structure of graphs representing the factors of same length of the word.

Example 3. We illustrate the previous notions on the Thue–Morse word \mathbf{t} , the fixed point of the morphism $0 \mapsto 01$ and $1 \mapsto 10$ starting with 0, i.e., $\mathbf{t} = 011010011001011010 \dots$. The word \mathbf{t} is closed under R and E . Let $H = \{\text{Id}, R, E, ER\}$. For the group H the value $\gamma_H(w) = 0$ for any $w \in \mathcal{A}^*$. Consider $w = 011010011001$, the prefix of \mathbf{t} of length 12. We have

$$\begin{aligned} \text{Pal}^R(w) &= \{\varepsilon, 0, 1, 11, 00, 101, 010, 0110, 1001, 001100, 10011001\}, \\ \text{Pal}^E(w) &= \{\varepsilon, 01, 10, 0011, 1100, 1010, 110100, 001100, 01101001\}, \\ \text{Pal}^H(w) &= \{\varepsilon, [0], [00], [01], [010], [0110], [0011], [1010], [110100], \\ &\quad [100110], [001100], [10011001], [01101001]\}. \end{aligned}$$

The corresponding defects of w are

$$\begin{aligned} D^R(w) &= |w| + 1 - \#\text{Pal}^R(w) = 2, \\ D^E(w) &= |w| - \#\text{Pal}^E(w) = 3, \\ D^H(w) &= |w| + 1 - \#\text{Pal}^H(w) = 0. \end{aligned}$$

In fact, the Thue–Morse word is H -rich, whereas its R -defect and E -defect are both infinite, see Example 8 later.

For G -richness, theorems analogous to the theorems for the classical richness can be stated, c.f. Pelantová and Starosta (2014). The list of known G -rich words with G having at least two antimorphisms is modest. It contains the generalized Thue–Morse words $\mathbf{t}_{b,m}$. The word $\mathbf{t}_{b,m}$ is defined on the alphabet $\{0, \dots, m-1\}$ for all $b \geq 2$ and $m \geq 2$ as

$$\mathbf{t}_{b,m} = (s_b(n) \bmod m)_{n=0}^{+\infty},$$

where $s_b(n)$ denotes the sum of digits in the base- b representation of the integer n . See for instance Allouche and Shallit (2000); Cusick and Ciungu (2011) where this class of words is studied. The language of $\mathfrak{t}_{b,m}$ is closed under a group isomorphic to the dihedral group of order $2m$, here denoted $I_2(m)$. In Section 5, we describe the group in details. In Starosta (2012), the second author proved that $\mathfrak{t}_{b,m}$ is $I_2(m)$ -rich for any parameters $b \geq 2$ and $m \geq 2$.

In Corollary 14 we add to the list of H -rich words also complementary-symmetric Rote words. As already mentioned in Introduction, an infinite binary word \mathbf{u} is a *complementary-symmetric Rote word* if its factor complexity satisfies $C_{\mathbf{u}}(n) = 2n$ for all $n \geq 1$ and its language is closed under the exchange of the two letters E .

In this article, we focus on groups G acting on \mathcal{A}^* for which the implication

$$\Psi_1(a) = \Psi_2(a) \implies \Psi_1 = \Psi_2$$

is true for any letter $a \in \mathcal{A}$ and any pair of antimorphisms $\Psi_1, \Psi_2 \in G$. In Pelantová and Starosta (2013), for such a group, the number 1 is called G -distinguishing, since the image of a single letter by an antimorphism from G allows to identify the antimorphism. For example, the number 1 is H -distinguishing for the group H used in Example 3. Also for the dihedral groups $I_2(m)$ studied in Section 5, the number 1 is $I_2(m)$ -distinguishing.

If an infinite word \mathbf{u} is closed under a group G and 1 is G -distinguishing, then

$$\Delta C_{\mathbf{u}}(n) + \#G \geq \sum_{\Psi \in G^{(2)}} \left(\mathcal{P}_{\mathbf{u}}^{\Psi}(n) + \mathcal{P}_{\mathbf{u}}^{\Psi}(n+1) \right) \quad \text{for all } n \in \mathbb{N}, n \geq 1, \quad (12)$$

where $G^{(2)}$ denotes the set of all involutory antimorphisms from G , see Pelantová and Starosta (2013). Clearly, if G is generated by one antimorphism, say Ψ , then $\#G = 2$ and $G^{(2)} = \{\Psi\}$. The inequality (1) is the special case of (12). Similarly, the inequality (2) can be obtained from (12) if we put $G = H = \{\text{Id}, R, E, ER\}$. The following G -analogue of the result obtained by Bucci, De Luca, Glen and Zamboni in Bucci et al. (2009) for the classical richness is proved in Pelantová and Starosta (2014).

Theorem 4. *Let an infinite word \mathbf{u} be closed under a group G such that the number 1 is G -distinguishing. The G -defect $D^G(\mathbf{u})$ is zero if and only if in (12) the equality holds for each $n \in \mathbb{N}, n \geq 1$.*

In Pelantová and Starosta (2013) we also introduced the notion almost G -rich word. A word \mathbf{u} closed under a group G is *almost G -rich* if there exists $N \in \mathbb{N}$ such that the equality in (12) takes place for all integers $n \geq N$. An infinite word \mathbf{u} is almost G -rich if and only if its G -defect

$$D^G(\mathbf{u}) = \sup\{D^G(w) : w \in \mathcal{L}(\mathbf{u})\}$$

is finite.

Remark 5. In fact, in Pelantová and Starosta (2014) the last statement is shown only for uniformly recurrent words. However, one can use the same argument we applied in proof of Theorem 2 in Balková et al. (2013) and show that $D^G(\mathbf{u})$ is finite if and only if in (12) the equality takes place from some N on.

4 Binary words invariant under two involutory antimorphisms

4.1 G -richness in binary alphabet

In this section we suppose $\mathcal{A} = \{0, 1\}$. On binary alphabet we have only two antimorphisms R and E . Therefore, only the groups

$$\{\text{Id}, R\}, \quad \{\text{Id}, E\}, \quad \text{and} \quad H = \{\text{Id}, R, E, ER\},$$

can be considered when inspecting the defect D^G . Let us start with examples of G -rich and almost G -rich words for these three groups.

Example 6. ($G = \{\text{Id}, R\}$)

The classical richness has been studied very intensively and thus there are known many examples of binary R -rich words including Sturmian words, see Droubay et al. (2001), Rote Words, see Blondin Massé et al. (2011), the period doubling word, see Balková (2008), etc. Plenty examples of binary almost R -rich words can be constructed by application of special standard P -morphisms to any rich word, see Glen et al. (2009) for the definition of standard P -morphism and a proof.

Example 7. ($G = \{\text{Id}, E\}$)

It can be easily seen, or shown using the results of Blondin Massé et al. (2008), that there exist only two E -rich infinite words, namely the periodic word $\mathbf{u} = (01)^\omega$ and its shift $(10)^\omega$. The two mentioned words are also R -rich and H -rich as the equalities hold in (1) and (2) for all $n \in \mathbb{N}, n \geq 1$.

Examples of infinite words with finite E -defect are E -standard words with seed (see Bucci et al. (2008) for their definition and Pelantová and Starosta (2012) for a proof). This class also includes very simple examples of words with finite E -defect: periodic words having the form w^ω with $w = E(w)$. One can easily show that in this case $D^E(w^\omega) = D^E(w^2)$ (see Corollary 8 in Brlek et al. (2004) for R -defect, a modification for E is straightforward).

Example 8. ($G = H = \{\text{Id}, R, E, ER\}$)

The only so far known examples of H -rich words are given in Starosta (2012): they are the generalized Thue–Morse words $\mathbf{t}_{b,2}$.

If b is odd, then $\mathbf{t}_{b,2} = (01)^\omega$ and hence $\mathbf{t}_{b,2}$ is also R -rich and E -rich.

If b is even, the word is aperiodic and $D^R(\mathbf{t}_{b,2}) = D^E(\mathbf{t}_{b,2}) = +\infty$. To prove it for any even b we use the fact that $\mathbf{t}_{b,2}$ is a fixed point of the morphism φ determined by

$$\varphi: \quad 0 \mapsto (01)^{\frac{b}{2}} \quad \text{and} \quad 1 \mapsto (10)^{\frac{b}{2}}.$$

It is readily seen that the factor $w = (01)^{\frac{b}{2}}$ is *strong*, i.e., its bilateral order $b(w)$ is positive, specifically $b(w) = 1$, as all four words $0w1$, $0w0$, $1w1$, and $1w0$ belong to $\mathcal{L}(\mathbf{u})$. Moreover w is an E -palindrome. The form of the morphism ensures that

- $b(\varphi(v)) = 1$ for any strong factor $v \neq \varepsilon$;
- if v is an R -palindrome, then $\varphi(v)$ is an E -palindrome,
- if v is an E -palindrome, then $\varphi(v)$ is an R -palindrome,

These properties imply that for any $k \in \mathbb{N}$, the factor $\varphi^{2k}(w)$ is an E -palindrome and hence it is not an R -palindrome. Thus there exist infinitely many non-palindromic bispecial factors with non-zero bilateral order. Using Theorem 1 one may see that $D^R(\mathfrak{t}_{b,2}) = +\infty$.

To prove that $D^E(\mathfrak{t}_{b,2}) = +\infty$ we may proceed analogously. The factors $\varphi^{2k+1}(w)$ are R -palindromes but they are not E -palindromes for all $k > 0$. These factors are bispecial with the same bilateral order 1. A modification of Theorem 1 for the antimorphism E (which can be found in full generality in Pelantová and Starosta (2014), Proposition 45) gives the result.

