Lyndon Words and Singular Factors of the Fibonacci Word GUY MELANÇON¹

LaBRI, UMR 5800 CNRS, Université Bordeaux I

Abstract. Two different factorizations of the Fibonacci infinite word were given independently in [7] and [5]. Our work started after a remark by J. Berstel, who observed (and consequently conjectured) that the words of the factorization in [5] were products of two consecutive words of the factorization in [7]. The results we give here confirm J. Berstel's claim and fully describe the links between the two factorizations.

Résumé. Deux factorisations du mot de Fibonacci ont été donné dans deux articles indépendants, [7] et [5]. Notre travail a pris sa source suite à une remarque de J. Berstel, qui avait observé (et par le fait même avait conjecturé) que les mots de la factorisation [5] étaient obtenus en concaténant deux mots consécutifs de la seconde factorisation [7]. Nos résultats confirment l'observation de J. berstel et détaillent les liens entre ces deux factorisations.

1 Introduction

The numerous aspects under which the combinatorial properties of the Fibonacci word have been studied are amazing. A huge set of different notions in *algebraic combinatorics on words*² may be illustrated by using this infinite word as an example. It is also of interest in other fields such as number theory, quasicrystals, computational complexity, to name only a few (see [1]).

The combinatorial structure of an infinite word is often revealed by the study of the set of its factors: that is, the finite words appearing within it. As far as the Fibonacci word is concerned, this is well illustrated by the recent work of Berstel and de Luca [2]. Wen and Wen [7] have looked at a particular set of factors of the Fibonacci word, they call singular factors. They are the consecutive factors of the Fibonacci word of lengths F_0 , F_1 , F_2 , etc, where $(F_n)_{n\geq 0}$ is the Fibonacci sequence given by $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$ $(n \geq 1)$.

Our work started from a remark by Jean Berstel linking the singular factors to the Lyndon words appearing in the Lyndon factorization of the Fibonacci word we gave in [5]. Our investigation not only confirmed the remark made by J. Berstel but also lead us to a full description of the link between the singular factors and the Lyndon factors of the Fibonacci word. This paper is the result of this work. We were also able to generalize our construction to characteristic sturmian words, a set of infinite words of which the Fibonacci word is the most famous example.

¹Work partially supported by Conseil Régional d'Aquitaine

²This is the title of the forthcoming book by Lothaire.

The paper is structured as follows. The first section briefly describes the results in [7] concerned with the present work. More precisely, we define the singular factors of the Fibonacci word, list some of their properties and state the two main theorems in [7]. The next section begins by recalling the Lyndon factorization of the Fibonacci word given in [5]. We then study the link between singular factors and Lyndon factors of the Fibonacci word. Expressing Lyndon words in terms of the singular factors leads to two theorems from which we are able to deduce the two main results in [7]. We conclude by sketching the generalization of our results to characteristic sturmian words.

2 Singular factors

Throughout the paper, we only consider the two letter alphabet $A = \{a, b\}$. We totally order A by a < b and extend this order to the set A^* of all words lexicographically. The notations we use are those usual in theoretical computer science (see [4]). We shall make great use of the notation wa^{-1} , denoting the word obtained from w by deleting an a at the end of w (if possible). Let us start by recalling the definition of the Fibonacci word.

Definition 2.1 Let $f_0 = b$, $f_1 = a$ and define $f_{n+1} = f_n f_{n-1}$, for $n \ge 1$. The words f_n $(n \ge 0)$ are usually called the finite Fibonacci words. Hence, e.g., $f_2 = ab$, $f_3 = aba$, $f_4 = abaab$, and so on. Since f_n is a left factor of f_{n+1} for all $n \ge 0$, we may consider the (right) infinite word

The word f is called the (infinite) Fibonacci word.

For more details on the Fibonacci word, the reader is referred to J. Berstel's recent survey on sturmian words [1].

- **Remarks 2.2** 1. The length of f_n is the *n*th Fibonacci number F_n (where the Fibonacci sequence is defined by $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$, for $n \ge 1$).
 - 2. Moreover, for all $n \ge 2$, we have $(|f_n|_a, |f_n|_b) = (F_{n-2}, F_{n-1})$.
 - 3. For all $n \ge 1$, the word f_{2n} ends with ab and the word f_{2n+1} ends with ba.

