Construction of words rich in palindromes and pseudopalindromes

E. Pelantová∗ Š. Starosta†

June 13, 2014

Abstract

First, we present an overview of results on words rich in palindromes and examples of rich words. We introduce $G$-richness, where $G$ is a finite group generated by involutory antimorphisms over $A^*$. We study $G$-analogue of results known for the classical richness. Finally, we concentrate on the question how to construct $G$-rich words. We recall the notion of generalized palindromic closure which uses multiple antimorphisms. Then we define an operation $S$, which (sometimes) maps $G'$-rich words to $G''$-rich words for (in general) distinct groups $G'$ and $G''$. This operation provides a tool for construction of $G$-rich words, in particular for construction of a new class of binary words rich in the classical sense.

1 Introduction

Words rich in palindromes attract attention of many authors. One of the reasons is a plenty of distinct ways to characterize rich words. In this sense rich words are similar to Sturmian words. Let us recall that a palindrome is a fixed point of the mapping $R$ which assigns to a word $w = w_0w_1\cdots w_n$ its reversal $R(w) = w_nw_{n-1}\cdots w_0$.

An infinite word $u$ over the alphabet $A$ is said to be rich if its language $L(u)$ is saturated by palindromes up to the highest possible level. To specify this bound, we denote by $	ext{Pal}(w)$ the set of all palindromes (including the empty word) occurring in the word $w$. As proven in [9] by Droubay and Pirillo, $\#\text{Pal}(w) \leq |w| + 1$ for any word $w$, where $|w|$ stands for the length of the word $w$. An infinite word $u$ is called rich (or full) if each factor $w \in L(u)$ satisfies $\#\text{Pal}(w) = |w| + 1$.

Prominent binary rich words are Sturmian words. Further examples of binary rich words are Rote words [4] and period doubling word [1]. Arnoux–Rauzy words and words coding interval exchange transformation under the symmetric permutation of intervals represent examples of rich words over a multiliteral alphabet.

For a word $u$ having its language closed under reversal, several equivalent definitions of richness were formulated. They use the concepts of complete return word occurring in $u$, factor complexity $C$ of $u$, palindromic complexity $P$ of $u$, graph of symmetries $\Gamma_n(u)$ of $u$, bilateral order $b(w)$ of factor $w$ of $u$, etc. The following theorem summarizes some properties characterizing rich words with language closed under reversal, their proofs and precise definitions of used notation can be found in [10, 8, 6, 2].

**Theorem 1.** For an infinite word $u$ with language closed under reversal the following statements are equivalent:

---

∗Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2, Czech Republic

†Department of Applied Mathematics, FIT, Czech Technical University in Prague, Thákurova 9, 160 00 Praha 6, Czech Republic
1. \( u \) is rich,
2. any complete return word of any palindromic factor of \( u \) is a palindrome,
3. for any factor \( w \) of \( u \), every factor of \( u \) that contains \( w \) only as its prefix and \( R(w) \) only as its suffix is a palindrome,
4. the longest palindromic suffix of any factor \( w \in \mathcal{L}(u) \) is unioccurrent in \( w \),
5. for each \( n \in \mathbb{N} \) the following equality holds
   \[ C(n + 1) - C(n) + 2 = P(n) + P(n + 1), \]
6. each graph of symmetries \( \Gamma_n(u) \) satisfies: all its loops are palindromes and the graph obtained from \( \Gamma_n(u) \) by removing loops is a tree,
7. the bilateral order \( b(w) \) of any bispecial factor \( w \) of \( u \) and its set of palindromic extensions \( \text{Pext}(w) = \{ awa \in \mathcal{L}(u) : awa = R(awa) \} \) satisfy:
   - if \( w \) is non-palindromic, then \( b(w) = 0 \),
   - if \( w \) is a palindrome, then \( b(w) = \#\text{Pext}(w) - 1 \).

A property weaker than richness is often considered in connection with the number of palindromes occurring in factors. More precisely, an infinite word \( u \) is almost rich if there exists a constant \( K \) such that
\[ |w| + 1 - K \leq \#\text{Pal}(w) \leq |w| + 1 \]
for any \( w \in \mathcal{L}(u) \). The minimal such \( K \) is usually referred to as the defect of \( u \).

