



**HAL**  
open science

## Look and Say Fibonacci

Patrice Séébold

► **To cite this version:**

Patrice Séébold. Look and Say Fibonacci. *RAIRO - Theoretical Informatics and Applications* (RAIRO: ITA), 2008, 42 (4), pp.729-746. 10.1051/ita:2007060 . lirmm-00184106

**HAL Id: lirmm-00184106**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00184106>**

Submitted on 24 Mar 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## LOOK AND SAY FIBONACCI

PATRICE SÉÉBOLD<sup>1,2</sup>

**Abstract.** The *LS* (*Look and Say*) derivative of a word is obtained by writing the number of consecutive equal letters when the word is spelled from left to right. For example,  $LS(1\ 1\ 2\ 3\ 3) = 2\ 1\ 1\ 2\ 2\ 3$  (two 1, one 2, two 3). We start the study of the behaviour of binary words generated by morphisms under the *LS* operator, focusing in particular on the Fibonacci word.

**Résumé.** La dérivée *LS* d'un mot est obtenue en décrivant les blocs de lettres qui apparaissent quand on épelle le mot. Par exemple,  $LS(1\ 1\ 2\ 3\ 3) = 2\ 1\ 1\ 2\ 2\ 3$  (deux 1, un 2, deux 3). Nous commençons l'étude de la transformation, par l'opération *LS*, des mots binaires engendrés par morphismes. Notre attention se porte ici en particulier sur le mot de Fibonacci.

**1991 Mathematics Subject Classification.** 68R15.

### 1. INTRODUCTION

Between remarkable words often studied for their combinatorial properties, the Fibonacci infinite word plays a central role. It is the basic model for Sturmian words, and almost all the nice properties of the Fibonacci numbers have a counterpart in the Fibonacci infinite word.

Here we study an operation on words which has no counterpart (to our knowledge) in numbers. The *Look and Say derivative* of a word is obtained by writing the number of consecutive equal letters when the word is spelled from left to right. It seems that this operation was first studied by Conway [9]. Other properties were

---

*Keywords and phrases:* Look and Say sequence, Conway, binary words, Fibonacci word, morphisms, Lyndon factorization

<sup>1</sup> LIRMM, Univ. Montpellier 2, CNRS - 161 rue Ada, 34392 Montpellier, France  
e-mail: [Patrice.Seebold@lirmm.fr](mailto:Patrice.Seebold@lirmm.fr)

<sup>2</sup> Département Mathématiques et Informatique Appliquées - UFR4, Université Paul Valéry, Route de Mende, 34199 Montpellier, France

given by Germain-Bonne in a series of three unpublished papers [10–12]. This operation may also be compared to run-length encoding (see, e.g., the recent work of Brlek *et al.* on smooth infinite words [7]).

This is an unusual operation whose behaviour is rather hard to predict in general. The goal of the present paper is to make the very beginning of the study of the behaviour of binary words generated by morphisms under the Look and Say operation. We will consider in particular the case of the Fibonacci word which is interesting and gives some surprising results.

The paper is organized as follows. After some preliminaries, section 3 is dedicated to a description of the Look and Say operation. Some elementary properties, given in section 4, allow to narrow the study down to the case of words over alphabets with two or three letters, and the important notions of chains and cycles are described in section 5. First general results are given in sections 6 and 7. Then the main part of the paper is section 8 which presents the beginning of a study of the behaviour of the Fibonacci infinite word under the action of the *LS* operator.

## 2. PRELIMINARIES

The terminology and notations are mainly those of Lothaire, 2002 [14] and Allouche and Shallit, 2003 [1].

Let  $A$  be an *alphabet*, finite set of *letters*, and  $A^*$  the set of (finite) *words* over  $A$ , free monoid generated by  $A$ . The identity element of  $A^*$  is the *empty word* denoted by  $\varepsilon$ . We denote by  $A^+$  the semigroup  $A^* \setminus \{\varepsilon\}$ .

The *length* of a word  $u$ , denoted by  $|u|$ , is the number of occurrences of letters in  $u$ . In particular  $|\varepsilon| = 0$ .

If  $n$  is a non-negative integer,  $u^n$  is the word obtained by concatenating  $n$  occurrences of the word  $u$ . Of course,  $|u^n| = n|u|$ . The cases  $n = 2$ ,  $n = 3$ , and  $n = 4$  deserve particular attention in what follows. A word  $u^2$  (resp.  $u^3$ ,  $u^4$ ), with  $u \neq \varepsilon$ , is called a *square* (resp. a *cube*, a *4-power*).

A word  $w$  is called a *factor* (resp. a *prefix*, resp. a *suffix*) of  $u$  if there exist words  $x, y$  such that  $u = xwy$  (resp.  $u = wy$ , resp.  $u = xw$ ). The factor (resp. the prefix, resp. the suffix) is *proper* if  $xy \neq \varepsilon$  (resp.  $y \neq \varepsilon$ , resp.  $x \neq \varepsilon$ ).

An *infinite word* (or *sequence*) over  $A$  is an application  $x : N \rightarrow A$ . It is written  $= x_0x_1 \cdots x_i \cdots$ ,  $i \in N$ ,  $x_i \in A$ . The set of infinite words over  $A$  is  $A^\omega$ , and the set of all the words (finite or infinite) over  $A$  is  $A^\infty$ .

The notion of factor is extended to infinite words as follows: a (finite, possibly empty) word  $u$  is a *factor* (resp. *prefix*) of an infinite word  $\mathbf{a}$  over  $A$  if there exist  $n \in N$  (resp.  $n = 0$ ) and  $m \in N$  ( $m = |u|$ ) such that  $u = a_n \cdots a_{n+m-1}$  (by convention  $a_n \cdots a_{n-1} = \varepsilon$ ).

A factor  $u$  of an infinite word  $x$  is *right special* if there exist two different letters of  $A$ , say  $a$  and  $b$ , such that both  $ua$  and  $ub$  are factors of  $x$ . An infinite word over a two-letter alphabet is *Sturmian* if it contains exactly one right special factor of each length.

A (finite or infinite) word  $w$  over  $A$  is *square-free* if it contains no factor  $uu$  with  $u \in A^+$ .

In what follows, we will consider morphisms on  $A$ . Let  $B$  be an alphabet (often,  $B = A$ ).

A *morphism on  $A$*  (in short a *morphism*) is an application  $f : A^* \rightarrow B^*$  such that  $f(uv) = f(u)f(v)$  for all  $u, v \in A^*$ . It is uniquely determined by its value on the alphabet  $A$ . A morphism  $f$  on  $A$  is *nonerasing* if  $f(a) \neq \varepsilon$  for all  $a \in A$ . When  $|f(a)| = 1$  for all  $a \in A$ ,  $f$  is said to be a *literal morphism*.

Now suppose  $A = B$ . A nonerasing morphism is *prolongable on  $x_0$* ,  $x_0 \in A^+$ , if there exists  $u \in A^+$  such that  $f(x_0) = x_0u$ . In this case, for all  $n \in \mathbb{N}$  the word  $f^n(x_0)$  is a proper prefix of the word  $f^{n+1}(x_0)$  and this defines a unique infinite word

$$x = x_0uf(u)f^2(u) \cdots f^n(u) \cdots$$

which is the limit of the sequence  $(f^n(x_0))_{n \geq 0}$ . We write  $x = f^\omega(x_0)$  and say that  $x$  is *generated* by  $f$ . A word  $w \in A^\infty$  is a *fixed point* of a morphism  $f$  on  $A$  if  $f(w) = w$ .

