On some problems related to palindrome closure*

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

and

Luca Q. Zamboni
Department of Mathematics, PO Box 311430
University of North Texas, Denton TX, USA

Abstract

In this paper we solve some open problems related to (pseudo)palindrome closure operators and to the infinite words generated by their iteration, that is, standard episturmian and pseudostandard words. We show that if ϑ is an involutory antimorphism of A^*, then both ϑ-palindromic closures of any factor of a ϑ-standard word are also factors of some ϑ-standard word. We also introduce the class of pseudostandard words with “seed”, obtained by iterated pseudopalindrome closure starting from a nonempty word. We prove that pseudostandard words with seed are morphic images of standard episturmian words. Moreover, for any given pseudostandard word s with seed, there exists an integer N such that for any $n \geq N$, s has at most one right (resp. left) special factor of length n.

1 Introduction

Sturmian words are a classical subject of combinatorics on words (see for instance [3]). By definition, they are infinite words having $n+1$ factors (i.e., blocks of consecutive symbols) of each length n; they enjoy many interesting characterizations and have a wide range of applications, from discrete geometry to crystallography.

Palindrome closure operators, introduced in [4], have had an important role in the study of Sturmian words. If w is a word, its *right* (resp. *left*)

---

*The work for this paper has been supported by the Italian Ministry of Education under Project COFIN 2005 – Automi e Linguaggi Formali: aspetti matematici e applicativi.
Email addresses: `{micbucci,aldo.deluca,alessandro.deluca}@unina.it` (M. Bucci, A. de Luca, and A. De Luca), `luca@unt.edu` (L. Q. Zamboni).

palindrome closure $w^{(+)}$ (resp. $w^{(-)}$) is the shortest palindrome having w as a prefix (resp. suffix). For any Sturmian word, there exists a (unique) *standard* Sturmian word having the same factors. Standard Sturmian words can be constructed by *iterated palindrome closure*, that is, by the following procedure. Start from the empty word, and successively add a letter from $\{a, b\}$ and apply the right palindrome closure operator. In this way one generates a sequence of palindromes, each one being a prefix of the next one, so that a limit is naturally defined. If both a and b are used infinitely many times during such process, the infinite word obtained as a limit is aperiodic, and is exactly a standard Sturmian word.

In recent years, many extensions of Sturmian words have appeared. In particular, in [6] *episturmian* words were introduced (see also [8]). They can be defined as words having the same set of factors of a *standard episturmian* word, which is just a word obtained by iterated palindrome closure over an arbitrary alphabet (and without the aperiodicity condition).

A further generalization was introduced in [5], by substituting palindrome closure with pseudopalindrome closure. A *pseudopalindrome* is a fixed point of some involutory antimorphism ϑ of a free monoid A^*. Thus ordinary palindromes are a special case of pseudopalindromes where the antimorphism is simply the reversal operator. We speak of ϑ-palindromes when a particular antimorphism ϑ is chosen. It is then natural to consider *ϑ-palindrome closure* operators, and to look at words obtained by iterated ϑ-palindrome closure, called *ϑ-standard* (or generally *pseudostandard*) words.

In this paper, we discuss some properties related to (pseudo)palindrome closure, episturmian and ϑ-standard words. In [5] it was proven that both palindromic closures $w^{(+)}$ and $w^{(-)}$ of a factor w of a Sturmian word are themselves factors of Sturmian words. In Section 3 this property is proved for episturmian words.

In Section 4, the closure property is extended to factors of ϑ-standard words too. We also show that there exists *one* ϑ-standard word having both closures as factors. Moreover, we prove that every left special factor of a ϑ-standard word t, whose length is at least 3, is a prefix of t. Recall that a factor u of a (finite or infinite) word w over an alphabet A is *left* (resp. *right*) *special* if there exist two distinct letters $a, b \in A$ such that both au and bu (resp. ua and ub) are factors of w.

In the last section we introduce the class of ϑ-standard words with *seed*. They are just infinite words obtained by iterated ϑ-palindrome closure, but starting from an arbitrary word u_0 (called seed) instead of the empty word. We show that every ϑ-standard word with seed is a morphic image of a standard episturmian word. More precisely, if $\Delta = xx_1x_2 \cdots x_n \cdots$ is the infinite sequence of letters which directs the construction of a ϑ-standard word t with a seed, then $t = \phi_x(s)$, where ϕ_x is a morphism depending on ϑ and u_0, and s is the standard episturmian word directed by $\Delta' = x_1x_2 \cdots x_n \cdots$.

Finally, we show that every sufficiently long left special factor of a ϑ-standard word with seed is a prefix of it, and give an upper bound for the minimal length from which this occurs, in terms of the length of the right ϑ-palindrome closure of u_0x. This solves a conjecture posed in [5].

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 $a_i \in A$, $i = 1, \ldots, n$. 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 $w = rvs$. If $w = vs$ for some s (resp. $w = rv$ for some word r), then v is called a *prefix* (resp. a *suffix*) of w. A word which is both a prefix and a suffix of w is called a *border* of 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.

For $X \subseteq A^*$ and $u \in A^*$, $u^{-1}X$ and Xu^{-1} denote respectively the sets

$$\{w \in A^* \mid uw \cap X \neq \emptyset\} \text{ and } \{w \in A^* \mid wu \cap X \neq \emptyset\} .$$

When X is a singleton $\{x\}$ and $u^{-1}X \neq \emptyset$ (resp. $Xu^{-1} \neq \emptyset$), the unique word $w \in u^{-1}\{x\}$ (resp. $w \in \{x\}u^{-1}$) is denoted by $u^{-1}x$ (resp. xu^{-1}).

If $w = a_1 \cdots a_n \in A^*$, $a_i \in A$, $i = 1, \ldots, n$, the *reversal*, or *mirror image*, of w is the word

$$\tilde{w} = a_n \cdots a_1 .$$

One sets $\tilde{\varepsilon} = \varepsilon$. A word is called *palindrome* if it is equal to its reversal. Any border of a palindrome is trivially a palindrome. We shall denote by $PAL(A)$, or simply PAL, the set of all palindromes on the alphabet A.

An infinite word (from left-to-right) x over the alphabet A is any map $x : \mathbb{N}_+ \longrightarrow A$ 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. Moreover, 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 $w = rvs$. If $w \in A^*$ and $a \in A$, $|w|_a$ denotes the number of distinct occurrences of a in w.

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. We recall that a factor v of a (finite or infinite) word w is called *right special* if it has two distinct right extensions in w, i.e., 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 both right and left special.

A factor w of a word $s \in A^\infty$ is called a *first return* to u if w contains exactly two occurrences of u, one as a prefix and the other as a suffix.

A word $s \in A^\omega$ is said to be *closed under reversal* if for any $w \in \mathrm{Fact}(s)$ one has $\tilde{w} \in \mathrm{Fact}(s)$. A word $w \in A^\omega$ is called *episturmian* if it is closed under reversal and it has at most one right (or equivalently, left) special factor of each length. We recall (see [6]) that every episturmian word is *uniformly recurrent*, i.e., every factor of an episturmian word occurs infinitely often, with bounded gaps.

An episturmian word w is called *standard* if every left special factor of w is a prefix of it. We denote by Ep_A, or simply Ep, the set of all episturmian words over A, and by SEp the set of standard ones.

Proposition 2.1 (cf. [6]). *For every episturmian word w, there exists a standard episturmian word s such that* $\mathrm{Fact}(s) = \mathrm{Fact}(w)$.

Thus $\mathrm{Fact}(Ep) = \mathrm{Fact}(SEp)$. The elements of $\mathrm{Fact}(Ep)$ are called *finite episturmian words.*

Given a word $w \in A^*$, we denote by $w^{(+)}$ its *right palindrome closure*, i.e., the shortest palindrome having w as a prefix. Similarly, $w^{(-)}$ is the left palindrome closure of w. For instance, if $w = abacbca$, then $w^{(+)} = abacbcaba$ and $w^{(-)} = acbcabacbca$.