Now we look at the question whether a word can be simultaneously (almost) G -rich for two groups on the binary alphabet. We will discuss the connection between finiteness of defects D^R , D^E and D^H . In what follows we will consider words invariant under R and E simultaneously. First we study the relationship between R - and E -palindromes.

Lemma 9. *Let $p, q \in \mathcal{A}^*$ be R -palindromes such that the word pq is an E -palindrome, i.e.,*

$$pq = E(q)E(p). \quad (13)$$

There exist $c \in \mathcal{A}^$ and $i, j \in \mathbb{N}$ such that $p = c(E(c)c)^i$ and $q = (E(c)c)^j E(c)$.*

Proof: We will induce on the difference of $|p|$ and $|q|$. First, suppose that $|p| = |q|$, then (13) implies that $q = E(p)$ and it suffices to set $c = p$ and $i = j = 0$.

Suppose now that $|p| \neq |q|$. We can suppose without loss of generality that $|p| < |q|$. Set $q = q_1q_2$ with $|p| = |q_2|$. It follows from (13) that $pq_1q_2 = E(q_2)E(q_1)E(p)$, thus $p = E(q_2)$ and $q_1 = E(q_1)$. Therefore, q_2 is a palindrome. Since q is a palindrome, we have $R(q_1q_2) = R(q_2)R(q_1) = q_1q_2 = q_2R(q_1)$. We get

$$q_1q_2 = q_2R(q_1). \quad (14)$$

This equation on words, written in general as $xz = zy$, has a well-known solution: there exist words $u, v \in \mathcal{A}^*$ and $k \in \mathbb{N}$ such that $x = uv, y = vu$ and $z = (uv)^k u$. If the word z is palindrome, then the form of z implies that u and v are palindromes as well. To use the solution of $xz = zx$ to solve (14), we set $z = q_2, x = q_1$ and $y = R(q_1)$ and we get the solutions $q_1 = uv = E(uv)$ and $q_2 = (uv)^k u$. Since $|q_1| = |q| - |p| = |u| + |v|$, it follows that the difference of $|u|$ and $|v|$ is less than $|q_1| = |q| - |p|$. We apply the induction hypothesis on the palindromes u and v satisfying $E(uv) = uv$ and we get that $u = d(E(d)d)^m$ and $v = (E(d)d)^n E(d)$ for some $d \in \mathcal{A}^*$. Substituting for p and q one can find that it suffices to set $c = E(d)$ and the claim is proved. \square

Corollary 10. *If p and q are palindromes such that $pq = E(pq)$, then there exists $c \in \mathcal{A}^*$ such that $pq = (cE(c))^j$ for some $j \in \mathbb{N}$.*

Proposition 11. *If an infinite recurrent word \mathbf{u} has finite R -defect and finite E -defect, then \mathbf{u} is periodic with a period conjugate to $rE(r)$, where r is an R -palindrome.*

Proof: Let \mathbf{u} be an infinite recurrent word with finite R - and E -defects. Using Proposition 5 in Pelantová and Starosta (2012), it follows that \mathbf{u} is closed under R and E and there exists an integer h such that

$$\begin{aligned} \Delta \mathcal{C}_{\mathbf{u}}(n) + 2 &= \mathcal{P}^R(n+1) + \mathcal{P}^R(n) && \text{and} \\ \Delta \mathcal{C}_{\mathbf{u}}(n) + 2 &= \mathcal{P}^E(n+1) + \mathcal{P}^E(n) \end{aligned}$$

for all $n \geq h$. Since \mathbf{u} is also closed under all elements of the group H , combining the two previous equalities with (2) we get $0 \geq \Delta\mathcal{C}_{\mathbf{u}}(n)$ for all $n \geq h$, i.e., the word \mathbf{u} is eventually periodic. Since \mathbf{u} is recurrent and closed under R , the word \mathbf{u} is purely periodic, i.e., $\mathbf{u} = w^\omega$. As \mathbf{u} is closed under E , the word $E(w)$ is a factor of ww . It implies that $w = w_1w_2$ with $E(w_1) = w_1$ and $E(w_2) = w_2$. As the length of any E -palindrome is even, the concatenation of two E -palindromes is conjugate to an E -palindrome, in other words, the word w is conjugate to an E -palindrome, say v . Thus $\mathbf{u} = w^\omega = v^\omega$ for some v . As v^ω has language closed under R as well, by the same reasoning we have $v = pq$, where $R(p) = p$ and $R(q) = q$. Applying Corollary 10 we get $v = pq = (cE(c))^j$ for some $j \in \mathbb{N}$. It is enough to set $r = c$. \square

The following proposition treats another combination of two G -defects.

Proposition 12. *Let $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$ be a word having its language closed under the group H and let $\Psi = R$ or $\Psi = E$. If $D^\Psi(\mathbf{u})$ is finite (resp. zero), then $D^H(\mathbf{u})$ is finite (resp. zero) as well.*

Before giving a proof of the last proposition, we recall Proposition 4.3 of Balková et al. (2011) which will be needed.

Proposition 13. *Let \mathbf{u} be an infinite word with language closed under reversal. Suppose that there exists an integer N such that for all $n \geq N$ the equality $\mathcal{P}_{\mathbf{u}}^R(n) + \mathcal{P}_{\mathbf{u}}^R(n+1) = \mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 2$ holds. The complete return words of any palindromic factor of length $n \geq N$ are palindromes.*

Proof of Proposition 12: Let us realize that closedness of \mathbf{u} under R and E ensures that the numbers $\mathcal{P}_{\mathbf{u}}^E(n)$ and $\mathcal{P}_{\mathbf{u}}^R(n)$ are even. Indeed, if $w \in \mathcal{L}(\mathbf{u})$ is an E -palindrome of length n , then $R(w)$ is an E -palindrome as well, and analogously for R -palindromes.

First we consider $\Psi = R$. Let us suppose that there exists a positive integer N such that

$$\begin{aligned} \Delta\mathcal{C}_{\mathbf{u}}(n) + 2 &= \mathcal{P}_{\mathbf{u}}^R(n) + \mathcal{P}_{\mathbf{u}}^R(n+1) && \text{for all } n \geq N \text{ and} \\ \Delta\mathcal{C}_{\mathbf{u}}(N) + 4 &> \mathcal{P}_{\mathbf{u}}^R(N) + \mathcal{P}_{\mathbf{u}}^R(N+1) + \mathcal{P}_{\mathbf{u}}^E(N) + \mathcal{P}_{\mathbf{u}}^E(N+1). \end{aligned}$$

We will show that this assumption leads to a contradiction.

In particular the assumption yields the inequality $2 > \mathcal{P}_{\mathbf{u}}^E(N) + \mathcal{P}_{\mathbf{u}}^E(N+1)$, which implies that there is no E -palindrome of length at least N . Let $w \in \mathcal{L}(\mathbf{u})$ be an R -palindrome of length at least N . We say that a factor f has Property π if it satisfies all of the following:

- 1) w occurs in f exactly once,
- 2) $E(w)$ occurs in f exactly once,
- 3) w is a suffix or a prefix of f ,
- 4) $E(w)$ is a suffix or a prefix of f .

Let u be a factor with Property π . Such factor must exist as $\mathcal{L}(\mathbf{u})$ is closed under E and thus $E(w) \in \mathcal{L}(\mathbf{u})$ as well. As w is an R -palindrome and \mathbf{u} is closed under reversal, the factor $R(u)$ has Property π as well. Since $ER(w) = E(w)$, we can assume without loss of generality that u is the factor starting in w and ending in $E(w)$. Let us look at the complete return word of w , say p , with prefix u . The fact that the equality $\Delta\mathcal{C}_{\mathbf{u}}(n) + 2 = \mathcal{P}_{\mathbf{u}}^R(n) + \mathcal{P}_{\mathbf{u}}^R(n+1)$ is valid for all $n \geq N$ implies according to Proposition 13

that the complete return word p of w is an R -palindrome. Thus the factor $R(u)$ is a suffix of p . Moreover p contains only two factors (namely u and $R(u)$) with Property π .

We have shown for every factor u' with Property π that its closest right neighbor in \mathbf{u} with Property π is its mirror image $R(u')$. Therefore, there exist only two factors with Property π , namely u and $R(u)$.

On the other hand, if u has Property π , then $E(u)$ has Property π as well and thus $E(u) \in \{u, R(u)\}$. As $E(w)$ is a suffix of u , the factor $E(u)$ has a prefix w . It implies that $E(u) = u$ which contradicts the fact that there is no E -palindrome longer than $|w|$.

We have shown that

$$\Delta C_{\mathbf{u}}(n) + 2 = \mathcal{P}_{\mathbf{u}}^R(n) + \mathcal{P}_{\mathbf{u}}^R(n+1) \quad \text{for all } n \geq N$$

implies

$$\Delta C_{\mathbf{u}}(n) + 4 = \mathcal{P}_{\mathbf{u}}^R(n) + \mathcal{P}_{\mathbf{u}}^R(n+1) + \mathcal{P}_{\mathbf{u}}^E(n) + \mathcal{P}_{\mathbf{u}}^E(n+1) \quad \text{for all } n \geq N.$$

If \mathbf{u} is R -rich, then $N = 1$ and thus \mathbf{u} is also H -rich. If its defect $D(\mathbf{u})$ is finite but nonzero, then $N > 1$ and \mathbf{u} has finite H -defect.