A *D0L-system* is a triple  $G = (A, f, u)$  where  $A$  is an alphabet,  $f$  a morphism on  $A$  and  $u \in A^*$ . An infinite word  $x$  is generated by  $G$  if  $x = (f^k)^\omega(u)$  for some  $k \in \mathbb{N}$ .

A *HD0L-system* is a quintuple  $T = (A, u, f, g, B)$  where  $A$  and  $B$  are alphabets,  $u \in A^+$ ,  $f$  is a nonerasing morphism on  $A$ , prolongable on  $u$ , and  $g$  is a morphism from  $A$  onto  $B$ . An infinite word  $y$  is generated by  $T$  if  $y = g(f^\omega(u))$ . When  $g$  is a literal morphism  $T$  is called a *tag-system*. The name of tag-system comes from the fundamental study of Cobham [8]. Chapter 5 of [18] is dedicated to a deep study of D0L-systems (see also Pansiot, 1983 [16] who used the terminology *extended tag-systems* for HD0L-systems).

### 3. THE LOOK AND SAY DERIVATIVE

Let  $A_k = \{1, \dots, k\}$  be the alphabet whose letters are the first  $k$  positive integers. In what follows we will consider finite or infinite words over  $A_k$ , but no word will have any infinite run of one given letter.

The *Look and Say derivative* of a (finite or infinite) word  $u$  over  $A_k$  is the (finite or infinite) word  $LS(u)$  obtained by describing the consecutive runs of letters in  $u$ .

**Example 3.1.** Let  $u = 3552$ . The description of  $u$  is "one three, two fives, one two" thus  $LS(u) = 132512$ .

Now, let  $u = (12)^\omega$ . The description of  $u$  is "one one, one two, one one, one two, ..." Here,  $LS(u) = (1112)^\omega$ .

**Remark 3.2.** The alphabet on which  $LS(u)$  is written is not necessarily  $A_k$ . For example, if  $u = 1112$  ( $u \in A_2^*$ ) then  $LS(u) = 3112 : LS(u) \in A_3^*$ .

Other clarifications must also be given here.

- When describing the word  $u$ , occurrences of one given letter are collected consecutively (in the order of a reading from left to right), not globally (commutatively) in the whole word.  
For example, if  $u = 3\ 5\ 5\ 2\ 5\ 5\ 5\ 3$  then  $LS(u) = 1\ 3\ 2\ 5\ 1\ 2\ 3\ 5\ 1\ 3$  ( $LS(u) \neq 1\ 2\ 2\ 3\ 5\ 5$ ).
- On the other hand it is always understood that repetitions are collected maximally: in the previous example the description cannot be "one three, one five, one five, one two, ...".

This leads to a formal definition of the Look and Say derivative.

**Definition 3.3.** Let  $(\alpha_j)_{j \geq 1}$  be a sequence of positive integers.

If  $u = a_1^{\alpha_1} a_2^{\alpha_2} \dots a_n^{\alpha_n} \dots$ , where  $a_{i-1} \neq a_i$  and  $a_i \neq a_{i+1}$  for each  $i \geq 2$ , then  $LS(u) = \alpha_1 a_1 \alpha_2 a_2 \dots \alpha_n a_n \dots$ .

As usually  $LS(u) = LS^1(u)$  and, for any positive integer  $k$ ,  $LS^{k+1}(u) = LS(LS^k(u))$ . They are all *descendants* of  $u$ . The word  $\lim_{n \rightarrow \infty} LS^n(u)$  is denoted by  $LS^\omega(u)$ .

## 4. BASIC PROPERTIES

The first basic property is obvious. It indicates that fixed points for the operation  $LS$  must be searched in the set of infinite words.

**Property 4.1.** [9, 10] *The empty word  $\varepsilon$  and the word 22 are the only finite fixed points of  $LS$ .*  $\square$

In the following,  $\varepsilon$  and 22 will be called *trivial words*. An important consequence of the above property is that every non trivial finite word  $u$  has an infinite number of descendants that are all different.

The next two properties indicate that, for the fixed points, the study of the Look and Say derivatives can be reduced to the case of words over a three-letter alphabet: indeed 4-powers can never be created by applying the operation  $LS$ , and the existing letters greater or equal to 4 are rejected to the end of the derivative.

**Property 4.2.** [9, 10] *For every positive integer  $k$  and for every word  $u \in A_k^\infty$ , the word  $LS(u) = a_1^{\alpha_1} a_2^{\alpha_2} \dots a_n^{\alpha_n} \dots$  is such that  $\alpha_i \in \{1, 2, 3\}$  for all  $i \geq 1$ .*

*Proof.* Let  $y^\alpha x^\beta z^\gamma$  be any factor of  $u$  where  $x$ ,  $y$ , and  $z$  are pairwise different letters. In  $LS(u)$  this factor gives  $\alpha y \beta x \gamma z$ . Since  $y \neq x$  and  $z \neq x$ , the greatest possible power is  $x^3$  when  $\beta = \gamma = x$ .  $\square$

A direct consequence of Property 4.2 is the following.

**Property 4.3.** *Let  $u \in A_k^\infty$  be such that  $LS(u) \in A_3^\infty$ . Then  $LS^{n+1}(u) \in A_3^\infty$  for every positive integer  $n$ .*  $\square$

**Corollary 4.4.** *For every positive integer  $k$  and for every non trivial word  $u \in A_k^\infty$ ,  $LS^\omega(u) \in A_3^\omega$ .*  $\square$

## 5. CHAINS AND CYCLES

All the results of this section are due to B. Germain-Bonne [11]. They are given without proof.

## 5.1. CHAINS

The operator  $LS$  is of course not a morphism: for example,  $LS(11) = 21$  but  $LS(1)LS(1) = 1111$ . Conway [9] and Germain-Bonne [11] have designed sets of elements on which  $LS$  acts as a morphism. The study of these sets (92 elements in Conway, 1987 [9], 91 elements in Germain-Bonne, 1993 [11]!) is out of our purpose. But in [11], Germain-Bonne has introduced the notions of sub-chains and half-chains on which  $LS$  acts partially as a morphism, and that will be useful in the present paper.

Let  $D$  be the following set of eight words

$$D = \{111, 123, 131, 132, 3112, 3113, 312, 321\}.$$

A *sub-chain* is any finite word over the alphabet  $A_3$  which starts with  $d \in D$  and ends by the letter 2.

A *half-chain* is any infinite word starting with a factor  $d \in D$ .

An important result is the following.

**Lemma 5.1.** *Let  $w$  be a word such that  $w = uv$  with  $u$  a sub-chain and  $v$  a half-chain. Then  $LS^3(u)$  is a sub-chain (longer than  $u$ ),  $LS^3(v)$  is a half-chain, and  $LS^3(w) = LS^3(u)LS^3(v)$ .  $\square$*

This result is useful because it indicates that, when a sub-chain  $u$  is followed by a half-chain  $v$ , the value of the image of  $u$  by  $LS^3$  does not depend on the value of  $v$ . Moreover, since this property is again true for  $LS^3(u)$ , the behaviour of  $LS^\omega(uv)$  is the same as this of  $LS^\omega(u)$ : for every positive integer  $n$ ,  $LS^{3n}(uv)$  starts with  $LS^{3n}(u)$ , which implies that  $LS^\omega(uv)$  and  $LS^\omega(u)$  have an infinitely long common prefix. We say that  $LS^\omega(uv)$  *tends towards*  $LS^\omega(u)$ .

**Remark 5.2.** The above property seems "natural" (since  $uv$  starts with  $u$  then  $LS^\omega(uv)$  starts with  $LS^\omega(u)$ ). However, since  $LS$  is not a morphism, this is not true in general. For example, consider the words 31 and 312.

