Pseudoepisturmian Words

(Extended abstract presented to the workshop *Recent Progress in Combinatorics on Words*, Montréal, March 11–16, 2007)

Michelangelo Bucci[1], Aldo de Luca[1], Alessandro De Luca[1], and Luca Q. Zamboni[2]

[1] 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
[2] Department of Mathematics, PO Box 311430
University of North Texas, Denton TX, USA
luca@unt.edu

Abstract. We study some classes of infinite words that generalize episturmian words, and analyse the relations occurring among such classes. In each case, the reversal operator R is replaced by an arbitrary involutory antimorphism of the free monoid A^*.

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, 2]). In such a frame, *Sturmian words* play a relevant role (see [1, Chap. 2]). Some natural extensions of Sturmian words to the case of an alphabet with more than two letters have been recently given in [3, 4], by introducing the class of the so called *episturmian words*.

In this paper we consider different extensions of episturmian words, all based on the replacement of the reversal operator by an arbitrary *involutory antimorphism* of the free monoid A^*. Involutory antimorphisms arise naturally also in some applications; a famous example is the Watson and Crick antimorphic involution in molecular biology (see for instance [5]).

We recall that for an infinite word $w \in A^\omega$, the following conditions are equivalent (see [3, 4]):

1. There exists an infinite word $\Delta = x_1x_2\cdots x_n\cdots \in A^\omega$ such that $w = \lim_{n\to\infty} u_n$, with $u_1 = \varepsilon$ and $u_{i+1} = (u_ix_i)^{(+)}$ for all $i \geq 1$, where $^{(+)}$ is the right palindrome closure operator.
2. w is closed under reversal, and each of its left special factors is a prefix of w.

An infinite word satisfying such conditions is called a *standard episturmian word.*

If the reversal operator is replaced by an arbitrary involutory antimorphism ϑ, then conditions 1 and 2 above are no longer equivalent, but each gives rise to a natural generalization (or extension) of the usual episturmian words.

Words generalizing condition 1 are called *ϑ-standard*, and were previously introduced in [6]. More precisely, a ϑ-standard word w is an infinite word over A obtained as a limit of a sequence $(u_n)_{n>0}$ of ϑ-palindromes, with $u_1 = \varepsilon$ and $u_{i+1} = (u_i x_i)^{\oplus_\vartheta}$ for a suitable directive word $\Delta = x_1 x_2 \cdots x_n \cdots$, where ${}^{\oplus_\vartheta}$ is the right ϑ-palindrome closure operator.

In this paper we introduce and study words generalizing condition 2 above, that we call *standard ϑ-episturmian*. Hence, a standard ϑ-episturmian word is any infinite word w which is closed under ϑ and such that each of its left special factors is a prefix of w.

The main purpose of this paper is to study various connections amongst these two families. We shall see that, in general, neither one is a subset of the other.

A further generalization of condition 1 is made by allowing the iterative ϑ-palindrome closure process to start from an arbitrary word u_0 (called *seed*). In [7] we called any word constructed in this way a *ϑ-standard word with seed*. This is a larger class, strictly containing not only ϑ-standard words (as is trivial by the definition), but also standard ϑ-episturmian words. Indeed, one of the main theorems of this paper shows that an infinite word s is ϑ-standard with seed if and only if it is closed under ϑ and there exists $N \geq 0$ such that any left special factor of s having length $n \geq N$ is a prefix of s.

In general, we shall refer to the words of these families as *pseudoepisturmian words*. In the next sections we shall analyse some properties of pseudoepisturmian words, and the relations existing among the above three classes of words.

For standard definitions and notations on words not explicitly given in the text, we refer to [1, 6].

2 Some Classes of Infinite Words

As is well known, an *involutory antimorphism* of the free monoid A^* is an arbitrary map $\vartheta : A^* \to A^*$ such that $\vartheta(uv) = \vartheta(v)\vartheta(u)$ for any $u, v \in A^*$, and $\vartheta \circ \vartheta = \mathrm{id}$. The reversal operator

$$R : w \in A^* \mapsto \tilde{w} \in A^*$$

is the basic example of involutory antimorphism of A^*. 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 shall denote by PAL_ϑ the set of ϑ-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. Some properties and results on ϑ-palindromes, relating ϑ-palindrome closure operators with periodicity and conjugacy, are in [6].

In the following, we shall fix an involutory antimorphism ϑ of A^*, and use the notation $\bar{w}$ for $\vartheta(w)$. We shall also drop the subscript ϑ from the ϑ-palindrome closure operator ${}^{\oplus_\vartheta}$ when no confusion arises.

If Q is the longest ϑ-palindromic suffix of w and $w = sQ$, then $w^{\oplus} = sQ\bar{s}$. In the special case $\vartheta = R$, we shall denote $w^{\oplus_R}$ by $w^{(+)}$, as usual.

Example 2.1. Let $A = \{a, b\}$ and $w = abaabbaa$. One has $w^{(+)} = abaabbaaba$. If $\vartheta = E \circ R$ where E is the interchange morphism defined by $E(a) = b$ and $E(b) = a$, one has $w^{\oplus} = abaabbaabbab$.

An infinite word s is said *closed under* ϑ if for any $w \in \mathrm{Fact}(s)$ one has $\bar{w} \in \mathrm{Fact}(s)$. One easily derives that if an infinite word is closed under ϑ, then it is recurrent.

2.1 ϑ-Standard Words with Seed

A wide class of infinite words over the alphabet A can be constructed by iterating the right ϑ-palindrome closure operator as follows (cf. [7, 6]). Let u_0 be a fixed word of A^* called *seed*, and $\hat{\psi}_\vartheta : A^* \to A^*$ be the map defined by $\hat{\psi}_\vartheta(\varepsilon) = u_0$ and

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

for $u \in A^*$ and $a \in A$. For any $u, v \in A^*$, one has $\hat{\psi}_\vartheta(uv) \in \hat{\psi}_\vartheta(u)A^* \cap A^*\hat{\psi}_\vartheta(u)$, so that the domain of $\hat{\psi}_\vartheta$ can be extended to infinite words too. More precisely, if $t \in A^\omega$, then

$$\hat{\psi}_\vartheta(t) = \lim_{n\to\infty} \hat{\psi}_\vartheta(w_n) \ ,$$

where $\{w_n\} = \mathrm{Pref}(t) \cap A^n$ for all $n \geq 0$. The word t is called the *directive word* of $\hat{\psi}_\vartheta(t)$, and denoted by $\Delta(\hat{\psi}_\vartheta(t))$. If $u_0 \neq \varepsilon$, then any word $\hat{\psi}_\vartheta(t)$ is called *ϑ-standard with seed*.

When the seed u_0 is empty, the map $\hat{\psi}_\vartheta$ is usually denoted by ψ_ϑ, and the corresponding infinite words are called *ϑ-standard words*. If $\vartheta = R$, the map ψ_R is simply written ψ, and R-standard words are exactly the *standard episturmian words*.

Example 2.2. Let $A = \{a, b, c\}$, ϑ be the involutory antimorphism exchanging b and c and fixing a, $u_0 = ab$, and $w = aac$. Then

$$\hat{\psi}_\vartheta(w) = \left(\hat{\psi}_\vartheta(aa)c\right)^{\oplus} = \left((abacaa)^{\oplus} c\right)^{\oplus} = abacaabacacbabacaabaca \ .$$

The following proposition holds (cf. [8]):

Proposition 2.3. *Let $s = \hat{\psi}_\vartheta(\Delta)$ be a ϑ-standard word with a seed u_0 of length k. The following hold:*

1. *A word w with $|w| > k$ is a prefix of s if and only if $w^{\oplus}$ is a prefix of s,*
2. *the set of all ϑ-palindromic prefixes of s is given by*

$$\hat{\psi}_\vartheta(\mathrm{Pref}(\Delta) \setminus \{\varepsilon\}) \cup (PAL_\vartheta \cap \mathrm{Pref}(u_0)) \ , \tag{1}$$

3. *s is closed under ϑ.*

By a generalization of an argument used in [3] for episturmian words, one can prove (see [8]) the following

Proposition 2.4. *Any ϑ-standard word s with seed is uniformly recurrent.*

Let $\hat{\psi}_\vartheta(\Delta)$ be a ϑ-standard word with seed u_0 and directive word $\Delta = xt_1t_2\cdots t_n\cdots$. Define the endomorphism ϕ_x of A^* by setting

$$\phi_x(a) = \hat{\psi}_\vartheta(xa)\hat{\psi}_\vartheta(x)^{-1}$$

for any letter $a \in A$. From the definition, one has that ϕ_x depends on ϑ and u_0; moreover, $\phi_x(a)$ ends with $\bar{a}$ for all $a \in A$, so that any word of the set $X = \phi_x(A)$ is uniquely determined by its last letter. Thus X is a suffix code, and ϕ_x is an injective morphism.

Example 2.5. Let A, ϑ, and u_0 be defined as in Example 2.2, and let $x = a$. Then

$$\begin{aligned}\phi_a(a) &= \hat{\psi}_\vartheta(aa)\hat{\psi}_\vartheta(a)^{-1} = abaca\ ,\\ \phi_a(b) &= \hat{\psi}_\vartheta(ab)\hat{\psi}_\vartheta(a)^{-1} = abac\ ,\\ \phi_a(c) &= \hat{\psi}_\vartheta(ac)\hat{\psi}_\vartheta(a)^{-1} = abacacb\ .\end{aligned}$$

The following important theorem on ϑ-standard words with seed, whose proof is in [7], shows that such words are morphic images of standard episturmian words.

Theorem 2.6. *Let $w \in A^\omega$ and $x \in A$. Then*

$$\hat{\psi}_\vartheta(xw) = \phi_x(\psi(w))\ ,$$

i.e., any ϑ-standard word s with seed is the image, by an injective morphism, of the standard episturmian word whose directive word is obtained by deleting the first letter of the directive word of s.

In general, a ϑ-standard word with seed (empty or not) can have left special factors which are not prefixes. However, the following noteworthy theorem, proven in [7], shows that all sufficiently long left special factors of a ϑ-standard word with seed are prefixes of it.

Theorem 2.7. *Let t be a ϑ-standard word with seed. Then there exists an integer $N \geq 0$ such that for every $n \geq N$, t has at most one left (resp. right) special factor of length n.*

An infinite word $s \in A^\omega$ is called a *ϑ-word with seed* if there exists a ϑ-standard word t with seed such that $\mathrm{Fact}(s) = \mathrm{Fact}(t)$.

2.2 ϑ-Standard Words

The class of ϑ-standard words was introduced in [6]. This is a (proper) subclass of ϑ-standard words with seed, obtained exactly by choosing the seed $u_0 = \varepsilon$. Similarly, a ϑ-word with seed ε will be called simply a ϑ-word. We recall the following theorem proved in [6]:

Theorem 2.8. *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 $\mu_\vartheta(a) = a^\oplus$.*

The preceding theorem is a stronger version of Theorem 2.6 since in the case of an empty seed, the morphism ϕ_x can be replaced by the simple morphism μ_ϑ, and moreover the ϑ-standard word $\psi_\vartheta(w)$ has the same directive word as $\psi(w)$.

The following theorem, whose proof is in [7], gives a noteworthy improvement of Theorem 2.7 in the case of ϑ-standard words:

Theorem 2.9. *Let w be a left special factor of a ϑ-standard word $t = \mu_\vartheta(s)$, with s a standard episturmian word. If $|w| \geq 3$, then w is a prefix of t.*

2.3 ϑ-Episturmian Words

As was previously mentioned in the introduction, another extension of episturmian words can be obtained by introducing infinite words w (called *standard ϑ-episturmian*) satisfying the two following requirements:

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

A word is called ϑ-episturmian if there exists a standard ϑ-episturmian word having the same set of factors. In the following we shall denote by Epi_ϑ the class of ϑ-episturmian words over A, and by SEpi_ϑ the set of standard ϑ-episturmian words. When $\vartheta = R$, Epi_R is just the class of episturmian words.

More generally, it will be useful to introduce for any $N \geq 0$ the family $SW_\vartheta(N)$ of all infinite words w which are closed under ϑ and such that every left special factor of w whose length is at least N is a prefix of w. Moreover, by $W_\vartheta(N)$ we denote the class of all infinite words having the same set of factors of some word in $SW_\vartheta(N)$. Thus $SW_\vartheta(0) = \mathrm{SEpi}_\vartheta$ and $W_\vartheta(0) = \mathrm{Epi}_\vartheta$.

By adapting some arguments used in [3], it is not difficult to prove (cf. [8]) the following

Proposition 2.10. *An infinite word s is in $W_\vartheta(N)$ if and only if s is closed under ϑ and it has at most one left special factor of any length greater than or equal to N.*

As an immediate consequence, one obtains:

Corollary 2.11. *An infinite word is ϑ-episturmian if and only if it is closed under ϑ and it has at most one left special factor of each length.*

3 General Properties of Pseudoepisturmian Words

Let us recall that $SW_\vartheta(N)$ is the family of all infinite words w which are closed under ϑ and such that, every left special factor of w whose length is at least N

is a prefix of w. Let us denote by SW_ϑ the class of words which are in $SW_\vartheta(N)$ for some $N \geq 0$, i.e.,

$$SW_\vartheta = \bigcup_{N \geq 0} SW_\vartheta(N) \ .$$

The next two propositions are simple extensions of analogous results for episturmian words (cf. [3, 8]).

Proposition 3.1. *Let $w \in SW_\vartheta(N)$ and u be a ϑ-palindromic factor of w such that $|u| \geq N$. Then the leftmost occurrence of u in w is a median factor of a ϑ-palindromic prefix of w, i.e., there exists a word λ such that $\lambda u \bar{\lambda} \in \mathrm{Pref}(w)$.*

Proposition 3.2. *Any word in $SW_\vartheta(N)$ has infinitely many ϑ-palindromic prefixes.*

For a (fixed and arbitrary) word $w \in SW_\vartheta(N)$ we denote by $(B_n)_{n \geq 1}$ the sequence of all ϑ-palindromic prefixes of w, ordered by increasing length. Moreover, for any $i > 0$ let x_i be the unique letter such that $B_i x_i$ is a prefix of w. The infinite word $x = x_1 x_2 \cdots x_n \cdots$ will be called the *subdirective word* of w. The proof of Proposition 3.2 shows that for any $i > 0$, B_{i+1} coincides with the prefix of w ending with the first occurrence of $\bar{x}_i B_i$. The next lemma shows that, under suitable circumstances, a stronger relation holds.

Lemma 3.3. *Let $w \in SW_\vartheta(N)$. With the above notation, let $n > 1$ be such that $x_n = x_k$ for some $k < n$ with $|B_k| \geq N - 2$. Then $B_{n+1} = (B_n x_n)^{\oplus}$.*

Proof. Let k be the greatest integer satisfying the hypotheses of the lemma. Let us first prove that $Q = \bar{x}_n B_k x_n$ does not occur in B_n. By contradiction, consider the rightmost occurrence of Q in B_n, i.e., let $Q\rho$ be a suffix of B_n such that Q does not occur in any shorter suffix. If $|\rho| \leq |B_k|$, then one can easily show that the suffix $Q\rho x_n$ of $B_n x_n$ is a ϑ-palindrome, which is absurd because its length is $|Q\rho x_n| > |Q|$.

Suppose then $Q\rho = \bar{x}_n B_k x_n v \bar{x}_n B_k$ for some $v \in A^*$. Since $Q\rho$ is a suffix of B_n, one has that $\bar{\rho} Q = B_k x_n \bar{v} Q$ is a prefix of B_n. Now there is no proper suffix u of $\bar{v}$ such that uQ is left special in w. Indeed, if such u existed, then uQ would be a prefix of B_n, and so $Q\bar{u}$ would be a suffix of B_n, contradicting (as $|u| < |\rho|$) the fact that $Q\rho$ begins with the rightmost occurrence of Q in B_n. Hence every occurrence of Q in w is preceded by $\bar{v}$. Since $\rho x_n = v \bar{x}_n B_k x_n$ is a factor of w, one obtains $v = \bar{v}$, so that $Q\rho x_n = \bar{x}_n B_k x_n v \bar{x}_n B_k x_n$ is a ϑ-palindromic suffix of $B_n x_n$ longer than Q, a contradiction.

Thus Q does not occur in B_n. Since Q is the longest ϑ-palindromic suffix of $B_n x_n$, we can write

$$w = B_n x_n w' = sQw' \ ,$$

where (s, w') is the leftmost occurrence of Q in w. By Proposition 3.1, $sQ\bar{s} = (B_n x_n)^{\oplus}$ is a prefix of w. From this one derives $B_{n+1} = (B_n x_n)^{\oplus}$. ☐

Theorem 3.4. *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.

Proof. *1.⇒2.* Let $s \in SW_\vartheta(N)$, and let B_i and x_i $(i > 0)$ be defined as above. We consider the minimal integer p such that $|B_p| \geq N - 2$. We set $x_{[p]} = x_p x_{p+1} \cdots x_n \cdots \in A^\omega$, and take the minimal m such that $\mathrm{alph}(x_p \cdots x_{p+m}) = \mathrm{alph}(x_{[p]})$. Let $h = p+m+1$. Then for all $n \geq h$, there exists k with $p \leq k \leq p+m$ such that $x_k = x_n$. Since $k \geq p$ one has $|B_k| \geq N - 2$, so that by Lemma 3.3, $B_{n+1} = (B_n x_n)^\oplus$.

2.⇒3. Let $\hat{\psi}_\vartheta(\Delta)$ be the ϑ-standard word with seed $u_0 = B_h$ and directive word $\Delta = x_h x_{h+1} \cdots x_n \cdots$. One has then $\hat{\psi}_\vartheta(\Delta) = s$.

3.⇒1. This follows from Theorem 2.7. □

Let us set
$$W_\vartheta = \bigcup_{N \geq 0} W_\vartheta(N) \ .$$
The following corollary is a direct consequence of the preceding theorem.

Corollary 3.5. *W_ϑ coincides with the set of all ϑ-words with seed.*

Let us observe that the proof of Theorem 3.4 shows that for any given ϑ-standard word s with seed there exists an integer h with the property that all $n \geq h$ satisfy the conditions of Lemma 3.3, i.e., there exists $k < n$ such that $|B_k| \geq N - 2$ and $x_n = x_k$. The minimal h satisfying the above property will be called the *critical integer* of s. Thus by Lemma 3.3 one has $B_{n+1} = (B_n x_n)^\oplus$ for all $n \geq h$.

Corollary 3.6. *Any standard ϑ-episturmian word is a ϑ-standard word with seed. Moreover, if $s \in \mathrm{SEpi}_\vartheta$ and $x = x_1 x_2 \cdots x_n \cdots$ is its subdirective word, then the critical integer of s is the minimal integer h such that* $\mathrm{alph}(x) = \mathrm{alph}(x_1 \cdots x_h)$.

Proof. It is sufficient to observe that a standard ϑ-episturmian word s is in $SW_\vartheta(0)$ as all its left special factors are prefixes of s. Therefore by Theorem 3.4, s is a ϑ-standard word with seed B_h. Since for all $n > 0$ one has trivially $|B_n| \geq N - 2$, the assertion follows from the definition of critical integer. □

We recall that if s is a standard episturmian word, any prefix p of s has a palindromic suffix which is unrepeated in p (cf. [3]). The following proposition (see [8] for a proof) shows that a similar but weaker result holds in the case of ϑ-standard words with seed.

Proposition 3.7. *Let s be a ϑ-standard word with seed and h be its critical integer. Any prefix p of s of length $> |B_h|$ has a ϑ-palindromic suffix with a unique occurrence in p.*

We call a factor w of $s \in A^\infty$ a *first return* to v if w contains exactly two occurrences of v, one as a prefix and the other as a suffix, so that $w = v\mu = \lambda v$. As a consequence of Proposition 3.7, one derives the following result (cf. [8]):

Proposition 3.8. *Let s be a ϑ-standard word with seed, and h be its critical integer. For any ϑ-palindromic factor P of length $|P| > |B_h|$, every first return to P in s is a ϑ-palindrome.*

In the case of episturmian words, one has the stronger result that *every* first return to a palindrome is a palindrome. This was proven in [9] (see also [7]). However this cannot be extended to ϑ-episturmian words. For instance, let s be the standard ϑ-episturmian word $(abaca)^\omega$, where $\vartheta(a) = a$ and $\vartheta(b) = c$. Then aba is a first return to a in s, but it is not a ϑ-palindrome.

4 Structure of ϑ-Episturmian Words

In this section we shall analyse in detail the class of ϑ-episturmian words, also by showing some relations with the other classes introduced so far.

From Corollary 3.6 and Theorem 2.6, one derives the following

Proposition 4.1. *Let s be a standard ϑ-episturmian word, h be its critical integer, and $x = x_1x_2\cdots x_n\cdots$ be the subdirective word of s. Then s is the image, by an injective morphism, of the standard episturmian word t whose directive word is $x_{h+1}x_{h+2}\cdots x_n\cdots$.*

However, this can be improved. In fact, the next results will show (cf. Theorem 4.4) that every $s \in \mathrm{SEpi}_\vartheta$ is a morphic image, by an injective morphism, of the standard episturmian word whose directive word is precisely x, the subdirective word of s.

In the following we shall denote by $\mathcal{P}$ the set of unbordered ϑ-palindromes. We remark that $\mathcal{P}$ is a *biprefix code*, i.e., none of its elements is a proper prefix or suffix of other elements of $\mathcal{P}$. The following lemma, whose proof we omit for the sake of brevity (cf. [8]), shows that any nonempty ϑ-palindrome admits a unique factorization in unbordered ϑ-palindromes.

Lemma 4.2. $PAL_\vartheta^* = \mathcal{P}^*$.

We remark that from the preceding lemma one derives that any standard ϑ-episturmian word s admits a (unique) infinite factorization in elements of $\mathcal{P}$, i.e., one can write

$$s = \pi_1\pi_2\cdots\pi_n\cdots\ , \qquad \text{with } \pi_i \in \mathcal{P} \text{ for all } i > 0\ . \tag{2}$$

Lemma 4.3. *Let $s \in \mathrm{SEpi}_\vartheta$, with $s = \pi_1\pi_2\cdots\pi_n\cdots$ as above. Let u be a nonempty and proper prefix of π_n, for some $n > 0$. Then u is not right special in s.*

Proof. By contradiction, assume that u is a right special factor of s. Then it is not left special; indeed, otherwise it would be a ϑ-palindrome since s is ϑ-episturmian, and this is clearly absurd as $\pi_n \in \mathcal{P}$.

Consider now the smallest integer h such that u is a prefix of π_h. If $h = 1$, then u would be a ϑ-palindrome, which is again a contradiction. Let then $h > 1$. Since u is not left special, $\bar{a}_{h-1}u$ is its unique left extension in s. One can keep extending to the left in a unique way, until one gets a left special factor, or reaches the beginning of the word. In either case, the factor q of s that one obtains is a prefix of s. Moreover it is right special in s, as every occurrence of the right special factor u extends to the left to q. Hence $\bar{q}$ is a left special factor of s, and then a prefix of s. Thus q is a ϑ-palindrome, and therefore it begins with $\bar{u}$. One has $|q| \geq 2|u|$, for otherwise there would be a nonempty word in $\mathrm{Pref}(u) \cap \mathrm{Suff}(\bar{u})$, that is, a nonempty ϑ-palindromic prefix of u, which contradicts the hypothesis that u is a proper prefix of π_h. Thus $q = \bar{u}q'u$ for some $q' \in PAL_\vartheta$.

We have $\pi_1 \cdots \pi_{h-1} \in \mathcal{P}^*$ and, by Lemma 4.2, $q' \in \mathcal{P}^*$. Since $\mathcal{P}$ is a biprefix code, this implies $\pi_1 \cdots \pi_{h-1}(q')^{-1} \in \mathcal{P}^*$, i.e., $q' = \pi_{h'} \cdots \pi_{h-1}$ for some $h' \leq h$ (if $h' = h$, then $q' = \varepsilon$). Then $\pi_1 \cdots \pi_{h'-1}$ has $\bar{u}$ as a suffix. As $\bar{u}$ has no nonempty ϑ-palindromic suffixes, it is a proper suffix of $\pi_{h'-1}$, which then begins in u, contradicting the minimality of h. □

Theorem 4.4. *Let $s \in A^\omega$ be a standard ϑ-episturmian word, Δ be its subdirective word, and $B = \mathrm{alph}(\Delta)$. There exists a morphism $\mu : B^* \to A^*$ such that $s = \mu(\psi(\Delta))$ and $\mu(B) \subseteq \mathcal{P}$.*

Proof. We can assume that s can be factorized as in (2). For any $n \geq 0$, let a_n be the first letter of π_n. We shall prove that if $n, m \geq 0$ are such that $a_n = a_m$, then $\pi_n = \pi_m$.

Let u be the longest common prefix of π_n and π_m, which is nonempty as $a_n = a_m$. By contradiction, suppose $\pi_n \neq \pi_m$. Then, as $\mathcal{P}$ is a biprefix code, u must be a *proper* prefix of both π_n and π_m, so that there exist two distinct letters b_n, b_m such that ub_n is a prefix of π_n and ub_m is a prefix of π_m. Hence u is a right special factor of s, but this contradicts the previous lemma.

We have shown that for any $n > 0$, π_n is determined by its first letter a_n. Thus, letting $C = \{a_n \mid n > 0\} \subseteq A$, it makes sense to define an injective morphism $\mu : C^* \to A^*$ by setting $\mu(a_n) = \pi_n$ for all $n > 0$. The word

$$t = \mu^{-1}(s) = a_1 a_2 \cdots a_n \cdots \in C^\omega$$

has infinitely many palindrome prefixes, corresponding to the inverse images of the ϑ-palindromic prefixes of s. Indeed, if $\pi_1 \cdots \pi_n$ is a ϑ-palindromic prefix of s, by the uniqueness of the factorization over $\mathcal{P}$ one obtains $\pi_i = \pi_{n+1-i}$ for $i = 1, \ldots, n$. Hence t is closed under reversal.

Let w be a left special factor of t, and let i, j be such that $a_i \neq a_j$ and $a_i w, a_j w \in \mathrm{Fact}(t)$. Then $\bar{a}_i\mu(w), \bar{a}_j\mu(w) \in \mathrm{Fact}(s)$, so that $\mu(w)$ is a left special factor of s, and hence a prefix of it. Again by the uniqueness of the factorization of s over $\mathcal{P}$, one derives $w \in \mathrm{Pref}(t)$. Therefore t is a standard episturmian word over C. Finally, since $\mu(PAL \cap \mathrm{Pref}(t)) = PAL_\vartheta \cap \mathrm{Pref}(s)$ as shown above, one easily obtains that the directive word of t is exactly Δ, so that $C = B$. □

Corollary 4.5. *A standard ϑ-episturmian word s is ϑ-standard if and only if $s = \mu_\vartheta(t)$ for some $t \in A^\omega$.*

Proof. If s is ϑ-standard, then by Theorem 2.8 there exists a standard episturmian word t such that $s = \mu_\vartheta(t)$. Conversely, if $t \in A^\omega$ and $s = \mu_\vartheta(t)$, then, since $\mu_\vartheta(a) \in \mathcal{P}$ for any $a \in A$, by the uniqueness of the factorization over $\mathcal{P}$ one has that μ_ϑ is the morphism μ considered in the preceding theorem. Thus $t = \mu_\vartheta^{-1}(s)$ is a standard episturmian word and s is ϑ-standard by Theorem 2.8. □

Proposition 4.6. *Let t be a standard episturmian word with* $\mathrm{alph}(t) = B$, $\mu : B^* \to A^*$ *be a morphism, and* $s = \mu(t)$. *If*

1. $\mu(x) \in PAL_\vartheta$ *for all* $x \in B$,
2. $\mathrm{card}(\mathrm{alph}(s)) = \sum_{x \in B} |\mu(x)|$,

then s is a standard ϑ-episturmian word.

Proof. From the first condition one obtains that μ sends palindromes into ϑ-palindromes, so that s has infinitely many ϑ-palindromic prefixes, and is therefore closed under ϑ.

Let w be a nonempty left special factor of s. Suppose first that w is a proper factor of $\mu(x)$ for some $x \in B$, and is not a prefix of $\mu(x)$. Let a be the first letter of w. The second condition in the hypothesis can be restated as follows: for every letter c of s there exists a unique $y \in B$ such that c occurs – exactly once – in $\mu(y)$. Thus our a occurs *only* in $\mu(x)$; since a is not a prefix of $\mu(x)$, it is always preceded in s by the letter which precedes a in $\mu(x)$. Hence a is not left special, a contradiction. Thus we can write w as $w_1\mu(u)w_2$, where w_1 is a proper suffix of $\mu(x_1)$ and w_2 is a proper prefix of $\mu(x_2)$, for some suitable $x_1, x_2 \in B$ such that $x_1ux_2 \in \mathrm{Fact}(t)$. One can prove that $w_1 = \varepsilon$ by showing, as done above, that otherwise its first letter, which would not be a prefix of $\mu(x_1)$, could not be left special in s.

Therefore $w = \mu(u)w_2$. Reasoning as above, one can prove that if $w_2 \neq \varepsilon$, then w is not right special, and more precisely that each occurrence of w can be extended on the right to an occurrence of $\mu(ux_2)$. Since w is left special in s, so is $\mu(ux_2)$. Without loss of generality, we can then suppose $w = \mu(u)$. Since w is left special in s, there exist two letters $a, b \in A$, $a \neq b$, such that $aw, bw \in \mathrm{Fact}(s)$. Hence there exist two (distinct) letters $x_a, x_b \in B$ such that $x_au, x_bu \in \mathrm{Fact}(t)$. Then u is a left special factor of t and hence a prefix of t, so that $w = \mu(u)$ is a prefix of s. □

Example 4.7. Consider the standard Sturmian word $t = aabaaabaaabaab\cdots$ having the directive word $(aab)^\omega$. Let $A = \{a, b, c, d, e\}$, and ϑ be the involutory antimorphism defined by $\bar{a} = b$, $\bar{c} = c$, $\bar{d} = e$. If μ is the morphism $\mu : \{a, b\}^* \to A^*$ defined by $\mu(a) = acb$ and $\mu(b) = de$, then the word

$$s = \mu(t) = acbacbdeacbacbacbde\cdots$$

is a standard ϑ-episturmian word. We observe that s is not ϑ-standard, as $ab = a^{\oplus}$ is not a prefix of s.

Remark 4.8. Any morphism satisfying the two conditions in the statement of Proposition 4.6 is such that $\mu(x) \in \mathcal{P}$ for any letter x. However there exist standard ϑ-episturmian words for which the morphism μ given by Theorem 4.4 does not satisfy such conditions. For instance, the standard ϑ-episturmian word $s = (abaca)^{\omega}$, with $\bar{a} = a$ and $\bar{b} = c$, is given by $s = \mu(t)$, where $t = \psi(aba^{\omega})$, $\mu(a) = a$, and $\mu(b) = bac$.

We say that a subset B of the alphabet A is *ϑ-skew* if $B \cap \vartheta(B) \subseteq PAL_{\vartheta}$, that is, if

$$x \in B,\ x \neq \bar{x} \Longrightarrow \bar{x} \notin B\ . \tag{3}$$

Proposition 4.9. *Let s be a standard ϑ-episturmian word and Δ be its subdirective word. Then $B = \mathrm{alph}(\Delta)$ is ϑ-skew.*

Proof. We can factorize s as in (2). By Theorem 4.4, it suffices to show that if $\pi_n = xw\bar{x}$ for some $n > 0$ and $w \in A^*$, then π_k does not begin with $\bar{x}$, for any $k > 0$. By contradiction, let k be the smallest integer such that $\bar{x} \in \mathrm{Pref}(\pi_k)$. Without loss of generality, we can assume $n < k$. By Lemma 4.3, no suffix of $w\bar{x}$ is a left special factor of s. Hence every occurrence of $\bar{x}$ in s is preceded by xw (or by a proper suffix of it, if the beginning of the word is reached). First suppose that π_k is preceded in s by xw. Then, since $w \in PAL_{\vartheta} \subseteq \mathcal{P}^*$ and $\mathcal{P}$ is a biprefix code, one has $w = \pi_{k'} \cdots \pi_{k-1}$ for some $k' \leq k$. Thus $\pi_{k'-1}$ ends in x and therefore begins with $\bar{x}$, contradicting the minimality of k.

If $\pi_1 \cdots \pi_{k-1} \in \mathrm{Suff}(w)$, from $n < k$ it follows that $\pi_n = xw\bar{x}$ is a proper factor of itself, which is trivially absurd. □

A ϑ-standard word s can have left special factors which are not prefixes of s. Such factors have length at most 2, by Theorem 2.9. For instance, consider the ϑ-standard word s with $\vartheta = E \circ R$ and $\Delta(s) = (ab)^{\omega}$. One has $s = abbaababbaabbaab\cdots$. As one easily verifies, b and ba are two left special factors which are not prefixes. Hence in general, a ϑ-standard word is not standard ϑ-episturmian. The next proposition gives a characterization of ϑ-standard words which are standard ϑ-episturmian.

Proposition 4.10. *A ϑ-standard word s is standard ϑ-episturmian if and only if $B = \mathrm{alph}(\Delta(s))$ is ϑ-skew.*

Proof. Let s be a ϑ-standard word such that B is ϑ-skew. By Theorem 2.8, one has $s = \mu_{\vartheta}(t)$, where $t = \psi(\Delta(s))$ is a standard episturmian word. The morphism μ_{ϑ} satisfies condition 1 in Proposition 4.6 by definition. By (3), one easily derives that the restriction of μ_{ϑ} to $\mathrm{alph}(t) = B$ satisfies also the second statement of Proposition 4.6, so that $s = \mu_{\vartheta}(t)$ is a standard ϑ-episturmian word.

The converse is a consequence of Proposition 4.9, as the subdirective word of a ϑ-standard word s is $\Delta(s)$. □

Example 4.11. 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.

Let us observe that in general a standard ϑ-episturmian word is not a ϑ-standard word. A simple example is given by the word $s = (abaca)^\omega$, where ϑ is the antimorphism which exchanges b with c and fixes a. One easily verifies that ε and a are the only left special factors of s, so that s is standard ϑ-episturmian. However (cf. Proposition 2.3) s is not ϑ-standard, since ab is a prefix of s, but $(ab)^\oplus = abca$ is not. Another example is the word s considered in Example 4.7: s is standard ϑ-episturmian, but it is not ϑ-standard because its first nonempty ϑ-palindromic prefix is acb and not $ab = a^\oplus$.

Although neither of the two classes (ϑ-standard and standard ϑ-episturmian words) is included in the other one, the following relation holds (cf. [8]).

Proposition 4.12. *Every ϑ-standard word is a morphic image, under a literal morphism, of a standard $\hat{\vartheta}$-episturmian word, where $\hat{\vartheta}$ is an extension of ϑ to a larger alphabet.*

Example 4.13. Let $A = \{a, b\}$, $\vartheta = E \circ R$ (i.e., $\bar{a} = b$), and s be the ϑ-standard word having the directive sequence $\Delta = (ab)^\omega$, so that $s = abbaababbaabbaab\cdots$. Then s is the image, under the literal morphism g defined by $g(a) = g(d) = a$ and $g(b) = g(c) = b$, of the $\hat{\vartheta}$-standard (and standard $\hat{\vartheta}$-episturmian) word

$$\hat{s} = abcdababcdabcdabab\cdots \quad ,$$

where $\hat{\vartheta}(a) = b$ and $\hat{\vartheta}(c) = d$.

References

1. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press (2002)
2. Lothaire, M.: Applied Combinatorics on Words. Cambridge University Press (2005)
3. Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy. Theoretical Computer Science **255** (2001) 539–553
4. Justin, J., Pirillo, G.: Episturmian words and episturmian morphisms. Theoretical Computer Science **276** (2002) 281–313
5. Kari, L., Konstantinidis, S., Sosík, P., Thierrin, G.: On hairpin-free words and languages. In: DLT. Number 3572 in Springer LNCS (2005) 296–307
6. de Luca, A., De Luca, A.: Pseudopalindrome closure operators in free monoids. Theoretical Computer Science **362** (2006) 282–300
7. Bucci, M., de Luca, A., De Luca, A., Zamboni, L.Q.: On some problems related to palindrome closure. Preprint n. 59 Dipartimento di Matematica e Applicazioni "R. Caccioppoli" (2006)
8. Bucci, M., de Luca, A., De Luca, A., Zamboni, L.Q.: On different generalizations of episturmian words. Preprint n. 5 Dipartimento di Matematica e Applicazioni "R. Caccioppoli" (2007)
9. Anne, V., Zamboni, L.Q., Zorca, I.: Palindromes and pseudo-palindromes in episturmian and pseudo-palindromic infinite words. In Brlek, S., Reutenauer, C., eds.: Words 2005. Number 36 in Publications du LaCIM (2005) 91–100