Standard Factors of Sturmian Words

Gwenaël Richomme, Kalle Saari, Luca Q. Zamboni

RAIRO-Theor. Inf. Appl. 44 (2010) 159–174
DOI: 10.1051/ita/2010011

Available online at:
www.rairo-ita.org

STANDARD FACTORS OF STURMIAN WORDS *, **

Gwénaël Richomme[1, 2], Kalle Saari[3]
and Luca Q. Zamboni[4, 5]

Abstract. Among the various ways to construct a characteristic Sturmian word, one of the most used consists in defining an infinite sequence of prefixes that are standard. Nevertheless in any characteristic word c, some standard words occur that are not prefixes of c. We characterize all standard words occurring in any characteristic word (and so in any Sturmian word) using firstly morphisms, then standard prefixes and finally palindromes.

Mathematics Subject Classification. 68R15.

1. Introduction

Morse and Hedlund [25] begun the study of Sturmian words in 1940 to help develop the theory of symbolic dynamical systems they initiated a couple of years earlier [24]. Since then, and especially during the last two decades, Sturmian words have been under intensive scrutiny, as indicated by the surveys [4,5,7]. A reason of

Keywords and phrases. Sturmian words, standard factors, morphisms, palindromes.

* *The second author was supported by the Finnish Academy under grant 8206039.*

** *The third author is partially supported by grant no. 090038011 from the Icelandic Research Fund.*

[1] Université de Picardie Jules Verne, Laboratoire MIS (Modélisation, Information, Systèmes), 33 rue Saint Leu, 80039 Amiens Cedex 1, France; gwenael.richomme@u-picardie.fr

[2] Université Paul-Valéry Montpellier 3, UFR 4, Dpt. MIAp, Route de Mende, 34199 Montpellier Cedex 5, gwenael.richomme@univ-montp3.fr

[3] Department of Mathematics and Turku Centre for Computer Science, University of Turku, 20014 Turku, Finland; kasaar@utu.fi

[4] Université de Lyon, Université Lyon 1, CNRS UMR 5208 Institut Camille Jordan, Bâtiment du Doyen Jean Braconnier, 43 bd. du 11 novembre 1918, 69622 Villeurbanne Cedex, France; luca.zamboni@wanadoo.fr

[5] Reykjavik University, School of Computer Science, Kringlan 1, 103 Reykjavik, Iceland; lqz@ru.is

Article published by EDP Sciences

this passion comes from the various connections Sturmian words have with other domains like for instance number theory [1,2], discrete geometry [8,20], crystallography [10] or scheduling [3].

Another reason is the richness of combinatorial properties of Sturmian words (see previous references) which concerns various aspects including, *e.g.*, palindromes, repetitions, unbordered or Lyndon words (see for instance [6,9,13,14,19, 22,23,26,27]). The problem considered in the current paper continues a theme of characterizing those factors of a Sturmian word that belong to some interesting class of finite words.

We consider standard words. These finite words can be used to define in a constructive way some Sturmian words (see Sect. 2 for more details) building arbitrarily long prefixes of so-called characteristic Sturmian words. Among these infinite words, the well-known Fibonacci word is maybe the one with a maximal number of extremal properties [11]. In his Ph.D. thesis [28], the second author observed that the Fibonacci word contains some factors that are standard but not prefixes of it, and he asked for a characterization of the set of standard words occurring as factors in the Fibonacci word.

Sections 3, 4 and 5 solve this problem in a much more general way since they provide, for any Sturmian word, characterizations of its factors that are standard. Section 3 uses existing links between standard words and Sturmian morphisms. The proof is rather technical but the result allows to derive in a simple way a characterization using standard words occurring in the construction of characteristic words mentioned above (Sect. 4). Previous characterizations are illustrated using the Fibonacci word as an example. This word is also used in Section 5 to show how to relate standard factors of a Sturmian word to palindromes. This needs a characterization based on directive sequences of standard words.

2. Standard and Sturmian words

We will follow the usual notation and terminology of combinatorics on words. For further information about the concepts and results mentioned in this section, we refer the reader to [12,21].

A *word* is a sequence, finite or infinite, of symbols drawn from a finite alphabet $\mathcal{A}$; in this paper we set $\mathcal{A} = \{0, 1\}$. The empty word is denoted by ε. The set of all finite words over $\mathcal{A}$ is denoted by $\mathcal{A}^*$; the set of all nonempty words over $\mathcal{A}$ is denoted by $\mathcal{A}^+$. A finite word u is a *factor* of a word w if we can write $w = xuy$. The length of a word u is denoted by $|u|$, and the number of occurrences of a letter a in u is denoted by $|u|_a$. For an integer $k \geq 2$, a *kth power* is a word of the form u^k, that is, a word obtained concatenating k occurrences of a given word u (we also denote $u^1 = u$ and $u^0 = \varepsilon$).

A *Sturmian word* is an infinite word with precisely $n+1$ factors for each length n. Equivalently [21], Theorem 2.1.13 and Proposition 2.1.18 a *Sturmian word* is an infinite word whose factors coincide with those of a characteristic word, where a *characteristic word* is an infinite word c_α, depending on an irrational α with

$0 < \alpha < 1$, such that

$$c_\alpha(n) = \lfloor \alpha(n+1) \rfloor - \lfloor \alpha n \rfloor$$

for all $n \geq 1$. When proving a property of the factors of a Sturmian word, it often suffices to prove it for a characteristic word.

A finite word s is called a *standard word* if there exist integers $n \geq -1$, $d_1 \geq 0$, $d_2, d_3, \ldots, d_n \geq 1$, and words $s_{-1}, s_0, s_1, \ldots, s_n$ with $s = s_n$ such that

$$s_{-1} = 1, \quad s_0 = 0, \quad s_k = s_{k-1}^{d_k} s_{k-2} \quad (k \geq 1).$$

Also, if $n \geq 1$, then $(d_1, d_2, \ldots, d_n)$ is called the *directive sequence* of the word s_n. Observe that the letter 1 has a directive sequence (0), but the letter 0 does not have any. Therefore we agree that the directive sequence of 0 is the empty sequence.

Two basic properties of standard words are that (1) they are primitive (that is, not a power of a shorter word) and (2) all standard words of length at least 2 have a suffix in $\{01, 10\}$.

In this paper, we are interested in factors of an infinite word that are standard words; we call such a factor briefly a *standard factor*.

Let us denote the simple continued fraction expansion of α by

$$\alpha = [\,0; d_1 + 1, d_2, d_3, \ldots\,].$$

The sequence $(d_1, d_2, d_3, \ldots)$ is called the *directive sequence* of c_α. This terminology is justified by the fact that c_α can be obtained as a limit of those standard words whose directive sequences are prefixes of the one of c_α, that is,

$$c_\alpha = \lim_{n \to \infty} s_n.$$

Many properties of Sturmian words can be dealt with using morphisms. Let us recall that a mapping $h\colon \mathcal{A}^* \to \mathcal{A}^*$ is called a *morphism* if it satisfies $h(uv) = h(u)h(v)$ for all $u, v \in \mathcal{A}^*$. Observe that a morphism is uniquely determined by how it maps letters.

