Fundamenta Informaticae XX (2012) 1–9 1
IOS Press

Some characterizations of Sturmian words in terms of the lexicographic order

Michelangelo Bucci[*]
Department of Mathematics, University of Turku
FI-20014 Turku, Finland
michelangelo.bucci@utu.fi

Luca Q. Zamboni[*†]
Université de Lyon, Université Lyon 1
CNRS UMR 5208 Institut Camille Jordan
Bâtiment du Doyen Jean Braconnier
43, blvd du 11 novembre 1918
F-69622 Villeurbanne Cedex, France

Department of Mathematics, University of Turku
FI-20014 Turku, Finland
luca.zamboni@utu.fi

Alessandro De Luca[*]
Dipartimento di Scienze Fisiche
Università degli Studi di Napoli Federico II
via Cintia, Monte S. Angelo
I-80126 Napoli, Italy

Department of Mathematics, University of Turku
FI-20014 Turku, Finland
alessandro.deluca@utu.fi

Abstract. In this paper we present three new characterizations of Sturmian words based on the lexicographic ordering of their factors.

Keywords: Sturmian words, lexicographic order

1. Introduction

Let $w \in A^\omega$ be an infinite word with values in a finite alphabet A. The *(block) complexity function* $p_w : \mathbb{N} \to \mathbb{N}$ assigns to each n the number of distinct factors of w of length n. A fundamental result due to Hedlund and Morse [**?**] states that a word w is ultimately periodic if and only if for some n the

[*]Partially supported by the FiDiPro grant "Words, Numbers, Tilings and Applications" from the Academy of Finland
[†]Partially supported by ANR grant *SUBTILE*

complexity $p_w(n) \leq n$. Sequences of complexity $p(n) = n + 1$ are called *Sturmian words*. The most studied Sturmian word is the so-called Fibonacci word

$$01001010010010100101001001010010\ldots$$

fixed by the morphism $0 \mapsto 01$ and $1 \mapsto 0$. In [**?**] Hedlund and Morse showed that each Sturmian word may be realized geometrically by an irrational rotation on the circle. More precisely, every Sturmian word is obtained by coding the symbolic orbit of a point x on the circle (of circumference one) under a rotation by an irrational angle α where the circle is partitioned into two complementary intervals, one of length α and the other of length $1 - \alpha$. And conversely each such coding gives rise to a Sturmian word. The irrational α is called the *slope* of the Sturmian word. An alternative characterization using continued fractions was given by Rauzy in [**?**] and [**?**], and later by Arnoux and Rauzy in [**?**]. Sturmian words admit various other types of characterizations of geometric and combinatorial nature (see for instance [**?**]). For example they are characterized by the following balance property: A word w is Sturmian if and only if w is a binary aperiodic (non-ultimately periodic) word with the property that for any two factors u and v of w of equal length, we have $-1 \leq |u|_i - |v|_i \leq 1$ for each letter i. Here $|u|_i$ denotes the number of occurrences of i in u. In this paper, we establish some new characterizations of Sturmian words in terms of the lexicographic order behavior of its factors. We prove:

Theorem 1.1. An infinite word w containing the letters 0 and 1 is Sturmian if and only if for every pair of lexicographically consecutive factors v, v' of the same length, there exist λ, μ such that v, v' either both belong to $\{\lambda 01\mu, \lambda 10\mu\}$ or both belong to $\{\lambda 0, \lambda 1\}$.

Actually our first main result is later formulated in more general terms. The fact that this property holds for Sturmian words has recently been shown in [**?**], and is a direct consequence of a result proved in [**?**].

Our second characterization requires the additional hypothesis of recurrence:

Theorem 1.2. Let w be a recurrent aperiodic binary word over the alphabet $\{0, 1\}$ and $v, v' \in \mathrm{Fact}(w)$. Then the following are equivalent:

1. w is Sturmian.
2. For all factors v, v' of w of equal length, if $v <_{\mathrm{lex}} v'$ then $|v|_1 \leq |v'|_1$.
3. For any pair of lexicographically consecutive factors v, v' of the same length, v and v' differ in at most two positions.

2. Preliminaries

In this section, we introduce the tools which will be used in the rest of the paper.

2.1. Standard notions in combinatorics on words

We will report here the standard notations and notions in combinatorics on words that will be used in the rest of the paper. For further results on the subject we refer the reader to [**?**].

By an *alphabet* we mean a finite non empty set A. The elements of A are called *letters*. We let A^* denote the free monoid over A, i.e. the set of finite sequences of elements of A equipped with the concatenation product. The neutral element of A^* will be called the *empty word* and is denoted ε. The set of nonempty words over A, i.e. the free semigroup over A, is denoted A^+. With the multiplicative notation, given a positive integer n and a word w, we let w^n denote the concatenation of n copies of w. For each word w, we put $w^0 = \varepsilon$.

Two words v, v' are said to be *conjugates* one of the other if there exist λ, μ such that $v = \lambda\mu$ and $v' = \mu\lambda$.

If a nonempty word x is such that $x = x_1 x_2 \cdots x_k$, with $x_i \in A$ for $1 \leq i \leq k$, then k is called the *length* of x and is denoted $|x|$. The length of the empty word is taken to be 0.

We say that a word v is a *factor* of another word w if there exist two words λ, μ such that $w = \lambda v \mu$. If $\lambda = \varepsilon$ (resp. $\mu = \varepsilon$) we call v a *prefix* (resp. a *suffix*) of w. If v is both a prefix and a suffix of w, we say that v is a *border*. A factor v of w is called *proper* if $|v| < |w|$. We denote with $\mathrm{Fact}(w)$ the set of all factors of the word w. A word w is said to be *unbordered* if the only borders of w are w and ε.

Most of the above definitions can be extended to the set A^ω of infinite words on the alphabet A. For $w, w' \in A^\omega$, we say w' is a *tail* of w if $w = vw'$ for some $v \in A^*$. If v is not empty, we call w' a *proper tail* of w.

We call an *occurrence* of v in w a word λ such that λv is a prefix of w. An infinite word w is said to be *recurrent* if each of its factors (or, equivalently, of its prefixes) has infinitely many occurrences in w. Given $v, w \in A^*$ we let $|w|_v$ denote the number of occurrences of v in w and set

$$\mathrm{Alph}(w) = \{x \in A \,|\, |w|_x > 0\}.$$

A factor v of w is *unioccurrent* if $|w|_v = 1$, i.e., if v occurs in w exactly once.

We say that an infinite word w is *periodic* if it can be expressed as an infinite concatenation of a finite word v, i.e. $w = v^\omega$. We say that an infinite word is *ultimately periodic* if it has a periodic tail. Otherwise we say w is *aperiodic*. It is easy to show that any infinite word that contains itself as a proper tail is periodic.

2.2. Lexicographic order

Let A be an alphabet equipped with a total order $<$. Then $<$ extends naturally to a partial order on A^*, denoted $<_{\mathrm{lex}}$, in the following way: We write $v <_{\mathrm{lex}} v'$ (and say v is *lexicographically* smaller than $v',)$ if $|v| = |v'|$ and there exists a word λ and two letters $a < b$ such that λa is a prefix of v and λb is a prefix of v'. Two words v, v' are said to be lexicographically *consecutive* or *adjacent* if $v <_{\mathrm{lex}} v'$ and there is no word w such that $v <_{\mathrm{lex}} w <_{\mathrm{lex}} v'$.

We say a factor v of a word w is *maximal* (resp. *minimal*) in w if there exists no factor v' such that $v <_{\mathrm{lex}} v'$ (resp. $v' <_{\mathrm{lex}} v$), thus omitting the sentence "with respect to the lexicographic order". We will say that v is *extremal* in w if it is either minimal or maximal.

Given two factors v, v' of a word w such that $v <_{\mathrm{lex}} v'$, we will write $v' = \mathrm{succ}_w(v)$ if there is no $f \in \mathrm{Fact}(w)$ such that $v <_{\mathrm{lex}} f <_{\mathrm{lex}} v'$. Notice that if $v \in \mathrm{Fact}(w)$ is non extremal, then there exist $f_1, f_2 \in \mathrm{Fact}(w)$ such that $f_1 = \mathrm{succ}_w(v)$ and $v = \mathrm{succ}_w(f_2)$.

Remark 2.1. It is easy to show that if v is a unioccurrent prefix of an infinite word w and v is extremal in w, then every prefix of w longer than v is unioccurrent, and extremal of the same kind.

We can extend the definition of lexicographic order to infinite words in a natural way, saying that the infinite word w is lexicographically smaller than w' if w has a prefix which is lexicographically smaller than a prefix (of the same length) of w'. The notion of extremality extends as well: we say that an infinite word w is minimal (resp. maximal) if it is lexicographically smaller (resp. larger) than all its tails.

Remark 2.2. It is clear that if aw and w are both extremal infinite binary words (and a is a letter), then they are extremal of the same kind (i.e. they are both minimal or maximal).

2.3. Sturmian words

Let v and v' be factors of w with $|v| = |v'|$. We say the pair (v, v') is *balanced* if $||v|_x - |v'|_x| \leq 1$ for each letter $x \in A$. Otherwise the pair (v, v') is said to be *imbalanced*. A word w is called balanced if all pairs of factors of w of the same length are balanced.

A binary word w is called *Sturmian* if w is aperiodic and balanced. As mentioned earlier, Sturmian words are also defined in terms of the block complexity function $p_w : \mathbb{N} \to \mathbb{N}$ which assigns to each n the number of distinct factors of w of length n: w is Sturmian if and only if $p_w(n) = n + 1$ for each $n \geq 0$.

For each Sturmian word $w \in \{0, 1\}^\omega$ we set

$$\Omega_w = \{w' \in \{0, 1\}^\omega \mid \mathrm{Fact}(w') = \mathrm{Fact}(w)\}.$$

Thus Ω_w is the *shift orbit closure* or *subshift* generated by w. The proof of the following proposition is in [**?**].

Proposition 2.3. Let w be a Sturmian word over the alphabet $\{0, 1\}$. Then there exists a unique word γ in Ω_w such that both 0γ and 1γ are in Ω_w.

Remark 2.4. The word γ in Proposition 2.3 is called the *characteristic* word of w and it is known that the prefixes of 0γ are lexicographically minimal among the factors of w, while the prefixes of 1γ are maximal.

We say that a factor v of a Sturmian word w is a *Christoffel word* if v is unbordered. We group into the next statement the well-known properties of Christoffel words that we will need in the rest of the paper (see for instance [**?**, **?**], [**?**, Prop. 5], [**?**, Prop. 6]).

Proposition 2.5. Let w be a Sturmian word over the alphabet $\{0, 1\}$ and let $v \in \mathrm{Fact}(w)$ be a Christoffel word such that $|v| > 1$. Then there exists u such that:

1. v is either $0u1$ or $1u0$, and they are both Christoffel words in w;
2. $0u1$ and $1u0$ are the only Christoffel words of length $|v|$ in w and are conjugates;
3. all conjugates of v are in $\mathrm{Fact}(w)$;
4. exactly one between $0u0$ and $1u1$ is a factor of w and is extremal in w;
5. the factors of w of length $|v|$ are either conjugates of v or of type xux.

Those factors of a Sturmian word having the same length as a Christoffel word, but not conjugate to a Christoffel word (i.e. the factors xux in the preceding proposition), are called *singular words* of the Sturmian word.

3. Main Result

We begin with the following key proposition:

Proposition 3.1. Let $w \in \{0,1\}^\omega$ be an imbalanced word. Then there exists a factor $u \in \mathrm{Fact}(w)$ of minimal length such that $0u0, 1u1$ are in $\mathrm{Fact}(w)$. Furthermore, either $10u0$ and $01u1$ are both factors of w or there exists a unique letter x such that xux is a prefix of w and occurs in w only finitely many times. In the latter case every prefix of w is extremal in w.

Proof:
Since w is not balanced, there exists an imbalanced pair (v, v') consisting of factors v and v' of w. It is well known (see [**?**]) that the imbalanced pair of minimal length is of the form $(0u0, 1u1)$ for some factor u of w and is unique. If both $10u0$, $01u1$ are factors of w we are done. So let us assume that there exists a letter $x \in \{0,1\}$ such that no occurrence of xux in w is preceded by $1-x$. Then every internal (non-prefix) occurrence of xux in w is preceded by x. We begin by showing that xux is a prefix of w from which it follows that x is unique. Without loss of generality we can assume that $x = 0$. Suppose that the first occurrence of $0u0$ in w occurs in position $n \geq 0$. If $n = 0$ we are done. So suppose $n > 0$. Then $00u$ is a factor of w occurring in position $n-1$ and the pair $(00u, 1u1)$ is imbalanced. By uniqueness of the shortest imbalanced pair we have that $00u = 0u0$ and hence $0u0$ also occurs in position $n-1$, a contradiction on the minimality of n. This also shows that if $0u0$ occurs in position t then it also occurs in each position r for $0 \leq r \leq t$. Thus $0u0$ occurs only finitely many times in w (for otherwise w would be 0^ω and thus not binary).

We next show that every prefix of w is minimal (if we had taken $x = 1$ then each prefix of w would be maximal). We proceed by contradiction. Let $n > 0$ be the least positive integer for which there exists a factor v' of w in position n which is lexicographically smaller than the corresponding prefix v of w of the same length. Then either there exists a proper prefix u' of u such that $0u'1$ is a prefix of v and $0u'0$ is a prefix of v', or v' begins in $0u0$. In the first case $0u'0$ and the prefix $1u'1$ of $1u1$ constitute a shorter imbalanced pair contradicting the minimality of $|u|$. In the second case v' is an internal occurrence of $0u0$ and is hence preceded by 0. Thus the factor v'' in position $n-1$ of length $|v'|$ is lexicographically smaller than v' and also smaller than v, contradicting the minimality of n. □

The next proposition introduces the main subject of this paper:

Proposition 3.2. Fix $k \geq 1$. Let $A = \{0, 1, \ldots, k\}$ be an ordered alphabet such that $0 < 1 < \cdots < k$ and w an infinite word such that $\mathrm{Alph}(w) = A$. The following are equivalent:

1. For every $v, v' \in \mathrm{Fact}(w)$ with $v' = \mathrm{succ}_w(v)$, there exist distinct letters $a < b$ in A and $\lambda, \mu \in A^*$ such that

$$\begin{cases} v &= \lambda ab\mu \\ v' &= \lambda ba\mu \end{cases} \quad \text{OR} \quad \begin{cases} v &= \lambda a \\ v' &= \lambda b \end{cases}$$

2. For every $v, v' \in \mathrm{Fact}(w)$ with $v' = \mathrm{succ}_w(v)$, there exist $m \in A$ and $\lambda, \mu \in A^*$ such that

$$\begin{cases} v &= \lambda m(m+1)\mu \\ v' &= \lambda(m+1)m\mu \end{cases} \quad \text{OR} \quad \begin{cases} v &= \lambda m \\ v' &= \lambda(m+1) \end{cases}$$

3. $A = \{0, 1\}$ and for every $v, v' \in \mathrm{Fact}(w)$ with $v' = \mathrm{succ}_w(v)$, there exist $\lambda, \mu \in A^*$ such that

$$\begin{cases} v &= \lambda 01\mu \\ v' &= \lambda 10\mu \end{cases} \quad \text{OR} \quad \begin{cases} v &= \lambda 0 \\ v' &= \lambda 1 \end{cases}$$

Proof:
Clearly $(3) \Rightarrow (2) \Rightarrow (1)$. To see that $(1) \Rightarrow (3)$ it suffices to show that (1) implies that $k = 1$. We first note that if $ab \in \mathrm{Fact}(w)$, then $b \in \{a-1, a, a+1\}$. In fact suppose that $a \neq b$. Then either $a < b$ or $b < a$. We consider the first case as the latter case is essentially identical. Let $x, y \in A$ such that ax is the greatest factor of length 2 beginning with a and $(a+1)y$ be the smallest factor of length 2 beginning with $(a+1)$. Clearly $(a+1)y = \mathrm{succ}_w(ax)$, which, from the hypothesis implies that $x = a+1$ and $y = a$. Thus ab is lexicographically smaller or equal to $a(a+1)$ from which it follows that $b = a+1$.

Now suppose to the contrary that $k > 1$, and consider the shortest factor v of w containing both 0 and 2. Then, from what we just proved, $v = 01^n2$ or $v = 21^n0$ for some $n > 0$. We will show that neither occurs in w. Suppose to the contrary that the first is a factor of w and consider the least $n > 0$ for which 01^n2 is a factor of w. Then as 01^n2 is not maximal, its successor is either of the form $101^{n-1}2$ or $01^{n-1}21$ or 01^nx for some $2 < x$. The first two cases contradict the minimality of n while the last case implies that $1x$ is a factor of w for some $2 < x$, again a contradiction. Similarly it is verified that $v = 21^n0$ is never a factor of w. Hence $k = 1$. □

Definition 3.3. We say that an infinite word w has the "Nice Factors Ordering property" (NFOp) if for w one of the equivalent conditions of Proposition 3.2 holds.

Remark 3.4. It is useful to stress that having the NFOp implies that the word w is actually binary. Also it is easy to see that NFOp actually characterizes the pairs of adjacent factors with respect to the lexicographic ordering, i.e., If w satisfies NFOp and v and v' are factors of w with $v = \lambda 01\mu$ and $v' = \lambda 10\mu$ or $v = \lambda 0$ and $v' = \lambda 1$, then $v' = \mathrm{succ}_w(v)$.

Lemma 3.5. If an infinite word w has the NFOp, then w is aperiodic.

Proof:
Let as assume by contradiction that there exist $w', v \in A^*$ with $w = w'v^\omega$. Then w has finitely many tails and it is readily proved that they must respect the NFOp, i.e. if x and y are two lexicographically consecutive tails of w then we can write

$$x = z01z' \quad y = z10z'.$$

In particular, this implies that every tail contains either 01 or 10, hence v cannot be a single letter. As v^ω contains both 01 and 10, v^ω contains tails of the form $(01v')^\omega$ and $\mu = (10v'')^\omega$ for some $v', v''$with $|v'| = |v''| = |v| - 2$. Clearly these two tails differ in an infinite number of positions. On the other hand w has only a finite number of tails and by assumption any two lexicographically consecutive tails differ in exactly two positions. Hence we obtain a contradiction. □

Lemma 3.6. Let w be an infinite word with the NFOp. Then there exists no factor u in $\mathrm{Fact}(w)$ such that $10u0$ and $01u1$ are both factors of w.

Proof:
Suppose to the contrary that there exists a shortest factor u such that both $10u0$ and $01u1$ are factors of w. Since $01u1 <_{\mathrm{lex}} 10u0$, but the two factors cannot be consecutive as they differ in at least three positions, the successor v of $01u1$ satisfies $01u1 <_{\mathrm{lex}} v <_{\mathrm{lex}} 10u0$. It follows that there exists a proper prefix λ of u such that $01\lambda0$ is a prefix of $01u1$ and $01\lambda1$ is a prefix of v (notice that v cannot begin with 1 since otherwise it would be $10u1$ and thus would be lexicographically larger than $10u0$). Since $10\lambda0$ is a prefix of $10u1$ the factors $01\lambda1$ and $10\lambda0$ contradict the minimality of $|u|$. □

The following result is a direct consequence of a result proved by the third author together with Jenkinson in [**?**] and, more recently, has appeared in [**?**]; we include it here with a different proof, for the sake of completeness.

Proposition 3.7. Let w be a Sturmian word on the alphabet $\{0,1\}$. Then w satisfies NFOp.

Proof:
Let $0u1$ be a Christoffel factor of w. As $1u0$ is a conjugate of $0u1$ it follows that each factoring $u = xy$ determines two conjugates of $0u1$, namely $v = y01x$ and $v' = y10x$. By Proposition 2.5, v and v' are factors of w; let $z = \mathrm{succ}_w(v)$. As $v <_{\mathrm{lex}} v'$, the longest common prefix of v and z is at least y. In fact it cannot be longer, otherwise we could write $v = y01x'0\lambda$ and $z = y01x'1\mu$ for some words x', λ, μ; as $v' = y10x'0\lambda$, we would have $0x'0, 1x'1 \in \mathrm{Fact}(w)$, a contradiction since w is balanced.

Similarly, y is also the longest common prefix between v' and the word z' such that $\mathrm{succ}_w(z') = v'$. It follows $z = v'$ and $v = z'$, i.e., v and v' are lexicographically consecutive. Thus any two consecutive conjugates of a Christoffel word in w satisfy the first condition in (3) of Proposition 3.2. More generally, if z and z' are lexicographically consecutive factors of w, then there exists a Christoffel factor $0u1$ and two consecutive conjugates v and v' of $0u1$ with z a prefix of v and z' a prefix of v'. The result now follows. □

Before proceeding to prove our main result, we need the following:

Lemma 3.8. Let $w \in \{0,1\}^\omega$ be a Sturmian word and $x \in \{0,1\}$. If xw satisfies NFOp then xw is Sturmian.

Proof:
We proceed by contradiction by supposing that w is Sturmian, xw satisfies NFOp and that xw is not Sturmian. Without loss of generality we can assume that $x = 0$. It follows that there exists u such that both $0u0$ and $1u1$ are factors of $0w$. Since w is Sturmian, it follows from Proposition 3.1 that $0u0$ is a unioccurrent prefix of $0w$ and every prefix of $0w$ is minimal in $0w$. On the other hand, if γ denotes the characteristic word of w (which has u as a prefix), then every prefix of 0γ is minimal in w. By NFOp it follows that for all $n > |u| + 2$ the prefixes of length n of $0w$ and 0γ can be written respectively as $0u01v_n$ and $0u10v_n$ for some word v_n. Hence there exists a tail v of w such that $0w = 0u01v$ and $0\gamma = 0u10v$. Thus $0v, 1v \in \Omega_w$, so that $v = \gamma$ and hence γ is a proper tail of itself, a contradiction. □

Theorem 3.9. Let w be an infinite word on the ordered alphabet $A = \{0, 1, \ldots, k\}$. The following statements are equivalent:

1. w is Sturmian over the alphabet $\{0, 1\}$.
2. w satisfies NFOp.

Proof:
That $(1) \Rightarrow (2)$ follows from Proposition 3.7. To see that $(2) \Rightarrow (1)$, we suppose that $w \in \{0,1\}^\omega$ satisfies NFOp and write $w = aw'$ with $a \in \{0, 1\}$. We need to show that w is Sturmian. By Lemma 3.5 w is aperiodic. If w is not Sturmian, then by Lemma 3.8 we deduce that w' is not Sturmian. Also, combining Proposition 3.1 and Lemma 3.6 we deduce that every prefix of w is extremal and hence w' also satisfies NFOp. In short, if $w = aw'$ satisfies NFOp and is not Sturmian, then every prefix of w is extremal and the tail w' satisfies NFOp and is not Sturmian. Thus writing $w' = bw''$ we deduce that every prefix of w' is extremal and w'' satisfies NFOp and is not Sturmian. Iterating this process indefinitely we deduce that for each tail v of w, each prefix of v is extremal in v. Since w is aperiodic it follows that there exists a tail v of w which begins in 01 and a tail v' of v which begins in 10. Since every prefix of v is minimal in v and every prefix of v' is maximal in v' it follows that 00 is not a factor of v and 11 is not a factor of v'. Hence $v' = (10)^\omega$, a contradiction. □

We next establish another characterization of Sturmian words based on the lexicographic order of their factors.

Theorem 3.10. Let w be a recurrent aperiodic binary word over the alphabet $\{0, 1\}$. Then the following are equivalent:

1. w is Sturmian.
2. For all factors and $v, v' \in \mathrm{Fact}(w)$ of equal length, if $v' = \mathrm{succ}_w(v)$ then v and v' differ in at most two positions.
3. For all factors and $v, v' \in \mathrm{Fact}(w)$ of equal length, if $v <_{\mathrm{lex}} v'$ then $|v|_1 \leq |v'|_1$.

Proof:
$(1) \Rightarrow (2)$: Since w is Sturmian, it has the NFOp by Theorem 3.9. The statement is clearly proven since the NFOp trivially implies condition (2) by definition.

$(2) \Rightarrow (3)$: Notice that condition (2) implies that if $f' = \mathrm{succ}_w(f)$, then there must exist $\lambda, \mu, \mu', x, x'$ with $|x| = |x'| \leq 1$ such that $f = \lambda 0 \mu x \mu'$ and $f' = \lambda 1 \mu x' \mu'$. Hence

$$|f|_1 = |\lambda|_1 + |\mu|_1 + |\mu'|_1 + |x|_1 \leq |\lambda|_1 + |\mu|_1 + |\mu'|_1 + 1 \leq |\lambda|_1 + |\mu|_1 + |\mu'|_1 + |x'|_1 + 1 = |f'|_1.$$

And thus, in particular $|f|_1 \leq |f'|_1$. Suppose $v <_{\mathrm{lex}} v'$. Then there must exist $v_0, \ldots, v_k$ such that $v = v_0$, $v' = v_k$ and for all $1 \leq n \leq k$, $v_n = \mathrm{succ}_w(v_{n-1})$, then

$$|v|_1 = |v_0|_1 \leq \cdots \leq |v_k|_1 = |v'|_1.$$

$(3) \Rightarrow (1)$: Assume w is not Sturmian; as w is aperiodic, it has to be imbalanced. Since w is recurrent, we have from Proposition 3.1 that there must exist u such that both $10u0$ and $01u1$ are factors of w. But clearly this is a contradiction, since $01u1 <_{\mathrm{lex}} 10u0$ and $|01u1|_1 = |10u0|_1 + 1$. This concludes the proof. □

Notice that, as opposed to Theorem 3.9, the above result actually needs the recurrence and aperiodicity hypotheses, as for example:

- the recurrent periodic word $(01)^\omega$ and the non-recurrent aperiodic word $00f$ (where f is the Fibonacci word) both respect condition (3), although neither is Sturmian,
- the non-recurrent ultimately periodic word 01^ω satisfies both (2) and (3), but it is not Sturmian.

Fundamenta Informaticae XX (2012) 1–9 1
IOS Press

Some characterizations of Sturmian words in terms of the lexicographic order

Michelangelo Bucci[*]
Department of Mathematics, University of Turku
FI-20014 Turku, Finland
michelangelo.bucci@utu.fi

Luca Q. Zamboni[*†]
Université de Lyon, Université Lyon 1
CNRS UMR 5208 Institut Camille Jordan
Bâtiment du Doyen Jean Braconnier
43, blvd du 11 novembre 1918
F-69622 Villeurbanne Cedex, France

Department of Mathematics, University of Turku
FI-20014 Turku, Finland
luca.zamboni@utu.fi

Alessandro De Luca[*]
Dipartimento di Scienze Fisiche
Università degli Studi di Napoli Federico II
via Cintia, Monte S. Angelo
I-80126 Napoli, Italy

Department of Mathematics, University of Turku
FI-20014 Turku, Finland
alessandro.deluca@utu.fi

Abstract. In this paper we present three new characterizations of Sturmian words based on the lexicographic ordering of their factors.

Keywords: Sturmian words, lexicographic order

1. Introduction

Let $w \in A^\omega$ be an infinite word with values in a finite alphabet A. The *(block) complexity function* $p_w : \mathbb{N} \to \mathbb{N}$ assigns to each n the number of distinct factors of w of length n. A fundamental result due to Hedlund and Morse [8] states that a word w is ultimately periodic if and only if for some n the

[*]Partially supported by the FiDiPro grant "Words, Numbers, Tilings and Applications" from the Academy of Finland
[†]Partially supported by ANR grant *SUBTILE*

complexity $p_w(n) \leq n$. Sequences of complexity $p(n) = n + 1$ are called *Sturmian words.* The most studied Sturmian word is the so-called Fibonacci word

$$01001010010010100101001001010010\ldots$$

fixed by the morphism $0 \mapsto 01$ and $1 \mapsto 0$. In [9] Hedlund and Morse showed that each Sturmian word may be realized geometrically by an irrational rotation on the circle. More precisely, every Sturmian word is obtained by coding the symbolic orbit of a point x on the circle (of circumference one) under a rotation by an irrational angle α where the circle is partitioned into two complementary intervals, one of length α and the other of length $1 - \alpha$. And conversely each such coding gives rise to a Sturmian word. The irrational α is called the *slope* of the Sturmian word. An alternative characterization using continued fractions was given by Rauzy in [11] and [12], and later by Arnoux and Rauzy in [1]. Sturmian words admit various other types of characterizations of geometric and combinatorial nature (see for instance [3]). For example they are characterized by the following balance property: A word w is Sturmian if and only if w is a binary aperiodic (non-ultimately periodic) word with the property that for any two factors u and v of w of equal length, we have $-1 \leq |u|_i - |v|_i \leq 1$ for each letter i. Here $|u|_i$ denotes the number of occurrences of i in u. In this paper, we establish some new characterizations of Sturmian words in terms of the lexicographic order behavior of its factors. We prove:

Theorem 1.1. An infinite word w containing the letters 0 and 1 is Sturmian if and only if for every pair of lexicographically consecutive factors v, v' of the same length, there exist λ, μ such that v, v' either both belong to $\{\lambda 01\mu, \lambda 10\mu\}$ or both belong to $\{\lambda 0, \lambda 1\}$.

Actually our first main result is later formulated in more general terms. The fact that this property holds for Sturmian words has recently been shown in [10], and is a direct consequence of a result proved in [5].

Our second characterization requires the additional hypothesis of recurrence:

Theorem 1.2. Let w be a recurrent aperiodic binary word over the alphabet $\{0, 1\}$ and $v, v' \in \mathrm{Fact}(w)$. Then the following are equivalent:

1. w is Sturmian.
2. For all factors v, v' of w of equal length, if $v <_{\mathrm{lex}} v'$ then $|v|_1 \leq |v'|_1$.
3. For any pair of lexicographically consecutive factors v, v' of the same length, v and v' differ in at most two positions.

2. Preliminaries

In this section, we introduce the tools which will be used in the rest of the paper.

2.1. Standard notions in combinatorics on words

We will report here the standard notations and notions in combinatorics on words that will be used in the rest of the paper. For further results on the subject we refer the reader to [6].

By an *alphabet* we mean a finite non empty set A. The elements of A are called *letters*. We let A^* denote the free monoid over A, i.e. the set of finite sequences of elements of A equipped with the concatenation product. The neutral element of A^* will be called the *empty word* and is denoted ε. The set of nonempty words over A, i.e. the free semigroup over A, is denoted A^+. With the multiplicative notation, given a positive integer n and a word w, we let w^n denote the concatenation of n copies of w. For each word w, we put $w^0 = \varepsilon$.

Two words v, v' are said to be *conjugates* one of the other if there exist λ, μ such that $v = \lambda\mu$ and $v' = \mu\lambda$.

If a nonempty word x is such that $x = x_1 x_2 \cdots x_k$, with $x_i \in A$ for $1 \leq i \leq k$, then k is called the *length* of x and is denoted $|x|$. The length of the empty word is taken to be 0.

We say that a word v is a *factor* of another word w if there exist two words λ, μ such that $w = \lambda v \mu$. If $\lambda = \varepsilon$ (resp. $\mu = \varepsilon$) we call v a *prefix* (resp. a *suffix*) of w. If v is both a prefix and a suffix of w, we say that v is a *border*. A factor v of w is called *proper* if $|v| < |w|$. We denote with $\mathrm{Fact}(w)$ the set of all factors of the word w. A word w is said to be *unbordered* if the only borders of w are w and ε.

Most of the above definitions can be extended to the set A^ω of infinite words on the alphabet A. For $w, w' \in A^\omega$, we say w' is a *tail* of w if $w = vw'$ for some $v \in A^*$. If v is not empty, we call w' a *proper tail* of w.

We call an *occurrence* of v in w a word λ such that λv is a prefix of w. An infinite word w is said to be *recurrent* if each of its factors (or, equivalently, of its prefixes) has infinitely many occurrences in w. Given $v, w \in A^*$ we let $|w|_v$ denote the number of occurrences of v in w and set

$$\mathrm{Alph}(w) = \{x \in A \,|\, |w|_x > 0\}.$$

A factor v of w is *unioccurrent* if $|w|_v = 1$, i.e., if v occurs in w exactly once.

We say that an infinite word w is *periodic* if it can be expressed as an infinite concatenation of a finite word v, i.e. $w = v^\omega$. We say that an infinite word is *ultimately periodic* if it has a periodic tail. Otherwise we say w is *aperiodic*. It is easy to show that any infinite word that contains itself as a proper tail is periodic.

2.2. Lexicographic order

Let A be an alphabet equipped with a total order $<$. Then $<$ extends naturally to a partial order on A^*, denoted $<_{\mathrm{lex}}$, in the following way: We write $v <_{\mathrm{lex}} v'$ (and say v is *lexicographically* smaller than $v',)$ if $|v| = |v'|$ and there exists a word λ and two letters $a < b$ such that λa is a prefix of v and λb is a prefix of v'. Two words v, v' are said to be lexicographically *consecutive* or *adjacent* if $v <_{\mathrm{lex}} v'$ and there is no word w such that $v <_{\mathrm{lex}} w <_{\mathrm{lex}} v'$.

We say a factor v of a word w is *maximal* (resp. *minimal*) in w if there exists no factor v' such that $v <_{\mathrm{lex}} v'$ (resp. $v' <_{\mathrm{lex}} v$), thus omitting the sentence "with respect to the lexicographic order". We will say that v is *extremal* in w if it is either minimal or maximal.

Given two factors v, v' of a word w such that $v <_{\mathrm{lex}} v'$, we will write $v' = \mathrm{succ}_w(v)$ if there is no $f \in \mathrm{Fact}(w)$ such that $v <_{\mathrm{lex}} f <_{\mathrm{lex}} v'$. Notice that if $v \in \mathrm{Fact}(w)$ is non extremal, then there exist $f_1, f_2 \in \mathrm{Fact}(w)$ such that $f_1 = \mathrm{succ}_w(v)$ and $v = \mathrm{succ}_w(f_2)$.

Remark 2.1. It is easy to show that if v is a unioccurrent prefix of an infinite word w and v is extremal in w, then every prefix of w longer than v is unioccurrent, and extremal of the same kind.

We can extend the definition of lexicographic order to infinite words in a natural way, saying that the infinite word w is lexicographically smaller than w' if w has a prefix which is lexicographically smaller than a prefix (of the same length) of w'. The notion of extremality extends as well: we say that an infinite word w is minimal (resp. maximal) if it is lexicographically smaller (resp. larger) than all its tails.

Remark 2.2. It is clear that if aw and w are both extremal infinite binary words (and a is a letter), then they are extremal of the same kind (i.e. they are both minimal or maximal).

2.3. Sturmian words

Let v and v' be factors of w with $|v| = |v'|$. We say the pair (v, v') is *balanced* if $||v|_x - |v'|_x| \leq 1$ for each letter $x \in A$. Otherwise the pair (v, v') is said to be *imbalanced*. A word w is called balanced if all pairs of factors of w of the same length are balanced.

A binary word w is called *Sturmian* if w is aperiodic and balanced. As mentioned earlier, Sturmian words are also defined in terms of the block complexity function $p_w : \mathbb{N} \to \mathbb{N}$ which assigns to each n the number of distinct factors of w of length n: w is Sturmian if and only if $p_w(n) = n + 1$ for each $n \geq 0$.

For each Sturmian word $w \in \{0, 1\}^\omega$ we set

$$\Omega_w = \{w' \in \{0, 1\}^\omega \mid \mathrm{Fact}(w') = \mathrm{Fact}(w)\}.$$

Thus Ω_w is the *shift orbit closure* or *subshift* generated by w. The proof of the following proposition is in [3].

Proposition 2.3. Let w be a Sturmian word over the alphabet $\{0, 1\}$. Then there exists a unique word γ in Ω_w such that both 0γ and 1γ are in Ω_w.

Remark 2.4. The word γ in Proposition 2.3 is called the *characteristic* word of w and it is known that the prefixes of 0γ are lexicographically minimal among the factors of w, while the prefixes of 1γ are maximal.

We say that a factor v of a Sturmian word w is a *Christoffel word* if v is unbordered. We group into the next statement the well-known properties of Christoffel words that we will need in the rest of the paper (see for instance [7, 2], [4, Prop. 5], [5, Prop. 6]).

Proposition 2.5. Let w be a Sturmian word over the alphabet $\{0, 1\}$ and let $v \in \mathrm{Fact}(w)$ be a Christoffel word such that $|v| > 1$. Then there exists u such that:

1. v is either $0u1$ or $1u0$, and they are both Christoffel words in w;
2. $0u1$ and $1u0$ are the only Christoffel words of length $|v|$ in w and are conjugates;
3. all conjugates of v are in $\mathrm{Fact}(w)$;
4. exactly one between $0u0$ and $1u1$ is a factor of w and is extremal in w;
5. the factors of w of length $|v|$ are either conjugates of v or of type xux.

Those factors of a Sturmian word having the same length as a Christoffel word, but not conjugate to a Christoffel word (i.e. the factors xux in the preceding proposition), are called *singular words* of the Sturmian word.

3. Main Result

We begin with the following key proposition:

Proposition 3.1. Let $w \in \{0,1\}^\omega$ be an imbalanced word. Then there exists a factor $u \in \mathrm{Fact}(w)$ of minimal length such that $0u0, 1u1$ are in $\mathrm{Fact}(w)$. Furthermore, either $10u0$ and $01u1$ are both factors of w or there exists a unique letter x such that xux is a prefix of w and occurs in w only finitely many times. In the latter case every prefix of w is extremal in w.

Proof:
Since w is not balanced, there exists an imbalanced pair (v, v') consisting of factors v and v' of w. It is well known (see [3]) that the imbalanced pair of minimal length is of the form $(0u0, 1u1)$ for some factor u of w and is unique. If both $10u0$, $01u1$ are factors of w we are done. So let us assume that there exists a letter $x \in \{0,1\}$ such that no occurrence of xux in w is preceded by $1-x$. Then every internal (non-prefix) occurrence of xux in w is preceded by x. We begin by showing that xux is a prefix of w from which it follows that x is unique. Without loss of generality we can assume that $x = 0$. Suppose that the first occurrence of $0u0$ in w occurs in position $n \geq 0$. If $n = 0$ we are done. So suppose $n > 0$. Then $00u$ is a factor of w occurring in position $n-1$ and the pair $(00u, 1u1)$ is imbalanced. By uniqueness of the shortest imbalanced pair we have that $00u = 0u0$ and hence $0u0$ also occurs in position $n-1$, a contradiction on the minimality of n. This also shows that if $0u0$ occurs in position t then it also occurs in each position r for $0 \leq r \leq t$. Thus $0u0$ occurs only finitely many times in w (for otherwise w would be 0^ω and thus not binary).

We next show that every prefix of w is minimal (if we had taken $x = 1$ then each prefix of w would be maximal). We proceed by contradiction. Let $n > 0$ be the least positive integer for which there exists a factor v' of w in position n which is lexicographically smaller than the corresponding prefix v of w of the same length. Then either there exists a proper prefix u' of u such that $0u'1$ is a prefix of v and $0u'0$ is a prefix of v', or v' begins in $0u0$. In the first case $0u'0$ and the prefix $1u'1$ of $1u1$ constitute a shorter imbalanced pair contradicting the minimality of $|u|$. In the second case v' is an internal occurrence of $0u0$ and is hence preceded by 0. Thus the factor v'' in position $n-1$ of length $|v'|$ is lexicographically smaller than v' and also smaller than v, contradicting the minimality of n. □

The next proposition introduces the main subject of this paper:

Proposition 3.2. Fix $k \geq 1$. Let $A = \{0, 1, \ldots, k\}$ be an ordered alphabet such that $0 < 1 < \cdots < k$ and w an infinite word such that $\mathrm{Alph}(w) = A$. The following are equivalent:

1. For every $v, v' \in \mathrm{Fact}(w)$ with $v' = \mathrm{succ}_w(v)$, there exist distinct letters $a < b$ in A and $\lambda, \mu \in A^*$ such that

$$\begin{cases} v &= \lambda ab\mu \\ v' &= \lambda ba\mu \end{cases} \quad \text{OR} \quad \begin{cases} v &= \lambda a \\ v' &= \lambda b \end{cases}$$

2. For every $v, v' \in \mathrm{Fact}(w)$ with $v' = \mathrm{succ}_w(v)$, there exist $m \in A$ and $\lambda, \mu \in A^*$ such that

$$\begin{cases} v &= \lambda m(m+1)\mu \\ v' &= \lambda(m+1)m\mu \end{cases} \quad \text{OR} \quad \begin{cases} v &= \lambda m \\ v' &= \lambda(m+1) \end{cases}$$

3. $A = \{0, 1\}$ and for every $v, v' \in \mathrm{Fact}(w)$ with $v' = \mathrm{succ}_w(v)$, there exist $\lambda, \mu \in A^*$ such that

$$\begin{cases} v &= \lambda 01\mu \\ v' &= \lambda 10\mu \end{cases} \quad \text{OR} \quad \begin{cases} v &= \lambda 0 \\ v' &= \lambda 1 \end{cases}$$

Proof:
Clearly $(3) \Rightarrow (2) \Rightarrow (1)$. To see that $(1) \Rightarrow (3)$ it suffices to show that (1) implies that $k = 1$. We first note that if $ab \in \mathrm{Fact}(w)$, then $b \in \{a-1, a, a+1\}$. In fact suppose that $a \neq b$. Then either $a < b$ or $b < a$. We consider the first case as the latter case is essentially identical. Let $x, y \in A$ such that ax is the greatest factor of length 2 beginning with a and $(a+1)y$ be the smallest factor of length 2 beginning with $(a+1)$. Clearly $(a+1)y = \mathrm{succ}_w(ax)$, which, from the hypothesis implies that $x = a+1$ and $y = a$. Thus ab is lexicographically smaller or equal to $a(a+1)$ from which it follows that $b = a+1$.

Now suppose to the contrary that $k > 1$, and consider the shortest factor v of w containing both 0 and 2. Then, from what we just proved, $v = 01^n2$ or $v = 21^n0$ for some $n > 0$. We will show that neither occurs in w. Suppose to the contrary that the first is a factor of w and consider the least $n > 0$ for which 01^n2 is a factor of w. Then as 01^n2 is not maximal, its successor is either of the form $101^{n-1}2$ or $01^{n-1}21$ or 01^nx for some $2 < x$. The first two cases contradict the minimality of n while the last case implies that $1x$ is a factor of w for some $2 < x$, again a contradiction. Similarly it is verified that $v = 21^n0$ is never a factor of w. Hence $k = 1$. □

Definition 3.3. We say that an infinite word w has the "Nice Factors Ordering property" (NFOp) if for w one of the equivalent conditions of Proposition 3.2 holds.

Remark 3.4. It is useful to stress that having the NFOp implies that the word w is actually binary. Also it is easy to see that NFOp actually characterizes the pairs of adjacent factors with respect to the lexicographic ordering, i.e., If w satisfies NFOp and v and v' are factors of w with $v = \lambda 01\mu$ and $v' = \lambda 10\mu$ or $v = \lambda 0$ and $v' = \lambda 1$, then $v' = \mathrm{succ}_w(v)$.

Lemma 3.5. If an infinite word w has the NFOp, then w is aperiodic.

Proof:
Let as assume by contradiction that there exist $w', v \in A^*$ with $w = w'v^\omega$. Then w has finitely many tails and it is readily proved that they must respect the NFOp, i.e. if x and y are two lexicographically consecutive tails of w then we can write

$$x = z01z' \quad y = z10z'.$$

In particular, this implies that every tail contains either 01 or 10, hence v cannot be a single letter. As v^ω contains both 01 and 10, v^ω contains tails of the form $(01v')^\omega$ and $\mu = (10v'')^\omega$ for some $v', v''$with $|v'| = |v''| = |v| - 2$. Clearly these two tails differ in an infinite number of positions. On the other hand w has only a finite number of tails and by assumption any two lexicographically consecutive tails differ in exactly two positions. Hence we obtain a contradiction. □

Lemma 3.6. Let w be an infinite word with the NFOp. Then there exists no factor u in $\mathrm{Fact}(w)$ such that $10u0$ and $01u1$ are both factors of w.

Proof:
Suppose to the contrary that there exists a shortest factor u such that both $10u0$ and $01u1$ are factors of w. Since $01u1 <_{\mathrm{lex}} 10u0$, but the two factors cannot be consecutive as they differ in at least three positions, the successor v of $01u1$ satisfies $01u1 <_{\mathrm{lex}} v <_{\mathrm{lex}} 10u0$. It follows that there exists a proper prefix λ of u such that $01\lambda0$ is a prefix of $01u1$ and $01\lambda1$ is a prefix of v (notice that v cannot begin with 1 since otherwise it would be $10u1$ and thus would be lexicographically larger than $10u0$). Since $10\lambda0$ is a prefix of $10u1$ the factors $01\lambda1$ and $10\lambda0$ contradict the minimality of $|u|$. □

The following result is a direct consequence of a result proved by the third author together with Jenkinson in [5] and, more recently, has appeared in [10]; we include it here with a different proof, for the sake of completeness.

Proposition 3.7. Let w be a Sturmian word on the alphabet $\{0,1\}$. Then w satisfies NFOp.

Proof:
Let $0u1$ be a Christoffel factor of w. As $1u0$ is a conjugate of $0u1$ it follows that each factoring $u = xy$ determines two conjugates of $0u1$, namely $v = y01x$ and $v' = y10x$. By Proposition 2.5, v and v' are factors of w; let $z = \mathrm{succ}_w(v)$. As $v <_{\mathrm{lex}} v'$, the longest common prefix of v and z is at least y. In fact it cannot be longer, otherwise we could write $v = y01x'0\lambda$ and $z = y01x'1\mu$ for some words x', λ, μ; as $v' = y10x'0\lambda$, we would have $0x'0, 1x'1 \in \mathrm{Fact}(w)$, a contradiction since w is balanced.

Similarly, y is also the longest common prefix between v' and the word z' such that $\mathrm{succ}_w(z') = v'$. It follows $z = v'$ and $v = z'$, i.e., v and v' are lexicographically consecutive. Thus any two consecutive conjugates of a Christoffel word in w satisfy the first condition in (3) of Proposition 3.2. More generally, if z and z' are lexicographically consecutive factors of w, then there exists a Christoffel factor $0u1$ and two consecutive conjugates v and v' of $0u1$ with z a prefix of v and z' a prefix of v'. The result now follows. □

Before proceeding to prove our main result, we need the following:

Lemma 3.8. Let $w \in \{0,1\}^\omega$ be a Sturmian word and $x \in \{0,1\}$. If xw satisfies NFOp then xw is Sturmian.

Proof:
We proceed by contradiction by supposing that w is Sturmian, xw satisfies NFOp and that xw is not Sturmian. Without loss of generality we can assume that $x = 0$. It follows that there exists u such that both $0u0$ and $1u1$ are factors of $0w$. Since w is Sturmian, it follows from Proposition 3.1 that $0u0$ is a unioccurrent prefix of $0w$ and every prefix of $0w$ is minimal in $0w$. On the other hand, if γ denotes the characteristic word of w (which has u as a prefix), then every prefix of 0γ is minimal in w. By NFOp it follows that for all $n > |u| + 2$ the prefixes of length n of $0w$ and 0γ can be written respectively as $0u01v_n$ and $0u10v_n$ for some word v_n. Hence there exists a tail v of w such that $0w = 0u01v$ and $0\gamma = 0u10v$. Thus $0v, 1v \in \Omega_w$, so that $v = \gamma$ and hence γ is a proper tail of itself, a contradiction. □

Theorem 3.9. Let w be an infinite word on the ordered alphabet $A = \{0, 1, \ldots, k\}$. The following statements are equivalent:

1. w is Sturmian over the alphabet $\{0, 1\}$.
2. w satisfies NFOp.

Proof:
That $(1) \Rightarrow (2)$ follows from Proposition 3.7. To see that $(2) \Rightarrow (1)$, we suppose that $w \in \{0,1\}^\omega$ satisfies NFOp and write $w = aw'$ with $a \in \{0, 1\}$. We need to show that w is Sturmian. By Lemma 3.5 w is aperiodic. If w is not Sturmian, then by Lemma 3.8 we deduce that w' is not Sturmian. Also, combining Proposition 3.1 and Lemma 3.6 we deduce that every prefix of w is extremal and hence w' also satisfies NFOp. In short, if $w = aw'$ satisfies NFOp and is not Sturmian, then every prefix of w is extremal and the tail w' satisfies NFOp and is not Sturmian. Thus writing $w' = bw''$ we deduce that every prefix of w' is extremal and w'' satisfies NFOp and is not Sturmian. Iterating this process indefinitely we deduce that for each tail v of w, each prefix of v is extremal in v. Since w is aperiodic it follows that there exists a tail v of w which begins in 01 and a tail v' of v which begins in 10. Since every prefix of v is minimal in v and every prefix of v' is maximal in v' it follows that 00 is not a factor of v and 11 is not a factor of v'. Hence $v' = (10)^\omega$, a contradiction. □

We next establish another characterization of Sturmian words based on the lexicographic order of their factors.

Theorem 3.10. Let w be a recurrent aperiodic binary word over the alphabet $\{0, 1\}$. Then the following are equivalent:

1. w is Sturmian.
2. For all factors and $v, v' \in \mathrm{Fact}(w)$ of equal length, if $v' = \mathrm{succ}_w(v)$ then v and v' differ in at most two positions.
3. For all factors and $v, v' \in \mathrm{Fact}(w)$ of equal length, if $v <_{\mathrm{lex}} v'$ then $|v|_1 \leq |v'|_1$.

Proof:
$(1) \Rightarrow (2)$: Since w is Sturmian, it has the NFOp by Theorem 3.9. The statement is clearly proven since the NFOp trivially implies condition (2) by definition.

$(2) \Rightarrow (3)$: Notice that condition (2) implies that if $f' = \mathrm{succ}_w(f)$, then there must exist $\lambda, \mu, \mu', x, x'$ with $|x| = |x'| \leq 1$ such that $f = \lambda 0 \mu x \mu'$ and $f' = \lambda 1 \mu x' \mu'$. Hence

$$|f|_1 = |\lambda|_1 + |\mu|_1 + |\mu'|_1 + |x|_1 \leq |\lambda|_1 + |\mu|_1 + |\mu'|_1 + 1 \leq |\lambda|_1 + |\mu|_1 + |\mu'|_1 + |x'|_1 + 1 = |f'|_1.$$

And thus, in particular $|f|_1 \leq |f'|_1$. Suppose $v <_{\mathrm{lex}} v'$. Then there must exist $v_0, \ldots, v_k$ such that $v = v_0$, $v' = v_k$ and for all $1 \leq n \leq k$, $v_n = \mathrm{succ}_w(v_{n-1})$, then

$$|v|_1 = |v_0|_1 \leq \cdots \leq |v_k|_1 = |v'|_1.$$

$(3) \Rightarrow (1)$: Assume w is not Sturmian; as w is aperiodic, it has to be imbalanced. Since w is recurrent, we have from Proposition 3.1 that there must exist u such that both $10u0$ and $01u1$ are factors of w. But clearly this is a contradiction, since $01u1 <_{\mathrm{lex}} 10u0$ and $|01u1|_1 = |10u0|_1 + 1$. This concludes the proof. □

Notice that, as opposed to Theorem 3.9, the above result actually needs the recurrence and aperiodicity hypotheses, as for example:

- the recurrent periodic word $(01)^\omega$ and the non-recurrent aperiodic word $00f$ (where f is the Fibonacci word) both respect condition (3), although neither is Sturmian,
- the non-recurrent ultimately periodic word 01^ω satisfies both (2) and (3), but it is not Sturmian.

References

[1] Arnoux, P., Rauzy, G.: Représentation géométrique de suites de complexité $2n+1$, *Bull. Soc. Math. France*, **119**, 1991, 199–215.

[2] Berstel, J., Lauve, A., Reutenauer, C., Saliola, F.: *Combinatorics on Words: Christoffel Words and Repetition in Words*, vol. 27 of *CRM monograph series*, American Mathematical Society, 2008.

[3] Berstel, J., Séébold, P.: Sturmian Words, in: *Algebraic Combinatorics on Words* (M. Lothaire, Ed.), Cambridge University Press, Cambridge UK, 2002, Chapter 2.

[4] Cao, W.-T., Wen, Z.-Y.: Some properties of the factors of Sturmian sequences, *Theor. Comput. Sci.*, **304**, 2003, 365–385.

[5] Jenkinson, O., Zamboni, L. Q.: Characterisations of balanced words via orderings, *Theor. Comput. Sci.*, **310**, 2004, 247–271.

[6] Lothaire, M.: *Combinatorics on Words*, Addison-Wesley, Reading MA, 1983, Reprinted by Cambridge University Press, Cambridge UK, 1997.

[7] Mignosi, F., Zamboni, L. Q.: A note on a conjecture of Duval and Sturmian words, *Theoret. Inform. