Some generalizations of episturmian words and morphisms

Michelangelo Bucci, Aldo de Luca, and Alessandro De Luca

Dipartimento di Matematica e Applicazioni, Università di Napoli Federico II
Via Cintia, Monte S. Angelo, I-80126 Napoli, Italy
{micbucci,aldo.deluca,alessandro.deluca}@unina.it

Abstract

We study some classes of infinite words that generalize standard episturmian words, defined by replacing the reversal operator with an arbitrary involutory antimorphism ϑ of A^*. An analysis of the relations occurring among such classes of words, and of the morphisms connecting them to standard episturmian words, is given. In particular, we analyse some structural properties of standard ϑ-episturmian words and their characteristic morphisms.

1 Introduction

The study of combinatorial and structural properties of finite and infinite words is a subject of great interest, with many applications in mathematics, physics, computer science, and biology (see for instance [10, 11]). In this framework, *Sturmian words* play a central role (see [10, Chap. 2]). Some natural extensions of Sturmian words to the case of an alphabet with more than two letters have been given in [5, 8], introducing the class of the so-called *episturmian words*.

We recall that for an infinite word $t \in A^\omega$, the following conditions are equivalent (see [5, 8]):

1. There exists an infinite word $\Delta = x_1 x_2 \cdots x_n \cdots \in A^\omega$ such that
$$t = \lim_{n\to\infty} \psi(x_1 \cdots x_n) ,$$
with ψ defined recursively as $\psi(\varepsilon) = \varepsilon$ and $\psi(wa) = (\psi(w)a)^{(+)}$ for all $w \in A^*$ and $a \in A$, where ${}^{(+)}$ is the right palindrome closure operator. We recall that for any $v \in A^*$, $v^{(+)}$ denotes the shortest palindrome having v as a prefix.

2. t is closed under reversal, and each of its left special factors is a prefix of t.

Infinite words satisfying such conditions are called *standard episturmian words.* The class of such words is denoted by $SEpi(A)$, or simply $SEpi$. An infinite word $s \in A^\omega$ is called *episturmian* if there exists a standard episturmian word t having the same set of factors as s. In this paper, we consider only extensions of *standard* episturmian words, since the corresponding generalizations of episturmian words can then be naturally obtained.

Several extensions of standard episturmian words are possible. For example, in [6] a generalization was obtained by making suitable hypotheses on the lengths of palindromic prefixes of an infinite word.

We will consider here different extensions, some of which were introduced in previous papers with L. Q. Zamboni [4, 3, 2]. All these generalizations are based on the replacement of the reversal operator R by an arbitrary *involutory antimorphism* ϑ of the free monoid A^*. By making such a replacement in definitions 1 and 2 above, one obtains two classes of infinite words, called respectively *ϑ-standard* and *standard ϑ-episturmian.* These classes, denoted by E1 and E2, are no longer equivalent nor comparable (i.e., neither of the two is contained in the other one). In Section 3, we give two natural generalizations of E1 and E2, and show that they coincide. The words in the larger class thus obtained are called *ϑ-standard with seed.* An analysis of the relations occurring among such classes of words, and of the morphisms connecting them to standard episturmian words, is given.

In Section 4, we consider some structural properties of standard ϑ-episturmian words, and introduce *ϑ-characteristic morphisms*, i.e., morphisms mapping every standard episturmian word on an alphabet X to standard ϑ-episturmian words over some alphabet A. Such morphisms are an obvious extension of episturmian morphisms and a powerful tool to construct nontrivial examples of standard ϑ-episturmian words (cf. [1]). We study some properties of ϑ-characteristic morphisms and give a characterization of them in a special case (cf. Theorem 4.6).

For definitions not given explicitly in the text, the reader is referred to [10, 4].

2 Different extensions of episturmian words

As is well known, an *involutory antimorphism* of the free monoid A^* is any map $\vartheta : A^* \to A^*$ such that $\vartheta(uv) = \vartheta(v)\vartheta(u)$ for any $u, v \in A^*$, and $\vartheta \circ \vartheta =$ id. Involutory antimorphisms naturally arise also in some applications; a famous example is the Watson and Crick antimorphic involution in molecular biology (see for instance [9]).