The first attempt to generalize the notion richness appeared in [3] and then in [17]. The reversal mapping \( R \) is replaced by an arbitrary involutory antimorphism \( \Psi : \mathcal{A}^* \rightarrow \mathcal{A}^* \), i.e., by the mapping with the following properties:
\[ \Psi^2 = \text{Id} \quad \text{and} \quad \Psi(uv) = \Psi(v)\Psi(u) \quad \text{for any} \quad u, v \in \mathcal{A}^*. \]
If \( w = \Psi(w) \), then \( w \) is a \( \Psi \)-palindrome or pseudopalindrome and \( \Psi \)-richness can be introduced analogously. As shown in [12], any uniformly recurrent \( \Psi \)-rich word is a morphic image of a rich word. In this sense, considering \( \Psi \) instead of \( R \) does not bring a broader variability into the concept of rich words.

## 2 G-richness

Recently we have introduced a further generalization of the notion of richness. The inspiration originates in the Thue–Morse word \( u_{TM} \in \{0, 1\}^\mathbb{N} \). Its language contains infinitely many palindromes and \( E \)-palindromes, where \( E \) is the antimorphism exchanging 0 and 1. Nevertheless, \( u_{TM} \) is neither rich nor \( E \)-rich. Our generalization of richness therefore consists in considering more antimorphisms simultaneously.

Let \( G \) be a finite group generated by involutory antimorphisms over \( \mathcal{A}^* \). A word \( w \in \mathcal{A}^* \) is a \( G \)-palindrome if \( w = \Psi(w) \) for some antimorphism \( \Psi \in G \). Denote the orbit of \( w \in \mathcal{A}^* \) under \( G \) by \( [w] = \{ \mu(w) : \mu \in G \} \). Since \( G \) is finite, all elements of \( [w] \) have the same length as \( w \) itself. Although we originally used for the definition of \( G \)-richness the \( G \)-analogue of the point 6 in Theorem 1 (see [13]), we demonstrated in [15] that the notion \( G \)-richness can be defined by many ways as well. We present here the definition based on the analogue with the point 3 in Theorem 1. Therefore we need to generalize the notion of return word.
Definition 2. Let $v, w \in A^*$. A word $v$ of length greater than $|w|$ is a complete $G$-return word of $[w]$ if a prefix and a suffix of $v$ belong to $[w]$ and $v$ contains no other occurrences of elements from $[w]$.

Example 3. Consider the binary alphabet $\{0, 1\}$, the group $H$ generated by the two involutory antimorphisms $R$ and $E$, and the Thue–Morse word

$$u_{TM} = 01101001100101101001011001101001100101 \ldots$$

which is a fixed point of the substitution $0 \mapsto 01$ and $1 \mapsto 10$.

The orbit of $w = 001 \in L(u_{TM})$ equals

$$[w] = \{001, R(001), E(001), ER(001)\} = \{001, 100, 011, 110\}.$$

Let us list all the complete $H$-return words of $[w]$ in $u_{TM}$:

$$v^{(1)} = 0110, \quad v^{(2)} = 1001, \quad v^{(3)} = 0011, \quad v^{(4)} = 1100, \quad v^{(5)} = 110100, \quad v^{(6)} = 001011.$$

All of the six complete $H$-return words are $H$-palindromes, since $v^{(1)}$ and $v^{(2)}$ are $R$-palindromes and the remaining four complete $H$-return words are $E$-palindromes.

Definition 4. Let $G$ be a finite group generated by involutory antimorphisms over $A^*$ and $u \in \mathcal{A}^N$ be an infinite word such that $L(u)$ is closed under $G$, i.e., $\mu(w)$ belongs to $L(u)$ for each $w \in L(u)$ and $\mu \in G$. We say that $u$ is $G$-rich if each complete $G$-return word of $[w]$ in $u$ is a $G$-palindrome for any $w \in L(u)$.

If the group $G$ contains only one antimorphism $R$ and the word $u$ has language closed under reversal, then the classical richness and $G$-richness coincide. To avoid confusion in the sequel, the classical richness will be referred to as $R$-richness.

The notion of almost $G$-richness can be introduced by using the definition of $G$-defect of a finite word. For more details the reader can consult [15].