The well-known Fibonacci word $\mathbf{f}$ is the characteristic word c_α with $\alpha = (3 - \sqrt{5})/2$, so that it has directive sequence $(1, 1, 1, \ldots)$. It is also the unique fixed point of the morphism

$$\varphi : \begin{cases} 0 & \mapsto & 01 \\ 1 & \mapsto & 0. \end{cases}$$

Next we define the morphisms

$$L_0 : \begin{cases} 0 & \mapsto & 0 \\ 1 & \mapsto & 01 \end{cases} \qquad L_1 : \begin{cases} 0 & \mapsto & 10 \\ 1 & \mapsto & 1. \end{cases}$$

The following lemma is well-known; it has a straightforward proof in the spirit of the proof of [7], Proposition 2.3.11, which we omit here.

Lemma 2.1. *A word $w \in \{0,1\}^*$ is standard if and only if there exists a morphism f in the monoid generated by L_0 and L_1 such that either $w = f(0)$ or $w = f(1)$.*

In addition to L_0 and L_1, we shall need a few other morphisms. For all integers $m \geq 1$, we define the morphisms

$$\theta_m : \begin{cases} 0 & \mapsto & 0^{m-1}1 \\ 1 & \mapsto & 0^{m-1}10. \end{cases}$$

We also define the morphism $E\colon 0 \mapsto 1, 1 \mapsto 0$, and denote (with $(d_1, d_2, d_3, \ldots)$ the directive sequence of c_α)

$$h_n = \theta_{d_1+1} \circ \theta_{d_2} \circ \ldots \circ \theta_{d_n} \quad (n \geq 1).$$

Morphisms h_n are handy because we have $h_n(0) = s_n$ and $h_n(1) = s_n s_{n-1}$ for all $n \geq 1$ (see the proof of [7], Prop. 2.2.24). Furthermore, they satisfy

$$h_{2k} = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}} \circ L_0, \text{ and} \tag{2.1}$$

$$h_{2k+1} = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_1^{d_{2k}} \circ L_0^{d_{2k+1}} \circ L_1 \circ E. \tag{2.2}$$

This follows from the two identities $\theta_m = L_0^{m-1} \circ E \circ L_0$ and $E \circ L_0 = L_1 \circ E$.

For a finite word x, the shortest palindrome that has x as a prefix is denoted by $x^{(+)}$; this is called the *(right) palindromic closure* of x due to de Luca [15]. We let Pal denote the operation of *iterated palindromic closure* defined by

$$Pal(a_1) = a_1 \qquad \text{and} \qquad Pal(a_1 a_2 \ldots a_n) = \big(Pal(a_1 a_2 \ldots a_{n-1}) a_n\big)^{(+)},$$

where a_i is a letter.

The characteristic word c_α can be also represented as follows. Let us denote $0^{d_1}1^{d_2}0^{d_3}\ldots = x_1 x_2 x_3 \ldots$, where $x_i \in \{0,1\}$. Then

$$c_\alpha = \lim_{n\to\infty} Pal(x_1 \ldots x_n).$$

The sequence $0^{d_1}1^{d_2}0^{d_3}\ldots$ is called the *directive word* of c_α. Both directive words and directive sequences will be used in this paper to represent a characteristic word.

The following relation (see [16], Thm. 9) will be useful. If $\mathbf{x}$ and $\mathbf{y}$ are characteristic words directed, respectively, by $a_1 a_2 a_3 \ldots$ and $a_2 a_3 \ldots$, where $a_i \in \{0,1\}$, then

$$\mathbf{x} = L_{a_1}(\mathbf{y}). \tag{2.3}$$

3. Characterization via morphisms L_0 and L_1

Here we present a characterization of standard factors of a Sturmian word using the morphisms L_0 and L_1. The proof can appear a bit laborious, but the result enables us to derive two other characterizations in a relatively simple manner.

First, however, we need to introduce an auxiliary notation. Let $n_1, n_2, n_3 \geq 1$ be integers. Then we let $\mathcal{S}(n_1, n_2, n_3)$ denote the union of the following five sets of words:

$$\big\{ L_0^i(0),\, L_0^i(1),\, L_0^i(10) \mid 0 \leq i < n_1 \big\},$$

$$\big\{ L_0^{n_1}(1^i 0) \mid 1 \leq i \leq n_2 \big\},$$

$$\big\{ L_0^{n_1-1}\big((10)^i 1\big) \mid 1 \leq i \leq n_2 \big\},$$

$$\big\{ L_0^{n_1}\big(1^{n_2-1}0(1^{n_2}0)^i 1^{n_2}0\big) \mid 1 \leq i \leq n_3 \big\},$$

$$\big\{ L_0^{n_1}\big(1^{i-1}01^i 0\big) \mid 1 \leq i \leq n_2 \big\}.$$

Using Lemma 2.1, it is readily checked that all words in $\mathcal{S}(n_1, n_2, n_3)$ are standard (note that $1^i 0 = L_1^i(0)$, $(10)^i 1 = L_1 \circ L_0^i(1)$, $1^{n_2-1}0(1^{n_2}0)^i 1^{n_2}0 = L_1^{n_2-1} \circ L_0 \circ L_1^{i+1}(0)$ and $1^{i-1}01^i 0 = L_1^{i-1} \circ L_0 \circ L_1(0)$).

It is instructive to observe that each word in $\mathcal{S}(n_1, n_2, n_3)$ is a factor of any characteristic word whose directive sequence starts with $0^{n_1}1^{n_2}0^{n_3}$. This follows from equation (2.3) when w belongs to one of the two first sets of $\mathcal{S}(n_1, n_2, n_3)$. When w belongs to one of the three last sets of $\mathcal{S}(n_1, n_2, n_3)$, we deduce the previous observation from the following three facts where $\mathbf{x}$ denotes an arbitrary non-ultimately-periodic infinite word over $\{0, 1\}$:

- 1^{n_2+1} is a factor of $L_1^{n_2}(\mathbf{x})$;
- $1^{n_2-1}0(1^{n_2}0)^{n_3}1^{n_2+1}$ is a factor of $L_1^{n_2} \circ L_0^{n_3}(\mathbf{x})$; and
- $1^{n_2-1}01^{n_2}0$ is a factor of $L_1^{n_2} \circ L_0^{n_3}(\mathbf{x})$.

Now we are ready for the main result of this section. In order to state it nicely, we extend notation $\mathcal{S}$ by letting $\mathcal{S}(0, n_2, n_3) = \emptyset$, the empty set, for all integers n_2 and n_3.