Definition 2.3 ([7, page 589])

Let $n \geq 2$ and suppose f_n ends with $\alpha\beta$ (with $\alpha, \beta \in A$ and $\alpha \neq \beta$). We define the word w_n by $w_n = \alpha f_n \beta^{-1}$.

The word w_n is a factor of the Fibonacci word f and is called the nth singular factor of f. We also define $w_0 = a$, $w_1 = b$; it is useful to set $w_{-1} = \epsilon$ (the empty word).

Hence, we have $w_2 = aa$, $w_3 = bab$, $w_4 = aabaa$, $w_5 = babaabab$, and so forth.

Remarks 2.4 We collect here some remarks from [7].

- 1. The length of w_n is the *n*th Fibonacci number F_n . Let us verify that w_n is indeed a factor of f. As is known, any conjugate of a factor of f is also a factor of f. Hence, the word $af_{2n}f_{2n+1}a^{-1}$ is a factor of f, since it is conjugated to f_{2n+2} . Thus w_{2n} and w_{2n+1} are factors of f since they are consecutive factors of $af_{2n}f_{2n+1}a^{-1} = (af_{2n}b^{-1})(bf_{2n+1}a^{-1})$.
- 2. Note that, for all $n \ge 0$, we have

$$(|w_n|_a, |w_n|_b) = \begin{cases} (|f_n|_a + 1, |f_n|_b - 1) & \text{if } n \text{ is odd} \\ (|f_n|_a - 1, |f_n|_b + 1) & \text{if } n \text{ is even} \end{cases}$$

3. As a consequence, w_n is not conjugated to f_n .

We are now able to formulate [7]'s first fundamental result:

Theorem 2.5 ([7, Theorem 1]) *We have:*

$$f = \prod_{j=0}^{\infty} w_j.$$

That is,

$$f = (a)(b)(aa)(bab)(aabaa)(babaabab)\cdots$$

The set of factors of the Fibonacci word has received great attention from a large number of authors (again, see [1]; see also [2]). From this point of view, the next fundamental result of [7] is the following:

Theorem 2.6 ([7, Theorem 2])

Two occurences of the singular factor w_m $(n \ge 0)$ never overlap. Denote these occurences by $w_{m,1}$, $w_{m,2}$, $w_{m,3}$, and so forth (from left to right). Then we have:

$$f = (\prod_{j=0}^{m-1} w_j)(w_{m,1} z_1 w_{m,2} z_2 w_{m,3} z_3 \cdots)$$

where $z_k \in \{w_{m+1}, w_{m-1}\}$, for all $k \ge 1$, and $z_1 z_2 z_3 \cdots$ is the Fibonacci word over the alphabet $\{w_{m+1}, w_{m-1}\}$.

For example, with m = 2, we have $w_m = aa$, $w_{m+1} = bab$ and $w_{m-1} = b$. Thus:

 $f = (a \ b)(\underline{aa} \ bab \ \underline{aa} \ b \ \underline{aa} \ b \ \underline{ab} \ \underline{aa} \ b \ \underline{ab} \ \underline{aa} \ \cdots)$

Note that the theorem is true also for m = 0 (recall that $w_{-1} = \epsilon$). The word $z_1 z_2 z_3 \cdots$ is then equal to the Fibonacci word over the jj alphabet $j_i \{b, \epsilon\}$.

3 Lyndon factorization

Lyndon words are words strictly smaller than their proper right factors. Although these may be defined over an arbitrary alphabet, we shall restrict ourselves to the two letter alphabet $A = \{a, b\}$. For instance, letters are Lyndon words. The words ab, abb, aab, aabb, etc, are Lyndon words. Denote by Lthe set of Lyndon words (over A). More generally, given $u, v \in L$, we have: $uv \in L \Leftrightarrow u < v$. Hence, e.g., aababb is a Lyndon word. For more details concerning Lyndon words, the reader is referred to [4, Chap. 5].

Any Lyndon word ℓ of length at least two is a product of two Lyndon words u, vwith u < v. For example, we have aababb = (a)(ababb), aababb = (aab)(abb)and aababb = (aabab)(b). The standard factorization of ℓ is obtained by taking v of maximal length. We usually denote the standard factorization of ℓ by $\ell = \ell'\ell''$. Hence, e.g., (aababb)' = a and (aababb)'' = ababb. The Lyndon tree associated with the Lyndon word ℓ is the (planar rooted binary) tree obtained by computing, recursively down to letters, the standard factorization of ℓ' and ℓ'' , and that of $(\ell')'$, and $(\ell')''$ and so on. Figure 1 shows the Lyndon tree associated with $\ell = aababb$. Note that each Lyndon tree is complete, that is, every interior vertex has both a right and left son. We will only deal with complete planar rooted binary trees, having their leaves labelled by letters of A, which will simply be called trees from now on.