$LS^7(31)$  starts with 1321132132,

$LS^7(312)$  starts with 312321.

Since the words 1321132 and 312 are sub-chains, and words starting with 132 and 321 are half-chains, this implies that 31 and 312 respectively tends towards the words  $X_1$  and  $X_3$  defined below, that are completely different.

## 5.2. CYCLES

An exhaustive study of the behaviour of all the beginnings of words over  $A_3$  indicates that the operators  $LS$  or  $LS^2$  admit no infinite fixed point. However  $LS^3$  admits twelve fixed points (divided into four cycles of order 3).

Let  $n$  be a positive integer. A *cycle of order  $n$*  is a set of  $n$  infinite words  $\{u, LS(u), \dots, LS^{n-1}(u)\}$  such that for all  $i$ ,  $0 \leq i \leq n-1$ ,  $LS^{n+i}(u) = LS^i(u)$ .

- Let  $x_1 = 1\ 3\ 2\ 1\ 1\ 3\ 2$ , and let  $y_1$  be such that  $LS^3(x_1) = x_1 y_1$  ( $y_1 = 1\ 3\ 2\ 2\ 1\ 1\ 3\ 3\ 1\ 1\ 2$ ).

$$\text{Let } X_1 = x_1 y_1 LS^3(y_1) LS^6(y_1) \dots = x_1 \prod_{i=0}^{\infty} LS^{3i}(y_1) \text{ and } X_3 = 2\ 2\ X_1.$$

The sets  $\{X_1, LS(X_1), LS^2(X_1)\}$  and  $\{X_3, LS(X_3), LS^2(X_3)\}$  are cycles (of order 3).

- Let  $x_2 = 3\ 1\ 2$ , and let  $y_2$  be such that  $LS^3(x_2) = x_2 y_2$  ( $y_2 = 3\ 2\ 1\ 1\ 2$ ).

$$\text{Let } X_2 = x_2 y_2 LS^3(y_2) LS^6(y_2) \dots = x_2 \prod_{i=0}^{\infty} LS^{3i}(y_2) \text{ and } X_4 = 2\ 2\ X_2.$$

The sets  $\{X_2, LS(X_2), LS^2(X_2)\}$  and  $\{X_4, LS(X_4), LS^2(X_4)\}$  are cycles (of order 3).

The following result indicates that these four cycles are the only one.

**Property 5.3.** *Let  $u$  be an infinite word over  $A_3$ . If there exists a positive integer  $n$  such that  $LS^n(u) = u$  then  $n = 3$  and  $u \in \bigcup_{j=1}^4 \{X_j, d(X_j), d^2(X_j)\}$ .  $\square$*

Two natural questions arise:

- does every word tend towards some cycle after applying  $LS$  an infinite number of times?
- in case of a positive answer to the previous question, in which cycle falls a given word after applying  $LS$  an infinite number of times?

**Remark 5.4.** Each of the four sets  $\{X_j, d(X_j), d^2(X_j)\}$  is a *cycle*. Since  $LS$  admits no infinite fixed points, the word  $LS^\omega(u)$  is never defined (except if  $u = \varepsilon$  or  $u = 2\ 2$ , see Property 4.1). But the words  $\lim_{n \rightarrow \infty} LS^{3n}(u)$ ,  $\lim_{n \rightarrow \infty} LS^{3n+1}(u)$ , and  $\lim_{n \rightarrow \infty} LS^{3n+2}(u)$  could be well defined, and if one of these limits is  $X_j$  ( $j \in \{1, 2, 3, 4\}$ ) then the two others are necessarily  $LS(X_j)$  and  $LS^2(X_j)$  for the same value of  $j$ . In this case we say that the limit of  $u$  is the cycle  $X_j$  and, since in what follows we only need to know in which cycle falls the words  $\lim_{n \rightarrow \infty} LS^{3n}(u)$ ,  $\lim_{n \rightarrow \infty} LS^{3n+1}(u)$ , and  $\lim_{n \rightarrow \infty} LS^{3n+2}(u)$ , we will write  $LS^\omega(u) = X_j$  to indicate that  $\{\lim_{n \rightarrow \infty} LS^{3n}(u), \lim_{n \rightarrow \infty} LS^{3n+1}(u), \lim_{n \rightarrow \infty} LS^{3n+2}(u)\} = \{X_j, LS(X_j), LS^2(X_j)\}$ .

In the next two sections we answer the above questions in the cases of unary and binary words. The general case (words over a three-letter alphabet) will be discussed later.

## 6. DESCENDANTS OF UNARY WORDS

Let  $A_1 = \{1\}$ . The only words over  $A_1$  are powers of the letter 1. Since our assumption is that no word contains any infinite run of one given letter we only consider, in this section, finite words  $u = 1^n$ ,  $n \geq 1$ .

We start with a general property.

**Lemma 6.1.** *For every integer  $n$ ,  $n \geq 2$ ,  $LS^\omega(1 n 1 1) = X_1$ .*

*Proof.* Since  $n \geq 2$ , the word  $LS^{12}(1 n 1 1)$  starts with  $x_1 y_1$ . But  $x_1$  is a sub-chain and  $y_1$  is the beginning of a half-chain. Thus by Lemma 5.1, for every non-negative integer  $n$ ,  $LS^{3n+12}(1 n 1 1)$  starts with  $LS^{3n}(x_1)$ . Since, by Corollary 4.4,  $LS^{3n}(x_1)$  tends towards  $X_1$ , the result is proved.  $\square$

The characterization of descendants of unary words is a direct consequence of this lemma.

**Proposition 6.2.** *For every word  $u \in A_1^+$ ,  $LS^\omega(u) = X_1$ .*

*Proof.* If  $u \in A_1^+$  then  $u = 1^n$  for some positive integer  $n$ .

- If  $n \geq 2$  then  $LS(u) = n 1$  and  $LS^2(u) = 1 n 1 1$ .
- If  $n = 1$  then  $LS(u) = 1 1$ ,  $LS^2(u) = 2 1$ , and  $LS^3(u) = 1 2 1 1$ .

In both cases, from Lemma 6.1,  $LS^\omega(u) = X_1$ .  $\square$

## 7. DESCENDANTS OF BINARY WORDS

Let  $A_2 = \{1, 2\}$ . Here words can be infinite and many different cases have to be considered. The following proposition is the equivalent, in the binary case, of Proposition 6.2.

**Proposition 7.1.** *Let  $u$  be a word over  $A_2$ .*

- *If  $u$  starts with  $1 1 1 2 2 2$ ,  $2 2 2 1 1 2 1$ ,  $2 2 1 1 2 1 2$ , or  $2 2 1 1 2 1 1 1$  then  $LS^\omega(u) = X_2$  ;*
- *if  $u$  starts with  $2 2 1 1 1 2 2 2$  then  $LS^\omega(u) = X_4$  ;*
- *if  $u$  starts with  $2 2 1 2$ ,  $2 2 1 1 1 1 1$ ,  $2 2 1 1 1 2 1$ , or  $2 2 1 1 1 2 2 1$  then  $LS^\omega(u) = X_3$  ;*
- *otherwise  $LS^\omega(u) = X_1$ .*

The whole proof of this proposition is long, tedious, and repetitive. We only give below the proof of the first two items. It is based on the following corollary of Lemma 5.1.

**Lemma 7.2.** *Let  $u \in A_3^\infty$ .*



- If  $u$  starts with 3 1 2 3 then  $LS^\omega(u) = X_2$ .
- If  $u$  starts with 2 2 3 1 2 3 then  $LS^\omega(u) = X_4$ .

