UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

Dottorato di Ricerca in Scienze Matematiche
XIX Ciclo

Combinatorial aspects of Sturmian sequences and their generalizations

Alessandro De Luca

Tesi

Tutore
Ch. Prof. A. de Luca

Coordinatore
Ch. Prof. S. Rionero

Contents

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 [38, 39, 40, 1]). In this framework, *Sturmian words* play a central role. They have been widely studied for their theoretical importance and their applications to various fields of science, such as crystallography, Diophantine approximation, or computer vision.

By definition, a Sturmian sequence is an infinite word which is not eventually periodic and has minimal *factor complexity*. This can be equivalently stated as follows: an infinite word is Sturmian if and only if it has $n+1$ distinct *factors* (blocks of consecutive symbols) of each length $n \geq 0$. In particular, Sturmian words are on a binary alphabet, say $\{a, b\}$. They also enjoy some remarkable characterizations of geometrical nature (*cutting sequences*, *mechanical words*). Several general surveys on this subject are available (see for instance [39, Chap. 2] and [1, Chap. 9–10]).

Infinite Sturmian words and their factors, called *finite Sturmian words*, enjoy many characteristic structural properties. Perhaps the most famous one is the so called *balance condition*: the numbers of a's in two factors of the same length can differ at most by 1. In recent years, some works have investigated Sturmian words by looking at their palindromic factors. A *palindrome* is a finite word which can be read without distinction from left to right or from right to left; more formally, it is a fixed point of the *reversal* operator, which maps each finite word to that obtained reversing the order of its letters.

Palindromes play an essential role in the structure of Sturmian words. In fact, an important theorem of X. Droubay and G. Pirillo [28] shows that an infinite word is Sturmian if and only if it has exactly *one* palindromic factor of length n for n even, and *two* for n odd. Moreover, A. de Luca and F. Mignosi

[26] proved that the set of palindromic prefixes of all standard Sturmian words is equal to the set of *central words*, i.e., words having two periods p and q which are coprime, and length $p+q-2$. Central words satisfy remarkable structural properties; for instance, a central word w is such that wab and wba can be factorized as a product of two palindromes. The set St of factors of all Sturmian words is equal to the set of factors of all central words (cf. [26]).

Palindromes, and more specifically *palindrome closure* operators, are also involved in some natural generalizations of Sturmian words. The right (resp. left) palindrome closure $w^{(+)}$ (resp. $w^{(-)}$) of a finite word w is the shortest palindrome having w as a prefix (resp. suffix). By iterating the operation of adding a letter (from $\{a, b\}$) to the right and then taking the palindrome closure, one obtains at the limit either a periodic word, or a standard Sturmian word (see [21]). As an example, taking as *directive word* (the infinite sequence of letters used in the iterated palindrome closure) the sequence $(ab)^\omega = ababab\cdots$, we get the sequence of central words $a, aba, abaaba, \ldots$, converging to the infinite *Fibonacci word*

$$f = abaababaabaababaababaab\cdots ,$$

which is arguably the most famous Sturmian word. The process of iterated palindrome closure, when extended to larger alphabets, produces *standard episturmian* words, first introduced by X. Droubay, J. Justin, and G. Pirillo in [27]. Standard episturmian words enjoy a famous characterization, often taken as definition: an infinite word s is standard episturmian if and only if it is closed under reversal and every *left special* factor of s is a prefix of it. We recall that a factor of a finite or infinite word w is said left special if it admits at least two different "extensions" in w: u is left special in w if there exist distinct letters a and b such that au and bu are factors of w.

The equivalence between the above definitions of standard episturmian words is not preserved if one substitutes the reversal operator by an arbitrary *involutory antimorphism* of A^*, i.e., a composition of the reversal with a permutation of the alphabet A. Indeed, as we shall see in later chapters, such substitution leads to two different extensions of episturmian words, namely *ϑ-standard* words and *standard ϑ-episturmian* words. Both families are included in the larger class of *ϑ-standard words with seed.* As in the Sturmian case, all these words have a "non-standard" counterpart; thus for instance, an infinite word is *ϑ-episturmian* if there exists a standard ϑ-episturmian words having the same set of factors. Most results about standard words have a natural extension to the non-standard case.

Overview

In this thesis, we consider several topics related to Sturmian words and their generalizations. In **Chapter 1** we recall some basic definitions and results concerning combinatorics on words, and introduce the central notion of involutory antimorphism of a free monoid. This allows to consider *ϑ-palindromes*, natural generalizations of palindromes: they are the fixed points of some involutory antimorphism ϑ of A^*, and will have a fundamental role throughout this work.

In **Chapter 2**, we devote our attention to Sturmian words and their factors. We first give some basic definitions and properties about standard and central words; then in Section 2.2 we provide two new characterizations (cf. Theorems 2.2.3 and 2.2.8) of factors of Sturmian words, both related to their periodical structure. More specifically, they are based on properties of the *fractional root* of the finite word w being considered, that is, the prefix z_w of w whose length equals the minimal period of w. From the applicative point of view, the interest of such characterizations lies in the possibility of implementing two new and simple algorithms recognizing whether a finite word is a factor of some Sturmian word, with linear time complexity. A simple formula enumerating the finite Sturmian words which are primitive is also derived. We then focus, in Section 2.4, on palindromic factors of Sturmian words, or *Sturmian palindromes*. Some structural and combinatorial properties of the language of Sturmian palindromes are presented. In particular, two new characterizations of central words are given, and a remarkable characterization of Sturmian palindromes is proved.

The last section of Chapter 2 deals with the enumeration of Sturmian palindromes. A main theorem (cf. Theorem 2.5.1) gives a simple formula which permits to count for any $n \geq 0$ the Sturmian palindromes of length n. As a consequence, an interesting relation between the numbers of Sturmian palindromes of odd and even length is found. Moreover, it is shown that the number

$g(n)$ of Sturmian palindromes of length n has, for all $n \geq 0$, a lower bound of the order $n^{1+\alpha}$, where $\alpha = \log_3 2$. From this we derive that the densities of central words with respect to Sturmian palindromes, and of Sturmian palindromes with respect to factors of Sturmian words, both vanish asymptotically.

In **Chapter 3** we introduce pseudopalindrome closure operators, and study the properties of Sturmian and episturmian words in relation to palindrome closure. In Section 3.1, we discuss some general properties of the ϑ-palindrome closure operators. It is shown that the right and left ϑ-palindrome closures of a word w have the same minimal period. The main result of the section is Theorem 3.1.6, which states that a nonempty word w has the same minimal period of its ϑ-palindromic closures if and only if its fractional root z_w is a product of two ϑ-palindromes. In Section 3.2, we introduce the notion of elementary ϑ-palindrome action, which consists in appending a letter to a word and then taking the right ϑ-palindrome closure. Such actions can be naturally extended from letters to a finite or infinite word w by an iterative composition of the elementary ϑ-palindrome actions corresponding to the successive letters of w. If w is an infinite word, then, starting from the empty word, one generates an infinite word called ϑ-standard. If ϑ is the reversal operator, one obtains a standard episturmian word.

In Sections 3.3 and 3.4, we consider Sturmian and episturmian words respectively. In Section 3.3 we prove that both closures $w^{(+)}$ and $w^{(-)}$ of a finite Sturmian word w are Sturmian themselves, and share the same minimal period of w since the fractional root of w is symmetric, i.e., the product of two palindromes. Moreover, there exists a standard Sturmian word s such that $w^{(+)}$ and $w^{(-)}$ are both factors of s. From the preceding results, a new characterization of finite Sturmian words can be given in terms of the minimal period and of the right special factors of its right palindrome closure (cf. Theorem 3.3.9). Some of the previous results can be extended to episturmian words. In Section 3.4, we show that if w is a factor of some episturmian word, then so are $w^{(+)}$ and $w^{(-)}$. However, in general, the minimal period of $w^{(+)}$ and $w^{(-)}$ is different from that of w, since the fractional root of w can be non-symmetric.

In **Chapter 4**, we analyse different possible generalizations of episturmian words, based on involutory antimorphisms. The first family is the one of ϑ-standard words, constructed by iterated ϑ-palindrome closure. The main result is that any ϑ-standard word is a morphic image, by an injective morphism (depending on ϑ), of the standard episturmian word having the same directive word. This allows to extend the closure property to factors of ϑ-standard words too: if w is a factor of some ϑ-standard word, then so are its left and right ϑ-palindrome closures, and there exists a 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. A generalization of the method for constructing ϑ-standard words is introduced in Section 4.2, by assuming that ϑ can vary among all involutory antimorphisms of A^* at each step of the iterating process, which is directed by a bi-sequence of letters and operators. In this way, one gets a wider family of infinite words, including the well-known Thue-Morse word on two symbols.

In Section 4.3 we introduce the class of ϑ-standard words with *seed*. They are infinite words obtained by iterated ϑ-palindrome closure, 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 $\Delta' = x_1x_2 \cdots x_n \cdots$. We also 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 result suggests another generalization of episturmian words, introduced in Section 4.4: the class of infinite words which are closed under ϑ and have all sufficiently long left special factors as prefixes. This turns out (cf. Theorem 4.4.6) to be the same as the family of ϑ-standard words with seed, and has the noteworthy subclass of *standard ϑ-episturmian* words, i.e., infinite words which are closed under ϑ and have *all* left special factors as prefixes.

The structure of such words is studied more in detail in Section 4.5. In particular, it is proved that every standard ϑ-episturmian word s can be uniquely factorized with unbordered ϑ-palindromes; As a consequence, it is proved that s is a morphic image, under an injective morphism, of the standard episturmian word whose directive word is the *subdirective word* of s, i.e., the infinite word formed by the letters immediately following the ϑ-palindromic prefixes of s. Finally, the intersection of the two families of ϑ-standard and of standard ϑ-episturmian words is fully characterized; it is a proper subclass of both families.

In conclusion, we mention that several results of this thesis were already published in [25, 23, 24, 12, 11].

Acknowledgments

I wish to express all my gratitude to my collaborators, who made this work possible. First of all to Prof. Aldo de Luca, my almost-homonymous supervisor, who introduced me to combinatorics on words and has been a wise and patient guide through all these years; then to Michelangelo Bucci, my dear friend and precious point of reference; and to our new coauthor, Luca Q. Zamboni from UNT, whose ideas and interesting point of view allowed a very stimulating research experience.

Many thanks to Jean Berstel, Arturo Carpi, and Maurizio Laporta for their useful comments and suggestions.

Finally, I owe everything to my friends, within and without C&C, and to my family. I thank them all whole-heartedly.

Chapter 1

Preliminaries

In this chapter we review the fundamental algebraic and combinatorial tools needed to state and prove the main results of this thesis.

1.1 Basic algebraic concepts

As is well known (see for instance [17]), a *semigroup* S is a set in which an associative binary operation (product) is defined.

A *monoid* M is a semigroup having an *identity* element 1_M such that $1_M x = x 1_M = x$ for all $x \in M$. A subsemigroup N of M is a *submonoid* if $1_M \in N$.

The product operation on a semigroup S can be naturally extended to the powerset $\mathcal{P}(S)$: given $X, Y \subseteq S$, we define

$$XY = \{xy \in S \mid x \in X \text{ and } y \in Y\} .$$

It is also common to define left and right *quotients*, by setting

$$X^{-1}Y = \{w \in S \mid Xw \cap Y \neq \emptyset\} ,$$

and

$$YX^{-1} = \{w \in S \mid wX \cap Y \neq \emptyset\} .$$

We shall often confuse singletons and their elements, when this does not lead to ambiguity. For instance, if $x \in S$ and $Y \subseteq S$, by xY we will mean the set $\{x\}Y$.

The *subsemigroup generated by* $X \subseteq S$ is the smallest subsemigroup of S containing X, and coincides with

$$X^+ = \bigcup_{n>0} X^n \; .$$

Similarly, the *submonoid generated by* $X \subseteq M$ is equal to

$$X^* = \bigcup_{n\geq 0} X^n \; ,$$

where conventionally $X^0 = \{1_M\}$.

Given two semigroups S, S', a *morphism* (resp. *antimorphism*) φ from S to S' is a map

$$\varphi : S \longrightarrow S'$$

such that $\varphi(xy) = \varphi(x)\varphi(y)$ (resp. $\varphi(xy) = \varphi(y)\varphi(x)$) for all $x, y \in S$. A monoid (anti-)morphism $\varphi : M \to M'$ is a semigroup (anti-)morphism such that $\varphi(1_M) = 1_{M'}$. An *isomorphism* is a bijective morphism, and an *automorphism* of M is an isomorphism between M and itself. When $\varphi : M \to M'$ is a morphism or antimorphism and $x \in M$, we shall often use the exponential notation x^φ for $\varphi(x)$.

A semigroup S (resp. monoid M) is *free over* $X \subseteq S$ (resp. $X \subseteq M$) if every element of X^+ admits a unique factorization over X, and $X^+ = S$ (resp. $X^* = M$). Free semigroups (monoids) over sets of the same cardinality are isomorphic.

1.2 Finite and infinite words

The free monoid of words

Let A be a nonempty finite set, or *alphabet*, whose elements are called *letters*. The set of finite sequences of letters, or *words* over A, can be naturally endowed with the binary operation of *concatenation*. The semigroup A^+ thus obtained is free over A: a word $w \in A^+$ can be written uniquely as a product of letters $w = a_1 a_2 \cdots a_n$, with $a_i \in A$, $i = 1, \ldots, n$. Therefore A^+ is called *the* free semigroup over A. The free monoid A^* is obtained by adding an identity element, the *empty word* $\varepsilon = 1_{A^*}$, to A^+: $A^* = A^+ \cup \{\varepsilon\}$.

Let $w = a_1 \cdots a_n \in A^+$, where $a_i \in A$ for $1 \leq i \leq n$. The integer n is the *length* of w, denoted by $|w|$. It is natural to set $|\varepsilon| = 0$.

A word u is a *factor* of $w \in A^*$ if $w = rus$ for some words r and s. In the special case $r = \varepsilon$ (resp. $s = \varepsilon$), u is called a *prefix* (resp. *suffix*) of w. A factor u of w is *proper* if $u \neq w$; it is *median* if $w = rus$ with $|r| = |s|$. We denote respectively by $\operatorname{Fact} w$, $\operatorname{Pref} w$, and $\operatorname{Suff} w$ the sets of all factors, prefixes, and suffixes of the word w.

A subset of A^* is often called a *language* over A. For $Y \subseteq A^*$, $\operatorname{Pref} Y$, $\operatorname{Suff} Y$, and $\operatorname{Fact} Y$ will denote respectively the languages of prefixes, suffixes, and factors of all the words of Y; in symbols,

$$\operatorname{Fact} Y = \bigcup_{w \in Y} \operatorname{Fact} w \ ,$$

and similarly for $\operatorname{Pref} Y$ and $\operatorname{Suff} Y$.

A *code* over A is a language $Z \subseteq A^*$ such that the monoid Z^* is free over Z. Thus Z is a code if and only if whenever $z_1, z_2, \ldots, z_n, z'_1, \ldots, z'_m \in Z$ are such that

$$z_1 \cdots z_n = z'_1 \cdots z'_m \ ,$$

then $n = m$ and $z_i = z'_i$ for $i = 1, \ldots, n$. A *prefix* (resp. *suffix*) *code* is a subset of A^+ with the property that none of its elements is a proper prefix (resp. suffix) of any other. Any prefix (or suffix) code is in fact a code. A *biprefix* code is a code which is both prefix and suffix.

Borders and periods

A factor of $w \in A^*$ is called a *border* of w if it is both a prefix and a suffix of w. A word is called *unbordered* if its only proper border is ε. Since the set of proper borders of the empty word is empty, coherently with the above definition we do not consider ε unbordered.

A positive integer p is a *period* of $w = a_1 \cdots a_n$ ($a_i \in A$, $i = 1, \ldots, n$) if whenever $1 \leq i, j \leq |w|$ one has that

$$i \equiv j \pmod p \Longrightarrow a_i = a_j \ .$$

Note that with this definition, any $n \geq |w|$ is a period of w. As is well known and quite evident (cf. [38]), a word w has a period $p \leq |w|$ if and only if it has a border of length $|w| - p$. We denote by π_w the minimal period of w, and set $\pi_\varepsilon = 1$. Thus a word w is unbordered if and only if $\pi_w = |w|$. If w is nonempty, then its *fractional root* z_w is its prefix of length $|z_w| = \pi_w$. We can write any nonempty word w as

$$w = z_w^k z'$$

where z_w is the fractional root of w, the integer $k \geq 1$ is sometimes called the *order* of w, and z' is a proper prefix of z_w.

We recall the following fundamental result about periodicity (cf. [38]):

Theorem 1.2.1 (Fine and Wilf). *If a word w has two periods p and q, and $|w| \geq p + q - \gcd(p, q)$, then w has also the period* $\gcd(p, q)$.

Infinite words and limits

An *infinite word* (from left to right) x over the alphabet A is just an infinite sequence of letters, i.e., a mapping $x : \mathbb{N}_+ \longrightarrow A$ where $\mathbb{N}_+$ is the set of positive integers. One can represent x as

$$x = x_1 x_2 \cdots x_n \cdots ,$$

where for any $i > 0$, $x_i = x(i) \in A$. A (finite) *factor* of x is either the empty word or any sequence $x_i \cdots x_j$ with $i \leq j$, i.e., any block of consecutive letters of x. If $i = 1$, then u is a *prefix* of x. We denote by $\operatorname{Fact} x$ and $\operatorname{Pref} x$ the sets of finite factors and prefixes of x respectively.

The product between a finite word w and an infinite one x is naturally defined as the infinite word wx having w as a prefix and $x_{j-|w|}$ as its j-th letter, for all $j > |w|$. The set of all infinite words over A is denoted by A^ω. We also set $A^\infty = A^* \cup A^\omega$.

A metric on A^ω can be defined by setting $d(x, x) = 0$ for $x \in A^\omega$, and

$$d(x, y) = 2^{-\ell}$$

for $y \neq x$, where $\ell = \max\{n \in \mathbb{N} \mid \operatorname{Pref} x \cap \operatorname{Pref} y \cap A^n \neq \emptyset\}$ is the length of the maximal common prefix of x and y. This metric induces the product topology on $A^\omega = A^{\mathbb{N}_+}$ (where A is discrete), making it a compact, perfect, and totally disconnected metric space, that is, a *Cantor space* (cf. [43]). The metric d can be "extended" to the whole A^∞ in the following way: define (as above) the metric d' on $(A')^\omega$, where $A' = A \cup \{\$\}$ and $\$ \notin A$; then identify any $w \in A^*$ with the infinite word $w\$^\omega$. In this way A^∞ is regarded as a subspace of $(A')^\omega$.

The main benefit of topology for our purposes is the possibility of taking limits of sequences. We recall that convergence with respect to the product topology is *pointwise*, so that a sequence of words $(z_m)_{m\geq 0}$ in A^∞ converges to an infinite word $x = x_1 \cdots x_n \cdots$ if and only if for any $k > 0$, there exists some $N \geq 0$ such that for all $n \geq N$, the k-th letter of z_n exists (i.e., $z_n \in A^\omega$ or $|z_n| \geq k$) and is equal to x_k. For instance, the sequence

$$(a^m b)_{m\geq 0}$$

converges to the infinite word $a^\omega = aaa\cdots$. A wide family of convergent sequences, which will appear frequently in the following chapters, is made of all sequences of finite words $(z_m)_{m\geq 0}$ such that for sufficiently large n, the word z_n is a prefix of z_{n+1}.

For any $Y \subseteq A^*$, Y^ω denotes the set of infinite words which can be factorized by the elements of Y. The above example shows that an infinite word which is the limit of a sequence of words of Y^* need not be in Y^ω (take $Y = a^*b$); however, it is in Y^ω if Y is finite.

Further definitions and properties

Let $w \in A^\infty$. An *occurrence* of a factor u in w is any pair $(\lambda, \rho) \in A^* \times A^\infty$ such that $w = \lambda u \rho$. If $a \in A$ and $w \in A^*$, $|w|_a$ denotes the number of occurrences of a in the word w; trivially we have

$$|w| = \sum_{a\in A} |w|_a \; .$$

For $w \in A^\infty$, $\operatorname{alph} w$ denotes the set of letters occurring in w, that is, $\operatorname{alph} w = \{a \in A \mid |w|_a > 0\}$.

Let $s \in A^\infty$ and $w, u \in \operatorname{Fact} s$. We call w a *first return* to u in s if w contains exactly two distinct occurrences of u, one as a prefix and the other as a suffix, i.e.,

$$w = u\lambda = \mu u \quad \text{with } \lambda, \mu \in A^+ \text{ and } w \notin A^+ u A^+ \; .$$

We observe that in such a case, $wu^{-1} = \mu$ is usually called a *return word over* u in s (see [29]). We call the integer $|\mu|$ the *shift* of the first return. An infinite word s is said *uniformly recurrent* if for any $v \in \mathrm{Fact}\, s$, the shifts of the 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. We recall that a factor v of a (finite or infinite) word w is called *right special* if it has at least two distinct right extensions in w, i.e., there exist at least 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.

We denote by R_w the smallest integer k, if it exists, such that w has no right special factor of length k (and we set $R_w = \infty$ otherwise, that is, when w is an infinite word having arbitrarily long right special factors). The following noteworthy inequality (cf. [22]) relates the minimal period π_w of a finite word w and R_w:

$$\pi_w \geq R_w + 1 . \tag{1.1}$$

Symmetrically, one can introduce the parameter L_w as the minimal length for which w has no *left* special factors; L_w satisfies $\pi_w \geq L_w + 1$ too.

A finite word w is *primitive* if it cannot be written as a power $w = u^k$ with $k > 1$. Clearly any unbordered word is primitive, but the converse is false: consider for instance the word aba. We denote by $\pi(A^*)$ the set of all primitive words over A. As is well known (cf. [38]), for any nonempty word w there exists a unique primitive word u such that $w = u^k$ for some $k \geq 1$. Such a u is usually called the *(primitive) root* of w and denoted by $\sqrt{w}$.

Two words $u, v \in A^*$ are *conjugate* if there exist $\lambda, \mu \in A^*$ such that $u = \lambda\mu$ and $v = \mu\lambda$. Conjugacy is an equivalence relation in A^*; we write $u \sim v$ if u and v are conjugate.

Suppose that $\leq$ is a total order on A. One can extend this order to the *lexicographic* order on A^* by letting, for all $v, w \in A^*$,

$$v \leq w \Longleftrightarrow (v \in \mathrm{Pref}\, w \;\text{ or }\; v = uav',\, w = ubw') ,$$

for some $u, v', w' \in A^*$ and $a, b \in A$ such that $a < b$.

A word is called a *Lyndon* (resp., *anti-Lyndon*) word if it is primitive and minimal (resp., maximal) in its conjugacy class, with respect to the lexicographic order. For instance, if $a < b$ then $w = aabab$ is a Lyndon word, for its conjugates ($ababa$, $babaa$, $abaab$, and $baaba$) are all lexicographically greater than w.

In the sequel, we shall need the two following simple lemmas; we report the proofs for the sake of completeness.

Lemma 1.2.2. *A word $w \in A^*$ has the period $p \leq |w|$ if and only if all its factors having length p are in the same conjugacy class.*

Proof. The case $w = \varepsilon$ is trivial. Then suppose that p is a period of $w = a_1 \cdots a_n$, $a_i \in A$, $i = 1, \ldots, n$. Let u be a factor of w of length p. By the definition of period, there exists a positive integer $i \leq p$ such that $u = a_i a_{i+1} \cdots a_p a_1 a_2 \cdots a_{i-1}$, so that u is a conjugate of $a_1 a_2 \cdots a_p$.

The converse is an easy consequence of the following fact: if $x, y \in A$ and $u \in A^*$, then $xu \sim uy$ if and only if $x = y$. Therefore, if all factors of w of length p are conjugate, one derives that $a_i = a_{i+p}$ for all i such that $1 \leq i \leq n-p$. □

Lemma 1.2.3. *A word $w \in A^*$ is primitive if and only if $\pi_{w^k} = |w|$ for any integer $k \geq 2$.*

Proof. Let w be a primitive word, and suppose that w^k has a period $q \leq |w|$. Since $|w|$ is a period of w^k and $|w^k| = k|w| > |w| + q$, by Theorem 1.2.1, w^k, as well as w, has also the period $d = \gcd(q, |w|)$. Thus $w = u^{|w|/d}$ for some u; this implies $|w|/d = 1$ and then $q = |w|$, as w is primitive.