For any $w \in A^*$, one has $w^{(-)} = \tilde{w}^{(+)}$. Moreover, if Q is the longest palindromic suffix of w and $w = sQ$, then $w^{(+)} = sQ\tilde{s}$.

Let $\psi : A^* \to A^*$ be defined by $\psi(\varepsilon) = \varepsilon$ and $\psi(va) = (\psi(v)a)^{(+)}$ for any $a \in A$ and $v \in A^*$. For any $u, v \in A^*$, one has $\psi(uv) \in \psi(u)A^* \cap A^*\psi(u)$. The map ψ can then be naturally extended to A^ω by setting, for any infinite word x,

$$\psi(x) = \lim_{n\to\infty} \psi(w_n) \ ,$$

where $\{w_n\} = \mathrm{Pref}(x) \cap A^n$ for all $n \geq 0$.

Proposition 2.2 (cf. [6]). *If u is a prefix of a standard episturmian word s, then also $u^{(+)}$ is a prefix of s. Equivalently, $s \in A^\omega$ is a standard episturmian word if and only if there exists $x \in A^\omega$ such that $s = \psi(x)$.*

Given a standard episturmian word s, the (unique) infinite word x such that $s = \psi(x)$ is called *directive word* of s and is denoted by $\Delta(s)$, or simply by Δ. A standard episturmian word s over the alphabet A is called a (standard) *Arnoux-Rauzy* word if every symbol of A occurs infinitely often in the associated directive word $\Delta(s)$. We will denote by AR_A, or simply AR, the set of Arnoux-Rauzy words over A. In the case of a binary alphabet, an AR-word is usually called *standard Sturmian word.*

Proposition 2.3. $\mathrm{Fact}(Ep) = \mathrm{Fact}(AR)$.

Proof. Let $u \in \mathrm{Fact}(Ep) = \mathrm{Fact}(SEp)$. Hence there exists $s \in SEp$ such that $u \in \mathrm{Fact}(s)$. Now let be $s = \psi(\Delta)$ where $\Delta = t_1t_2 \cdots t_n \cdots$. Therefore there exists a palindromic prefix p of s such that $u \in \mathrm{Fact}(p)$. Now $p = \psi(t_1 \cdots t_i)$ for some i. We can consider $\Delta' = t_1 \cdots t_i t$ with $t \in A^\omega$ such that any letter of A occurs infinitely many times in t. Hence $s' = \psi(\Delta') \in AR$ and contains p as a factor. Therefore $u \in \mathrm{Fact}(s')$. Hence, $\mathrm{Fact}(Ep) \subseteq \mathrm{Fact}(AR)$. Since the inverse inclusion is trivial, the result follows. □

The following proposition collects two properties of standard episturmian words (cf. Lemmas 1 and 4 in [6]) which will be useful in the sequel.

Proposition 2.4 (cf. [6])**.** *Let s be a standard episturmian word. The following hold:*

1. *Any prefix p of s has a palindromic suffix which has a unique occurrence in p.*
2. *The first letter a of s occurs in every factor of s having length 2.*

3 A closure property

We want to show that if $w \in \mathrm{Fact}(Ep)$, then also its right and left palindromic closures belong to $\mathrm{Fact}(Ep)$; since $w^{(-)} = \tilde{w}^{(+)}$, it suffices to prove only the right palindromic closure case. We have the following

Proposition 3.1. *Let u be a non-palindrome finite episturmian word; let Q be the longest palindromic suffix of u and write $u = saQ$ where $a \in A$ and $s \in A^*$ (s possibly empty). Then $ua = saQa$ is episturmian.*

Before proving the proposition we need some lemmas. The first lemma was proved in [1, Theorem 1.1]. We report here a different and simpler proof.

Lemma 3.2. *Let w be an episturmian word and $P \in PAL \cap \mathrm{Fact}(w)$. Then every first return to P in w is a palindrome.*

Proof. By Proposition 2.1, we may always suppose that w is a standard episturmian word. Let $u \in \mathrm{Fact}(w)$ be a first return to the palindrome 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 palindrome. Then let us suppose that $u = PvP$ with $v \in A^*$.

Now we consider the first occurrence of u or of $\tilde{u}$ in w. Without loss of generality, we may suppose that $w = \alpha u w'$, and $\tilde{u}$ does not occur in the prefix of w having length $|\alpha u| - 1$. Let Q be the palindromic suffix of αu of maximal length. If $|Q| > |u|$, then we have that $\tilde{u}$ 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, which is absurd in view of Proposition 2.4. The only remaining possibility is $Q = u$, i.e., u is a palindrome. ☐

The following lemma is well known. We report here a proof for the sake of completeness.

Lemma 3.3. *Let $w \in AR$ and s be the unique right special factor of length n. If $B_1, \ldots, B_m, \ldots$ are the bispecial factors of w ordered by increasing length, then s is a suffix of any B_m such that $|s| \leq |B_m|$ and, moreover, for any $x \in A$, $sx \in \mathrm{Fact}(w)$.*

Proof. Let us recall that bispecial factors of standard episturmian words coincide with their palindromic prefixes. Since s is right special factor of w, $\tilde{s}$ is left special and thus a prefix of w. Therefore, s is a suffix of any palidromic prefix B_m of w such that $|s| \leq |B_m|$. Since w is Arnoux-Rauzy, for any $x \in A$, $B_m x$ is a factor of w, and so $sx \in \mathrm{Fact}(w)$. ☐

Lemma 3.4. *Let w and w' be Arnoux-Rauzy words on the same alphabet A. If Q is their common (unique) right special factor of length $|Q| = n$, then w and w' share the same factors up to length $n + 1$.*

Proof. Trivial if $n = 0$. By induction, suppose we have proved the assertion for $n - 1 \geq 0$, i.e., w and w' have the same factors up to length n. Write $Q = aQ'$, $a \in A$, and notice that Q' is the only right special factor of length $n - 1$ of both w and w'.

By symmetry, it suffices to prove that any factor v of w, of length $|v| = n + 1$, is also a factor of w'. Let $v = v'b$, $b \in A$. Suppose first that $v' = Q$. By definition of Arnoux-Rauzy word, each right extension Qx, with $x \in A$, is a factor of both w and w'; in particular, v is a factor of w'.

Now assume that $v' \neq Q$. Let $v' = cv''$ with $c \in A$, and suppose that $v'' = Q'$. One has then $c \neq a$. In this case, since $v = cv''b$ and $Qb = av''b$ are different factors of w, one has that $v''b$ is left special in w. Since $|v''b| = n$, one derives that $v''b = \tilde{Q}$ is a left special factor of w' too, so that $v = cv''b$ is a factor of w'.

If $v'' \neq Q'$, then $v''b$ is the unique right extension of v'' in w. As $|v''b| = n$, it is also a factor of w', and no other letter x is such that $v''x \in \mathrm{Fact}(w')$. Hence $v = cv''b$ is the only right extension in w' of the factor $cv'' \neq Q$. ☐

We can now proceed to prove Proposition 3.1.

Proof of Proposition 3.1. We first observe that u contains a single occurrence of Q. Indeed, if u contained other occurrences of Q, by Lemma 3.2 the suffix of u beginning with the last of such occurrences would be a palindromic suffix of u strictly longer than Q, contradicting the hypothesis of maximality of the length of Q.

By Proposition 2.3 there exists an Arnoux-Rauzy word w such that $u \in \text{Fact}(w)$. We can assume that $ua \notin \text{Fact}(w)$ (otherwise ua is in $\text{Fact}(AR)$ as required); so there exist $b \in A$ such that $b \neq a$ and $ub \in \text{Fact}(w)$. Thus $aQb \in \text{Fact}(w)$; since Q is a palindrome and $w \in AR$, also $bQa \in \text{Fact}(w)$ and Q is a bispecial factor of w. Then it follows that every left special factor of w longer that Q must contain Q as a prefix, and since there is only a single occurrence of Q in u, Q itself is the longest suffix of u which is left special in w. Thus every occurrence of aQ in w must be "preceded" by s, i.e., if $w = \lambda aQ\mu$, then $w = \lambda' saQ\mu$, with $\lambda = \lambda' s$. In particular aQa is not a factor of w, for otherwise ua would be in $\text{Fact}(w)$, contradicting our assumption.