*Proof.* Let  $u = 3 1 2 3 u'$ .

- If  $u'$  does not start with 3 then
  - $LS(u)$  starts with 1 3 1 1 1 2 1 3
  - $LS^2(u)$  starts with 1 1 1 3 3 1 1 2 1 1
  - $LS^3(u)$  starts with 3 1 2 3 2 1 1 2
- If  $u'$  starts with 3 3 then
  - $LS(u)$  starts with 1 3 1 1 1 2  $a$ ,  $a \neq 2$
  - $LS^2(u)$  starts with 1 1 1 3 3 1 1 2
  - $LS^3(u)$  starts with 3 1 2 3 2 1
  - $LS^4(u)$  starts with 1 3 1 1 1 2 1 3 1 2
  - $LS^5(u)$  starts with 1 1 1 3 3 1 1 2 1 1 1 3 1 1
  - $LS^6(u)$  starts with 3 1 2 3 2 1 1 2 3 1 1 3
- Otherwise  $u'$  starts with  $3b$ ,  $b \neq 3$ , then
  - $LS(u)$  starts with 1 3 1 1 1 2 2 3
  - $LS^2(u)$  starts with 1 1 1 3 3 1 2 2
  - $LS^3(u)$  starts with 3 1 2 3 1 1
  - $LS^4(u)$  starts with 1 3 1 1 1 2 1 3
  - $LS^5(u)$  starts with 1 1 1 3 3 1 1 2 1 1
  - $LS^6(u)$  starts with 3 1 2 3 2 1 1 2

In both cases  $LS^6(u)$  starts with  $x_2y_2$ , and  $x_2$  is a sub-chain and  $y_2$  is the beginning of a half-chain. Thus by Lemma 5.1, for every non-negative integer  $n$ ,  $LS^{3n+6}(u)$  starts with  $LS^{3n}(x_2)$ . Since, by Corollary 4.4,  $LS^{3n}(x_2)$  tends towards  $X_2$ , the first part is proved.

For the second part, remark that in the previous case, for any positive integer  $n$ ,  $LS^n(u)$  never starts with the letter 2. Thus, since 2 2 is a fixed point of  $LS$ , if  $u$  starts with 2 2 3 1 2 3 then  $LS^{3n+6}(u)$  starts with 2 2  $LS^{3n}(x_2)$  which tends towards 2 2  $X_2 = X_4$ .  $\square$

*Proof of the first two items of Proposition 7.1.*

- If  $u = 1 1 1 2 2 2 u'$  then
  - $LS(u)$  starts with 3 1  $n$  2,  $n \geq 3$
  - $LS^2(u)$  starts with 1 3 1 1 1  $n$  (because  $n \neq 2$ )
  - $LS^3(u)$  starts with 1 1 1 3 3 1 (because  $n \neq 1$ )
  - $LS^4(u)$  starts with 3 1 2 3
- If  $u = 2 2 2 1 1 2 1 u'$  then
  - $LS(u)$  starts with 3 2 2 1 1 2
  - $LS^2(u)$  starts with 1 3 2 2 2 1
  - $LS^3(u)$  starts with 1 1 1 3 3 2
  - $LS^4(u)$  starts with 3 1 2 3
- If  $u = 2 2 1 1 2 1 2 u'$  or  $u = 2 2 1 1 2 1 1 1 u'$  then
  - $LS(u)$  starts with 2 2 2 1 1 2  $n$  1,  $n \neq 2$
  - $LS^2(u)$  starts with 3 2 2 1 1 2 (because  $n \neq 2$ )

$LS^3(u)$  starts with 1 3 2 2 2 1  
 $LS^4(u)$  starts with 1 1 1 3 3 2  
 $LS^5(u)$  starts with 3 1 2 3

In both cases  $LS^k(u)$  starts with 3 1 2 3 for some positive integer  $k$ . Then, by Lemma 7.2,  $LS^\omega(u) = X_2$ .

Now, if  $u$  starts with 2 2 1 1 2 2 2 then, from above,  $LS^4(u)$  starts with 2 2 3 1 2 3 which implies, from Lemma 7.2, that  $LS^\omega(u) = X_4$ .  $\square$

An interesting corollary of Proposition 7.1 will make the transition with the next section. It indicates that almost all the Sturmian words tend towards the same cycle.

**Corollary 7.3.** *Let  $u$  be a Sturmian word over  $A_2$ .*

- *If  $u$  starts with 2 2 1 then  $LS^\omega(u) = 2 2 X_1$ .*
- *Otherwise  $LS^\omega(u) = X_1$ .*

 $\square$ 

## 8. THE FIBONACCI CASE

After giving general properties about the behaviour of binary words under the action of the operator  $LS$ , we now turn to a classical particular case of binary word, the Fibonacci word  $F$ , and we give some interesting results about the descendants of  $F$ .

Let  $\varphi$  be the morphism on  $A_2$  defined by  $\varphi(1) = 12$ ,  $\varphi(2) = 1$ . The *Fibonacci word* is the infinite word  $F = \varphi^\omega(1)$ .

$$F = 1211212112112112112112112112112112112112\dots$$

From Corollary 7.3 we know that  $LS^\omega(F) = X_1$ .

Let  $F' = LS(F)$  be the first derivative of  $F$ . The word  $F'$  has interesting properties that we start studying here.

$$F' = 111221121112211221121112211211122112\dots$$

### 8.1. GENERATING $F'$

The first problem we consider is to determine how  $F'$  is generated.

Generating a word with a HD0L-system consists of applying a morphism to an infinite word generated by another morphism. Berstel [3] gave an example showing that the power of generation of HD0L-systems is greater than the power of generation of D0L-systems (only one morphism is applied): he proved that the Arshon word (a square-free word over a 3-letter alphabet, see Arshon, 1937 [2], Séébold, 2002 [20]) is generated by a tag-system (a particular case of an HD0L-system, see Cobham, 1972 [8]) whereas it cannot be obtained with a D0L-system. Here we prove that  $F'$  is a new example of this phenomenon.

We start by giving an HD0L-system to generate  $F'$ .

**Lemma 8.1.**  $F'$  is generated by the HD0L-system  $\langle A_2, \tilde{\varphi}\varphi, 2, d', A_2 \rangle$  where  $d'$  is the morphism defined on  $A_2$  by  $d'(1) = 2112$ ,  $d'(2) = 1112$  :  $F' = d'[(\tilde{\varphi}\varphi)^\omega(2)]$ .

*Proof.* Let us recall that  $\tilde{\varphi}$  is the morphism on  $A_2$  defined by  $\tilde{\varphi}(1) = 21$ ,  $\tilde{\varphi}(2) = 1$ , and consider the two morphisms on  $A_2$

$$\varphi\tilde{\varphi} : \begin{array}{l} 1 \mapsto 112 \\ 2 \mapsto 12 \end{array} \quad \text{and} \quad \tilde{\varphi}\varphi : \begin{array}{l} 1 \mapsto 211 \\ 2 \mapsto 21 \end{array}$$

The following equality is proved in Berstel and Séébold, 1994 [6].

$$2F = (\tilde{\varphi}\varphi)^\omega(2). \quad (8.1)$$

Now  $\varphi\varphi\tilde{\varphi} = \tilde{\varphi}\tilde{\varphi}\varphi$  thus, for every non-negative integer  $n$ ,  $\varphi\tilde{\varphi}(\tilde{\varphi}\varphi)^n = \varphi^{2n+1}\tilde{\varphi}$ .

$$\begin{aligned} \text{Consequently } \varphi\tilde{\varphi}(2F) &= \varphi\tilde{\varphi}[(\tilde{\varphi}\varphi)^\omega(2)] \\ &= \varphi\tilde{\varphi}(\lim_{n \rightarrow \infty} (\tilde{\varphi}\varphi)^n(2)) \\ &= \lim_{n \rightarrow \infty} \varphi\tilde{\varphi}(\tilde{\varphi}\varphi)^n(2) \\ &= \lim_{n \rightarrow \infty} \varphi^{2n+1}\tilde{\varphi}(2) \\ &= \varphi^\omega[\tilde{\varphi}(2)] \\ &= \varphi^\omega(1) \\ &= F \end{aligned}$$

which implies that  $F$  can be decomposed over  $\{12, 112\}$ .