The reversal operator is the basic example of involutory antimorphism of A^*. In fact, any involutory antimorphism is the composition $\vartheta = \tau \circ R = R \circ \tau$ where τ is an involutory permutation of the alphabet A. Thus it makes sense to call *ϑ-palindromes* the fixed points of an involutory antimorphism ϑ. We denote by PAL_ϑ the set of ϑ-palindromes over the alphabet A. The set PAL_R coincides with the set of usual palindromes over A.

Let ϑ be an involutory antimorphism of A^*. One can define the (right) ϑ-palindrome closure operator: for any $w \in A^*$, $w^{\oplus_\vartheta}$ denotes the shortest ϑ-palindrome having w as a prefix.

In the following, we shall fix an involutory antimorphism ϑ of A^*, and use the notation $\bar{w}$ for $\vartheta(w)$. We also drop the subscript ϑ from the ϑ-palindrome closure operator ${}^{\oplus_\vartheta}$ when no confusion arises. As can be easily verified (cf. [4]), if Q is the longest ϑ-palindromic suffix of w and $w = sQ$, then

$$w^{\oplus} = sQ\bar{s} \ .$$

We can naturally define the *iterated ϑ-palindrome closure* operator $\psi_\vartheta : A^* \to PAL_\vartheta$ by $\psi_\vartheta(\varepsilon) = \varepsilon$ and

$$\psi_\vartheta(ua) = (\psi_\vartheta(u)a)^{\oplus}$$

for $u \in A^*$, $a \in A$. For any $u, v \in A^*$ one has $\psi_\vartheta(uv) \in \psi_\vartheta(u)A^* \cap A^*\psi_\vartheta(u)$, so that ψ_ϑ can be extended to infinite words too. More precisely, if $\Delta = x_1x_2\cdots x_n\cdots \in A^\omega$ with $x_i \in A$ for $i \geq 1$, then

$$\psi_\vartheta(\Delta) = \lim_{n\to\infty} \psi_\vartheta(x_1\cdots x_n) \ .$$

The word Δ is called the *directive word* of $\psi_\vartheta(\Delta)$.

We give the following two natural extensions of standard episturmian words by replacing the operator R with an arbitrary ϑ in definitions 1 and 2 considered in the introduction:

- E1 – Let $\Delta \in A^\omega$. We call $s = \psi_\vartheta(\Delta)$ the *ϑ-standard word* directed by Δ.

- E2 – An infinite word over A satisfying the two following requirements:

  1. w is closed under ϑ,
  2. any left special factor of w is a prefix of w,

  is called a *standard ϑ-episturmian* word.

The class of standard ϑ-episturmian words is denoted by $SEpi_\vartheta$. Clearly, R-standard and standard R-episturmian words coincide with standard episturmian words.

Conditions E1 and E2 generalize respectively conditions 1 and 2 given in the introduction, but they are not equivalent when $\vartheta \neq R$, as shown in the following examples.

- $E1 \not\Rightarrow E2$: Let $A = \{a, b\}$, $\bar{a} = b$, $\Delta(s) = (ab)^\omega$, and

$$s = abbaababbaabbaab \cdots \ . \tag{1}$$

The words b and ba are two left special factors of s which are not prefixes.

- $E2 \not\Rightarrow E1$: Let $A = \{a, b, c, d, e\}$, $\bar{a} = b$, $\bar{c} = c$, $\bar{d} = e$, and $s = \mu(t)$, where $t = aabaaabaaabaab \cdots \in SEpi$, $\Delta(t) = (aab)^\omega$, $\mu(a) = acb$, and $\mu(b) = de$, so that

$$s = acbacbdeacbacbacbde \cdots \ . \tag{2}$$

As one can easily verify (for instance using Theorem 4.6), the word s is standard ϑ-episturmian but it is not ϑ-standard, as $a^\oplus = ab \notin \mathrm{Pref}(s)$.