Theorem 3.1. *Let $\alpha = [\,0\,; d_1 + 1, d_2, \ldots]$. A finite standard word u is a factor of c_α if and only if there exists an integer $k \geq 0$ such that one of the following holds:*

- *There exists $x \in \mathcal{S}(d_{2k+1}, d_{2k+2}, d_{2k+3})$ such that*

$$u = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}}(x).$$

- *There exists $x \in \mathcal{S}(d_{2k+2}, d_{2k+3}, d_{2k+4})$ such that*

$$u = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}} \circ L_0^{d_{2k+1}} \circ E(x).$$

Remark 3.2. By the definition of the continued fraction $[\,0\,; d_1 + 1, d_2, \ldots]$, we have $d_1 \geq 0$ and $d_k \geq 1$ for $k \geq 2$. When $d_1 = 0$, the first item provides no standard word for $k = 0$.

We will divide the proof of Theorem 3.1 into several lemmas. But let us see first what it says in the case of the Fibonacci word:

Example 3.3. Recall that the directive sequence of the Fibonacci word $\mathbf{f}$ is $(1, 1, \ldots)$. Since

$$\mathcal{S}(1,1,1) = \{0, 1, 10, 010, 101, 0010, 0010010\},$$

and since $L_0 \circ L_1 = \varphi^2$ and $L_0 \circ E = \varphi$, it follows from Theorem 3.1 that the set of standard factors of the Fibonacci word equals

$$\left\{ \varphi^n(1), \varphi^n(10), \varphi^n(101), \varphi^n(0010010) \mid n \geq 0 \right\}.$$

We will see other ways to characterize these standard factors in later examples.

Now we are ready to start proving Theorem 3.1.

Lemma 3.4. *Let $\mathbf{x}, \mathbf{y}$ be two infinite words over $\{0,1\}$ such that $\mathbf{x} = L_0(\mathbf{y})$. Suppose that the letter 1 occurs in $\mathbf{x}$. Then a word u is a standard factor of $\mathbf{x}$ if and only if one of the following holds:*

(1) $u \in \{0, 1, 10\}$;
(2) $u = L_0(v)$ *with v a standard factor of* $\mathbf{y}$;
(3) $u = L_0(w0)$ *with $w \neq \varepsilon$, the word $w0$ standard, and $w1$ a factor of* $\mathbf{y}$;
(4) $u = (10)^n 1$ *with $n \geq 1$ and 1^{n+1} a factor of* $\mathbf{y}$.

Proof. Using Lemma 2.1, it is verified that conditions (1)–(4) are sufficient, that is, if u fulfills one of these conditions then u is a standard factor of $\mathbf{x}$.

Conversely, suppose first that $|u| \leq 2$. As a standard word, u is primitive, and consequently $u \in \{0, 1, 01, 10\}$. We see that u satisfies condition (1), unless $u = 01 = L_0(1)$, in which case u satisfies condition (2).

Suppose now that $|u| \geq 3$. Lemma 2.1 implies that either $u = L_0(v)$ or $u = L_1(v)$ for some standard word v. Furthermore, the primitivity and length of u imply that both letters 0 and 1 occur in v, and hence $|v| < |u|$. We have two possibilities to consider:

If $u = L_0(v)$, then u begins with 0. If v is a factor of $\mathbf{y}$, then condition (2) is satisfied, and therefore we may suppose that v is not a factor of $\mathbf{y}$. It follows that u also ends with 0. Now it is easy to see that $v = w0$ for some word $w \neq \varepsilon$ such that $w1$ is a factor of $\mathbf{y}$. Hence condition (3) holds.

If $u = L_1(v)$, then the word 00 does not occur in u. Since u is a factor of $\mathbf{x} = L_0(\mathbf{y})$, the word 11 does not occur in u either. Consequently, we have $u = (10)^n$ or $u = (10)^n 1$ for some $n \geq 0$. Since u is primitive and $|u| \geq 3$, it follows that $u = (10)^n 1$ and $n \geq 1$. Since u is a factor of $L_0(\mathbf{y})$, we see that 1^{n+1} must be a factor of $\mathbf{y}$. Thus condition (4) holds. □

Lemma 3.5. *Let* $\mathbf{y}$ *be an infinite word over* $\{0,1\}$*, and let* $k \geq 1$ *be an integer. Then a standard word* u *is a factor of* $L_0^k \circ L_1(\mathbf{y})$ *if and only if one of the following conditions holds:*

(1) $u \in \{ L_0^i(0), L_0^i(1), L_0^i(10) \mid 0 \leq i < k \}$;
(2) $u = L_0^k(v)$ *with* v *a standard factor of* $L_1(\mathbf{y})$;
(3) $u = L_0^k(w0)$ *with* $w \neq \varepsilon$*,* $w0$ *is standard, and* $w1$ *is a factor of* $L_1(\mathbf{y})$;
(4) $u = L_0^{k-1}\big((10)^n 1\big)$ *with* $n \geq 1$ *and* 1^{n+1} *is a factor of* $L_1(\mathbf{y})$.

Proof. It is easy to see that conditions (1)–(4) are sufficient. We prove the necessity of the conditions by induction on k. The case $k = 1$ follows immediately from Lemma 3.4. So we may suppose that $k \geq 2$. Using Lemma 3.4, we have the following four cases to consider:

Case 1. If $u \in \{0, 1, 10\}$, then condition (1) holds.

Case 2. We have $u = L_0(v)$, where v is a standard factor of $L_0^{k-1}(\mathbf{y})$. By the induction assumption, one of the four conditions holds:

(1) $v \in \{ L_0^i(0), L_0^i(1), L_0^i(10) \mid 0 \leq i < k-1 \}$;
(2) $v = L_0^{k-1}(v')$ with v' a standard factor of $L_1(\mathbf{y})$;
(3) $v = L_0^{k-1}(w0)$ with $w \neq \varepsilon$, $w0$ standard, and $w1$ a factor of $L_1(\mathbf{y})$;
(4) $v = L_0^{k-2}\big((10)^n 1\big)$ with $n \geq 1$ and 1^{n+1} a factor of $L_1(\mathbf{y})$.

Therefore one of conditions (1)–(4) holds.

Case 3. We have $u = L_0(w0)$ with $w \neq \varepsilon$, the word $w0$ is a standard word, and $w1$ a factor of $L_0^{k-1}(\mathbf{y})$. Since $w0$ is standard, the word w must end in the letter 1. Consequently, the word 11 is a factor of $L_0^{k-1}(\mathbf{y})$. But this is not possible as $k - 1 \geq 1$.

Case 4. We have $u = (10)^n 1$ with $n \geq 1$ and 1^{n+1} a factor of $L_0^{k-1}(\mathbf{y})$. Again, since $k - 1 \geq 1$, this case is not possible. The proof is complete. □

Lemma 3.6. *If* u *is a standard word with* $u \in \{0,1\}^+0$*, then there exists an integer* $\ell \geq 1$ *such that either* $u = 1^\ell 0$ *or* $u \in 1^{\ell-1}0\{1^{\ell-1}0, 1^\ell 0\}^* 1^\ell 0$.