In  $F$ , each occurrence of the letter 2 is preceded and followed by the letter 1. So each occurrence of 2 in  $2F$  corresponds to an occurrence of 12 in  $F'$ . Thus, since  $F$  decomposes over  $\{12, 112\}$ , each occurrence of 12 in  $F$  corresponds to a factor 1112 in  $F'$ , and each occurrence of 112 in  $F$  corresponds to a factor 2112 in  $F'$ .

Consequently  $F' = d(F)$  where  $d$  is the substitution on  $A_2$  defined by  $d(12) = 1112$  and  $d(112) = 2112$ .

But, since from what precedes  $F = \varphi\tilde{\varphi}(2F)$ , the previous equality  $F' = d(F)$  implies that  $F' = d[\varphi\tilde{\varphi}(2F)]$ .

Let  $d' = d\varphi\tilde{\varphi} : d'(1) = 2112$ ,  $d'(2) = 1112$ .

We deduce, using equation (8.1), that  $F' = d'[(\tilde{\varphi}\varphi)^\omega(2)]$ .  $\square$

It is known from Cobham, 1972 [8] (for a proof, see Pansiot, 1983 [16], or Allouche and Shallit, 2003 [1]) that if an infinite word is generated by an HD0L-system then there exists a tag-system which generates this word. Here we prove a little more.

**Theorem 8.2.** *The word  $F'$  is generated by a tag-system, but it cannot be generated by a D0L-system.*

**Remark 8.3.** In [16], Pansiot gave an algorithm to transform an HD0L-system generating a given infinite word in a tag-system generating the same word. Here we do not use Pansiot's algorithm; we construct directly a (more simple) tag-system to generate  $F'$ .

*Proof of Theorem 8.2.*

We first give a tag-system generating  $F'$ .

Let  $a, b$  be two new letters and let  $A'_2 = \{1, 2, a, b\}$ . Let  $h$  be the morphism defined on  $A'_2$  by  $h(2) = 2aab1aab$ ,  $h(1) = 2aab1aab1aab$ , and  $h(a) = h(b) = \varepsilon$ . To end, let  $c$  be the morphism from  $A'_2$  onto  $A_2$  defined by  $c(2) = c(a) = 1$  and  $c(1) = c(b) = 2$ . We will prove that  $F'$  is generated by the tag-system  $\langle A'_2, h, 2, c, A_2 \rangle$ , i.e.,  $F' = c(h^\omega(2))$ .

First remark that for every non-negative integer  $n$ , since  $h^n(aab) = \varepsilon$ ,

$$h^n(2aab) = h^n(2) \text{ and } h^n(1aab) = h^n(1). \quad (8.2)$$

Let  $R$  be the morphism from  $A'_2$  onto  $A_2$  which erases the letters  $a$  and  $b$ :  $R(1) = 1$ ,  $R(2) = 2$ ,  $R(a) = R(b) = \varepsilon$ .

We first prove that  $R(h^\omega(2)) = 2F$ .

From equation (8.1), it is enough to prove that  $R(h^\omega(2)) = (\tilde{\varphi}\varphi)^\omega(2)$ .

To do it we prove by induction that, for every non-negative integer  $n$ ,

$$R(h^n(1)) = (\tilde{\varphi}\varphi)^n(1) \text{ and } R(h^n(2)) = (\tilde{\varphi}\varphi)^n(2).$$

This is of course true if  $n = 0$ .

$$\begin{aligned} \text{Now, } R(h^{n+1}(2)) &= R(h^n(2aab)) \\ &= R(h^n((2aab))R(h^n(1aab))) \\ &= R(h^n(2))R(h^n(1)) \text{ (from (8.2))} \\ &= (\tilde{\varphi}\varphi)^n(2)(\tilde{\varphi}\varphi)^n(1) \text{ (by induction)} \\ &= (\tilde{\varphi}\varphi)^n(21) \\ &= (\tilde{\varphi}\varphi)^n[(\tilde{\varphi}\varphi)(2)] \\ &= (\tilde{\varphi}\varphi)^{n+1}(2). \end{aligned}$$

The same is done for  $R(h^{n+1}(1)) = (\tilde{\varphi}\varphi)^{n+1}(1)$ .

Now, let  $S$  be the morphism from  $A_2$  onto  $A'_2$  which adds  $aab$  after each letter:  $S(1) = 1aab$ ,  $S(2) = 2aab$ .

Since  $RS = Id$  and  $R[h^\omega(2)] = 2F$ , one has  $R[S(2F)] = R[h^\omega(2)]$  so  $S(2F) = h^\omega(2)$ .

But, as we saw in the proof of Lemma 8.1,  $F' = d'(2F)$  where  $d'(1) = 2112$ ,  $d'(2) = 1112$ .

Since  $d' = cS$ , one has then  $F' = cS(2F) = c(h^\omega(2))$ .

Now, we prove that  $F'$  cannot be generated by a D0L-system.

Let us suppose that  $F'$  is generated by a D0L-system  $\langle A_2, f, u \rangle$ .

Since  $F'$  starts with the cube 111, it would start with  $f(1)f(1)f(1)$ .

If  $f(1) = 1$  then it is clear that  $f(2)$  must start with 2 and end with 22, and then  $f(21112)$  contains 221112 as a factor. This is impossible: indeed 221112 is not a factor of  $F'$  but, since 21112 is a factor of  $F'$ , if  $F'$  was generated by the D0L-system  $\langle A_2, f, u \rangle$  then  $f(21112)$  should be a factor of  $F'$ .

Thus  $|f(1)| \geq 2$  which means that  $F'$  would start with a cube  $X^3$  with  $|X| \geq 2$ .

In this case  $X$  starts with 111 and thus, necessarily,  $X \in \{1112, 2112\}^+$ . This implies (since  $F' = d(F)$  with  $d(12) = 1112$ ,  $d(112) = 2112$ ) that  $F$  would start with a factor  $Y^3$  with  $Y \in \{12, 112\}^+$ .

Let us suppose that  $Y$  is the smallest word over  $\{1\,2, 1\,1\,2\}^+$  such that  $Y^3$  is a prefix of  $F$ . Since  $1\,2 = \varphi(1)$  and  $1\,1\,2 = \varphi(2\,1)$ , there exists  $y \in A_2^*$  such that  $Y = \varphi(1\,y\,1)$ . By injectivity of  $\varphi$ , this implies that  $F$  starts with  $1\,y_1\,1\,1\,y_2\,1\,1\,y_3\,1$  ( $y_1 = y_2 = y_3 = y$ ).

Of course, since  $F$  does not start with  $1\,1\,1\,1\,1\,1$ , one has  $y \neq \varepsilon$  and, since  $F = \varphi(F)$  and by definition of  $\varphi$ , there exists  $z \in A_2^+$  such that  $1\,y_1\,1 = 1\,y_2\,1 = \varphi(z)$ .