The family of ϑ-standard words has been introduced in [4], and several results have been found. In particular, we recall the following two theorems (from [4] and [3] respectively):

Theorem 2.1. *For any $w \in A^\infty$, one has $\psi_\vartheta(w) = \mu_\vartheta(\psi(w))$, where μ_ϑ is the injective morphism defined for any letter $a \in A$ as $\mu_\vartheta(a) = a^\oplus$.*

By the preceding theorem, there exists an injective morphism μ_ϑ such that any ϑ-standard word is the morphic image under μ_ϑ of a standard episturmian word.

Theorem 2.2. *Let w be a left special factor of a ϑ-standard word t. If $|w| \geq 3$, then w is a prefix of t.*

3 Further extensions

In [2], two furher generalizations of standard episturmian words have been considered. They are defined by two properties which extend respectively conditions E1 and E2:

- F1 – We introduce the map $\hat{\psi}_\vartheta : A^* \to A^*$ defined as $\hat{\psi}_\vartheta(\varepsilon) = u_0$, where u_0 is an arbitrary fixed word of A^* called *seed*, and

$$\hat{\psi}_\vartheta(ua) = \left(\hat{\psi}_\vartheta(u)a\right)^\oplus$$

for $u \in A^*$ and $a \in A$. As usual, we can extend this definition to any infinite word $\Delta = x_1 x_2 \cdots x_n \cdots \in A^\omega$ (with $x_i \in A$ for $i \geq 1$) by:

$$\hat{\psi}_\vartheta(\Delta) = \lim_{n \to \infty} \hat{\psi}_\vartheta(x_1 \cdots x_n) \ .$$

The word $\hat{\psi}_\vartheta(\Delta)$ is called the *ϑ-standard word with seed* u_0 directed by Δ. When the seed u_0 is empty, one has $\hat{\psi}_\vartheta = \psi_\vartheta$ so that one obtains ϑ-standard words.

- F2 – We introduce for any $N \geq 0$, the class $SW_\vartheta(N)$ of all infinite words s such that:

  1. s is closed under ϑ,
  2. every left special factor of s whose length is at least N is a prefix of s.

  A word s satisfies F2 if $s \in SW_\vartheta = \bigcup_{n\geq 0} SW_\vartheta(N)$.

We observe that $SW_\vartheta(0) = SEpi_\vartheta$. Moreover, by Theorem 2.2, the class of ϑ-standard words is included in $SW_\vartheta(3)$.

The previous extensions of E1 and E2 are proper. In fact, for instance any R-standard word with seed abb is not standard episturmian, and the word s considered in (1) is in $SW_\vartheta(3)$ but it is not standard ϑ-episturmian.

The following theorem, proved in [2], shows that the two generalizations defined by F1 and F2 coincide.

Theorem 3.1. *Let $s \in A^\omega$. The following conditions are equivalent:*

1. *$s \in SW_\vartheta$,*
2. *s has infinitely many ϑ-palindromic prefixes, and if $(B_n)_{n>0}$ is the sequence of all its ϑ-palindromic prefixes ordered by increasing length, there exists an integer h such that*

$$B_{n+1} = (B_n x_n)^\oplus \ ,$$

   for all $n \geq h$, for a suitable letter x_n,

3. *s is a ϑ-standard word with seed.*

Figure 1 shows the relations among the classes introduced so far. A characterization of the intersection (that is, of words satisfying both E1 and E2) is given by the next proposition (cf. [2]):

Proposition 3.2. *Let s be a ϑ-standard word over A, and $B = \mathrm{alph}(\Delta(s))$. Then s is standard ϑ-episturmian if and only if*

$$x \in B,\, x \neq \bar{x} \implies \bar{x} \notin B \ .$$

Example 3.3. Let $A = \{a, b, c, d, e\}$, $\Delta = (acd)^\omega$, and ϑ be defined by $\bar{a} = b$, $\bar{c} = c$, and $\bar{d} = e$. The ϑ-standard word $\psi_\vartheta(\Delta) = abcabdeabcaba\cdots$ is standard ϑ-episturmian.

Figure 1: Generalized episturmian words

4 Standard ϑ-episturmian words and morphisms

We begin this section with some results concerning the structure of standard ϑ-episturmian words. We denote by $\mathcal{P}_\vartheta$, or simply $\mathcal{P}$, the set of unbordered ϑ-palindromes. We remark that $\mathcal{P}$ is a *biprefix code*. This means that every word of $\mathcal{P}$ is neither a prefix nor a suffix of any other element of $\mathcal{P}$. The following result was proved in [2]:

Proposition 4.1. $PAL_\vartheta^* = \mathcal{P}^*$.

This can be equivalently stated as follows: every ϑ-palindrome can be uniquely factorized by the elements of $\mathcal{P}$.

For any nonempty word w, we denote by w^f its first letter. Since $\mathcal{P}$ is a code, the map

$$\begin{aligned} \nu : \mathcal{P} &\longrightarrow A \\ \pi &\longmapsto \pi^f \end{aligned} \tag{3}$$

can be extended (uniquely) to a morphism $\nu : \mathcal{P}^* \to A^*$. Moreover, since $\mathcal{P}$ is a prefix code, any word in $\mathcal{P}^\omega$ can be uniquely factorized by the elements of $\mathcal{P}$, so that ν can be naturally extended to $\mathcal{P}^\omega$.

Every standard ϑ-episturmian word has infinitely many ϑ-palindromic prefixes. Let then $(B_n)_{n\geq 1}$ be the sequence of all ϑ-palindromic prefixes of a standard ϑ-episturmian word s, ordered by increasing length, and $\Delta = x_1x_2\cdots x_n\cdots$ ($x_i \in A$ for $i \geq 1$) be such that $B_nx_n \in \mathrm{Pref}(s)$ for all $n \geq 1$. Then Δ is called the *subdirective word* of s.

By Proposition 4.1, any standard ϑ-episturmian word s admits a (unique) *canonical* factorization by the elements of $\mathcal{P}$, that is,

$$s = \pi_1\pi_2\cdots\pi_n\cdots ,$$

where $\pi_i \in \mathcal{P}$ for $i \geq 1$. For instance, in the case of the standard ϑ-episturmian word s considered in (2), the canonical decomposition of s is

$$s = acb.acb.de.acb.acb.acb.de \cdots .$$

The following theorem was proved in [2]:

Theorem 4.2. *Let s be a standard ϑ-episturmian word, and*

$$\Pi_s = \{\pi_n \mid n \geq 1\}$$

be the set of words appearing in its canonical factorization $s = \pi_1 \pi_2 \cdots \pi_n \cdots$. Then $\nu(s)$ is a standard episturmian word, and the restriction of ν to Π_s is injective, i.e., if π_i and π_j occur in the factorization of s over $\mathcal{P}$, and $\pi_i^f = \pi_j^f$, then $\pi_i = \pi_j$.

As a consequence, we obtain:

Corollary 4.3. *Let $s \in A^\omega$ be a standard ϑ-episturmian word, Δ be its subdirective word, and $B = \mathrm{alph}(\Delta)$. There exists an injective morphism $\mu : B^* \to A^*$ such that $s = \mu(\psi(\Delta))$ and $\mu(B) \subseteq \mathcal{P}$.*

We observe that the morphism μ is just the inverse of the restriction of ν to Π_s and B.

The previous result shows that any standard ϑ-episturmian word s is a morphic image of a standard episturmian word. However, differently from the ϑ-standard case (see Theorem 2.1), the morphism μ depends on s. Hence, it is interesting to consider morphisms which map standard episturmian words on an alphabet X into standard ϑ-episturmian words over A. More precisely, we give the following definition: a morphism $\varphi : X^* \to A^*$ will be called *ϑ-characteristic* if

$$\varphi(SEpi(X)) \subseteq SEpi_\vartheta .$$

For each $a \in A$, let $\mu_a : A^* \to A^*$ be the morphism defined by $\mu_a(a) = a$ and $\mu_a(b) = ab$ for all $b \in A \setminus \{a\}$. If $w = a_1 \cdots a_n$, we set $\mu_w = \mu_{a_1} \circ \cdots \circ \mu_{a_n}$ (in particular, $\mu_\varepsilon = \mathrm{id}_A$). Any morphism μ_w with $w \in A^*$ is called *pure standard episturmian.* We recall (cf. [5, 7, 8]) that a *standard episturmian morphism* of A^* is any composition $\mu_w \circ \sigma$, with $w \in A^*$ and $\sigma : A^* \to A^*$ a morphism extending a permutation on the alphabet A.

