On different generalizations of episturmian words

Michelangelo Bucci, Aldo de Luca, 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`

and

Luca Q. Zamboni
Department of Mathematics, PO Box 311430
University of North Texas, Denton TX, USA
`luca@unt.edu`

Abstract

In this paper we study some classes of infinite words generalizing episturmian words, and analyse the relations occurring among such classes. In each case, the reversal operator R is replaced by an arbitrary involutory antimorphism ϑ of the free monoid A^*. In particular, we define the class of ϑ-words with seed, whose "standard" elements (ϑ-standard words with seed) are constructed by an iterative ϑ-palindrome closure process, starting from a finite word u_0 called the seed. When the seed is empty, one obtains ϑ-words; episturmian words are exactly the R-words. One of the main theorems of the paper characterizes ϑ-words with seed as infinite words closed under ϑ and having at most one left special factor of each length $n \geq N$ (where N is some nonnegative integer depending on the word). When $N = 0$ we call such words ϑ-episturmian. Further results on the structure of ϑ-episturmian words are proved. In particular, some relationships between ϑ-words (with or without seed) and ϑ-episturmian words are shown.

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 [8, 9]). In this framework, *Sturmian words* play a central role (see [8, Chap. 2]). Some natural extensions of Sturmian words to the case of an alphabet with more than two letters have been given in [4, 6], by introducing the class of the so-called *episturmian words*.

We recall that for an infinite word $w \in A^\omega$, the following conditions are equivalent (see [4, 6]):

1. There exists an infinite word $\Delta = x_1 x_2 \cdots x_n \cdots \in A^\omega$ such that $w = \lim_{n\to\infty} u_n$, with $u_1 = \varepsilon$ and $u_{i+1} = (u_i x_i)^{(+)}$ 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 by definition a standard episturmian word.

In this paper we consider different extensions of episturmian words, all based on the replacement of the reversal operator R by an arbitrary *involutory antimorphism* of the free monoid A^*. Involutory antimorphisms naturally arise also in some applications; a famous example is the Watson and Crick antimorphic involution in molecular biology (see for instance [7]).

If R is replaced by an involutory antimorphism ϑ, then conditions 1 and 2 above are no longer equivalent, and 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 [3]. 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 [2] 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.

We mention that a different generalization of episturmian words, not based on involutory antimorphisms, was obtained recently in [5] by making suitable hypotheses on the lengths of palindromic prefixes of an infinite word.

2 Preliminaries

Let A be a finite alphabet, A^* be the *free monoid* generated by A. The elements of A are usually called *letters* and those of A^* *words.* The identity element of A^* is called *empty word* and denoted by ε. A nonempty word w can be written uniquely as a sequence of letters $w = a_1a_2 \cdots a_n$, with $a_i \in A$, $i = 1, \dots, n$. The integer n is called the *length* of w and is denoted by $|w|$. The length of ε is conventionally 0.

Let $w \in A^*$. A word v is a *factor* of w if there exist words r and s such that $w = rvs$; v is a *proper* factor if $v \neq w$. If $|r| = |s|$ then the factor v is said *median.* If $w = vs$ for some s (resp. $w = rv$ for some word r), then v is called a *prefix* (resp. a *suffix*) of w. If v is both a prefix and a suffix of w, it is called a *border* of w. A word w is called *unbordered* if it has no proper and nonempty border. If v is a suffix of w, then wv^{-1} denotes the unique word u such that $uv = w$.

We shall denote respectively by $\mathrm{Fact}(w)$, $\mathrm{Pref}(w)$, and $\mathrm{Suff}(w)$ the sets of all factors, prefixes, and suffixes of the word w.

An infinite word (from left-to-right) x over the alphabet A is any map $x : \mathbb{N}_+ \to A$ where $\mathbb{N}_+$ is the set of positive integers. We can represent x as

$$x = x_1x_2 \cdots x_n \cdots \ ,$$

where for any $i > 0$, $x_i = x(i) \in A$. A (finite) *factor* of x is either the empty word or any sequence $u = x_i \cdots x_j$ with $i \leq j$, i.e., any block of consecutive letters of x. If $i = 1$, then u is a *prefix* of x. We shall denote by $\mathrm{Fact}(x)$ and $\mathrm{Pref}(x)$ the sets of finite factors and prefixes of x respectively. The set of all infinite words over A is denoted by A^ω. We set $A^\infty = A^* \cup A^\omega$.

The product between a finite word w and an infinite one x is naturally defined as the infinite word wx. An *occurrence* of the word v in $w \in A^\infty$ is any pair (r, s), with $r \in A^*$ and $s \in A^\infty$, such that $w = rvs$.

