On a Generalization of Standard Episturmian Morphisms

Michelangelo Bucci, Aldo de Luca, and Alessandro De Luca

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

Abstract. In a recent paper with L. Q. Zamboni the authors introduced the class of ϑ-*episturmian* words, where ϑ is an involutory antimorphism of the free monoid A^*. In this paper, we introduce and study ϑ-*characteristic morphisms*, that is, morphisms which map standard episturmian words into standard ϑ-episturmian words. They are a natural extension of standard episturmian morphisms. The main result of the paper is a characterization of these morphisms when they are injective.

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 [1–3]). In this framework, *Sturmian words* play a central role, as they are the aperiodic infinite words of minimal "complexity" (see [2, Chap. 2]). By definition, Sturmian words are on a binary alphabet; some natural extensions to the case of an alphabet with more than two letters have been given in [4, 5], introducing the class of the so-called *episturmian words.*

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; in [7–10] different extensions were introduced, all based on the replacement of the reversal operator $R : w \in A^* \mapsto \tilde{w} \in A^*$ by an arbitrary *involutory antimorphism* ϑ of the free monoid A^*. In particular, the so called ϑ-*standard* and *standard* ϑ-*episturmian* words were studied.

In this paper we focus on the study of ϑ-*characteristic morphisms*, a natural extension of standard episturmian morphisms, which map all standard episturmian words on an alphabet X to standard ϑ-episturmian words over some alphabet A. Beside being interesting by themselves, such morphisms are a powerful tool for constructing nontrivial examples of standard ϑ-episturmian words and for studying their properties. The main result of this paper is a characterization of injective ϑ-characteristic morphisms (cf. Theorem 3.2). For the sake of brevity, we shall prove here only this theorem; all the other proofs can be found in [11]. For notations and definitions not included here, the reader is referred to [1, 2, 7, 12].

1.1 Standard Episturmian Words and Morphisms

We recall (cf. [4, 5]) that an infinite word $t \in A^\omega$ is *standard episturmian* if it is *closed under reversal* (that is, if $w \in \mathrm{Fact}\, t$ then $\tilde{w} \in \mathrm{Fact}\, t$) and each of its left special factors is a prefix of t. We denote by $SEpi(A)$, or by $SEpi$ when there is no ambiguity, the set of all standard episturmian words over the alphabet A.

Given a word $w \in A^*$, we denote by $w^{(+)}$ its *right palindrome closure*, i.e., the shortest palindrome having w as a prefix (cf. [13]). We define the *iterated palindrome closure* operator $\psi : A^* \to A^*$ by setting $\psi(\varepsilon) = \varepsilon$ and $\psi(va) = (\psi(v)a)^{(+)}$ for any $a \in A$ and $v \in A^*$. From the definition, one easily obtains that the map ψ is injective. Furthermore, for any $u, v \in A^*$, one has $\psi(uv) \in \psi(u)A^* \cap A^*\psi(u)$. The operator ψ can then be naturally extended to A^ω. The following fundamental result was proved in [4]:

Theorem 1.1. *An infinite word t is standard episturmian over A if and only if there exists $\Delta \in A^\omega$ such that $t = \psi(\Delta)$.*

For any $t \in SEpi$, there exists a *unique* Δ such that $t = \psi(\Delta)$. This Δ is called the *directive word* of t. If every letter of A occurs infinitely often in Δ, the word t is called a (standard) *Arnoux-Rauzy word*. In the case of a binary alphabet, an Arnoux-Rauzy word is usually called a *standard Sturmian word* (cf. [2, Chap. 2]).

We report here some properties of the operator ψ which will be useful in the sequel. The first one is known (see for instance [13, 4]).

Proposition 1.2. *For all $u, v \in A^*$, u is a prefix of v if and only if $\psi(u)$ is a prefix of $\psi(v)$.*

Proposition 1.3. *Let $x \in A \cup \{\varepsilon\}$, $w' \in A^*$, and $w \in w'A^*$. Then $\psi(w'x)$ is a factor of $\psi(wx)$.*

The following proposition was proved in [4, Theorem 6].

Proposition 1.4. *Let $x \in A$, $u \in A^*$, and $\Delta \in A^\omega$. Then $\psi(u)x$ is a factor of $\psi(u\Delta)$ if and only if x occurs in Δ.*

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 $a_1, \ldots, a_n \in A$, we set $\mu_w = \mu_{a_1} \circ \cdots \circ \mu_{a_n}$ (in particular, $\mu_\varepsilon = \mathrm{id}_A$). The next formula, proved in [14], shows a connection between these morphisms, called *pure* standard episturmian morphisms (see [5]), and iterated palindrome closure.

Proposition 1.5. *For any $w, v \in A^*$, $\psi(wv) = \mu_w(\psi(v))\psi(w)$.*

By Theorem 1.1, there exists $v \in A^\omega$ such that $t = \psi(v)$, thus, from Proposition 1.5, one easily derives

$$\psi(wv) = \mu_w(\psi(v)) \ . \tag{1}$$

Furthermore, the following holds (cf. [11]):

Corollary 1.6. *For any $t \in A^\omega$ and $w \in A^*$, $\psi(w)$ is a prefix of $\mu_w(t)$.*

We recall (cf. [4, 14, 5]) that a *standard episturmian morphism* is an injective endomorphism φ of A^* such that $\varphi(SEpi) \subseteq SEpi$. As proved in [4], a morphism is standard episturmian if and only if can be written as $\mu_w \circ \sigma$, with $w \in A^*$ and $\sigma : A^* \to A^*$ a morphism extending a permutation on the alphabet A. All these morphisms are injective. The set of all standard episturmian morphisms is a monoid under map composition.

1.2 Involutory Antimorphisms and Pseudopalindromes

An *involutory antimorphism* of A^* is any antimorphism $\vartheta : A^* \to A^*$ such that $\vartheta \circ \vartheta = \mathrm{id}$. The simplest example is the reversal operator. Any involutory antimorphism ϑ satisfies $\vartheta = \tau \circ R = R \circ \tau$ for some morphism $\tau : A^* \to A^*$ extending an involution of A. Conversely, if τ is such a morphism, then $\vartheta = \tau \circ R = R \circ \tau$ is an involutory antimorphism of A^*.