The word  $z$  necessarily ends with  $2$  thus there exists  $z' \in A_2^+$  such that  $z = z'2$  and, since  $y_3 = y$ ,  $1\,y_1\,1\,1\,y_2\,1\,1\,y_3 = \varphi(z'2z'2z')$  and  $z'2z'2z'$  is a prefix of  $F$ .

It is easily verified that  $|z'| \geq 3$ . Moreover, since  $z'$  is a prefix of  $F$ ,  $z'$  starts with  $1\,2$ .

If the prefix  $z'2z'2z'$  is followed in  $F$  by the letter  $1$  then  $z'$  cannot end with  $1\,1$  (because  $F$  does not contain  $1\,1\,1$  as a factor) and, since  $2\,2$  is not a factor of  $f$ ,  $z'$  ends with  $2\,1$ . But in this case, since  $z'$  starts with  $1\,2$ ,  $z'2z'$  contains  $2\,1\,2\,1\,2$  as a factor which is impossible because  $2\,1\,2\,1\,2$  is not a factor of  $F$ .

Consequently  $F$  starts with  $z'2z'2z'2$ . But in this case  $Y^3 = \varphi(1\,y\,1\,1\,y\,1\,1\,y\,1) = \varphi^2(z'2z'2z'2)$  which implies that  $|z'2| < |Y|$ . Since  $F$  decomposes over  $\{1\,2, 1\,1\,2\}$ , we conclude that  $z'2 \in \{1\,2, 1\,1\,2\}^+$ . Thus  $z'2$  is a word of  $\{1\,2, 1\,1\,2\}^+$  such that  $(z'2)^3$  is a prefix of  $F$ , which contradicts the minimality of  $Y$ .

Consequently  $F'$  does not start with another cube than  $1\,1\,1$  which contradicts the existence of a morphism  $f \neq Id$  such that  $F'$  starts with  $f(1)f(1)f(1)$ :  $F'$  cannot be generated by a DOL-system.  $\square$

## 8.2. THE COMPLEXITY OF $F'$

Another interesting question is that of the subword complexity of an infinite word, i.e., the number of factors of each length in this word. It gives a sort of measure of the randomness of an infinite word: the lower is the growing of its subword complexity, the smaller is its randomness. For an overview of this notion see, e.g., Allouche and Shallit, 2003 [1, Chapter 10].

Formally, the *complexity function* of an infinite word  $u$  is the function  $P_u$  which gives for each non-negative integer  $n$  the number  $P_u(n)$  of different factors of length  $n$  in  $u$ . We will prove the following.

**Theorem 8.4.**  $P_{F'}(0) = 1$ ,  $P_{F'}(1) = 2$ ,  $P_{F'}(2) = 4$ ,  $P_{F'}(3) = 6$ ,  $P_{F'}(n) = n + 4$ ,  $n \geq 4$ .

In order to prove this, we state two intermediate results.

**Lemma 8.5.** *Every right special factor of  $F'$  of length greater than or equal to 4 ends with  $2\,1\,1\,2$ .*

*Proof.* First recall that  $F' = d'(2\,F)$  where  $d'(1) = 2\,1\,1\,2$  and  $d'(2) = 1\,1\,1\,2$ .

Let  $X' \in A_2^*$  be such that  $X'X$  is a prefix of  $F'$  with  $|X| = 4$ . Since  $F'$  does not contain the factors  $1\,1\,1\,1$ ,  $1\,2\,1\,2$ ,  $1\,2\,2\,2$ ,  $2\,1\,2\,1$ ,  $2\,1\,2\,2$ ,  $2\,2\,1\,2$ ,  $2\,2\,2\,1$ , and  $2\,2\,2\,2$ ,  $X$  can take only eight values.

- If  $X = 1112$  then necessarily  $X'X = d'(x2)$  where  $x2$  is a prefix of  $F$  and, since  $22$  is not a factor of  $F$ ,  $x2$  is followed in  $F$  by a  $1$  thus  $X$  is followed in  $F'$  by a  $2$ .
- If  $X = 1121$  or  $X = 1122$  then  $X' = d'(x)1$  or  $X' = d'(x)2$  because  $112$  can be only the end of  $d'(1)$  or  $d'(2)$ . Thus the last letter of  $X$  is necessarily the first letter of some  $d'(1)$  or  $d'(2)$ , and is then followed in  $F'$  by a  $1$ .
- If  $X = 1211$  or  $X = 1221$  then, as above,  $12$  is necessarily the end of some  $d'(x)$  and the last two letters of  $X$  are at the beginning of  $d'(1)$  or  $d'(2)$ , thus followed in  $F'$  by a  $1$ .
- If  $X = 2111$  or  $X = 2211$  then, as above,  $2$  is necessarily the end of some  $d'(x)$  and the last three letters of  $X$  are at the beginning of  $d'(1)$  or  $d'(2)$ , thus followed in  $F'$  by a  $2$ .

In both of these seven cases the factor  $X$  has only one possible extension thus  $X'X$  is not right special.

On the other hand, since  $21121$  and  $21122$  are both factors of  $F'$ ,  $2112$  is a right special factor of  $F'$ .

Consequently, the only possibility for a word to be a right special factor of  $F'$  is that this word ends with  $2112$ .  $\square$

The second result indicates that  $F'$  contains actually right special factors.

**Proposition 8.6.** *The word  $F'$  contains exactly one right special factor of each length  $n$ , for each integer  $n \geq 4$ .*

*Proof.* The word  $F$  contains exactly one right special factor of each length (Berstel, 1980 [4]).

Let  $x$  be such a factor:  $x1$  and  $x2$  are both factors of  $F$ . Thus  $d'(x1) = d'(x)2112$  and  $d'(x2) = d'(x)1112$  are both factors of  $F'$ , which implies that each suffix of  $d'(x)$  is a right special factor of  $F'$ . Consequently  $F'$  contains at least one right special factor of each length  $n$ ,  $n \geq 4$ .

Now, suppose that  $F'$  contains two different right special factors of length at least 4, say  $X$  and  $Y$  such that  $|X| = |Y|$ . From Lemma 8.5,  $X$  and  $Y$  both end with  $2112$  thus they are suffixes of some  $d'(x1)$  and  $d'(y1)$  where  $x1$  and  $y1$  are factors of  $F$ , and  $|x| = |y|$ . But in this case  $x1$  and  $y1$  are two right special factors of  $F$  of the same length, thus are equal which contradicts  $X \neq Y$ .  $\square$

*Proof of Theorem 8.4.*

Since  $F'$  decomposes over  $\{1112, 2112\}$  it contains all the possible factors of length smaller than or equal to 3, except  $212$  and  $222$ . Thus  $P_{F'}(0) = 1$ ,  $P_{F'}(1) = 2$ ,  $P_{F'}(2) = 4$ , and  $P_{F'}(3) = 6$ .

By Proposition 8.6,  $F'$  contains exactly one right special factor of length  $n$  for every integer  $n \geq 4$ , thus one has  $P_{F'}(n+1) = P_{F'}(n) + 1$  for every  $n \geq 4$ .

Since we have seen in the proof of Lemma 8.5 that  $F'$  contains exactly eight factors of length 4, one has  $P_{F'}(n) = n + 4$  when  $n = 4$ . Thus  $P_{F'}(n) = n + 4$  for every  $n \geq 4$ .  $\square$

8.3. LYNDON FACTORIZATIONS OF DESCENDANTS OF  $F$ 

In section 8.1, we have seen that for the generating process  $F'$  is different from  $F$  because  $F'$  cannot be generated by one morphism when  $F$  is.