Conversely, suppose $w \in A^*$ is not primitive. If $w = \varepsilon$, then

$$\pi_{w^k} = \pi_\varepsilon = 1 \neq 0 = |w| \; .$$

Let then $w \in A^+$ and let u be its primitive root. Clearly $|u|$ is a period of w^k, and $|u| < |w|$. □

We remark that also the *fractional* root z_w of a nonempty word w is trivially primitive. Hence, by Lemma 1.2.3 we obtain that for any $w \in A^+$ and $k \geq 2$,

$$\pi_w = \pi_{z_w^k} \; . \tag{1.2}$$

1.3 Antimorphisms of a free monoid

Uniqueness and involutions

We recall that any (anti-)morphism whose domain is the free monoid A^* is uniquely determined by the images of the letters. Formally, for any monoid M and any map $\varphi : A \to M$, there exists a unique morphism $\hat{\varphi} : A^* \to M$ (resp. a unique antimorphism $\bar{\varphi} : A^* \to M$) that extends φ, i.e., such that $\hat{\varphi}|_A = \varphi$ (resp. $\bar{\varphi}|_A = \varphi$). This property characterizes free monoids, and is usually taken as the definition of free objects in the frame of category theory (cf. [41]).

A morphism or antimorphism $\varphi : A^* \to A^*$ is *involutory* if it is an involution of A^*, that is, if $\varphi^2 = \mathrm{id}$.

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

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

One sets $\tilde{\varepsilon} = \varepsilon$. The map $R : A^* \to A^*$ defined by $w^R = \tilde{w}$ for any $w \in A^*$, called *reversal operator*, is clearly an involutory antimorphism of A^*.

Let τ be an involution of the alphabet A. Clearly, it can be regarded as a map $\tau : A \to A^*$, and then extended to a unique automorphism $\hat{\tau}$ of the free monoid A^*. The map $\vartheta = \hat{\tau} \circ R = R \circ \hat{\tau}$ is the unique involutory antimorphism of A^* extending the involution τ. One has, for $w = a_1 \cdots a_n$, $a_i \in A$, $i = 1, \ldots, n$,

$$w^{\vartheta} = a_n^{\tau} \cdots a_1^{\tau} \ .$$

Any involutory antimorphism of A^* can be constructed in this way; for example, the reversal R is obtained by extending the identity map of A.

If $A = \{a, b\}$, then there exist only two involutory antimorphisms, namely, the reversal R and the antimorphism $e = E \circ R$, called *exchange antimorphism*, extending the exchange map E defined on A as $E(a) = b$ and $E(b) = a$.

If the alphabet A has cardinality n, then the number of all involutory antimorphisms of A^* equals the number of the involutory permutations over n elements. As is well known, this number is given by

$$n! \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{1}{2^k (n-2k)! k!}$$

(sequence A000085 in [46]).

(Pseudo-)palindromes

Let ϑ be an involutory antimorphism of A^*. A word $w \in A^*$ is called *ϑ-palindrome* if it is a fixpoint of ϑ, i.e., $w = w^\vartheta$. The set of all ϑ-palindromes of A^* is denoted by $PAL_\vartheta(A)$ or simply PAL_ϑ when there is no ambiguity.

An R-palindrome is usually called *palindrome* and PAL_R is denoted by PAL. In less precise terms, a word which is a ϑ-palindrome with respect to a given but unspecified involutory antimorphism ϑ, is also called *pseudopalindrome*.

Examples 1.3.1. The English word *racecar* is a palindrome.

Let $A = \{a, b\}$, e be the exchange antimorphism, and $w = abaabb$. One has $w^e = aabbab$. The word $abbaab$ is an e-palindrome.

Let $A = \{a, b, c\}$ and τ be the involutory permutation defined as $\tau(a) = b$, $\tau(b) = a$, and $\tau(c) = c$. Setting $\vartheta = \tau \circ R$, the word $abcacbcab$ is a ϑ-palindrome.

A word is called *ϑ-symmetric* if it is the product of two ϑ-palindromes. An R-symmetric word is simply called *symmetric*. In particular, any ϑ-palindrome is ϑ-symmetric.

Some combinatorial properties of symmetric words were studied in [20], and more recently in [9], where the term symmetric was used. One easily verifies that all words on the alphabet $\{a, b\}$ of length ≤ 5 are symmetric. The word $w = abaabb$ is not symmetric but it is e-symmetric, because it is the product of the two words ab and $aabb$ which are e-palindromes.

In the remaining part of this section, we will assume that ϑ is a fixed involutory antimorphism of A^*. To simplify the notation, for any $w \in A^*$, we shall denote by $\bar{w}$ the word w^ϑ, so that for all $u, v \in A^*$ one has

$$|\bar{u}| = |u|\,, \quad \overline{uv} = \bar{v}\bar{u}\,, \quad \text{and} \quad \bar{\bar{u}} = u\,.$$

Lemma 1.3.2. *A word w is a conjugate of $\bar{w}$ if and only if it is ϑ-symmetric.*

Proof. If $w = \alpha\beta$ with $\alpha, \beta \in PAL_\vartheta$, then $\bar{w} = \beta\alpha$, so that $w \sim \bar{w}$. Conversely, suppose that w and $\bar{w}$ are conjugate. One can write $w = \lambda\mu$ and $\bar{w} = \mu\lambda$ for some $\lambda, \mu \in A^*$. Thus $w = \bar{\lambda}\bar{\mu} = \lambda\mu$. Since $|\lambda| = |\bar{\lambda}|$, one obtains $\lambda = \bar{\lambda}$ and $\mu = \bar{\mu}$. □

Lemma 1.3.3. *A ϑ-palindrome $w \in A^+$ has a period $p \leq |w|$ if and only if it has a ϑ-palindromic prefix (suffix) of length $|w| - p$.*

Proof. If w has a period $p \leq |w|$, then it has a border v of length $|w| - p$, so that we can write $w = \lambda v = v\mu$ for some words λ and μ. Since w is a ϑ-palindrome, one has

$$w = v\mu = \bar{v}\bar{\lambda} \ .$$

Therefore, $v = \bar{v}$. Conversely, if the ϑ-palindrome w has the ϑ-palindromic prefix v, one has

$$w = v\mu = \bar{\mu}v \ ,$$

so that v is a border of w and $|w| - |v|$ is a period of w. □

Lemma 1.3.4. *Let $w \in A^+$ and z_w be its fractional root. The word $z_{\bar{w}}$ is a conjugate of $\bar{z}_w$.*

Proof. Let w be a nonempty word. Since ϑ acts on the alphabet as a permutation, one derives that p is a period of w if and only if it is a period of $\bar{w}$. Therefore one has $\pi_w = \pi_{\bar{w}}$. We can write $w = z_w^k z'$ with $k \geq 1$ and z' a proper prefix of z_w, and

$$\bar{w} = \bar{z}' \bar{z}_w^k = z_{\bar{w}}^h z''$$

with $h \geq 1$ and z'' a proper prefix of $z_{\bar{w}}$. Since $|w| = |\bar{w}|$ and $|\bar{z}_w| = |z_{\bar{w}}| = \pi_{\bar{w}}$, one has $h = k$ and, by Lemma 2.4.13, $\bar{z}_w \sim z_{\bar{w}}$. □

Corollary 1.3.5. *Let $w \in A^+$ be a ϑ-palindrome having a period $p \leq |w|$. Any factor u of w of length p is ϑ-symmetric. In particular, z_w is ϑ-symmetric.*

Proof. Since $w = \bar{w}$ and $|u| = p$, by Lemma 2.4.13 one has $u \sim \bar{u}$. Hence, by Lemma 1.3.2 one obtains $u \in PAL_\vartheta^2$. As $|z_w| = \pi_w$, one derives $z_w \in PAL_\vartheta^2$. □

Chapter 2

Sturmian sequences

Sturmian words were first considered in the 18th century by J. Bernoulli III, in his astronomical studies. Several authors later developed the subject from different points of view, but the first systematic study was given in 1940 by M. Morse and G. A. Hedlund (cf. [44]). They were also the first to use the name Sturmian, in honor of C. F. Sturm.

By definition, an infinite word is *Sturmian* if for each $n \in \mathbb{N}$ it has $n+1$ distinct factors of length n. This implies that a Sturmian word is on a two-letter alphabet, that will be $\mathcal{A} = \{a, b\}$ for the rest of this chapter (we shall keep using a non-calligraphic A for a generic alphabet). As is well known [39], an infinite binary word x is Sturmian if and only if for any $n \geq 0$ there is only one right special factor of x of length n.

A famous theorem by Morse and Hedlund (cf. [43]) states that an infinite word s has less than $n+1$ factors for some $n \geq 0$ if and only if it is *eventually periodic*, that is, writable as $s = uv^\omega$ for some finite words u, v. Thus Sturmian words have the smallest possible number of factors of each length, among all infinite words which are not eventually periodic.

A first description of the structure of Sturmian words was given in [44], where the following well-known characterization is found: an infinite word $s \in \mathcal{A}^\omega$ is Sturmian if and only if it is not eventually periodic and it is *balanced*, i.e., it satisfies, for all $n \geq 0$ and $u, v \in \mathcal{A}^n \cap \operatorname{Fact} s$,

$$||u|_a - |v|_a| \leq 1 \,. \tag{2.1}$$

2.1 Standard and central Sturmian words

An equivalent geometrical definition of Sturmian words can be given in terms of *cutting sequences.* In fact, a Sturmian word can be defined by considering the sequence of cuts in a squared lattice ($\mathbb{N} \times \mathbb{N}$) made by a ray having a slope which is an irrational number α. A horizontal cut is denoted by the letter b, a vertical by a, and a cut with a corner by ab or ba.

A Sturmian word represented by a ray starting from the origin is usually called *standard* or *characteristic.* We shall denote by c_α the standard Sturmian word associated with the irrational slope α. Standard Sturmian words can be equivalently defined as follows. For any sequence $d_0, d_1, \ldots, d_n, \ldots$ of integers such that $d_0 \geq 0$ and $d_i > 0$ for $i > 0$, one defines, inductively, the sequence of words $(s_n)_{n \geq 0}$ where

$$s_0 = b,\ s_1 = a,\ \text{ and } s_{n+1} = s_n^{d_{n-1}} s_{n-1},\ \text{ for } n \geq 1\ . \tag{2.2}$$

The sequence $(s_n)_{n \geq 0}$ converges to a limit s which is an infinite standard Sturmian word. More precisely, one has $s = c_\alpha$, where the slope α is given by the continued fraction

$$\alpha = \cfrac{1}{d_0 + \cfrac{1}{d_1 + \cfrac{1}{\ddots}}} = [0; d_0, d_1, \ldots]$$

(see for instance [39]). Any standard Sturmian word can be generated in this way. If $d_i = 1$ for all $i \geq 0$, one obtains the famous *Fibonacci word*

$$f = abaababaabaababaababaa \cdots\ ,$$

whose slope is the inverse of the golden ratio.

We shall denote by *Stand* the set of all the words s_n, $n \geq 0$ of any sequence $(s_n)_{n \geq 0}$ constructed by the previous rule (2.2). Any word of *Stand* is called *finite standard (Sturmian) word.* We recall the following characterization of *Stand* given in [26]:

$$Stand = \mathcal{A} \cup (PAL^2 \cap PAL\{ab, ba\})\ , \tag{2.3}$$

i.e., a word $w \in \mathcal{A}^*$ is standard if and only if it is a letter or it satisfies the following equation:

$$w = \alpha\beta = \gamma xy \ ,$$

with $\alpha, \beta, \gamma \in PAL$ and $\{x, y\} = \mathcal{A}$.

A finite word w is called *central* if it has two periods p and q such that $\gcd(p, q) = 1$ and $|w| = p + q - 2$. Conventionally, the empty word ε is central (in this case, $p = q = 1$). Central words are over a two-letter alphabet. The set of all central words over $\mathcal{A} = \{a, b\}$ is usually denoted by PER. It is well known (see [26, 39]) that the set PER coincides with the set of palindromic prefixes of all standard Sturmian words. In the remaining part of this section we recall some properties of standard and central words which will be useful in the sequel.

The following important characterization of central words holds (see for instance [15]):

Proposition 2.1.1. *A word w is central over $\mathcal{A}$ if and only if w is a power of a letter of $\mathcal{A}$ or it satisfies the equation*

$$w = w_1 ab w_2 = w_2 ba w_1$$

for some words w_1 and w_2. Moreover, in this latter case, w_1 and w_2 are central words over $\mathcal{A}$, $p = |w_1| + 2$ and $q = |w_2| + 2$ are coprime periods of w, and $\min\{p, q\}$ is the minimal period of w.

Example 2.1.2. Let $w = aabaabaa \in PER$. We have

$$w = a(ab)aabaa = aabaa(ba)a \ ,$$

with $3 = \pi_w = |a| + 2$ and $7 = |aabaa| + 2$ being coprime periods of w, and $|w| = 8 = 3 + 7 - 2$.

From (2.3) and the preceding proposition, one easily derives (cf. [26]) that

$$Stand = \mathcal{A} \cup PER\{ab, ba\} \ , \tag{2.4}$$

i.e., any finite standard Sturmian word which is not a single letter is obtained by appending ab or ba to a central word. Conversely, any central word is obtained by deleting the last two letters of a standard word.

Let St be the set of finite Sturmian words, i.e., factors of infinite Sturmian words over the alphabet $\mathcal{A} = \{a, b\}$. We recall that for any infinite Sturmian word there exists an infinite standard Sturmian word having the same set of factors (cf. [39]). Therefore one easily derives that

$$St = \mathrm{Fact}(Stand) = \mathrm{Fact}(PER)\ . \tag{2.5}$$

Lemma 2.1.3 (see [15]). *If a central word w has the factor x^n, with $x \in \mathcal{A}$ and $n > 0$, then x^{n-1} is a prefix (and suffix) of w.*

Proposition 2.1.4 (see [45]). *A word w is central if and only if wab and wba are conjugate.*

Now let us suppose that the alphabet $\mathcal{A}$ is totally ordered by setting $a < b$.

Proposition 2.1.5 (see [3]). *The set $\mathcal{A} \cup aPERb$ is equal to the set of all Lyndon words which are Sturmian. Similarly, $\mathcal{A} \cup bPERa$ is the set of anti-Lyndon Sturmian words.*

Proposition 2.1.6 (see [31]). *A Sturmian word is unbordered if and only if it is a Lyndon or anti-Lyndon word.*

From Propositions 2.1.4 and 2.1.5, one derives the following interesting characterization of words conjugate of a standard word.

Proposition 2.1.7. *A primitive word $z \notin A$ is a conjugate of a standard word if and only if the Lyndon and the anti-Lyndon words in its conjugacy class have the same proper median factor of maximal length.*

Proof. Let z be a primitive word of length $|z| > 1$. Let s be a standard word conjugate to z. By (2.4), s can be written as $s = vxy$, with $v \in PER$ and $\{x, y\} = \mathcal{A}$. By Proposition 2.1.4, one derives that z is a conjugate of avb and bva. From Proposition 2.1.5, avb and bva are, respectively, a Lyndon and an anti-Lyndon word, so that the necessity is proved.

Conversely, let $z \in A^*$ and suppose that the Lyndon and the anti-Lyndon words in the conjugacy class of z can be written respectively as atb and bta, with $a, b \in A$ and $a < b$. By Proposition 2.1.4, one has that $t \in PER$, so that by (2.4), z is a conjugate of $tab \in Stand$. □

2.2 Finite Sturmian words and periodicity

In this section we give two characterizations of finite Sturmian words, based on properties of their fractional root. We need some preliminary propositions. The first one gives some characterizations of the words w such that $w^2 \in St$ (such words have been called *cyclic balanced* in [16]). The equivalence of some of the conditions in Proposition 2.2.1 has recently been proved in [16] (see also [32]). We report here a more direct and simple proof for the sake of completeness.

Proposition 2.2.1. *Let w be a word. The following conditions are equivalent:*

1. $w^2 \in St$,
2. $w^* \subseteq St$,
3. *every conjugate of w^2 is Sturmian,*
4. *every conjugate of w is Sturmian,*
5. *the primitive root of w is a conjugate of a standard Sturmian word.*

Proof. *1.* $\Rightarrow$ *2.* Let $n > 2$. Any two factors of w^n of length $k > |w|/2$ overlap, thus it suffices to verify the balance condition only for factors of w^n of length $k \leq |w|/2$, which is satisfied because such words are also factors of $w^2 \in St$.

2. $\Rightarrow$ *3.* This is trivial, since any conjugate of w^2 is a factor of w^3.

3. $\Rightarrow$ *4.* This is trivial too, because the square of a conjugate of w is just a conjugate of w^2.

4. $\Rightarrow$ *5.* Let u be the primitive root of w. If every conjugate of w is Sturmian, then so is every conjugate of u. Hence it suffices to prove that if w is primitive, then it has a conjugate which is a standard word. Indeed, there exists a unique conjugate of w which is a Lyndon word, say w'. Since w' is Sturmian, by Proposition 2.1.5 one has that w' is either a letter or a word avb with $v \in PER$. In the former case, the desired standard conjugate is w' itself; in the latter case, one can take vba.

5. $\Rightarrow$ *1.* Let u be the primitive root of $w = u^k$; if v is a standard word in its conjugacy class, from equations (2.2) and (2.5) one derives that $v^2 \in St$.

Since *1.* $\Rightarrow$ *3.* and u^2 is a conjugate of v^2, one has $u^2 \in St$. As *1.* $\Rightarrow$ *2.*, this implies $w^2 = u^{2k} \in St$. □

Let $w, u \in A^*$ with w unbordered; the word wu is called a *Duval extension* of w if no unbordered factor of wu is longer than w.

Proposition 2.2.2 (see [42]). *Every Duval extension wu of a Sturmian unbordered word w has the period $|w|$.*

We are now in the position of giving our first characterization of finite Sturmian words.

Theorem 2.2.3. *A nonempty word is Sturmian if and only if its fractional root is a conjugate of a standard word.*

Proof. Let w be a word. If its fractional root z_w is a conjugate of a standard word, then by Proposition 2.2.1, $z_w^* \subseteq St$, so that $w \in \operatorname{Fact} z_w^* \subseteq St$.

Conversely, let s be an unbordered factor of $w \in St$ of maximal length. One has $w = usv$ for suitable $u, v \in \mathcal{A}^*$. The word sv is a Duval extension of s, by the maximality of s. Since $\tilde{s}$ is unbordered too, and again by the maximality of s, the word $\tilde{s}\tilde{u} = \widetilde{us}$ is a Duval extension of $\tilde{s}$. From Proposition 2.2.2, one gets that both sv and $\widetilde{us}$ have the period $|s|$. This implies that also us has the period $|s|$.

By Lemma 1.2.2, all factors of us and sv having length $|s|$ are conjugates of s. Since any factor of w of length $|s|$ is either a factor of us or of sv, and s is a factor of both, we deduce from Lemma 1.2.2 that the whole w has the period $|s|$. Moreover, such period is minimal, because

$$|s| = \pi_s \leq \pi_w \leq |s| \ .$$

By Lemma 1.2.2, z_w is a conjugate of s; since s is an unbordered Sturmian word, by Proposition 2.1.6 it is a Lyndon (or anti-Lyndon) word, and therefore, by Proposition 2.1.5 it is in the set $\mathcal{A} \cup aPERb \cup bPERa$. Hence s, as well as z_w, is a conjugate of a standard word, which proves the assertion. □

Examples 2.2.4. Let w be the word $aababaa$. Its fractional root $z_w = aabab$ is a conjugate of the standard word $ababa$, so that w is Sturmian.

Let $r = baabb$. In the conjugacy class of its root $z_r = baab$ there is no standard word, so that r is not Sturmian.

Corollary 2.2.5. *Let w be a nonempty word and z_w be its fractional root. Then w is a finite Sturmian word if and only if so is z_w^2.*

Proof. This is a straightforward consequence of the preceding theorem and of Proposition 2.2.1. ☐

The following proposition improves upon a result in [22].

Proposition 2.2.6. *Let w be a word. If $\pi_w = R_w + 1$, then w is Sturmian.*

Proof. Let $w \in A^*$. If $\pi_w = 1$, the result is trivially true. Thus we assume $\pi_w = R_w + 1 > 1$, so that there exists a right special factor s of w such that $|s| = \pi_w - 2$. Hence, there exist letters $a, b \in A$ such that $a \neq b$ and $sa, sb \in \operatorname{Fact} w$. The words sa and sb cannot be both suffixes of w, so we suppose, without loss of generality, that sa is not. Therefore one has either $saa \in \operatorname{Fact} w$ or $sac \in \operatorname{Fact} w$ with $c \neq a$. Since $|saa| = |sac| = \pi_w$, these two possibilities imply, respectively:

$$w \in \operatorname{Fact}((saa)^*) \tag{2.6}$$

or

$$w \in \operatorname{Fact}((sac)^*) \,. \tag{2.7}$$

We first show that (2.6) cannot hold. By contradiction, assume that it does hold. Since sb is a factor of w, it has to be a factor of $saas$ as well. We clearly have $sb \neq sa$, thus there exist $u, v \in A^*$ and $x \in A$ such that $saas = uxsbv$. The words u and v are respectively a prefix and a suffix of s, and $|u| + |v| = |saas| - |xsb| = 2|s| + 2 - |s| - 2 = |s|$. Therefore $s = uv$ and $vaau = xuvb$. But this is a contradiction, because $|vaau|_a > |xuvb|_a$.

Equation (2.7) is then satisfied. Let $u = sacsa$. The word $sb \in \operatorname{Fact} w$ has to be a factor of u; since sb is not a suffix of u, one has either $sba \in \operatorname{Fact} u$ or $sbx \in \operatorname{Fact} u$, with $x \neq a$. By Lemma 1.2.2, the latter is impossible, because $|sac| = |sbx| = \pi_w$ is a period of u, and $|sac|_a > |sbx|_a$. Thus sba is a factor of u, and by Lemma 1.2.2 it is a conjugate of sac. Therefore $c = b$; by Proposition 2.1.4 and equation (2.4) one derives that sab is a standard word of length π_w. By Lemma 1.2.2, z_w is a conjugate of sab, so that by Theorem 2.2.3 one obtains $w \in St$. ☐

We recall that L_w denotes the minimal integer k for which w has no left special factor of length k. By symmetrical arguments, one can easily prove a result analogous to Proposition 2.2.6, namely, if $\pi_w = L_w + 1$, then $w \in St$.

Examples 2.2.7. The word $w = abbab$ has minimal period $\pi_w = 3$ and $R_w = 2$, therefore it is Sturmian. The word $v = aabba$ is not Sturmian, and indeed $\pi_v = 4 > 3 = R_v + 1 = L_v + 1$. However, for $u = aabab \in St$ one has $\pi_u = 5 > 4 = \max\{R_u, L_u\} + 1$.

Our second characterization of finite Sturmian words is a modification of Proposition 2.2.6:

Theorem 2.2.8. *A finite nonempty word w is Sturmian if and only if*

$$\pi_w = R_{z_w^2} + 1 . \tag{2.8}$$

Proof. Assume (2.8) holds. By Lemma 1.2.3, one has $\pi_{z_w^2} = |z_w| = \pi_w = R_{z_w^2} + 1$, so that from Proposition 2.2.6 it follows $z_w^2 \in St$. As $w \in \operatorname{Fact} z_w^*$, one obtains $w \in St$ by Proposition 2.2.1.

Conversely, let $w \in St$. The result is trivial if $\pi_w = 1$, so assume $|z_w| > 1$. By Theorem 2.2.3, z_w is a conjugate of a standard word. Since all conjugates of z_w are factors of z_w^2, by (2.4) and Proposition 2.1.4 there exists $v \in PER$ such that vab and vba are factors of z_w^2, of length π_w. This means that v is a right special factor of z_w^2 of length $\pi_w - 2$; thus $R_{z_w^2} \geq \pi_{z_w^2} - 1$. By (1.1), one has $\pi_{z_w^2} \geq R_{z_w^2} + 1$, hence $\pi_w = \pi_{z_w^2} = R_{z_w^2} + 1$ as desired. □

We remark that in the case of palindromes, condition (2.8) in the preceding theorem can be replaced by the equation $\pi_w = R_w + 1$. This will be proved in Theorem 2.4.18, as a consequence of Proposition 2.2.6 and of a property of Sturmian palindromes (cf. Proposition 2.4.13).