The *subword complexity* of a word $w \in A^\infty$ is the map $\lambda_w : \mathbb{N} \to \mathbb{N}$ defined as

$$\lambda_w(n) = \mathrm{card}(\mathrm{Fact}(w) \cap A^n)$$

for all $n \geq 0$.

An infinite word s is called *recurrent* if each of its factors occurs an infinite number of times in s. 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 $w = v\mu = \lambda v$. The integer $|\lambda|$ is called the *shift* of the two occurrences of v in w. An infinite word s is called *uniformly recurrent* if for any factor v of s the shifts of all first returns to v in s are bounded above by a constant c_v.

If $x \in A$ and vx (resp. xv) is a factor of $w \in A^\infty$, then vx (resp. xv) is called a *right* (resp. *left*) *extension* of v in w. A factor v of w is *right special* if it has two distinct right extensions in w, that is, there exist two distinct letters $a, b \in A$ such that both va and vb are factors of w. *Left special* factors are defined analogously. A factor of w is called *bispecial* if it is right and left special.

For any $w \in A^\infty$, we denote by $\mathrm{alph}(w)$ the set of the letters of A occurring in w.

If $w = a_1 \cdots a_n \in A^*$, $a_i \in A$ for $i = 1, \ldots, n$, the *reversal* $\tilde{w}$ is the word $a_n \cdots a_1$. Moreover one sets $\tilde{\varepsilon} = \varepsilon$. A word is called a *palindrome* if it is equal to its reversal. The set of all palindromes over A is denoted by PAL_A or simply PAL.

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.

3 Some classes of infinite words

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 shall also drop the subscript ϑ from the ϑ-palindrome closure operator ${}^{\oplus_\vartheta}$ when no confusion arises. As one easily verifies (cf. [3]), if Q is the longest ϑ-palindromic suffix of w and $w = sQ$, then

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

In a similar way, one can introduce a left ϑ-palindrome closure operator ${}^\ominus$; for any word w the following relation holds: $w^\ominus = \bar{w}^\oplus$. Some properties and results on ϑ-palindromes, relating ϑ-palindrome closure operators with periodicity and conjugacy, are in [3].

In the special case $\vartheta = R$, we shall always denote $w^{\oplus_R}$ by $w^{(+)}$, as usual.

Example 3.1. Let $A = \{a, b\}$ and $w = abaabbaa$. Then $w^{(+)} = abaabbaaba$. If $\vartheta = E \circ R$ where E is the interchange morphism defined by $E(a) = b$ and $E(b) = a$, one has $w^{\oplus} = abaabbaabbab$.

The following lemma, whose proof is in [3], will be useful in the sequel.

Lemma 3.2. *Let $u \in A^*$ and $w = (ux)^{\oplus}$, where $x \in A$. If p is any prefix of w of length $|p| > |u|$, then $p^{\oplus} = w$.*

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

3.1 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. [2, 3]). 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 $\{w_n\} = \text{Pref}(t) \cap A^n$ 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 3.3. Let $A = \{a, b, c\}$, ϑ be the involutory antimorphism exchanging b and c and fixing a, $u_0 = ab$, and $w = aac$. Then

$$\begin{aligned}\hat{\psi}_\vartheta(w) &= \left(\hat{\psi}_\vartheta(aa)c\right)^{\oplus} = \left(\left(\hat{\psi}_\vartheta(a)a\right)^{\oplus} c\right)^{\oplus} = \left((abacaa)^{\oplus} c\right)^{\oplus} \\ &= (abacaabacac)^{\oplus} = abacaabacacbabacaabaca \ .\end{aligned}$$

Proposition 3.4. *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.*

Proof. If $w^\oplus$ is a prefix of s, then trivially w is a prefix of s. Conversely, suppose that w is a prefix of s with $|w| > k$. If $\Delta = xt_1t_2 \cdots t_n \cdots$ with $x \in A$ and $t_i \in A$, $i > 0$. Let us set $u_1 = (u_0x)^\oplus = \hat{\psi}_\vartheta(x)$ and for $n > 1$, $u_{n+1} = \hat{\psi}_\vartheta(xt_1 \cdots t_n)$, so that $u_{n+1} = (u_nt_n)^\oplus$. We consider the least n such that $|u_n| < |w| \leq |u_{n+1}|$. By Lemma 3.2 one has $w^\oplus = u_{n+1} \in \mathrm{Pref}(s)$. This proves point 1.

By the definition of ϑ-standard words with seed, all the words in the set (1) are ϑ-palindromic prefixes of s. Conversely, let w be a ϑ-palindromic prefix of s. If $|w| \leq k$, then trivially $w \in PAL_\vartheta \cap \mathrm{Pref}(u_0)$. If $|w| > k$, then by following the same argument used for point 1, one has that there exists an integer $n > 0$ such that $w = w^\oplus = u_n \in \hat{\psi}_\vartheta(\mathrm{Pref}(\Delta))$. This proves point 2.