But, in section 8.2, we have seen that for the subword complexity  $F'$  is comparable to  $F$ : for  $n \geq 4$ ,  $P_{F'}(n+1) = P_{F'}(n) + 1$  and also  $P_F(n+1) = P_F(n) + 1$ .

In the present section, we will see that  $F'$  is again different from  $F$  when we deal with the lexicographic order  $\prec$ .

In what follows we consider a two-letter alphabet  $A = \{a, b\}$ , totally ordered by  $a \prec b$ .

Before examining  $F'$  we need to recall some definitions and results.

Let  $u, v$  be two finite words over  $A$ . The word  $u$  is *lexicographically smaller than* the word  $v$  ( $u \prec v$ ) if

- either  $u$  is a proper prefix of  $v$ ,
- or there exist words  $w, u', v' \in A^*$  such that  $u = wau'$  and  $v = wbv'$ .

For infinite words, only the second case is possible.

Let  $u, v$  be two infinite words over  $A$ . The word  $u$  is *lexicographically smaller than* the word  $v$  ( $u \prec v$ ) if there exists a word  $w \in A^*$  and two words  $u', v' \in A^\omega$  such that  $u = wau'$  and  $v = wbv'$ .

Regarding the lexicographic order, a very beautiful (and widely studied) notion is that of Lyndon words (for an introduction, see Lothaire, 1983 [13]).

A (finite or infinite) word  $u \in A^\omega$  is a *Lyndon word* if it is lexicographically smaller than all its proper suffixes.

Here we are more specifically interested in infinite words. The following well known result is fundamental.

**Theorem 8.7.** [21] *Any infinite word  $x$  may be uniquely expressed as a non-increasing product of Lyndon words, finite or infinite, in one of the two following forms:*

either  $x = \prod_{k \geq 0} l_k$  where  $(l_k)_{k \geq 0}$  is an infinite non-increasing sequence of finite Lyndon words (1)

or  $x = l_0 \cdots l_{m-1} y$  where  $l_0, \dots, l_{m-1}$ ,  $m \geq 0$ , is a (perhaps empty) finite non-increasing sequence of finite Lyndon words, and  $y$  is an infinite Lyndon word,  $y \prec l_{m-1}$ . (2)

□

In the first case the word  $x$  has a *Lyndon factorization of type (1)*, in the second case  $x$  has a Lyndon factorization of type (2).

An important question is then to know the type of the Lyndon factorization of a given infinite word.

Melançon [15] has proved the following beautiful result.

**Theorem 8.8.** *The Lyndon factorization of the Fibonacci word  $F$  is of type (1):*

$$F = \prod_{n \geq 0} (\varphi\tilde{\varphi})^n(12)$$

where, for every non-negative integer  $n$ ,  $(\varphi\tilde{\varphi})^n(12)$  is a finite Lyndon word and  $(\varphi\tilde{\varphi})^{n+1}(12) \prec (\varphi\tilde{\varphi})^n(12)$ .  $\square$

Here we prove that the situation is different for  $F'$ .

**Theorem 8.9.** *The Lyndon factorization of the word  $F'$  is of type (2):  $F'$  is an infinite Lyndon word.*

The proof of this theorem is based on two results.

The first one is a restriction in the case of a two-letter alphabet of a strong general result of Richomme [17] characterizing the morphisms which preserve infinite Lyndon words.

A morphism  $f$  on  $A$  preserves infinite Lyndon words if  $f(u)$  is an infinite Lyndon word whenever  $u \in A^\omega$  is an infinite Lyndon word.

**Theorem 8.10.** [17] *A non-erasing morphism  $f$  on  $A$  preserves infinite Lyndon words if and only if*

- $f(a) \prec f(b)$ ,
- $f(a)$  is a power of a Lyndon word,
- $f(ab)$  is a Lyndon word.  $\square$

It is important to note that no condition is needed on the alphabet on which are written  $f(a)$  and  $f(b)$ , except that this alphabet must be totally ordered. In particular, it is possible that this alphabet be also  $A$ , but with  $b \prec a$ .

The second useful result is a lemma about characteristic Sturmian words of which a proof can be found in Berstel and Séébold, 1993 [5].

A Sturmian word  $u$  over  $A$  is *characteristic* if both  $au$  and  $bu$  are Sturmian words.

**Lemma 8.11.** [5] *Let  $u \in A^\omega$ . If  $u$  is a characteristic Sturmian word then  $au$  is lexicographically smaller than all its proper suffixes.*  $\square$

*Proof of Theorem 8.9.*

Here, we order the alphabet  $A_2$  with  $2 \prec 1$  (thus, in all the results given above, the letter  $a$  is replaced by 2 and the letter  $b$  is replaced by 1).

- Since the words  $1F$  and  $2F$  are generated by Sturmian morphisms (see Berstel and Séébold, 1994 [6]), they are both Sturmian words thus  $F$  is a characteristic Sturmian word.

Thus, from Lemma 8.11, the word  $2F$  is lexicographically smaller than all its proper suffixes: it is a Lyndon word.

- Now, we have seen in the proof of Lemma 8.1 that  $F' = d'(2F)$  where  $d'$  is the morphism defined on  $A_2$  by  $d'(1) = 2112$  and  $d'(2) = 1112$ .



Here we consider that  $d'$  is a morphism from the alphabet  $A_2$  onto a two-letter alphabet  $A'_2$  where  $A'_2$  is a copy of  $A_2$  ( $A'_2 = \{1, 2\}$ ) such that  $1 \prec 2$ .

Then we have the following:

- $f(2) \prec f(1)$ ,
- $f(2)$  is a Lyndon word (thus a power of a Lyndon word),
- $f(21) = 111222112$  is a Lyndon word (over  $A'_2$ ).

Consequently, by Theorem 8.10,  $d'$  preserves infinite Lyndon words:  $F'$  is an infinite Lyndon word.  $\square$

We complete this section by stating without proof a result about  $F''$ , the second derivative of  $F$  ( $F'' = LS^2(F)$ ), which shows that the situation of  $F'$  regarding the type of its Lyndon factorization is rather "particular".

$$\begin{array}{rcl} F & = & 12\ 112 \quad 12\ 112\ 112 \quad 12\ 112 \quad 12\ \dots \\ F' & = & 1112\ 2112 \quad 1112\ 2112\ 2112 \quad 1112\ 2112 \quad 1112\ \dots \\ F'' & = & 31\ 22\ 21\ 12 \quad 31\ 22\ 21\ 22\ 21\ 12 \quad 31\ 22\ 21\ 12 \quad \dots \end{array}$$

As it is the case for  $F$  ( $F \in \{12, 112\}^\omega$ ) and for  $F'$  ( $F' \in \{1112, 2112\}^\omega$ ),  $F''$  decomposes over a two-element set of words (over  $A_3$ ):

$$F \in \{31222112, 312221222112\}^\omega.$$

Let  $g$  be the substitution from  $A_2$  onto  $A_3$  defined by  $g(21) = 31222112$  and  $g(211) = 312221222112$ , and let  $L$  be the application defined over the non-negative integers by

$$\begin{aligned} L(0) &= 3, \quad L(1) = 1222 \\ L(n) &= 112g[(\tilde{\varphi}\varphi)^{n-1}(1)](112)^{-1}, \quad \forall n \geq 2 \end{aligned}$$

**Theorem 8.12.** *The Lyndon factorization of  $F''$  is of type (1):*

$$F'' = \prod_{n \geq 0} L(n)$$

where, for every non-negative integer  $n$ ,  $L(n)$  is a finite Lyndon word and  $L(n+1) \prec L(n)$ .  $\square$

#### 8.4. WHY $F'$ IS PARTICULAR

In the previous section we have seen that  $F$  and  $F''$  both have a Lyndon factorization of type (2), while the Lyndon factorization of  $F'$  is of type (1). This particularity is perhaps in relation with another one:  $F'$  is the only descendant of  $F$  which is also written over  $A_2$ ; from  $F''$  all the other descendants of  $F$  are written over the alphabet  $A_3$ . This property ( $F' \in A_2^\omega$ ) is of course not true in general, even for the Sturmian words (of which  $F$  is however a prototype).