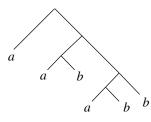


Figure 1: The Lyndon tree associated with $\ell = aababb$

The fundamental result concerning Lyndon words is the factorization theorem:

Theorem 3.1 ([3], see also [4])

Any non empty word is a unique product of non increasing Lyndon words. That is, given any non empty word $w \in A^*$, there exist $\ell_1, \ldots, \ell_n \in L$ $(n \ge 1)$, with $\ell_1 \ge \cdots \ge \ell_n$ such that $w = \ell_1 \cdots \ell_n$.

Proposition 3.2 ([5, Prop. 3.2]) Let $\varphi : A^* \to A^*$ be the morphism defined by $a \mapsto aab$ and $b \mapsto ab$. Define words by $\ell_0 = ab$ and $\ell_{n+1} = \varphi(\ell_n)$, for $n \ge 0$. Then $(\ell_n)_{n\ge 0}$ is a sequence of decreasing Lyndon words and we have

$$f = \prod_{n=0}^{\infty} \ell_n.$$
(1)

Thus, we have $f = (ab)(aabab)(aabaababaabab)\cdots$

- **Remarks 3.3** 1. The length of ℓ_n is F_{2n+2} . This is easily verified by using the morphism φ and by noting that $|\varphi(w)|_a = 2|w|_a + |w|_b$ and $|\varphi(w)|_b = |w|_a + |w|_b$.
 - 2. J. Berstel had pointed out that the Lyndon words ℓ_0 , ℓ_1 , ℓ_2 , ... were concatenation of two consecutive singular factors, i.e. $\ell_0 = ab = (a)(b) = w_0w_1$, $\ell_1 = abbab = (aa)(bab) = w_2w_3$, etc. This is in accordance with the fact that $|\ell_n| = F_{2n+2} = F_{2n+1} + F_{2n} = |w_{2n+1}| + |w_{2n}|$. Now, the word ℓ_n is also equal to $af_{2n}f_{2n+1}a^{-1}$, as may be verified by induction. Hence, J. Berstel's claim is correct since $af_{2n}f_{2n+1}a^{-1} = (af_{2n}b^{-1})(bf_{2n+1}a^{-1})$.
 - 3. As a consequence, Eq. (1) reproves Thm. 2.5 ([7, Theorem 1]).
 - 4. Note that, from the definition of the words ℓ_n , we find: $w_{2n+2} = \varphi(w_{2n})b^{-1}$ and $w_{2n+3} = b\varphi(w_{2n+1})$ $(n \ge 0)$.
 - 5. The morphism φ preserves the standard factorization of the words ℓ_n . More precisely, we have $\ell'_{n+1} = \varphi(\ell'_n)$ and $\ell''_{n+1} = \varphi(\ell''_n)$. This property has a geometrical interpretation: to obtain the Lyndon tree of ℓ_{n+1} one only needs to replace in that of ℓ_n the leaves labelled by a by the Lyndon subtree (a(a,b)) and those labelled by b by the Lyndon subtree (a,b). See figure 2.

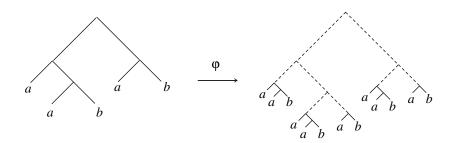


Figure 2: The tree structure of ℓ_n is preserved by φ

3.1 L-R operators

Rem. 3.3.2, proving $\ell_n = w_{2n}w_{2n+1}$, may be refined. For this, we need to define operators L and R corresponding to paths in a tree. The idea we describe here

is intuitively clear and is best described with pictures (see the figures), although we do need to tranlate it with proper notations. Let it be understood that Land R act on a given tree and let x be an interior vertex of that tree. Then, we denote by L.x (resp. R.x) the left (resp. right) son of x.