In the case $\Psi = E$ the proof is analogous. \square

Corollary 14. *Every complementary-symmetric Rote word is H -rich.*

Proof: In Blondin Massé et al. (2011), it is proved that Rote words are R -rich. Since a complementary-symmetric Rote word is closed under H , the previous theorem proves the statement. \square

Remark 15. Let us stress that the reverse implication in Proposition 12 does not hold. As shown in Example 8, the Thue–Morse word has $D^H(\mathbf{t}) = 0$, whereas $D^R(\mathbf{t}) = D^E(\mathbf{t}) = \infty$.

According to Proposition 11, the finiteness of both defects $D^E(\mathbf{u})$ and $D^R(\mathbf{u})$ forces the word \mathbf{u} to be periodic. The Rote words illustrate that there exist aperiodic words with finite $D^H(\mathbf{u})$ and $D^R(\mathbf{u})$.

4.2 The mapping S on binary words

In this section we introduce and study the basic properties of the mapping $S : \mathcal{A}^* \setminus \{\varepsilon\} \rightarrow \mathcal{A}^*$ that is given by

$$S(u_0 \cdots u_n) = v_1 \cdots v_n, \quad \text{where } v_i = (u_{i-1} + u_i) \bmod 2 \quad \text{for } i = 1, \dots, n.$$

In particular, $S(a) = \varepsilon$ for every $a \in \mathcal{A}$. The following list contains some elementary properties of S .

- I. $SR = RS$, and $SE = SR$.
- II. $S(w) = S(u)$ if and only if $w = u$ or $w = ER(u)$.
- III. $S(w)$ is an R -palindrome if and only if w is an R -palindrome or an E -palindrome.

Proof: Points I and II give

$$S(w) = R(S(w)) \iff S(w) = S(R(w)) \iff w = R(w) \text{ or } w = ER(R(w)) = E(w).$$

\square

The operation S is naturally extended to $\mathcal{A}^{\mathbb{N}}$ by setting

$$S(u_0u_1u_2\dots) = v_1v_2\dots, \quad \text{where } v_i = (u_{i-1} + u_i) \bmod 2 \quad \text{for } i \geq 1.$$

To describe the factor complexity of $S(\mathbf{u})$ we study special factors in

$$\mathcal{L}(S(\mathbf{u})) = \{S(v) : v \in \mathcal{L}(\mathbf{u})\}.$$

Lemma 16. *Let $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$. A factor $S(v)$ is right special in $\mathcal{L}(S(\mathbf{u}))$ if and only if one of the following occurs:*

- a) v or $ER(v)$ is right special in $\mathcal{L}(\mathbf{u})$,
- b) $\{v, ER(v)\} \subset \mathcal{L}(\mathbf{u})$, and $\{va, ER(va)\} \not\subset \mathcal{L}(\mathbf{u})$ for both $a \in \{0, 1\}$.

Proof: Let $S(v)$ be right special in $\mathcal{L}(S(\mathbf{u}))$. Then $S(v0)$ and $S(v1)$ belong to $\mathcal{L}(S(\mathbf{u}))$. It may happen that either both $v0$ and $v1$ belong to $\mathcal{L}(\mathbf{u})$, which means that v is right special in $\mathcal{L}(\mathbf{u})$, or both $ER(v0)$ and $ER(v1)$ belong to $\mathcal{L}(\mathbf{u})$, which means that $ER(v)$ is right special in $\mathcal{L}(\mathbf{u})$.

Otherwise v and $ER(v)$ are not right special in $\mathcal{L}(\mathbf{u})$, but necessarily both belong to $\mathcal{L}(\mathbf{u})$. Let va and $ER(v)b$ be the unique right prolongations in $\mathcal{L}(\mathbf{u})$ of v and $ER(v)$ respectively. Since $S(va)$ and $S(ER(v)b)$ must be distinct right prolongations of $S(v) = S(ER(v))$, we have $a \neq ER(b)$, i.e., $a = b$. Since $ER(v)$ has a unique extension to the right $ER(v)a$, we get $ER(v)(1-a) = ER(va) \notin \mathcal{L}(\mathbf{u})$. \square

Lemma 17. *Let $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$. The word \mathbf{u} is uniformly recurrent if and only if $S(\mathbf{u})$ is uniformly recurrent.*

Proof: (\Rightarrow): Let w be a factor of $S(\mathbf{u})$. Then $w = S(v)$ for some $v \in \mathcal{L}(\mathbf{u})$. The gaps between the neighboring occurrences of v in \mathbf{u} are bounded by some constant. The gaps between the occurrences of w in $S(\mathbf{u})$ are bounded by the same constant.

(\Leftarrow): Let v be a factor of \mathbf{u} . Then $w = S(v)$ is a factor of $S(\mathbf{u})$ and the gaps between the occurrences of w are bounded, say by K . If v is the only factor of \mathbf{u} such that $w = S(v)$, i.e., v is the only preimage of w by S in \mathbf{u} , then the occurrences of v in \mathbf{u} are bounded by K as well. Let us suppose that w has more preimages in \mathbf{u} . According to Property II, there are only two preimages of w , namely v and $ER(v)$. Let f be a factor of \mathbf{u} such that v is a prefix of f and $ER(v)$ is a suffix of f and v and $ER(v)$ occur in f only once. Then $S(f)$ is a complete return word of $w = S(v) = S(ER(v))$. As $S(\mathbf{u})$ is uniformly recurrent, the gaps between the occurrences of the factor $S(f)$ are bounded, say by C . Both possible preimages of $S(f)$ in \mathbf{u} , namely f and $ER(f)$, contain v either as its prefix or its suffix. Thus the gaps between the occurrences of v in \mathbf{u} are bounded by C as well. \square

Lemma 18. *Let $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$. If $S(\mathbf{u})$ is closed under R , then \mathbf{u} is closed under R or under E .*

Proof: Let v be a prefix of \mathbf{u} . The word $S(v)$ is a factor of $S(\mathbf{u})$. According to the assumption, $RS(v) = SR(v)$ belongs to $\mathcal{L}(S(\mathbf{u}))$ as well. Due to Property II, either $R(v)$ or $E(RR(v)) = E(v)$ belong to $\mathcal{L}(\mathbf{u})$. Thus

- a) either there exist infinitely many prefixes $v \in \mathcal{L}(\mathbf{u})$ such that $R(v) \in \mathcal{L}(\mathbf{u})$;
- b) or there exist infinitely many prefixes $v \in \mathcal{L}(\mathbf{u})$ such that $E(v) \in \mathcal{L}(\mathbf{u})$.

Let us suppose that a) happens. For any $w \in \mathcal{L}(\mathbf{u})$ we may find a prefix v such that $R(v) \in \mathcal{L}(\mathbf{u})$ and w is a factor of v . Thus, $R(w) \in \mathcal{L}(\mathbf{u})$ and we can conclude that \mathbf{u} is closed under R .

The case b) is analogous. □

Example 19. The period doubling word is the fixed point of the primitive morphism

$$\varphi_{PD} : 0 \mapsto 11 \quad \text{and} \quad 1 \mapsto 10.$$

Thus

$$\mathbf{u}_{PD} = 10111010101110111011101010\dots$$

It is well-known that the period doubling word is the image of the Thue–Morse word \mathbf{t} by S .

The word $\mathbf{u}_{PD} = S(\mathbf{t})$ is closed under R , the word \mathbf{t} is closed under R and E . It illustrates that in the previous lemma the simultaneous closedness under R and E is not excluded.

The previous lemma guarantees that \mathbf{u} is closed at least under one of the antimorphisms E and R . We now focus on a property of $S(\mathbf{u})$ that ensures that \mathbf{u} is closed under both of them.

Lemma 20. *Let $\mathbf{v} = S(\mathbf{u}) \in \{0, 1\}^{\mathbb{N}}$. The language $\mathcal{L}(\mathbf{u})$ contains infinitely many E -palindromes and R -palindromes if and only if $\mathcal{L}(\mathbf{v})$ contains infinitely many R -palindromes centered at the letter 1 and infinitely many R -palindromes not centered at the letter 1.*

Proof: Let u be a finite non-empty word and let $v = S(u)$. It suffices to realize the following:

1. u is an E -palindrome if and only if v is an R -palindrome centered at the letter 1;
2. u is an R -palindrome of even length if and only if v is an R -palindrome centered at the letter 0;
3. u is an R -palindrome of odd length if and only if v is an R -palindrome of even length, i.e., centered at ε .

□

Corollary 21. *Let $\mathbf{v} = S(\mathbf{u}) \in \{0, 1\}^{\mathbb{N}}$ be uniformly recurrent. If $\mathcal{L}(\mathbf{v})$ contains infinitely many R -palindromes centered at the letter 1 and infinitely many R -palindromes not centered at the letter 1, then \mathbf{u} is closed under all elements of H .*

Proof: The previous lemma implies that $\mathcal{L}(\mathbf{u})$ contains infinitely many E -palindromes and R -palindromes. Let $w \in \mathcal{L}(\mathbf{u})$. Since $\mathcal{L}(\mathbf{u})$ contains R -palindromes of arbitrary length and \mathbf{u} is uniformly recurrent by Lemma 17, the factor w is a factor of an R -palindromic factor of \mathbf{u} , thus $R(w)$ also occurs in \mathbf{u} . Analogously, $E(w)$ is factor of an E -palindromic factor and thus $E(w) \in \mathcal{L}(\mathbf{u})$. □

An example of application of the last corollary are Sturmian words. It is known that they contain infinitely many R -palindromes centered at 1 and 0, which implies that their preimages by S , namely the complementary-symmetric Rote words, have their language closed under H .