Let ϑ be an involutory antimorphism of A^*. We call *ϑ-palindrome* any fixed point of ϑ, i.e., any word w such that $w = \vartheta(w)$, and denote by PAL_ϑ the set of all ϑ-palindromes. We observe that $\varepsilon \in PAL_\vartheta$ by definition, and that R-palindromes are exactly the usual palindromes. If one makes no reference to the antimorphism ϑ, a ϑ-palindrome is often called a *pseudopalindrome*. Some general properties of pseudopalindromes, have been studied in [7].

In the following, we shall fix an involutory antimorphism ϑ of A^*, and use the notation $\bar{w}$ for $\vartheta(w)$. We denote by $\mathcal{P}_\vartheta$ the set of *unbordered* ϑ-palindromes (i.e., ϑ-palindromes without nontrivial ϑ-palindromic prefixes). We remark that $\mathcal{P}_\vartheta$ is a *biprefix code* (cf. [12]) and that $\mathcal{P}_R = A$. The following result was proved in [9]:

Proposition 1.7. $PAL_\vartheta^* = \mathcal{P}_\vartheta^*$.

This can be equivalently stated as follows: every ϑ-palindrome can be uniquely factorized by the elements of $\mathcal{P}_\vartheta$. For instance, if $\bar{a} = b$ and $\bar{c} = c$, the ϑ-palindrome $abacabcbab$ can be factorized as $ab \cdot acabcb \cdot ab$.

For any nonempty word w, we will denote, from now on, by w^f and w^ℓ respectively the first and the last letter of w. Since $\mathcal{P}_\vartheta$ is a code, the map

$$f : \pi \in \mathcal{P}_\vartheta \longmapsto \pi^f \in A \tag{2}$$

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

Proposition 1.8. *Let $\varphi : X^* \to A^*$ be an injective morphism such that $\varphi(X) \subseteq \mathcal{P}_\vartheta$. Then, for any $w \in X^*$:*

1. $\varphi(\tilde{w}) = \overline{\varphi(w)}$,
2. $w \in PAL \Longleftrightarrow \varphi(w) \in PAL_\vartheta$,

1.3 Overlap-Free and Normal Codes

We say that a code Z over A is *overlap-free* if no two of its elements overlap properly, i.e., if for all $u, v \in Z$, $\mathrm{Suff}\, u \cap \mathrm{Pref}\, v \subseteq \{\varepsilon, u, v\}$.

For instance, let $Z_1 = \{a, bac, abc\}$ and $Z_2 = \{a, bac, cba\}$. One has that Z_1 is an overlap-free suffix code, and Z_2 is a prefix code which is not overlap-free.

Let Z be a subset of A^*; we denote by $LS\, Z$ (resp. $RS\, Z$) the set of all words $u \in \mathrm{Fact}\, Z$ which are *left special* (resp. *right special*) in Z, i.e., such that there exist two distinct letters a and b for which $au, bu \in \mathrm{Fact}\, Z$ (resp. $ua, ub \in \mathrm{Fact}\, Z$).

A code $Z \subseteq A^+$ will be called *right normal* if it satisfies the following condition:

$$(\mathrm{Pref}\, Z \setminus Z) \cap RS\, Z \subseteq \{\varepsilon\} \ , \tag{3}$$

i.e., any proper and nonempty prefix u of any word of Z such that $u \notin Z$ is not right special in Z. In a symmetric way, a code Z is called *left normal* if it satisfies the condition

$$(\mathrm{Suff}\, Z \setminus Z) \cap LS\, Z \subseteq \{\varepsilon\} \ . \tag{4}$$

A code Z is called *normal* if it is right and left normal.

As an example, the code $Z_1 = \{a, ab, bb\}$ is right normal but not left normal; the code $Z_2 = \{a, aba, aab\}$ is normal.

Proposition 1.9. *Let Z be a biprefix, overlap-free, and right normal (resp. left normal) code. Then:*

1. *if $z \in Z$ is such that $z = \lambda v \rho$, with $\lambda, \rho \in A^*$ and v a nonempty prefix (resp. suffix) of $z' \in Z$, then $\lambda z'$ (resp. $z'\rho$) is a prefix (resp. suffix) of z, proper if $z \neq z'$.*
2. *for $z_1, z_2 \in Z$, if $z_1^f = z_2^f$ (resp. $z_1^\ell = z_2^\ell$), then $z_1 = z_2$.*

From the preceding proposition, a biprefix, overlap-free, and normal code satisfies both properties 1 and 2 and their symmetrical statements; all the statements of the following propositions can also be applied to codes which are biprefix, overlap-free, and normal.

Proposition 1.10. *Let Z be a suffix, left normal, and overlap-free code over A, and let $a, b \in A$, $v \in A^*$, $\lambda \in A^+$ be such that $a \neq b$, $va \notin Z^*$, $va\lambda \in \mathrm{Pref}\, Z^*$, and $b\lambda \in \mathrm{Fact}\, Z^*$. Then $a\lambda \in \mathrm{Fact}\, Z$.*

Proposition 1.11. *Let Z be a biprefix, overlap-free, and right normal code over A. If $\lambda \in \mathrm{Pref}\, Z^* \setminus \{\varepsilon\}$, then there exists a unique word $u = z_1 \cdots z_k$ with $k \geq 1$ and $z_i \in Z$, $i = 1, \ldots, k$, such that*

$$u = z_1 \cdots z_k = \lambda \zeta, \quad z_1 \cdots z_{k-1} \delta = \lambda \ , \tag{5}$$

where $\delta \in A^+$ and $\zeta \in A^*$.