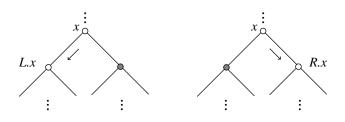


Figure 3: Operators R and L acting on trees

We will use sequences of operators L and R always acting from the root of Lyndon trees. Figure 4 illustrates the effect of the operator RLR over the Lyndon tree associated with the Lyndon word *aabaabab*. Note that any sequence of L - R operators is of the form $L^{a_{2n}}R^{a_{2n-1}}\cdots R^{a_1}L^{a_0}$ (with $a_0, a_{2n} \ge 0$ and $a_i > 0$ for all $1 \le i \le 2n - 1$) and acts from the right; that is,

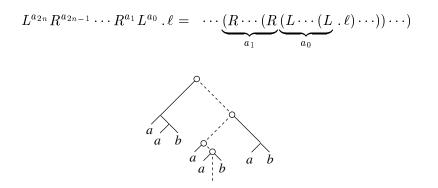


Figure 4: The decomposition induced by an R - L operator

Let T be the Lyndon tree associated with the word ℓ . For any vertex x of a tree, there is a unique path going from the root downto x described by a unique sequence of operators $L^{a_{2n}} R^{a_{2n-1}} \cdots R^{a_1} L^{a_0}$. Suppose that x is an interior vertex; then, the L - R path going from the root downto x determines a unique decomposition of ℓ as a product $\ell = uv$, with $u, v \in A^*$ non empty. We write $L^{a_{2n}} R^{a_{2n-1}} R^{a_1} L^{a_0} \cdot \ell = (u, v)$. The decomposition illustrated in figure 4 is precisely (RLR).aabaabab = (aabaa, bab). Note that, with this convention, the identity operator gives the decomposition (ℓ', ℓ'') cutting the tree of ℓ at its root.

Proposition 3.4 We have:

$$(RL)^n \cdot \ell_n = (w_{2n}, w_{2n+1}).$$

Moreover,

$$(RL)^{n-1}R \cdot \ell'_n = (w_{2n}, w_{2n-1}), \qquad (RL)^{n-1} \cdot \ell''_n = (w_{2n-2}, w_{2n-1}).$$

The proof is best understood using pictures; see the figures. We proceed by induction. Suppose $(RL)^n \cdot \ell_n = (w_{2n}, w_{2n+1})$ and that, moreover, the left and right sons of the vertex $(RL)^n \cdot \ell_n$ are leaves (figure 5).

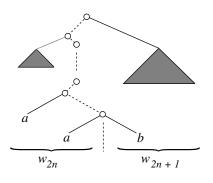


Figure 5: $(RL)^n \cdot \ell_n = (w_{2n}, w_{2n+1})$

The tree associated with ℓ_{n+1} is obtained from that associated with ℓ_n by replacing the leaves labelled by *a*'s with (a, (a, b)) and those labelled by *b*'s with (a, b) (cf Rem. 3.3.5). Hence, the factorization induced by the operator $(RL)^n$ on ℓ_{n+1} is $(RL)^n \cdot \ell_{n+1} = (\varphi(w_{2n}), \varphi(w_{2n+1}))$ (figure 6). Now, observe that w_{2n} ends with *aa*; thus, the left subtree of the vertex $(RL)^n \cdot \ell_{n+1}$ is (a, (a, b)).

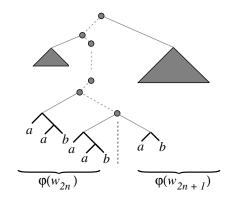
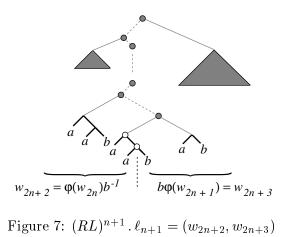


Figure 6: $(RL)^n \cdot \ell_{n+1} = (RL)^n \cdot \varphi(\ell_n) = (\varphi(w_{2n-1}), \varphi(w_{2n}))$

Since $w_{2n+2} = \varphi(w_{2n})b^{-1}$ and $w_{2n+3} = b\varphi(w_{2n+1})$ (cf Rem. 3.3.4), we see that the decomposition (w_{2n+2}, w_{2n+3}) is obtained by going down this left subtree following the path *RL*. Thus $(w_{2n+2}, w_{2n+3}) = RL (RL)^n \cdot \ell_{n+1} = (RL)^{n+1} \cdot \ell_{n+1}$ (figure 7).