Proposition 2.2.9. *Let w be a word having minimal period $\pi_w > 1$ and v be its shortest prefix such that $\pi_v = \pi_w$. Let ux ($x \in A$) be the suffix of v of length $\pi_w - 1$. One has $w \in St$ if and only if there exists a letter $y \neq x$ such that uy is a factor of z_w^2.*

Proof. If $uy \in \operatorname{Fact} z_w^2$, then u is a right special factor of z_w^2 of length $\pi_w - 2$, so that $\pi_w \leq R_{z_w^2} + 1$. By (1.1) one has $\pi_w = \pi_{z_w^2} \geq R_{z_w^2} + 1$; thus $\pi_w = R_{z_w^2} + 1$ and by Theorem 2.2.8 it follows $w \in St$.

Conversely, as shown in the proof of Theorem 2.2.3, any word of St has an unbordered factor of maximal length, whose value is the minimal period of the word. Therefore, one can write v as $v = tx$ with $x \in \mathcal{A}$ and $\pi_t < \pi_w$ and t cannot have unbordered factors of length π_w since the maximal length of these factors is π_t. Since $v \in St$, it has an unbordered factor r of maximal length $|r| = \pi_v = \pi_w$. This factor has to be necessarily a suffix of v. Since r is unbordered and $|r| = \pi_w > 1$, from Propositions 2.1.5 and 2.1.6 one has $r = yux$ with $u \in PER$ and $\{x, y\} = \mathcal{A}$. By Lemma 1.2.2, z_w is conjugate of yux and, by Proposition 2.1.4, of xuy. Since $xuy \in \operatorname{Fact} z_w^2$, the result follows. ☐

Examples 2.2.10. Let $w = aababaa \in St$. One has $\pi_w = 5$, $z_w^2 = aababaabab$, and $R_{z_w^2} = 4$, so that $\pi_w = R_{z_w^2} + 1$. The shortest prefix v of w such that $\pi_v = \pi_w$ is $v = aabab$. Its suffix of length $\pi_w - 1$ is $ub = abab$, and $ua = abaa$ is a factor of z_w^2.

Let $r = baabb \notin St$. One has $\pi_r = 4$, $z_r^2 = baabbaab$, and $R_{z_r^2} = 2$, so that $\pi_r > R_{z_r^2} + 1$. In this case, the shortest prefix v such that $\pi_v = \pi_r$ is $v = r$. The suffix ub of v of length 3 is abb, and $aba \notin \operatorname{Fact} z_r^2$.

Enumeration of primitive Sturmian words

As an application of preceding results, we give a formula which counts for any $n > 1$ the finite primitive Sturmian words of length n. We need the following:

Lemma 2.2.11. *The number of words of length $n > 0$ which are conjugate of standard Sturmian words is 2 if $n = 1$ and $n\phi(n)$ for $n > 1$, where ϕ is Euler's totient function.*

Proof. For $n = 1$ the result is trivial since the only two words conjugate of standard words are a and b. Let us suppose $n > 1$. As is well known (see for instance [39, Chap. 2]), the number of standard words of length $n > 1$ is given by $2\phi(n)$. If s is a standard word, by (2.4) we can write $s = vxy$ with $\{x, y\} = \{a, b\}$ and $v \in PER$. By Proposition 2.1.4, $s' = vyx \in Stand$ is a conjugate of s. In the conjugacy class of s there is no other standard word. Indeed, if $t = uxy$ is a conjugate of s, with $u \in PER$, then $|t|_a = |s|_a$ and $|t|_b = |s|_b$, so that t and s have the same "slope"; from this it follows that $u = v$

(see for instance [3, 39]). Hence, in each conjugacy class of a standard word of length $n > 1$ there are exactly two standard words. Thus, the number of these conjugacy classes is $\phi(n)$. Since any standard word is primitive, in any class there are n words. From this the assertion follows. □

Proposition 2.2.12. *For any $n > 1$, the number of primitive finite Sturmian words of length n is given by:*

$$\sum_{i=1}^{n}(n+1-i)\phi(i) - \sum_{\substack{d|n \\ d\neq n}} d\phi(d) \, .$$

Proof. Let w be a non-primitive Sturmian word of length $n > 1$. The word w can be written uniquely as $w = u^k$, with $u \in \pi(\mathcal{A}^*)$ and $k > 1$. Moreover, from Lemma 1.2.3 one has $z_w = u$; by Theorem 2.2.3, u is a conjugate of a standard word. Since $|w| = k|u|$, the integer $|u|$ is a proper divisor of n. Conversely, if u is a conjugate of a standard word, then by Proposition 2.2.1 one has that $u^k \in St$ for any k.

The number of primitive Sturmian words of length n is then obtained by subtracting from $\mathrm{card}(St \cap \mathcal{A}^n)$ the number of words conjugate of a standard word whose length is a proper divisor of n. It is well known (see for instance [39, Chap. 2]) that the number of all finite Sturmian words of length n is given by the following formula:

$$\mathrm{card}(St \cap \mathcal{A}^n) = 1 + \sum_{i=1}^{n}(n+1-i)\phi(i) \, .$$

From Lemma 2.2.11 it follows

$$\mathrm{card}(St \cap \pi(\mathcal{A}^*) \cap \mathcal{A}^n) = 1 + \sum_{i=1}^{n}(n+1-i)\phi(i) - \left(\sum_{\substack{d|n \\ d\neq n}} d\phi(d) + 1 \right)$$

which proves the assertion. □

2.3 New algorithms for the recognition of finite Sturmian words

The problem of finding efficient algorithms for testing whether a finite word is Sturmian is of fundamental importance in discrete geometry for several applications such as pattern recognition, image processing, and computer graphics.

Several linear-time algorithms have been found by different authors, using various concepts and techniques (cf. [6] and references therein). In particular, in [8] a linear algorithm which uses methods of elementary number theory is given, and in [5, 36] linear algorithms based on methods of discrete geometry are provided. In these latter works an essential role is played by a suitable representation of finite Sturmian words by triplets of integers introduced in [37].

In this section, we give two new and simple linear algorithms for the recognition of Sturmian words, which are based on the combinatorial results on words obtained in the previous section.

A first algorithm to recognize whether a word w of length n is Sturmian can be carried out, by Proposition 2.1.7 and Theorem 2.2.3, in the following three steps.

1. Determine the fractional root z_w of w.
2. Compute the Lyndon word ℓ and the anti-Lyndon word ℓ' in the conjugacy class of z_w.
3. Compare ℓ and ℓ' and check whether they have the same proper median factor of maximal length.

Step 1 can be executed in linear time; in fact, there exists an algorithm to determine the minimal period π_w (as well as the minimal periods of all prefixes of w) which runs in linear time [40]. Therefore, also the fractional root z_w can be generated in linear time. As regards to step 2, to determine the Lyndon word in the conjugacy class of z_w requires $O(|z_w|)$ time (see [40]). The same occurs for the anti-Lyndon word. Step 3 trivially requires $O(|z_w|)$ time. In conclusion, the preceding algorithm allows one to recognize whether a word is Sturmian or not in linear time.

A second algorithm can be developed as follows, by using Proposition 2.2.9.

1. Determine the fractional root z_w of w.
2. If $|z_w| = 1$, then $w \in St$; if alph(z_w) contains more than two letters, then $w \notin St$.
3. Find the shortest prefix v of w such that $\pi_v = \pi_w$.

4. Take the suffix ux of v of length $\pi_w - 1$, with $x \in \mathrm{alph}(z_w)$.

5. Verify if uy (with $y \in \mathrm{alph}(z_w)$, $y \neq x$) is a factor of z_w^2.

As we have already discussed, steps 1 and 3 can be executed in linear time. Steps 2 and 4 trivially require $O(n)$ time, and step 5 can be carried out by a linear-time pattern matching procedure (see for instance [40]). In conclusion, the proposed algorithm runs in linear time.

2.4 Sturmian palindromes: structural properties

In the remaining part of this chapter we shall be interested in the set $St \cap PAL$, whose elements will be called *Sturmian palindromes*.

One has that $PER \subseteq St \cap PAL$. However, the previous inclusion is strict since there exist non-central Sturmian palindromes, for instance *abba*.

We have seen that $St = \mathrm{Fact}(PER)$. We shall prove (cf. Corollary 2.4.2) a similar property for Sturmian palindromes.

Theorem 2.4.1. *Every palindromic factor of a standard Sturmian word c_α is a median factor of a palindromic prefix of c_α.*

The result is attributed to A. de Luca [21] by J.-P. Borel and C. Reutenauer, who gave a geometrical proof in [7]. Theorem 2.4.1 can be also obtained as a consequence of a more general result of X. Droubay, J. Justin, and G. Pirillo [27]. We shall report later a direct proof for the sake of completeness.

Corollary 2.4.2. *A word is a Sturmian palindrome if and only if it is a median factor of some central word.*

Proof. Trivially, every median factor of a palindrome is itself a palindrome. Since $St = \mathrm{Fact}(PER)$, it follows that a median factor of an element of PER is a Sturmian palindrome.

Conversely, let u be in $St \cap PAL$. By definition, there exists an infinite (standard) Sturmian word s such that $u \in \mathrm{Fact}\, s$. By Theorem 2.4.1, u is a median factor of a palindromic prefix of s. Since palindromic prefixes of standard Sturmian words are exactly the elements of PER, the result follows. □

Our proof of Theorem 2.4.1, which follows a simple argument suggested by A. Carpi [14], is based on the following results (see [21]):

Proposition 2.4.3. *If $w \in \operatorname{Fact} x$, where x is an infinite Sturmian word, then the reversal $\tilde{w}$ is a factor of x too. Moreover, if x is standard, then w is a right special factor of x if and only if $\tilde{w}$ is a prefix of x.*

Corollary 2.4.4. *A palindromic factor of an infinite standard Sturmian word x is a right special factor of x if and only if it is a palindromic prefix of x.*

Proof of Theorem 2.4.1. By contradiction, let $c_\alpha = \lambda u x$, where u is a palindrome that is not a median factor of any palindromic prefix of c_α, and $\lambda \in \mathcal{A}^*$ has minimal length for such condition. Since u cannot be a prefix of c_α, we have $|\lambda| \geq 1$. Thus we can assume, without loss of generality, $\lambda = \lambda' a$. Now let z be the first letter of x, so that $x = zx'$. Suppose first $z = a$. The palindrome aua is not a median factor of a palindromic prefix of c_α, otherwise so would be u. But $c_\alpha = \lambda' auax'$ with $|\lambda'| < |\lambda|$, and this contradicts the minimality of $|\lambda|$. Therefore $z = b$, and then aub and $bua = \widetilde{aub}$ are factors of c_α, in view of Proposition 2.4.3. This means in particular that u is a right special factor of c_α. Corollary 2.4.4 then implies that u is a prefix of c_α, a contradiction. □

We recall some basic facts (see [26, 21]):

Proposition 2.4.5. *Let w be a word. The following conditions are equivalent:*

1. *$w \in PER$,*
2. *awb and bwa are Sturmian,*
3. *awa, awb, bwa, and bwb are all Sturmian.*

Proposition 2.4.6. *If wa and wb are Sturmian words, then there exists a letter $x \in \mathcal{A}$ such that xwa and xwb are both Sturmian.*

We now prove two easy consequences (see also [21]):

Proposition 2.4.7. *Let $w \in \mathcal{A}^*$ be a palindrome. If wa and wb are Sturmian, then w is central.*

Proof. From the previous proposition, there exists a letter $x \in \mathcal{A}$ such that xwa and xwb are both Sturmian. Without loss of generality, we may suppose $x = a$, so that $awb \in St$. Therefore $\widetilde{awb} = bwa$ is Sturmian too, thus by Proposition 2.4.5, w is central. □

Lemma 2.4.8. *Let w be a Sturmian palindrome. If w is not central, then there exists a unique letter $x \in \mathcal{A}$ such that xwx is Sturmian.*

Proof. If awa and bwb are both Sturmian, then $w \in PER$ by Proposition 2.4.7, a contradiction. Since by Corollary 2.4.2 the word w is a (proper) median factor of some central word, there exists a unique letter $x \in \mathcal{A}$ such that xwx is Sturmian. □

We have seen with Corollary 2.4.2 that a Sturmian palindrome is a median factor of a central word. We will now give some further results concerning the structure of Sturmian palindromes.

Proposition 2.4.9. *A palindrome $w \in A^*$ with minimal period $\pi_w > 1$ can be uniquely represented as*

$$w = w_1xyw_2 = w_2yx\tilde{w}_1$$

with $x, y \in A$, w_2 the longest proper palindromic suffix of w, and $|w_1xy| = \pi_w$. The word w is not central if and only if either $w_1 \notin PAL$ or $w = (w_1xx)^kw_1$ where $k \geq 1$ is the order of w.

Proof. Since $\pi_w > 1$, it follows by Lemma 1.3.3 that w can be uniquely factorized as $w = w_1xyw_2$ where w_2 is the longest proper palindromic suffix of w, $x, y \in A$, and $|w_1xy| = \pi_w$. Since w is a palindrome, we can write

$$w = w_1xyw_2 = w_2yx\tilde{w}_1\ .$$

When $\pi_w > 1$, by Proposition 2.1.1, w is central if and only if $w_1 \in PAL$ and $x \neq y$. Therefore, in the case $w_1 \in PAL$, w is not central if and only if $w = w_1xxw_2 = w_2xxw_1$. The word w has the two periods

$$\pi_w = |w_1xx| \text{ and } q = |w_2xx| \tag{2.9}$$

and length $\pi_w + q - 2$. Thus $w \notin PER$ if and only if $d = \gcd(\pi_w, q) > 1$. Since $|w| \geq \pi_w + q - d$, by Theorem 1.2.1 the word w has the period $d = \pi_w$. This occurs if and only if $q = k\pi_w$ with $k \geq 1$. From (2.9) this condition is equivalent to the statement $w_2xx = (w_1xx)^k$, i.e., $w = (w_1xx)^k w_1$. □

Example 2.4.10. Let $w = aaabaaaaabaaa \in St \cap PAL$, with $\pi_w = 7$. The word w can be factorized as $(aaaba)aa(aaabaaa)$, where $aaabaaa$ is the longest proper palindromic suffix of w, $|aaaba| = \pi_w - 2 = 5$. The prefix $aaaba$ is not a palindrome, thus w is not central.

Let $v = abaababababaaba \in St \cap PAL$. We factorize v as

$$v = (abaabab)ab(abaaba)$$

where $abaaba$ is the longest proper palindromic suffix of v. Also in this case $abaabab$ is not a palindrome, so that $w \notin PER$.

Let $u = abbabbabba \in St \cap PAL$. We factorize u as $(a)bb(abbabba)$, where $abbabba$ is the longest palindromic suffix of u. In this case, the prefix a is a palindrome, and $u = (abb)^3a$. Hence u is not central.

Lemma 2.4.11. *If $w = w_1xyw_2 = w_2yx\tilde{w}_1$, where w_2 is the longest proper palindromic suffix of w and $x, y \in A$, then $w' = ywy$ has the minimal period $\pi_{w'} = \pi_w$.*

Proof. Since w is a factor of w', one has $\pi_{w'} \geq \pi_w$. The word yw_2y is a palindromic proper suffix of $w' = yw_1xyw_2y$, so that by Lemma 1.3.3 the word w' has the period $|yw_1x|$. Hence, $\pi_{w'} \leq |yw_1x| = |w_1xy| = \pi_w$. Thus $\pi_w = \pi_{w'}$. □

The next lemma is essentially a restatement of Lemma 2 in [19]. Note that its first part is an obvious consequence of Lemma 2.4.11.

Lemma 2.4.12. *Let $w = w_1xyw_2 = w_2yxw_1 \in PER$, with $|w_2| > |w_1|$ and $\{x, y\} = \mathcal{A}$. The word $v = ywy$ has minimal period $\pi_v = \pi_w$, whereas $v' = xwx = xw_1xyw_2x$ has minimal period $\pi_{v'} = |w_2| + 2 = |w| - \pi_w + 2$.*

Let $w \in (St \cap PAL) \setminus PER$. We denote by u the (unique) *shortest* median extension of w in PER, and by v the *longest* central median factor of w. Note that also v is unique. For instance, for the Sturmian palindrome $w = baaabaaab$ one has $u = aawaa$ and $v = aaabaaa$.

Theorem 2.4.13. *Let $w \in (St \cap PAL) \setminus PER$. With the preceding notation, one has $\pi_u = \pi_w$. Moreover, either $\pi_w = \pi_v$ or $\pi_w = |v| - \pi_v + 2$.*

Proof. We consider first the case that $\pi_v = 1$, so that $v = x^n$ with $x \in \mathcal{A}$ and $n = |v|$. In such a case w has also the median palindromic factor $v_1 = yx^ny$, where $\{x, y\} = \mathcal{A}$ (recall that v is the longest central median factor of w). Moreover, $n = |v|$ is at least 2, otherwise v_1 would be equal to $yxy \in PER$. One has $\pi_{v_1} = |yx^n| = n + 1 = |v| - \pi_v + 2$. Now we define, for $2 \leq i \leq n$:

$$v_i = xv_{i-1}x = x^{i-1}yx^nyx^{i-1} = (x^{i-1}yx^{n-i+1})(x^{i-1}yx^{i-1}) . \tag{2.10}$$

The word $v_n = x^{n-1}yx^nyx^{n-1}$ is central, whereas by Lemma 2.1.3 we have $v_i \notin PER$. From Lemma 2.4.8 it follows that the words v_i are the *only* Sturmian extensions of v_1 which are median factors of v_n. Since for $i < n$ one has $v_i \notin PER$, one derives that $w = v_k$ for some $1 \leq k < n$, and $u = v_n$. As shown in (2.10), by Lemma 2.4.11 all the v_i's have the same minimal period, for $1 \leq i \leq n$. The result in this case follows: $\pi_w = \pi_u = |v| - \pi_v + 2$.

Now let us assume $\pi_v > 1$. One has $v = w_1xyw_2 = w_2yxw_1$, with $w_1, w_2 \in PAL$ and $x \neq y$. We suppose $|w_1| < |w_2|$, so that $\pi_v = |w_1| + 2$. From the definition of v, it follows that there exists a letter $z \in \mathcal{A}$ such that $v_1 = zvz$ is a median factor of w which is not central. By Lemma 2.4.12, we have $\pi_{v_1} = \pi_v$ if $z = y$, or else $\pi_{v_1} = |v| - \pi_v + 2$ if $z = x$.

Using Lemma 2.4.11, we shall now define a sequence of palindromes with the same minimal period as v_1. Let us first suppose that $z = y$, so that $v_1 = yw_1xyw_2y$. We set $v_2 = xv_1x = (xyw_1)(xyw_2yx)$. Moreover, if $w_1 = p_1p_2 \cdots p_k$ with $p_j \in \mathcal{A}$ for $1 \leq j \leq k$, we set $v_i = p_{k-i+3}v_{i-1}p_{k-i+3}$ for $i \geq 3$, so that

$$\begin{aligned} v_3 &= p_kv_2p_k = (p_kxyp_1 \cdots p_{k-1})(p_kxyw_2yxp_k) , \\ &\vdots \\ v_{k+2} &= p_1v_{k+1}p_1 = p_1 \cdots p_kxyw_1xyw_2yxp_k \cdots p_1 = w_1xyw_1xyw_2yx\tilde{w}_1 . \end{aligned}$$

Since $w_1 = \tilde{w}_1$, the last equation can be written as

$$v_{k+2} = (w_1)xy(w_1xyw_2yxw_1) = (w_1xyw_2yxw_1)yx(w_1)$$

showing, by Proposition 2.4.9, that the word v_{k+2} is central, so that for any $i \leq \pi_v = k + 2$ one has $v_i \in St \cap PAL$.

Let $s \leq k+2$ be the minimal integer such that $v_s \in PER$. Since for $i < s$ one has $v_i \notin PER$, one derives from Lemma 2.4.8 that $u = v_s$ and $w = v_r$ for some integer $r < s$. Hence $\pi_w = \pi_{v_s} = \pi_u$, and in this case $\pi_w = \pi_v$.

The case $z = x$ is similarly dealt with, but interchanging the roles of w_1 and w_2. Thus one assumes $w_2 = q_1 \cdots q_k$, $q_j \in \mathcal{A}$, $1 \leq j \leq k$, and defines v_i as $q_{k-i+3} v_{i-1} q_{k-i+3}$ for $i \geq 3$, starting from $v_2 = y v_1 y = (yxw_2)(yxw_1xy)$ and ending with

$$v_{k+2} = w_2 yxw_2 yxw_1 xyw_2 \in PER \ .$$

Therefore there exist integers r and s such that $1 \leq r < s \leq k+2 = |v| - \pi_v + 2$, $w = v_r$, and $u = v_s$, so that $\pi_w = \pi_u$ and $\pi_w = \pi_{v_1} = |v| - \pi_v + 2$. □

Example 2.4.14. Let $w = baaabaaab \in St \cap PAL$. Following the notations of Theorem 2.4.13, one has $v = aaabaaa$, $v_1 = w$, and $u = v_3 = aabaaabaaabaa$. Thus $\pi_w = \pi_u = \pi_v = 4$.

Let $w = babbbbab$. In this case we have $v = bbbb$, $w = v_2$, and $u = v_4 = bbbabbbbabbb$, so that $\pi_w = \pi_u = 5 = |v| + 1 = |v| - \pi_v + 2$.

For any word $w \in A^*$, we denote by K_w the length of the shortest unrepeated suffix of w. Conventionally, one assumes $K_\varepsilon = 0$. There exist some relations among the parameters R_w, K_w, π_w, and $|w|$; the following lemma synthesizes some results proved in [22, Corollary 5.3, Propositions 4.6 and 4.7] which will be useful in the sequel.

Lemma 2.4.15. *For any $w \in A^*$, one has*

$$|w| \geq R_w + K_w \ .$$

Moreover, the following holds:

- *if $\pi_w = R_w + 1$, then $|w| = R_w + K_w$,*
- *if $|w| = R_w + K_w$, then for any n there exists at most one right special factor of w of length n.*

The following theorem gives a further criterion, different from Proposition 2.4.9, to discriminate whether a palindrome is central or not.

Theorem 2.4.16. *Let $w \in A^*$ be a palindrome with $\pi_w > 1$. Then w is central if and only if its prefix of length $\pi_w - 2$ is a right special factor of w.*

Proof. From Proposition 2.4.9, we can write

$$w = w_1xyw_2 = w_2yx\tilde{w}_1 \tag{2.11}$$

where $x, y \in A$, w_2 is the longest proper palindromic suffix of w, $|w_1| = \pi_w - 2$, and w is central if and only if $w_1 \in PAL$ and $x \neq y$. Therefore we have to prove that w_1 is a right special factor of w if and only if $w_1 = \tilde{w}_1$ and $x \neq y$.

Indeed, assume that these two latter conditions are satisfied. Since $\tilde{w}_1 = w_1$ and w_2 is the longest proper palindromic suffix (and prefix) of w, one has that w_1 is a border of w_2. This implies, from (2.11), that w_1 is a right special factor of w.

Conversely, suppose w_1 is a right special factor of w. Let us first prove that $w_1 \in PAL$. By hypothesis, we have $\pi_w - 2 = |w_1| \leq R_w - 1$, that is $R_w \geq \pi_w - 1$. By Lemma 2.4.15 one has $\pi_w \geq R_w + 1$, so that $\pi_w = R_w + 1$. This implies $|w| = R_w + K_w$, again by Lemma 2.4.15. The suffix $\tilde{w}_1$ of w is repeated, because w_1 is a right special factor of w, which is a palindrome. This leads to

$$\pi_w - 2 = |\tilde{w}_1| \leq K_w - 1$$

and thus to $|w| = R_w + K_w \geq 2\pi_w - 2$. If $|w| = 2\pi_w - 2$, then $|w_1| = |w_2|$ so that one derives $w_1 = w_2 \in PAL$. If $|w| \geq 2\pi_w - 1$, then w has the prefix w_1xyw_1x, so that $yw_1x \in \operatorname{Fact} w$. Since w_1 is a right special factor of w, there exists a letter $z \neq x$ such that $w_1z \in \operatorname{Fact} w$. Moreover, since w_1z is not a prefix, there exists a letter y' such that $y'w_1z \in \operatorname{Fact} w$. One has $y \neq y'$, for otherwise yw_1 would be a right special factor of w of length $\pi_w - 1 = R_w$, which is a contradiction. As w is a palindrome, the words $x\tilde{w}_1y$ and $z\tilde{w}_1y'$ are factors of w too, so that $\tilde{w}_1$ is a right special factor of w. By Lemma 2.4.15, one obtains $w_1 = \tilde{w}_1$. Therefore we get $w_1 \in PAL$ again.