In conclusion of this section, we report (cf. [11]) the following simple general lemma on prefix codes, which will be useful in the next sections:

Lemma 1.12. *Let $g : B^* \to A^*$ be an injective morphism such that $g(B) = Z$ is a prefix code. Then for all $p \in B^*$ and $q \in B^\infty$, one has that p is a prefix of q if and only if $g(p)$ is a prefix of $g(q)$.*

1.4 Standard ϑ-Episturmian Words

In [9] *standard ϑ-episturmian* words were naturally defined by substituting, in the definition of standard episturmian words, the closure under reversal with the *closure under ϑ*. Thus an infinite word s is standard ϑ-episturmian if it satisfies the following two conditions:

1. for any $w \in \operatorname{Fact} s$, one has $\bar{w} \in \operatorname{Fact} s$,
2. for any left special factor w of s, one has $w \in \operatorname{Pref} s$.

We denote by $SEpi_\vartheta$ the set of all standard ϑ-episturmian words over A.

The following proposition, proved in one direction in [9] and completely in [11], is a first tool for constructing nontrivial standard ϑ-episturmian words.

Proposition 1.13. *Let $g : X^* \to A^*$ be an injective morphism such that $g(X) \subseteq \mathcal{P}_\vartheta$ for a fixed ϑ. Then $g(SEpi(X)) \subseteq SEpi_\vartheta(A)$ if and only if each letter of* $\operatorname{alph} g(X)$ *appears exactly once in $g(X)$.*

Example 1.14. Let $A = \{a, b, c, d, e\}$, $\bar{a} = b$, $\bar{c} = c$, $\bar{d} = e$, $X = \{a, b\}$, and $s = g(t)$, where $t = aabaaabaaabaab \cdots \in SEpi(X)$, $\Delta(t) = (aab)^\omega$, $g(a) = acb$, and $g(b) = de$, so that

$$s = acbacbdeacbacbacbde \cdots \quad . \tag{6}$$

Proposition 1.13 ensures that g maps $SEpi(X)$ into a subset of $SEpi_\vartheta(A)$, thus s is standard ϑ-episturmian.

In the following, for a given standard ϑ-episturmian word s we shall denote by

$$\Pi_s = \{\pi_n \mid n \geq 1\} \tag{7}$$

the set of words of $\mathcal{P}_\vartheta$ appearing in its unique factorization $s = \pi_1 \pi_2 \cdots$ in unbordered ϑ-palindromes.

The details of the proof of the following useful theorem can be found in [11].

Theorem 1.15. *Let $s \in SEpi_\vartheta$. Then Π_s is an overlap-free and normal code.*

Since for $s \in SEpi_\vartheta$, Π_s is a biprefix, overlap-free, and normal code, by Proposition 1.8, Proposition 1.9, and Lemma 1.12 one can derive the following theorem, proved in [9, Theorem 5.5] in a different way, which shows in particular that any standard ϑ-episturmian word is a morphic image, by a suitable injective morphism, of a standard episturmian word.

Theorem 1.16. *Let s be a standard ϑ-episturmian word. Then $f(s)$ is a standard episturmian word, and the restriction of f to Π_s is injective, i.e., if π_i and π_j occur in the factorization of s over $\mathcal{P}_\vartheta$, and $\pi_i^f = \pi_j^f$, then $\pi_i = \pi_j$.*

2 Characteristic Morphisms

Let X be a finite alphabet. A morphism $\varphi : X^* \to A^*$ will be called *ϑ-characteristic* if $\varphi(SEpi(X)) \subseteq SEpi_\vartheta$, i.e., φ maps any standard episturmian word over the alphabet X in a standard ϑ-episturmian word on the alphabet A. With this terminology, we observe that an injective morphism $\varphi : X^* \to X^*$ is standard episturmian if and only if it is R-characteristic. A trivial example of a non-injective ϑ-characteristic morphism is the constant morphism $\varphi : x \in X \mapsto a \in A$, where a is a fixed ϑ-palindromic letter; furthermore Proposition 1.13 provides an easy way of constructing injective ϑ-characteristic morphisms, like the one used in Example 1.14.

Let $X = \{x, y\}$, $A = \{a, b, c\}$, ϑ defined by $\bar{a} = a$, $\bar{b} = c$, and $\varphi : X^* \to A^*$ be the injective morphism such that $\varphi(x) = a$, $\varphi(y) = bac$. If t is any standard episturmian word beginning in y^2x, then $s = \varphi(t)$ begins with $bacbaca$, so that a is a left special factor of s which is not a prefix of s. Thus s is not ϑ-episturmian and therefore φ is not ϑ-characteristic.

A first result (cf. [11]) on the structure of ϑ-characteristic morphisms is given by the following:

Proposition 2.1. *Let $\varphi : X^* \to A^*$ be a ϑ-characteristic morphism. For each x in X, $\varphi(x) \in PAL_\vartheta^2$.*

Let $\varphi : X^* \to A^*$ be a morphism such that $\varphi(X) \subseteq \mathcal{P}_\vartheta^*$. For any $x \in X$, let $\varphi(x) = \pi_1^{(x)} \cdots \pi_{r_x}^{(x)}$ be the unique factorization of $\varphi(x)$ by the elements of $\mathcal{P}_\vartheta$. Set

$$\Pi(\varphi) = \{\pi \in \mathcal{P}_\vartheta \mid \exists x \in X, \exists i : \ 1 \leq i \leq r_x \text{ and } \pi = \pi_i^{(x)}\} \ . \tag{8}$$

If φ is a ϑ-characteristic morphism, then by Propositions 2.1 and 1.7, we have $\varphi(X) \subseteq PAL_\vartheta^2 \subseteq \mathcal{P}_\vartheta^*$, so that $\Pi(\varphi)$ is well defined.

The following important theorem provides a useful decomposition of injective ϑ-characteristic morphisms (see [11] for a proof).

