Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Combinatorics 25 (2004) 685–705
European Journal of Combinatorics
www.elsevier.com/locate/ejc

Harmonic and gold Sturmian words

Arturo Carpi[a,c], Aldo de Luca[b,c]

[a]Dipartimento di Matematica e Informatica dell'Università di Perugia, Via Vanvitelli 1, 06123 Perugia, Italy
[b]Dipartimento di Matematica e Applicazioni dell'Università di Napoli "Federico II", Via Cintia, Complesso Universitario di Monte S. Angelo, 80126 Napoli, Italy
[c]Istituto di Cibernetica del C. N. R. "E. Caianiello", Pozzuoli (NA), Italy

Received 10 October 2003; accepted 17 October 2003

Abstract

In the combinatorics of Sturmian words an essential role is played by the set *PER* of all finite words w on the alphabet $\mathcal{A} = \{a, b\}$ having two periods p and q which are coprime and such that $|w| = p + q - 2$. As is well known, the set *St* of all finite factors of all Sturmian words equals the set of factors of *PER*. Moreover, the elements of *PER* have many remarkable structural properties. In particular, the relation $Stand = \mathcal{A} \cup PER\{ab, ba\}$ holds, where *Stand* is the set of all finite standard Sturmian words. In this paper we introduce two proper subclasses of *PER* that we denote by *Harm* and *Gold*. We call an element of *Harm* a harmonic word and an element of *Gold* a gold word. A harmonic word w beginning with the letter x is such that the ratio of two periods p/q, with $p < q$, is equal to its slope, i.e., $(|w|_y + 1)/(|w|_x + 1)$, where $\{x, y\} = \{a, b\}$. A gold word is an element of *PER* such that p and q are primes. Some characterizations of harmonic words are given and the number of harmonic words of each length is computed. Moreover, we prove that *St* is equal to the set of factors of *Harm* and to the set of factors of *Gold*. We introduce also the classes **Harm** and **Gold** of all infinite standard Sturmian words having infinitely many prefixes in *Harm* and *Gold*, respectively. We prove that **Gold** ∩ **Harm** contain continuously many elements. Finally, some conjectures are formulated.

1. Introduction

Sturmian words have been extensively studied by several authors for at least two centuries. They have many applications in various different fields like algebra, theory of numbers, physics (symbolic dynamics, crystallography), and computer science

E-mail addresses: carpi@dipmat.unipg.it (A. Carpi), deluca@mat.uniroma1.it (A. de Luca).

0195-6698/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejc.2003.10.007

(computer graphics and pattern matching); the study of the structure and combinatorics of these words has become a subject of the greatest interest, with a large literature on it (see, for instance, the excellent overview by Berstel and Séébold [2]). For all definitions and notations not explicitly given in the paper, the reader is referred to [9].

Sturmian words can be defined in several different but equivalent ways. Some definitions are 'geometrical' and others of 'combinatorial' nature. A 'geometrical' definition is the following: a Sturmian word can be defined by considering the sequence of the cuts (*cutting sequence*) in a squared-lattice made by a ray having a slope which is an irrational number. A horizontal cut is denoted by the letter b, a vertical cut by a and a cut with a corner by ab or ba. Sturmian words represented by a ray starting from the origin are usually called *standard* or *characteristic*.

A combinatorial definition of Sturmian words can be given in terms of *subword complexity*. We recall that if w is an infinite word on the alphabet $\mathcal{A}$, then the subword complexity of w is the map $f_w : \mathbb{N} \to \mathbb{N}$, defined as: for each $n \geq 0$

$$f_w(n) = \mathrm{Card}(\mathrm{Fact}(w) \cap \mathcal{A}^n),$$

where $\mathrm{Fact}(w)$ is the set of all factors, or subwords, of w. In other terms for each n, $f_w(n)$ counts the number of factors of w of length n. Sturmian words are infinite words w whose subword complexity f_w is such that

$$f_w(n) = n + 1$$

for all $n \geq 0$. As is well known [2] this is also equivalent to saying that Sturmian words have the minimal possible value for subword complexity without being ultimately periodic. Moreover, since $f_w(1) = 2$ one has that these words are in a two letter alphabet. From now on, we shall take the alphabet $\mathcal{A}$ equal to the binary alphabet $\{a, b\}$.

The most famous Sturmian word is the *Fibonacci word*

$$f = abaababaabaababaababaabaababaabaab\ldots$$

which is the limit, according to a suitable topology, of the sequence of words $(f_n)_{n\geq 0}$, inductively defined as:

$$f_0 = b, \qquad f_1 = a, \qquad \text{and} \qquad f_{n+1} = f_n f_{n-1} \qquad \text{for } n \geq 1.$$

The words f_n of this sequence are called *finite Fibonacci words*. The name Fibonacci is due to the fact that for each n, the length $|f_n|$ of the word f_n is equal to the $(n+1)$th term of the Fibonacci series:

$$1, 1, 2, 3, 5, 8, 13, \ldots.$$

Standard Sturmian words can be defined in the following way which is a natural generalization of the definition of the Fibonacci word. Let $c_0, c_1, \ldots, c_n, \ldots$ be any sequence of integers such that $c_0 \geq 0$ and $c_i > 0$ for $i > 0$. We define, inductively, the sequence of words $(s_n)_{n\geq 0}$, where

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

The sequence $(s_n)_{n\geq 0}$ converges to a limit s which is an infinite standard Sturmian word. Any standard Sturmian word is obtained in this way. The sequence $(s_n)_{n\geq 0}$ is called the *approximating sequence* of s and $(c_0, c_1, c_2, \ldots)$ the *directive sequence* of s. The Fibonacci word f is the standard Sturmian word whose directive sequence is $(1, 1, \ldots, 1, \ldots)$. We shall denote by **Stand** the set of all infinite standard Sturmian words and by *Stand* the set of all the words s_n, $n \geq 0$ of any standard sequence $(s_n)_{n\geq 0}$. Any word of *Stand* is called a *finite standard word*, or a *generalized Fibonacci word*.

2. Finite standard words

Infinite standard Sturmian words are of great interest since [2] for any Sturmian word t there exists an infinite standard Sturmian word $s \in \mathbf{Stand}$ such that $\mathrm{Fact}(t) = \mathrm{Fact}(s)$. If one is interested in the study of the language *St* of the factors of all Sturmian words, one can limit oneself to consider only infinite standard Sturmian words. Indeed, one has:

$$St = \bigcup_{s \in \mathbf{Stand}} \mathrm{Fact}(s).$$

Since any element of **Stand** is the limit of a sequence of finite standard words, one has that any factor of a word of **Stand** is a factor of a word of *Stand*. Hence,

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

The set *Stand* has several characterizations based on quite different concepts [3, 4]. One of the characterizations is based on periods of words. More precisely, consider the set *PER* of all words w having two periods p, q which are coprime and such that $|w| = p + q - 2$. Thus a word w belongs to *PER* if it is a power of a single letter or is a word of maximal length for which the theorem of Fine and Wilf (cf. [9]) does not apply. In the sequel we assume that $\epsilon \in PER$ (this is, formally, coherent with the above definition if one takes $p = q = 1$). In [6] it has been proved that:

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

Thus, any element of *Stand* which is not a single letter can be obtained by appending ab or ba to a word of *PER*. Another characterization is based on palindromes. Let *PAL* be the set of all palindromes on $\mathcal{A}$. The set Σ is the subset of $\mathcal{A}^*$ defined as:

$$\Sigma = \mathcal{A} \cup (PAL^2 \cap PAL\{ab, ba\}).$$

Thus a word w belongs to Σ if and only if w is a single letter or satisfies the equation:

$$w = AB = Cxy,$$

where $A, B, C \in PAL$ and $\{x, y\} = \{a, b\}$. It was proved in [6] that

$$Stand = \Sigma.$$

From this one can easily derive the following remarkable structure result on the set *PER* [3]. For a word w we denote by $\mathrm{alph}(w)$ the set of the letters occurring in w.