Let w be a factor of s. Since there are infinitely many ϑ-palindromic prefixes of s, there exists a ϑ-palindromic prefix u having w as a factor. Therefore, also $\bar{w}$ is a factor of u and of s. This concludes the proof. ☐

By a generalization of an argument used in [4] for episturmian words, one can prove the following:

Proposition 3.5. *Any ϑ-standard word s with seed is uniformly recurrent.*

Proof. Let $\Delta(s) = xt_1 \cdots t_n \cdots$ be the directive word of $s = \lim_{n\to\infty} u_n$, where $u_1 = (u_0x)^\oplus$ and $u_{n+1} = (u_nt_n)^\oplus$ for $n > 0$. The word s is trivially recurrent. We shall prove that the shifts of the first returns to any factor v of s are bounded by a constant. Let m be the smallest integer such that $v \in \mathrm{Fact}(u_m)$. Let us set $p = u_m$ and let ρ_n be the maximal shift of all first returns to p in u_n, for all $n > m$. Since $u_{n+1} = (u_nt_n)^\oplus$, one has $|u_{n+1}| \leq 2|u_n| + 2$, where such upper bound is reached if and only if $u_{n+1} = u_nt_n\bar{t}_nu_n$. This implies that $\rho_{m+1} \leq |p| + 2$. Since $|u_{n+1}| \leq 2|u_n| + 2$ for all n, one easily derives that $\rho_{n+1} \leq \max\{\rho_n, |p|+2\}$. From this it follows that $\rho_n \leq |p| + 2$ for all $n > m$.

Since v is a factor of u_m, the shifts of all first returns of v in s are upper limited by $|p| + 2 = |u_m| + 2$. ☐

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 3.6. Let A, ϑ, and u_0 be defined as in Example 3.3, 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 [2], shows that such words are morphic images of standard episturmian words.

Theorem 3.7. *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.

Proposition 3.8. *If s is a ϑ-standard word with seed and two letters of A occur infinitely often in $\Delta(s)$, then any prefix of s is a left special factor of s.*

Proof. A prefix p of s is also a prefix of any ϑ-palindromic prefix B of s such that $|p| \leq |B|$. Since there exist two distinct letters, say a and b, which occur infinitely often in $\Delta(s)$, one has $Ba, Bb \in \mathrm{Fact}(s)$. Therefore, $\bar{p}a, \bar{p}b \in \mathrm{Fact}(s)$, i.e., $\bar{p}$ is right special. Since by Proposition 3.4, s is closed under ϑ, one has $\bar{a}p, \bar{b}p \in \mathrm{Fact}(s)$; as $\bar{a} \neq \bar{b}$, p is left special. □

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 [2], shows that all sufficiently long left special factors of a ϑ-standard word with seed are prefixes of it.

Theorem 3.9. *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.*

One of the main results in this paper shows that the previous property on left special factors, along with closure under ϑ, characterizes ϑ-standard words with seed.

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 $\mathrm{Fact}(s) = \mathrm{Fact}(t)$.

3.2 ϑ-standard words

The class of ϑ-standard words was introduced in [3]. 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 [3]:

Theorem 3.10. *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 stronger than Theorem 3.7, since in the case of an empty seed the morphism ϕ_x is formally replaced by the simpler 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 [2], gives a noteworthy improvement of Theorem 3.9 in the case of ϑ-standard words:

Theorem 3.11. *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.*

3.3 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 $SEpi_\vartheta$ 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 as some word in $SW_\vartheta(N)$. Thus $SW_\vartheta(0) = SEpi_\vartheta$ and $W_\vartheta(0) = Epi_\vartheta$. By Theorem 3.11, the class of ϑ-standard words is included in $SW_\vartheta(3)$.

Proposition 3.12. *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.*

Proof. The "only if" part follows immediately from the fact that $\mathrm{Fact}(s) = \mathrm{Fact}(t)$ for some $t \in SW_\vartheta(N)$. Let us prove the "if" part. Let us first suppose that s has infinitely many left special factors. Hence s has exactly one left special factor for each length $n \geq N$, say v_n. Then for any $n \geq N$, v_n is a prefix of v_{n+1}, so that

$$t = \lim_{n\to\infty} v_n$$

is a well-defined infinite word. Trivially $\mathrm{Fact}(t) \subseteq \mathrm{Fact}(s)$; thus to prove that $\mathrm{Fact}(t) = \mathrm{Fact}(s)$ it suffices to show that any given factor w of s with $|w| \geq N$ is a factor of some v_n, $n \geq N$. Since s is closed under ϑ, $\bar{w}$ is a factor of s. Let p be a prefix of s ending in $\bar{w}$. Since s is recurrent, we can consider a prefix of s of the kind pup for some $u \in A^*$. Then there exists $v \in A^*$ such that pv is a right special factor of s, for otherwise one would have $s = (pu)^\omega$, contradicting the fact that s has infinitely many left special factors. Hence $\bar{w}v$ is a right special factor of s, so that $\bar{v}w$ is a left special factor of s. Since $|w| \geq N$, we have $|\bar{v}w| \geq N$ and therefore $\bar{v}w \in \mathrm{Pref}(t)$; thus $\mathrm{Fact}(t) = \mathrm{Fact}(s)$ as desired. This implies that any left special factor of t is also left special in s. It follows that $t \in SW_\vartheta(N)$.