Theorem 2.2. *Let $\varphi : X^* \to A^*$ be an injective ϑ-characteristic morphism. Then $\Pi(\varphi)$ is an overlap-free and normal code. Furthermore φ can be decomposed as*

$$\varphi = g \circ \mu_w \circ \eta \ , \tag{9}$$

where $\eta : X^* \to B^*$ is an injective literal morphism, $B \subseteq A$, $\mu_w : B^* \to B^*$ is a pure standard episturmian morphism (with $w \in B^*$), and $g : B^* \to A^*$ is an injective morphism such that $g(B) = \Pi(\varphi)$. The above decomposition can always be chosen so that $B = \eta(X) \cup \operatorname{alph} w \subseteq A$ and $g(b) \in bA^* \cap \mathcal{P}_\vartheta$ for each $b \in B$.

Example 2.3. Let X, A, ϑ, and g be defined as in Example 1.14 and let $\eta(a) = a, \eta(b) = b$. The morphism φ defined by $\varphi(a) = acbdeacb, \varphi(b) = acbde$ is decomposable as

$$\varphi = g \circ \mu_{ab} \circ \eta \ .$$

Since g is ϑ-characteristic and $\mu_{ab} \circ \eta = \mu_{ab}$ is R-characteristic, it follows that φ is ϑ-characteristic.

Example 2.4. Let $X = \{x, y\}$, $A = \{a, b, c\}$, and ϑ be the antimorphism of A^* such that $\bar{a} = a$ and $\bar{b} = c$. The morphism $\varphi : X^* \to A^*$ defined by $\varphi(x) = a$ and $\varphi(y) = abac$ is ϑ-characteristic (this will be clear after Theorem 3.2, see Example 3.3). It can be decomposed as $\varphi = g \circ \mu_a \circ \eta$, where $\eta : X^* \to B^*$ (with $B = \{a, b\}$) is the morphism such that $\eta(x) = a$ and $\eta(y) = b$, while $g : B^* \to A^*$ is defined by $g(a) = a$ and $g(b) = bac$. We remark that $(\mu_a \circ \eta)(SEpi(X)) \subseteq SEpi(B)$, but from Proposition 1.13 it follows that $g(SEpi(B)) \not\subseteq SEpi_\vartheta$.

Proposition 2.5 (cf. [11]). *Let $\varphi : X^* \to A^*$ be an injective ϑ-characteristic morphism, decomposed as in* (9)*. The word $u = g(\psi(w))$ is a ϑ-palindrome such that for each $x \in X$, $\varphi(x)$ is either a prefix of u or equal to $ug(\eta(x))$.*

3 Main Result

Before proceeding with the main theorem, which gives a characterization of all injective ϑ-characteristic morphisms, we need the following lemma, again proved in [11].

Lemma 3.1. *Let $t \in SEpi(B)$ with $\operatorname{alph} t = B$, and let $s = g(t)$ be a standard ϑ-episturmian word over A, with $g : B^* \to A^*$ an injective morphism such that $g(B) \subseteq \mathcal{P}_\vartheta$. Suppose that $b, c \in A \setminus \operatorname{Suff} \Pi_s$ and $v \in \Pi_s^*$ are such that $bv\bar{c} \in \operatorname{Fact} \Pi_s$. Then there exists $\delta \in B^*$ such that $v = g(\psi(\delta))$.*

Theorem 3.2. *Let $\varphi : X^* \to A^*$ be an injective morphism. Then φ is ϑ-characteristic if and only if it is decomposable as*

$$\varphi = g \circ \mu_w \circ \eta$$

as in (9)*, with $B = \eta(X) \cup \operatorname{alph} w$ and $g(B) = \Pi \subseteq \mathcal{P}_\vartheta$ satisfying the following conditions:*

1. *Π is an overlap-free and normal code,*
2. *$LS\,(\{g(\psi(w))\} \cup \Pi) \subseteq \operatorname{Pref} g(\psi(w))$,*
3. *if $b, c \in A \setminus \operatorname{Suff} \Pi$ and $v \in \Pi^*$ are such that $bv\bar{c} \in \operatorname{Fact} \Pi$, then $v = g(\psi(w'x))$, with $w' \in \operatorname{Pref} w$ and $x \in \{\varepsilon\} \cup (B \setminus \eta(X))$.*

Example 3.3. Let $A = \{a, b, c\}$, $X = \{x, y\}$, $B = \{a, b\}$, and let ϑ and $\varphi : X^* \to A^*$ be defined as in Example 2.4, namely $\bar{a} = a$, $\bar{b} = c$, and $\varphi = g \circ \mu_a \circ \eta$, where $\eta(x) = a$, $\eta(y) = b$, and $g : B^* \to A^*$ is defined by $g(a) = a$ and $g(b) = bac$. Then $\Pi = g(B) = \{a, bac\}$ is an overlap-free, normal code which satisfies condition 2 of Theorem 3.2. The only word verifying the hypotheses of condition 3 is bac, and $bac = ba\bar{b} = g(b) \in \Pi$, with $a \in \Pi^*$ and $b \notin \operatorname{Suff} \Pi$. Since $a = g(\psi(a))$ and $B \setminus \eta(X) = \emptyset$, also condition 3 of Theorem 3.2 is satisfied. Hence φ is ϑ-characteristic.

Example 3.4. Let $X = \{x, y\}$, $A = \{a, b, c\}$, ϑ be such that $\bar{a} = a$, $\bar{b} = c$, and the morphism $\varphi : X^* \to A^*$ be defined by $\varphi(x) = a$ and $\varphi(y) = abaac$. In this case we have $\varphi = g \circ \mu_a \circ \eta$, where $B = \{a, b\}$, $g(a) = a$, $g(b) = baac$,

$\eta(x) = a$, and $\eta(y) = b$. Then the morphism φ is not ϑ-characteristic. Indeed, if t is any standard episturmian word starting with yxy, then $\varphi(t)$ has the prefix $abaacaabaac$, so that aa is a left special factor of $\varphi(t)$ but not a prefix of it. In fact, condition 3 of Theorem 3.2 is not satisfied in this case, since $baac = baa\bar{b} = g(b)$, $b \notin \operatorname{Suff} \Pi$, $aa \in \Pi^*$, $B \setminus \eta(X) = \emptyset$, and

$$aa \notin \{g(\psi(w')) \mid w' \in \operatorname{Pref} a\} = \{\varepsilon, a\} \ .$$

If we choose $X' = \{y\}$ with $\eta'(y) = b$, then

$$g(\mu_a(\eta'(y^\omega))) = (abaac)^\omega \in SEpi_\vartheta \ ,$$

so that $\varphi' = g \circ \mu_a \circ \eta'$ is ϑ-characteristic. In this case $B = \eta'(X') \cup \operatorname{alph} a$, $B \setminus \eta'(X') = \{a\}$, and $aa = g(\psi(aa)) = g(aa)$, so that condition 3 is satisfied.