Proposition 2.1. *Let w be a word such that* $\mathrm{Card}(\mathrm{alph}(w)) > 1$. *Then $w \in PER$ if and only if w can be uniquely represented as:*

$$w = PxyQ = QyxP,$$

with x, y fixed letters in $\{a, b\}$, $x \neq y$, $P, Q \in PAL$, and $|P| < |Q|$. Moreover, $p = |P| + 2$ and $q = |Q| + 2$ are periods of w such that p is the minimal period and $\gcd(p, q) = 1$.

Since any element of *Stand* is a prefix of an element of *PER* one has that

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

There exist several methods to generate standard words [2, 3, 5, 13]. In [3] the following method to generate the set *PER* was given. Let us introduce in $\mathcal{A}^*$ the map $(-): \mathcal{A}^* \to PAL$ which associates with any word $w \in \mathcal{A}^*$ the word $w^{(-)}$ defined as the shortest palindrome having the suffix w. We call $w^{(-)}$ the *palindromic left-closure of* w. If P is the longest palindromic prefix of $w = Pu$, then one has

$$w^{(-)} = u^{\sim} Pu,$$

where $u^{\sim}$ is the *reversal* of u, i.e., the sequence of the symbols of u taken in reverse order. If X is a subset of $\mathcal{A}^*$ we denote by $X^{(-)}$ the set

$$X^{(-)} = \{w^{(-)} \in \mathcal{A}^* \mid w \in X\}.$$

The following holds [3]:

Lemma 2.1. *Let $w \in PER$ and $x \in \{a, b\}$. Then $(xw)^{(-)} \in PER$. Moreover, if $w = PxyQ$, with $P, Q \in PAL$ and $\{x, y\} = \{a, b\}$, then one has*

$$(xw)^{(-)} = QyxPxyQ, \qquad (yw)^{(-)} = PxyQyxP.$$

Now, let us define the map

$$\psi: \mathcal{A}^* \to PER,$$

as follows: $\psi(\epsilon) = \epsilon$ and for all $v \in \mathcal{A}^*$, $x \in \mathcal{A}$,

$$\psi(vx) = (x\psi(v))^{(-)}.$$

The word v is called the *generating word* of $\psi(v)$. One has that for all $v, u \in \mathcal{A}^*$

$$\psi(vu) \in \mathcal{A}^*\psi(v) \cap \psi(v)\mathcal{A}^*. \tag{1}$$

For instance, in the case of the generating word $v = a^2ba$ one has $\psi(v) = aabaaabaa$.

Let E be the automorphism of $\mathcal{A}^*$ defined by $E(a) = b$, $E(b) = a$. For any $v \in \mathcal{A}^*$, one has $|E(v)| = |v|$ and, moreover, the set of periods of $E(v)$ is equal to the set of periods of v. Therefore, $v \in PER$ if and only if $E(v) \in PER$. One easily verifies that for all $v \in \mathcal{A}^*$

$$\psi(E(v)) = E(\psi(v)).$$

It has been proved in [3] that the map $\psi: \mathcal{A}^* \to PER$ is a bijection and that the restriction of ψ to $a\mathcal{A}^*$ is a bijection of $a\mathcal{A}^*$ onto $PER_a = PER \cap a\mathcal{A}^*$. Hence one has:

$$PER_a = \psi(a\mathcal{A}^*) = \bigcup_{n\geq 0} \psi(a\mathcal{A}^n).$$

Setting for each $n \geq 0$

$$X_n = \psi(a\mathcal{A}^n),$$

it follows

$$PER_a = \bigcup_{n\geq 0} X_n,$$

where $X_0 = \{a\}$ and for any $n > 0$ one has:

$$X_n = (\mathcal{A}X_{n-1})^{(-)}.$$

Let us observe that the set PER can be expressed as:

$$PER = \{\epsilon\} \cup PER_a \cup E(PER_a),$$

where $E(PER_a)$ is equal to PER_b, i.e., the set of elements of PER beginning with the letter b.

A different, and in some respects dual, way of generating the elements of *Stand* is obtained by introducing the *Fibonacci morphism* $F: \mathcal{A}^* \to \mathcal{A}^*$. defined as: $F(a) = ab$, $F(b) = a$. The following proposition holds [5]:

Proposition 2.2. *Stand is the smallest subset of $\mathcal{A}^*$ containing $\mathcal{A}$ and closed under the Fibonacci morphism F and the automorphism E.*

In the following we shall mainly refer to the set PER_a, i.e., the set of all elements of PER beginning with the letter a. One can introduce some suitable maps which are bijections of the set PER_a and the set $\mathcal{I}$ of all irreducible fractions p/q with $0 < p < q$. These representations of PER_a are of great interest and have remarkable applications. Moreover, they are related to each other in a very natural way.

The first map $\theta: PER_a \to \mathcal{I}$, that we call *ratio of periods*, is defined as follows. For all $n > 0$, we set

$$\theta(a^n) = \frac{1}{n+1}.$$

If $w \in PER_a$ and $\mathrm{Card}(\mathrm{alph}(w)) > 1$, then from Proposition 2.1, w can be uniquely factorized as:

$$w = QyxP = PxyQ$$

with $\{x, y\} = \{a, b\}$, $P, Q \in PAL$ and $|P| < |Q|$. If $w \in PER_a$ then we set:

$$\theta(w) = \frac{|P| + 2}{|Q| + 2} = \frac{p}{q}.$$

The meaning of $\theta(w)$ when $w \in PER_a$, is the ratio of the *minimal* period of w and the period q such that $\gcd(p, q) = 1$ and $|w| = p + q - 2$.

The second map $\zeta : PER_a \to \mathcal{I}$ is defined as follows: for any $w \in PER_a$

$$\zeta(w) = \frac{|w|_b + 1}{|w|_a + 1},$$

where $|w|_a$ and $|w|_b$ denote, respectively, the number of occurrences of the letters a and b in w. For any $w \in PER_a$, we call $\zeta(w)$ the *slope* of w.

One can prove that the maps ζ and θ are *bijections*. In order to see this and the relationships existing among them we recall some basic representations of binary words by irreducible fractions.

We introduce a suitable representation of the set $\mathcal{I}$ by continued fractions. Let $(\alpha_1, \dots, \alpha_n)$ be a sequence of $n > 0$ positive integers. We shall denote by

$$\langle \alpha_1, \dots, \alpha_n \rangle$$

the continued fraction $[0; \alpha_1, \dots, \alpha_{n-1}, \alpha_n + 1]$. It is trivial to verify that $\mathcal{I}$ is faithfully represented by the set of all such continued fractions $\langle \alpha_1, \dots, \alpha_n \rangle$.

Any word $v \in a\mathcal{A}^*$ can be uniquely represented as:

$$v = a^{\alpha_1} b^{\alpha_2} \cdots a^{\alpha_{n-1}} b^{\alpha_n},$$

where n is an even integer, $\alpha_i > 0$, $i = 1, \dots, n - 1$, and $\alpha_n \geq 0$. We call the sequence of integers $(\alpha_1, \dots, \alpha_m)$, where $m = n$ if $\alpha_n > 0$ and $m = n - 1$ otherwise, the *integral representation* of the word v. The following holds [1, 3]:

Proposition 2.3. *Let $(\alpha_1, \dots, \alpha_n)$ be the integral representation of a word $v \in a\mathcal{A}^*$ and $w = \psi(v) \in PER_a$. Then one has*

$$\zeta(w) = \langle \alpha_1, \dots, \alpha_n \rangle \qquad \textit{and} \qquad \theta(w) = \langle \alpha_n, \dots, \alpha_1 \rangle.$$

Example 2.1. Let v be the word $v = ab^2ab$ having the integral representation $(1, 2, 1, 1)$. One has:

$$\psi(v) = w = ababaababababaababa$$

and

$$\zeta(w) = \langle 1, 2, 1, 1 \rangle = \tfrac{8}{11} \qquad \text{and} \qquad \theta(w) = \langle 1, 1, 2, 1 \rangle = \tfrac{7}{12}.$$