Set $\Delta(w) = t_1t_2\cdots$. Let $\varepsilon = B_1, B_2, \ldots$ be the sequence of all bispecial factors of w, ordered by increasing length, i.e., $|B_i| < |B_{i+1}|$ for all $i > 0$. Thus for each i we have that t_iB_i is right special or equivalently B_it_i is left special. By assumption $Q = B_m$ for some $m > 1$. Let $|Q| = n - 1$ for $n \geq 2$. We then have that t_mQ is right special in w and, from Lemma 3.3, $t_mQx \in \text{Fact}(w)$ for all $x \in A$. It is clear that $t_m \neq a$ since $aQa \notin \text{Fact}(w)$ and $t_mQa \in \text{Fact}(w)$, then we have that aQb and t_mQb are distinct factors of w, thus Qb is left special and bQ is the unique right special factor of w of length n. So $t_m = b$.

Let w' be any Arnoux-Rauzy sequence over A, whose directive word $\Delta(w') = t'_1t'_2\cdots$ satisfies $t'_i = t_i$ for $0 < i \leq m - 1$ and $t'_m = a$. Since Q is the unique right special factor of w and w' of length $n-1$, from Lemma 3.4, we obtain that w and w' have the same factors of length k for each $k \leq n$, but differ on some factors of length $n+1$. Indeed, from the definition of w', we have that aQ is its unique right special factor of length n, thus for all $x \in A$ we have that $aQx \in \text{Fact}(w')$.

Now let us prove that, as in w, each occurrence of aQ in w' is preceded by s. Let $p \in A^*$ be such that $|p| = |s|$ and $paQ \in \text{Fact}(w')$. Let then S be the largest common suffix of paQ and saQ and Q' its prefix of length $n - 1$. Clearly $Q \neq Q'$ since there is only one occurrence of Q in saQ. Assume that $S \neq paQ$, then there exist $x, y \in A$ such that $x \neq y$, $xS \in \text{Suff}(saQ)$ and $yS \in \text{Suff}(paQ)$; then xQ' and yQ' are both factors of w and w' since these latter words have the same factors of length n. Thus Q' is a left special factor of w and w', and that is a contradiction, since the only left special factor of length $n-1$ in w and in w' is Q. Thus $p = s$ and so every occurrence of aQ in w' is preceded by s.

Since aQa is a factor of w', it follows that $saQa = ua$ is a factor of w'. Hence ua is in $\mathrm{Fact}(AR)$ as required. ☐

From the preceding proposition one derives the following theorem, announced without proof in [5].

Theorem 3.5. *If w is a finite episturmian word, then so is each of $w^{(+)}$ and $w^{(-)}$.*

Proof. Trivial if $w \in PAL$. Let then $w = a_1 \cdots a_n Q$, where $a_i \in A$ for $i = 1, \ldots, n$ and Q is the longest palindromic suffix of w. By Proposition 3.1, $wa_n = a_1 \cdots a_n Q a_n$ is a finite episturmian word; since its longest palindromic suffix is $a_n Q a_n$, also $wa_n a_{n-1}$ is episturmian. In this way, by applying Proposition 3.1 exactly n times, one eventually obtains that

$$a_1 a_2 \cdots a_n Q a_n \cdots a_2 a_1 = w^{(+)}$$

is episturmian. Since $w^{(-)} = \tilde{w}^{(+)}$, the assertion follows. ☐

Corollary 3.6. *Let $a \in A$ and $u \in A^*$. If au is a finite episturmian word, then so is $au^{(+)}$.*

Proof. If au is not a palindrome, then by Theorem 3.5, $(au)^{(+)} = au^{(+)}a$ is an episturmian word and therefore so is $au^{(+)}$. Let us then suppose that au is a palindrome.

By Theorem 3.5 one has $u^{(+)} \in \mathrm{Fact}(s)$ for a suitable $s \in AR$. Since s is recurrent there exist letters $x, y \in A$ such that

$$xu^{(+)}y \in \mathrm{Fact}(s)\ .$$

If $x \neq y$, then, since s is closed under reversal, one has also $yu^{(+)}x \in \mathrm{Fact}(s)$. Hence, $u^{(+)}$ is bispecial so that it follows $au^{(+)} \in \mathrm{Fact}(s)$. Let us now consider the case $x = y$. If $x = a$, then the assertion is trivially verified.

Suppose then $x \neq a$. As au is a palindrome, we can write $u = u'a$ with $u' \in PAL$. Hence,

$$x(u'a)^{(+)}x \in \mathrm{Fact}(s)\ .$$

Since $(u'a)^{(+)}$ begins with $u'a$ and terminates with au', one has that $xu'a$ and $au'x$ are factors of s, so that u' is bispecial and, therefore, a palindrome prefix of s.

Let $\Delta(s) = t_1 t_2 \cdots t_n \cdots$ be the directive word of s. There exists an integer k such that $u' = \psi(t_1 t_2 \cdots t_k)$. We consider any AR word s' whose directive word $\Delta(s')$ has the prefix $t_1 t_2 \cdots t_k a$. Thus $u'a = u$ is a prefix of s'. This implies, by Proposition 2.2, that $u^{(+)}$ is a bispecial prefix of s'. From this one derives $au^{(+)} \in \mathrm{Fact}(s')$. ☐

4 Pseudostandard words

An *involutory antimorphism* of the free monoid A^* is a 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. One can then define the ϑ-palindrome closure operators: $w^{\oplus_\vartheta}$ (resp. $w^{\ominus_\vartheta}$) denotes the shortest ϑ-palindrome having w as a prefix (resp. suffix).

Some properties and results on ϑ-palindromes, relating ϑ-palindrome closure operators with periodicity and conjugacy, are in [5]. Further interesting combinatorial properties of ϑ-palindromes, motivated by problems of molecular biology, have been recently studied in [9].

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. We observe that from the definition it follows

$$w^{\ominus} = \bar{w}^{\oplus}$$

for any $w \in A^*$.

For example, when $A = \{a, b\}$, $\vartheta = E \circ R$ where E is the interchange morphism defined by $E(a) = b$ and $E(b) = a$, one has $(aabab)^{\oplus} = aababb$ and $(aabab)^{\ominus} = ababbaabab$.

The following lemma will be useful in the sequel.

Lemma 4.1. *For any $u \in PAL_\vartheta \setminus \{\varepsilon\}$ and $a \in A$, $(ua)^{\oplus}$ is a first return to u, i.e., if $(ua)^{\oplus} = \lambda u \rho$ with $\lambda, \rho \in A^*$, then either $\lambda = \varepsilon$ or $\rho = \varepsilon$.*

Proof. By contradiction, let $\lambda, \rho \in A^+$ be such that

$$(ua)^{\oplus} = \lambda u \rho \; . \tag{1}$$

Clearly $|\lambda| + |u| + |\rho| = |(ua)^{\oplus}| \le 2|u| + 2$, which implies $|\lambda| \le |u| + 2 - |\rho| \le |u| + 1$. Actually, one has $|\lambda| \le |u|$, for if $\lambda = ua$ then from (1) one derives $|(ua)^{\oplus}| = 2|u| + 2$, and then $(ua)^{\oplus} = ua\bar{a}u = ua u \rho$, so that $u\rho = \bar{a}u$. This implies that for some $k > 0$, $u = \bar{a}^k \notin PAL_\vartheta$, a contradiction.