Example 3.5. Let $X = \{x, y\}$, $A = \{a, b, c, d, e, h\}$, and ϑ be the antimorphism over A defined by $\bar{a} = a$, $\bar{b} = c$, $\bar{d} = e$, $\bar{h} = h$. Let also $w = adb \in A^*$, $B = \{a, b, d\} = \operatorname{alph} w$, and $\eta : X^* \to B^*$ be defined by $\eta(x) = a$ and $\eta(y) = b$. Finally, set $g(a) = a$, $g(d) = dahae$, and $g(b) = badahaeadahaeac$, so that the morphism $\varphi = g \circ \mu_w \circ \eta$ is such that

$$\varphi(y) = adahaeabadahaeadahaeac \quad \text{and} \quad \varphi(x) = \varphi(y)\, adahaea \ .$$

Then φ is ϑ-characteristic, as the code $\Pi(\varphi) = g(B)$ and the word $u = g(\psi(w)) = g(adabada) = \varphi(x)$ satisfy all three conditions of Theorem 3.2.

Note that Proposition 1.13 can be derived as a corollary of Theorem 3.2 (cf. [11]).

Remark 3.6. Let us observe that Theorem 3.2 gives an effective procedure to decide whether, for a given ϑ, an injective morphism $\varphi : X^* \to A^*$ is ϑ-characteristic. The procedure runs in the following steps:

1. Check whether $\varphi(X) \subseteq \mathcal{P}_\vartheta^*$.
2. If the previous condition is satisfied, then compute $\Pi = \Pi(\varphi)$.
3. Verify that Π is overlap-free and normal.
4. Compute $B = f(\Pi)$ and the morphism $g : B^* \to A^*$ given by $g(B) = \Pi$.
5. Since $\varphi = g \circ \zeta$, verify that ζ is R-characteristic, i.e., there exists $w \in B^*$ such that $\zeta = \mu_w \circ \eta$, where η is a literal morphism from X^* to B^*.
6. Compute $g(\psi(w))$ and verify that conditions 2 and 3 of Theorem 3.2 are satisfied. This can be effectively done.

Proof (Theorem 3.2). **Necessity:** From Theorem 2.2, we obtain the decomposition (9) where $B = \eta(X) \cup \operatorname{alph} w$ and $g(B) = \Pi(\varphi) \subseteq \mathcal{P}_\vartheta$ is an overlap-free and normal code.

Let us set $u = g(\psi(w))$, and prove that condition 2 holds. We first suppose that $\operatorname{card} X \geq 2$, and that $a, a' \in \eta(X)$ are distinct letters. Let Δ be an infinite word such that $\operatorname{alph} \Delta = \eta(X)$. Setting $t_a = \psi(wa\Delta)$ and $t_{a'} = \psi(wa'\Delta)$, by (1) we have

$$t_a = \mu_w(\psi(a\Delta)) \quad \text{and} \quad t_{a'} = \mu_w(\psi(a'\Delta)) \ ,$$

so that, setting $s_y = g(t_y)$ for $y \in \{a, a'\}$, we obtain

$$s_y = g(\mu_w(\psi(y\Delta))) \in SEpi_\vartheta$$

as $\psi(y\Delta) \in \eta(SEpi(X)) \subseteq SEpi(B)$ and $\varphi = g \circ \mu_w \circ \eta$ is ϑ-characteristic. By Corollary 1.6 and (1), one obtains that the longest common prefix of t_a and $t_{a'}$ is $\psi(w)$. As $\operatorname{alph} \Delta = \eta(X)$ and $B = \eta(X) \cup \operatorname{alph} w$, we have $\operatorname{alph} t_a = \operatorname{alph} t_{a'} = B$, so that $\Pi_{s_a} = \Pi_{s_{a'}} = \Pi$. Since g is injective, by Theorem 1.16 we have $g(a)^f \neq g(a')^f$, so that the longest common prefix of s_a and $s_{a'}$ is $u = g(\psi(w))$. Any word of $LS(\{u\} \cup \Pi)$, being a left special factor of both s_a and $s_{a'}$, has to be a common prefix of s_a and $s_{a'}$, and hence a prefix of u.

Now let us suppose $X = \{z\}$ and denote $\eta(z)$ by a. In this case we have

$$\varphi(SEpi(X)) = \{g(\mu_w(a^\omega))\} = \{(g(\mu_w(a)))^\omega\} \ .$$

Let us set $s = (g(\mu_w(a)))^\omega \in SEpi_\vartheta$. By Corollary 1.6, $u = g(\psi(w))$ is a prefix of s. Let $\lambda \in LS(\{u\} \cup \Pi)$. Since $\Pi = \Pi_s$, the word λ is a left special factor of the ϑ-episturmian word s, so that we have $\lambda \in \operatorname{Pref} s$.

If $a \in \operatorname{alph} w$, then $B = \{a\} \cup \operatorname{alph} w = \operatorname{alph} w = \operatorname{alph} \psi(w)$, so that $\Pi \subseteq \operatorname{Fact} u$. This implies $|\lambda| \leq |u|$ and then $\lambda \in \operatorname{Pref} u$ as desired.

If $a \notin \operatorname{alph} w$, then by Proposition 2.5 we obtain $\varphi(z) = g(\mu_w(a)) = u\, g(a)$, because $\varphi(z) \notin \operatorname{Pref} u$ otherwise by Lemma 1.12 we would obtain $\mu_w(a) \in \operatorname{Pref} \psi(w)$, that implies $a \in \operatorname{alph} w$. Hence $s = (u\, g(a))^\omega$. Since $\Pi \subseteq (\operatorname{Fact} u) \cup \{g(a)\}$, we have $|\lambda| \leq |u\, g(a)|$, so that $\lambda \in \operatorname{Pref}(u\, g(a))$. Again, if λ is a proper prefix of u we are done, so let us suppose that $\lambda = u\lambda'$ for some $\lambda' \in \operatorname{Pref} g(a)$, and that λ is a left special factor of $g(a)$. Then the prefix λ' of $g(a)$ is repeated in $g(a)$. The longest repeated prefix p of $g(a)$ is either a right special factor or a border of $g(a)$. Both possibilities imply $p = \varepsilon$, since $g(a)$ is unbordered and Π is a biprefix and normal code. As $\lambda' \in \operatorname{Pref} p$, it follows $\lambda' = \varepsilon$. This proves condition 2.

Finally, let us prove condition 3. Let $b, c \in A \setminus \operatorname{Suff} \Pi$, $v \in \Pi^*$, and $\pi \in \Pi$ be such that $bv\bar{c} \in \operatorname{Fact} \pi$. Let $t' \in SEpi(X)$ with $\operatorname{alph} t' = X$, and set $t = \mu_w(\eta(t'))$, $s_1 = g(t)$. Since φ is ϑ-characteristic, $s_1 = \varphi(t')$ is standard ϑ-episturmian. By Lemma 3.1, we have $v = g(\psi(\delta))$ for some $\delta \in B^*$. If $\delta = \varepsilon$ we are done, as condition 3 is trivially satisfied for $w' = x = \varepsilon$; let us then write $\delta = \delta' a$ for some $a \in B$. The words $bg(\psi(\delta'))$ and $g(a\psi(\delta'))$ are both factors of the ϑ-palindrome π; indeed, $\psi(\delta' a)$ begins with $\psi(\delta')a$ and terminates with $a\psi(\delta')$. Hence $g(\psi(\delta'))$ is left special in π as $b \notin \operatorname{Suff} \Pi$ is different from $(g(a))^\ell \in \operatorname{Suff} \Pi$. Therefore $g(\psi(\delta'))$ is a prefix of $g(\psi(w))$, as we have already proved condition 2. Since g is injective and Π is a biprefix code, by Lemma 1.12 it follows $\psi(\delta') \in \operatorname{Pref} \psi(w)$, so that $\delta' \in \operatorname{Pref} w$ by Proposition 1.2. Hence, we can write $\delta = w'x$ with $w' \in \operatorname{Pref} w$ and x either equal to a (if $\delta' a \notin \operatorname{Pref} w$) or to ε. It remains to show that if $w'x \notin \operatorname{Pref} w$, then $x \notin \eta(X)$.

Let us first assume that $\eta(X) = \{x\}$. In this case we have $s_1 = g(\mu_w(\eta(t'))) = g(\psi(wx^\omega))$ by (1). Since $bv = bg(\psi(w'x)) \in \operatorname{Fact} \pi$, $g(x)$ is a proper factor of π. Then, as $B = \{x\} \cup \operatorname{alph} w$ and $g(x) \neq \pi$, we must have $\pi \in g(\operatorname{alph} w)$, so that $bv \in \operatorname{Fact} g(\psi(w))$ as $\operatorname{alph} w = \operatorname{alph} \psi(w)$. By Proposition 1.3, $\psi(w'x)$ is a factor of $\psi(wx)$. We can then write $\psi(wx) = \zeta\psi(w'x)\zeta'$ for some $\zeta, \zeta' \in B^*$. If ζ were empty, by Proposition 1.2 we would obtain $w'x \in \mathrm{Pref}(wx)$; since $w'x \notin \mathrm{Pref}\, w$, we would derive $w = w'$, which is a contradiction since we proved that $bv = bg(\psi(w'x)) \in \mathrm{Fact}\, g(\psi(w))$. Therefore $\zeta \neq \varepsilon$, and v is left special in s, being preceded both by $(g(\zeta))^{\ell}$ and by $b \notin \mathrm{Suff}\, \Pi$. This implies that v is a prefix of s and then of $g(\psi(w))$ as $|v| \leq |g(\psi(w))|$. By Lemma 1.12, it follows $\psi(w'x) \in \mathrm{Pref}\, \psi(w)$ and then $w'x \in \mathrm{Pref}\, w$ by Proposition 1.2, which is a contradiction.

Suppose now that there exists $y \in \eta(X) \setminus \{x\}$, and let $\Delta \in \eta(X)^{\omega}$ with $\mathrm{alph}\, \Delta = \eta(X)$. The word $s_2 = g(\psi(wyx\Delta))$ is equal to $g(\mu_w(\psi(yx\Delta)))$ by (1), and is then standard ϑ-episturmian since $\varphi = g \circ \mu_w \circ \eta$ is ϑ-characteristic. By applying Proposition 1.3 to w' and $wy \in w'A^*$, we obtain $\psi(w'x) \in \mathrm{Fact}\, \psi(wyx)$. We can write $\psi(wyx) = \zeta\psi(w'x)\zeta'$ for some $\zeta, \zeta' \in B^*$. As $w'x \notin \mathrm{Pref}\, w$ and $x \neq y$, we have by Proposition 1.2 that $\psi(w'x) \notin \mathrm{Pref}\, \psi(wy)$, so that $\zeta \neq \varepsilon$. Hence $v = g(\psi(w'x))$ is left special in s_2, being preceded both by $(g(\zeta))^{\ell}$ and by $b \notin \mathrm{Suff}\, \Pi$. This implies that v is a prefix of s_2 and then of $g(\psi(wy))$; by Lemma 1.12, this is absurd since $\psi(w'x) \notin \mathrm{Pref}\, \psi(wy)$.