The maps θ and ζ can be naturally extended to maps from PER to $\mathcal{I} \cup \{1/1\}$ by setting

$$\theta(\epsilon) = \zeta(\epsilon) = \tfrac{1}{1}$$

and for $w \in PER_b$,

$$\theta(w) = \theta(E(w)) \qquad \text{and} \qquad \zeta(w) = \zeta(E(w)).$$

We note that if $w \in PER_b$, then the slope of w is given by:

$$\zeta(w) = \frac{|w|_a + 1}{|w|_b + 1}.$$

Table 1
Values of $\zeta(w)$, for $w \in$ *Harm*, $\mathrm{alph}(w) = \mathcal{A}$, $|w| \leq 100$ (gold–harmonic fractions in bold)

$\zeta(w)$	$\|w\|$	$\zeta(w)$	$\|w\|$	$\zeta(w)$	$\|w\|$	$\zeta(w)$	$\|w\|$	$\zeta(w)$	$\|w\|$
2/3	3	10/23	31	23/30	51	**29/41**	68	13/72	83
3/5	6	13/21	32	21/34	53	35/37	70	28/59	85
3/7	8	6/29	33	27/29	54	**19/53**	70	43/45	86
5/7	10	**17/19**	34	15/41	54	17/55	70	23/65	86
5/8	11	6/31	35	**13/43**	54	27/46	71	21/67	86
4/11	13	14/25	37	20/37	55	**31/43**	72	34/55	87
7/9	14	19/21	38	**17/41**	56	26/49	73	**19/71**	88
4/13	15	**11/29**	38	**29/31**	58	37/39	74	27/64	89
9/11	18	9/31	38	**19/41**	58	34/43	75	45/47	90
8/13	19	9/32	39	11/49	58	25/53	76	32/61	91
11/13	22	**13/29**	40	11/50	59	39/41	78	39/56	93
7/17	22	21/23	42	8/55	61	31/49	78	47/49	94
5/19	22	19/26	43	31/33	62	9/71	78	31/65	94
7/18	23	23/25	46	18/47	63	9/73	80	**17/79**	94
5/21	24	**17/31**	46	14/51	63	**41/43**	82	22/75	95
13/15	26	**7/41**	46	8/57	63	29/55	82	10/89	97
12/17	27	**7/43**	48	**23/43**	64	**13/71**	82	49/51	98
11/19	28	16/35	49	33/35	66	38/47	83	10/91	99
15/17	30	25/27	50	22/47	67	16/69	83	35/67	100

3. Harmonic words

A word $w \in PER$ will be called *harmonic* if its slope is equal to the ratio of its periods, i.e.,

$$\zeta(w) = \theta(w).$$

We shall denote by *Harm* the set of harmonic words.

Example 3.1. The words $w_1 = abaaba$ and $w_2 = aabaaabaaabaa$ are harmonic. Indeed,

$$\zeta(w_1) = \frac{3}{5} = \theta(w_1) \qquad \text{and} \qquad \zeta(w_2) = \frac{4}{11} = \theta(w_2).$$

The word ϵ is harmonic since $\theta(\epsilon) = \zeta(\epsilon) = 1$. The word $w_3 = babbab$ is harmonic since $w_3 = E(w_1)$.

The value of $\zeta(w) = \theta(w)$ for the words $w \in Harm$ with $\mathrm{Card}(\mathrm{alph}(w)) > 1$ and $|w| \leq 100$ are reported in Table 1.

A word $v \in \mathcal{A}^*$ is called a *sesquipalindrome* if $v = E(v^\sim) = (E(v))^\sim$. For instance, the word $v = a^2b^3a^3b^2$ is a sesquipalindrome. We note that if v is a sesquipalindrome, then $|v|$ is an even integer.

Proposition 3.1. *A word of PER is harmonic if and only if its generating word is a palindrome or a sesquipalindrome.*

Proof. The result is trivial if the harmonic word w is ϵ. In such a case the generating word of w is ϵ which is a palindrome. Therefore, we assume that $w \neq \epsilon$. We first suppose that $w \in PER_a$. Let $w = \psi(v)$ where the generating word v has the integral representation $(\alpha_1, \ldots, \alpha_n)$. If $w \in Harm$, then $\zeta(w) = \theta(w)$ and by Proposition 2.3, $\langle \alpha_1, \ldots, \alpha_n \rangle = \langle \alpha_n, \ldots, \alpha_1 \rangle$, i.e.,

$$[0; \alpha_1, \ldots, \alpha_{n-1}, \alpha_n + 1] = [0; \alpha_n, \ldots, \alpha_2, \alpha_1 + 1].$$

Hence, one has for $1 \leq i \leq n$,

$$\alpha_i = \alpha_{n-i+1}. \tag{2}$$

If n is odd, then

$$v = a^{\alpha_1} b^{\alpha_2} \cdots a^{\alpha_n},$$

so that by Eq. (2), v is a palindrome. If n is even, then

$$v = a^{\alpha_1} b^{\alpha_2} \cdots a^{\alpha_{n-1}} b^{\alpha_n}$$

so that by Eq. (2), v is a sesquipalindrome.

Conversely, if $v \in a\mathcal{A}^*$ is a palindrome or a sesquipalindrome, then its integral representation $(\alpha_1, \ldots, \alpha_n)$ is such to satisfy Eq. (2). This implies that the word $w = \psi(v)$ is such that $\zeta(w) = \theta(w)$, i.e., w is harmonic.

If $w \in PER_b$, then $E(w) = \psi(E(v))$ with $E(v) \in a\mathcal{A}^*$. Therefore, since $E(w) \in PER_a$, from what we have proved before one has that $E(w) \in Harm$ if and only if its generating word $E(v)$ is a palindrome or a sesquipalindrome. This occurs if and only if v is a palindrome or a sesquipalindrome. Since $w \in Harm$ if and only if $E(w) \in Harm$, we conclude the proof. $\square$

As a consequence we derive the following:

Proposition 3.2. *Any word of PER can be extended on the right (on the left) to a harmonic word.*

Proof. Let $w \in PER$ and v be its generating word, i.e., $w = \psi(v)$. Let v' be any right extension of v to a palindrome or sesquipalindrome. One has $v' = v\lambda$ with $\lambda \in \mathcal{A}^*$. By Proposition 3.1

$$w' = \psi(v') \in Harm.$$

Moreover, by Eq. (1),

$$\psi(v') = \xi w = w\xi^{\sim}.$$

This proves our assertion. $\square$

Proposition 3.3. $St = \mathrm{Fact}(Harm)$.

Proof. From Proposition 3.2 one has that

$$PER \subseteq \mathrm{Fact}(Harm) \subseteq St.$$

Since $St = \mathrm{Fact}(PER)$ it follows that $St = \mathrm{Fact}(Harm)$. $\square$

Let us recall [3] that any $s \in \mathbf{Stand}$, can be generated as follows: let $\mathcal{A}^\omega$ be the set of all infinite words on the alphabet $\mathcal{A}$. One considers any infinite word $x \in \mathcal{A}^\omega \backslash \mathcal{A}^*\{a^\omega, b^\omega\}$, i.e., $x \neq vz^\omega$, with $v \in \mathcal{A}^*$ and $z \in \mathcal{A}$. Any such infinite word x can be uniquely written as:

$$x = b^{\alpha_0} a^{\alpha_1} b^{\alpha_2} \cdots b^{\alpha_{2n}} a^{\alpha_{2n+1}} \cdots$$

with $\alpha_0 \geq 0$ and $\alpha_i > 0$ for $i > 0$. Then one constructs recursively the sequence of words $(\sigma_n)_{n\geq 0}$, where $\sigma_0 = \epsilon$ and for all $n \geq 1$

$$\sigma_n = (x_n \sigma_{n-1})^{(-)} = \psi(x_1 \cdots x_n),$$

where x_n denotes, for all $n > 0$, the nth letter of x. One has that

$$s = \lim_{n\to\infty} \sigma_n.$$

We shall set $s = \psi(x)$ and call x the *generating word* of s.