We shall now prove that $x \neq y$. By contradiction, suppose w has the factorization

$$w = (w_1xx)^k w_1\,, \text{ with } k \geq 1$$

as granted by Proposition 2.4.9. Since w_1 is a right special factor of w, one has $w_1z \in \operatorname{Fact} w$ for a suitable letter $z \neq x$. Thus we have either $w_1z = xw_1$ or $w_1z = v_2xxv_1z$, where v_1z is a prefix of w_1 and v_2 is a suffix of w_1. Since $|w_1| = |w_1z| - 1$, we can write $w_1 = v_1z\alpha v_2$, with $\alpha \in A$. The first case is impossible since w_1 is a palindrome and $x \neq z$. In the latter case, one obtains:

$$v_1 z \alpha v_2 = w_1 = \tilde{w}_1 = \tilde{v}_1 x x \tilde{v}_2$$

which is absurd again, because $x \neq z$. □

Example 2.4.17. The word $w = baab$ is a Sturmian palindrome of minimal period $\pi_w = 3$. Its prefix of length 1 is not a right special factor, hence $w \notin PER$. The word $v = abababbababa$ is a Sturmian palindrome having minimal period 7, and its prefix $ababa$ of length 5 is not right special. Therefore $v \notin PER$. On the contrary, the word $u = aabaabaa$ has minimal period 3, and its prefix of length 1 is a right special factor, so that u is central.

In Proposition 2.2.6 we have proved that any finite word w such that $\pi_w = R_w + 1$ is Sturmian. The converse does not hold in general, as shown in Examples 2.2.7 and 2.4.19. However, the result is true for Sturmian palindromes, as the next theorem shows.

Theorem 2.4.18. *A palindrome $w \in A^*$ is Sturmian if and only if $\pi_w = R_w + 1$.*

Proof. By Proposition 2.2.6, the condition is sufficient. Necessity is trivially true if $\pi_w = 1$. By (1.1), one has $\pi_w \geq R_w + 1$. Hence, if $\pi_w > 1$ the condition $\pi_w = R_w + 1$ is equivalent to the existence of a right special factor s of w of length $|s| = \pi_w - 2$.

We prove that every Sturmian palindrome w such that $\pi_w \geq 2$ has such a factor. If w is central, the result follows directly from Theorem 2.4.16. Thus we suppose $w \notin PER$, and as in Theorem 2.4.13 we denote by v the central median factor of w of maximal length.

If $\pi_v = 1$, then there exists a letter $x \in \mathcal{A}$ and an integer $n \geq 1$ such that $v = x^n$. From the maximality condition, one derives that $n > 1$. In this case, by Theorem 2.4.13 one derives $\pi_w = |v| + 1 = n + 1$ and $yx^n y \in \operatorname{Fact} w$, where $\{x, y\} = \mathcal{A}$; therefore x^{n-1} is the desired right special factor of w, of length $n - 1 = \pi_w - 2$.

If $\pi_v > 1$, by using Proposition 2.1.1 we can write v as $v_1 x y v_2 = v_2 y x v_1$, with $\pi_v = |v_1 x y|$. By Theorem 2.4.13, one has either $\pi_w = \pi_v$ or $\pi_w = |v| - \pi_v + 2$. In the first case, the result is a consequence of Theorem 2.4.16. Indeed, the prefix v_1 of the central word v, whose length is $\pi_v - 2 = \pi_w - 2$, is a right special factor of v, and then of w. In the latter case, one derives that the word $xvx = xv_1xyv_2x = xv_2yxv_1x$ is a factor of w, so that v_2 is a right special factor of w, of length $|v| - \pi_v = \pi_w - 2$. ☐

Example 2.4.19. The word $u = ababaa$ is not a palindrome, but $\pi_u = 5 = R_u + 1$, thus it is Sturmian. However, the word $v = aabab \in St$ has $\pi_v = 5 > 3 = R_v + 1$. Let $w = abba \in St \cap PAL$. One has $\pi_w = 3 = R_w + 1$. The palindrome $s = aabbaa$ is not Sturmian. One has $\pi_s = 4 > 3 = R_s + 1$.

We remark that, by symmetrical arguments, one can prove results analogous to Proposition 2.2.6 and Theorem 2.4.18, namely, *if* $\pi_w = L_w + 1$, *then* $w \in St$, and *a palindrome* $w \in A^*$ *is Sturmian if and only if* $\pi_w = L_w + 1$.

2.5 Enumeration of Sturmian palindromes

In this section we shall give an explicit formula for the enumeration function of $St \cap PAL$, that is, the function $g : \mathbb{N} \to \mathbb{N}$ defined for all $n \geq 0$ as

$$g(n) := \text{card}(St \cap PAL \cap \mathcal{A}^n) .$$

For any $n \geq 0$, $g(n)$ gives the number of Sturmian palindromes of length n.

Theorem 2.5.1. *For any $n \geq 0$, the number $g(n)$ of Sturmian palindromes of length n is given by*

$$1 + \sum_{i=0}^{\lceil n/2 \rceil - 1} \phi(n - 2i) , \tag{2.12}$$

where ϕ is Euler's totient function. Equivalently, for any $n \geq 0$

$$g(2n) = 1 + \sum_{i=1}^{n} \phi(2i) \quad \text{and} \quad g(2n+1) = 1 + \sum_{i=0}^{n} \phi(2i+1) .$$

Proof. Given $w \in St \cap PAL$, at least one of its extensions awa and bwb is Sturmian. Indeed, according to Lemma 2.4.8, if $w \notin PER$, then exactly one of these extensions is in St. If $w \in PER$, then from Proposition 2.4.5, both awa and bwb are Sturmian palindromes. Since the number of central words of length n is $\phi(n+2)$ (see [26]), we get:

$$g(n+2) = g(n) + \phi(n+2)$$

and this implies the desired formula, because $g(0) = 1$ and $g(1) = 2$. □

We define a function $f : \mathbb{N} \to \mathbb{N}$ by setting for $n \geq 0$:

$$f(2n) = 1 + \frac{n(n+1)}{2} \quad \text{and} \quad f(2n+1) = 2 + n(n+1) \ .$$

It is easy to verify that $g(n) \leq f(n)$ for all $n \geq 0$. Moreover, for any $n \geq 0$ we set

$$h(n) = \text{card}(PER \cap \mathcal{A}^n) = \phi(n+2) \ .$$

In Table 1 we list the values of the functions g, f, and h for $0 \leq n \leq 17$. As an example, in Table 2 we list all 14 Sturmian palindromes of length 7. The six central words in it are underlined.

Table 2.1: The functions g, f, and h.

n	$g(n)$	$f(n)$	$h(n)$	n	$g(n)$	$f(n)$	$h(n)$
0	1	1	1	9	20	22	10
1	2	2	2	10	14	16	4
2	2	2	2	11	30	32	12
3	4	4	4	12	18	22	6
4	4	4	2	13	42	44	8
5	8	8	6	14	24	29	8
6	6	7	4	15	50	58	16
7	14	14	6	16	32	37	6
8	10	11	4	17	66	74	18

The following proposition relates the numbers of Sturmian palindromes of odd and even length.

Proposition 2.5.2. *For any $n > 0$ one has*

$$g(2n-1) = g(4n) - 2g(2n) + 2 \ .$$

Proof. From Theorem 2.5.1 one has

$$g(4n) = 1 + \sum_{i=1}^{2n} \phi(2i) \ .$$

Table 2.2: Sturmian palindromes of length 7 (central words are underlined).

$\underline{aaaaaaa}$	$\underline{bbbbbbb}$
$\underline{aaabaaa}$	$\underline{bbbabbb}$
$aababaa$	$bbababb$
$abaaaba$	$babbbab$
$\underline{abababa}$	$\underline{bababab}$
$abbabba$	$baabaab$
$abbbbba$	$baaaaab$

As is well known (see for instance [30]), for any $n > 0$ one has $\phi(2n) = \phi(n)$ for odd n and $\phi(2n) = 2\phi(n)$ for even n. Thus we can write

$$\begin{aligned} g(4n) &= 1 + 2 \sum_{\substack{i \text{ even} \\ i \leq 2n}} \phi(i) + \sum_{\substack{i \text{ odd} \\ i < 2n}} \phi(i) \\ &= 1 + 2 \sum_{k=1}^{n} \phi(2k) + \sum_{k=0}^{n-1} \phi(2k+1) \\ &= g(2n-1) + 2(g(2n) - 1) , \end{aligned}$$

which concludes the proof. □

Now we consider the problem of finding lower bounds for the number of Sturmian palindromes of any length. We premise the following simple lemma:

Lemma 2.5.3. *The totient function ϕ has the following lower bounds:*

$$\phi(n) \geq n^{\alpha} \text{ for odd } n \text{ and } \phi(n) \geq 2^{-\alpha} n^{\alpha} \text{ for even } n ,$$

where $\alpha = \log_3 2 = 0.6309\ldots$.

Proof. The case $n = 1$ is trivial. Let us factorize an integer $n > 1$ uniquely as $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$, where for $1 \leq i \leq r$ the p_i are primes, $k_i \geq 1$, and $p_1 < p_2 < \cdots < p_r$. As is well known (cf. [30]), Euler's function ϕ is related to the primes p_i by the following relation:

$$\phi(n) = n \prod_{i=1}^{r} \frac{p_i - 1}{p_i} . \tag{2.13}$$

Let us first suppose that n is *odd*, so that $p_1 \geq 3$. By (2.13) one derives

$$\phi(n) \geq n\left(\frac{2}{3}\right)^r \quad \text{and } n \geq 3^r ,$$

so that $r \leq \log_3 n$ and $\phi(n) \geq n(2/3)^{\log_3 n} = n^{\log_3 2} = n^\alpha$.

Now suppose n is *even*. We can write $n = 2^k m$ with m odd and $k \geq 1$. From the multiplicative property of ϕ, one has $\phi(n) = \phi(2^k)\phi(m) = 2^{k-1}\phi(m)$. From the preceding result, one has

$$\phi(m) = \phi\left(\frac{n}{2^k}\right) \geq \left(\frac{n}{2^k}\right)^\alpha ,$$

so that

$$\phi(n) \geq 2^{k-1}\left(\frac{n}{2^k}\right)^\alpha = 2^{k(1-\alpha)-1} n^\alpha \geq 2^{-\alpha} n^\alpha .$$

□

Proposition 2.5.4. *Let $\beta = \frac{1}{2(1+\alpha)}$. For $n \geq 0$ one has:*

$$g(2n+1) \geq (2-\beta) + \beta(2n+1)^{1+\alpha} \tag{2.14}$$

and

$$g(2n) \geq 1 + \frac{1}{\alpha} n^{1+\alpha} . \tag{2.15}$$

Proof. By Lemma 2.5.3, we can write

$$g(2n+1) = 1 + \sum_{i=0}^{n} \phi(2i+1) \geq 2 + \sum_{i=1}^{n} (2i+1)^\alpha .$$

Approximating the sum with an integral, one has

$$\sum_{i=1}^{n} (2i+1)^\alpha \geq \int_0^n (2x+1)^\alpha \mathrm{d}x = \beta(2n+1)^{1+\alpha} - \beta ,$$

so that (2.14) follows.

By Lemma 2.5.3, we can write

$$g(2n) = 1 + \sum_{i=1}^{n} \phi(2i) \geq 1 + \sum_{i=1}^{n} i^\alpha \geq 1 + \int_0^n x^\alpha \mathrm{d}x = 1 + \frac{1}{\alpha} n^{1+\alpha} ,$$

so that (2.15) follows. □

As a consequence, one derives that

$$g(n) = \Omega(n^{1+\alpha}) .$$

From this result we can prove that the *density* $h(n)/g(n)$ of central words of length n with respect to all Sturmian palindromes of length n, vanishes when n diverges.

Proposition 2.5.5. *The following holds:*

$$\lim_{n\to\infty} \frac{h(n)}{g(n)} = 0 \ .$$

Proof. We recall that $h(n) = \phi(n+2) \leq n+1$ for all $n \geq 0$. Since $g(n) = \Omega(n^{1+\alpha})$, i.e., $g(n) \geq dn^{1+\alpha}$ for all $n \geq 0$ and some positive constant d, it follows that for any $n > 0$ one has

$$\frac{h(n)}{g(n)} \leq \frac{n+1}{dn^{1+\alpha}} \ .$$

As the right hand side of last equation vanishes when n diverges, the assertion follows. □

Now let us recall (cf. [39]) that for any $n \geq 0$ the number $st(n) = \mathrm{card}(St \cap \mathcal{A}^n)$ of all finite Sturmian words of length n is given by the following formula:

$$st(n) = 1 + \sum_{i=1}^{n} (n-i+1)\phi(i) \ .$$

We shall prove that the *density* $g(n)/st(n)$ of Sturmian palindromes of length n with respect to all Sturmian words of length n, tends to 0 when n tends to infinity. The proof is based on the following lemma.

Lemma 2.5.6. $st(n) = \Omega(n^{2+\alpha})$.

Proof. By Lemma 2.5.3 one has $\phi(n) \geq 2^{-\alpha} n^{\alpha}$ for any $n > 0$, so that

$$st(n) \geq 1 + 2^{-\alpha} \sum_{i=1}^{n} (n-i+1) i^{\alpha} = 1 + 2^{-\alpha}(n+1) \sum_{i=1}^{n} i^{\alpha} - 2^{-\alpha} \sum_{i=1}^{n} i^{1+\alpha} \ .$$

Since

$$\sum_{i=1}^{n} i^{\alpha} \geq \int_0^n x^{\alpha} \mathrm{d}x \quad \text{and} \quad \sum_{i=1}^{n} i^{1+\alpha} \leq \int_0^{n+1} x^{1+\alpha} \mathrm{d}x \ ,$$

one obtains

$$st(n) \geq 2^{-\alpha} \frac{n+1}{\alpha+1} n^{1+\alpha} - 2^{-\alpha} \frac{(n+1)^{2+\alpha}}{\alpha+2} = 2^{-\alpha}(n+1) n^{1+\alpha} r(n)$$

where

$$r(n) = \frac{1}{\alpha+1} - \frac{1}{\alpha+2} \left(\frac{n+1}{n} \right)^{2+\alpha} \ .$$

The function r is strictly increasing and it satisfies, for all $n \geq 3$, the inequality

$$0 < r(n) < \frac{1}{(\alpha+1)(\alpha+2)} \ ,$$

so that $st(n) \geq 2^{-\alpha}(n+1)n^{1+\alpha} r(3)$. Therefore there exists a constant $d > 0$ such that for all $n \geq 0$, $st(n) \geq dn^{2+\alpha}$, i.e., $st(n) = \Omega(n^{2+\alpha})$. □

Proposition 2.5.7. *The following holds:*

$$\lim_{n\to\infty} \frac{g(n)}{st(n)} = 0 .$$

Proof. From the definition one has that for any n, $g(n) \leq 1 + \Phi(n)$, where

$$\Phi(n) = \sum_{i=1}^{n} \phi(i) .$$

As is well known (cf. [30]), $\Phi(n) = O(n^2)$, so that by the previous lemma one has

$$\frac{g(n)}{st(n)} \leq \frac{cn^2}{dn^{2+\alpha}}$$

for all $n > 0$ and some constants $c, d > 0$. Since the right hand side in the previous equation vanishes when n diverges, the result follows. □

Let us observe that, by using the well known result (see for instance [30, Theorem 327]),

$$\lim_{n\to\infty} \frac{\phi(n)}{n^{1-\delta}} = \infty \quad \text{for any } \delta > 0 ,$$

one easily derives a sharper asymptotic lower bound for the function g, i.e.,

$$g(n) = \Omega(n^{2-\delta})$$

when n diverges.

Chapter 3

Palindrome closure and episturmian words

As shown in the previous chapter, palindromes play a decisive role in the study of Sturmian words. In this chapter we introduce palindrome closure operators, which are involved in a famous characterization of standard Sturmian words (see Proposition 3.3.1), as well as in a natural generalization to larger alphabets, namely episturmian words.

We begin by defining the operators and prove some preliminary results, all in a slightly generalized setting that will be useful from the point of view of Chapter 4. Then we move to the main goal of this chapter, which is to prove that the classes of finite Sturmian and finite episturmian words are both closed under palindrome closure operators.

3.1 Pseudopalindrome closure operators

Let ϑ be an involutory antimorphism of A^*. We define in A^* two closure operators associating to each word w respectively the shortest ϑ-palindrome having w as a prefix, and the shortest ϑ-palindrome having w as a suffix. We prove that the minimal periods of these two ϑ-palindrome closures of w are equal, and moreover, if w is nonempty, their fractional roots are conjugate. The main result of the section is that the minimal period of the ϑ-palindrome closures of a nonempty word w is equal to the minimal period of w if and only if the fractional root of w is ϑ-symmetric.

Lemma 3.1.1. *For any word $w \in A^*$, there exists a unique shortest ϑ-palindrome having w as a prefix (resp. suffix).*

Proof. Let us observe that certainly there exists a ϑ-palindrome having w as a prefix, namely $w\bar{w}$. Now suppose that $w\lambda_1$ and $w\lambda_2$ ($\lambda_1, \lambda_2 \in A^*$) are two ϑ-palindromes having w as a prefix, both of length $k \leq 2|w|$. One has $0 \leq |\lambda_1| = |\lambda_2| = k - |w| \leq |w|$. Hence, if u is the prefix of w of length $k - |w|$, one derives that $\lambda_1 = \lambda_2 = \bar{u}$.

In a similar way, one proves that there exists a unique shortest ϑ-palindrome having w as a suffix. □

For any word $w \in A^*$, we denote by $w^{\oplus_\vartheta}$ (resp. $w^{\ominus_\vartheta}$) the shortest ϑ-palindrome in A^* having w as a prefix (resp. suffix). We call $w^{\oplus_\vartheta}$ (resp. $w^{\ominus_\vartheta}$) the *right* (resp. *left*) *ϑ-palindrome closure* of w. To simplify the notation, we shall write $w^{\oplus}$ and $w^{\ominus}$ for $w^{\oplus_\vartheta}$ and $w^{\ominus_\vartheta}$ respectively, when no confusion arises.

When ϑ is the reversal operator R, $w^{\oplus}$ and $w^{\ominus}$ are respectively the shortest palindrome having w as a prefix and the shortest palindrome having w as a suffix. As usual, they will be denoted by $w^{(+)}$ and $w^{(-)}$ (cf. [21]).

For any word w, we denote by $P_\vartheta(w)$ (resp. $Q_\vartheta(w)$) the longest ϑ-palindromic prefix (resp. ϑ-palindromic suffix) of w. When there is no ambiguity, we shall simply write P and Q instead of $P_\vartheta(w)$ and $Q_\vartheta(w)$, respectively.

Proposition 3.1.2. *With the above notation, if w is a word and $w = sQ = Pt$, then $w^{\oplus} = sQ\bar{s}$ and $w^{\ominus} = \bar{t}Pt$.*

Proof. Let $w = sQ$ and $w^{\oplus} = sQ\lambda$ with $\lambda \in A^*$. Since $w^{\oplus}$ is a ϑ-palindrome, one has

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

If $|s| \geq |\lambda|$, then $s = \bar{\lambda}\delta$, $\delta \in A^*$. Since $w^{\oplus} = \bar{\lambda}\delta Q\lambda$, it follows that δQ is a ϑ-palindrome. One can write $w = sQ = \bar{\lambda}\delta Q$, so that $\delta = \varepsilon$ by the definition of Q. Hence $\lambda = \bar{s}$. In a similar way, one proves that $w^{\ominus} = \bar{t}Pt$. □

As a consequence of the definition, one derives that

$$(\bar{w})^{\oplus} = w^{\ominus}\ . \tag{3.1}$$

Example 3.1.3. Let $w = abaabb$. One has $P_R(w) = aba$, $Q_R(w) = bb$, $w^{(+)} = abaabbaaba$, and $w^{(-)} = bbaabaabb$. If ϑ is the exchange antimorphism $e = E \circ R$, one has $P_e(w) = ab$, $Q_e(w) = aabb$, $w^{\oplus} = abaabbab$, and $w^{\ominus} = aabbabaabb$.

Proposition 3.1.4. *Let $w \in A^+$ and $w = sQ = Pt$. One has that $z_{w^{\oplus}} \sim z_{w^{\ominus}}$ so that $\pi_{w^{\oplus}} = \pi_{w^{\ominus}}$. If $w \notin PAL_{\vartheta}$, then*

$$z_{w^{\oplus}} = s\bar{t} \text{ and } z_{w^{\ominus}} = \bar{t}s \ .$$

Proof. If w is a ϑ-palindrome, then the result is trivial. Let us then suppose $w = sQ = Pt \notin PAL_{\vartheta}$, so that $s, t \in A^+$. By Proposition 3.1.2, one has:

$$w^{\oplus} = sQ\bar{s} = Pt\bar{s} = s\bar{t}P \ , \tag{3.2}$$

$$w^{\ominus} = \bar{t}Pt = \bar{t}sQ = Q\bar{s}t \ . \tag{3.3}$$

Since P and Q are proper ϑ-palindromic prefixes and suffixes of $w^{\oplus}$ and $w^{\ominus}$ respectively, by Lemma 1.3.3 one has that $p = |s\bar{t}| = |t\bar{s}| > 0$ is a period of $w^{\oplus}$ and of $w^{\ominus}$.

Let us now prove that P is the longest proper ϑ-palindromic prefix (suffix) of $w^{\oplus}$. By contradiction, suppose that T is a ϑ-palindromic prefix of $w^{\oplus}$ of length greater than $|P|$. If $|T| \leq |Pt|$, then T would be a ϑ-palindromic prefix of w longer than P, which is absurd. If $|Pt| < |T| < |w^{\oplus}|$, then one would contradict the fact that $w^{\oplus}$ is the shortest ϑ-palindrome having w as a prefix. Therefore, by Lemma 1.3.3, $p = \pi_{w^{\oplus}}$. Since by (3.2), $s\bar{t}$ is a prefix of $w^{\oplus}$ and $|s\bar{t}| = \pi_{w^{\oplus}}$, one has $z_{w^{\oplus}} = s\bar{t}$. In a similar way, one shows that Q is the longest proper ϑ-palindromic prefix (suffix) of $w^{\ominus}$, so that $p = \pi_{w^{\ominus}}$ and $z_{w^{\ominus}} = \bar{t}s$. From this it follows $z_{w^{\oplus}} \sim z_{w^{\ominus}}$. ☐

Example 3.1.5. Let $w = z_w = abaabb$ (see Example 3.1.3). If $\vartheta = R$, then $s = abaa$, $t = abb$, $z_{w^{(+)}} = abaabba = s\tilde{t}$, and $z_{w^{(-)}} = bbaabaa = \tilde{t}s$, so that $z_{w^{(+)}} \sim z_{w^{(-)}}$.

If $\vartheta = e = E \circ R$, one has $s = ab = \bar{s}$, $t = aabb = \bar{t}$, $z_{w^{\oplus}} = abaabb = s\bar{t}$, and $z_{w^{\ominus}} = aabbab = \bar{t}s$, so that $z_{w^{\oplus}} \sim z_{w^{\ominus}}$.

Theorem 3.1.6. *Let $w \in A^+$. One has $\pi_w = \pi_{w^{\oplus}}$ if and only if z_w is ϑ-symmetric.*

Proof. We first prove the "if" part. Suppose $z_w = \alpha\beta$ with $\alpha, \beta \in PAL_\vartheta$, so that

$$w = (\alpha\beta)^n z'$$

where $n \geq 1$ and $z' \in \mathrm{Pref}(\alpha\beta)$. Moreover, let $w = Pt = sQ$ as before, so that by Proposition 3.1.4 one has $z_{w^\oplus} = s\bar{t}$. Since $\pi_{w^\oplus} \geq \pi_w$, it suffices to show that $|s\bar{t}| \leq \pi_w$. We distinguish two cases, depending on the length of z'.

The first possibility is $z' \in \mathrm{Pref}\,\alpha$. Let $u \in A^*$ be such that $\alpha = z'u = \bar{u}\bar{z}'$. Then the word $\bar{z}'\beta(\alpha\beta)^{n-1}z'$ is a ϑ-palindromic suffix of w, and therefore a suffix of Q. This implies $|s| \leq |u|$, because $w = sQ = \bar{u}\,\bar{z}'\beta(\alpha\beta)^{n-1}z'$. In a similar way, since $(\alpha\beta)^{n-1}\alpha$ is a ϑ-palindromic prefix of w (and then of P), one has $|t| \leq |\beta z'|$ because $w = Pt = (\alpha\beta)^{n-1}\alpha\,\beta z'$. In conclusion, one gets $|s\bar{t}| \leq |u| + |\beta z'| = |\alpha\beta| = \pi_w$, as desired.