Another example is the period doubling word defined in Example 19. One can easily see that given an R -palindrome w centered at $x \in \{0, 1\}$, the word $\varphi_{PD}(w)1$ is also an R -palindrome centered at $1 - x$. Therefore, the period doubling word satisfies the assumptions of the corollary and it follows that the language of one of its preimage by S , namely the Thue–Morse word, is closed under H .

The following definition is inspired by Balková et al. (2016). Given a finite word v and a letter a , the notation $|w|_a$ stands for the number of occurrences of the letter a in v .

Definition 22. We say that an infinite word $\mathbf{v} \in \{0, 1\}^{\mathbb{N}}$ has *well distributed occurrences modulo 2* (denoted WELLDOC(2)) if for every factor $w \in \mathcal{L}(\mathbf{v})$ we have

$$\left\{ (|v|_0, |v|_1) \bmod 2 : vw \text{ is a prefix of } \mathbf{v} \right\} = \mathbb{Z}_2^2.$$

Proposition 23. Let $\mathbf{v} \in \{0, 1\}^{\mathbb{N}}$ have WELLDOC(2) and be closed under reversal. If \mathbf{u} is a word such that $\mathbf{v} = S(\mathbf{u})$, then \mathbf{u} is closed under all elements of H .

Proof: Denote $\mathbf{v} = v_1v_2\dots$ and $\mathbf{u} = u_0u_1\dots$. Since $S(\mathbf{u}) = \mathbf{v}$, it follows that $u_0 + u_1 = v_1 \pmod 2$, $u_1 + u_2 = v_2 \pmod 2, \dots$. Summing first k equations we get

$$u_k = u_0 + \sum_{i=1}^k v_i.$$

It follows that

$$u_{k+j} = u_{k-1} + \sum_{i=k}^{k+j} v_i$$

for all $k > 0$ and $j \in \mathbb{N}$. Suppose that s and ℓ are two distinct occurrences of a factor $f \in \mathcal{L}(\mathbf{v})$ of length n . We have $\sum_{i=s}^{s+j} v_i = \sum_{i=\ell}^{\ell+j} v_i$ for all $j \in \{0, \dots, n-1\}$. If $u_{s-1} = u_{\ell-1}$, then we have $u_{s-1} \cdots u_{s+n} = u_{\ell-1} \cdots u_{\ell+n}$. On the other hand if $u_{s-1} \neq u_{\ell-1}$, then $u_{s-1} \cdots u_{s+n} = ER(u_{\ell-1} \cdots u_{\ell+n})$. Note that $u_{s-1} = u_0 + \sum_{i=1}^{s-1} v_i$ and $v_1 \cdots v_{s-1}f$ is a prefix of \mathbf{v} , and analogously for the index ℓ . Since \mathbf{v} has WELLDOC(2), we may choose the indices s and ℓ such that

$$\sum_{i=1}^{s-1} v_i = 0 \quad \text{and} \quad \sum_{i=1}^{\ell-1} v_i = 1.$$

It implies that with every factor $w \in \mathcal{L}(\mathbf{u})$, the factor $ER(w)$ also occurs in \mathbf{u} . As $\mathbf{v} = S(\mathbf{u})$ is closed under reversal, Lemma 18 implies that \mathbf{u} is closed under R or E . This together with the closedness under ER already implies that \mathbf{u} is closed under all elements of H . \square

4.3 Richness of \mathbf{u} versus richness of $S(\mathbf{u})$

This section is devoted to the study of images and preimages of almost rich words by the mapping S .

Theorem 24. Let $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$ be closed under all elements of $H = \{\text{Id}, E, R, ER\}$. The word \mathbf{u} is H -rich (resp. almost H -rich) if and only if $S(\mathbf{u})$ is R -rich (resp. almost R -rich).

Proof: Let $\mathbf{v} = S(\mathbf{u})$. Since \mathbf{u} is closed under ER , any factor of $\mathcal{L}(\mathbf{v})$ has two preimages in $\mathcal{L}(\mathbf{u})$. Moreover, v is right special in $\mathcal{L}(\mathbf{u})$ if and only if $ER(v)$ is right special in $\mathcal{L}(\mathbf{u})$ as well. Thus by Lemma 16, any right special factor in $\mathcal{L}(\mathbf{v})$ of length n is image of two right special factors of length $n+1$. According to (3) we get

$$2\Delta\mathcal{C}_{\mathbf{v}}(n) = \Delta\mathcal{C}_{\mathbf{u}}(n+1). \quad (15)$$

Analogously, v is an R - or E -palindrome in $\mathcal{L}(\mathbf{u})$ if and only if $ER(v)$ is an R - or E -palindrome in $\mathcal{L}(\mathbf{u})$. According to Property III we have

$$2\mathcal{P}_{\mathbf{v}}^R(n) = \mathcal{P}_{\mathbf{u}}^R(n+1) + \mathcal{P}_{\mathbf{u}}^E(n+1).$$

Thus the equality

$$\Delta\mathcal{C}_{\mathbf{v}}(n) + 2 = \mathcal{P}_{\mathbf{v}}^R(n+1) + \mathcal{P}_{\mathbf{v}}^E(n),$$

testifying that \mathbf{v} is R -rich, holds if and only if the equality

$$\Delta\mathcal{C}_{\mathbf{u}}(n+1) + 4 = \mathcal{P}_{\mathbf{u}}^R(n+1) + \mathcal{P}_{\mathbf{u}}^E(n+1) + \mathcal{P}_{\mathbf{u}}^R(n+2) + \mathcal{P}_{\mathbf{u}}^E(n+2),$$

testifying that \mathbf{u} is H -rich, is satisfied. \square

As already noted above, the word $\mathbf{t}_{b,2}$ is H -rich. Thus, using the last theorem with $\mathbf{u} = \mathbf{t}_{b,2}$ and (15) together with the equality

$$\mathcal{C}_{\mathbf{w}}(n) = 1 + \sum_{i=0}^{n-1} \Delta\mathcal{C}_{\mathbf{w}}(i)$$

valid for any infinite word \mathbf{w} , we obtain the following corollary:

Corollary 25. *For every integer b greater than 1 the word $\mathbf{v} = S(\mathbf{t}_{b,2})$ is R -rich. Its factor complexity satisfies*

$$\mathcal{C}_{\mathbf{v}}(n) = \frac{1}{2}(\mathcal{C}_{\mathbf{t}_{b,2}}(n) - 1).$$

Using the factor complexity of the word $\mathbf{t}_{b,2}$ described in Starosta (2012), one can see that the binary R -rich word $S(\mathbf{t}_{b,2})$ is not Sturmian.

Remark 26. Since a complementary-symmetric Rote word \mathbf{u} is closed under all elements of H and $S(\mathbf{u})$ is Sturmian, which is R -rich, Theorem 24 provides an alternative proof of Corollary 14 without exploiting the result that every Rote word \mathbf{u} is R -rich.

Theorem 27. *Let $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$ be a uniformly recurrent word. If \mathbf{u} is almost R -rich, then the word $S(\mathbf{u})$ is almost R -rich.*

Proof: Since \mathbf{u} is almost rich, its language contains infinitely many palindromes. This fact for uniformly recurrent words implies that \mathbf{u} is closed under reversal. If \mathbf{u} is closed under E as well, then according to Proposition 12 the word \mathbf{u} is almost H -rich, and the claim follows from Theorem 24.

It is enough to consider \mathbf{u} that is not closed under E . We will show that the set $\{w \in \mathcal{L}(\mathbf{u}) : ER(w) \in \mathcal{L}(\mathbf{u})\}$ is finite. Assume the opposite. Let v be a factor of length n . As \mathbf{u} is uniformly recurrent there exists a number $r(n)$ such that any factor of \mathbf{u} longer than $r(n)$ contains all factors of length n . Since $\{w \in \mathcal{L}(\mathbf{u}) : ER(w) \in \mathcal{L}(\mathbf{u})\}$ is not finite, there exists w belonging to this set and being longer than $r(n)$. And thus the factor v of length n occurs in w and $ER(v)$ occurs in $ER(w)$. Since both w and $ER(w)$ belong to $\mathcal{L}(\mathbf{u})$, the factor v and $ER(v)$ belongs to $\mathcal{L}(\mathbf{u})$ as well — a contradiction with assumption that \mathbf{u} is not closed under E .

Let N be the maximal length of an element of the finite set $\{w \in \mathcal{L}(\mathbf{u}) : ER(w) \in \mathcal{L}(\mathbf{u})\}$. Any factor of $S(\mathbf{u})$ longer than N has unique preimage in $\mathcal{L}(\mathbf{u})$. According to Lemma 16 for any $n > N$ there is one-to-one correspondence between $\mathcal{L}_n(S(\mathbf{u}))$ and $\mathcal{L}_{n+1}(\mathbf{u})$. Thus,

$$\Delta\mathcal{C}_{\mathbf{u}}(n+1) = \Delta\mathcal{C}_{S(\mathbf{u})}(n) \quad \text{for all } n > N. \quad (16)$$

Moreover, there exists no E -palindrome of length $n > N$ and thus we have one-to-one correspondence between the set of all R -palindromes in $\mathcal{L}(S(\mathbf{u}))$ of length n and the set of all R -palindromes in $\mathcal{L}(\mathbf{u})$ of length $n + 1$. It gives

$$\mathcal{P}_{\mathbf{u}}^R(n + 1) = \mathcal{P}_{S(\mathbf{u})}^R(n) \quad \text{for all } n > N. \quad (17)$$

Since \mathbf{u} has language closed under reversal and is almost rich using (5) there exists a constant M such that

$$\Delta\mathcal{C}_{\mathbf{u}}(n) + 2 = \mathcal{P}_{\mathbf{u}}(n) + \mathcal{P}_{\mathbf{u}}(n + 1) \quad \text{for all } n \geq M.$$

This equality and equalities (16) and (17) imply

$$\Delta\mathcal{C}_{S(\mathbf{u})}(n) + 2 = \mathcal{P}_{S(\mathbf{u})}(n) + \mathcal{P}_{S(\mathbf{u})}(n + 1) \quad \text{for all } n > \max\{N, M\}.$$

It follows that the word $S(\mathbf{u})$ is almost R -rich. \square

Corollary 28. *If \mathbf{u} is a Sturmian word, then $S^k(\mathbf{u})$ is almost R -rich for all $k > 0$.*

We add two more examples related to images (Example 29) and preimages (Example 30) of words constructed by iterated operation S . However, we do not give any proofs of their properties and we just state them as hypotheses given by computer evidence.