We observe that an injective morphism $\varphi : A^* \to A^*$ is standard episturmian (cf. [5, 8]) if and only if it is R-characteristic.

A ϑ-characteristic morphism $\varphi : X^* \to A^*$ satisfies several necessary conditions. The next two propositions summarize some of them (proved in [1]).

Proposition 4.4. *Let $\varphi : X^* \to A^*$ be a ϑ-characteristic morphism. The following holds:*

1. $\varphi(X) \subseteq PAL_{\vartheta}^2$,

2. *if there exist two letters* $x, y \in X$ *such that* $\varphi(x)^f \neq \varphi(y)^f$, *then* $\varphi(X) \subseteq PAL_{\vartheta}$.

Moreover, in the case of an injective morphism we have:

Proposition 4.5. *Let* $\varphi : X^* \to A^*$ *be an injective* ϑ*-characteristic morphism. Then:*

1. $\varphi(X)$ *is a suffix code,*

2. *if* $\varphi(X) \subseteq PAL_{\vartheta}$, *then for each* $x, y \in X$ *with* $x \neq y$ *one has*

$$\mathrm{alph}(\varphi(x)) \cap \mathrm{alph}(\varphi(y)) = \emptyset \ ,$$

3. *if* $\varphi(X) \subseteq PAL_{\vartheta}$ *and* $\mathrm{card}\, X \geq 2$, *then* $\varphi(X) \subseteq \mathcal{P}$.

As a consequence of the previous propositions, and from a result proved in [2], we obtain the following theorem giving a characterization of injective ϑ-characteristic morphisms $\varphi : X^* \to A^*$ such that $\varphi(X) \subseteq \mathcal{P}$.

Theorem 4.6. *Let* $\varphi : X^* \to A^*$ *be an injective morphism such that for any* $x \in X$, $\varphi(x) \in \mathcal{P}$. *Then* φ *is* ϑ*-characteristic if and only if the following two conditions hold:*

1. $\mathrm{alph}(\varphi(x)) \cap \mathrm{alph}(\varphi(y)) = \emptyset$, *for any* x, y *in* X *with* $x \neq y$.

2. *for any* $x \in X$ *and* $a \in A$, $|\varphi(x)|_a \leq 1$.

Remark. In the "if" part of the above theorem we can replace the requirement $\varphi(X) \subseteq \mathcal{P}$ by $\varphi(X) \subseteq PAL_{\vartheta}$, as this, together with condition 2, implies $\varphi(X) \subseteq \mathcal{P}$. In the "only if" part, in view of Proposition 4.5, one can replace $\varphi(X) \subseteq \mathcal{P}$ by $\varphi(X) \subseteq PAL_{\vartheta}$ under the hypothesis that $\mathrm{card}\, X \geq 2$.

Example 4.7. Let $X = \{x, y\}$, $A = \{a, b, c, d, e\}$, and ϑ be the antimorphism of A^* such that $\bar{a} = b$, $\bar{c} = c$, $\bar{d} = e$. Then the morphism $\varphi : X^* \to A^*$ defined by $\varphi(x) = acb$ and $\varphi(y) = de$ is ϑ-characteristic.

As an immediate consequence of the Theorem 4.6, we obtain:

Corollary 4.8. *Let* $\zeta : X^* \to B^*$ *be an R-characteristic morphism, and* $\mu : B^* \to A^*$ *be an injective morphism satisfying* $\mu(B) \subseteq \mathcal{P}$ *and the two conditions in the statement of Theorem 4.6. Then* $\varphi = \mu \circ \zeta$ *is* ϑ*-characteristic.*

Example 4.9. Let X, A, ϑ, and φ be defined as in Example 4.7, and let ζ be the endomorphism of X^* such that $\zeta(x) = xy$ and $\zeta(y) = xyx$. Since $\zeta = \mu_{xy} \circ \sigma$, where $\sigma(x) = y$ and $\sigma(y) = x$, the morphism ζ is standard episturmian. Hence the morphism $\alpha : X^* \to A^*$ given by

$$\alpha(x) = acbde, \quad \alpha(y) = acbdeacb$$

is ϑ-characteristic, as $\alpha = \varphi \circ \zeta$.

