



HAL
open science

Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words

Gwenaël Richomme

► **To cite this version:**

Gwenaël Richomme. Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words. 2005, pp.341-351. hal-00013457

HAL Id: hal-00013457

<https://hal.science/hal-00013457v1>

Submitted on 8 Nov 2005

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.

LaRIA : Laboratoire de Recherche en Informatique d'Amiens
Université de Picardie Jules Verne – CNRS FRE 2733
33, rue Saint Leu, 80039 Amiens cedex 01, France
Tel : (+33)03 22 82 88 77
Fax : (+33)03 22 82 54 12
<http://www.laria.u-picardie.fr>

**Conjugacy of morphisms
and Lyndon decomposition of
standard Sturmian words**

G. Richomme^a

LaRIA RESEARCH REPORT : LRR 2005-09
(October 2005)

^a LaRIA, Université de Picardie Jules Verne, gwenael.richomme@u-picardie.fr

Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words*

G. Richomme
LaRIA, Université de Picardie Jules Verne
33, Rue Saint Leu,
F-80039 Amiens cedex 1
(gwenael.richomme@u-picardie.fr)

October 28, 2005

Abstract

Using the notions of conjugacy of morphisms, we answer a question of G. Melançon concerning the decomposition in Lyndon words of standard Sturmian words. We show some connections with morphisms preserving Lyndon words

1 Introduction

Finite (or infinite) Lyndon words can be encountered in many studies (see for instance [8, 9, 10]). They are the nonempty words which are smaller in lexicographic order than all their proper suffixes. The Lyndon factorization theorem [4] states that any finite word can be decomposed uniquely in a product of nonincreasing (in lexicographic order) Lyndon words. This result was extended to infinite words [19] (In such a case, the decomposition can end with an infinite Lyndon word). Thus some works concern the decomposition in Lyndon words of some infinite words (see for instance [3, 5, 11, 12, 18] for such results).

In [12], G. Melançon gives a decomposition in Lyndon words of standard Sturmian words. He asks the following question: in which cases, the sequence of nonincreasing Lyndon words appearing in the decomposition of a standard Sturmian word can be written $(g^n(\ell_0))_{n \geq 0}$ with ℓ_0 a Lyndon word and g a morphism. In Section 5, we answer this question.

For this, we use results about morphisms preserving Lyndon words [14] and about conjugacy of morphisms [13]. In particular, we show that when a positive answer exists to the previous question, g preserves Lyndon words and is the conjugate of a morphism f that generates the decomposed standard Sturmian word.

In Section 2, we recall notions on Sturmian words and morphisms. Section 3 recalls both the decomposition in Lyndon words of standard Sturmian words obtained by G. Melançon, and his question. This section also recall notions on morphisms preserving Lyndon words. Section 4 presents notions on conjugacy of morphisms and introduces a new particular case, namely the strong

*This paper was presented at the 5th International Conference on Words which held in Montréal on september 2005 (Publications du LaCIM numéro 36, page 341-351 (S. Brlek, C. Reutenauer eds.).)

conjugacy. Using it in conjunction with morphisms preserving Lyndon words, we give a new proof that for any standard Sturmian words w over $\{a < b\}$, aw is an infinite Lyndon word [3]. Finally, in Section 5, we answer G. Melançon. Note that at a first step, we express the decomposition of a standard Sturmian word using only morphisms.

2 Sturmian words and morphisms

We recall here notions on words (see for instance [8, 9] for more details).

An *alphabet* A is a set of symbols called *letters*. Here we consider only finite alphabets. A *word over* A is a sequence of letters from A . The *empty word* ε is the empty sequence. Equipped with the concatenation operation, the set A^* of finite words over A is a free monoid with neutral element ε and set of generators A . We denote by A^ω the set of infinite words over A . As usually, for a finite word u and an integer n , the n^{th} power of u , denoted u^n , is the word ε if $n = 0$ and the word $u^{n-1}u$ otherwise. If u is not the empty word, u^ω denotes the infinite word obtained by infinitely repeating u . A finite word w is said *primitive* if for any word u , the equality $w = u^n$ (with n an integer) implies $n = 1$. Any word is the power of a unique primitive word called the *primitive root* of w .

Given a nonempty word $u = u_1 \dots u_n$ with $u_i \in A$, the *length* $|u|$ of u is the integer n . One has $|\varepsilon| = 0$. If for some words u, v, p, s (possibly empty), $u = pvs$, then v is a *factor* of u , p is a *prefix* of u and s is a *suffix* of u . When $p \neq u$ (resp. $s \neq u$), we say that p is a *proper prefix* (resp. s is a *proper suffix*) of u . By $|u|_a$ we denote the number of occurrences of the letter a in the word u .

Sturmian words may be defined in many equivalent ways (see [1] for instance). They are infinite binary words. Here we will consider them as the infinite balanced non ultimately periodic words. We recall that a (finite or infinite) word w over $\{a, b\}$ is *balanced* if for any factors u and v of same length $||u|_a - |v|_a| \leq 1$, and that an infinite word w is *ultimately periodic* if $w = uv^\omega$ for some finite words u and v .

Many studies of Sturmian words use Sturmian morphisms. Let A, B be two alphabets. A *morphism* (*endomorphism* if $A = B$) f from A^* to B^* is a mapping from A^* to B^* such that for all words u, v over A , $f(uv) = f(u)f(v)$. We also say that f is a morphism on A or that f is defined on A (without any other precision when B has no importance). A morphism on A is entirely known by the images of the letters of A . A morphism extends naturally on infinite words. We denote just by juxtaposition the composition of morphisms. Given an endomorphism f , if $\lim_{n \rightarrow \infty} f^n(a)$ exists, then this limit is denoted $f^\omega(a)$ and is a fixed point of f : the word $f^\omega(a)$ is said generated by f .

Sturmian morphisms are the morphisms in $\{E, L_a, L_b, R_a, R_b\}^*$ where E, L_a, L_b, R_a, R_b are the endomorphisms defined on $\{a, b\}$ by $E(a) = b, E(b) = a, L_a(a) = a, L_a(b) = ab, L_b(a) = ba, L_b(b) = b, R_a(a) = a, R_a(b) = ba, R_b(a) = ab, R_b(b) = b$. Many relations exists between Sturmian words and Sturmian morphisms. For instance, it is known [2, 6] that any Sturmian word can be defined as an infinite product of Sturmian morphisms.

A particular case of Sturmian words is the standard (or characteristic) one. For any standard Sturmian words, there exists a sequence $(d_n)_{n \geq 0}$ of integers, called *the directive sequence* verifying $d_1 \geq 0$ and $d_k \geq 1$ for all $k \geq 2$, such that

$$w = \lim_{n \rightarrow \infty} s_n$$

where the sequence $(s_n)_{n \geq -1}$ of words is defined by : $s_{-1} = b$, $s_0 = a$ and $s_n = s_{n-1}^{d_n} s_{n-2}$ for $n \geq 1$. Let us observe that for every $n \geq 0$, s_{2n} ends with a . Moreover [1],

$$\begin{aligned} s_{2n} &= L_a^{d_1} L_b^{d_2} \dots L_a^{d_{2n-1}} L_b^{d_{2n}}(a) \\ &= L_a^{d_1} L_b^{d_2} \dots L_a^{d_{2n-1}} L_b^{d_{2n}} L_a^{d_{2n+1}}(a) \\ s_{2n+1} &= L_a^{d_1} L_b^{d_2} \dots L_a^{d_{2n-1}} L_b^{d_{2n}} L_a^{d_{2n+1}}(b) \\ &= L_a^{d_1} L_b^{d_2} \dots L_a^{d_{2n-1}} L_b^{d_{2n}} L_a^{d_{2n+1}} L_b^{d_{2n+2}}(b) \end{aligned}$$

3 Lyndon words and morphisms

From now on we consider ordered alphabets. We denote $\{\alpha_1 < \dots < \alpha_n\}$ the n -letter alphabet $\{\alpha_1, \dots, \alpha_n\}$ with order $\alpha_1 < \dots < \alpha_n$. Given an ordered alphabet A , we denote by \preceq the lexicographic order whenever used on A^* or on A^ω . Let recall that for two different (finite or infinite) words u and v , $u \prec v$ if and only if $u = xay$, $v = xbz$ with $a, b \in A$, $a < b$, $x \in A^*$, $y, z \in A^* \cup A^\omega$, or if (when u is finite) u is a prefix of v .

A nonempty finite word w is a *Lyndon word* if for all nonempty words u and v , $w = uv$ implies $w \prec vu$. Equivalently [4, 8], a nonempty word w is a Lyndon word if all its nonempty proper suffixes are greater than it for the lexicographic order. For instance, on the one-letter alphabet $\{a\}$, only a is a Lyndon word. On $\{a < b\}$ the Lyndon words of length at most 5 are a , b , ab , aab , abb , $aaaab$, $aabb$, $abbb$, $aaaab$, $aaabb$, $aabab$, $aabbb$, $abbbb$. Lyndon words are primitive.

The second definition of Lyndon words extends to infinite words: An infinite word is an *infinite Lyndon word* if all its proper suffixes are greater than it for the lexicographic order. A useful result of G. Melançon [12] states that an infinite word is a Lyndon word if and only if it has an infinity of prefixes that are Lyndon words. See for instance [7] for a recent example of infinite Lyndon word.

Any nonempty finite or infinite Lyndon words can be decomposed as a nonincreasing product of Lyndon words. First, R.C. Lyndon proved (see [8] for instance):

Any word $w \in A^+$ may be written uniquely as a nonincreasing product of Lyndon words: $w = \ell_1 \ell_2 \dots \ell_n$ where for each i , ℓ_i is a Lyndon word and $\ell_n \preceq \ell_{n-1} \preceq \dots \preceq \ell_1$.

This result was generalized to infinite words [19]:

Any right infinite word w may be uniquely expressed as a nonincreasing product of Lyndon words, finite or infinite, in one of the two following forms: either there exists an infinite nonincreasing sequence of finite Lyndon words $(\ell_k)_{k \geq 0}$ such that

$$w = \prod_{n \geq 0} \ell_n = \ell_0 \ell_1 \dots$$

or there exist finite Lyndon words $\ell_0, \dots, \ell_{m-1}$ ($m \geq 0$) and an infinite word ℓ_m such that $\ell_m \prec \ell_{m-1} \preceq \ell_{m-2} \preceq \dots \preceq \ell_0$ and

$$w = \ell_0 \dots \ell_{m-1} \ell_m.$$

As already said in the introduction, many works concern the decomposition in Lyndon words of some infinite words. In [12], G. Melançon obtains the decomposition of standard Sturmian words. We consider these words here on the alphabet $\{a < b\}$. For any word w ending with the letter a , let us denote \bar{w} the word such that $w = \bar{w}a$.

Theorem 3.1 [12] *Let s be a standard Sturmian word with directive sequence $(d_n)_{n \geq 1}$. Let $\ell_n = as_{2n}^{d_{2n+1}-1} s_{2n-1} \bar{s}_{2n}$. (if $d_1 = 0$ then $\ell_0 = b$).*

The words $(\ell_n)_{n \geq 0}$ form a strictly decreasing sequence of Lyndon words and the unique factorization of s as a nonincreasing product of Lyndon words is

$$s = \prod_{n \geq 0} \ell_n^{d_{2n+1}}.$$

G. Melançon wrote [12, Remark 3.7] :

When is the sequence $(\ell_n)_{n \geq 0}$ morphic ? More precisely, is it possible to give a morphism $\varphi : \{a, b\}^ \rightarrow \{a, b\}^*$ and a Lyndon word ℓ_0 such that $\ell_{n+1} = \varphi(\ell_n)$? This question has a positive answer in the case where the directive sequence is constant. For instance, if $d_n = 2$ for all $n \geq 0$, then we may set $\ell_0 = aab$ and use the morphism mapping $a \mapsto aaabaab$ and $b \mapsto aab$.*

A characteristic Sturmian word may be itself morphic. That is, may be the limit $\lim_n \varphi^n(a)$ of a (nonerasing) morphism (satisfying $\varphi(a) \in aA^$). It is known that this is essentially equivalent to the fact that its directive sequence is periodic. Unfortunately, even when a characteristic Sturmian word s has a periodic directive sequence, it seems that the sequence $(\ell_n)_{n \geq 0}$ is not always morphic, although it is possible to describe patterns in the factorization.*

The aim of this paper is to answer this question. The main ideas of our proof are generalizations of the following remarks: the morphism $a \mapsto aaabaab$ and $b \mapsto aab$ is the Sturmian morphism $L_a^2 R_b^2$ and preserves Lyndon words. Moreover $L_a^2 L_b^2$ is a conjugate of $L_a^2 R_b^2$, $L_a^2 L_b^2(a) = (l_0)^2 a$ and $L_a^2 R_b^2(a) = a(l_0)^2$. Let us note that, in [14], similar remarks are made about the decomposition of the Fibonacci word (the standard Sturmian word of directive sequence $(1)_{n \geq 0}$). In Section 4, we recall notions on conjugacy of morphisms.

Let us now recall some results on morphisms preserving (finite) Lyndon words. These morphisms are studied in [14]. By definition, a morphism f preserves Lyndon word if for each Lyndon word w , $f(w)$ is a Lyndon word. Effective characterizations of such morphisms are given in [14]. Consequently Sturmian words preserving Lyndon words are known:

Proposition 3.2 [14] *A Sturmian morphism on $\{a < b\}$ is a Lyndon morphism if and only if it belongs to $\{L_a, R_b\}^*$.*

To end this section, let us observe that a study of morphisms preserving infinite Lyndon words is given in [15].

4 Strong Conjugacy

In this section, we recall the notion of conjugacy (see, e.g., [9, 13]). We also introduce the particular case of strong conjugacy which will be useful to answer G. Melançon.

Let A and B be two alphabets and let f and g be two morphisms from A^* to B^* . The morphism g is a (right) *conjugate* of f if there exists a word u such that for any word x over A , $f(x)u = ug(x)$. We will also say that f and g are *u -conjugated*, and we will denote $f \triangleleft_u g$. Moreover if $f(a) = ua$ and $g(a) = au$ for a letter a , f and g will be called *strongly (on a) u -conjugated*.

Let us recall that any morphism f has at least one conjugate: itself ($f \triangleleft_\epsilon f$). The Fibonacci morphism $\varphi = L_a E$ defined by $\varphi(a) = ab$ and $\varphi(b) = a$ has exactly two conjugates, itself and the morphism $\tilde{\varphi} = R_a E$ ($\tilde{\varphi}(a) = ba$, $\tilde{\varphi}(b) = a$). A lot of relations between conjugacy of morphisms and Sturmian morphisms were given by P. Séébold [17] and generalized to a larger family of morphisms in [13].

Since $\varphi(a)$ does not end with the letter a , no morphism is strongly conjugate (on a) to the Fibonacci morphism. Nevertheless we can observe that φ^2 ($a \mapsto aba$, $b \mapsto ab$) is strongly ab -conjugated to $\varphi\tilde{\varphi}$ ($a \mapsto aab$, $b \mapsto ab$). More generally, for all integers x and y ($y \neq 0$), the morphism $L_a^x L_b^y$ is strongly conjugated to the morphism $L_a^x R_b^y$. This follows immediately the formulas: $L_a^x L_b^y(a) = (a^x b)^y a$, $L_a^x L_b^y(b) = a^x b$, $L_a^x R_b^y(a) = a(a^x b)^y$, $L_a^x R_b^y(b) = a^x b$ ($L_a^x L_b^y \triangleleft_{(a^x b)^y} L_a^x R_b^y$).

A basic property of conjugacy is [9, 13]: for morphisms f, f', g, g' , and words u, u' , if $f \triangleleft_u g$ and $f' \triangleleft_{u'} g'$ then $ff' \triangleleft_{f(u')u} gg'$ (of course $f(u')u = ug(u')$). This property extends to strong conjugacy:

Lemma 4.1 *Let f, f', g, g' , (a a letter) and u, u' words such that f is strongly (on a) u -conjugated to g and f' is strongly (on a) u' -conjugated to g' . Then ff' is strongly (on a) $[f(u')u]$ -conjugated to gg' .*

Proof. We already know $ff' \triangleleft_{f(u')u} gg'$. By hypothesis, $f(a) = ua$, $g(a) = au$, $f'(a) = u'a$ et $g'(a) = au'$. Thus $ff'(a) = f(u'a) = f(u')ua$ and $gg'(a) = g(au') = aug(u') = af(u')u$. So ff' is strongly $[f(u')u]$ -conjugated to gg' . \square

We end this section with a first use of strong conjugacy concerning Sturmian words. One particular property of any standard Sturmian word w over $\{a < b\}$ is that both aw and bw are Sturmian words [16]. Words aw (with w standard Sturmian) are also known as Christoffel words. In [3], it is shown, that Christoffel words are infinite Lyndon words:

Proposition 4.2 [3] *For any standard Sturmian word w over $\{a < b\}$, aw is an infinite Lyndon word.*

Proof. Let w be a standard word with directive sequence $(d_n)_{n \geq 1}$. We have already said that a standard word can be viewed as $w = \lim_{n \rightarrow \infty} s_n$ for some words s_n defined in Section 2. In fact, we can verify that then $w = \lim_{n \rightarrow \infty} s_{2n}$. Let $n \geq 1$. We know that $s_{2n} = L_a^{d_1} L_b^{d_2} \dots L_a^{d_{2n-1}} L_b^{d_{2n}}(a)$. As a consequence of Lemma 4.1 and of the fact that for all integers x and y , the morphism $L_a^x L_b^y$ is strongly conjugated to the morphism $L_a^x R_b^y$, we can verify that $L_a^{d_1} L_b^{d_2} \dots L_a^{d_{2n-1}} L_b^{d_{2n}}$ is strongly conjugated to $L_a^{d_1} R_b^{d_2} \dots L_a^{d_{2n-1}} R_b^{d_{2n}}$.

In particular, $aL_a^{d_1} L_b^{d_2} \dots L_a^{d_{2n-1}} L_b^{d_{2n}}(a) = L_a^{d_1} R_b^{d_2} \dots L_a^{d_{2n-1}} R_b^{d_{2n}}(a)a$. By Proposition 3.2, the morphism $L_a^{d_1} R_b^{d_2} \dots L_a^{d_{2n-1}} R_b^{d_{2n}}$ preserves Lyndon words. Hence $L_a^{d_1} R_b^{d_2} \dots L_a^{d_{2n-1}} R_b^{d_{2n}}(a)$ is a Lyndon word. Consequently the word w has an infinity of Lyndon words as prefixes. It is a Lyndon word. \square

Let us note that the previous proof technique can be used to state other results. For instance, we let the reader prove:

Proposition 4.3 *Let A be an alphabet and a a letter in A . Let f, g be two nonerasing endomorphisms on A and let u be a word over A such that f is u -strongly conjugate to g . Then $f^\omega(a)$ and $g^\omega(a)$ exist and $af^\omega(a) = g^\omega(a)$.*

Thus if g generates on a an infinite Lyndon word (which is the case if it preserves Lyndon words or if it preserves infinite Lyndon words (see [15])), $af^\omega(a)$ is an infinite Lyndon word.

The situation of Proposition 4.3 can be met for morphisms that are not Sturmian. For instance, this is the case with the morphisms:

$$f : \begin{cases} a \mapsto aba \\ b \mapsto abb \end{cases} \qquad g : \begin{cases} a \mapsto aab \\ b \mapsto bab \end{cases}$$

Moreover one can see that g preserves infinite Lyndon words and generates an infinite Lyndon word.

5 An answer to G. Melançon

In this section, we consider a standard Sturmian word w over the ordered alphabet $\{a < b\}$ with directive sequence $(d_n)_{n \geq 1}$ (Let recall that $d_1 \geq 0$ and $d_n \geq 1$ for all $n \geq 2$). The sequence of words $(s_n)_{n \geq 0}$ and $(\ell_n)_{n \geq 0}$ are those defined respectively at the end of Section 2 and in Theorem 3.1. In particular, $w = \lim_{n \rightarrow \infty} \prod_{n \geq 0} \ell_n^{d_{2n}}$ is the decomposition in Lyndon words of w (for each $n \geq 0$, ℓ_n is a Lyndon word and $\ell_{n+1} \preceq \ell_n$). Our result is:

Theorem 5.1 *With the hypotheses of this section, there exists a morphism g such that for all $n \geq 0$, $\ell_{n+1} = g(\ell_n)$ if and only if one of the two following cases hold:*

- $1 \leq d_1 \leq d_3$, and for all $n \geq 1$, $d_{2n} = d_2$ and $d_{2n+1} = d_3$. In this case, $\ell_0 = a^{d_1}b$ and $g = L_a^{d_1} R_b^{d_2} L_a^{d_3-d_1}$.
- $d_1 = 0$, $1 \leq d_2 \leq d_4$, and for all $n \geq 1$, $d_{2n+2} = d_4$ and $d_{2n+1} = d_3$. In this case, $\ell_0 = b$ and $g = R_b^{d_2} L_a^{d_3} R_b^{d_4-d_2}$.

We observe that in each case, the morphism g is a Sturmian morphism that preserves Lyndon words (see Proposition 3.2). Moreover the word w is generated by a Sturmian morphism $(L_a^{d_1} L_b^{d_2} L_a^{d_3-d_1}$ or $L_b^{d_2} L_a^{d_3} L_b^{d_4-d_2})$.

In order to prove the previous theorem, using the strong conjugacy, we first express each Lyndon word ℓ_n with morphisms. For $n \geq 0$, we denote:

$$f_n = (L_a^{d_1} L_b^{d_2}) \dots (L_a^{d_{2n-1}} L_b^{d_{2n}})$$

$$g_n = (L_a^{d_1} R_b^{d_2}) \dots (L_a^{d_{2n-1}} R_b^{d_{2n}})$$

The interest of the morphisms f_n is immediate since we have already seen relations between them and the words s_n ($s_{2n} = f_n(a)$, $s_{2n+1} = f_{n+1}(b)$). We also observe that each g_n is a morphism that preserves Lyndon words. As a consequence of Lemma 4.1 and of the fact that for all integers x and y , the morphism $L_a^x L_b^y$ is strongly conjugated to the morphism $L_a^x R_b^y$, we have:

Lemma 5.2 *For all $n \geq 1$, f_n is strongly (on a) conjugated to g_n .*

Now we give a new formula for the words $(\ell_n)_{n \geq 0}$:

Lemma 5.3 For all $n \geq 0$, $\ell_n = g_n L_a^{d_{2n+1}}(b)$

Proof.

$$\begin{aligned} \ell_n a &= a s_{2n}^{d_{2n+1}-1} s_{2n-1} \overline{s_{2n}} a \\ &= a s_{2n}^{d_{2n+1}-1} s_{2n-1} s_{2n} \\ &= a f_n(a^{d_{2n+1}-1} b a). \end{aligned}$$

If $n = 0$, $\ell_n a = a^{d_1} b a = L_a^{d_1}(b) a = g_0 L_a^{d_1}(b) a$.

When $n \geq 1$, let u_n be the word such that $f \triangleleft_{u_n} g_n$. By Lemma 5.2, $f_n(a) = u_n a$, $g_n(a) = a u_n$. Thus

$$\begin{aligned} \ell_n a &= a f_n(a^{d_{2n+1}-1} b) u_n a \\ &= a u_n g_n(a^{d_{2n+1}-1} b) a \\ &= g_n(a^{d_{2n+1}} b) a \\ &= g_n L_a^{d_{2n+1}}(b) a \end{aligned}$$

Consequently for all $n \geq 0$, $\ell_n = g_n L_a^{d_{2n+1}}(b)$. \square

Let us observe that Lemma 5.2 allows to give a new proof of the fact that the words $(\ell_n)_{n \geq 0}$ form a strictly decreasing sequence of Lyndon words. Indeed, by Proposition 3.2, each morphism $g_n L_a^{d_{2n+1}}$ is a Lyndon morphism, hence $g_n L_a^{d_{2n+1}}$ is a Lyndon word. Moreover $R_b^{d_{2n}} L_a^{d_{2n+1}}(b)$ for each $n \geq 1$, then $R_b^{d_{2n}} L_a^{d_{2n+1}}(b) \prec b$ which implies $\ell_n = g_n L_a^{d_{2n+1}}(b) \prec g_{n-1} L_a^{d_{2n-1}}(b) = \ell_{n-1}$ (since any morphism preserving Lyndon words also strictly preserves the lexicographic order on finite words [14]).

Proof of Theorem 5.1. Note that the “if” part of the theorem is immediate. Assume the sequence $(\ell_n)_{n \geq 0}$ is morphic. Let g be the morphism such that, for all $n \geq 0$, $g(\ell_n) = \ell_{n+1}$. Observe that the morphism g cannot be erasing since otherwise this contradicts the fact that ℓ_2 is a primitive word (as a Lyndon word).

We first consider the case $d_1 \geq 2$. Observe $\ell_0 = a^{d_1} b$ and

$$g(a^{d_1} b) = \ell_1 = [a(a^{d_1} b)^{d_2}]^{d_3} a^{d_1} b.$$

Assume $g(a) = a$, and so $g(b) = ab(a^{d_1} b)^{d_2-1} [a(a^{d_1} b)^{d_2}]^{d_3-1} a^{d_1} b$. The word $\ell_2 = g(\ell_1)$ has $g(a^{d_1+1} b)$ as prefix. Thus the words a^{d_1+2} and $ba^{d_1} b$ are factors of ℓ_2 . This contradicts the fact that ℓ_2 , as a factor of a Sturmian word, is balanced. Hence $g(a) \neq a$.

Since $d_1 \geq 2$ and $g(a^{d_1} b)$ starts with $a^{d_1+1} b$, the word $a^{d_1+1} b$ is a prefix of $g(a)$. More precisely, $a(a^{d_1} b)^{d_2}$ must be a prefix of $g(a)$. Finally, we can verify that $g(a) = (a(a^{d_1} b)^{d_2})^k$ for an integer $k \geq 1$. It follows $g(b) = (a(a^{d_1} b)^{d_2})^{d_3-kd_1} a^{d_1} b$ which implies $d_3 \geq kd_1$.

Assume $k \geq 2$. The word $\ell_2 = g(\ell_1)$ contains $g(ba^{d_1} b)$ and $g(a^{d_1+1} b)$ as factors. The word $g(ba^{d_1} b)$ ends with bub where $u = (a^{d_1} b)^{d_2} [a(a^{d_1} b)^{d_2}]^{d_3} a^{d_1}$. Furthermore the word $g(a^{d_1+1} b) = [a(a^{d_1} b)^{d_2}]^{d_3+k} a^{d_1} b$ starts with aua . This contradicts the fact that ℓ_2 is balanced.

Hence $k = 1$, $d_3 \geq d_1$, $g(a) = a(a^{d_1}b)^{d_2}$, $g(b) = [a(a^{d_1}b)^{d_2}]^{d_3-d_1}a^{d_1}b$. We observe that $g = L_a^{d_1}R_b^{d_2}L_a^{d_3-d_1}$ and that it is an injective morphism.

Now we can prove that, for all $n \geq 1$, $d_{2n} = d_2$ and $d_{2n+1} = d_3$. We act by induction on n . There is nothing to do for $n = 1$. Let $n \geq 1$. Assume that we have already proved $d_{2p} = d_2$ and $d_{2p+1} = d_3$ for all integers p with $1 \leq p \leq n$. We have $\ell_{n+1} = g_{n+1}L_a^{d_{2n+3}}(b) = L_a^{d_1}(R_b^{d_2}L_a^{d_3})^n R_b^{d_{2n+2}}L_a^{d_{2n+3}}(b) = (L_a^{d_1}R_b^{d_2}L_a^{d_3-d_1})^n L_a^{d_1}R_b^{d_{2n+2}}L_b^{d_{2n+3}}(b) = g^n(L_a^{d_1}R_b^{d_{2n+2}}L_b^{d_{2n+3}}(b))$. Moreover $\ell_{n+1} = g^n(\ell_1)$. Since g is injective, $\ell_1 = L_a^{d_1}R_b^{d_{2n+2}}L_b^{d_{2n+3}}(b)$. This implies $d_{2n+2} = d_2$ and $d_{2n+3} = d_3$.

Now we consider the case $d_1 = 1$. We have $\ell_0 = ab$ and $\ell_1 = [a(ab)^{d_2}]^{d_3}ab$. As in case $d_1 \geq 2$, we cannot have $g(a) = a$. Hence $g(a)$ starts with aa . We observe that $g(a)$ cannot ends with a , since otherwise the balanced word $\ell_2 = g(\ell_1)$ contains aaa and bab . We observe also that $g(a) \neq [a(ab)^{d_2}]^i a(ab)^k$ for any integer k, i such that $1 \leq k < d_2$ and $i \geq 0$. Indeed otherwise the word ℓ_2 containing both $g(aa)$ and $g(ab)$ should contains the factors $a(ab)^k aa$ and $b(ab)^k ab$ (since $(ab)^{d_2+1}$ ends $g(ab)$): this contradicts the fact that ℓ_2 is balanced. It follows that $g(a) = [a(ab)^{d_2}]^k$ with $1 \leq k \leq d_3$ and $g(b) = [a(a^{d_1}b)^{d_2}]^{d_3-k}a^{d_1}b$. Exactly as in case $d_1 \geq 2$, we can then prove that $k = 1$, $g = L_a R_b^{d_2} L_a^{d_3-1}$ and for all integers $n \geq 1$, $d_{2n} = d_2$ and $d_{2n+1} = d_3$.

From now on, we consider the case $d_1 = 0$. we have $\ell_0 = b$ and so $g(b) = \ell_1 = (ab^{d_2})^{d_3}b$. Moreover $\ell_2 = R_b^{d_2}L_a^{d_3}R_b^{d_4}L_a^{d_5}(b)$, that is

$$\ell_2 = [ab^{d_2}[(ab^{d_2})^{d_3}b]^{d_4}]^{d_5}(ab^{d_2})^{d_3}b.$$

Furthermore $\ell_2 = g^2(b) = g((ab^{d_2})^{d_3}b)$. It follows that

$$g((ab^{d_2})^{d_3}) = [ab^{d_2}g(b)^{d_4}]^{d_5}$$

Since the word $ab^{d_2}g(b)^{d_4} = ab^{d_2}[(ab^{d_2})^{d_3}b]^{d_4}$ is a primitive word, $g(ab^{d_2}) = [ab^{d_2}g(b)^{d_4}]^x$ and $xd_3 = d_5$ for an integer $x \geq 1$. Since ab^{d_2} is not a suffix of $g(b)$, $d_2 \leq d_4$.

Let us prove that $x = 1$, that is, $d_3 = d_5$. Assume by contradiction that $x \geq 2$. The word ℓ_2 has $(ab^{d_2})^{d_3+1}$ as a prefix and $[(ab^{d_2})^{d_3}b]^2$ as a suffix. Let $u = ab^{d_2}g(b)^{d_4}$: $g(ab^{d_2}) = u^x$. The word $\ell_3 = g(\ell_2)$ contains the factor $g((ab^{d_2})^{d_3+1}) = u^{(d_3+1)x} = uu^{d_5}uu^{x-2}$ which contains the factor $ab^{d_2}g(b)^{d_4}u^{d_5}ab^{d_2}(ab^{d_2})^{d_3}b$ which starts with $ab^{d_2}g(b)^{d_4}u^{d_5}(ab^{d_2})^{d_3}a$. Observe now that $g((ab^{d_2})^{d_3}b) = [ab^{d_2}g(b)^{d_4}]^{d_5}g(b)$ ends with $b^{d_2+1}g(b)^{d_4}$. Consequently the word ℓ_3 also contains the factor $b^{d_2+1}g(b)^{d_4}g(((ab^{d_2})^{d_3}b) = bb^{d_2}g(b)^{d_4}u^{d_5}(ab^{d_2})^{d_3}b$. We have a contradiction with the fact that ℓ_3 is a balanced word.

From what precedes, $g(ab^{d_2}) = ab^{d_2}g(b)^{d_4}$ and so $g(a) = ab^{d_2}g(b)^{d_4-d_2} = ab^{d_2}((ab^{d_2})^{d_3}b)^{d_4-d_2}$. Moreover $g(b) = (ab^{d_2})^{d_3}b$. We observe $g = R_b^{d_2}L_a^{d_3}R_b^{d_4-d_2}$. As in case $d_1 \geq 2$, we can state that, for all integers $n \geq 2$, $d_{2n} = d_4$ and $d_{2n-1} = d_3$. \square

6 Conclusion

This paper shows the interest of conjugacy of morphisms and of morphisms preserving Lyndon words as tools to tackle problems concerning Sturmian words and/or Lyndon words. We are now working to find other situations where these tools can be useful. In particular, we are looking for the decomposition in Lyndon words of any Sturmian words.

Acknowledgement

Many thanks to P. Séébold for useful discussions on strong conjugacy.

References

- [1] J. Berstel and P. Séébold. Sturmian words, chapter 2 in [9].
- [2] V. Berthé, C. Holton, and L. Q. Zamboni. Initial powers of sturmian sequences. To appear in *Acta Informatica*, see www.lirmm.fr/~berthe.
- [3] J.P. Borel and F. Laubie. Quelques mots sur la droite projective réelle. *Journal de Théorie des Nombres de Bordeaux*, 5:23–51, 1993.
- [4] K.T. Chen, R.H. Fox, and R.C. Lyndon. Free differential calculus IV – the quotient groups of the lower central series. *Ann. Math.* 68, 68:81–95, 1958.
- [5] A. Ido and G. Melançon. Lyndon factorization of the Thue-Morse word and its relatives. *Discret. Math. and Theoret. Comput. Sci.*, 1:43–52, 1997.
- [6] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoretical Computer Science*, 276(1-2):281–313, 2002.
- [7] J. Justin and G. Pirillo. On a characteristic property of Arnoux-Rauzy sequences. *RAIRO Theoretical Informatics and Applications*, 36:385–388, 2002.
- [8] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, 1983. Reprinted in 1997 by Cambridge University Press in the Cambridge Mathematical Library, Cambridge, UK, 1997.
- [9] M. Lothaire. *Algebraic Combinatorics on words*, volume 90 of *Encyclopedia of Mathematics*. Cambridge University Press, Cambridge, UK, 2002.
- [10] M. Lothaire. *Applied Combinatorics on Words*. To appear. (see www-igm.univ-mlv.fr/~berstel).
- [11] G. Melançon. Lyndon factorization of infinite words. In *STACS'96*, volume 1046 of *Lect. Notes in Comp. Sci.*, pages 147–154, 1996.
- [12] G. Melançon. Lyndon factorization of Sturmian words. *Discrete Mathematics*, 210:137–149, 2000.
- [13] G. Richomme. Conjugacy and episturmian morphisms. *Theoretical Computer Science*, 302:1–34, 2003.
- [14] G. Richomme. Lyndon morphisms. *Bulletin of the Belgian Mathematical Society*, 10:761–785, 2003.

- [15] G. Richomme. On morphisms preserving infinite lyndon words. to appear in *Journal of Automata, Languages and Combinatorics*. An extended abstract appeared in the proceedings of the conference "Journées Montoises d'Informatique Théorique", Liège, Belgium, 2004 (Prépublication 04.006, Institut de Mathématique, Université de Liège), p325-333.
- [16] P. Séébold. Fibonacci morphisms and Sturmian words. *Theoretical Computer Science*, 88:365–384, 1991.
- [17] P. Séébold. On the conjugation of standard morphisms. *Theoretical Computer Science*, 195:91–109, 1998.
- [18] P. Séébold. Lyndon factorization of the Prouhet words. *Theoretical Computer Science*, 307:179–197, 2003.
- [19] R. Siromoney, L. Mathew, V. R. Dare, and K. G. Subramanian. Infinite Lyndon words. *Information Processing Letters*, 50:101–104, 1994.