The second part of the proposition is proved in a similar manner.



3.2 Invariance Property

In this section, we prove an invariance property of the factorization (1) which leads, as a corollary, to a new proof of Thm. 2.6 ([7, Theorem 2]).

Theorem 3.5 We have:

(i)

$$f = (\prod_{j=0}^{n-1} \ell_j) \varphi^n(f).$$
(2)

Moreover, $\varphi^n(f)$ is equal to the Fibonacci word over the alphabet $\{\ell'_n, \ell''_n\}$.

(ii) Furthermore, $\varphi^n(f)$ is also equal to the Fibonacci word over the alphabet $\{\ell_n, \ell'_n\}$.

We claim: for any $n \ge 0$, $\varphi^n(a) = \ell'_n$ and $\varphi^n(b) = \ell''_n$; consequently, the words $\ell'_m, \ell''_m \ (m \ge n)$ are Lyndon words over the alphabet $\{\ell'_n, \ell''_n\}$, which shall prove the theorem.

That $\{\ell'_n, \ell''_n\}$ and $\{\ell'_n, \ell_n\}$ may be considered as alphabets (or, more precisely, as *codes*) follows from the fact that ℓ'_n is not a suffix of ℓ''_n nor ℓ_n ; indeed, no word is both a suffix and a prefix of the same Lyndon word.

We see at once that ℓ_n is the image of the word ab under φ^n ; that is, it is a Lyndon word over $\{\ell'_n, \ell''_n\}$. Similarly, one sees that ℓ_m (m > n) is the image of ℓ_{m-n} under the morphism φ^n . That shows part (i). Part (ii) follows from the fact that $\varphi^n(ab) = \ell_n$, again, and that the morphism $a \mapsto ab, b \mapsto a$ leaves f invariant (that is, f is equal to the Fibonacci word over the alphabet $\{ab, a\}$).

3.3 A New Proof of Thm. 2.6

Recall that $\ell_n = w_{2n}w_{2n+1}$ and that, by Prop. 3.4, we have $\ell'_n = w_{2n}w_{2n-1}$ and $\ell''_n = w_{2n-2}w_{2n-1}$. Thus, in case (i), Eq. (2) reads:

$$f = (\prod_{j=0}^{n-1} w_{2j} w_{2j+1}) f_{\{w_{2n} w_{2n-1}, w_{2n-2} w_{2n-1}\}}$$

where $f_{\{w_{2n}w_{2n-1},w_{2n-2}w_{2n-1}\}}$ denotes the Fibonacci word over the alphabet $\{w_{2n}w_{2n-1}, w_{2n-2}w_{2n-1}\}$. Note that this is also equal to:

$$f = (\prod_{j=0}^{2n-2} w_j) w_{2n-1} f_{\{w_{2n}w_{2n-1}, w_{2n-2}w_{2n-1}\}}.$$

In other words, $f_{\{w_{2n}w_{2n-1},w_{2n-2}w_{2n-1}\}}$ is obtained by first forming the Fibonacci word over the alphabet $\{w_{2n}, w_{2n-2}\}$ and then inserting *before* each occurence of w_{2n} or w_{2n-2} the word w_{2n-1} . This is precisely what says Thm. 2.6, for m = 2n - 1 odd.

In case (ii), we find:

$$f = (\prod_{j=0}^{2n-1} w_j) f_{w_{2n} w_{2n+1}}, \{w_{2n} w_{2n-1}, w_{2n-1},$$

which provides a proof for m = 2n even.

4 Generalization to characteristic sturmian words

In this section, we give results generalizing Prop. 3.4 and Thm. 3.5. Their proofs are technical; we shall limit ourselves to their statements.

Let us first recall some definitions.

Definition 4.1 Let $(c_n)_{n\geq 0}$ be a sequence of integers satisfying $c_0 \geq 0$ and $c_n > 0$, for n > 0. Define $s_0 = b$, $s_1 = a$ and $s_{n+1} = s_n^{c_{n-1}}s_{n-1}$. Then $s = \lim_{n \to \infty} s_n$ is a well defined infinite word.

The sequence $(c_n)_{n\geq 0}$ is called the directive sequence of s. Moreover, s is a characteristic sturmian word; for more details, see [1].