Now suppose that s has only finitely many left special factors. As is well known, this implies that s is eventually periodic, and hence periodic since it is recurrent. Let then w be the longest left special factor of s, and let $s = \lambda w s'$ for some $\lambda \in A^*$ and $s' \in A^\omega$. Then $t = ws'$ has the same set of factors as s. This implies that t is a word of $SW_\vartheta(N)$. ☐

As an immediate consequence, one obtains:

Corollary 3.13. *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.*

Remark. In the case of a binary alphabet $A = \{a, b\}$, by definition any word $s \in Epi_\vartheta$ has a subword complexity λ_s such that $\lambda_s(n) \leq n + 1$ for all $n \geq 0$. It follows that any word in Epi_ϑ is either Sturmian or periodic. In particular, if $\vartheta = E \circ R$, then the word s cannot be Sturmian, since any Sturmian word has either aa or bb as a factor, but not both, whereas s, being closed under ϑ, does not satisfy this requirement. Thus Epi_ϑ contains only the two periodic words $(ab)^\omega$ and $(ba)^\omega$, whereas Epi_R contains all Sturmian words.

4 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. Trivially, we have $SW_\vartheta(N) \subseteq SW_\vartheta(N + 1)$. Let us denote by SW_ϑ the class of words which are in $SW_\vartheta(N)$ for some $N \geq 0$, i.e.,

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

One of the main results is the proof that SW_ϑ coincides with the class of ϑ-standard words with seed (cf. Theorem 4.4). As a corollary, we will derive that any standard ϑ-episturmian word is a ϑ-standard word with seed.

For the sake of clarity, we report in Table 1 the definitions and the notations of the different classes of words introduced so far. We consider only the standard case, since the "non-standard" words of a given class are defined by the property of having the same set of factors as a standard one.

Table 1: Summary of the generalizations of standard episturmian words

Name	**Symbol**	**Definition**
ϑ-standard with seed	SW_ϑ [a]	Generated by iterated ϑ-palindrome closure, starting from any seed
ϑ-standard		Generated by iterated ϑ-palindrome closure, starting from ε
Standard ϑ-episturmian	$SEpi_\vartheta = SW_\vartheta(0)$	Closed under ϑ, and all left special factors are prefixes
	$SW_\vartheta(N)$	Closed under ϑ, and all left special factors of length at least N are prefixes

---

[a] After Theorem 4.4

In order to prove the main theorem, we need some preliminary results.

Lemma 4.1. *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.*

Proof. By contradiction, suppose that $w = \lambda x v u \bar{v} \bar{y} w'$, for some letters $x, y \in A$ with $x \neq y$, and words $\lambda, v \in A^*$, $w' \in A^\omega$. Since w is closed under ϑ, both $xvu\bar{v}$ and $yvu\bar{v}$ are factors of w, so that $vu\bar{v}$ is a left special factor of w of length $|vu\bar{v}| \geq N$, and hence a prefix of it. This leads to a contradiction, because we have found an occurrence of u in w before the leftmost one. □

Proposition 4.2. *Any word in SW_ϑ has infinitely many ϑ-palindromic prefixes.*

Proof. Let $w \in SW_\vartheta(N)$ for a suitable $N \geq 0$, and u be a prefix of w, with $|u| \geq N$. We shall prove that w has a ϑ-palindromic prefix whose length is at least $|u|$, from which the assertion will follow.

Let $\alpha\bar{u}$ ($\alpha \in A^*$) be the prefix of w ending with the first occurrence of $\bar{u}$. Since u is a prefix of w, one has $\alpha\bar{u} = u\beta$ for a suitable $\beta \in A^*$. If $\beta = \varepsilon$, then $\alpha = \varepsilon$ and $u = \bar{u}$, so that $\alpha\bar{u} = u$ is the desired ϑ-palindromic prefix.