The second case occurs when z' is not a prefix of α, so that $z' = \alpha z''$ with $z'' \in \mathrm{Pref}\,\beta$. Let v be the word such that $\beta = z''v = \bar{v}\bar{z}''$. Then $\bar{z}''(\alpha\beta)^{n-1}\alpha z''$ is a ϑ-palindromic suffix of w, so that one derives $|s| \leq |\alpha v|$ following the above arguments. Moreover, since $(\alpha\beta)^n\alpha \in PAL_\vartheta \cap \mathrm{Pref}\,w$, one obtains $|t| \leq |z''|$, which implies $|s\bar{t}| \leq |\alpha v| + |z''| = |\alpha\beta| = \pi_w$.

Let us now prove the "only if" part. If $\pi_w = \pi_{w^\oplus}$, then $z_w = z_{w^\oplus}$. Moreover, since $w^\oplus$ is a ϑ-palindrome beginning with z_w, it has the suffix $\bar{z}_w$. As $|z_w| = |\bar{z}_w| = \pi_w$, one has by Lemma 2.4.13 that $z_w \sim \bar{z}_w$. By Lemma 1.3.2 it follows $z_w \in PAL_\vartheta^2$. □

Corollary 3.1.7. *Let $L_\vartheta = \{w \in A^* \mid z_w \in PAL_\vartheta^2\}$. If $w \in L_\vartheta$, then $w^\oplus, w^\ominus \in L_\vartheta$ and $\pi_{w^\oplus} = \pi_{w^\ominus} = \pi_w$.*

Proof. Let $w \in L_\vartheta$. By Corollary 1.3.5, $z_{w^\oplus}, z_{w^\ominus} \in PAL_\vartheta^2$ so that $w^\oplus, w^\ominus \in L_\vartheta$. By Theorem 3.1.6 and Proposition 3.1.4, $\pi_w = \pi_{w^\oplus} = \pi_{w^\ominus}$. □

Corollary 3.1.8. *Let $w \in A^+$. If z_w is ϑ-symmetric, then $z_w^2 A^* \cap w^\oplus A^* \neq \emptyset$.*

Proof. Since $z_w \in PAL_\vartheta^2$, by the previous theorem one has $\pi_w = \pi_{w^\oplus}$ so that $z_w = z_{w^\oplus}$. One can write

$$w^\oplus = z_{w^\oplus}^k\, z' = z_w^k\, z' \ ,$$

with $k \geq 1$ and $z' \in \mathrm{Pref}\, z_w$. If $k > 1$, one has that z_w^2 is a prefix of $w^\oplus$. If $k = 1$, then $w^\oplus = z_w z'$, so that $w^\oplus \in \mathrm{Pref}\, z_w^2$. □

Let us remark that the converse of the statement of the preceding corollary does not hold in general, as shown in the last example reported below.

Examples 3.1.9. Let $w = abaabb$ (see Examples 3.1.3 and 3.1.5). One has that $w = z_w \notin PAL^2$, so that $\pi_{w^{(+)}} = \pi_{w^{(-)}} = 7 \neq 6 = \pi_w$. For $\vartheta = e$, since $z_w \in PAL_e^2$, one has $\pi_{w^\oplus} = \pi_{w^\ominus} = \pi_w = 6$.

Let $w = aabaa$. One has that $z_w = aab \notin PAL_e^2$. One has $w^\oplus = aabaabbabb$ and $\pi_{w^\oplus} = 10 \neq 3 = \pi_w$.

Let $w = abccbab$. One has $z_w = abccb \in PAL^2$ and $\pi_w = 5$. Thus $w^{(+)} = abccbabccba$ and $w^{(-)} = babccbab$. Moreover, $z_w = z_{w^{(+)}} \sim z_{w^{(-)}} = babcc$. In this case z_w^2 is a prefix of $w^{(+)}$.

Let $w = babaab$. One has $z_w = babaa \in PAL^2$, $w^{(+)} = babaabab$, $w^{(-)} = baababaab$, so that $\pi_{w^{(+)}} = \pi_{w^{(-)}} = \pi_w = 5$. In this case $w^{(+)}$ is a prefix of z_w^2.

Let $w = (aba^2b^2)^2a$, whose fractional root is $z_w = aba^2b^2 \notin PAL^2$. One has that z_w^2 is a prefix of w, and then of $w^{(+)}$.

We conclude this section with three lemmas that will be useful in the sequel.

Lemma 3.1.10. *If a word $u \in A^*$ and a letter $x \in A$ are ϑ-palindromes, then the fractional root of ux is ϑ-symmetric.*

Proof. If ux is unbordered, then $z_{ux} = ux \in PAL_\vartheta^2$, so that z_{ux} is ϑ-symmetric.

If $|z_{ux}| < |ux|$, then z_{ux} is a prefix of u and $\pi_{ux} = |z_{ux}| \leq |u|$ is a period of u. By Corollary 1.3.5, it follows that z_{ux} is ϑ-symmetric. ☐

We remark that the preceding result does not hold in general when the letter x is not a ϑ-palindrome. For instance, let $A = \{a, b\}$ and $\vartheta = e$. The word $w = aabb$ is an e-palindrome, but the word $wb = aabbb = z_{wb}$ is not e-symmetric.

Lemma 3.1.11. *Let $u \in A^*$ and $w = (ux)^\oplus$, where $x \in A$. If p is any prefix of w of length $|p| > |u|$, then $p^\oplus = w$.*

Proof. The word w is a ϑ-palindrome having p as a prefix, so that $|p^\oplus| \leq |w|$. Moreover, p has the prefix ux, so that

$$\left|(ux)^\oplus\right| \leq \left|p^\oplus\right| \leq |w| \ .$$

Therefore, $|p^\oplus| = |w|$. Since w is a ϑ-palindrome of minimal length having p as a prefix, it follows by Lemma 3.1.1 that $w = p^\oplus$. ☐

Lemma 3.1.12. *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

$$(ua)^\oplus = \lambda u \rho \, . \tag{3.4}$$

Clearly $|\lambda| + |u| + |\rho| = |(ua)^\oplus| \leq 2|u| + 2$, which implies $|\lambda| \leq |u| + 2 - |\rho| \leq |u| + 1$. Let us show that actually one has $|\lambda| \leq |u|$. Indeed, if $\lambda = ua$ then from (3.4) one derives $|(ua)^\oplus| = 2|u| + 2$; this implies that $a \notin PAL_\vartheta$ and $(ua)^\oplus = ua\bar{a}u = uau\rho$, so that $u\rho = \bar{a}u$. It follows 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. □

3.2 Iterated pseudopalindrome closures

Let ϑ be a fixed involutory antimorphism of A^* and $^\oplus$ the right ϑ-palindrome closure operator. For any letter $a \in A$ we denote by D_a^ϑ, or simply D_a, the map $D_a : A^* \to PAL_\vartheta$ defined as: for all $v \in A^*$

$$D_a(v) = (va)^\oplus \, .$$

We call the operators D_a, $a \in A$, the *elementary ϑ-palindrome (right) actions* of the letters of A on A^*. One can extend inductively the definition of the operators D_a from the letters of the alphabet A to the words of A^* by setting $D_\varepsilon =$ id and for any $a \in A$ and $w \in A^*$,

$$D_{wa} = D_a \circ D_w \, .$$

Hence, if $w = a_1a_2\cdots a_n$, $a_i \in A$, $i = 1, \ldots, n$, one has:

$$D_w = D_{a_n} \circ D_{a_{n-1}} \circ \cdots \circ D_{a_1} .$$

Thus the action of the operator D_w on the words of A^* is obtained by successive elementary ϑ-palindrome actions with an iterated process which is directed by the word w. Since for any $w, u \in A^*$ and $a \in A$ the word $D_w(u)$ is a prefix of $D_{wa}(u)$, we can define for an infinite word x an operator $D_x : A^* \to A^\omega$ by setting, for any $u \in A^*$:

$$D_x(u) = \lim_{n\to\infty} D_{w_n}(u) , \tag{3.5}$$

where $\{w_n\} = \operatorname{Pref} x \cap A^n$ for $n \geq 0$.

The words u and x are called, respectively, the *seed* and the *directive word* of $D_x(w)$. Until Section 4.3, we shall consider mainly the case when the seed u is equal to the empty word. Therefore, we set for any $w \in A^\infty$

$$\psi_\vartheta(w) = D_w(\varepsilon) . \tag{3.6}$$

From this definition one has $\psi_\vartheta(\varepsilon) = \varepsilon$ and, for any $w \in A^*$ and $a \in A$,

$$\psi_\vartheta(wa) = (\psi_\vartheta(w)a)^\oplus .$$

For any $w, v \in A^*$, one has:

$$\psi_\vartheta(wv) \in \psi_\vartheta(w)A^* \cap A^*\psi_\vartheta(w) .$$

If $x = x_1x_2\cdots x_n\cdots \in A^\omega$, from (3.5) and (3.6) it follows

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

The infinite word $\psi_\vartheta(x)$ will be called the *ϑ-standard (infinite) word* directed by x. The directive word of a ϑ-standard word t will be also denoted by $\Delta(t)$. A ϑ-standard word will be called, without specifying the antimorphism ϑ, a *pseudostandard* word.

Examples 3.2.1. Let $A = \{a, b\}$ and $x = (ab)^\omega$. If $\vartheta = R$, one obtains

$$\psi_R(a) = a , \quad \psi_R(ab) = aba , \quad \psi_R(aba) = abaaba , \ldots$$

In this case, $\psi_R((ab)^\omega) = abaababaabaab\cdots$ is the Fibonacci word f.

If $\vartheta = e$, one has

$$\psi_e(a) = ab\,, \quad \psi_e(ab) = abbaab\,, \quad \psi_e(aba) = abbaababbaab\,, \ldots$$

In this case, $\psi_e\left((ab)^\omega\right) = \mu(f)$, where μ is the Thue-Morse morphism (cf. [38]) defined as $\mu(a) = ab$, $\mu(b) = ba$.

Proposition 3.2.2. *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 $\bar{w} \in \mathrm{Fact}\, s$.*

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 and that $\Delta(s) = x = x_1x_2\cdots x_n\cdots$ with $x_i \in A$, $i > 0$. Let us set $u_1 = \varepsilon$ and for $n > 1$, $u_{n+1} = \psi_\vartheta(x_1\cdots x_n)$, so that $u_{n+1} = (u_nx_n)^\oplus$. If $w = \varepsilon$, then trivially $w^\oplus = \varepsilon \in \mathrm{Pref}\, s$. If $w \neq \varepsilon$, we consider the least n such that $|u_n| < |w| \leq |u_{n+1}|$. By Lemma 3.1.11 one has $w^\oplus = u_{n+1} \in \mathrm{Pref}\, s$. This proves point 1.

By definition of ϑ-standard words, all the words in the set $\psi_\vartheta(\mathrm{Pref}\, x)$ are ϑ-palindromic prefixes of s. Conversely, if w is a ϑ-palindromic prefix of s, then by following the same argument used for point 1, one has that there exists an integer n such that $w = w^\oplus = u_n \in \psi_\vartheta(\mathrm{Pref}\, x)$. 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. □

Proposition 3.2.3. *Let t be a ϑ-standard word. If w is a factor of t, then either $w^\oplus$ or $w^\ominus$ are factors of t.*

Proof. We suppose that $w \notin PAL_\vartheta$, otherwise the result is trivial. By Proposition 3.2.2, $\mathrm{Fact}\, t$ is closed under ϑ, so that also $\bar{w}$ is a factor of t. Let p be a prefix of t such that $p = \lambda u$, where u is either w or $\bar{w}$ and λ is of minimal length. If $\lambda = \varepsilon$, then u is a prefix of t, so that by the preceding proposition, $u^\oplus$ is a factor of t. Suppose $\lambda \neq \varepsilon$ and let Q be the longest ϑ-palindromic suffix of p. One can write $p = \lambda u = sQ$ with $s \in A^*$. We now show that Q is the longest ϑ-palindromic suffix of u. Indeed, otherwise one would have $Q = \mu u = \bar{u}\bar{\mu}$ with $\mu \in A^+$, so that

$$p = \lambda u = s\mu u = s\bar{u}\bar{\mu} \,.$$

Since $\lambda = s\mu$ and $|\mu| > 0$, one has $|s| < |\lambda|$ and this contradicts the minimality of $|\lambda|$. Hence we can write $p = \lambda u = \lambda s'Q$, where $u = s'Q$ and Q is the longest ϑ-palindromic suffix of u. Thus $p^{\oplus} = \lambda s'Q\bar{s'}\bar{\lambda} = \lambda u^{\oplus}\bar{\lambda}$. Since $p^{\oplus}$ is a ϑ-palindromic prefix of t by the preceding proposition, it follows that $u^{\oplus} \in \operatorname{Fact} t$.

We have proved that in all cases, $u^{\oplus}$ is a factor of t. Therefore, if $u = w$, one has $w^{\oplus} \in \operatorname{Fact} t$; if $u = \bar{w}$, by (3.1) one has $w^{\ominus} \in \operatorname{Fact} t$. □

A stronger version of the preceding Proposition will be given with Theorem 4.1.15.

R-standard words were introduced in [27] as *standard episturmian words*. In the next two sections, we consider Sturmian and episturmian words, and give some combinatorial results which are mainly concerned with palindrome closures of their factors. In the next chapter we will consider again general pseudostandard words, as well as different generalizations of standard episturmian words.

3.3 Palindrome closure in Sturmian words

The link between palindrome closure and Sturmian words is expressed by the following well known proposition (see for instance [21]):

Proposition 3.3.1. *An infinite word w over $\mathcal{A} = \{a, b\}$ is standard Sturmian if and only if it is R-standard and its directive word contains infinitely many occurrences of both a and b. Moreover, the directive word of the Sturmian word c_α, with $\alpha = [0; d_0, d_1, \ldots]$, is $\Delta(w) = a^{d_0}b^{d_1}a^{d_2}b^{d_3}\cdots$.*

We now consider factors of Sturmian words in relation with palindrome closure. We need some preliminary results; the first one is a simple lemma on standard words.

Lemma 3.3.2. *If $s \in Stand$, then $s \sim \tilde{s}$.*

Proof. The result is trivial if $s \in \mathcal{A}$. If s is not a letter, then by (2.4), $s \in PAL^2$ and the result follows from Lemma 1.3.2. □

Corollary 3.3.3. *If a word w is a conjugate of a standard word, then $w \sim \tilde{w}$.*

Proof. Let s be a standard word such that $w \sim s$. One has $\tilde{w} \sim \tilde{s}$. Since by the preceding lemma $s \sim \tilde{s}$, the result follows. □

We are now in the position of stating the first main result of this section.

Theorem 3.3.4. *If $w \in St$, then $w^{(+)}, w^{(-)} \in St$ and $\pi_w = \pi_{w^{(+)}} = \pi_{w^{(-)}}$.*

Proof. Let w be a finite Sturmian word. The result is trivial when $w = \varepsilon$; let us then suppose $w \in A^+$. By Theorem 2.2.3, $z_w \sim s$ for some $s \in Stand$. By Corollary 3.3.3, $z_w \sim \tilde{z}_w$ and by Lemma 1.3.2, $z_w \in PAL^2$. By Theorem 3.1.6 and Proposition 3.1.4, one has $\pi_{w^{(+)}} = \pi_{w^{(-)}} = \pi_w$. This implies that $z_{w^{(+)}} = z_w \sim s$, so that by Theorem 2.2.3 it follows $w^{(+)} \in St$. Since by Proposition 3.1.4, $z_{w^{(-)}} \sim z_{w^{(+)}} = z_w \sim s$, by applying again Theorem 2.2.3, one obtains $w^{(-)} \in St$. □

From the previous results and from Corollary 3.1.8 one derives the following:

Corollary 3.3.5. *Let w be a nonempty Sturmian word. One has*

$$z_w^2 \mathcal{A}^* \cap w^{(+)} \mathcal{A}^* \neq \emptyset \, .$$

The following proposition shows that the left and right palindromic closures of a finite Sturmian word are factors of a suitable infinite standard Sturmian word.

Proposition 3.3.6. *Let $w \in St$. There exists an infinite standard Sturmian word s such that $w^{(+)}, w^{(-)} \in \mathrm{Fact}\, s$.*

Proof. If $w \in St$, then by (1.2) one has that for any $k > 1$, the fractional root of z_w^k is z_w. By Theorem 2.2.3, z_w is a conjugate of a standard word; therefore, by using again Theorem 2.2.3, one has $z_w^k \in St$. By Proposition 3.1.4 and Theorem 3.3.4, one has $z_w = z_{w^{(+)}} \sim z_{w^{(-)}}$. This implies $z_{w^{(-)}} \in \mathrm{Fact}\, z_w^2$. Therefore, there exists an integer $m > 1$ such that $w^{(+)}, w^{(-)} \in \mathrm{Fact}\, z_w^m$. Since $z_w^m \in St$, there exists an infinite standard Sturmian word s such that $z_w^m \in \mathrm{Fact}\, s$. Hence, $w^{(+)}, w^{(-)} \in \mathrm{Fact}\, s$. □

Let us observe that in general, if s is an infinite Sturmian word, then $w^{(+)} \in \operatorname{Fact} s$ does not imply $w^{(-)} \in \operatorname{Fact} s$. For instance, in the case of the Fibonacci word f, one has that $(abab)^{(+)} = ababa$ is a factor of f, whereas $(abab)^{(-)} = babab$ cannot be a factor of f. Indeed otherwise, since $aabaa \in \operatorname{Fact} f$, the "balance" condition for Sturmian words (cf. [39, Chap. 2]) would not be satisfied.

Proposition 3.3.7. *Let w be a nonempty word. The following conditions are equivalent:*

1. *w is a prefix of a standard Sturmian word,*
2. *$w^{(+)}$ is central,*
3. *the fractional root z_w is a standard word.*

Proof. *1.* $\Leftrightarrow$ *2.* This is a consequence of Proposition 3.2.2. Indeed, w is a prefix of a standard Sturmian word if and only if $w^{(+)}$ is a prefix of a standard Sturmian word, and this occurs if and only if $w^{(+)}$ is a central word.

2. $\Rightarrow$ *3.* Trivial if $\pi_w = \pi_{w^{(+)}} = 1$. Then assume by Proposition 2.1.1 that $w^{(+)} = w_1xyw_2$, with $\{x, y\} = \{a, b\}$ and $|w_1| < |w_2|$, so that by (2.4) one has $z_{w^{(+)}} = w_1xy \in Stand$. Since $z_w = z_{w^{(+)}}$, the result follows.

3. $\Rightarrow$ *2.* Since $z_w \in PAL^2$, by Theorem 3.1.6 one has $z_{w^{(+)}} = z_w$. The word z_w is standard, so that, as a consequence of the construction via directive sequences, one derives that for any $k \geq 1$, $z_w^k \in \operatorname{Pref} s$, where s is an infinite standard Sturmian word. Now

$$w^{(+)} = z_{w^{(+)}}^k z' = z_w^k z' \in \operatorname{Pref}(z_w^{k+1})$$

for some $z' \in \operatorname{Pref} z_w$. Hence $w^{(+)}$ is a palindromic prefix of a standard word, so that $w^{(+)} \in PER$. ☐

From Theorem 3.3.4 a new characterization of finite Sturmian words can be given. We need the following lemma that summarizes some results proved in [23]:

Lemma 3.3.8. *Let $w \in A^*$. If $\pi_w = R_w + 1$, then w is Sturmian. Conversely, if w is a Sturmian palindrome, then $\pi_w = R_w + 1$.*

Theorem 3.3.9. *A word w is Sturmian if and only if*

$$\pi_{w^{(+)}} = R_{w^{(+)}} + 1 .$$

Proof. By Theorem 3.3.4, w is Sturmian if and only if $w^{(+)}$ is Sturmian. By the previous lemma, the result follows. □

In a perfectly symmetric way, one derives that a word w is Sturmian if and only if $\pi_{w^{(-)}} = R_{w^{(-)}} + 1$.

We observe that if $w \in St$, then from the preceding proposition and Theorem 3.3.4 one derives $\pi_w = R_{w^{(+)}} + 1$. However, this condition does not assure in general that w is Sturmian, as shown by the following example: let $w = abaabb \notin St$; one has $\pi_w = 6$, $\pi_{w^{(+)}} = 7$, and $R_{w^{(+)}} = 5$.

A characterization of finite Sturmian words similar to Theorem 3.3.9 is given by Theorem 2.2.8, of which we will now give a different proof, based on Theorem 3.3.9 and on the following lemma.

Lemma 3.3.10. *If $w \in A^+$ and $\pi_w = R_w + 1$, then $R_w = R_{z_w^2}$.*

Proof. By (1.2) and (1.1) one has that for any $k > 1$

$$|z_w| = \pi_w = \pi_{z_w^k} \geq R_{z_w^k} + 1 .$$

Since $\operatorname{Fact} z_w^2 \subseteq \operatorname{Fact} z_w^k$ and $\pi_w = R_w + 1$, one has that for all $k > 1$

$$|z_w| - 1 \geq R_w \geq R_{z_w^k} \geq R_{z_w^2} .$$

As any factor of z_w^k of length at most $|z_w| - 1$ is also a factor of z_w^2, it follows that $R_{z_w^k} = R_{z_w^2}$ for any $k > 1$. By the definition of fractional root, there exists an integer $h \geq 1$ such that $w \in \operatorname{Fact} z_w^h$, so that $R_w \leq R_{z_w^h} = R_{z_w^2}$. From this $R_w = R_{z_w^2}$. □

Proof of Theorem 2.2.8. Let w be a nonempty Sturmian word. By Theorem 3.3.9, $\pi_{w^{(+)}} = R_{w^{(+)}} + 1$. From Theorem 3.3.4, $\pi_{w^{(+)}} = \pi_w$ and $z_{w^{(+)}} = z_w$. By the preceding lemma, one derives $R_{w^{(+)}} = R_{z_w^2}$ and $\pi_w = R_{z_w^2} + 1$.

Conversely, if $\pi_w = R_{z_w^2} + 1$, then $\pi_{z_w^2} = R_{z_w^2} + 1$, so that by Lemma 3.3.8 we have $z_w^2 \in St$. By Corollary 2.2.5, the result follows. □

3.4 Episturmian words

Episturmian words are a natural generalization of infinite Sturmian words in the case of alphabets with more than two letters. They have been introduced in [27] and their theory has been developed in several papers (see for instance [33, 35]).

As stated above, an infinite word $t \in A^\omega$ is a *standard episturmian* word if it is an R-standard word over A. In the following, we shall denote the operator ψ_R simply by ψ.

The following proposition was proved in [27]:

Proposition 3.4.1. *Let $s \in A^\omega$. The following statements are equivalent:*

1. *s is a standard episturmian word.*
2. *s is closed under reversal, and every left special factor of it is a prefix of s.*

As we shall see in the next chapter, the two above conditions are no longer equivalent when R is substituted by an arbitrary involutory antimorphism ϑ. The first condition gives rise to ϑ-standard words, as we have already seen; the second one leads to what we call *standard ϑ-episturmian* words instead (see Section 4.5). The class of standard ϑ-episturmian words is neither a superclass nor a subclass of that of ϑ-standard words.

An infinite word $s \in A^\omega$ is called *episturmian* if there exists a standard episturmian word $t \in A^\omega$ such that $\mathrm{Fact}\, s = \mathrm{Fact}\, t$. Hence, an infinite word s is episturmian if and only if it has at most one right special factor of each length and $\mathrm{Fact}\, s$ is closed under reversal.

We will use the symbols *Epi* and *SEpi* respectively for the sets of episturmian and standard episturmian words over A. By definition, we have $\mathrm{Fact}(\mathit{Epi}) = \mathrm{Fact}(\mathit{SEpi})$. Of course, any (standard) Sturmian word is a (standard) episturmian word over a two-letter alphabet.