Let then $v, w \in A^*$ be such that $u = \lambda v$ and $(ua)^{\oplus} = uw = \bar{w}u$, whence $\lambda u \rho = uw = \lambda v w$. Thus $u\rho = vw$, so that v is also a prefix of u and therefore a border of u. Since u is a ϑ-palindrome, v is a ϑ-palindrome too, so that $u = \lambda v = v\bar{\lambda}$. Therefore

$$(ua)^{\oplus} = \lambda u \rho = \lambda v \bar{\lambda} \rho \; .$$

Thus $\lambda v \bar{\lambda}$ is a ϑ-palindrome beginning with ua and strictly shorter than $(ua)^{\oplus}$, which is a contradiction. $\square$

We can naturally define a map $\psi_\vartheta : A^* \to A^*$ by $\psi_\vartheta(\varepsilon) = \varepsilon$ and

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

for $u \in A^*$, $a \in A$. For any $u, v \in A^*$ one has $\psi_\vartheta(uv) \in \psi_\vartheta(u)A^* \cap A^*\psi_\vartheta(u)$, so that as done for the iterated palindromic closure, the domain of ψ_ϑ can be extended to infinite words too. More precisely, if $t \in A^\omega$, then

$$\psi_\vartheta(t) = \lim_{n\to\infty} \psi_\vartheta(w_n) \ ,$$

where $\{w_n\} = \mathrm{Pref}(t) \cap A^n$ for all $n \geq 0$. The word t is called the *directive word* of $\psi_\vartheta(t)$ and is denoted by $\Delta(\psi_\vartheta(t))$. The images of infinite words over A by ψ_ϑ have been called *ϑ-standard words* in [5]. If $\vartheta = R$, then $\psi_R = \psi$, where ψ is the iterated palindrome closure operator introduced in Section 2, so that an R-standard word is a standard episturmian word. A ϑ-standard word, without specifying the antimorphism ϑ, has been called *pseudostandard word.*

The following theorem, proven in [5], shows that any ϑ-standard word is a morphic image of the standard episturmian word having the same directive word.

Theorem 4.2. *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 $\mu_\vartheta(a) = a^{\oplus}$.*

A new proof of the preceding theorem will be given in Section 5, as a consequence of a more general result.

Some general properties of ϑ-standard words have been considered in [5]. In particular, we recall that

Proposition 4.3. *Let $s = \psi_\vartheta(x)$ be a ϑ-standard word. The following hold:*

1. *w 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 $\psi_\vartheta(\mathrm{Pref}(x))$,*
3. *s is closed under ϑ, i.e., if $w \in \mathrm{Fact}(s)$, then $\bar{w} \in \mathrm{Fact}(s)$.*

Moreover, the following holds:

Proposition 4.4. *If s is a ϑ-standard word over A 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 the preceding proposition 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. ☐

For the converse of the previous proposition, we observe that a ϑ-standard word s can have left special factors which are not prefixes of s. As an example, 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 of s, which are not prefixes.

However, we will show that if a left special factor w of a ϑ-standard word s is not a prefix of s, then $|w| \leq 2$. For a proof of this we need a couple of lemmas. We denote by $A' = A \setminus PAL_\vartheta$ the set of letters of A that are not ϑ-palindromic.

Lemma 4.5. *The following holds:*

$$A'\mu_\vartheta(A^*) \cap \mu_\vartheta(A^*) = \mu_\vartheta(A^*)A' \cap \mu_\vartheta(A^*) = \emptyset\ .$$

Proof. It is sufficient to observe that any word in $\mu_\vartheta(A^*)$ has an even number of occurrences of letters in A'. ☐

Lemma 4.6. *Let $b, c \in A'$, and let $f = \bar{b}\mu_\vartheta(u)$ and $g = \mu_\vartheta(v)c$ be factors of a ϑ-standard word $t = \mu_\vartheta(s)$, with $s \in SEp$. Then:*

1. *If $bu, vc \in \mathrm{Fact}(s)$ and $|f| > 1$, then $f \neq g$.*
2. *If $u \in \mathrm{Fact}(s)$ and $|f| > 3$, then $bu \in \mathrm{Fact}(s)$.*

Proof. (1). Since $|f| > 1$, one has $u \neq \varepsilon$. By contradiction, if $f = g$, one has also $v \neq \varepsilon$, so that $\bar{b}b$ is a prefix of $\mu_\vartheta(v)$. Then $b\bar{b}$ is a prefix of $\mu_\vartheta(u)$, and so on; therefore, $f = \bar{b}(b\bar{b})^k = (\bar{b}b)^k\bar{b}$ for $k = |u| = |v| \geq 1$. Hence $c = \bar{b}$, $u = b^k$, and $v = \bar{b}^k$. As $k \geq 1$, by Proposition 2.4, $bu = b^{k+1}$ and $vc = \bar{b}^{k+1}$ cannot be both factors of the episturmian word s, a contradiction.

(2). Since $|f| > 3$, one derives $|u| > 1$. By contradiction, suppose $bu \notin \mathrm{Fact}(s)$. By the preceding lemma and by Theorem 4.2, one derives $f = \mu_\vartheta(v)c$ for some suitable $v \in A^*$ and $c \in A'$ such that $vc \in \mathrm{Fact}(s)$. As done before, one then obtains $f = (\bar{b}b)^k\bar{b}$ so that $b^k, \bar{b}^k \in \mathrm{Fact}(s)$, which is absurd by Proposition 2.4, as $k \geq 2$. ☐

Theorem 4.7. *Let w be a left special factor of a ϑ-standard word $t = \mu_\vartheta(s)$, with $s \in SEp$. If $|w| \geq 3$, then w is a prefix of t.*

Proof. By Theorem 4.2, w can be written in one of the following ways:

1. $w = \mu_\vartheta(u)$, with $u \in \mathrm{Fact}(s)$,

2. $w = \bar{b}\mu_\vartheta(u)$, with $bu \in \mathrm{Fact}(s)$ and $b \in A'$,
3. $w = \mu_\vartheta(u)c$, with $uc \in \mathrm{Fact}(s)$ and $c \in A'$,
4. $w = \bar{b}\mu_\vartheta(u)c$, with $buc \in \mathrm{Fact}(s)$ and $b, c \in A'$.

In case 1, let $xw, yw \in \mathrm{Fact}(t)$ with $x \neq y$ letters of A. If x is ϑ-palindromic, then clearly one must have $xu \in \mathrm{Fact}(s)$. If $x \in A'$, then by the preceding lemma one has $\bar{x}u \in \mathrm{Fact}(s)$, since $|xw| > 3$. Since the same holds for y, u is a left special factor of the episturmian word s, and therefore a prefix of it. Thus $w = \mu_\vartheta(u)$ is a prefix of t.

Cases 2 and 4 are absurd; indeed, by the preceding lemma one derives that every occurrence of w is preceded by b.

Finally, in case 3, by the preceding lemma one derives that every occurrence of w is followed by $\bar{c}$. Hence $\mu_\vartheta(uc)$ is a left special factor of t and one can apply the same argument as in case 1 to show that it is a prefix of t. ☐

An infinite word t is a *ϑ-word* if there exists a ϑ-standard word s such that $\mathrm{Fact}(w) = \mathrm{Fact}(s)$. An R-word is an episturmian word.

Proposition 3.1 and Theorem 3.5 can be extended to the class of ϑ-words, showing that if w is a factor of a ϑ-word, then $w^\oplus$ and $w^\ominus$ are also factors of ϑ-words. A proof can be obtained as a consequence of Theorems 3.5 and 4.2 and of Corollary 3.6. However, we need the following lemma (cf. [5]):

Lemma 4.8. *Let $u \in A^*$ and $x \in A \cup \{\varepsilon\}$. Then*

$$(\mu_\vartheta(u)x)^\oplus = \mu_\vartheta\left((ux)^{(+)}\right) .$$

Theorem 4.9. *Let w be a factor of a ϑ-standard word. Then each of $w^\oplus$ and $w^\ominus$ is a factor of a ϑ-standard word.*

Proof. We shall suppose $w \notin PAL_\vartheta$, otherwise the result is trivial. Since $w^\ominus = \bar{w}^\oplus$, it suffices to prove the result for $w^\oplus$. Let $A' = A \setminus PAL_\vartheta$ as above. From Theorem 4.2, one derives that w can be written in one of the following ways:

1. $w = \mu_\vartheta(u)x$, with $x \in A \cup \{\varepsilon\}$ and $ux \in \mathrm{Fact}(Ep)$,
2. $w = \bar{a}\mu_\vartheta(u)b$, with $a, b \in A'$ and $aub \in \mathrm{Fact}(Ep)$,
3. $w = \bar{a}\mu_\vartheta(u)$, with $a \in A'$ and $au \in \mathrm{Fact}(Ep)$.