Then suppose $\beta = x_1x_2 \cdots x_n$ with $x_i \in A$ for $i = 1, \ldots, n$. As $|\alpha| = |\beta|$, one has $\alpha = y_n \ldots y_1$ for some $y_i \in A$, $i = 1, \ldots, n$. Since $\alpha \neq \varepsilon$, one has $u \neq \bar{u}$, so that $\bar{u}$ is not left special in w. Hence $y_1\bar{u}$ is the only left extension of $\bar{u}$ in w. As w is closed under ϑ, $u\bar{y}_1$ is the only right extension of u in w. This implies $y_1 = \bar{x}_1$.

Since $\alpha\bar{u} = y_n \cdots y_2\bar{x}_1\bar{u}$ ends with the first occurrence of $\bar{u}$ (and hence with the first occurrence of $\bar{x}_1\bar{u}$), one can apply the same argument as above to the prefix ux_1, in order to show that $y_2 = \bar{x}_2$. Continuing this way, one eventually obtains $y_i = \bar{x}_i$ for all $i = 1, \ldots, n$, so that $\alpha = \bar{\beta}$ and $\alpha\bar{u}$ is again the desired ϑ-palindromic prefix of w. □

For a (fixed but arbitrary) word $w \in SW_\vartheta$ 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_ix_i is a prefix of w. The infinite word $x = x_1x_2 \cdots x_n \cdots$ will be called the *subdirective word* of w. The proof of Proposition 4.2 shows that for any $i > 0$, B_{i+1} coincides with the prefix of w ending with the first occurrence of $\bar{x}_iB_i$.

The next lemma shows that, under suitable circumstances, a stronger relation holds.

Lemma 4.3. *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 $B_{n+1} = (B_nx_n)^\oplus$.*

Proof. Let k be the greatest integer satisfying the hypotheses of the lemma. Let us first prove that $Q = \bar{x}_nB_kx_n$ 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_nx_n is a ϑ-palindrome, which is absurd because its length is $|Q\rho x_n| > |Q|$.

Suppose then $Q\rho = \bar{x}_nB_kx_nv\bar{x}_nB_k$ for some $v \in A^*$. Since $Q\rho$ is a suffix of B_n, one has that $\bar{\rho}Q = B_kx_n\bar{v}Q$ is a prefix of B_n (see Figure 1). 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 $\rho x_n = v\bar{x}_nB_kx_n$ is a factor of w, one obtains $v = \bar{v}$, so that $Q\rho x_n = \bar{x}_nB_kx_nv\bar{x}_nB_kx_n$ is a ϑ-palindromic suffix of B_nx_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 Lemma 4.1, $sQ\bar{s} = (B_n x_n)^{\oplus}$ is a prefix of w. From this one derives $B_{n+1} = (B_n x_n)^{\oplus}$. □

Figure 1: Lemma 4.3

Theorem 4.4. *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 4.3, $B_{n+1} = (B_n x_n)^{\oplus}$.

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 3.9. □

Let us set

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

The following corollary is a straightforward consequence of the preceding theorem.

Corollary 4.5. *W_ϑ coincides with the set of all ϑ-words with seed.*

Let us observe that the proof of Theorem 4.4 shows that for any given $s \in SW_\vartheta(N)$, there exists an integer h with the property that for all $n \geq h$, $B_{n+1} = (B_n x_n)^\oplus$. The minimal h satisfying this property will be called the *critical integer* of s. In view of Lemma 4.3, the critical integer of s is less than or equal to the minimal integer p with the property that for all $n \geq p$ there exists $k < n$ such that $|B_k| \geq N - 2$ and $x_n = x_k$.

Corollary 4.6. *Any standard ϑ-episturmian word is a ϑ-standard word with seed. Moreover, if $s \in SEpi_\vartheta$ and $x = x_1 x_2 \cdots x_n \cdots$ is its subdirective word, then the critical integer h of s is less than or equal to the minimal integer p such that* $\mathrm{alph}(x) = \mathrm{alph}(x_1 \cdots x_p)$.

Proof. It is sufficient to observe that a standard ϑ-episturmian word s is in $SW_\vartheta(0)$ because all its left special factors are prefixes of s. Therefore by Theorem 4.4, s is a ϑ-standard word with seed B_h. Since for all $n > 0$ one has $|B_n| \geq N - 2$, it follows trivially that $h \leq p$. $\square$

Proposition 4.7. *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.*

Proof. Since $|p| > |B_h|$ there exists $n \geq h$ such that

$$|B_n x_n| \leq |p| < |B_{n+1}| ,$$

with $B_{n+1} = (B_n x_n)^\oplus$ by the definition of h.

We can write $B_n x_n = vQ$, where Q is the longest ϑ-palindromic suffix of $B_n x_n$, which is nonempty, and, as shown in the proof of Lemma 4.3, has a unique occurrence in $B_n x_n$. Since $B_{n+1} = vQ\bar{v}$, we can write $p = vQ\bar{v}_2$, where $v = v_1 v_2$ for some $v_1, v_2 \in A^*$ and $|v_2| < |v|$. Now $v_2 Q \bar{v}_2$ is a ϑ-palindromic suffix of p which has a unique occurrence in p, for otherwise Q would be repeated in $B_n x_n$. This concludes the proof. $\square$

Let us observe that in the case of a standard episturmian word s, a stronger result holds: any prefix p of s has a palindromic suffix which is unrepeated in p (cf. [4]).