Proof. The claim is clearly true if $|u| = 2$. If $|u| \geq 3$, then Lemma 2.1 implies that $u = L_0(v)$ or $u = L_1(v)$ for some standard word v with $2 \leq |v| < |u|$. Furthermore, also v satisfies $v \in \{0,1\}^+0$, and therefore either v is of the form $1^\ell 0$ or it is in $1^{\ell-1}0\{1^{\ell-1}0, 1^\ell 0\}^* 1^\ell 0$ for some $\ell \geq 1$. A straightforward computation shows that u is of one of the attested forms. □

Lemma 3.7. *Let* $\mathbf{y}$ *be a Sturmian word over* $\{0,1\}$*. Suppose that* $w \neq \varepsilon$ *is a finite word such that* $w0$ *is a standard word and* $w1$ *is a factor of* $\mathbf{y}$*. Then either* $w = 1^\ell$ *or* $w = 1^{\ell-1}0(1^\ell 0)^i 1^\ell$*, where* $\ell \geq 1$ *and* $i \geq 0$ *are integers.*

Proof. Since $w \neq \varepsilon$ and $w0$ is standard, the word w ends in the letter 1. Let $\ell \geq 1$ denote the largest integer such that 1^ℓ is a suffix of w. If $w = 1^\ell$, we are done, so we may suppose that this is not the case. Then by Lemma 3.6, we have $w0 \in 1^{\ell-1}0\{1^{\ell-1}0, 1^\ell 0\}^* 1^\ell 0$.

Since the works of Morse and Hedlund [25] it is well-known that all Sturmian words have the balance property, that is, for all factors u, v of the same length

of a Sturmian word, we have $\big||u|_0 - |v|_0\big| \le 1$. Hence, since $w1$ ends in $1^{\ell+1}$, the balance property of $\mathbf{y}$ implies that $01^{\ell-1}0$ cannot occur in w. Consequently, we have $w0 \in 1^{\ell-1}0\{1^\ell 0\}^* 1^\ell 0$, and the proof is complete. □

Using the previous result, we may rewrite Lemma 3.5 as follows.

Lemma 3.8. *Let* $\mathbf{y}$ *be an infinite word over* $\{0,1\}$*, and let* $k \ge 1$ *be an integer. Then a standard word* u *is a factor of* $L_0^k \circ L_1(\mathbf{y})$ *if and only if one of the following conditions holds:*

(1) $u \in \{ L_0^i(0), L_0^i(1), L_0^i(10) \mid 0 \le i < k \}$;
(2) $u = L_0^k(v)$ *with* v *a standard factor of* $L_1(\mathbf{y})$;
(3) $u = L_0^k(1^\ell 0)$ *with* $\ell \ge 1$ *and* $1^{\ell+1}$ *is a factor of* $L_1(\mathbf{y})$;
(4) $u = L_0^k\big(1^{\ell-1}0(1^\ell 0)^i 1^\ell 0\big)$*, where* $\ell \ge 1$*,* $i \ge 0$*, and* $1^{\ell-1}0(1^\ell 0)^i 1^{\ell+1}$ *is a factor of* $L_1(\mathbf{y})$;
(5) $u = L_0^{k-1}\big((10)^n 1\big)$ *with* $n \ge 1$ *and* 1^{n+1} *a factor of* $L_1(\mathbf{y})$.

Proof. This statement is obtained by applying Lemma 3.7 to condition (3) of Lemma 3.5, which then splits into conditions (3) and (4) of the current lemma. Therefore we only need to make sure that conditions (3) and (4) are sufficient. But this is immediately clear. □

Finally, before proving Theorem 3.1, we formulate the previous lemma into a more suitable form:

Lemma 3.9. *Let* $\mathbf{y}$ *be an infinite word over* $\{0,1\}$ *such that* 0 *occurs in* $\mathbf{y}$*. Let* $d_1, d_2, d_3 \ge 1$ *be integers. A standard word* u *is a factor of* $L_0^{d_1} \circ L_1^{d_2} \circ L_0^{d_3} \circ L_1(\mathbf{y})$ *if and only if one of the following conditions holds:*

(1) $u \in \{ L_0^i(0), L_0^i(1), L_0^i(10) \mid 0 \le i < d_1 \}$;
(2) $u = L_0^{d_1}(v)$ *with* v *a standard factor of* $L_1^{d_2} \circ L_0^{d_3} \circ L_1(\mathbf{y})$;
(3) $u \in \{L_0^{d_1}(1^i 0) \mid 1 \le i \le d_2\}$;
(4) $u \in \{L_0^{d_1}\big(1^{\ell-1}0(1^\ell 0)^i 1^\ell 0\big) \mid (i = 0$ *and* $1 \le \ell \le d_2)$ *or* $(1 \le i \le d_3$ *and* $\ell = d_2)\}$;
(5) $u \in \{L_0^{d_1-1}\big((10)^i 1\big) \mid 1 \le i \le d_2\}$.

Proof. That conditions (1)–(3) and (5) are sufficient is a direct consequence of the sufficiency of corresponding conditions in Lemma 3.8. For condition (4), the same holds since either 01 or 00 occurs in $\mathbf{y}$ and thus the word $1^{d_2-1}0(1^{d_2}0)^{d_3}1^{d_2+1}$ is a factor of $L_1^{d_2} \circ L_0^{d_3} \circ L_1(\mathbf{y})$.

Conversely, suppose u is a standard factor of $L_0^{d_1} \circ L_1^{d_2} \circ L_0^{d_3} \circ L_1(\mathbf{y})$. In what follows, we denote $\mathbf{z} = L_1^{d_2} \circ L_0^{d_3} \circ L_1(\mathbf{y})$. Lemma 3.8 implies that one of the following five cases holds:

Case 1. Condition (1) of Lemma 3.8 holds with $k = d_1$. This is condition (1) of the present lemma.

Case 2. Condition (2) in Lemma 3.8 holds, and we have $u = L_0^{d_1}(v)$ with v a standard factor of $\mathbf{z}$.

Case 3. Condition (3) in Lemma 3.8 holds, that is, we have

$$u = L_0^{d_1}(1^\ell 0)$$

with $\ell \geq 1$ and $1^{\ell+1}$ is a factor of $\mathbf{z}$. It is easy to see that 1^{d_2+2} is not a factor of $\mathbf{z}$; therefore $\ell \leq d_2$.

Case 4. Next we suppose that condition (4) in Lemma 3.8 holds. Then

$$u = L_0^{d_1}\big(1^{\ell-1}0(1^\ell 0)^i 1^\ell 0\big),$$

where $\ell \geq 1$, $i \geq 0$, and $1^{\ell-1}0(1^\ell 0)^i 1^{\ell+1}$ is a factor of $\mathbf{z}$. Since the word 1^{d_2+2} is not a factor $\mathbf{z}$, we have $\ell \leq d_2$. We may thus assume that $i \geq 1$.

Now the word $01^\ell 0$ is a factor of $\mathbf{z}$. Since $\mathbf{z}$, as any word of the form $L_1^{d_2}(\mathbf{x})$, can be factorized over $\{1^{d_2}0, 1\}$, and since $\ell \leq d_2$, it follows that $\ell = d_2$. To finish proving this case, we only need to show that $i \leq d_3$. Indeed, otherwise the word

$$1^{d_2-1}0(1^{d_2}0)^{d_3+1}1^{d_2+1}$$

is a factor of $\mathbf{z}$, and it follows that 0^{d_3+2} is a factor of $L_0^{d_3} \circ L_1(\mathbf{y})$, which is not possible because 00 is not factor of $L_1(\mathbf{y})$.

Case 5. Finally, suppose that condition (5) in Lemma 3.8 holds, that is,

$$u = L_0^{d_1-1}\big((10)^n 1\big)$$

with $n \geq 1$ and 1^{n+1} a factor of $\mathbf{z}$. Since 1^{d_2+2} is not a factor of $\mathbf{z}$, we have $n \leq d_2$. □

Now we are ready to finish proving Theorem 3.1.

Proof of Theorem 3.1. From Lemma 2.1, a finite word u is a standard factor of c_α if and only if $E(u)$ is a standard factor of $E(c_\alpha)$. When $d_1 = 0$, $E(c_\alpha)$ is the word c_β with $\beta = 1 - \alpha = [0; d_2 + 1, d_3, \ldots]$ (see [7], Cor. 2.2.20). Then since $E \circ L_0 = L_1 \circ E$ (and $E \circ L_1 = L_0 \circ E$), it is straightforward that if Theorem 3.1 holds for c_β then it holds for c_α. So we may assume that $d_1 \geq 1$.

Suppose that

$$u = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}}(x) \quad \text{with } x \in \mathcal{S}(d_{2k+1}, d_{2k+2}, d_{2k+3}).$$

As noted earlier, x is a standard factor of any characteristic word whose directive word starts with $0^{d_{2k+1}}1^{d_{2k+2}}0^{d_{2k+3}}\ldots$ Consequently, by Equation (2.3) and Lemma 2.1, u is a standard factor of c_α. Analogous reasoning works also in the case when

$$u = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}} \circ L_0^{d_{2k+1}} \circ E(x) \quad \text{with } x \in \mathcal{S}(d_{2k+2}, d_{2k+3}, d_{2k+4})$$

because then $E(x)$ is a standard factor of any characteristic word whose directive word starts with $1^{d_{2k+2}}0^{d_{2k+3}}1^{d_{2k+4}}\ldots$

Conversely, suppose that u is a standard factor of c_α. By equation (2.3), $c_\alpha = L_0^{d_1} \circ L_1^{d_2} \circ L_0^{d_3} \circ L_1(\mathbf{y})$ for a word $\mathbf{y}$. Consequently, since we are assuming that $d_1 \geq 1$, Lemma 3.9 applies, and so u satisfies one of conditions (1)–(5) of Lemma 3.9. If u satisfies condition (1), (3), (4), or (5), then $u \in \mathcal{S}(d_1, d_2, d_3)$, so we may assume that condition (2) holds. Then $u = L_0^{d_1}(u_1)$, where u_1 is a standard factor of the characteristic word $L_1^{d_2} \circ L_0^{d_3} \circ L_1^{d_4} \circ \dots$

Denote $u_1' = E(u_1)$. Then $u = L_0^{d_1} \circ E(u_1')$, and u_1' is a standard factor of the characteristic word $L_0^{d_2} \circ L_1^{d_3} \circ L_0^{d_4} \circ \dots$ Consequently, Lemma 3.9 applies, and so either $u_1' \in \mathcal{S}(d_2, d_3, d_4)$ or $u_1' = L_0^{d_2}(u_2)$, where u_2 is a standard factor of the characteristic word $L_1^{d_3} \circ L_0^{d_4} \circ L_1^{d_5} \circ \dots$ In the first case, we have

$$u = L_0^{d_1} \circ E(u_1') \quad \text{and} \quad u_1' \in \mathcal{S}(d_2, d_3, d_4),$$

and the claim holds. In the second case, we denote $u_2' = E(u_2)$, whence u_2' is a standard factor of the characteristic word $L_0^{d_3} \circ L_1^{d_4} \circ L_0^{d_5} \circ \dots$ Again, Lemma 3.9 implies that either $u_2' \in \mathcal{S}(d_3, d_4, d_5)$ or $u_2' = L_0^{d_3}(u_3)$, where u_3 is a standard factor of the characteristic word $L_1^{d_4} \circ L_0^{d_5} \circ L_1^{d_6} \circ \dots$ In the first case, we have

$$u = L_0^{d_1} \circ E \circ L_0^{d_2}(u_2) = L_0^{d_1} \circ L_1^{d_2}(u_2') \quad \text{and} \quad u_2' \in \mathcal{S}(d_3, d_4, d_5),$$

and the claim holds. In the second case, we continue the process.

Since $d_i \geq 1$ for $i \geq 2$, the sequence of lengths of words $u_1, u_2, u_3, \dots$ is strictly decreasing, and hence the procedure described above cannot continue forever. When it stops, we arrive at the form given in the claim. □

4. Characterization via standard words

Here we reformulate Theorem 3.1 to a more constructive form.

Theorem 4.1. *Let $\alpha = [0; d_1 + 1, d_2, \dots]$. Then a finite standard word u is a factor of c_α if and only if there exists an integer $k \geq 0$, or $k \geq 1$ if $d_1 = 0$, such that u belongs to one of the following sets (where the s_i's correspond to those defined in Sect. 2):*

$$\begin{aligned}
&\{\, s_k,\ s_k^i s_{k-1},\ s_k^i s_{k-1} s_k \mid 0 \leq i < d_{k+1} \,\};\\
&\{\, s_{k+1}^{i-1} s_k s_{k+1}^i s_k \mid 1 \leq i \leq d_{k+2} \,\};\\
&\{\, s_{k+1}^{d_{k+2}-1} s_k s_{k+2}^{i+1} \mid 1 \leq i \leq d_{k+3} \,\};\\
&\{\, s_k^{d_{k+1}-1} s_{k-1} s_{k+1}^i \mid 1 \leq i \leq d_{k+2} \,\}.
\end{aligned}$$