Proposition 3.4.2. *Let w be a nonempty prefix of a standard episturmian word. The fractional root z_w is symmetric, so that $\pi_w = \pi_{w^{(+)}}$.*

Proof. Let u be the longest palindromic prefix of s whose length is less than $|w|$. We can write $w = ux\xi$ with $\xi \in A^*$, so that by Lemma 3.1.11 one obtains

$w^{(+)} = (ux)^{(+)}$. One has

$$\pi_{ux} \leq \pi_w \leq \pi_{w^{(+)}} = \pi_{(ux)^{(+)}} \ . \tag{3.7}$$

Since u is a palindrome, by Lemma 3.1.10 one has that z_{ux} is symmetric, so that by Theorem 3.1.6, $\pi_{ux} = \pi_{(ux)^{(+)}}$. By (3.7), $\pi_w = \pi_{w^{(+)}}$, that is equivalent to $z_w \in PAL^2$ by Theorem 3.1.6. □

Example 3.4.3. Let t be the standard episturmian word, called *Tribonacci word*, $t = \psi\left((abc)^\omega\right)$,

$$t = abacabaabacababac \cdots \ .$$

The fractional roots of the nonempty prefixes of t are the symmetric words

$$a,\ ab,\ abac,\ abacaba,\ abacabaabacab, \ldots$$

Let us observe that in the case of a ϑ-standard word s, the fractional root of a prefix of s is not in general ϑ-symmetric. For instance, consider in the case of $A = \{a, b\}$ and $\vartheta = e$, any e-standard word s having a directive word beginning with a^2b. The word s has the prefix $ababbaabab = (ababb)^{\oplus}$. Let $w = ababb$. One has $z_w = w \notin PAL_e^2$. In fact, one has $\pi_w = 5$ and $\pi_{w^{\oplus}} = 6$.

The finite factors of (standard) episturmian words are also called *finite episturmian words*. Differently from the Sturmian case, the fractional root of a finite episturmian word can be non-symmetric, as shown in the following example.

Example 3.4.4. The word $v = aabaacaabaacaaba$ is a prefix of a standard episturmian word. The word $w = z_w = baac$ is a non-symmetric factor of v. However, $w^{(+)} = baacaab$ and $w^{(-)} = caabaac$ are factors of v.

Let us observe that Corollary 2.2.5 cannot be extended to the case of episturmian finite words, since there exist finite episturmian words w such that z_w^2 is not a finite episturmian word, as shown by the following:

Example 3.4.5. The word $w = bac = z_w$ is a finite episturmian word. However, $z_w^2 = (bac)^2$ is not factor of any episturmian word. Indeed, as shown in [27], the number of all palindromic factors in a finite episturmian word u has to be equal to $|u| + 1$. The number of palindromic factors of $(bac)^2$ is 4, and $|z_w^2| + 1 = 7$.

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. By Proposition 3.3.1, the families of standard Sturmian words and of binary AR-words coincide.

Example 3.4.6. Let $A = \{a, b\}$ and $x = (ab)^\omega$. One has that

$$f = \psi(x) = abaababaabaababa\cdots$$

is the Fibonacci word, a standard Sturmian word. On an alphabet with three letters $A = \{a, b, c\}$, if we take $x = (abc)^\omega$ as a directive word, then

$$\tau = \psi(x) = abacabaabacababacabaabac\cdots$$

is a standard Arnoux-Rauzy word, often called *Tribonacci* word. The word $s = cabaabacababacabaab\cdots$ such that $abas = \tau$ is an example of an episturmian word which is not standard, as a is a left special factor of s but not a prefix of it.

The periodic word $s = (abac)^\omega$ is standard episturmian, but not Arnoux-Rauzy. Its directive word is $\Delta(s) = abc^\omega$.

The following proposition can be easily proved using well-known results on episturmian words (see [27]).

Proposition 3.4.7. *Let s be a standard episturmian word. Any bispecial factor of s is a palindromic prefix of s. If s is not periodic, the converse holds too.*

Proposition 3.4.8. $\mathrm{Fact}(Epi) = \mathrm{Fact}(AR)$.

Proof. Let $u \in \mathrm{Fact}(Epi) = \mathrm{Fact}(SEpi)$. Hence there exists $s \in SEpi$ such that $u \in \mathrm{Fact}\, s$. Now let be $s = \psi(\Delta)$ where $\Delta = t_1 t_2 \cdots t_n \cdots$, with $t_i \in A$ for $i \geq 1$. 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, so that $u \in \mathrm{Fact}\, s'$. Therefore, $\mathrm{Fact}(Epi) \subseteq \mathrm{Fact}(AR)$. 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 [27]) which will be useful in the sequel.

Proposition 3.4.9 (cf. [27]). *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 of s occurs in every factor of s having length 2.*

Clearly, if p is a prefix of a standard episturmian word, then the palindromic suffix of p which has a unique occurrence in p is the longest palindromic suffix of p. We want to show that if $w \in \mathrm{Fact}(Epi)$, then also its right and left palindrome closures belong to $\mathrm{Fact}(Epi)$; since episturmian words are closed under reversal, and $w^{(-)} = \tilde{w}^{(+)}$, it suffices to prove only the right palindrome closure case. We have the following

Proposition 3.4.10. *Let u be a non-palindromic 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 a finite episturmian word.*

Before proving the proposition we need some lemmas. The first lemma was proved in [2, Theorem 1.1]. We report here a different and simpler proof.

Lemma 3.4.11. *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. We may 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 that $\tilde{u}$ 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 3.4.9. 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 3.4.12. *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, for any $x \in A$, $sx \in \operatorname{Fact} w$.*

Proof. Since w is not periodic, by Proposition 3.4.7 the bispecial factors B_i, $i > 0$, are its palindromic prefixes. Moreover, if $t = t_1 t_2 \cdots t_n \cdots \in A^\omega$ is the directive word of w, then $B_{i+1} = (B_i t_i)^{(+)}$ for any $i > 0$. Since s is a 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 $|s| \leq |B_m|$. As $w \in AR$, any letter $x \in A$ occurs infinitely often in t; hence there exists $k \geq m$ such that $x = t_k$, so that $B_k x$ is a factor of w. Since B_m is a suffix of B_k, it follows $sx \in \operatorname{Fact} w$. □

Lemma 3.4.13. *Let w and w' be Arnoux-Rauzy words on the alphabet A. If w and w' have the same right special factor of length n, then they share the same factors up to length $n+1$.*

Proof. Trivial if $n = 0$. By induction, suppose we have proved the assertion for the integer $n - 1 \geq 0$. Let Q be the common right special factor of w and w' of length n. If we write $Q = aQ'$, with $a \in A$, then Q' is the only right special factor of length $n-1$ of both w and w'. Hence w and w' have the same factors up to length n.

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 Lemma 3.4.12, 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' as a consequence of Lemma 3.4.12.

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 $v''x \in \operatorname{Fact} w'$. Hence $v = cv''b$ is the only right extension in w' of the factor $cv'' \neq Q$. □

We can now proceed to prove Proposition 3.4.10.

Proof of Proposition 3.4.10. We first observe that u contains a single occurrence of Q. Indeed, if u contained other occurrences of Q, then by Lemma 3.4.11 the suffix of u beginning with the penultimate occurrence would be a palindromic suffix of u strictly longer than Q, contradicting the hypothesis of maximality of the length of Q.

By Proposition 3.4.8 there exists an Arnoux-Rauzy word w such that $u \in \operatorname{Fact} w$. We can assume that $ua \notin \operatorname{Fact} w$ (otherwise ua is in $\operatorname{Fact}(AR)$ as required); so there exist $b \in A$ such that $b \neq a$ and $ub \in \operatorname{Fact} w$. Thus $aQb \in \operatorname{Fact} w$; since Q is a palindrome and $w \in AR$, also $bQa \in \operatorname{Fact} w$ 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 $\operatorname{Fact} w$, contradicting our assumption.

Set $\Delta(w) = t_1 t_2 \cdots$. Let $B_1 = \varepsilon, B_2, \ldots$ be the sequence of all bispecial factors of w, ordered by increasing length, i.e., $|B_i| < |B_{i+1}|$ for all $i > 0$. By Proposition 3.4.7, they are the palindromic prefixes of w as w is not periodic. Moreover, for each $i > 0$ we have $B_{i+1} = (B_i t_i)^{(+)}$, so that $B_i t_i$ is left special and $t_i B_i$ is right special.

Since Q is a bispecial factor of w, one has $Q = B_m$ for some $m > 1$. Let $|Q| = n - 1$ for $n \geq 2$. We then have that $t_m Q$ is right special in w and, from Lemma 3.4.12, $t_m Qx \in \operatorname{Fact} w$ for all $x \in A$. It is clear that $t_m \neq a$ since $aQa \notin \operatorname{Fact} w$ and $t_m Qa \in \operatorname{Fact} w$, then we have that aQb and $t_m Qb$ 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 $\Delta(w') = t'_1 t'_2 \cdots$ 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 3.4.13, we obtain that w and w' have the same factors of length k for each $k \leq n$. However, they 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, so that by Lemma 3.4.12, for all $x \in A$ we have that $aQx \in \operatorname{Fact} w'$. Therefore $aQa \in \operatorname{Fact} w' \setminus \operatorname{Fact} w$.

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 $paQ \in \operatorname{Fact} w'$. 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. If we assume that $S \neq paQ$, then there exist $x, y \in A$ such that $x \neq y$, $xS \in \operatorname{Suff}(saQ)$ and $yS \in \operatorname{Suff}(paQ)$; 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 $\operatorname{Fact}(AR)$ as required. □

From the preceding proposition we get the announced result:

Theorem 3.4.14. *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 3.4.10, $wa_n = a_1 \cdots a_n Q a_n$ 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 3.4.10 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 3.4.15. *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 3.4.14, $(au)^{(+)} = au^{(+)}a$ is an episturmian word and therefore so is $au^{(+)}$. Let us then suppose that au is a palindrome.

By Theorem 3.4.14 one has $u^{(+)} \in \text{Fact}\, s$ for a suitable $s \in AR$. Since s is recurrent there exist letters $x, y \in A$ such that

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

If $x \neq y$, then, since s is closed under reversal, one has also $yu^{(+)}x \in \text{Fact}\, s$. Hence $u^{(+)}$ is bispecial, so that it follows $au^{(+)} \in \text{Fact}\, s$. 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,

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

Since $(u'a)^{(+)}$ begins with $u'a$ and ends with au', one has that $xu'a$ and $au'x$ are factors of s, so that u' is bispecial and then a palindromic prefix of s by Proposition 3.4.7.

Let $\Delta(s) = t_1t_2\cdots t_n\cdots$ be the directive word of s. There exists an integer k such that $u' = \psi(t_1t_2\cdots t_k)$. We consider any AR word s' whose directive word $\Delta(s')$ has the prefix $t_1t_2\cdots t_ka$. Thus $u'a = u$ is a prefix of s'. This implies, by Propositions 3.2.2 and 3.4.7, that $u^{(+)}$ is a bispecial prefix of s'. From this one derives $au^{(+)} \in \text{Fact}\, s'$. ☐

Chapter 4

Extensions via involutory antimorphisms

In this chapter we shall analyse several different generalizations of episturmian words, all based on the replacement of the reversal operator by a generic involutory antimorphism. The first generalization was already introduced in the preceding chapter.

4.1 Pseudostandard words

Let ϑ be an involutory antimorphism of A^*. We recall (cf. Section 3.2) that the map ψ_ϑ defined by (3.6) satisfies, for any $x \in A^\omega$,

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

where $\{w_n\} = A^n \cap \operatorname{Pref} x$ for any $n \geq 0$. The word $\psi_\vartheta(x)$ is ϑ-standard, $\psi_\vartheta(A^\omega)$ is the set of all ϑ-standard infinite words, and $\psi_\vartheta(A^*)$ is the set of their ϑ-palindromic prefixes.

As we have seen in Section 1.3, the reversal operator R is the basic involutory antimorphism of A^*, because any other is obtained by composing R with an involutory permutation. Therefore, it is natural to ask whether any pseudostandard word can be obtained, by a suitable morphism, from a standard episturmian word. As we shall see later, the answer to this problem is positive (cf. Theorem 4.1.2). To this end, we introduce the endomorphism μ_ϑ of A^* by setting $\mu_\vartheta(a) = a^\oplus$ for each $a \in A$. Thus for every letter a one has:

$$\mu_\vartheta(a) = \begin{cases} a & \text{if } a = \bar{a} \\ a\bar{a} & \text{if } a \neq \bar{a} \end{cases} .$$

We observe that μ_ϑ is injective, since $\mu_\vartheta(A)$ is a prefix code.

Example 4.1.1. If $\vartheta = R$, then $\mu_R = \mathrm{id}$. If $\vartheta = e$ is the "exchange" antimorphism of $\{a, b\}^*$, then $\mu_e(a) = ab$ and $\mu_e(b) = ba$, i.e., μ_e is the Thue-Morse morphism.

The main result of this section is the following:

Theorem 4.1.2. *For any $w \in A^\infty$, one has*

$$\psi_\vartheta(w) = \mu_\vartheta(\psi(w)) . \tag{4.1}$$

By this theorem, any ϑ-standard word is a morphic image (by μ_ϑ) of the standard episturmian word having the same directive word. Moreover, the set of palindromic prefixes of ϑ-standard words over A is a morphic image of the palindromic prefixes of standard episturmian words. In particular, the Thue-Morse morphism sends standard Sturmian words to words constructible via iterated e-palindromic closure: $\mu_\vartheta(\psi(x)) = \psi_e(x)$. For instance, $\psi_e\left((ab)^\omega\right) = \mu_\vartheta(f)$ where f is the Fibonacci word.

To prove Theorem 4.1.2, we need some lemmas and propositions concerning the morphism μ_ϑ and the antimorphism ϑ.

It is easy to verify that for any $a \in A$, one has

$$\overline{a^\oplus} = a^\oplus \quad \text{and} \quad (\bar{a})^\oplus = \widetilde{a^\oplus} . \tag{4.2}$$

Lemma 4.1.3. *For all $w \in A^*$, $\mu_\vartheta(\tilde{w}) = \overline{\mu_\vartheta(w)}$.*

Proof. Let $w = a_1 \cdots a_n$, with $a_i \in A$, $i = 1, \ldots, n$. By (4.2),

$$\mu_\vartheta(\tilde{w}) = a_n^\oplus \cdots a_1^\oplus = \overline{a_n^\oplus} \cdots \overline{a_1^\oplus} = \overline{a_1^\oplus \cdots a_n^\oplus} = \overline{\mu_\vartheta(w)} .$$ □

Corollary 4.1.4. *The morphism μ_ϑ sends palindromes into ϑ-palindromes and vice-versa. Formally, for any $w \in A^*$,*

$$w \in PAL \iff \mu_\vartheta(w) \in PAL_\vartheta , \tag{4.3}$$

$$w \in PAL_\vartheta \iff \mu_\vartheta(w) \in PAL . \tag{4.4}$$

Proof. From the previous lemma, since μ_ϑ is injective one immediately obtains

$$w = \tilde{w} \Longleftrightarrow \mu_\vartheta(w) = \mu_\vartheta(\tilde{w}) = \overline{\mu_\vartheta(w)}\ ,$$

proving (4.3).

Let $w = a_1 \cdots a_n$, $a_i \in A$, $i = 1, \ldots, n$. By (4.2), $w = \bar{w}$ is equivalent to:

$$\mu_\vartheta(w) = \mu_\vartheta(\bar{w}) = \mu_\vartheta(\bar{a}_n \cdots \bar{a}_1) = (\bar{a}_n)^\oplus \cdots (\bar{a}_1)^\oplus = \widetilde{a_n^\oplus} \cdots \widetilde{a_1^\oplus} = \widetilde{\mu_\vartheta(w)}$$

as desired. □

Let $|| \, ||_\vartheta : A^* \to \mathbb{Z}_2$ be the morphism of A^* in the additive group $\mathbb{Z}_2$ of the integers mod 2, defined by the rule: for all $a \in A$,

$$||a||_\vartheta = \begin{cases} 0 & \text{if } a = \bar{a} \\ 1 & \text{if } a \neq \bar{a} \end{cases} .$$

In other terms, for any $w \in A^*$, $||w||_\vartheta$ counts, modulo 2, the occurrences of letters in w which are not ϑ-palindromes. Note that one has obviously $||w||_\vartheta = ||\bar{w}||_\vartheta$ for any word w. Let us observe that if $\vartheta = R$, then $||w|| = 0$ for all $w \in A^*$; if $\vartheta = e$, then $||w||_e = (|w| \mod 2)$ for all $w \in \{a, b\}^*$. In the following, we shall denote $|| \, ||_\vartheta$ simply by $|| \, ||$ when there is no ambiguity.

Lemma 4.1.5. *If $w \in \mu_\vartheta(A^*) \cup PAL_\vartheta$, then $||w|| = 0$.*

Proof. It is clear from the definition that $||\mu_\vartheta(u)|| = 0$ for all $u \in A^*$. Indeed, any letter which is not a ϑ-palindrome is sent by μ_ϑ in two non-ϑ-palindromic letters. Let $w = a_1 a_2 \cdots a_n \in PAL_\vartheta$, $a_i \in A$, $i = 1, \ldots, n$. Since $a_i = \bar{a}_{n+1-i}$ for $1 \leq i \leq n$, it follows that:

- if $n = |w|$ is even, then $w = v\bar{v}$,
- if n is odd, then $w = vc\bar{v}$,

where $v = a_1 \cdots a_{\lfloor n/2 \rfloor}$ and $c \in A \cap PAL_\vartheta$. In both cases,

$$||w|| = ||v|| + ||\bar{v}|| = 2||v|| = 0\ .$$ □

Proposition 4.1.6. *Let $w \in A^*$. Then $PAL_\vartheta \cap \operatorname{Suff} \mu_\vartheta(w) = \mu_\vartheta(PAL \cap \operatorname{Suff} w)$.*

Proof. The "$\supseteq$" inclusion is a consequence of (4.3). Now we prove the inverse inclusion. Let s be a suffix of $\mu_\vartheta(w)$ which is not in $\mu_\vartheta(\operatorname{Suff} w)$. If $w = a_1 \cdots a_n$, with $a_i \in A$ for $1 \le i \le n$, then $\mu_\vartheta(w) = a_1^\oplus \cdots a_n^\oplus$, so that s has to be of the form $s = \bar{a}_i \mu_\vartheta(u)$ for some $i \in \{1, \ldots, n\}$ (such that $a_i \neq \bar{a}_i$) and $u \in \operatorname{Suff} w$. Hence, by Lemma 4.1.5, $\|s\| = \|a_i\| + \|\mu_\vartheta(u)\| = 1$, and therefore $s \notin PAL_\vartheta$, again by Lemma 4.1.5. □

Theorem 4.1.7. *For all $w \in A^*$, one has*

$$(\mu_\vartheta(w))^\oplus = \mu\left(w^{(+)}\right) , \tag{4.5}$$
$$(\mu_\vartheta(w))^\ominus = \mu\left(w^{(-)}\right) . \tag{4.6}$$

Proof. Let $w = sQ$ with $Q = Q_R(w)$. Then by Proposition 3.1.2, $w^{(+)} = sQ\tilde{s}$, so that by Lemma 4.1.3,

$$\mu\left(w^{(+)}\right) = \mu_\vartheta(s)\mu(Q)\mu_\vartheta(\tilde{s}) = \mu(s)\mu_\vartheta(Q)\overline{\mu_\vartheta(s)} .$$

By Corollary 4.1.4, $\mu_\vartheta(Q)$ is a ϑ-palindromic suffix of $\mu_\vartheta(w)$. Let us prove that it is the longest one. Indeed, suppose by contradiction that λ is a ϑ-palindromic suffix of $\mu_\vartheta(w)$, with $|\lambda| > |\mu_\vartheta(Q)|$. By Proposition 4.1.6, $\lambda = \mu_\vartheta(v)$ for some $v \in PAL \cap \operatorname{Suff} w$. This is a contradiction, because $|v| > |Q|$. Thus (4.5) is proved.

By (3.1), $w^{(-)} = \tilde{w}^{(+)}$ so that by (4.5) one has

$$\mu\left(w^{(-)}\right) = \mu\left(\tilde{w}^{(+)}\right) = (\mu_\vartheta(\tilde{w}))^\oplus .$$

By Lemma 4.1.3, $\mu_\vartheta(\tilde{w}) = \overline{\mu_\vartheta(w)}$. Therefore, since by (3.1)

$$\left(\overline{\mu_\vartheta(w)}\right)^\oplus = (\mu_\vartheta(w))^\ominus ,$$

equation (4.6) is proved. □

Corollary 4.1.8. *Let $w \in A^*$ and $a \in A$. The following holds:*

$$(\mu_\vartheta(w)a)^\oplus = \mu\left((wa)^{(+)}\right) .$$

Proof. From the preceding theorem, one has $(\mu_\vartheta(wa))^\oplus = \mu\left((wa)^{(+)}\right)$. Therefore, it suffices to prove that

$$(\mu_\vartheta(w)a)^\oplus = (\mu_\vartheta(wa))^\oplus = (\mu(w)\mu_\vartheta(a))^\oplus . \tag{4.7}$$

If $a \in PAL_\vartheta$, then $a = \mu_\vartheta(a)$ and (4.7) follows. Then assume $a \notin PAL_\vartheta$, so that (4.7) can be rewritten as

$$(\mu_\vartheta(w)a)^\oplus = (\mu_\vartheta(w)a\bar{a})^\oplus \ .$$

In view of Lemma 3.1.11, it suffices to show that $\mu_\vartheta(w)a\bar{a}$ is a prefix of $(\mu_\vartheta(w)a)^\oplus$.

Suppose first that $\bar{a}PAL_\vartheta \cap \operatorname{Suff} \mu_\vartheta(w) = \emptyset$. Then $Q_\vartheta(\mu_\vartheta(w)a) = \varepsilon$, so that by Proposition 3.1.2,

$$(\mu_\vartheta(w)a)^\oplus = \mu_\vartheta(w)a\bar{a}\overline{\mu_\vartheta(w)}$$

and we are done.

If $\bar{a}PAL_\vartheta \cap \operatorname{Suff} \mu_\vartheta(w)$ is nonempty, then let $\bar{a}\lambda$ be its longest element. It is easy to see that $\bar{a}\lambda a$ is the longest ϑ-palindromic suffix of $\mu_\vartheta(w)a$. Moreover, by Proposition 4.1.6 there exists $v \in PAL \cap \operatorname{Suff} w$ such that $\lambda = \mu_\vartheta(v)$. If $w = uv$, since $\bar{a}\mu_\vartheta(v)$ is a suffix of $\mu_\vartheta(w) = \mu(u)\mu_\vartheta(v)$, one derives that $u = u'a$ for some word u'. Hence

$$(\mu_\vartheta(w)a)^\oplus = \mu_\vartheta(u')a\bar{a}\mu_\vartheta(v)a\bar{a}\overline{\mu_\vartheta(u')} = \mu_\vartheta(w)a\bar{a}\overline{\mu_\vartheta(u')} \ ,$$

which concludes the proof. □

We are in the position of proving the main theorem.

Proof of Theorem 4.1.2. Equation (4.1) is trivially satisfied for $w = \varepsilon$. By induction, let us assume (4.1) holds for some $w \in A^*$, and prove it for wa with $a \in A$. Indeed,

$$\psi_\vartheta(wa) = (\psi_\vartheta(w)a)^\oplus = (\mu_\vartheta(\psi(w))a)^\oplus = \mu\left((\psi(w)a)^{(+)}\right) = \mu_\vartheta(\psi(wa)) \ ,$$

where the third equality is a consequence of Corollary 4.1.8.

The case $w \in A^\omega$ is easily dealt with. □

For any letter $a \in A$, we define the morphism $\mu_a : A^* \to A^*$ by $\mu_a(a) = a$ and $\mu_a(b) = ab$, for any $b \neq a$. Moreover, we set $\mu_\varepsilon = \mathrm{id}$ and, for any $w = a_1a_2\cdots a_n \in A^+$,

$$\mu_w = \mu_{a_1} \circ \mu_{a_2} \circ \cdots \circ \mu_{a_n} \ .$$

As a consequence of Theorem 4.1.2 and of a result of Justin [34], we derive the following proposition which allows one to compute $(\psi_\vartheta(w)a)^\oplus$ for any $w \in A^*$ and $a \in A$, starting from its prefix (suffix) $\psi_\vartheta(w)$, by using the morphisms μ_ϑ and μ_w.