In the first case, by Theorem 3.5 there exists a standard episturmian word $s = \psi(\Delta)$ such that $(ux)^{(+)} \in \mathrm{Fact}(s)$. Thus, by Lemma 4.8 and Theorem 4.2, $w^\oplus = \mu_\vartheta\left((ux)^{(+)}\right)$ is a factor of the ϑ-standard word $\psi_\vartheta(\Delta) = \mu_\vartheta(s)$.

In the second case, if $\bar{a}\mu_\vartheta(u)b$ is a ϑ-palindrome the result is obvious. Otherwise one has, by using Lemma 4.8,

$$w^\oplus = \bar{a}\left(\mu_\vartheta(u)b\right)^\oplus a = \bar{a}\mu_\vartheta\left((ub)^{(+)}\right)a \in \operatorname{Fact}\left(\mu_\vartheta\left(a(ub)^{(+)}a\right)\right) .$$

Moreover, aub is not a palindrome, since otherwise one would derive, for instance using Lemma 4.8, that $\bar{a}\mu_\vartheta(u)b$ is a ϑ-palindrome. Thus $(aub)^{(+)} = a(ub)^{(+)}a$ and the result is a consequence of Theorem 4.2.

In the third case, since w is not a ϑ-palindrome, by Lemma 4.8 one obtains

$$w^\oplus = \bar{a}\mu_\vartheta(u)^\oplus a \in \operatorname{Fact}\left(\mu_\vartheta(au^{(+)}a)\right) .$$

If $u = a^k$ for some $k \geq 0$, then $au^{(+)}a = a^{k+2} \in \operatorname{Fact}(Ep)$; otherwise $au^{(+)}$ is not a palindrome and $au^{(+)}a = (au^{(+)})^{(+)}$, so that $au^{(+)}a$ is episturmian by Corollary 3.6 and Theorem 3.5. Once again, the assertion follows from Theorem 4.2. ☐

Corollary 4.10. *Let w be a factor of a ϑ-standard word. Then there exists a ϑ-standard word having both $w^\oplus$ and $w^\ominus$ as factors.*

Proof. Trivial if $w \in PAL_\vartheta$. Let then $w = Pbt = saQ$, where P (resp. Q) is the longest ϑ-palindromic prefix (resp. suffix) of w, and $a, b \in A$. Thus $w\bar{a}$ and $\bar{b}w$, being respectively factors of $w^\oplus = saQ\bar{a}\bar{s}$ and $w^\ominus = \bar{t}\bar{b}Pbt$, are factors of ϑ-standard words by Theorem 4.9.

Suppose $w\bar{a} \notin PAL_\vartheta$. Then $(w\bar{a})^\ominus = aw^\ominus\bar{a}$, so that $w^\ominus\bar{a}$ is a factor of some ϑ-standard word, by Theorem 4.9. Consider the word

$$(w^\ominus\bar{a})^\oplus = (\bar{t}\bar{b}Pbt\bar{a})^\oplus = (\bar{t}\bar{b}saQ\bar{a})^\oplus ,$$

and call Q' the longest ϑ-palindromic suffix of $w^\ominus\bar{a}$; then $Q' = aQ\bar{a}$. Indeed, since $aQ\bar{a}$ is a ϑ-palindrome, one has $|Q'| \geq |aQ\bar{a}|$; but $|aQ\bar{a}| < |Q'| \leq |saQ\bar{a}|$ is absurd, for Q would not be the longest ϑ-palindromic suffix of w, and $|Q'| > |saQ\bar{a}|$ cannot happen, for otherwise there would exist a ϑ-palindromic proper suffix of $w^\ominus$ having w as a suffix, contradicting the definition of $w^\ominus$. Thus

$$(w^\ominus\bar{a})^\oplus = \bar{t}\bar{b}saQ\bar{a}\bar{s}bt = \bar{t}\bar{b}Pbt\bar{a}\bar{s}bt$$

is a factor of some ϑ-standard word, again by Theorem 4.9, and it contains both $w^\oplus$ and $w^\ominus$ as factors.

If $w\bar{a} \in PAL_\vartheta$ but $\bar{b}w \notin PAL_\vartheta$, one can prove by a symmetric argument that $(\bar{b}w^\oplus)^\ominus$ is a factor of some ϑ-standard word having both $w^\oplus$ and $w^\ominus$ as factors. Let then $w\bar{a}, \bar{b}w \in PAL_\vartheta$, so that

$$w^\oplus = w\bar{a} = a\bar{w} \ \text{ and } w^\ominus = \bar{b}w = \bar{w}b . \tag{2}$$

If w is a single letter, one derives $w = a = b$, so that $w^{\oplus} = a\bar{a}$ and $w^{\ominus} = \bar{a}a$. Therefore $w^{\oplus}$ and $w^{\ominus}$ are factors of any ϑ-standard word whose directive word begins with a^2. Let us then suppose $|w| > 1$. From (2) it follows $w = aRb$ for some $R \in A^*$ such that $aR = \bar{R}\bar{a} = P$ and $Rb = \bar{b}\bar{R} = Q$. Moreover,

$$w = aRb = a\bar{b}\bar{R} = \bar{R}\bar{a}b \; , \tag{3}$$

showing that $\bar{R}$ is a border of w. Therefore one has either $w = (a\bar{b})^k$ or $w = (a\bar{b})^k a$, for some $k > 0$. In the first case, from (3) one derives $a = \bar{a}$ and $b = \bar{b}$, so that any ϑ-standard word whose directive word begins with ab^{k+1} contains both $w^{\oplus} = (ab)^k a$ and $w^{\ominus} = b(ab)^k$ as factors. In the latter case, by (3) one obtains $a = b$, so that any ϑ-standard word whose directive word begins with a^{k+1} contains both $w^{\oplus} = (a\bar{a})^k$ and $w^{\ominus} = (\bar{a}a)^k$ as factors. ☐

For a finite episturmian word w, the above result can be proven directly; in fact, by Corollary 3.6, we have that $(bw^{(+)})^{(-)}$ is a finite episturmian word having both $w^{(+)}$ and $w^{(-)}$ as factors. In the case of Sturmian words, results analogous to Theorem 4.9 and Corollary 4.10 were proven in [5] with a different and simpler technique based on the structure of finite Sturmian words.

5 Words generated by nonempty seeds

We now consider a generalization of the construction of ϑ-standard words. Define $\hat{\psi}_\vartheta(\varepsilon) = u_0$ with u_0 a fixed word of A^* called *seed*, and

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

for $u \in A^*$ and $a \in A$. As usual, we can extend this definition to infinite words $t \in A^\omega$ by:

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

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 $\Delta(\hat{\psi}_\vartheta(t))$. When the seed u_0 is empty, one has $\hat{\psi}_\vartheta = \psi_\vartheta$ so that one obtains ϑ-standard words. If $u_0 \neq \varepsilon$, then any word $\hat{\psi}_\vartheta(t)$ is called *ϑ-standard with seed*.