3 Examples of $G$-rich words

3.1 Generalized Thue–Morse words

For a given integer base $b > 1$ and an integer $m > 1$, the number $s_b(n)$ denotes the digit sum of the expansion of number $n$ in the base $b$. The generalized Thue–Morse word $t_{b,m}$ is defined by

$$t_{b,m}(n) = (s_b(n) \mod m)^\infty_{n=0}.$$

Thus, the alphabet of $t_{b,m}$ is $A = \mathbb{Z}_m = \{0, \ldots, m-1\}$.

In this notation the classical Thue-Morse word $u_{TM}$ equals $t_{2,2}$. It is readily seen that $t_{b,m}$ is a fixed point of the substitution $\varphi$ determined by

$$k \mapsto k(k+1)(k+2) \ldots (k+b-1) \quad \text{for any } k \in \mathbb{Z}_m,$$

where letters are expressed modulo $m$. Moreover, $t_{b,m}$ is periodic if and only if $b = 1 \pmod{m}$.

The language of $t_{b,m}$ is closed under a finite group containing $m$ involutory antimorphisms. This group is the dihedral group $I_2(m)$ and it is generated by antimorphisms $\Psi_k$ defined for every $x \in \mathbb{Z}_m$ by

$$\Psi_k(x) = x - k \quad \text{for any } k \in \mathbb{Z}_m .$$

In [18], the second author proved that $t_{b,m}$ is $I_2(m)$-rich.
3.2 Complementary-symmetric Rote words

An infinite sequence $u$ over binary alphabet $\{0, 1\}$ is a complementary-symmetric Rote word (CSR word) if it has its factor complexity $C(n) = 2^n$ and its language $\mathcal{L}(u)$ is closed under the exchange $0 \leftrightarrow 1$. In [4], the authors proved that Rote words are rich in the classical sense, or shortly $R$-rich.

The definition of CSR word forces $\mathcal{L}(u)$ to be closed under both antimorphisms over the binary alphabet $R$ and $E$ and thus Rote words are closed under the group $H$ generated by $E$ and $R$. We may apply the following proposition proved in [14].

**Proposition 5.** Let $u \in \{0, 1\}^\mathbb{N}$ be a word having its language closed under the group $H$. If $u$ is $R$-rich, then $u$ is $H$-rich as well.

Consequently, any CSR word is $H$-rich.

4 Palindromic closures using multiple antimorphisms

Sturmian words belong to words rich in the classical sense. Among them, the most important role is played by the standard Sturmian words. It is well known that each standard Sturmian word can be constructed by iterations of the operation of palindromic closure which depends on the slope of the Sturmian word. The same property holds for standard Arnoux–Rauzy words as well.

In [7], de Luca and De Luca generalized the notion of standard words considering the set $I$ of all involutory antimorphisms on $A^*$ instead of just one fixed antimorphism. We will denote by $I^\mathbb{N}$ the set of all infinite sequences over $I$.

For $v \in A^*$ and $\vartheta \in I$, we denote $v^\vartheta$ the shortest $\vartheta$-palindrome having a prefix $v$. For example, $(01101)^E = 01101001$, where $E$ is the antimorphism on binary alphabet defined by $E(0) = 1$ and $E(1) = 0$.

**Definition 6.** Let $\Theta = \vartheta_1 \vartheta_2 \vartheta_3 \ldots \in I^\mathbb{N}$ and $\Delta = \delta_1 \delta_2 \delta_3 \ldots \in A^\mathbb{N}$. Denote

$$w_0 = \varepsilon \quad \text{and} \quad w_n = (w_{n-1} \delta_n)^{\vartheta_n} \quad \text{for any} \quad n \in \mathbb{N}, n \geq 1.$$  

The word $$u_\Theta(\Delta) = \lim_{n \to \infty} w_n$$ is called a generalized pseudostandard word with the directive sequence of letters $\Delta$ and the directive sequence of antimorphisms $\Theta$.

Let us stress that the definition of $u_\Theta(\Delta)$ is correct as $w_n$ is a prefix of $w_{n+1}$ for any $n$. Moreover, if $\Theta = R^\omega$, the word $u_\Theta(\Delta)$ is standard in the usual sense.