Example 29. As stated in Corollary 25, the word $S(\mathbf{t}_{b,2})$ is R -rich. Theorem 27 then implies that $S^k(\mathbf{t}_{b,2})$ is almost R -rich for all $k > 0$. Our computer experiments suggest that in this case the word $S^k(\mathbf{t}_{b,2})$ is in fact R -rich.

As we have already mentioned, the list of known H -rich words is very modest: complementary-symmetric Rote words and generalized binary Thue–Morse words $\mathbf{t}_{b,2}$. Theorem 24, Proposition 23 and Corollary 21 give us a recipe for construction of an (almost) H -rich word: take a binary (almost) R -rich word with property WELLDOC(2) or a binary (almost) R -rich word with suitable structure of palindromes and find its preimage by the operation S . The complementary-symmetric Rote words were obtained by this procedure applied to the Sturmian words. The Thue–Morse word $\mathbf{t} = \mathbf{t}_{2,2}$ can be obtained by this procedure applied to the period doubling word.

Example 30. Let \mathbf{u} be a Sturmian word. Let $\mathbf{u}^{(k)}$ be an infinite word such that $S^k(\mathbf{u}^{(k)}) = \mathbf{u}$ for all $k \in \mathbb{N}$. The word $\mathbf{u}^{(1)}$ is a complementary-symmetric Rote word which is, as already mentioned, H -rich and R -rich. According to our computer experiments, so is the word $\mathbf{u}^{(2)}$. The word $\mathbf{u}^{(3)}$ is not R -rich, but it is still H -rich. The word $\mathbf{u}^{(k)}$ for $k > 3$ is not H -rich nor R -rich. However, the symmetries of $\mathbf{u}^{(k)}$ are preserved: $\mathbf{u}^{(k)}$ is closed under all elements of H . This is witnessed by the following difference of its factor complexity

$$\Delta\mathcal{C}_{\mathbf{u}^{(k)}}(n) = 2^{n-1} \quad \text{for } 0 < n \leq k \quad \text{and} \quad \Delta\mathcal{C}_{\mathbf{u}^{(k)}}(n) = 2^k \quad \text{for } n > k$$

which is suggested by our experiments.

5 The mapping S on multiliteral alphabets

In this section we study the mapping S acting on a larger alphabet $\mathbb{Z}_m = \{0, \dots, m-1\}$. The mapping S is defined for every word $w = w_0 \cdots w_n$ with $w_i \in \mathbb{Z}_m$ by

$$S(w_0 w_1 \cdots w_n) = v_1 \cdots v_n, \quad (18)$$

where $v_i = (w_{i-1} + w_i) \bmod m$ for every $i \in \{1, \dots, n\}$.

The alphabet \mathbb{Z}_m allows many finite groups generated by involutory antimorphisms. We restrict our attention to groups isomorphic to groups of symmetries of a regular polyhedron. The reason is simple: we have examples of G -rich words only for such groups, namely the generalized Thue–Morse words. We demonstrate that at least for these words the mapping S transforms a G -rich word to an almost G' -rich word (cf. Theorems 24 and 27 for an analogue on the binary alphabet).

Let us describe the elements of the mentioned group explicitly. For all $x \in \mathbb{Z}_m$ denote by Ψ_x the antimorphism given by

$$\Psi_x(k) = x - k \quad \text{for all } k \in \mathbb{Z}_m$$

and by Π_x the morphism given by

$$\Pi_x(k) = x + k \quad \text{for all } k \in \mathbb{Z}_m.$$

The group $I_2(m)$ is the union of these antimorphisms and morphisms:

$$I_2(m) = \{\Psi_x : x \in \mathbb{Z}_m\} \cup \{\Pi_x : x \in \mathbb{Z}_m\}.$$

The definition of the generalized Thue–Morse words is recalled in Preliminaries. It is known that the word $\mathbf{t}_{b,m}$ is a fixed point of the morphism $\varphi_{b,m} : \mathbb{Z}_m^* \rightarrow \mathbb{Z}_m^*$ defined by

$$\varphi_{b,m} : a \mapsto a(a+1)(a+2) \cdots (a+b-1) \quad \text{for all } a \in \mathbb{Z}_m.$$

Let us stress that all operations on letters in this section are taken modulo m . As shown in Starosta (2012), the word $\mathbf{t}_{b,m}$ is closed under all elements of $I_2(m)$ and moreover $\mathbf{t}_{b,m}$ is $I_2(m)$ -rich. We will focus on images of $\mathbf{t}_{b,m}$ by S with parameters $b \geq 3$ and $m \geq 3$. Let $I'_2(m)$ denote the group generated by antimorphisms $\{\Psi_{2y} : y \in \mathbb{Z}_m\}$; it can be easily seen that

$$I'_2(m) = \{\Psi_{2x} : x \in \mathbb{Z}_m\} \cup \{\Pi_{2x} : x \in \mathbb{Z}_m\}. \quad (19)$$

If m is odd, then $I'_2(m) = I_2(m)$, if m is even, then $I'_2(m)$ is isomorphic to $I_2(\frac{m}{2})$.

The aim of this section is to prove the following theorem.

Theorem 31. *Let $m, b \in \mathbb{Z}$ such that $m \geq 3$ and $b \geq 3$.*

1. *The word $S(\mathbf{t}_{b,m})$ is almost $I'_2(m)$ -rich.*
2. *If m or b is odd, the word $S(\mathbf{t}_{b,m})$ is $I'_2(m)$ -rich.*

The first part of Theorem 31 is a direct consequence of Proposition 34, the second part follows from Lemma 35 and the description of factors of $S(\mathbf{t}_{b,m})$ up to the length 3 presented at the end of this section.

Example 32. Let us consider the word $\mathbf{t}_{4,4}$. It starts with 0 and it is a fixed point of the morphism

$$\varphi_{4,4}: 0 \mapsto 0123, \quad 1 \mapsto 1230, \quad 2 \mapsto 2301 \text{ and } 3 \mapsto 3012.$$

$$\begin{aligned} \text{Thus } \mathbf{t}_{4,4} &= 01231230230130121230230130120123230130120123\dots \\ \text{and } S(\mathbf{t}_{4,4}) &= 1310313213103133313213103132131113103132131\dots \end{aligned}$$

Now we consider the word $\mathbf{t}_{3,4}$. Its fixing morphism is

$$\varphi_{3,4}: 0 \mapsto 012, \quad 1 \mapsto 123, \quad 2 \mapsto 230 \text{ and } 3 \mapsto 301.$$

$$\begin{aligned} \text{Thus } \mathbf{t}_{3,4} &= 01212323012323030123030101212323030123030101\dots \\ \text{and } S(\mathbf{t}_{3,4}) &= 13331113131113331313331113331113331313331113\dots \end{aligned}$$

We start with a list of observations concerning properties of the mapping S . To deduce some of the observations we exploit a peculiar property of generalized Thue–Morse words. The form of morphism $\varphi_{b,m}$ forces the language of $\mathbf{u} = \mathbf{t}_{b,m}$ to have the following property

$$u_0u_1u_2u_3 \in \mathcal{L}(\mathbf{u}) \Rightarrow u_i - u_{i-1} = 1 \text{ for at least two indices } i \in \{1, 2, 3\}. \quad (20)$$

(A) $S\Psi_y = \Psi_{2y}S$ for any $y \in \mathbb{Z}_m$. If m is even, then $S\Psi_y = S\Psi_{y+\frac{m}{2}}$ for any $y \in \mathbb{Z}_m$.

(B) If $S(u_0 \cdots u_n) = S(v_0 \cdots v_n)$ with $u_i, v_j \in \mathbb{Z}_m$, then there exists $x \in \mathbb{Z}_m$ such that

$$v_0 \cdots v_n = (u_0 + x)(u_1 - x) \cdots (u_n + (-1)^n x).$$

(C) Let \mathbf{u} be closed under all elements of $I_2(m)$ and satisfy (20). Consider $w = S(v)$ for some $v = v_0v_1 \cdots v_n \in \mathcal{L}(\mathbf{u})$ with $n \geq 3$.

If m is odd, then v is the only preimage of w by S in \mathbf{u} .

If m is even, then w has exactly two preimages by S in \mathbf{u} , namely $v_0v_1 \cdots v_n$ and $(v_0 + \frac{m}{2})(v_1 + \frac{m}{2}) \cdots (v_n + \frac{m}{2})$.