Example 5.1. Let $A = \{a, b, c\}$, ϑ be the involutory antimorphism exchanging a and b and fixing c, $u_0 = acbbc$, and $w = abc$. Then

$$\begin{aligned}\hat{\psi}_\vartheta(w) &= \left(\hat{\psi}_\vartheta(ab)c\right)^{\oplus} = \left(\left(\hat{\psi}_\vartheta(a)b\right)^{\oplus} c\right)^{\oplus} = \left(\left((acbbca)^{\oplus}b\right)^{\oplus} c\right)^{\oplus} \\ &= \left((acbbcaacbb)^{\oplus}c\right)^{\oplus} = acbbcaacbbcaacbcacbbcaacbbcaacb \; .\end{aligned}$$

Let $t = xt_1t_2\cdots$, with $x \in A$ and $t_i \in A$ for $i \geq 1$. We remark that the set of ϑ-palindromic prefixes of the word $w = \hat{\psi}_\vartheta(t)$ is

$$(PAL_\vartheta \cap \mathrm{Pref}(u_0)) \cup \{u_n \mid n \geq 1\} ,$$

where $u_1 = (u_0x)^\oplus$ and $u_{i+1} = (u_it_i)^\oplus$ for $i \geq 1$.

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 5.2. Let A, ϑ, and u_0 be defined as in Example 5.1, and let $x = a$. Then

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

To simplify the notation, in the following we shall often omit in the proofs the subscript x from ϕ_x, when no confusion arises.

Theorem 5.3. *Fix $x \in A$ and $u_0 \in A^*$. Let $\hat{\psi}_\vartheta$ and ϕ_x be defined as above. Then for any $w \in A^*$, the following holds:*

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

Proof. In the following we shall often use the property that if γ is an endomorphism of A^* and v is a suffix of $u \in A^*$, then $\gamma(uv^{-1}) = \gamma(u)\gamma(v)^{-1}$.

We will prove the theorem by induction on $|w|$. It is trivial that for $w = \varepsilon$ the claim is true since $\psi(\varepsilon) = \varepsilon = \phi(\varepsilon)$. Suppose that for all the words shorter than w, the statement holds. For $|w| > 0$, we set $w = vy$ with $y \in A$.

First we consider the case $|v|_y \neq 0$. We can then write $v = v_1yv_2$ with $|v_2|_y = 0$, so that

$$\hat{\psi}_\vartheta(xv) = \hat{\psi}_\vartheta(xv_1yv_2) = \hat{\psi}_\vartheta(xv_1)y\lambda = \bar{\lambda}\bar{y}\hat{\psi}_\vartheta(xv_1) ,$$

for a suitable $\lambda \in A^*$. Note that $\hat{\psi}_\vartheta(xv_1)$ is the largest ϑ-palindromic prefix (resp. suffix) followed (resp. preceded) by y (resp. $\bar{y}$) in $\hat{\psi}_\vartheta(xv)$. Therefore,

$$\hat{\psi}_\vartheta(xvy) = \bar{\lambda}\bar{y}\hat{\psi}_\vartheta(xv_1)y\lambda = \hat{\psi}_\vartheta(xv)\hat{\psi}_\vartheta(xv_1)^{-1}\hat{\psi}_\vartheta(xv) . \tag{4}$$

By a similar argument one has:

$$\psi(vy) = \psi(v)\psi(v_1)^{-1}\psi(v) . \tag{5}$$

By induction we have:

$$\hat{\psi}_\vartheta(xv) = \phi(\psi(v))\hat{\psi}_\vartheta(x)\ , \quad \hat{\psi}_\vartheta(xv_1) = \phi(\psi(v_1))\hat{\psi}_\vartheta(x)\ .$$

Replacing in (4), and by (5), we obtain

$$\begin{aligned}\hat{\psi}_\vartheta(xvy) &= \phi(\psi(v))\phi(\psi(v_1))^{-1}\phi(\psi(v))\hat{\psi}_\vartheta(x)\\ &= \phi(\psi(v)\psi(v_1)^{-1}\psi(v))\hat{\psi}_\vartheta(x)\\ &= \phi(\psi(vy))\hat{\psi}_\vartheta(x)\ ,\end{aligned}$$

which was our aim.

Now suppose that $|v|_y = 0$ and $PAL_\vartheta \cap \mathrm{Pref}(u_0x)y^{-1} \neq \emptyset$. Let α_y be the longest word in $PAL_\vartheta \cap \mathrm{Pref}(u_0x)y^{-1}$, that is the longest ϑ-palindromic prefix of u_0x which is followed by y. Since $|v|_y = 0$, one derives that the longest ϑ-palindromic suffix of $\hat{\psi}_\vartheta(xv)y$ is $\bar{y}\alpha_y y$, whence

$$\hat{\psi}_\vartheta(xvy) = \left(\hat{\psi}_\vartheta(xv)y\right)^{\oplus} = \hat{\psi}_\vartheta(xv)\alpha_y^{-1}\hat{\psi}_\vartheta(xv)\ . \tag{6}$$

By induction, this implies

$$\hat{\psi}_\vartheta(xvy) = \phi(\psi(v))\hat{\psi}_\vartheta(x)\alpha_y^{-1}\phi(\psi(v))\hat{\psi}_\vartheta(x)\ . \tag{7}$$

By using (6) for $v = \varepsilon$, one has $\hat{\psi}_\vartheta(xy) = \hat{\psi}_\vartheta(x)\alpha_y^{-1}\hat{\psi}_\vartheta(x)$, and

$$\phi(y) = \hat{\psi}_\vartheta(xy)\left(\hat{\psi}_\vartheta(x)\right)^{-1} = \hat{\psi}_\vartheta(x)\alpha_y^{-1}\ .$$

Moreover, since $\psi(v)$ has no palindromic prefix (resp. suffix) followed (resp. preceded) by y one has

$$\psi(vy) = \psi(v)y\psi(v)\ . \tag{8}$$

Thus from (7) we obtain

$$\begin{aligned}\hat{\psi}_\vartheta(xvy) &= \phi(\psi(v))\phi(y)\phi(\psi(v))\hat{\psi}_\vartheta(x)\\ &= \phi(\psi(v)y\psi(v))\hat{\psi}_\vartheta(x)\\ &= \phi(\psi(vy))\hat{\psi}_\vartheta(x)\ .\end{aligned}$$

Finally we consider $|v|_y = 0$ and $PAL_\vartheta \cap \mathrm{Pref}(u_0x)y^{-1} = \emptyset$. In this case, since $\hat{\psi}_\vartheta(xv)$ has no ϑ-palindromic suffix preceded by $\bar{y}$ (has no ϑ-palindromic prefix followed by y), we have

$$\hat{\psi}_\vartheta(xvy) = \hat{\psi}_\vartheta(xv)y^{\oplus}\hat{\psi}_\vartheta(xv)\ . \tag{9}$$

By induction we then obtain

$$\begin{aligned}\hat{\psi}_\vartheta(xvy) &= \hat{\psi}_\vartheta(xv)y^{\oplus}\hat{\psi}_\vartheta(xv) \qquad (10)\\ &= \phi(\psi(v))\hat{\psi}_\vartheta(x)y^{\oplus}\phi(\psi(v))\hat{\psi}_\vartheta(x)\ .\end{aligned}$$

In particular, if $v = \varepsilon$,

$$\hat{\psi}_\vartheta(xy) = \hat{\psi}_\vartheta(x)y^\oplus\hat{\psi}_\vartheta(x) \ ,$$

so

$$\hat{\psi}_\vartheta(xy)\hat{\psi}_\vartheta(x)^{-1} = \hat{\psi}_\vartheta(x)y^\oplus = \phi(y) \ .$$

Then from (10) and (8) one derives

$$\begin{aligned}\hat{\psi}_\vartheta(xvy) &= \phi(\psi(v))\phi(y)\phi(\psi(v))\hat{\psi}_\vartheta(x) \\ &= \phi(\psi(v)y\psi(v))\hat{\psi}_\vartheta(x) \\ &= \phi(\psi(vy))\hat{\psi}_\vartheta(x) \ ,\end{aligned}$$

which completes the proof. □

From the previous theorem, in the case that w is an infinite word, we obtain:

Theorem 5.4. *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.