Proof. We start with the following four identities; they follow immediately from equations (2.1) and (2.2). For all $k \geq 0$, we have

$$L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}}(0) = s_{2k}, \tag{4.1}$$

$$L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}}(1) = s_{2k-1}, \tag{4.2}$$

$$L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}} \circ L_0^{d_{2k+1}}(0) = s_{2k}, \tag{4.3}$$

$$L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}} \circ L_0^{d_{2k+1}}(1) = s_{2k+1}. \tag{4.4}$$

Next, if $x \in \mathcal{S}(d_{2k+1}, d_{2k+2}, d_{2k+3})$, then x belongs to one of the following sets:

$$\{\, L_0^i(0),\, L_0^i(1),\, L_0^i(10) \mid 0 \leq i < d_{2k+1} \,\} = \{\, 0,\, 0^i1,\, 0^i10 \mid 0 \leq i < d_{2k+1} \,\};$$

$$\{\, L_0^{d_{2k+1}}(1^i0) \mid 1 \leq i \leq d_{2k+2} \,\};$$

$$\{\, L_0^{d_{2k+1}}\big(1^{i-1}01^i0\big) \mid 1 \leq i \leq d_{2k+2} \,\};$$

$$\{\, L_0^{d_{2k+1}}\big(1^{d_{2k+2}-1}0(1^{d_{2k+2}}0)^i1^{d_{2k+2}}0\big) \mid 1 \leq i \leq d_{2k+3} \,\} =$$
$$\{\, L_0^{d_{2k+1}}\big(1^{d_{2k+2}-1}0\big(L_1^{d_{2k+2}}(0)\big)^{i+1}\big) \mid 1 \leq i \leq d_{2k+3} \,\};$$

$$\{\, L_0^{d_{2k+1}-1}\big((10)^i1\big) \mid 1 \leq i \leq d_{2k+2} \,\} =$$
$$\{\, 0^{d_{2k+1}-1}1\big(L_0^{d_{2k+1}}(1)\big)^i \mid 1 \leq i \leq d_{2k+2} \,\}.$$

Consequently, if

$$u = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}}(x) \quad \text{with} \quad x \in \mathcal{S}(d_{2k+1}, d_{2k+2}, d_{2k+3}),$$

the equations above imply that u is in one of the sets

$$\{\, s_{2k},\, s_{2k}^i s_{2k-1},\, s_{2k}^i s_{2k-1} s_{2k} \mid 0 \leq i < d_{2k+1} \,\},$$
$$\{\, s_{2k+1}^i s_{2k} \mid 1 \leq i \leq d_{2k+2} \,\},$$
$$\{\, s_{2k+1}^{i-1} s_{2k} s_{2k+1}^i s_{2k} \mid 1 \leq i \leq d_{2k+2} \,\}.$$
$$\{\, s_{2k+1}^{d_{2k+2}-1} s_{2k} s_{2k+2}^{i+1} \mid 1 \leq i \leq d_{2k+3} \,\},$$
$$\{\, s_{2k}^{d_{2k+1}-1} s_{2k-1} s_{2k+1}^i \mid 1 \leq i \leq d_{2k+2} \,\}.$$

Similarly, if

$$u = L_0^{d_1} \circ L_1^{d_2} \circ \ldots \circ L_0^{d_{2k-1}} \circ L_1^{d_{2k}} \circ L_0^{d_{2k+1}} \circ E(x) \quad \text{with} \quad x \in \mathcal{S}(d_{2k+2}, d_{2k+3}, d_{2k+4}),$$

then it follows from the equations above that u is in one of the sets

$$\begin{aligned}
&\{\, s_{2k+1},\ s_{2k+1}^i s_{2k},\ s_{2k+1}^i s_{2k} s_{2k+1} \mid 0 \le i < d_{2k+2} \,\},\\
&\{\, s_{2k+2}^i s_{2k+1} \mid 1 \le i \le d_{2k+3} \,\},\\
&\{\, s_{2k+2}^{i-1} s_{2k+1} s_{2k+2}^i s_{2k+1} \mid 1 \le i \le d_{2k+3} \,\},\\
&\{\, s_{2k+2}^{d_{2k+3}-1} s_{2k+1} s_{2k+3}^{i+1} \mid 1 \le i \le d_{2k+4} \,\},\\
&\{\, s_{2k+1}^{d_{2k+2}-1} s_{2k} s_{2k+2}^i \mid 1 \le i \le d_{2k+3} \,\}.
\end{aligned}$$

Applying Theorem 3.1, it follows that u is a standard factor of c_α if and only if there exists an integer $k \ge 0$, or $k \ge 1$ if $d_1 = 0$, such that u is in one of the following sets

$$\begin{aligned}
&\{\, s_k,\ s_k^i s_{k-1},\ s_k^i s_{k-1} s_k \mid 0 \le i < d_{k+1} \,\},\\
&\{\, s_{k+1}^i s_k \mid 1 \le i \le d_{k+2} \,\},\\
&\{\, s_{k+1}^{i-1} s_k s_{k+1}^i s_k \mid 1 \le i \le d_{k+2} \,\},\\
&\{\, s_{k+1}^{d_{k+2}-1} s_k s_{k+2}^{i+1} \mid 1 \le i \le d_{k+3} \,\},\\
&\{\, s_k^{d_{k+1}-1} s_{k-1} s_{k+1}^i \mid 1 \le i \le d_{k+2} \,\}.
\end{aligned}$$

Now the formulation of the statement is obtained by observing that sets of the form

$$\{\, s_{k+1}^i s_k \mid 1 \le i \le d_{k+2} \,\}$$

are included in the union of sets of the form

$$\{\, s_k,\ s_k^i s_{k-1},\ s_k^i s_{k-1} s_k \mid 0 \le i < d_{k+1} \,\}.$$

This completes the proof. □

Example 4.2. Let us denote $f_n = \varphi^n(0)$ for $n \ge 0$ and $f_{-1} = 1$. Then the characterization of the standard factors of the Fibonacci word $\mathbf{f}$ given by Theorem 4.1 is

$$\{\, f_{n-1},\quad f_{n-1}f_n,\quad f_n f_{n+2} f_{n+2},\quad f_{n-1}f_{n+1} \mid n \ge 0 \,\},$$

which is clearly the same set we obtained in Example 3.3.

5. Characterization via the directive sequence

In this section, we characterize the standard factors of a Sturmian word in terms of their directive sequences, and show how this can be used for characterizing standard factors in terms of palindromic closure.

Theorem 5.1. *Let $\alpha = [\,0\,; d_1 + 1, d_2, \dots]$. A finite standard word u is a factor of c_α if and only if it has one of the following directive sequences, where $k \geq 0$, or $k \geq 1$ if $d_1 = 0$:*

1.	*empty sequence,* (0), $(0,1)$;		
2.	$(d_1, d_2, \dots, d_{k-1}, d_k, i)$,	$1 \leq i \leq d_{k+1} + 1$;	
3.	$(d_1, d_2, \dots, d_{k-1}, d_k, i, 1)$,	$1 \leq i \leq d_{k+1} + 1$;	
4.	$(d_1, d_2, \dots, d_{k-1}, d_k, d_{k+1}, i, 1, 1)$,	$1 \leq i < d_{k+2}$;	
5.	$(d_1, d_2, \dots, d_{k+1}, d_{k+2} - 1, 1, i + 1)$,	$1 \leq i \leq d_{k+3}$	*if* $d_{k+2} \neq 1$;
6.	$(d_1, d_2, \dots, d_{k-1}, d_k, d_{k+1} + 1, i + 1)$,	$1 \leq i \leq d_{k+3}$	*if* $d_{k+2} = 1$;
7.	$(d_1, d_2, \dots, d_k, d_{k+1} - 1, 1, i)$,	$1 \leq i \leq d_{k+2}$	*if* $d_{k+1} \neq 1$;
8.	$(d_1, d_2, \dots, d_{k-1}, d_k + 1, i)$,	$1 \leq i \leq d_{k+2}$	*if* $d_{k+1} = 1$ *and* $k \neq 0$;
9.	$(0, 1, i)$,	$1 \leq i \leq d_2$	*if* $d_1 = 1$.

Remark 5.2. In the previous theorem, some items (*e.g.*, 5 and 7) give the same directive sequences for some instances of d_k and i. Nevertheless, no set of directive sequences corresponding to an item is included in another.

Proof of Theorem 5.1. Theorem 5.1 is a direct consequence of Theorem 4.1. Below we show the correspondence between words occurring in Theorem 4.1 and their directive sequences. More precisely, using the characterization of standard factors of c_α in Theorem 4.1, we show that each of the standard factors has a directive sequence listed above. The converse can be seen with a similar construction.

By [form n] we indicate that the directive sequence is of the form of the nth item in Theorem 5.1. Thus for each item, looking at all directive sequences on this form for all $k \geq 0$, one can verify the possible values for i.

- The directive sequence of the word s_k is

$$\begin{cases} (d_1, \dots, d_k) & \text{if } k \neq 0, \text{ [form 2][form 3 when } d_k = 1 \text{ and } k \geq 2] \\ \text{the empty sequence} & \text{if } k = 0. \text{ [form 1]} \end{cases}$$

- The directive sequence of the word $s_k^i s_{k-1}$ for $k \geq 0$ and $0 \leq i < d_{k+1}$ is

$$\begin{cases} (d_1, \dots, d_{k-1}, d_k, i) & \text{if } i \neq 0, \text{ [form 2] [form 3 when } i = 1 \text{ and } k \geq 1] \\ (d_1, \dots, d_{k-1}) & \text{if } i = 0 \text{ and } k \geq 2, \text{ [form 2]} \\ \text{the empty sequence} & \text{if } i = 0 \text{ and } k = 1, \text{ [form 1]} \\ (0) & \text{if } i = 0 \text{ and } k = 0. \text{ [form 1]} \end{cases}$$

- The directive sequence of the word $s_k^i s_{k-1} s_k$ for $k \geq 0$ and $0 \leq i < d_{k+1}$ is

$$\begin{cases} (d_1, \dots, d_{k-1}, d_k, i, 1) & \text{if } i \neq 0 \text{ (sequence } (i,1) \text{ when } k = 0), \text{ [form 3]} \\ (d_1, \dots, d_{k-1}, d_k + 1) & \text{if } i = 0 \text{ and } k \neq 0, \text{ [form 2]} \\ (0, 1) & \text{if } i = 0 \text{ and } k = 0. \text{ [form 1]} \end{cases}$$

  For the second case, note that $s_{k-1} s_k = s_{k-1}^{d_k+1} s_{k-2}$.

- The directive sequence of the word $s_{k+1}^{i-1}s_k s_{k+1}^i s_k$ for $k \geq 0$ and $1 \leq i \leq d_{k+2}$ is

$$\begin{cases} (d_1, \ldots, d_k, d_{k+1}, i-1, 1, 1) & \text{if } i \neq 1, \text{ [form 4]} \\ (d_1, \ldots, d_k, d_{k+1}+1, 1) & \text{if } i = 1. \text{ [form 3]} \end{cases}$$

  For the second case, note that $s_k s_{k+1} s_k = s_k^{d_{k+1}+1} s_{k-1} s_k$.

- The directive sequence of the word $s_{k+1}^{d_{k+2}-1} s_k s_{k+2}^{i+1}$ for $k \geq 0$ and $1 \leq i \leq d_{k+3}$ is

$$\begin{cases} (d_1, \ldots, d_k, d_{k+1}, d_{k+2}-1, 1, i+1) & \text{if } d_{k+2} \neq 1, \text{ [form 5]} \\ (d_1, \ldots, d_k, d_{k+1}+1, i+1) & \text{if } d_{k+2} = 1. \text{ [form 6]} \end{cases}$$

  In the first case we used $s_{k+1}^{d_{k+2}-1} s_k s_{k+2}^{i+1} = (s_{k+1}^{d_{k+2}-1} s_k s_{k+1})^{i+1} s_{k+1}^{d_{k+2}-1} s_k$; the second case follows from $s_k s_{k+2}^{i+1} = s_k (s_{k+1} s_k)^{i+1} = (s_k^{d_{k+1}+1} s_{k-1})^{i+1} s_k$.

- The directive sequence of the word $s_k^{d_{k+1}-1} s_{k-1} s_{k+1}^i$ for $k \geq 0$ and $1 \leq i \leq d_{k+2}$ is

$$\begin{cases} (d_1, \ldots, d_k, d_{k+1}-1, 1, i) & \text{if } d_{k+1} \neq 1, \text{ [form 7]} \\ (d_1, \ldots, d_k+1, i) & \text{if } d_{k+1} = 1,\ k \neq 0, \text{ [form 8]} \\ (0, 1, i) & \text{if } d_{k+1} = 1,\ k = 0. \text{ [form 9].} \end{cases}$$

□

Next we will recall a result by de Luca connecting the directive sequence of a standard word and palindromic closures ([15], p. 66).

Each word w over $\{0,1\}$ is uniquely determined by a finite sequence $(h_1, h_2, \ldots, h_n)$ of integers, where $h_1 \geq 0$, $h_i > 0$ for $1 < i < n$ and $w = 0^{h_1}1^{h_2}0^{h_3}\ldots$; such a representation of w is called its *integral representation.*

Proposition 5.3 (de Luca)**.** *Let $w \in \{0,1\}^*$, and let $(h_1, h_2, \ldots, h_n)$ be its integral representation. The standard words $Pal(w)01$ and $Pal(w)10$ have the directive sequences*

$$(h_1, \ldots, h_n, 1) \text{ and } (h_1, \ldots, h_n + 1)$$

if n is even, and

$$(h_1, \ldots, h_n + 1) \text{ and } (h_1, \ldots, h_n, 1)$$

if n is odd, respectively.