Proposition 4.8. *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.*

Proof. Let P be a ϑ-palindromic factor of s, with $|P| > |B_h|$. Let $u \in \mathrm{Fact}(s)$ be a first return to P, i.e., $u = P\lambda = \rho P$, $\lambda, \rho \in A^*$, and the only two occurrences of P in u are as a prefix and as a suffix of u. If $|P| > |\rho|$, then the prefix P of u overlaps with the suffix P in u and this implies, as is easily to verify, that u is a ϑ-palindrome. Then let us suppose that $u = PvP$ with $v \in A^*$.

Now we consider the first occurrence of u or of $\bar{u}$ in s. Without loss of generality, we may suppose that $s = \alpha u s'$, and $\bar{u}$ does not occur in the prefix of s having length $|\alpha u| - 1$. Let Q be the ϑ-palindromic suffix of αu of maximal length. If $|Q| > |u|$, then we have that $\bar{u}$ occurs in αu before u, which is absurd. Then suppose $|Q| \leq |u|$. If $|u| > |Q| > |P|$, then one contradicts the hypothesis that u is a first return to P. If $|Q| = |P|$, then $Q = P$ has more than one occurrence in αu. Since $|\alpha u| > |B_h|$, one reaches a contradiction by Proposition 4.7. Thus the only remaining possibility is $Q = u$, i.e., u 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 [1] (see also [2]). 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.

5 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 4.6 and Theorem 3.7, one derives the following

Proposition 5.1. *Let s be a standard ϑ-episturmian word, h be its critical integer, and $x = x_1 x_2 \cdots x_n \cdots$ 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 5.5) that every $s \in SEpi_\vartheta$ 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 $\mathcal{P}_\vartheta$, or simply $\mathcal{P}$, 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 $\mathcal{P}$.

Proposition 5.2. $PAL_\vartheta^* = \mathcal{P}^*$.

Proof. Since $\mathcal{P} \subseteq PAL_\vartheta$, one has $\mathcal{P}^* \subseteq PAL_\vartheta^*$. Thus it suffices to show that every nonempty ϑ-palindrome admits a factorization in unbordered ϑ-palindromes, i.e., is in $\mathcal{P}^*$. Note that such a factorization is necessarily unique, as $\mathcal{P}$ is a code.

Let $w \in PAL_\vartheta$. If $|w| = 1$, then clearly w is unbordered, so that $w \in \mathcal{P}$. Let then $|w| > 1$ and suppose, by induction, that every ϑ-palindrome which is shorter than w can be factorized in elements of $\mathcal{P}$. If w is unbordered, then we are done. Let then u be the longest proper border of w. Since w is a ϑ-palindrome, so is u.

If $|w| \geq 2|u|$, then $w = uvu$ for some $v \in \mathit{PAL}_\vartheta$, so that both $u, v \in \mathcal{P}^*$ by induction. This implies the assertion in this case.

If $|w| < 2|u|$, then there exists a border β of u such that $w = u_1\beta\bar{u}_1$, where $u = u_1\beta = \beta\bar{u}_1$. By induction, both β and $u = u_1\beta$ are in $\mathcal{P}^*$; since $\mathcal{P}$ is a biprefix code, this implies that $u_1 = u\beta^{-1}$ is in $\mathcal{P}^*$ too. Hence $w = u_1 u \in \mathcal{P}^*$ as requested. ☐

Example 5.3. Let $A = \{a, b, c, d, e\}$ and ϑ be the antimorphism defined by $\bar{a} = a$, $\bar{b} = c$, and $\bar{d} = e$. The word $acbdaaecba.abaca \in \mathit{PAL}_\vartheta^2$ can be uniquely factorized in unbordered ϑ-palindromes as:

$$a.cb.daae.cb.a.a.bac.a\ .$$

We remark that from the preceding proposition 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 5.4. *Let $s \in \mathit{SEpi}_\vartheta$, 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 $q' \in \mathit{PAL}_\vartheta$.

We have $\pi_1\cdots\pi_{h-1} \in \mathcal{P}^*$ and, by Proposition 5.2, $q' \in \mathcal{P}^*$. 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 5.5. *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* $\mu(B) \subseteq \mathcal{P}$*.*

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, each being the inverse image of a ϑ-palindromic prefix 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 $i = 1, \ldots, n$; conversely, if $w \in PAL$, then trivially $\mu(w) \in PAL_\vartheta$. 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 the prefix code $\mathcal{P}$, one derives $w \in \mathrm{Pref}(t)$. Therefore t is a standard episturmian word over C.

Let $\Delta = x_1 x_2 \cdots x_n \cdots$, and let $B_n = \mu(a_1) \cdots \mu(a_{r_n})$ be the n-th ϑ-palindromic prefix of s for any $n > 1$. Then, as shown above, $a_1 \cdots a_{r_n}$ is exactly the n-th palindromic prefix of t. Since the only word occurring in the factorization (2) and beginning with x_n is $\mu(x_n)$, we have $B_n \mu(x_n) \in \mathrm{Pref}(s)$, so that $x_n = a_{r_n+1}$ for all $n > 1$. This proves that the directive word of t is exactly Δ, and hence $C = B$. □

Corollary 5.6. *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 3.10 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 3.10. $\square$

Proposition 5.7. *Let* $\mu : B^* \to A^*$ *be an injective morphism such that*

1. $\mu(x) \in PAL_{\vartheta}$ *for all* $x \in B$,
2. $\mathrm{alph}(\mu(x)) \cap \mathrm{alph}(\mu(y)) = \emptyset$ *if* $x, y \in B$ *and* $x \neq y$,
3. $|\mu(x)|_a \leq 1$ *for all* $x \in B$ *and* $a \in A$.

Then, for any standard episturmian word $t \in B^{\omega}$, $s = \mu(t)$ *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. By the second condition, $\mu(x)$ is the only word in $\mu(B)$ containing the letter a; by condition 3, a occurs exactly once 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 μ is injective, u is uniquely determined. As 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. $\square$

Example 5.8. 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, since it does not begin with $ab = a^{\oplus}$.

Remark. Any morphism satisfying the three conditions in the statement of Proposition 5.7 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 5.5 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 5.9. *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 5.5, 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 5.4, 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 3.11. 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 5.10. *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 3.10, 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 5.7 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 5.7, so that $s = \mu_\vartheta(t)$ is a standard ϑ-episturmian word.

The converse is a consequence of Proposition 5.9, as the subdirective word of a ϑ-standard word s is $\Delta(s)$. ☐

Example 5.11. 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 3.4) 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 5.8: 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.

Proposition 5.12. *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.*

Proof. Let $s = \psi_\vartheta(\Delta)$ be a ϑ-standard word, $B \subseteq A$ be the set of letters occurring in Δ, and $A' = A \setminus PAL_\vartheta$. Moreover, let us set

$$C = \{c \in B \cap A' \mid \exists r \in (B \setminus \{c, \bar{c}\})^* : r\bar{c} \in \mathrm{Pref}(\Delta)\} ,$$

i.e., let C be the set of letters c occurring in Δ and such that $\bar{c}$ occurs before the first occurrence of c. If $C = \emptyset$, then by the previous proposition s is a standard ϑ-episturmian word, so that the assertion is trivially verified. Let us explicitly note that if $c \in C$, then $\bar{c} \notin C$.

Suppose then C nonempty, and let $C' = \{c' \mid c \in C\}$ and $\hat{C} = \{\hat{c} \mid c \in C\}$ be two sets having the same cardinality as C, both disjoint from A. One can then naturally define the bijective map $\varphi : B \to (B \setminus C) \cup C'$ such that $\varphi(a) = a$ if $a \notin C$, and $\varphi(a) = a'$ otherwise. Set $\hat{A} = A \cup C' \cup \hat{C}$, and define an involutory antimorphism $\hat{\vartheta}$ over $\hat{A}$ by setting $\hat{\vartheta}_{|A} = \vartheta$ and $\hat{\vartheta}(c') = \hat{c}$ for any $c' \in C'$.

Extending φ to a morphism from B^* to $\hat{A}^*$, it makes sense to consider the infinite word $\hat{\Delta} = \varphi(\Delta)$ over $\hat{A}$. Thus we can define as well the $\hat{\vartheta}$-standard word $\hat{s}$ directed by $\hat{\Delta}$. Since $\mathrm{alph}(\hat{\Delta})$ is $\hat{\vartheta}$-skew, by the previous proposition $\hat{s}$ is also $\hat{\vartheta}$-episturmian.

By Theorem 3.10, one has $s = \mu_\vartheta(\psi(\Delta))$ and $\hat{s} = \mu_{\hat{\vartheta}}(\psi(\hat{\Delta}))$. Since φ is injective on B, it follows $\psi(\hat{\Delta}) = \varphi(\psi(\Delta))$, so that

$$\hat{s} = \mu_{\hat{\vartheta}}(\varphi(\psi(\Delta))) . \tag{4}$$

Let $g : \hat{A}^* \to A^*$ be the literal morphism defined as follows:

$$g_{|C'} = \varphi^{-1}, \quad g_{|\hat{C}} = \vartheta \circ \varphi^{-1} \circ \hat{\vartheta}, \text{ and } g_{|A} = \mathrm{id} ,$$

i.e., let $g(a) = a$ if $a \in A$, and for all $c \in C$, let $g(c') = c$ and $g(\hat{c}) = \bar{c}$. We want to show that $g(\hat{s}) = s = \mu_\vartheta(\psi(\Delta))$. In view of (4), it suffices to prove that $g \circ \mu_{\hat{\vartheta}} \circ \varphi = \mu_\vartheta$ over B. Indeed, by the definitions, if $c \in C$ then

$$g(\mu_{\hat{\vartheta}}(\varphi(c))) = g(c'\hat{c}) = c\bar{c} = \mu_\vartheta(c) ,$$

whereas if $a \in B \setminus C$, then

$$g(\mu_{\hat{\vartheta}}(\varphi(a))) = g(a^{\oplus}) = a^{\oplus} = \mu_\vartheta(a) . \qquad \square$$

Example 5.13. 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

$$s = abbaababbaabbaab \cdots .$$

In this case $A' = A = B$, $C = \{b\}$, $C' = \{b'\}$, and $\hat{C} = \{\hat{b}\}$. We set $c = b'$ and $d = \hat{b}$, so that $\hat{A} = \{a, b, c, d\}$, $\hat{\vartheta}(a) = b$, and $\hat{\vartheta}(c) = d$. The morphism φ in this case is defined by $\varphi(a) = a$ and $\varphi(b) = c$. Hence $\hat{\Delta} = \varphi(\Delta) = (ac)^\omega$. The $\hat{\vartheta}$-standard (and standard $\hat{\vartheta}$-episturmian) word $\hat{s}$ directed by $\hat{\Delta}$ is

$$\hat{s} = abcdababcdabcdabab \cdots .$$

The literal morphism g is defined by $g(a) = g(d) = a$, and $g(b) = g(c) = b$. One has $g(\hat{s}) = s$.

6 Concluding remarks and open problems

In the previous sections we have introduced some extensions of episturmian words obtained by replacing the reversal operator R by an arbitrary involutory antimorphism ϑ. More precisely, these words are defined by natural generalizations of some conditions, each of which characterizes standard episturmian words; these are no longer equivalent in the case of an arbitrary ϑ. In this way we have obtained the class of ϑ-standard words, which are generated by iteration of the ϑ-palindrome closure operator, and the class of standard ϑ-episturmian words, which are infinite words closed under ϑ and whose left special factors are prefixes.

Neither of these two classes of words is included in the other. A characterization of the words belonging to the intersection of the two classes has been given (see Corollary 5.6 and Proposition 5.10). Moreover, the two preceding classes are strictly included in the class of ϑ-standard words with seed (see Fig. 2).

Figure 2: Generalized episturmian words

A basic theorem (see Theorem 4.4) shows that this larger class coincides with the set of infinite words which are closed under ϑ and whose sufficiently long left special factors are prefixes. This deep result proves that these two further natural generalizations (i.e., iterated ϑ-palindrome closure starting from any seed, and closure under ϑ with the requirement that all sufficiently long left special factors are prefixes) of the above mentioned conditions are once again equivalent.

The link existing between episturmian words and all these generalizations has been given by some theorems (see Theorems 3.7, 3.10, and 5.5) showing that the words of such families are suitable morphic images of standard episturmian words.

Finally, we mention two interesting open problems. A first task is to study the morphisms $\phi : X^* \to A^*$ such that the image under ϕ of any standard episturmian word over an alphabet X is a standard ϑ-episturmian word on A. Proposition 5.7 gives a sufficient condition, which is not necessary. It would be interesting to find a characterization of such morphisms. In the case $\vartheta = R$ and $X = A$ the injective morphisms of this family are the standard episturmian morphisms introduced in [4, 6].

A second problem is to determine whether morphisms $\phi : X^* \to A^*$ of the previous class are able to generate, when applied to all standard episturmian words over X, all standard ϑ-episturmian words over A. We observe that both these questions are already settled in the case of ϑ-standard words (see Theorem 3.10).

References

[1] V. Anne, L. Q. Zamboni, and I. Zorca. Palindromes and pseudo-palindromes in episturmian and pseudo-palindromic infinite words. In

S. Brlek and C. Reutenauer, editors, *Words 2005*, number 36 in Publications du LaCIM, pages 91–100, 2005.

[2] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On some problems related to palindrome closure. Preprint n. 59 Dipartimento di Matematica e Applicazioni "R. Caccioppoli", 2006.

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

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

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

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

[7] 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*, number 3572 in Springer LNCS, pages 296–307, 2005.

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

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