Proof. Let $w \in A^\omega$, $t = \psi(w)$, and $w_n = \mathrm{Pref}(w) \cap A^n$ for all $n \geq 0$. From the preceding theorem, for all $n \geq 0$, $\hat{\psi}_\vartheta(xw_n) = \phi(\psi(w_n))\hat{\psi}_\vartheta(x)$. Since $\psi(w_{n+1}) = \psi(w_n)\xi_n$ with $\xi_n \in A^+$, one has $\phi(\psi(w_{n+1})) = \phi(\psi(w_n))\phi(\xi_n)$. Hence, $\hat{\psi}_\vartheta(xw_{n+1})$ has the same prefix of $\hat{\psi}_\vartheta(xw_n)$ of length $|\phi(\psi(w_n))|$, which diverges with n. Since

$$\lim_{n\to\infty} \phi(\psi(w_n)) = \phi\left(\psi(w)\right) \ ,$$

the result follows. □

In the case of an empty seed, from Theorem 5.3 one has

$$\psi_\vartheta(xw) = \phi_x(\psi(w))\psi_\vartheta(x) = \phi_x(\psi(w))x^\oplus \ . \tag{11}$$

Moreover, one easily derives that

$$\phi_x(x) = x^\oplus, \quad \phi_x(y) = x^\oplus y^\oplus \ \text{ for } y \neq x \ .$$

In the case $\vartheta = R$, the morphism ϕ_x reduces to μ_x defined as $\mu_x(y) = xy$ for $y \neq x$ and $\mu_x(x) = x$. Since $x^\oplus = x$, from (11) one obtains the following formula due to Justin [7]:

$$\psi(xw) = \mu_x(\psi(w))x \ . \tag{12}$$

It is noteworthy that Theorem 5.3 provides an alternate proof of Theorem 4.2:

Proof of Theorem 4.2. It is sufficient to observe that, in the case of an empty seed, $x^{\oplus} = \mu_\vartheta(x)$ and $\phi_x = \mu_\vartheta \circ \mu_x$, so that by (11) and (12) one derives:

$$\psi_\vartheta(xw) = (\mu_\vartheta \circ \mu_x)(\psi(w))\mu_\vartheta(x) = \mu_\vartheta(\mu_x(\psi(w))x) = \mu_\vartheta(\psi(xw)) \ ,$$

as desired. ☐

Our next goal is to prove a result analogous to Theorem 4.7 for words generated by nonempty seeds. However, because of the presence of an arbitrary seed, one cannot hope to prove exactly the same assertion; thus in Theorem 5.8 we shall prove that any *sufficiently long* left special factor of a ϑ-standard word with seed is a prefix of it, and that the minimal length from which this occurs depends on the seed in general.

In the following, we shall set

$$u_1 = \hat{\psi}_\vartheta(x) = (u_0 x)^{\oplus} \ ,$$

so that $\phi_x(a) = (u_1 a)^{\oplus} u_1^{-1}$ and $|\phi_x(a)| \leq |u_1| + 2$ for any $a \in A$.

For any letter a, u_a will denote (if it exists) the longest ϑ-palindromic suffix (resp. prefix) of u_1 preceded (resp. followed) by $\bar{a}$ (resp. by a). One has then $u_1 = \phi_x(a) u_a$ for any a such that u_a is defined, and $\phi_x(a) = u_1 a^{\oplus}$ otherwise.

Lemma 5.5. *Let* $\phi_x(A) = X$. *If* $w \in X^*$, *then* $u_1 \in \mathrm{Pref}(wu_1)$.

Proof. Trivial if $w = \varepsilon$. We shall prove by induction that for all $n \geq 1$, if $w \in X^n$, then $u_1 \in \mathrm{Pref}(wu_1)$. Let $w \in X$. Then there exists $a \in A$ such that $w = \phi(a) = (u_1 a)^{\oplus} u_1^{-1}$. Thus $wu_1 = (u_1 a)^{\oplus}$, so that the statement holds for $n = 1$.

Let us suppose the statement is true for n, we will prove it for $n + 1$. If $w \in X^{n+1}$, there exist $a \in A$ and $v \in X^n$ such that $w = \phi(a)v$. By induction, vu_1 can be written as $u_1 v'$ for some $v' \in A^*$. Then one has $wu_1 = \phi(a) u_1 v'$ and, as shown above, u_1 is a prefix of $\phi(a) u_1$, which concludes the proof. ☐

Recall (cf. [2]) that a pair $(p, q) \in A^* \times A^*$ is *synchronizing* for the code X over the alphabet A if for all $\lambda, \rho \in A^*$,

$$\lambda pq\rho \in X^* \Longrightarrow \lambda p, q\rho \in X^* \ .$$

Proposition 5.6. *The pair* (ε, u_1) *is synchronizing for* $X = \phi_x(A)$.

Proof. Since X is a suffix code, it suffices to show that for any $\lambda, \rho \in A^*$,

$$\lambda u_1 \rho \in X^* \Longrightarrow u_1 \rho \in X^* \ .$$

This is trivial if $\lambda = \varepsilon$. Let us factorize $\lambda u_1 \rho$ by the elements of X. Then we can write $\lambda = \lambda' p$ and $u_1 \rho = s \rho'$, where $\lambda', \rho' \in X^*$, and $ps = \phi(a) \in X$

for some letter a (see Figure 1). If $p = \varepsilon$, then trivially $u_1\rho \in X^*$. Suppose then $p \neq \varepsilon$, so that $s \notin X$.

Since $ps \in X$, it follows $|s| \leq |u_1| + 1$. Let us prove that $|s| \leq |u_1|$. By contradiction, suppose $|s| = |u_1| + 1$. Then $\phi(a) = ps = u_1 a\bar{a}$ and $s = u_1\bar{a}$. Therefore $ps = u_1 a\bar{a} = pu_1\bar{a}$, so that $u_1 a = pu_1$. This implies $a = p$ and $u_1 = a^k$ for a suitable $k > 0$. Since a is not a ϑ-palindrome, it follows $u_1 \notin PAL_\vartheta$, a contradiction.

Thus one has $u_1 = sw$ for some $w \in \mathrm{Pref}(\rho')$. By Lemma 5.5, u_1 is a prefix of $\rho' u_1$; clearly, w is a prefix of $\rho' u_1$ too. Therefore w is a prefix of u_1, as $|w| = |u_1| - |s|$. Thus $u_1 = w\bar{s}$, and

$$(u_1 a)^{\oplus} = \phi(a)u_1 = psu_1 = psw\bar{s} = pu_1\bar{s} \ .$$

Since $p \neq \varepsilon$, by Lemma 4.1 one obtains $\bar{s} = \varepsilon$. Hence $u_1\rho = \rho' \in X^*$. ☐

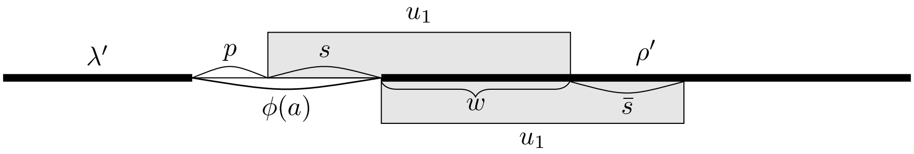

Figure 1: Proposition 5.6

In the following, if Z is a finite subset of A^*, we shall denote by Z^ω the set of all infinite words which can be factorized by the elements of Z. As is well known (cf. [2]) a word $t \in Z^\omega$ has a unique factorization by means of the elements of Z if and only if Z is a code having *finite deciphering delay.* From Proposition 5.6, the code $X = \phi_x(A)$ has a *bounded synchronization delay*, and therefore a finite deciphering delay.