Proposition 4.1.9. *For any $w \in A^*$ and $a \in A$,*

$$\psi_\vartheta(wa) = (\mu \circ \mu_w)(a)\, \psi_\vartheta(w)\ .$$

Proof. We use the result of Justin [34] stating that for any $v, w \in A^*$,

$$\psi(wv) = \mu_w(\psi(v))\psi(w)\ .$$

Therefore, for $v = a \in A$ one gets $\psi(wa) = \mu_w(a)\psi(w)$. By Theorem 4.1.2,

$$\begin{aligned} \psi_\vartheta(wa) &= \mu_\vartheta(\psi(wa)) \\ &= \mu_\vartheta(\mu_w(a))\, \mu_\vartheta(\psi(w)) \\ &= (\mu \circ \mu_w)(a)\psi_\vartheta(w) \end{aligned}$$

as desired. □

Example 4.1.10. Let $A = \{a, b\}$, $\vartheta = e$, and $w = aba$. One has $\psi_\vartheta(aba) = abbaababbaab$ and $\mu_w(a) = aba$. Hence, $\mu(\mu_w(a)) = abbaab$ and

$$\psi_\vartheta(abaa) = (abbaab)(abbaababbaab)\ .$$

We have seen that ϑ-standard words are morphic images (under μ_ϑ) of standard episturmian words. This allows to extend many properties of standard episturmian words to general ϑ-standard words.

Proposition 4.1.11. *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 B is a suffix of any ϑ-palindromic prefix of s whose length is at least $|B|$, and there exist two distinct letters (say a and b) which occur infinitely often in $\Delta(s)$, by Proposition 3.2.2 one derives $Ba, Bb \in \operatorname{Fact} s$. Therefore, as $\bar{p} \in \operatorname{Suff} B$, we have $\bar{p}a, \bar{p}b \in \operatorname{Fact} s$, i.e., $\bar{p}$ is right special. Since by Proposition 3.2.2 s is closed under ϑ, one has $\bar{a}p, \bar{b}p \in \operatorname{Fact} s$; 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. For instance, consider the e-standard word s in Example 3.2.1. 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 4.1.12. *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 4.1.13. *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 $s \in SEpi$. Then:*

1. *If $bu, vc \in \operatorname{Fact} s$ and $|f| > 1$, then $f \neq g$.*

2. *If $u \in \operatorname{Fact} s$ and $|f| > 3$, then $bu \in \operatorname{Fact} s$.*

Proof. (1). Since $|f| > 1$, one has $u \neq \varepsilon$. By contradiction, if $f = g$, one has also $v \neq \varepsilon$, so that, from the definition of μ_ϑ, $\bar{b}b$ 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 3.4.9, $bu = b^{k+1}$ 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 $bu \notin \operatorname{Fact} s$. By the preceding lemma and by Theorem 4.1.2, one derives $f = \mu_\vartheta(v')c'$ for some suitable $v' \in A^*$ and $c' \in A'$ such that $v'c' \in \operatorname{Fact} s$. As done before, one then obtains $f = (\bar{b}b)^k\bar{b}$ so that $b^k, \bar{b}^k \in \operatorname{Fact} s$, which is absurd by Proposition 3.4.9, as $k \geq 2$. □

Theorem 4.1.14. *Let w be a left special factor of a ϑ-standard word $t = \mu_\vartheta(s)$, with $s \in SEpi$. If $|w| \geq 3$, then w is a prefix of t.*

Proof. By Theorem 4.1.2, w can be written in one of the following ways:

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

2. $w = \bar{b}\mu_\vartheta(u)$, with $bu \in \operatorname{Fact} s$ and $b \in A'$,

3. $w = \mu_\vartheta(u)c$, with $uc \in \operatorname{Fact} s$ and $c \in A'$,

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

In case 1, let $xw, yw \in \operatorname{Fact} t$ with $x \neq y$ letters of A. If x is ϑ-palindromic, then clearly one must have $xu \in \operatorname{Fact} s$. If $x \in A'$, then by the preceding lemma one has $\bar{x}u \in \operatorname{Fact} s$, as $|xw| > 3$. 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 $\operatorname{Fact} t = \operatorname{Fact} s$. An R-word is an episturmian word.

Proposition 3.4.10 and Theorem 3.2.3 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 3.2.3 and 4.1.2 and of Corollary 3.4.15.

Theorem 4.1.15. *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 4.1.2, 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 $ux \in \operatorname{Fact}(Epi)$,

2. $w = \bar{a}\mu_\vartheta(u)b$, with $a, b \in A'$ and $aub \in \operatorname{Fact}(Epi)$,

3. $w = \bar{a}\mu_\vartheta(u)$, with $a \in A'$ and $au \in \operatorname{Fact}(Epi)$.

In the first case, by Theorem 3.2.3 there exists a standard episturmian word $s = \psi(\Delta)$ such that $(ux)^{(+)} \in \mathrm{Fact}\, s$. Thus, by (4.5), Corollary 4.1.8 and Theorem 4.1.2, $w^{\oplus} = \mu_{\vartheta}\left((ux)^{(+)}\right)$ is a factor of the ϑ-standard word $\psi_{\vartheta}(\Delta) = \mu_{\vartheta}(s)$.

In the second case, by using Corollary 4.1.8, one has:

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

Moreover, aub is not a palindrome, since otherwise one would derive, for instance using Corollary 4.1.4, that $w = \bar{a}\mu_{\vartheta}(u)b$ is a ϑ-palindrome, which contradicts our assumption. Thus $(aub)^{(+)} = a(ub)^{(+)}a$ and the result is a consequence of Theorem 4.1.2.

In the third case, since w is not a ϑ-palindrome, by (4.5) one obtains

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

If $u = a^k$ for some $k \geq 0$, then $au^{(+)}a = a^{k+2} \in \mathrm{Fact}(Epi)$; otherwise $au^{(+)}$ is not a palindrome and $au^{(+)}a = (au^{(+)})^{(+)}$, so that $au^{(+)}a$ is episturmian by Corollary 3.4.15 and Theorem 3.2.3. Once again, the assertion follows from Theorem 4.1.2. $\square$

Corollary 4.1.16. *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 4.1.15.

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 4.1.15. 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 4.1.15, 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

$$w^{\oplus} = w\bar{a} = a\bar{w} \text{ and } w^{\ominus} = \bar{b}w = \bar{w}b \, . \tag{4.8}$$

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 (4.8) 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,

$$w = aRb = a\bar{b}\bar{R} = \bar{R}\bar{a}b \, , \tag{4.9}$$

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 (4.9) 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 (4.9) 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. □

Remark. For a finite episturmian word w, the proof of the preceding result can be simplified by using Theorem 3.2.3 and Corollary 3.4.15. Indeed, if w is not a palindrome, we can write $w = Pbt = saQ$, where P and Q are respectively the longest palindromic prefix and suffix of w, and $a, b \in A$. By Theorem 3.2.3, $w^{(+)}$ and $w^{(-)}$ are finite episturmian words; moreover bw is a factor of $w^{(-)}$, so that by Corollary 3.4.15, $bw^{(+)}$ is a finite episturmian word. By Theorem 3.2.3, $\left(bw^{(+)}\right)^{(-)}$ is a finite episturmian word, which has also $w^{(-)}$ as a factor, as one can prove similarly as in the proof of Corollary 4.1.16.

Example 4.1.17. Let τ be the Tribonacci word

$$\tau = \psi\left((abc)^{\omega}\right) = abacabaabacababacabaabacabac \cdots \, .$$

If $w = bac \in \operatorname{Fact}\tau$, one has that $w^{(+)} = bacab$ and $w^{(-)} = cabac$ are factors of τ. However, in the case of the factor $v = abacabab$, one has $v^{(+)} = abacababacaba \in \operatorname{Fact}\tau$, whereas $v^{(-)} = babacabab$ is not a factor of τ, since otherwise v would be a left special factor of τ, which is a contradiction as $v \notin \operatorname{Pref} \tau$. Nevertheless, both $v^{(+)}$ and $v^{(-)}$ are factors of any episturmian word whose directive word begins with $abcbb$. Indeed, $v = Pb$ where $P = abacaba$ is the longest palindromic prefix of v, and

$$\left(bv^{(+)}\right)^{(-)} = abacababacababacaba = \psi(abcbb)\ .$$

4.2 More antimorphisms simultaneously

Let $\mathcal{I}$ be the set of all involutory antimorphisms of A^*, and $\mathcal{I}^\omega$ be the set of infinite sequences over $\mathcal{I}$.

Let $\Theta = \vartheta_1\vartheta_2\cdots\vartheta_n\cdots \in \mathcal{I}^\omega$ and let ${}^{\oplus_i}$ be the ϑ_i-palindromic closure operator, for all $i \geq 1$. We define inductively an operator ψ_Θ by setting $\psi_\Theta(\varepsilon) = \varepsilon$, and

$$\psi_\Theta(x_1x_2\cdots x_{n+1}) = (\psi_\Theta(x_1\cdots x_n)x_{n+1})^{\oplus_{n+1}}$$

whenever $x_i \in A$ for $i \geq 1$. With this notation, ψ_{ϑ^ω} is just the operator ψ_ϑ considered in the preceding section.

If $x = x_1x_2\cdots x_n\cdots \in A^\omega$, $x_i \in A$ for $i \geq 1$, then $\psi_\Theta(x_1\cdots x_i)$ is a prefix of $\psi_\Theta(x_1\cdots x_{i+1})$ for any i, so that the infinite word

$$\psi_\Theta(x) = \lim_{n\to\infty} \psi_\Theta(x_1\cdots x_n)$$

is well defined. We call $\psi_\Theta(x)$ a *generalized pseudostandard word*. The pair (x, Θ) which determines $\psi_\Theta(x)$ can be called the *directive bi-sequence* of $\psi_\Theta(x)$. With a suitable choice of the Θ-sequences one can construct all standard episturmian words ($\Theta = R^\omega$), as well as all ϑ-standard words ($\Theta = \vartheta^\omega$). Theorem 4.2.1 below shows a less trivial example.

In the following, we shall assume $A = \{a, b\}$, ${}^\oplus = {}^{\oplus_e}$, and $\mu = \mu_e$, where e is the exchange antimorphism of A^*.

Theorem 4.2.1. *The following holds:*

$$\psi_{(eR)^\omega}(ab^\omega) = \mu^\omega(a)\ ,$$

i.e., the Thue-Morse word can be obtained via a ψ_Θ operator.

We need two lemmas.

Lemma 4.2.2. $PAL \cap b\mu(A^*) = b(ab)^*$.

Proof. The "$\supseteq$" inclusion is trivial. Let us prove the inverse inclusion. Since $PAL \cap b\mu(A^0) = \{b\} \subseteq b(ab)^*$, we assume by induction that

$$PAL \cap b\mu(A^k) \subseteq b(ab)^* \tag{4.10}$$

for all k less than some $n > 0$, and prove (4.10) for $k = n$.

Let $w \in PAL \cap b\mu(A^n)$. Since $n > 0$, w has to end with b and therefore with ab. Thus $w = bw'b$ with $w' \in PAL \cap \mu(A^{n-1})a$. If $n = 1$, then $w' = a$ and so $w = bab \in b(ab)^*$. If $n > 1$, w' has to begin with ab, so that $w' = aw''a$ with $w'' \in PAL \cap b\mu(A^{n-2}) \subseteq b(ab)^*$. Hence $w = baw''ab \in b(ab)^*$. □

Lemma 4.2.3. *For any* $n \geq 0$,

$$PAL \cap \text{Suff}\left(\mu^{2n+1}(a)\right) = \{\varepsilon\} \cup \left\{\mu^{2k}(b) \mid 0 \leq k \leq n\right\} .$$

Proof. Since $PAL \cap \text{Suff}\, \mu(a) = \{\varepsilon, b\}$, it suffices to show that for any $n > 0$,

$$PAL \cap \text{Suff}\left(\mu^{2n+1}(a)\right) = \{b\} \cup \mu^2\left(PAL \cap \text{Suff}\left(\mu^{2n-1}(a)\right)\right) . \tag{4.11}$$

Since $\mu^{2n+1}(a)$ ends with aab for all $n > 0$, the preceding lemma shows that all palindromic suffixes of $\mu^{2n+1}(a)$ different from b have even length. Indeed, suppose that q is a palindromic suffix of $\mu^{2n+1}(a)$ of odd length. Since q has to begin with b, one can write $q = b\mu(u)$ with $u \in \text{Suff}\, \mu^{2n}(a)$. From the preceding lemma, $q \in b(ab)^*$ so that if $q \neq b$, q and $\mu^{2n+1}(a)$ end with bab, which is a contradiction. Therefore, all palindromic suffixes of $\mu^{2n+1}(a)$ different from b are in $\mu\left(\text{Suff}\left(\mu^{2n}(a)\right)\right)$.

If w is a word with odd length, then $\mu(w)$ cannot be a palindrome, because its minimal (nonempty) median factor is ab or ba. This implies

$$\mu\left(\text{Suff}\left(\mu^{2n}(a)\right)\right) \cap PAL = \mu^2\left(\text{Suff}\left(\mu^{2n-1}(a)\right)\right) \cap PAL .$$

By Corollary 4.1.4, $w \in PAL \iff \mu^2(w) \in PAL$, so that

$$\mu^2\left(\text{Suff}\left(\mu^{2n-1}(a)\right)\right) \cap PAL = \mu^2\left(\text{Suff}\left(\mu^{2n-1}(a) \cap PAL\right)\right) .$$

This proves (4.11). □

Proof of Theorem 4.2.1. It suffices to show that, for any $n \geq 0$,

$$\mu^{2n+2}(a) = \left(\mu^{2n+1}(a)b\right)^{(+)} , \quad (4.12)$$

$$\mu^{2n+1}(a) = \left(\mu^{2n}(a)b\right)^{\oplus} . \quad (4.13)$$

Let us first prove that (4.12) is equivalent to the statement

$$Q_R\left(\mu^{2n+1}(a)b\right) = bb . \quad (4.14)$$

Indeed, suppose that (4.12) is satisfied. Since $|\mu^{2n+2}(a)| = 2|\mu^{2n+1}(a)|$, one derives that (4.14) holds. Conversely, suppose that (4.14) is satisfied. Since $\mu^{2n+1}(a)$ ends with b, one can write $\mu^{2n+1}(a) = ub$ with $u \in A^*$, so that

$$\left(\mu^{2n+1}(a)b\right)^{(+)} = ubb\tilde{u} = \mu^{2n+1}(a)\widetilde{\mu^{2n+1}(a)} .$$

As is well known (cf. [38]), for all $n \geq 0$ one has $\widetilde{\mu^{2n+1}(a)} = \mu^{2n+1}(b)$. Therefore,

$$\left(\mu^{2n+1}(a)b\right)^{(+)} = \mu^{2n+1}(a)\mu^{2n+1}(b) = \mu^{2n+2}(a) .$$

Equation (4.14) can be equivalently restated saying that any nonempty palindromic suffix of $\mu^{2n+1}(a)$ is preceded by a. By Lemma 4.2.3, the set of nonempty palindromic suffixes of $\mu^{2n+1}(a)$ is $\{\mu^{2k}(b) \mid 0 \leq k \leq n\}$. Since

$$\mu^{2n+1}(a) = \mu^{2n}(a)\mu^{2n}(b) = \mu^{2n}(a)\mu^{2n-1}(b)\mu^{2n-1}(a) ,$$

by iterating this formula one has that for any $k \leq n$ the suffix $\mu^{2k}(b)$ is preceded by the word $\mu^{2k}(a)$, which ends with a. This proves (4.12).

By Corollary 4.1.8 and equation (4.12), one has

$$\left(\mu^{2n}(a)b\right)^{\oplus} = \mu\left(\left(\mu^{2n-1}(a)b\right)^{(+)}\right) = \mu\left(\mu^{2n}(a)\right) = \mu^{2n+1}(a)$$

which proves (4.13). □

4.3 Words generated by nonempty seeds

We now consider a generalization of the construction of ϑ-standard words. We recall that the operator ψ_ϑ was defined in Section 3.2 by setting $\psi_\vartheta(w) = D_w(\varepsilon)$

for any $w \in A^*$. We now consider a fixed arbitrary seed $u_0 \in A^*$ instead of ε and therefore set

$$\hat{\psi}_\vartheta(w) = D_w(u_0) \ .$$

Equivalently, the map $\hat{\psi}_\vartheta : A^* \to A^*$ can be defined by setting $\hat{\psi}_\vartheta(\varepsilon) = u_0$, and

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

for $u \in A^*$ and $a \in A$. 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 $\{w_n\} = \operatorname{Pref} t \cap A^n$ 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 4.3.1. Let $A = \{a, b, c\}$, ϑ be the involutory antimorphism exchanging a and b and fixing c, $u_0 = acbbc$, and $w = abc$. Then

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

Proposition 4.3.2. *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*

$$\hat{\psi}_\vartheta(\operatorname{Pref} \Delta \setminus \{\varepsilon\}) \cup (PAL_\vartheta \cap \operatorname{Pref} u_0) \ , \tag{4.15}$$

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 $|w| > k$, and let $\Delta = xt_1t_2\cdots t_n\cdots$ 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 3.1.11 one has $w^{\oplus} = u_{n+1} \in \operatorname{Pref} s$. This proves point 1.

By the definition of ϑ-standard words with seed, all the words in the set (4.15) are ϑ-palindromic prefixes of s. Conversely, let w be a ϑ-palindromic prefix of s. If $|w| \leq k$, then trivially $w \in PAL_{\vartheta} \cap \operatorname{Pref} u_0$. If $|w| > k$, then by following the same argument used for point 1, one has that there exists an integer $n > 0$ such that $w = w^{\oplus} = u_n \in \hat{\psi}_{\vartheta}(\operatorname{Pref} \Delta)$. 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 [27] for episturmian words, one can prove the following:

Proposition 4.3.3. *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 $v \in \operatorname{Fact}(u_m)$. 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 $\rho_{m+1} \leq |p| + 2$. Moreover, for all $n > m$ we have $\rho_{n+1} \leq \max\{\rho_n, |p| + 2\}$. Indeed, let w be a first return to p in u_{n+1} of maximal length, so that its shift is ρ_{n+1}. If $w \in \operatorname{Fact}(u_n)$, then $\rho_{n+1} = \rho_n$. Let us suppose that w is not a factor of u_n. We set $u_n = \lambda p = p\bar{\lambda}$ and $u_{n+1} = \alpha w\beta$ with $\alpha, \beta, \lambda \in A^*$. Then $|\alpha| \geq |\lambda|$ and $|\beta| \geq |\lambda|$, otherwise w would be a factor of u_n. Therefore, as $|u_{n+1}| \leq 2|u_n| + 2$, we obtain

$$|w| \leq |u_{n+1}| - 2|\lambda| = |u_{n+1}| - 2|u_n| + 2|p| \leq 2|p| + 2 \ ,$$

so that $\rho_{n+1} \leq |p| + 2$. Thus in any case $\rho_{n+1} \leq \max\{\rho_n, |p| + 2\}$. As $\rho_{m+1} \leq |p| + 2$, 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 $|p| + 2 = |u_m| + 2$. □

The following result generalizes Proposition 4.1.11, and can be proved analogously.

Proposition 4.3.4. *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.*

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 $\operatorname{Fact} s = \operatorname{Fact} t$.

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 4.3.5. Let A, ϑ, and u_0 be defined as in Example 4.3.1, and let $x = a$. Then

$$\begin{aligned} \phi_a(a) &= \hat{\psi}_\vartheta(aa)\hat{\psi}_\vartheta(a)^{-1} = acbbcaacb\ , \\ \phi_a(b) &= \hat{\psi}_\vartheta(ab)\hat{\psi}_\vartheta(a)^{-1} = acbbca\ , \\ \phi_a(c) &= \hat{\psi}_\vartheta(ac)\hat{\psi}_\vartheta(a)^{-1} = acbbcaacbc\ . \end{aligned} \tag{4.16}$$

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 4.3.6. *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,

$$\hat{\psi}_\vartheta(xvy) = \bar{\lambda}\bar{y}\hat{\psi}_\vartheta(xv_1)y\lambda = \hat{\psi}_\vartheta(xv)\hat{\psi}_\vartheta(xv_1)^{-1}\hat{\psi}_\vartheta(xv) . \tag{4.17}$$

By a similar argument one has:

$$\psi(vy) = \psi(v)\psi(v_1)^{-1}\psi(v) . \tag{4.18}$$

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 (4.17), and by (4.18), 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

$$\hat{\psi}_\vartheta(xvy) = \left(\hat{\psi}_\vartheta(xv)y\right)^{\oplus} = \hat{\psi}_\vartheta(xv)\alpha_y^{-1}\hat{\psi}_\vartheta(xv) . \tag{4.19}$$

By induction, this implies

$$\hat{\psi}_\vartheta(xvy) = \phi(\psi(v))\hat{\psi}_\vartheta(x)\alpha_y^{-1}\phi(\psi(v))\hat{\psi}_\vartheta(x) . \tag{4.20}$$

By using (4.19) 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

$$\psi(vy) = \psi(v)y\psi(v)\ . \tag{4.21}$$

Thus from (4.20) 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

$$\hat{\psi}_\vartheta(xvy) = \hat{\psi}_\vartheta(xv)y^\oplus\hat{\psi}_\vartheta(xv)\ . \tag{4.22}$$

By induction we then obtain

$$\begin{aligned}\hat{\psi}_\vartheta(xvy) &= \hat{\psi}_\vartheta(xv)y^\oplus\hat{\psi}_\vartheta(xv) \qquad (4.23)\\ &= \phi(\psi(v))\hat{\psi}_\vartheta(x)y^\oplus\phi(\psi(v))\hat{\psi}_\vartheta(x)\ .\end{aligned}$$

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 (4.23) and (4.21) 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. □

Example 4.3.7. Let us refer to Example 4.3.1. We have $w = abc$, $u_0 = acbbc$, and ϑ defined by $\bar{a} = b$, $\bar{c} = c$. By the preceding theorem, one has

$$\hat{\psi}_\vartheta(abc) = \phi_a(\psi(bc))\hat{\psi}_\vartheta(a)\ .$$

Since $\psi(bc) = bcb$, $\phi_a(bcb) = \phi_a(b)\phi_a(c)\phi_a(b)$, and $\hat{\psi}_\vartheta(a) = (u_0a)^\oplus = acbbcaacb$, by using (4.16) we obtain

$$\hat{\psi}_\vartheta(abc) = acbbcaacbbcaacbcacbbcaacbbcaacb \ ,$$

as already shown in Example 4.3.1.

From Theorem 4.3.6, in the case that w is an infinite word, we obtain:

Theorem 4.3.8. *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 $w_n = \operatorname{Pref} w \cap A^n$ for all $n \geq 0$. From Theorem 4.3.6, 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 4.3.6 one has

$$\psi_\vartheta(xw) = \phi_x(\psi(w))\psi_\vartheta(x) = \phi_x(\psi(w))x^\oplus \ . \tag{4.24}$$

Moreover, one easily derives that

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

When $u_0 = \varepsilon$ and $\vartheta = R$, 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 (4.24) one obtains the following formula due to Justin [34]:

$$\psi(xw) = \mu_x(\psi(w))x \ . \tag{4.25}$$

It is noteworthy that Theorem 4.3.6 provides an alternate proof of Theorem 4.1.2:

Proof of Theorem 4.1.2. 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 (4.24) and (4.25) 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 4.1.14 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 4.3.12 we shall prove that any *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 $(u_0x)^{\oplus}$.

In the following, we shall set

$$u_1 = \hat{\psi}_{\vartheta}(x) = (u_0x)^{\oplus} ,$$

so that $\phi_x(a) = (u_1a)^{\oplus}u_1^{-1}$ 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 $\phi_x(a) = u_1a^{\oplus}$ otherwise.

Lemma 4.3.9. *Let $X = \phi_x(A)$. 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 $w = \phi(a) = (u_1a)^{\oplus}u_1^{-1}$. Thus $wu_1 = (u_1a)^{\oplus}$, 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_1v' for some $v' \in A^*$. Then one has $wu_1 = \phi(a)u_1v'$ and, as shown above, u_1 is a prefix of $\phi(a)u_1$, which concludes the proof. □

Recall (cf. [4]) 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 4.3.10. *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 4.1). 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 $ps = u_1 a \bar{a} = p u_1 \bar{a}$, so that $u_1 a = p u_1$. 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 $w \in \operatorname{Pref} \rho'$. By Lemma 4.3.9, 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

$$(u_1 a)^\oplus = \phi(a) u_1 = p s u_1 = p s w \bar{s} = p u_1 \bar{s} .$$

Since $p \neq \varepsilon$, by Lemma 3.1.12 one obtains $\bar{s} = \varepsilon$. Hence $u_1 \rho = \rho' \in X^*$. □

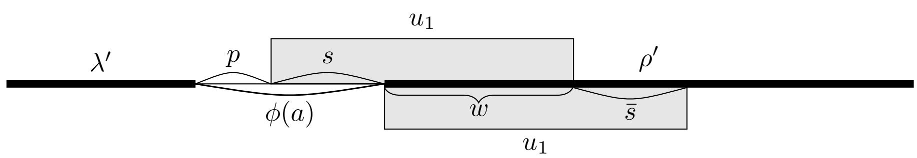

Figure 4.1: Proposition 4.3.10

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. [4]) 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*. By Lemma 4.3.9, the code $X = \phi_x(A)$ has the property that there exists an integer $n > 0$ such that $u_1 \in \operatorname{Pref} v$ for all $v \in X^n$; from Proposition 4.3.10 it follows that all pairs of $X^n \times X^n$ are synchronizing for X, so that X has a *bounded synchronization delay* and therefore a finite deciphering delay.