We have the following immediate corollary of the previous proposition.

Corollary 5.4. *If s is a standard word of length at least 2 with directive sequence $(d_1, \ldots, d_n)$ then*

$$\begin{aligned} s &= Pal(0^{d_1}1^{d_2}\ldots 0^{d_{n-1}}1^{d_n-1})10, \quad \text{when } n \text{ is even,} \\ s &= Pal(0^{d_1}1^{d_2}\ldots 1^{d_{n-1}}0^{d_n-1})01, \quad \text{when } n \text{ is odd.} \end{aligned}$$

Now one can use Corollary 5.4 to present the standard factors of a Sturmian word by using the directive sequences given in Theorem 5.1 and the palindromic closure

operation. We did not do this since the statement of this result involves a lot of cases (more than in Thm. 5.1) and so does not seem interesting.

However, the Fibonacci word is a special case, as shown in the next example.

Example 5.5. In this example we present a characterization of the standard factors of the Fibonacci word that is essentially different from the ones in Examples 3.3 and 4.2.

The directive sequence of the Fibonacci word is $(1, 1, 1, \ldots)$. Thus Theorem 5.1 says that the directive sequences of the standard factors of the Fibonacci word are of the following form:

- empty sequence, (0), $(0, 1)$,
- $(1, 1, \ldots, 1, 1)$,
- $(1, 1, \ldots, 1, 2)$, possibly with no preceding 1's; that is $(2), (1, 2), (1, 1, 2), \ldots$,
- $(1, 1, \ldots, 1, 2, 1)$, possibly with no preceding 1's; that is $(2, 1), (1, 2, 1), \ldots$,
- $(1, 1, \ldots, 1, 2, 2)$, possibly with no preceding 1's; that is $(2, 2), (1, 2, 2), \ldots$,
- $(0, 1, 1)$.

Therefore, by Corollary 5.4, the standard factors of the Fibonacci word are

- $0, 1, 01, 10, 101$;
- For each nonempty prefix u of $(01)^\infty$, the words $Pal(u)01$ and $Pal(u)10$;
- For each nonempty prefix u of $(01)^\infty$ of odd length, the words $Pal(u0)10$ and $Pal(u01)10$;
- For each nonempty prefix u of $(01)^\infty$ of even length, the words $Pal(u1)01$ and $Pal(u10)01$.

To conclude let us mention that most of the previous characterizations can certainly be extended to the more general case of episturmian words [16,17] using links between episturmian morphisms and directive sequences of these words defined over arbitrary alphabet (see, *e.g.*, [18]). Nevertheless the main problem should be a combinatorial explosion.

References

[1] J.-P. Allouche and V. Berthé, *Words in number theory*, edited by M. Lothaire. *Applied Combinatorics on Words*, Cambridge University Press (2005) 520–574.

[2] J.-P. Allouche and J. Shallit, *Automatic sequences*. Cambridge University Press (2003).

[3] E. Altman, B. Gaujal and A. Hordijk, Balanced sequences and optimal routing. *J. ACM* **47** (2000) 752–775.

[4] P. Arnoux, *Sturmian sequences*, edited by P. Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, Lect. Notes Math. **1794** (2002) 143–198.

[5] J. Berstel, *Sturmian and episturmian Words (a survey of some recent results)*, edited by S. Bozapalidis and G. Rahonis. *Conference on algebraic informatics* (CAI'07), *Lect. Notes Comput. Sci.* **4728** (2007) 23–47.

[6] J. Berstel and A. de Luca, Sturmian words. Lyndon words and trees. *Theoret. Comput. Sci.* **178** (1997) 171–203.

[7] J. Berstel and P. Séébold, *Sturmian words*, edited by M. Lothaire. *Algebraic combinatorics on words*, Cambridge University Press (2002) 45–110.

[8] J. Berstel, A. Lauve, C. Reutenauer and F. Saliola, Combinatorics on words: Christoffel words and repetition in words. CRM Monograph Series, Vol. 27, CRM-AMS Montréal (2008).
[9] V. Berthé, C. Holton and L.Q. Zamboni, Initial powers of Sturmian sequences. *Acta Inform.* **122** (2006) 315–347.
[10] E. Bombieri and J.E. Taylor, Which distributions of matter diffract? An initial investigation. *J. Phys.* **47** (1986) 19–28.
[11] J. Cassaigne, On extremal properties of the Fibonacci word. *RAIRO-Theor. Inf. Appl.* **42** (2008) 701–715.
[12] C. Choffrut and J. Karhumäki, Combinatorics of words, edited by G. Rozenberg and A. Salomaa. *Handbook of Formal Languages* **1**, Springer (1997).
[13] W.-f. Chuan, Unbordered factors of the characteristic sequences of irrational numbers. *Theoret. Comput. Sci.* **205** (1998) 337–344.
[14] J.D. Currie and K. Saari, Least periods of factors of infinite words. *RAIRO-Theor. Inf. Appl.* **43** (2009) 165–178.
[15] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics. *Theoret. Comput. Sci.* **183** (1997) 45–82.
[16] X. Droubay, J. Justin and G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.* **255** (2001) 539–553.
[17] A. Glen and J. Justin, Episturmian words: a survey. *RAIRO-Theor. Inf. Appl.* **43** (2009) 403–442.
[18] A. Glen, F. Levé and G. Richomme, Directive words of episturmian words: equivalence and normalization. *RAIRO-Theor. Inf. Appl.* **43** (2009) 299–319.
[19] T. Harju and D. Nowotka, Minimal Duval Extensions. *Int. J. Found. Comput. Sci.* **15** (2004) 349–354.
[20] R. Klette and A. Rosenfeld, Digital straightness – a review. *Discrete Appl. Math.* **139** (2004) 197–230.
[21] M. Lothaire, *Algebraic Combinatorics on Words.* Cambridge University Press (2002).
[22] G. Melançon, Lyndon factorization of Sturmian words. *Discrete Math.* **210** (2000) 137–149.
[23] F. Mignosi and L.Q. Zamboni, A note on a conjecture of Duval and Sturmian words. *RAIRO-Theor. Inf. Appl.* **36** (2002) 1–3.
[24] M. Morse and G.A. Hedlund, Symbolic dynamics. *Amer. J. Math.* **60** (1938) 815–866.
[25] M. Morse and G.A. Hedlund, Symbolic dynamics II: Sturmian trajectories. *Amer. J. Math.* **62** (1940) 1–42.
[26] G. Richomme, Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words. *Theoret. Comput. Sci.* **380** (2007) 393–400.
[27] K. Saari, Everywhere α-repetitive sequences and Sturmian words, in *Proc. CSR 2007. Lect. Notes Comput. Sci.* **4649** (2007) 363–372.
[28] K. Saari, *On the frequency and periodicity of infinite words.* Ph.D. Thesis, University of Turku, TUCS Dissertations 97 (2008).