The Fibonacci word is a special case of a sturmian word, with $c_n = 1$, for all $n \ge 0$. Hence, it seems natural to look for a generalization of the results in the previous sections. In [5], we gave the Lyndon factorization of any general characteristic sturmian word s, namely we proved:

$$s = \prod_{n=0}^{\infty} [(as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1}]^{c_{2n+1}}$$

where $((as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1})_{n\geq 0}$ is a sequence of strictly decreasing Lyndon words. We write $\ell_n = (as_{2n+1}a^{-1})^{c_{2n}-1}as_{2n}s_{2n+1}a^{-1}$.

Definition 4.2 Suppose the sequence $(c_n)_{n\geq 0}$ is given. Let $n\geq 2$ and suppose s_n ends with $\alpha\beta$ (with $\alpha, \beta \in A$ and $\alpha \neq \beta$). We define the word w_n by $w_n = \alpha f_n \beta^{-1}$. We also define $w_0 = a, w_1 = b$.

Hence, e.g., $w_2 = a^{c_0+1}$, $w_3 = b(a^{c_0}b)^{c_1}$, and so on (since $s_2 = s_1^{c_0}s_0 = a^{c_0}b$, $s_3 = s_2^{c_1}s_1 = (a^{c_0}b)^{c_1}a$, etc). Those words w_n are factors of the word s associated with the sequence $(c_n)_{n\geq 0}$ and play a role analog to the singular factors studied in the preceding sections. Write $u_n = as_{2n}s_{2n+1}a^{-1}$; then u_n is a Lyndon word. Moreover, we have $u'_n = as_{2n+1}a^{-1}$. This a key fact in proving that $(\ell_n)_{n\geq 0}$ is a sequence of decreasing Lyndon words (see [5]). Observe that $u_n = w_{2n}w_{2n+1}$. Many other properties satisfied by the words w_n confirm them as the proper generalization of the singular factors for the characteristic sturmian word sassociated to $(c_n)_{n\geq 0}$. We shall only state one of these properties generalizing Prop. 3.4; it is illustrated in figure 8.

Proposition 4.3 We have:

$$R^{c_0}L^{c_1}\cdots R^{c_{2n}}L^{c_{2n+1}}$$
. $u_n = (w_{2n}, w_{2n+1})$.

We may formulate an invariance property analog to Thm. 3.5.

Theorem 4.4 We have:

$$s = \left(\prod_{j=0}^{n-1} \ell_j^{c_{2j+1}}\right) \times \bar{s}$$

where \bar{s} is the sturmian word with directive sequence $(d_m)_{m\geq 0}$ over the alphabet $\{u'_n, u''_n\}$, where $d_m = c_{m+2n}$.

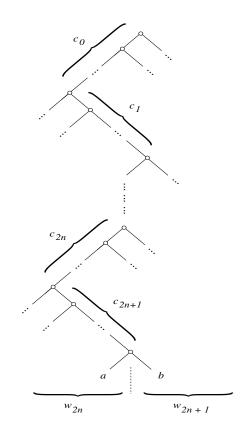


Figure 8: The R - L path giving (w_{2n}, w_{2n+1}) in the general case

References

- BERSTEL J. Recent Results in Sturmian Words. Invited paper to DLT'95. To be published by World Scientific, http://www-igm.univ-mlv.fr/~berstel/, 1996.
- [2] BERSTEL J. and de LUCA A. Sturmian Words, Lyndon Words and Trees. *Theoretical Computer Science*, to appear.
- CHEN K. T., FOX R. H., and LYNDON R. C. Free Differential Calculus, IV

 The Quotient Groups of the Lower Central Series. Annals of Mathematics, 68:81 – 95, 1958.
- [4] LOTHAIRE M. Combinatorics on Words. Addison Wesley, 1983.
- [5] MELANÇON G. Lyndon Factorization of Sturmian Words. Discrete Mathematics, to appear. Special Issue for FPSAC'96, 8th international Conference on Formal Power Series and Algebraic Combinatorics, STANTON D. and LEROUX P., eds.

- [6] SIROMONEY R., MATTHEW L., DARE V. R., and SUBRAMANIAN K.
 G. Infinite Lyndon Words. *Information Processing Letters*, 50:101 104, 1994.
- [7] WEN Z.-X. and WEN Z.-Y. Some Properties of the Singular Words of the Fibonacci Word. *European Journal of Combinatorics*, 15:587–598, 1994.