Lemma 4.3.11. *Let $X = \phi_x(A)$ and $w = ru_1azs \in X^*$, with $a, z \in A$ and $r, s \in A^*$. If we 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.*

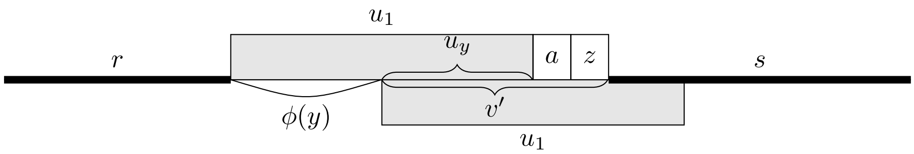

Figure 4.2: Lemma 4.3.11

Proof. Let $w \in X^*$ be such that $w = ru_1azs$, with $z \in A$. From Proposition 4.3.10 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 4.2). Since $v = v's \in X^*$, from Lemma 4.3.9 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 4.3.12. *Let $t = \hat{\psi}_\vartheta(x\Delta)$ be a ϑ-standard word with seed. Then there exists an integer $N \geq 0$ such that any left special factor of t of length greater than or equal to N is a prefix of t.*

Proof. Set $z = \psi(\Delta) = z_1z_2\cdots z_n\cdots$, where $z_i \in A$ for all $i \geq 1$. From Theorem 4.3.6 we have that $t = \phi(z)$, so that t can be factorized uniquely as

$$t = \phi(z_1)\phi(z_2)\cdots\phi(z_n)\cdots \in X^\omega ,$$

where $X = \phi_x(A)$. 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

$$w = p\phi(z_{i+1}\cdots z_{i+h})s = p'\phi(z_{j+1}\cdots z_{j+k})s' , \tag{4.26}$$

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 (4.26) 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 4.3.9, 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 4.3.10, $(p', \phi(z_{j+1} \cdots z_{j+k})s')$ 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 $u_1 c$ is a prefix of both $\phi(z_{i+1} \cdots z_{i+h})$ and $\phi(z_{j+1} \cdots z_{j+k})$. By Lemma 4.3.11 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 4.3.9, there exists a letter d such that $u_1 d$ is a prefix of w', so, by applying Lemma 4.3.11 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 4.3.9, 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 4.3.9, 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. □

We observe that the proof of the preceding theorem shows that for a ϑ-standard word s with seed u_0, all left special factors of length greater than or equal to $N = 2|u_1| + 3$ are prefixes of s. However, this bound is not tight. In fact, for instance, if $u_0 = \varepsilon$ then $N = 5$, whereas from Theorem 4.1.14 one has that all left special factors of a ϑ-standard word s, having length at least 3, are prefixes of s.

4.4 The class SW_ϑ and ϑ-episturmian words

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 4.1.14, the class of ϑ-standard words is included in $SW_\vartheta(3)$.

Proposition 4.4.1. *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 4.4.2. *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.

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 4.4.6). 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 4.1 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 4.1: 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 4.4.6

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

Lemma 4.4.3. *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 4.4.4. *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 $\beta = x_1 x_2 \cdots x_n$ 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 $B_i x_i$ is a prefix of w. The infinite word $x = x_1 x_2 \cdots x_n \cdots$ will be called the *subdirective word* of w. The proof of Proposition 4.4.4 shows that for any $i > 0$, B_{i+1} coincides with the prefix of w ending with the first occurrence of $\bar{x}_i B_i$.

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

Lemma 4.4.5. *Let $w \in SW_\vartheta(N)$. With the above notation, let $n > 1$ be such that $x_n = x_k$ for some $k < n$ with $|B_k| \geq N-2$. Then $B_{n+1} = (B_n x_n)^\oplus$.*

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

Suppose then $Q\rho = \bar{x}_n B_k x_n v \bar{x}_n B_k$ for some $v \in A^*$. Since $Q\rho$ is a suffix of B_n, one has that $\bar{\rho}Q = B_k x_n \bar{v} Q$ is a prefix of B_n (see Figure 4.3). Now there is no proper suffix u of $\bar{v}$ such that uQ is left special in w. Indeed, if such u existed, then uQ would be a prefix of B_n, and so $Q\bar{u}$ would be a suffix of B_n, contradicting (as $|u| < |\rho|$) the fact that $Q\rho$ begins with the rightmost occurrence of Q in B_n. Hence every occurrence of Q in w is preceded by $\bar{v}$. Since $\rho x_n = v \bar{x}_n B_k x_n$ is a factor of w, one obtains $v = \bar{v}$, so that $Q\rho x_n = \bar{x}_n B_k x_n v \bar{x}_n B_k x_n$ is a ϑ-palindromic suffix of $B_n x_n$ longer than Q, a contradiction.

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

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

where (s, w') is the leftmost occurrence of Q in w. By Lemma 4.4.3, $sQ\bar{s} = (B_n x_n)^\oplus$ is a prefix of w. From this one derives $B_{n+1} = (B_n x_n)^\oplus$. □

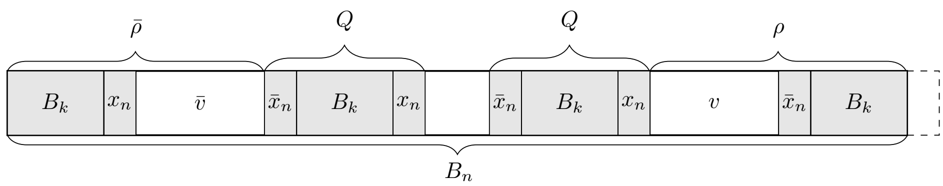

Figure 4.3: Lemma 4.4.5

Theorem 4.4.6. *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 $s \in SW_\vartheta(N)$, $x_1x_2\cdots x_n\cdots$ be its subdirective word, and $(B_i)_{i\geq 0}$ the sequence of all ϑ-palindromic prefixes of s. 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 4.4.5, $B_{n+1} = (B_n x_n)^{\oplus}$.

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 4.3.12. □

Let us set

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

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

Corollary 4.4.7. W_ϑ *coincides with the set of all* ϑ*-words with seed.*

Let $s \in SW_\vartheta(N)$. We call *critical integer* h of s 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 $x_n = x_k$. We observe that the proof of Theorem 4.4.6 shows that for any given $s \in SW_\vartheta(N)$ having critical integer h, one has that for all $n \geq h$, $B_{n+1} = (B_n x_n)^{\oplus}$.

Corollary 4.4.8. *Any standard* ϑ*-episturmian word is a* ϑ*-standard word with seed. Moreover, if* $s \in SEpi_\vartheta$ *and* $x = x_1x_2\cdots x_n\cdots$ *is its subdirective word, then the critical integer* h *of* s *is equal to the minimal integer* p *such that* $\mathrm{alph}\, x = \mathrm{alph}(x_1\cdots x_{p-1})$.

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 4.4.6, 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 $h = p$. □

Proposition 4.4.9. *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 4.4.5, 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. □

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. [27]).

Proposition 4.4.10. *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 $u \in \operatorname{Fact} s$ 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 4.4.9. 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 [2] (see also [12]). However this cannot be extended to ϑ-episturmian words. For instance, let s be the standard ϑ-episturmian word $(abaca)^\omega$, where $\vartheta(a) = a$ and $\vartheta(b) = c$. Then aba is a first return to a in s, but it is not a ϑ-palindrome.

4.5 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 4.4.8 and Theorem 4.3.6, one derives the following

Proposition 4.5.1. *Let s be a standard ϑ-episturmian word, h be its critical integer, and $x = x_1x_2\cdots x_n\cdots$ be the subdirective word of s. Then s is the image, by an injective morphism, of the standard episturmian word t whose directive word is $x_{h+1}x_{h+2}\cdots x_n\cdots$.*

However, this can be improved. In fact, the next results will show (cf. Theorem 4.5.5) 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 4.5.2. $PAL_\vartheta^* = \mathcal{P}^*$.

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 $v \in PAL_\vartheta$, 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 4.5.3. 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 $acbdaaecba.abaca \in PAL_\vartheta^2$ 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

$$s = \pi_1\pi_2\cdots\pi_n\cdots\,, \qquad \text{with } \pi_i \in \mathcal{P} \text{ for all } i > 0\,. \tag{4.27}$$

Lemma 4.5.4. *Let $s \in SEpi_\vartheta$, 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 $\operatorname{Pref} u \cap \operatorname{Suff} \bar{u}$, that is, a nonempty ϑ-palindromic prefix of u, which contradicts the hypothesis that u is a proper prefix of π_h. Thus $q = \bar{u}q'u$ for some $q' \in PAL_\vartheta$.

We have $\pi_1 \cdots \pi_{h-1} \in \mathcal{P}^*$ and, by Proposition 4.5.2, $q' \in \mathcal{P}^*$. 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 4.5.5. *Let $s \in A^\omega$ be a standard ϑ-episturmian word, Δ be its subdirective word, and $B = \operatorname{alph} \Delta$. 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 (4.27). 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 $a_i w, a_j w \in \operatorname{Fact} t$. Then $\bar{a}_i \mu(w), \bar{a}_j \mu(w) \in \operatorname{Fact} s$, 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 $w \in \operatorname{Pref} t$. 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 (4.27) and beginning with x_n is $\mu(x_n)$, we have $B_n \mu(x_n) \in \operatorname{Pref} s$, 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 4.5.6. *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 4.1.2 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 4.1.2. □

Proposition 4.5.7. *Let $\mu : B^* \to A^*$ be an injective morphism such that*

1. *$\mu(x) \in PAL_\vartheta$ for all $x \in B$,*

2. *$\operatorname{alph} \mu(x) \cap \operatorname{alph} \mu(y) = \emptyset$ 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 $x_1ux_2 \in \operatorname{Fact} t$. 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 $aw, bw \in \operatorname{Fact} s$. Hence there exist two (distinct) letters $x_a, x_b \in B$ such that $x_au, x_bu \in \operatorname{Fact} t$. Then u is a left special factor of t and hence a prefix of t, so that $w = \mu(u)$ is a prefix of s. ☐

Example 4.5.8. 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 4.5.7 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 4.5.5 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

$$x \in B,\ x \neq \bar{x} \Longrightarrow \bar{x} \notin B\ . \tag{4.28}$$

Proposition 4.5.9. *Let s be a standard ϑ-episturmian word and Δ be its subdirective word. Then $B = \operatorname{alph}\Delta$ is ϑ-skew.*

Proof. We can factorize s as in (4.27). By Theorem 4.5.5, 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 $\bar{x} \in \operatorname{Pref}\pi_k$. Without loss of generality, we can assume $n < k$. By Lemma 4.5.4, 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 $\pi_1 \cdots \pi_{k-1} \in \operatorname{Suff} w$, 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 4.1.14. For instance, consider the ϑ-standard word s with $\vartheta = E \circ R$ and $\Delta(s) = (ab)^\omega$. One has $s = abbaababbaabbaab \cdots$. As one easily verifies, b and ba are two left special factors which are not prefixes. Hence in general, a ϑ-standard word is not standard ϑ-episturmian.

The next proposition gives a characterization of ϑ-standard words which are standard ϑ-episturmian.

Proposition 4.5.10. *A ϑ-standard word s is standard ϑ-episturmian if and only if $B = \operatorname{alph}(\Delta(s))$ is ϑ-skew.*

Proof. Let s be a ϑ-standard word such that B is ϑ-skew. By Theorem 4.1.2, one has $s = \mu_\vartheta(t)$, where $t = \psi(\Delta(s))$ is a standard episturmian word. The morphism μ_ϑ satisfies condition 1 in Proposition 4.5.7 by definition. By (4.28), one easily derives that the restriction of μ_ϑ to alph $t = B$ satisfies also the second statement of Proposition 4.5.7, so that $s = \mu_\vartheta(t)$ is a standard ϑ-episturmian word.

The converse is a consequence of Proposition 4.5.9, as the subdirective word of a ϑ-standard word s is $\Delta(s)$. □

Example 4.5.11. 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 3.2.2) s is not ϑ-standard, since ab is a prefix of s, but $(ab)^\oplus = abca$ is not. Another example is the word s considered in Example 4.5.8: 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 4.5.12. *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

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

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 $\operatorname{alph} \hat{\Delta}$ is $\hat{\vartheta}$-skew, by the previous proposition $\hat{s}$ is also $\hat{\vartheta}$-episturmian.

By Theorem 4.1.2, 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

$$\hat{s} = \mu_{\hat{\vartheta}}(\varphi(\psi(\Delta))) \, . \tag{4.29}$$

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

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

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 (4.29), 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 4.5.13. 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 $g(\hat{s}) = s$.

Chapter 5

Conclusions

5.1 Summary

This work is about Sturmian words and their generalizations, an important topic in combinatorics on words.

Among other well known notions, we have defined the fractional root z_w of a word w, that is, the prefix of w whose length is the minimal period π_w of w. Another new fundamental notion is that of ϑ-palindromes, i.e., fixed points of an involutory antimorphism ϑ of the free monoid of words A^*.

We have given an analysis of the periodical structure of factors of Sturmian words, which has led to two new characterizations, showing that the property of being Sturmian (or not) for a finite word is completely determined by its fractional root:

1. $w \in A^+$ is Sturmian if and only if z_w is standard (Theorem 2.2.3).

2. $w \in A^+$ is Sturmian if and only if $|z_w| = R_{z_w^2} + 1$ (Theorem 2.2.8).

Both characterizations naturally produce linear-time algorithms for the recognition of finite Sturmian words (Section 2.3), which is an important problem also for matters of discrete geometry and computer vision.

As a byproduct, the following formula counting the number $p(n)$ of primitive Sturmian words of any length $n \geq 2$ is found (Proposition 2.2.12):

$$p(n) = \sum_{i=1}^{n}(n+1-i)\phi(i) - \sum_{\substack{d|n \\ d\neq n}} d\phi(d) \; .$$

We have then focused on the set $St \cap PAL$ of Sturmian palindromes, showing that a Sturmian palindrome is necessarily a median factor of a central word (Corollary 2.4.2) having the same minimal period (Theorem 2.4.13). Among other noteworthy structural results, we have proved that a palindrome w is Sturmian if and only if $\pi_w = R_w + 1$ (Theorem 2.4.18), after a similar characterization for central words (Theorem 2.4.16).

A formula for the enumeration function $g(n)$ of Sturmian palindromes has been found for all $n \geq 0$:

$$g(n) = 1 + \sum_{i=0}^{\lceil n/2 \rceil - 1} \phi(n - 2i)$$

(see Theorem 2.5.1). This has allowed to prove that the asymptotic density of central words in $St \cap PAL$ vanishes (Proposition 2.5.5), and so does the asymptotic density of Sturmian palindromes in St. (Proposition 2.5.7).

Next, we have introduced the important pseudopalindrome closure operators. We have proved several properties linking ϑ-palindrome closure and periodical structure. For instance, given any $w \in A^*$, the minimal periods of the right closure $w^{\oplus}$ and of the left closure $w^{\ominus}$ coincide (Proposition 3.1.4), and are the same if and only if z_w is the product of two ϑ-palindromes (Theorem 3.1.6).

The iteration of palindrome closure operators produces standard Sturmian words, and more generally standard episturmian words. The properties of the factors of such words, in relation to palindrome closure, have been analysed. Slightly stronger results have been found in the Sturmian case; anyway the main result is that both closures $w^{(+)}$ and $w^{(-)}$ of a finite episturmian word are episturmian themselves (Theorems 3.3.4 and 3.4.14), and even factors of a common episturmian word (Proposition 3.3.6 and Corollary 4.1.16).

In the last chapter 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.

We have studied several structural properties of these words. In the ϑ-standard case, this has been done mainly in relation with ϑ-palindrome closure, whereas for standard ϑ-episturmian words, the main tool is the factorization (4.27) of such words with unbordered ϑ-palindromes.

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 4.5.6 and Proposition 4.5.10). Moreover, the two preceding classes are strictly included in the class of ϑ-standard words with seed (see Fig. 5.1).

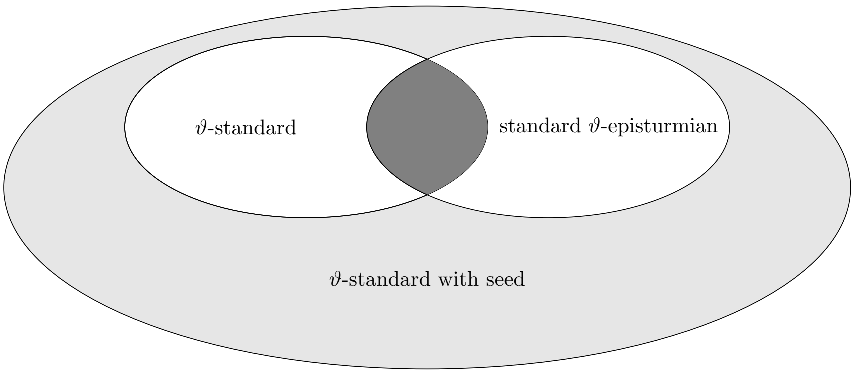

Figure 5.1: Generalized episturmian words

A basic theorem (see Theorem 4.4.6) 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 4.3.6, 4.1.2, and 4.5.5) showing that the words of such families are suitable morphic images of standard episturmian words.

5.2 Further research and open problems

Sturmian and episturmian words are the subject of much ongoing research within the scientific community. Here we list some possible developments which are related to our work, as presented in the preceding chapters.

- Very recently, J. Currie and K. Saari [18] have improved upon Theorem 2.2.3, by characterizing the finite standard words which are conjugate to the fractional roots of the factors of a given Sturmian word. This could possibly lead to the development of algorithms testing whether a word is a factor of a Sturmian word of given slope (up to the first partial quotients in its continued fraction expansion).

- It could be interesting to give a deeper analysis of the structure of generalized pseudostandard words, as defined in Section 4.2.

- The introduction of involutory antimorphisms in our work was motivated for example by biology (a common example of involution is the Watson-Crick one, for the DNA "alphabet" of four bases). Recently (cf. [13]), the following result was proved, showing the occurrence of involutory antimorphisms from purely combinatorial conditions:

  Theorem. *Let w be an infinite word over A satisfying the following three conditions:*

  1. *every left special factor of w is a prefix of it,*
  2. *w has at most one right special factor of each length,*
  3. *for some constant k and all $n \geq 1$ one has*

  $$\operatorname{card}(\operatorname{Fact} w \cap A^{n+1}) - \operatorname{card}(\operatorname{Fact} w \cap A^{n}) = k\,. \tag{5.1}$$

  Then there exists an involutory antimorphism ϑ of A^* such that w is standard ϑ-episturmian.

  We remark that by requiring that (5.1) holds also for $n = 0$, a well-known characterization of Arnoux-Rauzy words is obtained.

  The above theorem is not a characterization of infinite words which are standard ϑ-episturmian for some ϑ; this could be achieved by substituting (5.1) with a suitable weaker condition.

- 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 4.5.7 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 [27, 35].

  A second problem is to determine whether morphisms $\varphi : X^* \to A^*$ 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 4.1.2). Both questions are being addressed in [10].

Bibliography

[1] J.-P. Allouche and J. Shallit. *Automatic Sequences*. Cambridge University Press, Cambridge UK, 2003.

[2] 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.

[3] J. Berstel and A. de Luca. Sturmian words, Lyndon words and trees. *Theoretical Computer Science*, 178:171–203, 1997.

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

[5] J. Berstel and M. Pocchiola. Random generation of finite Sturmian words. *Discrete Mathematics*, 153:29–39, 1996.

[6] P. Bhattacharya, R. A. Melter, and A. Rosenfeld, editors. *Vision Geometry*, volume 119 of *Contemporary Mathematics*. AMS, Providence, RI, 1991.

[7] J.-P. Borel and C. Reutenauer. Palindromic factors of billiard words. *Theoretical Computer Science*, 340:334–348, 2005.

[8] M. Boshernitzan and A. S. Fraenkel. A linear algorithm for nonhomogeneous spectra of numbers. *J. Algorithms*, 5:187–198, 1984.

[9] S. Brlek, S. Hamel, M. Nivat, and C. Reutenauer. On the palindromic complexity of infinite words. *International Journal of Foundations of Computer Science*, 15:293–306, 2004.

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

[11] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni. On different generalizations of episturmian words. *Theoretical Computer Science*, to appear, 2007.

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

[13] M. Bucci, A. De Luca, and L. Q. Zamboni. On θ-episturmian words. Manuscript in progress, 2007.

[14] A. Carpi. Private communication, 2005.

[15] A. Carpi and A. de Luca. Codes of central Sturmian words. *Theoretical Computer Science*, 340:220–239, 2005.

[16] W.-F. Chuan. Moments of conjugacy classes of binary words. *Theoretical Computer Science*, 310:273–285, 2004.

[17] P. M. Cohn. *Algebra*, volume 1–2. John Wiley & Sons, Chichester, 2nd edition, 1982.

[18] J. D. Currie and K. Saari. Least periods of factors of infinite words. Manuscript in progress, 2007.

[19] F. D'Alessandro. A combinatorial problem on trapezoidal words. *Theoretical Computer Science*, 273:11–33, 2002.

[20] A. de Luca. On some combinatorial problems in free monoids. *Discrete Mathematics*, 38:207–225, 1982.

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

[22] A. de Luca. On the combinatorics of finite words. *Theoretical Computer Science*, 218:13–39, 1999.

[23] A. de Luca and A. De Luca. Combinatorial properties of Sturmian palindromes. *International Journal of Foundations of Computer Science*, 17:557–574, 2006.

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

[25] A. de Luca and A. De Luca. Some characterizations of finite Sturmian words. *Theoretical Computer Science*, 356:118–125, 2006.

[26] A. de Luca and F. Mignosi. Some combinatorial properties of Sturmian words. *Theoretical Computer Science*, 136:361–385, 1994.

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

[28] X. Droubay and G. Pirillo. Palindromes and Sturmian words. *Theoretical Computer Science*, 223:73–85, 1999.

[29] F. Durand. A characterization of substitutive sequences using return words. *Discrete Mathematics*, 179:89–101, 1998.

[30] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, Oxford UK, 1979.

[31] T. Harju and D. Nowotka. Minimal Duval extensions. *Internat. J. Found. Comput. Sci.*, 15:349–354, 2004.

[32] O. Jenkinson and L. Q. Zamboni. Characterisations of balanced words via orderings. *Theoretical Computer Science*, 310:247–271, 2004.

[33] J. Justin. On a paper by Castelli, Mignosi, Restivo. *Theoretical Informatics and Applications*, 34:373–377, 2000.

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

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

[36] M. Lindenbaum and A. Bruckstein. On recursive $o(n)$ partitioning of a digitized curve into digital straight segments. *IEEE Trans. Pattern Anal. Machine Intelligence*, 15:949–953, 1993.

[37] M. Lindenbaum and J. Koplowitz. A new parametrization of digital straight lines. *IEEE Trans. Pattern Anal. Machine Intelligence*, 13:847–852, 1991.

[38] M. Lothaire. *Combinatorics on Words*. Addison-Wesley, Reading MA, 1983. Reprinted by Cambridge University Press, Cambridge UK, 1997.

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

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

[41] S. Mac Lane. *Categories for the Working Mathematician*. Springer-Verlag, New York, 2nd edition, 1998.

[42] F. Mignosi and L. Q. Zamboni. A note on a conjecture of Duval and Sturmian words. *Theoret. Inform. Appl.*, 36:1–3, 2002.

[43] M. Morse and G. A. Hedlund. Symbolic dynamics. *American Journal of Mathematics*, 60:815–866, 1938.

[44] M. Morse and G. A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *American Journal of Mathematics*, 62:1–42, 1940.

[45] G. Pirillo. A curious characteristic property of standard Sturmian words. In G.-C. Rota, H. Crapo, and D. Senato, editors, *Algebraic combinatorics and computer science*, pages 541–546. Springer Italia, Milan, 2001.

[46] N. J. A. Sloane. *The On-Line Encyclopedia of Integer Sequences*. AT&T Research, `http://www.research.att.com/~njas/sequences/`.