**Example 7.** Consider the directive sequence of letters $\Delta = 0(101)^\omega$ and the directive sequence of antimorphisms $\Theta = (RE)^\omega$. Then

$$w_0 = \varepsilon$$
$$w_1 = 0^R = 0$$
$$w_2 = (01)^E = 01$$
$$w_3 = (010)^R = 010$$
$$w_4 = (0101)^E = 0101$$
$$w_5 = (01011)^R = 01011010$$
\[w_6 = (010110100)^E = 010110100101\]

The authors of [7] proved that the famous Thue–Morse word \(u_{TM}\) is a generalized pseudostandard word with directive sequences
\[\Delta = 01^\omega \quad \text{and} \quad \Theta = (ER)^\omega.\]

The article [11] is devoted to the generalized Thue–Morse words. As already mentioned, they are \(I_2(m)\)-rich, where \(I_2(m)\) is the group generated by anitmorphisms \(\Psi_x\) described by (1). We proved the following theorem.

**Theorem 8.** The generalized Thue-Morse word \(t_{b,m}\) is a generalized pseudostandard word if and only if \(b \leq m\) or \(b \mod m = 1\). If this is the case, then the corresponding sequences \(\Delta\) and \(\Theta\) are
\[\Delta = 0(12\ldots(b-1))^\omega \in \mathbb{Z}_m^\omega \quad \text{and} \quad \Theta = (\Psi_0\Psi_1\ldots\Psi_{m-1})^\omega \in I_2(m)^\omega.\]

In [5] pseudostandard words over binary alphabet are studied and then the results are applied to standard CSR words. Let us recall that a CSR word \(u\) is standard, if both \(0u\) and \(1u\) are CRS words too. The authors of [5] proved

**Theorem 9.** For any standard complementary-symmetric Rote word \(u\) there exist \(\Theta \in \{E, R\}^\mathbb{N}\) and \(\Delta \in \{0,1\}^\mathbb{N}\) such that \(u = u_0(\Delta)\).

The key idea used in the proof of the previous theorem is the connection between standard CSR words and standard Sturmian words. Rote in [16] deduced the following theorem.

**Theorem 10.** An infinite word \(u = u_0u_1u_2\ldots\) over the alphabet \(\{0,1\}\) is a complementary-symmetric Rote word if and only if the word \(v = v_1v_2v_3\ldots\) defined by \(v_i = u_{i-1} + u_i \mod 2\) for each \(i = 1,2,\ldots\) is a standard Sturmian word.

The operation \(v_i = u_{i-1} + u_i \mod 2\) applied to a \(H\)-rich Rote word gives us \(R\)-rich Sturmian word. The same observation holds true for the Thue-Morse word \(u_{TM}\), which is \(H\)-rich. After application of the operation we get the period doubling word \(u_{PD}\), which is \(R\)-rich. The period doubling word can be defined as the fixed point of the substitution \(\varphi_{PD}\) determined by \(0 \mapsto 11\) and \(1 \mapsto 10\).

The previous examples inspired us to investigate relationship between the generalized richness of words and their images under the operation, which we extended to a larger alphabet as well.

### 5 The operation \(S\)

Let \(m\) be an integer greater than 1. In this section we consider the alphabet \(A = \mathbb{Z}_m = \{0,\ldots,m-1\}\) and the operation \(S : A^\mathbb{N} \rightarrow A^\mathbb{N}\) defined by
\[S(w_0w_1w_2\cdots) = v_1v_2\ldots, \quad \text{where} \quad v_i = (w_{i-1} + w_i) \mod m \quad \text{for every} \quad i \in \mathbb{N}, i \geq 1. \quad (2)\]

First we concentrate on the binary alphabet \(\{0,1\}\). In [14], we showed the two following statements.

**Theorem 11.** Let the language of \(u \in \{0,1\}^\mathbb{N}\) be closed under the group \(H = \{\text{Id}, E, R, ER\}\). The word \(u\) is \(H\)-rich if and only if \(S(u)\) is \(R\)-rich.
Theorem 12. Let $u \in \{0, 1\}^\mathbb{N}$ be a uniformly recurrent word. If $u$ is almost $R$-rich, then for all $k > 0$ the word $S^k(u)$ is almost $R$-rich.

The multiliteral alphabet $\mathbb{Z}_m$ allows many finite groups generated by involutory antimorphisms. We restrict our attention to the groups $I_2(m)$. 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.