Proof: Let $S(u) = S(v)$ for a factor $u = u_0u_1 \cdots u_n \in \mathcal{L}(\mathbf{u})$. As $n \geq 3$, property (20) implies that there exists $j \in \{1, 2, 3\}$ such that $u_j - u_{j-1} = v_j - v_{j-1} = 1$. From Property (B), we obtain $v_j - v_{j-1} = u_j - u_{j-1} + (-1)^j 2x$, thus $2x = 0$. If m is odd, then necessarily $x = 0$. If m is even, then also $x = \frac{m}{2}$ satisfies $2x = 0$. As the morphism $\Psi_0\Psi_{\frac{m}{2}}$ maps a to $a + \frac{m}{2}$, the language of \mathbf{u} is closed under addition of $\frac{m}{2}$ to all letters of any factor of \mathbf{u} , i.e., $(v_0 + \frac{m}{2})(v_1 + \frac{m}{2}) \cdots (v_n + \frac{m}{2}) \in \mathcal{L}(\mathbf{u})$. \square

(D) Let \mathbf{u} be closed under all elements of $I_2(m)$ and satisfy (20). Consider $u \in \mathcal{L}(\mathbf{u})$.

(a) If u is a Ψ_y -palindrome, then $S(u)$ is a Ψ_{2y} -palindrome.

(b) If $S(u)$ is a Ψ_{2y} -palindrome with $|S(u)| \geq 3$ and m is odd, then u is a Ψ_y -palindrome.

(c) If $S(u)$ is a Ψ_{2y} -palindrome with $|S(u)| \geq 3$ and m is even, then u is a Ψ -palindrome for $\Psi = \Psi_y$ and $\Psi = \Psi_{y+\frac{m}{2}}$.

Proof: (Da): Applying S to $u = \Psi_y(u)$ and using (A), one has $S(u) = S(\Psi_y(u)) = \Psi_{2y}(S(u))$, i.e., $S(u)$ is a Ψ_{2y} -palindrome.

(Db) and (Dc): Using Property (A), we obtain $S(u) = \Psi_{2y}(S(u)) = S(\Psi_y(u))$. As $|S(u)| \geq 3$ implies $|u| \geq 4$ and (20) is satisfied, we may apply Property (C). For odd m , it implies $\Psi_y(u) = u$ as we want to show. For even m , we have also the second possibility $u = \Psi_y(\Psi_0\Psi_{\frac{m}{2}}(u))$. It is easy to check that $\Psi_y\Psi_0\Psi_{\frac{m}{2}} = \Psi_{y+\frac{m}{2}}$. \square

To prove Theorem 31 we use the notion of complete G -return word of an orbit, as introduced in Pelantová and Starosta (2014). Let us recall that the orbit $[w]$ of a factor $w \in \mathcal{L}(\mathbf{u})$ is defined by (11).

A factor $v \in \mathcal{L}(\mathbf{u})$ is a *complete G -return word of $[w]$ in \mathbf{u}* if

- $|v| > |w|$,
- a prefix and a suffix of v belong to $[w]$, and
- v contains no other elements of $[w]$.

Theorem 33 (Pelantová and Starosta (2014)). *If \mathbf{u} is an infinite word closed under all elements of G , then*

1. \mathbf{u} is G -rich if and only if for all $w \in \mathcal{L}(\mathbf{u})$ every complete G -return word of $[w]$ is a G -palindrome.
2. \mathbf{u} is almost G -rich if and only if there exists an integer N such that for all $w \in \mathcal{L}(\mathbf{u})$ longer than N every complete G -return word of $[w]$ is a G -palindrome.

Using the previous theorem, we can easily prove the proposition which directly implies the validity of the first part of Theorem 31 because the generalized Thue–Morse words satisfy its assumption.

Proposition 34. *Let $\mathcal{A} = \mathbb{Z}_m$ and let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ be closed under all elements of $I_2(m)$. If $\mathcal{L}(\mathbf{u})$ satisfies (20) and \mathbf{u} is $I_2(m)$ -rich, then the word $S(\mathbf{u})$ is almost $I_2(m)$ -rich.*

Proof: To ease the notation put $G = I_2(m)$ and $G' = I_2'(m)$. If m is odd then $G = G'$. Otherwise, $\#G = 2\#G'$.

As \mathbf{u} satisfies (20), we can apply Property (D) to each palindrome $S(u)$ in $S(\mathbf{u})$ with length $|S(u)| \geq 3$. Property (D) implies that $u \in \mathcal{L}(\mathbf{u})$ is a G -palindrome if and only if $S(u)$ is a G' -palindrome in $\mathcal{L}(S(\mathbf{u}))$. Let $S(u)$ be a G' -palindrome in $S(\mathbf{u})$ and $S(v)$ be a complete G' -return word of $[S(u)]$ in $S(\mathbf{u})$. Then v is a complete G -return word in \mathbf{u} of $[u]$. The word \mathbf{u} is G -rich and due to Theorem 33, the factor v is a G -palindrome. According to Property (D), the complete G' -return word $S(v)$ is a G' -palindrome. Thus $S(\mathbf{u})$ is almost G' -rich. \square

In the remaining part of this section we focus on G' -richness of $S(\mathbf{t}_{b,m})$ in the case when m or b is odd.

Lemma 35. *Let $G' = I_2'(m)$ and $\mathbf{v} = S(\mathbf{t}_{b,m})$. If for $n = 1$ and $n = 2$ the equality*

$$\Delta\mathcal{C}_{\mathbf{v}}(n) + \#G' = \sum_{\substack{\Psi \in G' \\ \Psi \text{ is an antimorphism}}} \left(\mathcal{P}_{\mathbf{v}}^{\Psi}(n) + \mathcal{P}_{\mathbf{v}}^{\Psi}(n+1) \right) \quad (21)$$

holds, then $\mathbf{v} = S(\mathbf{t}_{b,m})$ is G' -rich.

Proof: We combine the results of Pelantová and Starosta (2014).

It is easy to see that $\Psi_{2x}(a) \neq \Psi_{2y}(a)$ for any $a \in \mathbb{Z}_m$ and any pair of distinct antimorphisms Ψ_{2x} and Ψ_{2y} from G' . This property guarantees that the number 1 is G' -distinguishing in the sense of Definition 7 of Pelantová and Starosta (2014). In the proof of Proposition 34 we verify that any complete G' -return word of $[w]$ in \mathbf{v} is a G' -palindrome for each G' -palindrome $w \in \mathcal{L}(\mathbf{v})$ of length at least 3. As 1 is G' -distinguishing, Lemma 28 of Pelantová and Starosta (2014) says that this fact implies equality (21) for all $n \geq 3$. According to Proposition 42 of Pelantová and Starosta (2014) the word \mathbf{v} is G' -rich if the equality in (21) holds for each $n \in \mathbb{N}, n \geq 1$. \square

In the case m or b odd and $n = 1$ and $n = 2$, we will confirm the equality (21) in the next lemma. To ease the notation we put

$$F(n) = \sum_{\substack{\Psi \in G' \\ \Psi \text{ is an antimorphism}}} \mathcal{P}_{\mathbf{v}}^{\Psi}(n)$$

and $\rho : \mathbb{Z}_m \rightarrow \mathbb{Z}_m$ denotes the permutation

$$\rho(k) = k + b - 1 = \Pi_{b-1}(k).$$

We will use the following statements from Starosta (2012). Let q denote the order of ρ , i. e., the least positive integer q such that $q(b-1) \equiv 0 \pmod{m}$. The factors of length 2 of $\mathbf{t}_{b,m}$ are

$$\mathcal{L}_2(\mathbf{t}_{b,m}) = \{\rho^k(r-1)r : r \in \mathbb{Z}_m, 0 \leq k \leq q-1\}$$

and of length 3

$$\mathcal{L}_3(\mathbf{t}_{b,m}) = \{\rho^k(r-1)r(r+1) : r \in \mathbb{Z}_m, 0 \leq k \leq q-1\} \cup \{(r-1)r\rho^{-k}(r+1) : r \in \mathbb{Z}_m, 0 \leq k \leq q-1\}.$$

It is easy to deduce the set of all factors of length 4:

$$\begin{aligned} \mathcal{L}_4(\mathbf{t}_{b,m}) = & \{\rho^k(r-1)r(r+1)(r+2) : r \in \mathbb{Z}_m, 0 \leq k \leq q-1\} \\ & \cup \{(r-2)(r-1)r\rho^{-k}(r+1) : r \in \mathbb{Z}_m, 0 \leq k \leq q-1\} \\ & \cup \{(r-1)r\rho^{-k}(r+1)(\rho^{-k}(r+1)+1) : r \in \mathbb{Z}_m, 0 \leq k \leq q-1\}. \end{aligned}$$

Lemma 36. *Let $\mathbf{v} = S(\mathbf{t}_{b,m})$. The numbers of factors of \mathbf{v} of length n satisfy the following:*

n	m odd	m even, b odd	m even, b even
1	m	$\frac{m}{2}$	m
2	qm	$\frac{qm}{2}$	$\frac{3qm}{4}$
3	$3qm - 2m$	$\frac{3qm}{2} - m$	$\frac{3qm}{2} - m$

Proof: $n = 1$

It follows from the form of $\mathcal{L}_2(\mathbf{t}_{b,m})$ that $\mathcal{L}_1(\mathbf{v}) = \{\rho^k(r-1) + r : r \in \mathbb{Z}_m, 0 \leq k \leq q-1\}$. We have $\rho^k(r-1) + r = r-1 + k(b-1) + r = 2r-1 + k(b-1)$

If m is odd, then we have directly that $\mathcal{L}_1(\mathbf{v}) = \mathbb{Z}_m$ (for $k = 0$).