Indeed  $F' \in A_2^\omega$  because  $F$  does not contain 111, nor 222 as a factor, and it is a general property: if  $u$  is any word over  $A_2$  then  $LS(u)$  is also a word over  $A_2$  if and only if  $u$  does not contain any factor 111, nor 222. A natural question is to characterize, among Sturmian words generated by morphisms, those having the above property. This is done in the following Proposition 8.13.

Let  $St = \{\varphi, \tilde{\varphi}, E\}^+$  be the set of all the Sturmian morphisms ( $E : 1 \mapsto 2, 2 \mapsto 1$ ). Except  $\varphi$ , are there some other morphisms from  $St$  generating words without 111 and 222? The answer is of course yes since it is known that compositions of  $\varphi$  and  $\tilde{\varphi}$  generate infinite words having the same set of factors as the Fibonacci word  $F$  (see, e.g., Berstel and Séébold, 1994 [6]). Thus all these words do not contain any factor 111, nor 222. On the other hand, it is also already known that morphisms from the set  $T = \{\varphi E, \tilde{\varphi} E\}^+ \cup \{E\varphi, E\tilde{\varphi}\}^+$  generate words with arbitrarily large powers of one single letter (see, e.g., Séébold, 1998 [19]).

The result below is based on the observation that, for a morphism to generate a Sturmian word having no factor 111, nor 222, it is necessary that this word has the same factors of length 4 as the Fibonacci word  $F$  (or as its inverse  $E(F)$ ).

**Proposition 8.13.** *A Sturmian morphism generates words containing neither 111 nor 222 if and only if either  $f = \varphi$  or  $f = \tilde{\varphi}$ , or  $f$  or  $Ef$  starts with  $\varphi^2$ ,  $\tilde{\varphi}^2$ ,  $\varphi\tilde{\varphi}$ , or  $\tilde{\varphi}\varphi$ .*

*Proof.* It is an easy task to verify that morphisms starting with the required factors generate words containing neither 111, nor 222.

Conversely, if a morphism has not the required form then it is a basic element of the set  $T$  described above, or its decomposition starts with

$$\varphi E\varphi, \varphi E\tilde{\varphi} = \tilde{\varphi} E\varphi, \tilde{\varphi} E\tilde{\varphi}, E\varphi E\varphi, E\varphi E\tilde{\varphi} = E\tilde{\varphi} E\varphi, \text{ or } E\tilde{\varphi} E\tilde{\varphi}.$$

In other words,

$$f \in \{\varphi E, \tilde{\varphi} E, E\varphi, E\tilde{\varphi}, g, Eg\} \text{ where } g \in \{\varphi E\varphi, \varphi E\tilde{\varphi} = \tilde{\varphi} E\varphi, \tilde{\varphi} E\tilde{\varphi}\}.St.$$

It is again an easy task to verify that such a morphism cannot generate an infinite word without 111 or 222.  $\square$

## REFERENCES

- [1] J.-P. ALLOUCHE, J. SHALLIT, *Automatic sequences: theory, applications, generalizations*, Cambridge University Press, Cambridge, UK, 2003.
- [2] S. ARSHON, Démonstration de l'existence de suites asymétriques infinies, *Mat. Sb.* **44** (1937), 769–777 (in Russian), 777–779 (French summary).
- [3] J. BERSTEL, Mots sans carré et morphismes itérés, *Discrete Math.* **29** (1980), 235–244.
- [4] J. BERSTEL, Fibonacci words – a survey, *in*: G. Rozenberg, A. Salomaa (ed), *The Book of L*, Springer-Verlag 1986, 13–27.
- [5] J. BERSTEL, P. SÉÉBOLD, A characterization of Sturmian morphisms, *MFCS'93*, Gdansk (Poland), *Lect. Notes Comp. Sci.* **711** (1993), 281–290.
- [6] J. BERSTEL, P. SÉÉBOLD, A remark on morphic Sturmian words, *Informatique théorique et applications* **28** (1994), 255–263 (Special Issue in Honor of K. Culik II).
- [7] S. BRLEK, S. DULUCQ, A. LADOUCEUR, L. VUILLON, Combinatorial properties of smooth infinite words, *Theoret. Comput. Sci.* **352** (2006), 306–317.
- [8] A. COBHAM, Uniform tag sequences, *Math. Systems Th.* **6** (1972), 164–192.

- [9] J. H. CONWAY, The weird and wonderful chemistry of audioactive decay, *in*: T. M. Cover, B. Gopinath (eds), *Open problems in communication and computation*, Springer-Verlag, New-York 1987, 173–188.  
See also *Eureka* **46** (1986), 5–18.
- [10] B. GERMAIN-BONNE, À propos d’une itération sur chaînes de caractères numériques, *Laboratoire d’Analyse Numérique et d’Optimisation, Université Lille 1*, Research Report ANO 293, january 1993, 8 pp.
- [11] B. GERMAIN-BONNE, Chaînes alphanumériques ; cycles et points fixes, *Laboratoire d’Analyse Numérique et d’Optimisation, Université Lille 1*, Research Report ANO 301, march 1993, 23 pp.
- [12] B. GERMAIN-BONNE, Mots autodescriptifs et co-descriptifs, *Laboratoire d’Analyse Numérique et d’Optimisation, Université Lille 1*, Research Report ANO 332, june 1994, 32 pp.
- [13] M. LOTHAIRE, *Combinatorics on Words*, vol. 17 of *Encyclopedia of Mathematics and Applications*, Addison-Wesley, Reading, Mass., 1983.  
Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, Cambridge, UK, 1997.
- [14] M. LOTHAIRE, *Algebraic Combinatorics on Words*, vol. 90 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, UK, 2002.
- [15] G. MELANÇON, Lyndon factorization of sturmian words, *Discrete Math.* **210** (2000), 137–149.
- [16] J.-J. PANSIOT, Hiérarchie et fermeture de certaines classes de tag-systèmes, *Acta Informatica* **20** (1983), 179–196.
- [17] G. RICHOMME, On morphisms preserving infinite Lyndon words, *Discrete Math. and Theoret. Comput. Sci.* **9** (2007), 89–108.
- [18] G. ROZENBERG, A. SALOMAA (Eds), *Handbook of Formal Languages* Vol. 1, Springer, 1997.
- [19] P. SÉÉBOLD, On the conjugation of standard morphisms, *Theoret. Comput. Sci.* **195** (1998), 91–109.
- [20] P. SÉÉBOLD, About some overlap-free morphisms on a  $n$ -letter alphabet, *Journal of Automata, Languages and Combinatorics* **7** (2002), 579–597.
- [21] R. SIROMONEY, L. MATHEW, V.R. DARE, K.G. SUBRAMANIAN, Infinite Lyndon words, *Inform. Proc. Letters* **50** (1994), 101–104.

Communicated by (The editor will be set by the publisher).  
(The dates will be set by the publisher).