Sufficiency: Let $t' \in SEpi(\eta(X))$ and $t = \mu_w(t') \in SEpi(B)$. Since $g(B) = \Pi \subseteq \mathcal{P}_{\vartheta}$, by Proposition 1.8 it follows that $g(t)$ has infinitely many ϑ-palindromic prefixes, so that its set of factors is closed under ϑ.

Thus, in order to prove that $g(t) \in SEpi_{\vartheta}$, it is sufficient to show that any nonempty left special factor λ of $g(t)$ is in $\mathrm{Pref}\, g(t)$. Since λ is left special, there exist $a, a' \in A$, $a \neq a'$, $v, v' \in A^*$, and $r, r' \in A^{\omega}$, such that

$$g(t) = va\lambda r = v'a'\lambda r' \quad . \tag{10}$$

The word $g(t)$ can be uniquely factorized by the elements of Π. Therefore, $va\lambda$ and $v'a'\lambda$ are in $\mathrm{Pref}\, \Pi^*$. We consider three different cases.

Case 1: $va \notin \Pi^*$, $v'a' \notin \Pi^*$.

Since Π is a biprefix (as it is a subset of $\mathcal{P}_{\vartheta}$), overlap-free and normal code, by Proposition 1.10 we have $a\lambda, a'\lambda \in \mathrm{Fact}\, \Pi$. Therefore, by condition 2 of Theorem 3.2, it follows $\lambda \in LS\, \Pi \subseteq \mathrm{Pref}\, g(\psi(w))$, so that it is a prefix of $g(t)$ since by Corollary 1.6, $\psi(w)$ is a prefix of $t = \mu_w(t')$.

Case 2: $va \in \Pi^*$, $v'a' \in \Pi^*$.

From (10), we have $\lambda \in \mathrm{Pref}\, \Pi^*$. By Proposition 1.11, there exists a unique word $\lambda' \in \Pi^*$ such that $\lambda' = \pi_1 \cdots \pi_k = \lambda\zeta$ and $\pi_1 \cdots \pi_{k-1}\delta = \lambda$, with $k \geq 1$, $\pi_i \in \Pi$ for $i = 1, \ldots, k$, $\delta \in A^+$, and $\zeta \in A^*$.

Since g is injective, there exist and are unique the words $\tau, \gamma, \gamma' \in B^*$ such that $g(\tau) = \lambda', g(\gamma) = va, g(\gamma') = v'a'$. Moreover, we have $g(\gamma\tau) = va\lambda' = va\lambda\zeta \in \mathrm{Pref}\, g(t)$ and $g(\gamma'\tau) = v'a'\lambda' = v'a'\lambda\zeta \in \mathrm{Pref}\, g(t)$. By Lemma 1.12, we derive $\gamma\tau, \gamma'\tau \in \mathrm{Pref}\, t$. Setting $\alpha = \gamma^{\ell}, \alpha' = \gamma'^{\ell}$, we obtain $\alpha\tau, \alpha'\tau \in \mathrm{Fact}\, t$, and $\alpha \neq \alpha'$ as $a \neq a'$. Hence τ is a left special factor of t; since $t \in SEpi(B)$, we have $\tau \in \mathrm{Pref}\, t$, so that $g(\tau) = \lambda' \in \mathrm{Pref}\, g(t)$. As λ is a prefix of λ', it follows $\lambda \in \mathrm{Pref}\, g(t)$.

Case 3: $va \notin \Pi^*$, $v'a' \in \Pi^*$ (resp. $va \in \Pi^*$, $v'a' \notin \Pi^*$).

We shall consider only the case when $va \notin \Pi^*$ and $v'a' \in \Pi^*$, as the symmetric case can be similarly dealt with.

Since $v'a' \in \Pi^*$, by (10) we have $\lambda \in \operatorname{Pref} \Pi^*$. By Proposition 1.11, there exists a unique word $\lambda' \in \Pi^*$ such that $\lambda' = \pi_1 \cdots \pi_k = \lambda\zeta$ and $\pi_1 \cdots \pi_{k-1}\delta = \lambda$, with $k \geq 1$, $\pi_i \in \Pi$ for $i = 1, \ldots, k$, $\delta \in A^+$, and $\zeta \in A^*$. By the uniqueness of λ', $v'a'\lambda'$ is a prefix of $g(t)$.

By (10) we have $va\pi_1 \cdots \pi_{k-1}\delta \in \operatorname{Pref} g(t)$. By Proposition 1.10, $a\lambda \in \operatorname{Fact} \Pi$, so that there exist $\xi, \xi' \in A^*$ and $\pi \in \Pi$ such that

$$\xi a\lambda\xi' = \xi a\pi_1 \cdots \pi_{k-1}\delta\xi' = \pi \in \Pi \ .$$

Since δ is a nonempty prefix of π_k, it follows from Proposition 1.9 that $\pi = \xi a\pi_1 \cdots \pi_k\xi'' = \xi a\lambda'\xi''$, with $\xi'' \in A^*$.

Let p (resp. q) be the longest word in $\operatorname{Suff}(\xi a) \cap \Pi^*$ (resp. in $\operatorname{Pref} \xi'' \cap \Pi^*$), and write $\pi = \xi a\lambda'\xi'' = zp\lambda'qz'$, with $z, z' \in A^*$.

Since λ' and zp are nonempty and Π is a biprefix code, one derives that z and z' cannot be empty. Moreover, $b = z^\ell \notin \operatorname{Suff} \Pi$ and $\bar{c} = (z')^f \notin \operatorname{Pref} \Pi$, for otherwise the maximality of p and q could be contradicted using Proposition 1.9.

By condition 3, we have $p\lambda'q = g(\psi(w'x))$ for some $w' \in \operatorname{Pref} w$ and $x \in \{\varepsilon\} \cup (B \setminus \eta(X))$. Since $p, \lambda', q \in \Pi^*$ and g is injective, we derive $\lambda' = g(\tau)$ for some $\tau \in \operatorname{Fact} \psi(w'x)$. We will show that λ' is a prefix of $g(t)$, which proves the assertion as $\lambda \in \operatorname{Pref} \lambda'$.