If m is even and b is odd, then $2r - 1 + k(b - 1)$ is odd for every r and k , and thus $\mathcal{L}_1(\mathbf{v}) = \{2i + 1 : 0 \leq i < \frac{m}{2}\}$.

If m is even and b is even, then for $k = 0$ the number $2r - 1$ is odd, and for $k = 1$ the number $2r - 1 + (b - 1)$ is even, and we have $\mathcal{L}_1(\mathbf{v}) = \mathbb{Z}_m$.

$n = 2$

The structure of $\mathcal{L}_3(\mathbf{t}_{b,m})$ implies that the factors of \mathbf{v} of length 2 are of the following forms:

1. $(\rho^k(r - 1) + r)(2r + 1) = (2r - 1 + k(b - 1))(2r + 1)$ for $r \in \mathbb{Z}_m$ and $0 \leq k < q$, and
2. $(2r' - 1)(\rho^{k'}(r' + 1) + r') = (2r' - 1)(2r' + 1 + k'(b - 1))$ for $r' \in \mathbb{Z}_m$ and $0 \leq k' < q$.

First, let us see the number of factors of type 1. Fix $r \in \mathbb{Z}_m$. Suppose $2r + 1 = 2\tilde{r} + 1$ for some \tilde{r} . This equation has 1 solution for m odd and 2 solutions for m even. It is easy to see that if m is odd, there are qm distinct factors of type 1 and if m is even their number is $\frac{qm}{2}$. The counts for the second type are exactly the same.

Let us now look how the two types of factors overlap. Fix r and k and suppose

$$\begin{aligned} 2r - 1 + k(b - 1) &= 2r' - 1, \\ 2r + 1 &= 2r' + 1 + k'(b - 1). \end{aligned}$$

It follows that $2r' = 2r + k(b - 1)$ which may not have a solution only if m is even and b is even, otherwise it has a solution and the two types overlap completely. If m is even and b is even, the two types overlap only if k is even. Thus, they overlap in $\frac{qm}{4}$ cases.

$n = 3$

It follows from $\mathcal{L}_4(\mathbf{t}_{b,m})$ that there are the following 3 types of factors of length 3:

1. $(\rho^k(r - 1) + r)(2r + 1)(2r + 3)$ for $r \in \mathbb{Z}_m$, $0 \leq k < q$;
2. $(2r' - 3)(2r' - 1)(r' + \rho^{-k'}(r' + 1))$ for $r' \in \mathbb{Z}_m$, $0 \leq k' < q$;
3. $(2r'' - 1)(r'' + \rho^{-k''}(r'' + 1))(\rho^{-k''}(r'' + 1) + \rho^{-k''}(r'' + 1) + 1)$ for $r'' \in \mathbb{Z}_m$, $0 \leq k'' < q$.

Analogously to the previous case $n = 2$, it can be shown that there are qm distinct factors of each type if m is odd, and $\frac{qm}{2}$ distinct factors of each type if m is even.

When investigating the common factors of each type, one can show that each pair has m common factors if m is odd, and $\frac{m}{2}$ common factors otherwise. Overall, we find that $\#\mathcal{L}_3(\mathbf{v}) = 3qm - 2m$ if m is odd and $\#\mathcal{L}_3(\mathbf{v}) = \frac{3qm}{2} - m$ if m is even. \square

Lemma 37. Let $G' = I'_2(m)$ and $\mathbf{v} = S(\mathbf{t}_{b,m})$. The values of $F(n)$ for $n \in \{1, 2, 3\}$ are as follows:

n	m odd	m even, b odd	m even, b even
1	m	$\frac{m}{2}$	m
2	qm	$\frac{qm}{2}$	$\frac{qm}{4}$
3	qm	$\frac{qm}{2}$	$\frac{qm}{2}$

Proof: $n = 1$:

It is not hard to show that every factor of length 1 is a Ψ -palindrome for a unique $\Psi \in G'$.

$n = 2$:

Suppose that the first type of factor of length 2 from the proof of Lemma 36 is a Ψ_ℓ -palindrome. We have

$$2r - 1 + k(b - 1) = \Psi_\ell(2r + 1) = \ell - 2r - 1,$$

which leads to $\ell = 4r - 2 + k(b - 1)$, i.e., every factor of length 2 is a Ψ -palindrome for some $\Psi \in G'$ except for the case of m even and b even where one may find such ℓ only if k is even. As one can see in the proof of Lemma 36, the case of k even is when the two types of factors of length 2 overlap, thus the total number of G' -palindromes is $\frac{qm}{4}$ in this case.

$n = 3$:

We will refer to the 3 types of factors of length 3 as given in the proof of Lemma 36 above. The first two types can be a Ψ -palindrome for some $\Psi \in I_2(m)$ if and only if $k = 0$ or $k' = 0$. This case is also included in the third type for $k'' = 0$, so we need just to check this type.

Suppose that a factor of the third type is a Ψ_ℓ -palindrome:

$$\begin{aligned} 2r'' - 1 &= \Psi_\ell(\rho^{-k''}(r'' + 1) + \rho^{-k''}(r'' + 1) + 1) = \ell - (2r'' + 3 - 2k''(b - 1)), \\ r'' + \rho^{-k''}(r'' + 1) &= 2r'' + 1 - k''(b - 1) = \Psi_\ell(r'' + \rho^{-k''}(r'' + 1)) = \ell - (2r'' + 1 - k''(b - 1)). \end{aligned}$$

Both equalities yield $\ell = 4r'' + 2 - 2k''(b - 1)$. Thus, every factor of type 3 is a Ψ -palindrome for some $\Psi \in I_2(m)$. According to the proof of Lemma 36, there are qm such factors if m is odd, and $\frac{qm}{2}$ such factors otherwise. \square

Proof of the second part of Theorem 31: Results of Lemmas 36 and 37 can be summarized into the following table:

	m odd	m even, b odd	m even, b even
$\Delta\mathcal{C}_\vee(1)$	$(q - 1)m$	$(q - 1)\frac{m}{2}$	$\frac{3qm}{4} - m$
$F(1) + F(2)$	$(q + 1)m$	$\frac{(q+1)m}{2}$	$\frac{qm}{4} + m$
$\Delta\mathcal{C}_\vee(2)$	$2qm - 2m$	$qm - m$	$\frac{3qm}{4} - m$
$F(2) + F(3)$	$2qm$	qm	$\frac{3qm}{4}$
$\#G'$	$2m$	m	m

Therefore the assumption of Lemma 35 is satisfied in the case when m or b is odd. Consequently, $S(\mathbf{t}_{b,m})$ is G' -rich. \square

Corollary 38. Let $b \in \mathbb{N}, b \geq 1$ and S be the operation defined by (18) for the alphabet \mathbb{Z}_4 . We have

- $S(\mathbf{t}_{2b+1,4})$ is an infinite word over the binary alphabet $\{1, 3\}$ and it is H -rich (here H stands for the group generated by the both involutory antimorphisms over the binary alphabet $\{1, 3\}$).
- $S^2(\mathbf{t}_{2b+1,4})$ is an infinite word over the binary alphabet $\{0, 2\}$ and it is R -rich (R stands for the reversal mapping over the binary alphabet $\{0, 2\}$).

- $S^k(\mathbf{t}_{2b+1,4})$ is an infinite word over the binary alphabet $\{0, 2\}$ and it is almost R -rich for any $k \in \mathbb{N}, k \geq 2$.

Proof: The fact that the alphabet of $S(\mathbf{t}_{2b+1,4})$ is $\{1, 3\}$ is shown in the proof of Lemma 36, where $\mathcal{L}_1(\mathbf{v})$ is described for any generalized Thue–Morse word \mathbf{v} . According to the second part of Theorem 31, the word $S(\mathbf{t}_{2b+1,4})$ is $I'_2(4)$ -rich. The group $I'_2(4)$ defined by (19) is isomorphic to H .

It is easy to see that the operation S assigns to any word $\mathbf{u} \in \{1, 3\}^{\mathbb{N}}$ the word over the alphabet $\{0, 2\}$. The second part of the corollary follows from Theorem 24, the third one from Theorem 27. \square

6 Comments and open questions

For infinite words over the binary alphabet $\{0, 1\}$, we illustrated that the operation S puts into a broader context the classical richness (here usually referred to as R -richness) and H -richness. The main open question is which other operation acting on infinite words behaves analogously. Let us mention here some open questions connected with S .

- The operation S on $\{0, 1\}$ applied to an almost R -rich word gives an almost R -rich word, see Theorem 27. In particular, any iteration of S applied to a Sturmian word \mathbf{u} gives an almost rich word, cf. Corollary 28. Is the R -defect of $S^k(\mathbf{u})$ zero as suggested by our computer experiments?
- The operation S on $\{0, 1\}$ applied to an H -rich word gives an R -rich word. In particular, $S(\mathbf{t}_{b,2})$ is rich for any generalized Thue–Morse word $\mathbf{t}_{b,2}$. Our computational experiments suggest that $S^k(\mathbf{t}_{b,2})$ is R -rich for any $k \in \mathbb{N}, k \geq 1$, see Example 29. Is it true?
- On the other hand, any preimage by S of each Sturmian word \mathbf{u} is H -rich and R -rich simultaneously, in fact it is a complementary-symmetric Rote word. Our computer experiments suggest that even the second preimage $S^{-2}(\mathbf{u})$ is simultaneously H - and R -rich, whereas $S^{-3}(\mathbf{u})$ is only H -rich, but not R -rich, see Example 30. Is it true?