We focus on images of $t_{b,m}$ by $S$ with parameters $b \geq 3$ and $m \geq 3$. We denote by $I_2'(m)$ the group generated by the antimorphisms $\{\Psi_{2y}: y \in \mathbb{Z}_m\}$.

If $m$ is odd, then $I_2'(m) = I_2(m)$, if $m$ is even, then $I_2'(m)$ is isomorphic to $I_2(m_2)$.

In [14] we demonstrated a weaker multiliteral analogue of Theorem 11.

Theorem 13. Let $m, b \in \mathbb{Z}$ such that $m \geq 3$ and $b \geq 3$.

1. The word $S(t_{b,m})$ is almost $I_2'(m)$-rich.
2. If $m$ or $b$ is odd, the word $S(t_{b,m})$ is $I_2'(m)$-rich.

Iterating the operation $S$ one can produce (almost) $G$-rich words for different groups $G$ as illustrated by the following examples.

Example 14. Let $b \in \mathbb{N}, b \geq 2$ and $S$ be the operation defined for the alphabet $\mathbb{Z}_4$. Then

- $t_{2b+1,4}$ is an infinite word over the alphabet $\{0, 1, 2, 3\}$ and it is $I_2(4)$-rich.
- $S(t_{2b+1,4})$ 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 $\{1, 3\}$).
- $S^2(t_{2b+1,4})$ is an infinite word over the binary alphabet $\{0, 2\}$ and it is $R$-rich.
- $S^k(t_{2b+1,4})$ is an infinite word over the binary alphabet $\{0, 2\}$ and it is almost $R$-rich for any $k \in \mathbb{N}, k \geq 2$.

Example 15. Let $b \in \mathbb{N}, b \geq 2$ and $S$ be the operation defined for the alphabet $\mathbb{Z}_2$. Then

- $t_{2b+1,2}$ is an infinite binary word and it is $H$-rich.
- $S(t_{2b+1,2})$ is an infinite binary word and it is $R$-rich.
- $S^k(t_{2b+1,2})$ is an infinite binary word and it is almost $R$-rich for any $k \in \mathbb{N}, k \geq 2$.

6 Binary projections

The examples concluding the previous section provided us a new class of binary words which are $H$-rich and also a new class of binary $R$-rich words, i.e., words rich in the classical sense. These words originated in the generalized Thue–Morse words. In this section we restrict ourselves again to a binary alphabet and we describe a procedure how to construct a new class of $H$-rich and $R$-rich words. Paradoxically, the procedure exploits ternary Arnoux-Rauzy words.
Definition 16. Let $A = \{A, B, C\}$ be an alphabet and $x \in A$. A morphism $\zeta : A \to \{0, 1\}$ defined by

$$
\zeta : a \mapsto \begin{cases} 
0 & \text{if } a \neq x, \\
1 & \text{otherwise.}
\end{cases}
$$

is called binary projection over $A$.

As a binary projection depends on the choice of $x \in A$, we have 3 distinct binary projections over $A$. The procedure we promised to give is based on the following statements which can be found in [14].

Theorem 17. Let $u$ be an Arnoux–Rauzy word over the ternary alphabet $A = \{A, B, C\}$, $\zeta$ be a binary projection over $A$ and $S$ be the operation defined by (2) for the parameter $m = 2$. Then

1. the binary word $\zeta(u)$ is $R$-rich;
2. any preimage of $\zeta(u)$ by $S$ is $H$-rich;
3. the image of $\zeta(u)$ by $S$ is $R$-rich.

7 Conclusions

We introduced the classical palindromic richness in the broader context of $G$-richness and demonstrated that it can provide other fruitful points of view at this problem. However, it remains to solve the main question: Does there exist a $G$-rich word for any finite group $G$ generated by involutory antimorphisms over $A^*$? We expect a positive answer.

Moreover, we expect that there exist more operations analogous to the operation $S$ which map $G'$-rich words to $G''$-rich words. To find them we need to have at our disposal more examples of $G'$-rich words for different groups $G$. Because of this task, we are convinced that the generalized palindromic closure as introduced by de Luca and De Luca deserves much more attention. Although language of generalized pseudostandard word is closed under a group $G$ generated by antimorphisms occurring in the directive sequence $\Theta$, unlike the classical richness and surprisingly, these words are not necessarily $G$-rich.

Acknowledgments

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.

References