It has been proved [6] that the set of all palindromic prefixes of s is equal to the set $\{\sigma_n \mid n \geq 0\}$.

Lemma 3.1. *Let* $x = x_1 x_2 \cdots x_n \cdots$ *be an infinite word with* $\mathrm{alph}(x) = \{a, b\}$. *The word* x *satisfies the condition:*

(C) *Any prefix of* x *is either a palindrome or a sesquipalindrome*

if and only if $x = (ab)^\omega$ *or* $x = (ba)^\omega$.

Proof. Without loss of generality we may suppose that $x_1 = a$. Let k be the minimal integer such that $x_k = b$. Hence, $x_1 \cdots x_k = a^{k-1}b$. From condition (C) one derives $k = 2$ and $x_1 x_2 = ab$. We prove now by induction that $x_i = a$ for all odd integers and $x_i = b$ for all even integers. We have already proved the base of the induction. Suppose the statement is true up to n and prove it for $n + 1$. If n is even, then $x_1 \cdots x_n = (ab)^{n/2}$, so that $x_1 \cdots x_n x_{n+1}$ satisfies (C) if and only if $x_{n+1} = a$. If n is odd, then $x_1 \cdots x_n = (ab)^{\lfloor n/2 \rfloor} a$, so that $x_1 \cdots x_n x_{n+1}$ satisfies (C) if and only if $x_{n+1} = b$. From this one derives that $x = (ab)^\omega$ which proves the assertion.

Conversely, it is trivial to verify that the words $(ab)^\omega$ and $(ba)^\omega$ satisfy condition (C). $\square$

Proposition 3.4. *An infinite standard Sturmian word is such that all palindromic prefixes are harmonic if and only if it is the Fibonacci word* f *or* $E(f)$.

Proof. Let s be an infinite standard Sturmian word having the generating infinite word x, so that $s = \psi(x)$ and $s = \lim_{n\to\infty} \sigma_n$, where $\sigma_0 = \epsilon$ and for all $n > 0$, $\sigma_n = \psi(x_1 \cdots x_n)$. The word s has all palindromic prefixes which are harmonic if and only if for all $n \geq 0$, $\sigma_n \in$ *Harm*. By Proposition 3.1 this latter condition is satisfied if and only if any prefix of x is either a palindrome or a sesquipalindrome. By Lemma 3.1 this occurs if and only if $x = (ab)^\omega$ or $x = (ba)^\omega$. In the first case $s = f$ and in the second case $s = E(f)$. $\square$

We say that an infinite standard Sturmian word $s \in \mathbf{Stand}$ is *harmonic* if s has infinitely many palindromic prefixes which are harmonic. From Proposition 3.4 the infinite standard

Sturmian words f and $E(f)$ are harmonic. We shall denote by **Harm** the set of all harmonic infinite standard Sturmian words.

Proposition 3.5. *An infinite standard Sturmian word is harmonic if and only if its generating word has infinitely many prefixes which are palindromes or sesquipalindromes.*

Proof. Let $s \in$ **Stand** and $x \in \mathcal{A}^\omega$ be its generating word, i.e., $s = \psi(x)$. Let us suppose that x has infinitely many prefixes

$$x[h_i] = x_1 \cdots x_{h_i}, i \geq 1,$$

with $h_1 < h_2 < \cdots < h_n < \cdots$, which are palindromes or sesquipalindromes. By Proposition 3.1 one has that for all $i \geq 1$

$$\psi(x[h_i]) \in Harm.$$

Since for all $i \geq 1$, $\psi(x[h_i])$ is a prefix of s it follows that s is an infinite harmonic standard Sturmian word.

Conversely, suppose that s is an infinite harmonic standard Sturmian word and let x be its generating word, so that $s = \psi(x)$ and $s = \lim_{n\to\infty} \sigma_n$, where $\sigma_0 = \epsilon$ and for all $n > 0$, $\sigma_n = \psi(x_1 \cdots x_n)$. By hypothesis there exist integers $k_1 < k_2 < \cdots < k_n < \cdots$ such that $s[k_i] \in Harm$. This implies that there exists a sequence of integers $h_1 < h_2 < \cdots < h_n < \cdots$ such that for all $i \geq 1$

$$s[k_i] = \sigma_{h_i} = \psi(x_1 \cdots x_{h_i}) = \psi(x[h_i]).$$

Since $s[k_i] \in Harm$ one obtains by Proposition 3.1 that its generating word $x[h_i]$ is either a palindrome or a sesquipalindrome. Thus x has infinitely many prefixes which are palindromes or sesquipalindromes which concludes the proof. □

From the preceding proposition one derives that if an infinite standard Sturmian word t has a generating word which is an infinite standard Sturmian word, then t is harmonic.

Corollary 3.1. *If $s \in$* **Stand**, *then $\psi(s) \in$* **Harm**.

Proof. It is sufficient to observe that any infinite standard Sturmian word has infinitely many palindromic prefixes, so that the result follows from Proposition 3.5. □

Corollary 3.2. *There exist continuously many infinite harmonic words.*

Proof. As is well known, there exist continuously many infinite standard Sturmian words. Indeed, with each positive irrational number α one can injectively associate the standard Sturmian word having slope α. Since the map ψ is injective the result follows from the preceding corollary. □

Corollary 3.3. *There exist infinite standard Sturmian words which are not harmonic.*

Proof. It is sufficient to consider an infinite standard Sturmian word s having a generating word of the kind $x = ab^2a^3b^4a^5 \cdots b^{2n}a^{2n+1} \cdots$ or $x' = a(ab)^\omega$. These words have only finitely many prefixes which are palindromes or sesquipalindromes, so that by Proposition 3.5 it follows that $s = \psi(x)$ and $s' = \psi(x')$ are not harmonic. □

Let μ be the Thue-Morse endomorphism of $\mathcal{A}^*$ defined by $\mu(a) = ab$ and $\mu(b) = ba$.

Lemma 3.2. *If $w \in \mathcal{A}^+$ is a palindrome (resp. a sesquipalindrome), then $\mu(w)$ is a sesquipalindrome (resp. a palindrome).*

Proof. The proof is by induction on the length of w. If the length of w is equal to 1 or 2 the result is trivially verified. Let us then suppose that $|w| > 2$. We first suppose that w is a palindrome. We can write $w = xux$, with $x \in \mathcal{A}$ and u a palindrome. One has $\mu(w) = \mu(x)\mu(u)\mu(x)$. By induction the word $\mu(u)$ is a sesquipalindrome. If $x = a$, one obtains $\mu(w) = ab\mu(u)ab$ which is a sesquipalindrome. If $x = b$, one obtains $\mu(w) = ba\mu(u)ba$ which is a sesquipalindrome.

In a similar way if w is a sesquipalindrome, then we can write $w = xuy$, with $x \neq y$ and u a sesquipalindrome. By induction one easily derives that $\mu(w)$ is a palindrome. $\square$

Proposition 3.6. *The infinite standard Sturmian word having as generating word the Thue-Morse word in two symbols is harmonic.*

Proof. As is well known [9] the Thue-Morse word t in two symbols can be defined as

$$t = \lim_{n\to\infty} \mu^n(a),$$

where $\mu^0(a) = a$ and for all $n > 0$, $\mu^n(a) = \mu(\mu^{n-1}(a))$. From the preceding lemma it follows that for all $n \geq 0$, $\mu^{2n}(a)$ is a palindrome, whereas $\mu^{2n+1}(a)$ is a sesquipalindrome. From Proposition 3.5 it follows that the infinite standard Sturmian word $\psi(t)$ is harmonic. $\square$

Let F be the Fibonacci morphism and D the morphism $D = E \circ F$, so that $D(a) = a$ and $D(b) = ab$. One can easily prove (cf. [5]) that if $w \in PER_a$, then $F(w)a, D(w)a \in PER_a$. We shall denote by $Harm_a$ the set of harmonic words beginning with the letter a.