Lemma 5.7. *Let w be in X^*, $a \in A$, and let (r, zs), with $r, s \in A^*$ and $z \in A$, be an occurrence of $u_1 a$ in w, i.e., $w = ru_1azs$. Set $v' = \phi_x(a)^{-1}u_1az$. Then $(r, v's)$ is in $X^* \times X^*$ and it is an occurrence of $\phi_x(a)$ in w.*

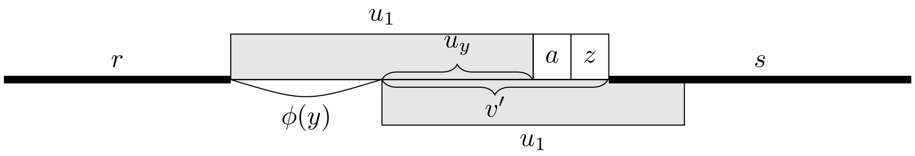

Figure 2: Lemma 5.7

Proof. Let $w \in X^*$ be such that $w = ru_1azs$, with $z \in A$. From Proposition 5.6 we have that r and u_1azs are in X^*. Let $y \in A$ be a letter such that $v = \phi(y)^{-1}u_1azs$ is in X^* and set $v' = \phi(y)^{-1}u_1az$. It is clear from the definition of ϕ that either $v' = \varepsilon$, $v' = z$ or $v' = u_yaz$, where u_y is the longest ϑ-palindromic suffix of u_1 preceded by $\bar{y}$. In the first two cases, it must be $\phi(y) = u_1a^{\oplus}$, so that $a = y$; let then $v' = u_yaz$ (see Figure 2). Since $v = v's \in X^*$, from Lemma 5.5 it follows that u_1 is a prefix of $v'su_1$, so u_ya, whose length is less than $|u_1|$, is a prefix of u_1. By definition, u_y is a prefix of u_1 followed by y, hence $u_yy = u_ya$ and $a = y$. Thus $(r, v's) \in X^* \times X^*$ is an occurrence of $\phi(a)$ in w. □

Theorem 5.8. *Let $t = \hat{\psi}_\vartheta(x\Delta)$ 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.*

Proof. Set $z = \psi(\Delta) = z_1z_2\cdots z_n\cdots$, where each z_i is a letter, for all $i \geq 1$. From Theorem 5.3 we have that $t = \phi(z)$, so that t can be factorized uniquely as $t = \phi(z_1)\phi(z_2)\cdots\phi(z_n)\cdots \in X^\omega$. We shall prove that each left special factor w of t longer than $2\,|u_1| + 2$ is also a prefix of t. Since w is left special, there exist two different occurrences of w in t preceded by distinct letters, say a and b. Moreover, since $|w| > 2|u_1| + 2$, we can write

$$w = p\phi(z_{i+1}\cdots z_{i+h})s = p'\phi(z_{j+1}\cdots z_{j+k})s' \,, \tag{13}$$

where $\phi(z_i) = rap$, $\phi(z_j) = r'bp'$, $\phi(z_{i+h+1}) = s\lambda$, and $\phi(z_{j+k+1}) = s'\lambda'$, with $\lambda, \lambda' \in A^+$ and i, j, h, k positive integers. Thus one can rewrite t as

$$t = \phi(z_1\cdots z_{i-1})raw\lambda\phi(z_{i+h+2}\cdots) = \phi(z_1\cdots z_{j-1})r'bw\lambda'\phi(z_{j+k+2}\cdots).$$

Without loss of generality, we can suppose $|p| \leq |p'|$. From (13) and from the preceding equation, we have

$$rap'\phi(z_{j+1}\cdots z_{j+k})s'\lambda\phi(z_{i+h+2}\cdots) \in X^\omega.$$

Since $|w| > 2\,|u_1| + 2$ and $p' \leq |u_1| + 1$, one has $|\phi(z_{j+1}\cdots z_{j+k})s'| > |u_1| + 1$, so that from Lemma 5.5, u_1 is a prefix of $\phi(z_{j+1}\cdots z_{j+k})s'\lambda'u_1$ and then of $\phi(z_{j+1}\cdots z_{j+k})s'$.

By Proposition 5.6, $(p', \phi(z_{j+1}\cdots z_{j+k})s')$ is a synchronizing pair for X, so that rap' is in X^*. If $p' \neq \varepsilon$, then $r'bp'$ is the only word of the code X having p' as a suffix (recall that any codeword of X is determined by its last letter); hence it should be a suffix of rap', which is clearly a contradiction as $a \neq b$. Then $p' = \varepsilon$, that implies also $p = \varepsilon$. Thus, we can write

$$t = \phi(z_1\cdots z_i)w\lambda\phi(z_{i+h+2}\cdots) = \phi(z_1\cdots z_j)w\lambda'\phi(z_{j+k+2}\cdots),$$

and $z_i \neq z_j$, as w is left special.

Since $w = \phi(z_{i+1}\cdots z_{i+h})s = \phi(z_{j+1}\cdots z_{j+k})s'$ is longer than $2\,|u_1| + 2$ and $|s|, |s'| \leq |u_1| + 1$, there exists a letter $c \in A$ such that u_1c is a prefix of both $\phi(z_{i+1}\cdots z_{i+h})$ and $\phi(z_{j+1}\cdots z_{j+k})$. By Lemma 5.7 one has $\phi(z_{i+1}\cdots z_{i+h}) = \phi(c)\rho$ and $\phi(z_{j+1}\cdots z_{j+k}) = \phi(c)\rho'$ with $\rho, \rho' \in X^*$, so that $z_{i+1} = z_{j+1} = c$ since X is a code.

Let l be the greatest integer such that $z_{i+m} = z_{j+m}$ for all $m \leq l$. Then both $z_i z_{i+1}\cdots z_{i+l}$ and $z_j z_{j+1}\cdots z_{j+l} = z_j z_{i+1}\cdots z_{i+l}$ are factors of z. Since $z_i \neq z_j$, $z_{i+1}\cdots z_{i+l}$ is a left special factor of the episturmian word z, thus a prefix of z, i.e., $z_{i+1}\cdots z_{i+l} = z_1\cdots z_l$. Hence $\phi(z_{i+1}\cdots z_{i+l})$ is a prefix of t.

Now let us suppose that $w' = \phi(z_{i+l+1}\cdots z_{i+h})s = \phi(z_{j+l+1}\cdots z_{j+k})s'$ is strictly longer than u_1. By Lemma 5.5, there exists a letter d such that $u_1 d$ is a prefix of w', so, by applying Lemma 5.7 to $w'\lambda \in X^*$ and to $w'\lambda' \in X^*$ one derives $\phi(z_{i+l+1}) = \phi(z_{j+l+1}) = \phi(d)$, contradicting the fact that $i + l$ was the largest of such indexes. Then $|w'| \leq |u_1|$. By Lemma 5.5, u_1 is a prefix of $w'\lambda u_1$. Thus w' is a prefix of u_1 and $w = \phi(z_{i+1}\cdots z_{i+l})w'$ is a prefix of $\phi(z_{i+1}\cdots z_{i+l})u_1 = \phi(z_1\cdots z_l)u_1$. Let m be an integer such that $|u_1| \leq |\phi(z_{l+1}\cdots z_{l+m})|$. By Lemma 5.5, u_1 is a prefix of $\phi(z_{l+1}\cdots z_{l+m})$ and $\phi(z_1\cdots z_l)u_1$ is a prefix of $\phi(z_1\cdots z_{l+m})$ which is a prefix of t. In conclusion, we obtain that w is a prefix of t. ☐

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] J. Berstel and D. Perrin. *Theory of Codes*. Academic Press, 1985.

[3] J. Berstel and P. Séébold. Sturmian words. In M. Lothaire, editor, *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge UK, 2002. Chapter 2.

[4] A. de Luca. Sturmian words: structure, combinatorics, and their arithmetics. *Theoretical Computer Science*, 183:45–82, 1997.

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

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

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

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

[9] L. Kari and K. Mahalingam. Involution conjugate and commutative words. Preprint, 2006.