We have introduced the operation S over the alphabet \mathbb{Z}_m with $m \geq 3$ as well. But our results on multiliteral alphabet are restricted to special groups and words.

- We have considered G -richness for $G = I_2(m)$ only. Proposition 34 connects $I_2(m)$ -richness of \mathbf{u} and $I'_2(m)$ -richness of $S(\mathbf{u})$ for words \mathbf{u} satisfying the assumption (20). Is the proposition valid without the assumption?
- It would be interesting to study behaviour of ternary episturmian words with respect to operation S on \mathbb{Z}_3 . For example, which group of symmetries G has the preimage of the Tribonacci word by S ? Is the preimage G -rich? Are images of the Tribonacci word by S still R -rich?
- Corollary 38 illustrates that the operation S over the alphabet \mathbb{Z}_4 can produce binary almost R -rich words as well. What is the R -defect of the words $S^k(\mathbf{t}_{2b+1,4})$?

The last comment we want to state here concerns the palindromic closure operator. It is used for construction of standard episturmian words. The construction is governed by a directive sequence of letters Δ . Any episturmian word \mathbf{u} is closed under reversal and \mathbf{u} is rich in the classical sense. In de

Luca and De Luca (2006), the authors introduced the concept of generalized pseudopalindromic closure operator, where multiple involutory antimorphisms are used. It means that the construction is governed by two sequences: a directive sequence of letters Δ and a directive sequence of antimorphisms Θ . Let us denote the resulting infinite word by $\mathbf{u}(\Delta, \Theta)$.

In general, $\mathbf{u}(\Delta, \Theta)$ is closed under the group G generated by the involutory antimorphisms occurring infinitely many times in the directive sequence Θ , but the word $\mathbf{u}(\Delta, \Theta)$ need not to be G -rich. Nevertheless, several examples of G -rich words constructed by generalized pseudopalindromic closure operator are already known. De Luca and de Luca showed that the Thue–Morse word $\mathbf{t} = \mathbf{t}_{2,2}$ can be constructed in this way. In Jajcayová et al. (2014) the generalized Thue–Morse words $\mathbf{t}_{b,m}$ with the same property are characterized. The concept of generalized pseudopalindromic closure on binary alphabet is systematically studied by Blondin Massé, Paquin, Tremblay and Vuillon in Blondin Massé et al. (2013). In particular, they proved that any standard complementary-symmetric Rote word can be constructed by using generalized pseudopalindromic closure operator. Nevertheless, the question which pairs (Δ, Θ) produce H -rich words is open and requires a deeper study.

Acknowledgements

The first author acknowledges financial support from the Czech Science Foundation grant GAČR 13-03538S and the second author acknowledges financial support from the Czech Science Foundation grant GAČR 13-35273P. The computer experiments were performed using the open-source computer algebra system SageMath The Sage Developers (2015). We would like to thank the anonymous referee who provided us with useful hints to improve the article.

References

- J.-P. Allouche and J. Shallit. Sums of digits, overlaps, and palindromes. *Discrete Math. Theor. Comput. Sci.*, 4:1–10, 2000.
- P. Baláži, Z. Masáková, and E. Pelantová. Factor versus palindromic complexity of uniformly recurrent infinite words. *Theor. Comput. Sci.*, 380(3):266–275, 2007. doi: 10.1016/j.tcs.2007.03.019.
- L. Balková, M. Bucci, A. D. Luca, J. Hladký, and S. Puzynina. Aperiodic pseudorandom number generators based on infinite words. *Theor. Comput. Sci.*, 647:85–100, 2016. ISSN 0304-3975. doi: 10.1016/j.tcs.2016.07.042.
- L. Balková. *Beta-integers and Quasicrystals*. PhD thesis, Czech Technical University in Prague and Université Paris Diderot-Paris 7, 2008.
- L. Balková, E. Pelantová, and Š. Starosta. Palindromes in infinite ternary words. *RAIRO-Theor. Inf. Appl.*, 43(4):687–702, oct 2009. doi: 10.1051/ita/2009016.
- L. Balková, E. Pelantová, and Š. Starosta. Sturmian jungle (or garden?) on multiliteral alphabets. *RAIRO-Theor. Inf. Appl.*, 44:443–470, 2010. doi: 10.1051/ita/2011002.
- L. Balková, E. Pelantová, and Š. Starosta. Infinite words with finite defect. *Adv. in Appl. Math.*, 47(3): 562–574, 2011. ISSN 0196-8858. doi: 10.1016/j.aam.2010.11.006.

- L. Balková, E. Pelantová, and Š. Starosta. Proof of the Brlek-Reutenauer conjecture. *Theor. Comput. Sci.*, 475:120–125, 2013. ISSN 0304-3975. doi: 10.1016/j.tcs.2012.12.024.
- A. Blondin Massé, S. Brlek, A. Garon, and S. Labbé. Combinatorial properties of f -palindromes in the Thue-Morse sequence. *Pure Math. Appl.*, 19(2-3):39–52, 2008.
- A. Blondin Massé, S. Brlek, S. Labbé, and L. Vuillon. Palindromic complexity of codings of rotations. *Theor. Comput. Sci.*, 412(46):6455–6463, 2011. ISSN 03043975. doi: 10.1016/j.tcs.2011.08.007.
- A. Blondin Massé, G. Paquin, H. Tremblay, and L. Vuillon. On generalized pseudostandard words over binary alphabets. *J. Integer Seq.*, 16, 2013. ISSN 1530-7638.
- S. Brlek and C. Reutenauer. Complexity and palindromic defect of infinite words. *Theor. Comput. Sci.*, 412(4-5):493–497, 2011. ISSN 0304-3975. doi: 10.1016/j.tcs.2010.11.025.
- S. Brlek, S. Hamel, M. Nivat, and C. Reutenauer. On the palindromic complexity of infinite words. *Int. J. Found. Comput. Sci.*, 15(2):293–306, 2004. doi: 10.1142/S012905410400242X.
- M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On different generalizations of episturmian words. *Theor. Comput. Sci.*, 393(1-3):23–36, 2008. doi: 10.1016/j.tcs.2007.10.043.
- M. Bucci, A. De Luca, A. Glen, and L. Q. Zamboni. A connection between palindromic and factor complexity using return words. *Adv. in Appl. Math.*, 42(1):60–74, 2009. ISSN 0196-8858. doi: 10.1016/j.aam.2008.03.005.
- J. Cassaigne. Complexity and special factors. *Bull. Belg. Math. Soc. Simon Stevin* 4, 1:67–88, 1997.
- T. W. Cusick and L. C. Ciungu. Sum of digits sequences modulo m . *Theor. Comput. Sci.*, 412(35):4738–4741, 2011. ISSN 0304-3975. doi: 10.1016/j.tcs.2011.05.030.
- A. de Luca and A. De Luca. Pseudopalindrome closure operators in free monoids. *Theor. Comput. Sci.*, 362(1–3):282–300, 2006. ISSN 0304-3975. doi: 10.1016/j.tcs.2006.07.009.
- X. Droubay and G. Pirillo. Palindromes and Sturmian words. *Theor. Comput. Sci.*, 223(1-2):73–85, 1999. doi: 10.1016/S0304-3975(97)00188-6.
- X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theor. Comput. Sci.*, 255(1-2):539–553, 2001. doi: 10.1016/S0304-3975(99)00320-5.
- A. Glen, J. Justin, S. Widmer, and L. Q. Zamboni. Palindromic richness. *Eur. J. Combin.*, 30(2):510–531, 2009. doi: 10.1016/j.ejc.2008.04.006.
- A. Hof, O. Knill, and B. Simon. Singular continuous spectrum for palindromic Schrödinger operators. *Comm. Math. Phys.*, 174:149–159, 1995. doi: 10.1007/bf02099468.
- T. Jajcayová, E. Pelantová, and Š. Starosta. Palindromic closures using multiple antimorphisms. *Theor. Comput. Sci.*, 533:37–45, 2014. ISSN 0304-3975. doi: 10.1016/j.tcs.2014.03.020.
- E. Pelantová and Š. Starosta. Almost rich words as morphic images of rich words. *Int. J. Found. Comput. Sci.*, 23(05):1067–1083, Aug. 2012. ISSN 0129-0541, 1793-6373. doi: 10.1142/S012905411240045X.

- E. Pelantová and Š. Starosta. Languages invariant under more symmetries: overlapping factors versus palindromic richness. *Discrete Math.*, 313:2432–2445, 2013. doi: 10.1016/j.disc.2013.07.00.
- E. Pelantová and Š. Starosta. Palindromic richness for languages invariant under more symmetries. *Theor. Comput. Sci.*, 518:42–63, 2014. doi: 10.1016/j.tcs.2013.07.021.
- G. Rote. Sequences with subword complexity $2n$. *J. Number Th.*, 46:196–213, 1993. doi: 10.1006/jnth.1994.1012.
- Š. Starosta. On theta-palindromic richness. *Theor. Comput. Sci.*, 412(12-14):1111–1121, 2011. ISSN 0304-3975. doi: 10.1016/j.tcs.2010.12.011.
- Š. Starosta. Generalized Thue-Morse words and palindromic richness. *Kybernetika*, 48(3):361–370, 2012.
- Š. Starosta. Morphic images of episturmian words having finite palindromic defect. *Eur. J. Combin.*, 51: 359–371, 2016. doi: 10.1016/j.ejc.2015.07.001.
- The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 6.10)*, 2015. <http://www.sagemath.org>.