Proposition 3.7. *Let $w \in Harm_a$. One has that $F(w)a \in Harm_a$ if and only if w is a palindromic prefix of the Fibonacci word f. Moreover, $D(w)a \notin Harm_a$.*

Proof. Let $(\alpha_1, \ldots, \alpha_n)$ be the integral representation of the generating word of w. Since $w \in Harm_a$, one has $\alpha_i = \alpha_{n-i+1}$ for all $i = 1, \ldots, n$. The slope of $F(w)a$ is given by

$$\zeta(F(w)a) = \frac{|F(w)a|_b + 1}{|F(w)a|_a + 1} = \frac{|w|_a + 1}{|w|_a + |w|_b + 2} = \frac{1}{1 + \zeta(w)} = \langle 1, \alpha_1, \ldots, \alpha_n \rangle.$$

By Proposition 2.3 one has that $\theta(F(w)a) = \langle \alpha_n, \ldots, \alpha_1, 1 \rangle$. Therefore, $F(w)a$ is in $Harm_a$ if and only if $\alpha_n = 1$ and for all $i = 1, \ldots, n-1$ one has $\alpha_i = \alpha_{n-i}$. One derives that $F(w)a$ is in $Harm_a$ if and only if for all $i = 1, \ldots, n$ one has $\alpha_i = 1$, which concludes the proof of the first assertion.

The slope of $D(w)a$ is given by:

$$\zeta(D(w)a) = \frac{|D(w)a|_b + 1}{|D(w)a|_a + 1} = \frac{|w|_b + 1}{|w|_a + |w|_b + 2} = \frac{\zeta(w)}{1 + \zeta(w)} = \langle 1 + \alpha_1, \alpha_2, \ldots, \alpha_n \rangle,$$

whereas by Proposition 2.3 one has $D(F(w)a) = \langle \alpha_n, \ldots, \alpha_2, \alpha_1 + 1 \rangle$. If one imposes that $D(F(w))a$ is harmonic one obtains $1 + \alpha_1 = \alpha_n$. Since $\alpha_1 = \alpha_n$ one reaches a contradiction. $\square$

4. Combinatorics of harmonic words

In this section we give a combinatorial characterization of harmonic words which allows us to obtain a formula counting for any integer n the number of harmonic words of length n.

We say that a rational number $\alpha > 1$ has a *symmetric development in continued fractions* if $\alpha = [q_0; q_1, \ldots, q_n]$ with $q_i = q_{n-i}$ for $i = 0, \ldots, n$.

Lemma 4.1. *A word $w \in PER$ is harmonic if and only if $1 + 1/\zeta(w)$ (or $1 + 1/\theta(w)$) has a symmetric development in continued fractions.*

Proof. If $\zeta(w) = \langle \alpha_1, \ldots, \alpha_{n-1}, \alpha_n \rangle$, then

$$1 + \frac{1}{\zeta(w)} = [\alpha_1 + 1; \alpha_2, \ldots, \alpha_{n-1}, \alpha_n + 1] \tag{3}$$

and

$$1 + \frac{1}{\theta(w)} = [\alpha_n + 1; \alpha_{n-1}, \ldots, \alpha_2, \alpha_1 + 1].$$

If $w \in Harm$, then by Eq. (2) one has $\alpha_i = \alpha_{n-i+1}$ for $1 \leq i \leq n$ so that $1 + 1/\zeta(w)$, as well as $1 + 1/\theta(w)$, have a symmetric development in continued fractions. Conversely, suppose that $1 + 1/\zeta(w)$ has a symmetric development in continued fractions. Necessarily, this is given by Eq. (3) with $\alpha_i = \alpha_{n-i+1}$ for $1 \leq i \leq n$. By Proposition 2.3 one derives $\zeta(w) = \theta(w)$ so that $w \in Harm$. Similarly, one reaches the same result if one supposes that $1 + 1/\theta(w)$ has a symmetric development in continued fractions. □

Proposition 4.1. *A word $w \in PER$ is harmonic if and only if*

$$p^2 \equiv \pm 1 \pmod{|w| + 2},$$

where p is the minimal period of w.

Proof. Let $w \in PER$ be such that $|w| = p + q - 2$, where p is the minimal period of w. We observe that

$$1 + \frac{1}{\theta(w)} = \frac{p+q}{p} = \frac{|w| + 2}{p}.$$

From the previous proposition $w \in Harm$ if and only if $(|w| + 2)/p$ has a symmetric development in continued fractions. By a classical result (cf. [10, Chapter 24]) this occurs if and only if $p + q = |w| + 2$ divides $p^2 \pm 1$, i.e., $p^2 \equiv \pm 1 \pmod{|w| + 2}$. □

Proposition 4.2. *The number of harmonic words of length $n \geq 0$ is equal to the number of roots of the equation $x^2 \equiv \pm 1 \pmod{n + 2}$.*

Proof. The statement is trivially true in the case $n = 0$, since there exists a unique harmonic word of length 0, namely the empty word, and a unique root of the equation $x^2 \equiv \pm 1 \pmod 2$. For any $n > 0$, we consider the sets

$$H_n = Harm_a \cap A^n$$

and

$$K_n = \left\{ p \in \mathbb{N} \;\middle|\; 0 \le p < \frac{n+2}{2} \qquad \text{and} \qquad p^2 \equiv \pm 1 \pmod{n+2} \right\}.$$

We prove that there exists a bijection $\lambda_n : H_n \to K_n$. For any $w \in Harm_a \cap A^n$, let $\lambda_n(w)$ be the minimal period p of w. We know that $|w| = p + q - 2$ with $q > p$. Therefore, $p < (n+2)/2$. Since w is harmonic, by Proposition 4.1, $p^2 \equiv \pm 1 \pmod{n+2}$. Thus $\lambda_n(w) \in K_n$. The map λ_n is injective. Indeed, if $v \in Harm_a \cap A^n$ and $\lambda_n(v) = \lambda_n(w) = p$, then $\theta(v) = \theta(w) = p/q$ so that $v = w$. Now we verify that λ_n is surjective. Indeed, let $p \in K_n$ and set $n = p+q-2$. Since $p^2 \equiv \pm 1 \pmod{n+2}$, one derives that $\gcd(p, q) = 1$. Moreover, since $p < (n+2)/2$, one has $p < q$. Hence there exists a unique element w of $PER_a \cap A^n$ having periods p and q. As $p^2 \equiv \pm 1 \pmod{n+2}$ by Proposition 4.1 it follows that $w \in Harm_a$ and $\lambda_n(w) = p$. Since λ_n is bijective, it follows $\mathrm{Card}(H_n) = \mathrm{Card}(K_n)$.

Now, we prove that $\mathrm{Card}(K_n) = \mathrm{Card}(K'_n)$, where

$$K'_n = \left\{ p \in \mathbb{N} \;\middle|\; \frac{n+2}{2} \le p < n+2 \qquad \text{and} \qquad p^2 \equiv \pm 1 \pmod{n+2} \right\}.$$

The map defined on K_n as $\varrho_n(p) = n + 2 - p$ is a bijection onto K'_n. Indeed, since $(\varrho_n(p))^2 \equiv p^2 \equiv \pm 1 \pmod{n+2}$ and $(n+2)/2 \le \varrho_n(p) < n+2$, one has $\varrho_n(p) \in K'_n$. The map $\varrho_n : K_n \to K'_n$ is trivially injective. Let us prove that it is surjective. Let $q \in K'_n$ and suppose that n is odd or $q \ne (n+2)/2$. If we set $p = n+2-q$, then $0 \le p < (n+2)/2$ and $p^2 \equiv q^2 \equiv \pm 1 \pmod{n+2}$, so that $p \in K_n$ and $q = \varrho_n(p)$. Now we suppose that n is even and $q = (n+2)/2$. Since $2q \equiv 0 \pmod{n+2}$ it follows $2q^2 \equiv 0 \pmod{n+2}$ so that $q^2 \not\equiv \pm 1 \pmod{n+2}$, i.e., $q \notin K'_n$.

Since $\mathrm{Card}(Harm_a \cap A^n) = \mathrm{Card}(Harm_b \cap A^n)$, one derives

$$\mathrm{Card}(Harm \cap A^n) = 2\,\mathrm{Card}(Harm_a \cap A^n) = \mathrm{Card}(K_n \cup K'_n),$$

which proves the assertion. □

For any $n \ge 0$ the set of roots of the equation $x^2 \equiv \pm 1 \pmod{n+2}$ is a finite 2-group G since, for any root x, $x^4 \equiv 1 \pmod{n+2}$ so that any element of G has an order which is a power of two. As is well known from group theory, the order of G is a power of 2. Hence, from Proposition 4.2 one derives that the number of harmonic words of length n is a power of 2. The following theorem gives for any $n > 0$ the exact number of harmonic words of length n.

Theorem 4.1. *Let n be a positive integer and factorize $n + 2$ as:*

$$n + 2 = 2^{\alpha} p_1^{\beta_1} \cdots p_r^{\beta_r}$$

where $p_1, \ldots, p_r$, $r \ge 0$, are distinct odd primes, $\alpha \ge 0$, and $\beta_i > 0$, $1 \le i \le r$. The number of harmonic words of length n is given by

- 2^{r+1} *if $\alpha \le 1$ and for $1 \le i \le r$, $p_i \equiv 1 \pmod 4$,*
- 2^{r} *if $\alpha \le 1$ and there exists i, $1 \le i \le r$ such that $p_i \equiv 3 \pmod 4$,*
- 2^{r+1} *if $\alpha = 2$,*
- 2^{r+2} *if $\alpha > 2$.*

Proof. By the previous proposition, we have to count for each n the number of roots of the equation $x^2 \equiv 1 \pmod{n+2}$ and of the equation $x^2 \equiv -1 \pmod{n+2}$. From the theory of congruential equations (cf. [7, 11]) the number of roots of the equations above is equal, respectively, to the product of the numbers of roots of the equations

$$x^2 \equiv 1 \pmod{2^\alpha}, \qquad x^2 \equiv 1 \pmod{p_i}, \qquad i = 1, \ldots, r,$$

and

$$x^2 \equiv -1 \pmod{2^\alpha}, \qquad x^2 \equiv -1 \pmod{p_i}, \qquad i = 1, \ldots, r.$$

Now, the number of roots of the equation $x^2 \equiv 1 \pmod{2^\alpha}$ is 1 if $\alpha \le 1$, 2 if $\alpha = 2$, and 4 for all $\alpha > 2$. The number of roots of the equation $x^2 \equiv 1 \pmod{p_i}$ is exactly 2 for any $i = 1, \ldots, r$. The number of roots of the equation $x^2 \equiv -1 \pmod{2^\alpha}$ is 1 if $\alpha \le 1$, and 0 for all $\alpha > 1$. Finally, for any $i = 1, \ldots, r$ the number of roots of the equation $x^2 \equiv -1 \pmod{p_i}$ is 2 if $p_i \equiv 1 \pmod 4$ and 0 otherwise (i.e., $p_i \equiv 3 \pmod 4$). From this, one derives the statement. □

Example 4.1. Let $n = 8$. Since $n + 2 = 10 = 2 \cdot 5$ and $5 \equiv 1 \pmod 4$, there are 4 harmonic words of length 8, namely,

$$a^8, \qquad b^8, \qquad (a^2b)^2a^2, \qquad (b^2a)^2b^2.$$

From Theorem 4.1 one has in particular:

Corollary 4.1. *If $n = p^\beta - 2$ with $\beta > 0$ and p odd prime, then there are 4 harmonic words if $p \equiv 1 \pmod 4$ and 2 harmonic words otherwise.*

A consequence of the previous corollary is that there are infinitely many integers n such that $\mathrm{Card}(Harm \cap A^n) = 2$.

Proposition 4.3. *For any $p > 0$ there exists at least one harmonic word having minimal period p.*

Proof. If $p \le 2$ the result is trivial, as, for instance, the words a and aba are harmonic words having minimal periods 1 and 2, respectively. Thus we suppose $p > 2$ and take $q = p^2 - p \pm 1$. One has $q > p$ and since $p^2 \equiv \pm 1 \pmod{p+q}$ a word $w \in PER$ such that $\theta(w) = p/q$ is harmonic by Proposition 4.1. □

Proposition 4.4. *Let w be a harmonic word of minimal period $p > 1$. Then*

$$|w| < p^2.$$

Moreover, the upper bound is optimal.

Proof. By Proposition 4.1 one has $p^2 \equiv \pm 1 \pmod{|w|+2}$ so that $|w| + 2$ divides $p^2 \pm 1$. Since $p > 1$, $|w| + 2 \le p^2 + 1$, i.e., $|w| < p^2$.

This bound is optimal since a word $w \in PER$ such that $\theta(w) = p/(p^2 - p + 1)$ is harmonic and such that $|w| = p^2 - 1$. □

5. Gold words

A word $w \in PER$ will be called a *gold word* if the ratio of its periods $\theta(w) = p/q$ is an irreducible fraction with p and q prime numbers. The set of all gold words will be denoted by *Gold*.

For instance, the words $w_1 = aabaabaa$ and $w_2 = aaaaabaaaaaabaaaaa$ are gold since $\theta(w_1) = 3/7$ and $\theta(w_2) = 7/13$, whereas $w_3 = aaabaaaabaaa$ is not gold since $\theta(w_3) = 5/9$.

By the definition if $g \in Gold$, then $E(g) \in Gold$. Moreover, the length $n = p + q - 2$ of a gold word is either an odd prime or an even integer. Indeed, if n is an odd integer, then the minimal period p of g has to be equal to 2 and n has to be equal to q which is an odd prime. In such a case there exist two gold words $g \in PER_a$ and $E(g)$ such that $|g| = n$ and $\theta(g) = 2/n$. The generating word of g is

$$ab^{\lfloor n/2 \rfloor}$$

and

$$g = (ab)^{\lfloor n/2 \rfloor} a.$$

If n is an even integer, the periods p and q of g have to be odd primes with $p \neq q$. As a consequence there do not exist gold words for $n = 2$ and $n = 4$. One can easily verify by a computer that there exist gold words for all even lengths $n > 4$ and less than very large integers. Therefore, it is natural to set the following:

Conjecture 5.1. *For any even integer $n \geq 6$ there exists a gold word of length n.*

We remark that this conjecture is equivalent to the statement that every even integer n greater than 6 is the sum of two different odd primes. This is a statement stronger than the famous Goldbach's conjecture (see, for instance, [7]) stating that any even integer $\geq$6 is the sum of two odd primes. Hence, if Conjecture 5.1 is true also Goldbach's conjecture will be true.

Now, we shall prove that any element of *PER* can be extended on the right (left) to a gold word. To this end, we recall the following proposition proved in [3]:

Proposition 5.1. *Let w be an element of PER having the generating word v and ratio of periods p/q. If x is the last letter of v and $y \neq x$ is the other letter of $\mathcal{A}$, one has*

$$\theta(\psi(vx^k)) = \frac{p}{q + kp} \qquad \textit{for all } k \geq 0$$

and

$$\theta(\psi(vy^k)) = \frac{q}{p + kq} \qquad \textit{for all } k > 0.$$

Proof. We suppose without loss of generality that v begins with the letter a. We can write $v = a^{\alpha_1} b^{\alpha_2} \cdots x^{\alpha_n}$, where $x = a$ if n is odd and $x = b$ otherwise. Let $\theta(\psi(v)) = p/q = \langle \alpha_n, \ldots, \alpha_1 \rangle$. By Proposition 2.3 one has that for all $k > 0$,

$$\theta(\psi(vy^k)) = \langle k, \alpha_n, \ldots, \alpha_1 \rangle = \frac{1}{k + \frac{p}{q}}.$$