By making use of Corollary 4.3, we proved in [1] the following:

Theorem 4.10. *Let $\varphi : X^* \to A^*$ be a ϑ-characteristic morphism. Then there exist $B \subseteq A$, a morphism $\zeta : X^* \to B^*$, and a morphism $\mu : B^* \to A^*$ such that:*

1. $\zeta(SEpi(X)) \subseteq SEpi(B)$,
2. $\mu(B) \subseteq \mathcal{P}$,
3. $\varphi = \mu \circ \zeta$.

We observe that the morphism ζ appearing in the previous theorem is R-characteristic. The next proposition gives a full characterization, when $\operatorname{card} X \geq 2$, of R-characteristic morphisms:

Proposition 4.11. *Let $\zeta : X^* \to A^*$ be an injective morphism, with $\operatorname{card} X \geq 2$. Then ζ is R-characteristic if and only if it can be decomposed as $\zeta = \mu_w \circ \eta$, where $\mu_w : A^* \to A^*$ is a pure standard episturmian morphism ($w \in A^*$) and $\eta : X^* \to A^*$ is an injective literal morphism.*

From the preceding proposition and Theorem 4.10, it follows that every injective ϑ-characteristic morphism $\varphi : X^* \to A^*$ (with $\operatorname{card} X \geq 2$) can be decomposed as

$$\varphi = \mu \circ \mu_w \circ \eta ,$$

where $\eta : X^* \to B^*$ is an injective literal morphism (with $B \subseteq A$), $\mu_w : B^* \to B^*$ is a pure standard episturmian morphism (with $w \in B^*$), and $\mu : B^* \to A^*$ is an injective morphism such that $\mu(b) \in bA^* \cap \mathcal{P}$ for each $b \in B$. In [1], we gave a characterization of injective ϑ-characteristic morphisms of the kind $\varphi = \mu \circ \mu_w$, that is, when $\eta = \mathrm{id}$ in the above formula.

In conclusion, we mention two important problems which remain open:

1. To give a characterization of morphisms which are ϑ-characteristic, in the general case.
2. To determine whether ϑ-characteristic morphisms $\varphi : X^* \to A^*$ are able to produce all standard ϑ-episturmian words, when applied to standard episturmian words over X.

References

[1] M. Bucci, A. de Luca, and A. De Luca. Characteristic morphisms of generalized episturmian words. Manuscript, 2007.

[2] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On different generalizations of episturmian words. Preprint n. 5 Dipartimento di Matematica e Applicazioni "R. Caccioppoli", 2007.

[3] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On some problems related to palindrome closure. *Theoretical Informatics and Applications*, to appear, 2007.

[4] A. de Luca and A. De Luca. Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science*, **362**:282–300, 2006.

[5] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoretical Computer Science*, **255**:539–553, 2001.

[6] S. Fischler. Palindromic prefixes and episturmian words. *Journal of Combinatorial Theory, Series A*, **113**:1281–1304, 2006.

[7] J. Justin. Episturmian morphisms and a Galois theorem on continued fractions. *Theoretical Informatics and Applications*, **39**:207–215, 2005.

[8] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoretical Computer Science*, **276**:281–313, 2002.

[9] L. Kari, S. Konstantinidis, P. Sosík, and G. Thierrin. On hairpin-free words and languages. In C. De Felice and A. Restivo, editors, *Developments in Language Theory*, volume 3572 of *Springer LNCS*, pages 296–307, 2005.

[10] M. Lothaire. *Algebraic Combinatorics on Words.* Cambridge University Press, 2002.

[11] M. Lothaire. *Applied Combinatorics on Words.* Cambridge University Press, 2005.