Suppose first that $p = \varepsilon$, so that $a = b$ and $\tau \in \operatorname{Pref} \psi(w'x)$. If $\tau \in \operatorname{Pref} \psi(w')$, then $\lambda' \in g(\operatorname{Pref} \psi(w')) \subseteq \operatorname{Pref} g(\psi(w')) \subseteq \operatorname{Pref} g(\psi(w))$, and we are done as $g(\psi(w)) \in \operatorname{Pref} g(t)$. Let us then assume $x \neq \varepsilon$, so that $x \in B \setminus \eta(X)$, and $\psi(w')x \in \operatorname{Pref} \tau$. Moreover, we can assume $w'x \notin \operatorname{Pref} w$, for otherwise we would derive $\lambda' \in \operatorname{Pref} g(\psi(w))$ again. Let $\Delta \in \eta(X)^\omega$ be the directive word of t', so that by (1) we have $t = \psi(w\Delta)$. Since $w' \in \operatorname{Pref} w$, we can write $w\Delta = w'\Delta'$ for some $\Delta' \in B^\omega$, so that $t = \psi(w'\Delta')$.

We have already observed that $v'a'\lambda'$ is a prefix of $g(t)$; as $v'a' \in \Pi^*$, by Lemma 1.12 one derives $\tau \in \operatorname{Fact} t$. Since $\psi(w')x \in \operatorname{Pref} \tau$, it follows $\psi(w')x \in \operatorname{Fact} \psi(w'\Delta')$; by Proposition 1.4, we obtain $x \in \operatorname{alph} \Delta'$. This implies, since $x \notin \eta(X)$, that $w \neq w'$, and we can write $w = w'\sigma x\sigma'$ for some $\sigma, \sigma' \in B^*$. By Proposition 1.3, $\psi(w'x)$ is a factor of $\psi(w'\sigma x)$ and hence of $\psi(w)$, so that, since $\tau \in \operatorname{Pref} \psi(w'x)$, we have $\tau \in \operatorname{Fact} \psi(w)$. Hence we have either $\tau \in \operatorname{Pref} \psi(w)$, so that $\lambda' \in \operatorname{Pref} g(\psi(w))$ and we are done, or there exists a letter y such that $y\tau \in \operatorname{Fact} \psi(w)$, so that $d\lambda' \in \operatorname{Fact} g(\psi(w))$ with $d = (g(y))^\ell \in \operatorname{Suff} \Pi$. In the latter case, since $a = b \notin \operatorname{Suff} \Pi$ and $a\lambda' \in \operatorname{Fact} \Pi$, we have by condition 2 that $\lambda' \in \operatorname{Pref} g(\psi(w))$. Since $g(\psi(w))$ is a prefix of $g(t)$, in the case $p = \varepsilon$ the assertion is proved.

If $p \neq \varepsilon$, we have $a \in \operatorname{Suff} \Pi$. Let then $\alpha, \alpha' \in B$ be such that $(g(\alpha))^\ell = a$ and $(g(\alpha'))^\ell = a'$; as $a \neq a'$, we have $\alpha \neq \alpha'$. Since $p\lambda'$ is a prefix of $g(\psi(w'x))$, $p \in \Pi^*$, and $p^\ell = (g(\alpha))^\ell = a$, by Lemma 1.12 one derives that $\alpha\tau$ is a factor of $\psi(w'x)$. Moreover, as $v'a'\lambda' \in \operatorname{Pref} g(t)$ and $v'a' \in \Pi^*$, we derive that $\alpha'\tau$ is a factor of t.

Let then δ' be any prefix of the directive word Δ of t', such that $\alpha'\tau \in \operatorname{Fact}\psi(w\delta')$. By Proposition 1.3, $\psi(w\delta' x)$ contains $\psi(w'x)$, and hence $\alpha\tau$, as a factor. Thus τ is a left special factor of $\psi(w\delta' x)$ and then of the standard episturmian word $\psi(w\delta' x^\omega)$; as $|\tau| < |\psi(w\delta')|$, it follows $\tau \in \operatorname{Pref}\psi(w\delta')$ and then $\tau \in \operatorname{Pref} t$, so that $\lambda' \in \operatorname{Pref} g(t)$. The proof is now complete. □

References

1. Lothaire, M.: Combinatorics on Words. Addison-Wesley, Reading MA (1983)
2. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press, Cambridge UK (2002)
3. Lothaire, M.: Applied Combinatorics on Words. Cambridge University Press, Cambridge UK (2005)
4. Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy. Theoretical Computer Science **255** (2001) 539–553
5. Justin, J., Pirillo, G.: Episturmian words and episturmian morphisms. Theoretical Computer Science **276** (2002) 281–313
6. Fischler, S.: Palindromic prefixes and episturmian words. Journal of Combinatorial Theory, Series A **113** (2006) 1281–1304
7. de Luca, A., De Luca, A.: Pseudopalindrome closure operators in free monoids. Theoretical Computer Science **362** (2006) 282–300
8. Bucci, M., de Luca, A., De Luca, A., Zamboni, L.Q.: On some problems related to palindrome closure. Theoretical Informatics and Applications (2008) DOI: 10.1051/ita:2007064, to appear.
9. Bucci, M., de Luca, A., De Luca, A., Zamboni, L.Q.: On different generalizations of episturmian words. Theoretical Computer Science **393** (2008) 23–36
10. Bucci, M., de Luca, A., De Luca, A., Zamboni, L.Q.: On θ-episturmian words. European Journal of Combinatorics (2008) to appear.
11. Bucci, M., de Luca, A., De Luca, A.: Characteristic morphisms of generalized episturmian words. Preprint n. 18, DMA "R. Caccioppoli" (2008)
12. Berstel, J., Perrin, D.: Theory of Codes. Academic Press, New York (1985)
13. de Luca, A.: Sturmian words: structure, combinatorics, and their arithmetics. Theoretical Computer Science **183** (1997) 45–82
14. Justin, J.: Episturmian morphisms and a Galois theorem on continued fractions. Theoretical Informatics and Applications **39** (2005) 207–215