In a similar way one has that for all $k \geq 0$,

$$\theta(\psi(vx^k)) = \langle k + \alpha_n, \alpha_{n-1}, \ldots, \alpha_1 \rangle = \frac{1}{k + \frac{q}{p}},$$

which concludes the proof. □

Proposition 5.2. *Any element w of PER can be extended on the right (left) to a gold word g such that $|g| > |w|$. Moreover, if v is the generating word of w, then there exist arbitrarily large positive integers k_0, h_0 such that $va^{k_0}b^{h_0}$ is the generating word of g.*

Proof. Let w be an element of *PER* having the generating word v and ratio of periods p/q. First we suppose that the last letter of v is a. Since $\gcd(p, q) = 1$, by the famous theorem of Dirichlet on primes in arithmetical progressions (see, for instance, [12]) one has that there exist infinitely many positive integers k such that $q + kp$ is a prime number. Let k_0 be an integer such that $q + k_0 p = \pi_1$ is a prime number. By the preceding proposition

$$\theta(\psi(va^{k_0})) = \frac{p}{\pi_1}.$$

Since $\gcd(p, \pi_1) = 1$ by using again Dirichlet's theorem it follows that there exist infinitely many positive integers h such that $p + h\pi_1$ is a prime number. Let h_0 be an integer such that $p + h_0\pi_1$ is a prime number π_2. By using again Proposition 5.1 one derives

$$\theta(\psi(va^{k_0}b^{h_0})) = \frac{\pi_1}{\pi_2}.$$

Therefore, the word $g = \psi(va^{k_0}b^{h_0})$ is gold. Moreover,

$$g = \psi(v)\xi = w\xi = \xi^{\sim}w$$

with $\xi \neq \epsilon$ which proves our assertion in this case.

Now suppose that the last letter of v is b, and set $v' = va$. By the previous argument, there are arbitrarily large positive integers k_0', h_0' such that $\psi(v'a^{k_0'}b^{h_0'}) = \psi(va^{1+k_0'}b^{h_0'})$ is gold, concluding the proof. □

Example 5.1. The word $w = (a^2ba)^4a \in PER$ has the generating word $v = a^2ba^3$ and ratio of periods $4/15$. It can be extended on the right (left) to the word $g = (a^2ba)^5(aba^2)^5 = waba(aba^2)^5 = (a^2ba)^5abaw$ which is gold since $\theta(g) = 19/23$. The generating word of g is a^2ba^4b. In this case $k_0 = h_0 = 1$.

We remark that following the same argument of the proof of Proposition 5.2, if $w \in$ *PER* has a minimal period p which is prime, then w can be extended on the right (left) to a gold word having the same minimal period.

From Proposition 5.2 one derives:

Proposition 5.3. $St = \mathrm{Fact}(Gold)$.

Proof. From Proposition 5.2 one has that

$$PER \subseteq \mathrm{Fact}(Gold) \subseteq St.$$

Since $St = \mathrm{Fact}(PER)$ it follows that $St = \mathrm{Fact}(Gold)$. $\square$

An infinite standard Sturmian word will be called *gold* if it has infinitely many prefixes which are gold. We shall denote by **Gold** the class of all infinite gold standard Sturmian words. The following lemma will be useful in the sequel.

Lemma 5.1. *Let* $u = x^{\alpha_1} y^{\alpha_2} x^{\alpha_3} y^{\alpha_4} \cdots$ *with* $\{x, y\} = \{a, b\}$ *and* $\alpha_i > 0, i = 1, \ldots, n, \ldots$ *be the generating word of an infinite standard Sturmian word. Let* $(q_k)_{k\geq -1}$ *be the sequence of integers defined as:*

$$q_{-1} = 0, \quad q_0 = 1, \quad q_1 = \alpha_1 + 1, \quad q_k = \alpha_k q_{k-1} + q_{k-2} \quad \text{for } k > 1. \tag{4}$$

For any $n > 0$ *let* v_n *be the word* $v_n = x^{\alpha_1} y^{\alpha_2} \cdots z^{\alpha_n}$ *where* $z = x$ *if* n *is odd and* $z = y$, *otherwise. Then the ratio of periods of* $\psi(v_n)$ *is given by:*

$$\theta(\psi(v_n)) = \frac{q_{n-1}}{q_n}.$$

Proof. By Proposition 2.3 one has that

$$\theta(\psi(v_n)) = \langle \alpha_n, \ldots, \alpha_1 \rangle = [0; \alpha_n, \ldots, \alpha_2, \alpha_1 + 1].$$

From the theory of continued fractions [8], one has that this continued fraction is equal to q_{n-1}/q_n where for any $n > 0$, q_n is the denominator of the nth convergent p_n/q_n of the continued fraction $[0; \alpha_1 + 1, \alpha_2, \ldots, \alpha_n, \ldots]$. $\square$

Proposition 5.4. *The class* **Gold** *contains continuously many elements.*

Proof. Let us consider the sequence of integers as defined in Eq. (4), i.e., $q_1 = 0$, $q_0 = 1, q_1 = \alpha_1 + 1, q_k = \alpha_k q_{k-1} + q_{k-2}$ for $k > 1$. There exist infinitely many values of α_1 for which q_1 is prime. Moreover, since for all $k > 0$, $\gcd(q_k, q_{k-1}) = 1$, by Dirichlet's theorem one has that there exist infinitely many values of α_k, $k > 1$, for which $q_k = \alpha_k q_{k-1} + q_{k-2}$, is a prime number. For any such choice consider the word

$$u = a^{\alpha_1} b^{\alpha_2} a^{\alpha_3} b^{\alpha_4} \cdots$$

which generates an infinite standard Sturmian word s. By Lemma 5.1 for any $n > 0$ the prefix $v_n = a^{\alpha_1} b^{\alpha_2} \cdots z^{\alpha_n}$ of u, where $z = a$ if n is odd and $z = b$ otherwise, generates the element $\psi(v_n)$ of PER such that

$$\theta(\psi(v_n)) = \frac{q_{n-1}}{q_n}.$$

Therefore, $\psi(v_n) \in Gold$ and $s \in$ **Gold**. Since there exist continuously many such generating infinite words u, the result follows. $\square$

Proposition 5.5. *There exist infinite standard Sturmian words which have no gold palindromic prefixes.*

Proof. Consider the sequence of integers as defined in Eq. (4). One can always choose the integers $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ in such a way that for any $k > 0$, the integer q_k is a composite number. Now, without loss of generality, consider the infinite word

$$u = a^{\alpha_1} b^{\alpha_2} a^{\alpha_3} b^{\alpha_4} \cdots$$

and take any prefix t of u. We have to consider two cases:

Case 1. $|t| \leq |\alpha_1|$. One has $\theta(\psi(t)) = 1/(|t|+1)$. Therefore, $\psi(t) \notin Gold$.

Case 2. $|t| > |\alpha_1|$. There exists an integer $n \geq 1$ for which

$$t = a^{\alpha_1} b^{\alpha_2} a^{\alpha_3} b^{\alpha_4} \cdots x^{\alpha_n} y^k,$$

where $k > 0$, $y \neq x$, and $x = a$ if n is odd, $x = b$ otherwise. One has by Lemma 5.1

$$\theta(\psi(t)) = \langle k, \alpha_n, \ldots, \alpha_1 \rangle = \frac{1}{k + \frac{q_{n-1}}{q_n}} = \frac{q_n}{kq_n + q_{n-1}}.$$

Since for all $n > 0$ the integers q_n are composite, one derives $\psi(t) \notin Gold$ which concludes the proof. □

Example 5.2. Let u be the infinite word

$$u = a^3 b^2 (a^2 b^3)^\omega.$$

According to Eq. (4) one has $q_1 = 4$, $q_2 = 9$ and for $n > 2$, $q_n = 2q_{n-1} + q_{n-2}$ if n is odd and $q_n = 3q_{n-1} + q_{n-2}$ if n is even. One easily derives that q_n is a multiple of 2 if n is odd and a multiple of 3 if n is even. Therefore, the word $\psi(u)$ has no gold palindromic prefix.

The following proposition shows, in particular, that the Fibonacci word f is not gold.

Proposition 5.6. *The only gold prefixes of the Fibonacci word f are aba and $abaaba$.*

Proof. Let $(F_n)_{n\geq 1}$ be the Fibonacci series. As is well known, u is a palindromic prefix of f, if and only if $|u| = F_n - 2$ for a suitable $n \geq 3$. In this case, $\theta(u) = F_{n-2}/F_{n-1}$.

We recall [7] that if $n \neq 4$ is a composite number, then F_n is composite. Hence, for $n > 6$ at least one of the two integers F_{n-2}, F_{n-1} has to be composite. Thus the only cases where F_{n-2} and F_{n-1} are both primes are when $n = 5$ or $n = 6$. The corresponding palindromic prefixes of f are aba and $abaaba$, respectively, which concludes the proof. □

Now, we consider the class **Gold**∩**Harm**, i.e., the class of all infinite standard Sturmian words having infinitely many gold prefixes as well as infinitely many harmonic prefixes.

Proposition 5.7. *The class* **Gold** ∩ **Harm** *contains continuously many elements.*

Proof. We shall prove that there exist continuously many infinite words which generate elements of **Gold** ∩ **Harm**. In fact, each of these words will have infinitely many palindromic prefixes which generate harmonic words as well as infinitely many prefixes generating gold words.

Table 2
Gold–harmonic fractions p/q with $p + q - 2 \leq 500$

2/3	29/31	23/109	107/109	41/239	179/181	107/317
3/5	23/43	43/89	89/131	17/271	61/311	109/331
3/7	29/41	17/127	73/149	97/197	67/307	199/241
5/7	19/53	71/73	97/127	73/223	151/229	199/251
5/19	31/43	53/103	37/191	101/199	127/257	37/419
7/17	13/71	29/139	59/173	149/151	191/193	227/229
11/13	41/43	41/127	53/181	113/191	79/311	43/419
11/19	19/71	71/97	41/199	79/233	89/307	157/317
17/19	17/79	13/157	89/151	131/181	197/199	31/449
11/29	23/83	71/109	83/163	107/211	101/307	79/401
13/29	41/71	47/137	71/181	89/241	137/271	239/241
7/41	11/109	89/109	79/181	131/199	73/337	83/409
17/31	19/101	67/137	23/241	97/239	79/337	167/331
7/43	31/89	101/103	67/197	113/223	41/379	
13/43	41/79	29/181	47/229	101/239	71/349	
17/41	59/61	71/139	137/139	89/271	139/281	
19/41	47/83	53/163	29/251	109/251	181/239	

Let us consider any sequence $(v_n)_{n\geq 0}$ of finite words where $v_0 = b$ and, for $n > 0$

$$v_n = v_{n-1}a^{k_n}b^{h_n}a^{k_n}\tilde{v}_{n-1},$$

where (k_n, h_n) is an arbitrary pair of positive integers such that $v_{n-1}a^{k_n}b^{h_n}$ is the generating word of a gold word. By Proposition 5.2 there exist infinitely many such pairs.

The sequence $(v_n)_{n\geq 0}$ has a limit v which is the generating word of an infinite standard Sturmian word $\psi(v) \in$ **Gold** $\cap$ **Harm** since for any $n > 0$, v_n is a palindrome and $\psi(v_{n-1}a^{k_n}b^{h_n}) \in Gold$.

As different choices of the values of k_n or h_n, $n \geq 0$, lead to different generating words and ψ is an injective map, the result follows. $\square$

It is interesting to consider the class $Gold \cap Harm$, i.e., the class of all finite words which are both harmonic and gold. A word $w \in PER$ belongs to $Gold \cap Harm$ if $\theta(w) = \zeta(w) = p/q$ with p and q primes.

Example 5.3. The words $w_1 = aabaabaa$ and $w_2 = a^3b(a^4b)^3a^3$ belong to the class $Gold \cap Harm$ as $\theta(w_1) = \zeta(w_1) = 3/7$ and $\theta(w_2) = \zeta(w_2) = 5/19$.

An irreducible fraction p/q with $p < q$ will be called *gold–harmonic* if p and q are both primes and $p^2 \equiv \pm 1 \pmod{p+q}$. By Proposition 4.1 it is clear that there exists a bijection of the set $Gold \cap Harm_a$ and the set of gold–harmonic fractions. Table 2 gives the set of all gold–harmonic fractions p/q with $p + q - 2 \leq 500$.

We observe that there exist only 2 gold–harmonic words having an odd length, namely aba and bab. Indeed, a gold–harmonic word having an odd length has minimal period $p = 2$ so that, by Proposition 4.4, $|w| \leq 3$ and from this the assertion follows.

Let us state the following:

Conjecture 5.2. *The set $Gold \cap Harm$ is infinite.*

The preceding conjecture is trivially equivalent to the statement that the set of gold–harmonic fractions is infinite. Let us remark that if p and q are *twin* odd primes, i.e., p and q are odd primes such that $q = p + 2$, then the fraction p/q is gold–harmonic as $p^2 \equiv 1 \pmod{2(p+1)}$. In such a case one has

$$\frac{p}{q} = \left\langle 1, \frac{p-1}{2}, 1 \right\rangle.$$

The corresponding word w of $Gold \cap Harm$, beginning with the letter a, has the generating word

$$ab^{\frac{p-1}{2}}a$$

so that $w = ((ab)^{(p-1)/2}a)^2$. It follows that if there exist infinitely many twin primes (and this is a classic conjecture of number theory), then Conjecture 5.2 has a positive answer. However, by inspecting Table 2, it is noteworthy that the great majority of gold–harmonic fractions consists of fractions p/q with p and q primes such that $q > p + 2$, at least when $p + q - 2 \leq 500$.

By Proposition 4.1, if p and q are primes such that $q = p^2 - p \pm 1$, then the fraction p/q is gold–harmonic. Therefore, if there exist infinitely many pairs of primes of the form $(p, p^2 - p \pm 1)$, then Conjecture 5.2 has a positive answer.

Finally, we remark that, as one can easily verify by a computer, for any odd prime $p < 2693$ there exists a prime $q > p$ such that the fraction p/q is gold–harmonic. However, the prime $p = 2693$ is the least odd prime such that this property is not verified.

Acknowledgements

The authors are indebted to Prof. Jean Berstel for some stimulating discussions and suggestions. The work for this paper has been supported by the Italian Ministry of Education under Project COFIN 2001 - Linguaggi Formali e Automi: teoria ed applicazioni. An extended abstract of this paper was presented at the conference WORDS' 03, held in Turku (FI) on September 10–13, 2003.

References

[1] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees, Theoretical Computer Science 178 (1997) 171–203.

[2] J. Berstel, P. Séébold, Sturmian words, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002, pp. 45–110.

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

[4] A. de Luca, Combinatorics of standard Sturmian words, Lecture Notes in Computer Science, vol. 1075, Springer, Berlin, 1997, pp. 249–267.

[5] A. de Luca, Standard Sturmian morphisms, Theoretical Computer Science 178 (1997) 205–224.

[6] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoretical Computer Science 136 (1994) 361–385.

[7] G.H. Hardy, E.M. Wright, An Introduction to the Theory of Numbers, Clarendon Press, Oxford, 1979.
[8] A. Ya Khinchin, Continued Fractions, The University of Chicago Press, Chicago, 1964.
[9] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading, MA, 1983.
[10] É. Lucas, Théorie des Nombres, vol. I, Gauthier-Villars, Paris, 1891.
[11] O. Ore, Number Theory and Its History, Dover Publications, New York, 1988.
[12] H. Rademacher, Lectures on Elementary Number Theory, R. E. Krieger Publishing Company, Huntington, NY, 1977.
[13] G. Rauzy, Mots infinis en arithmétique, in: M. Nivat, D. Perrin (Eds.), Automata on infinite words, Lecture Notes in Computer Science, vol. 192, Springer, Berlin, 1984, pp. 165–171.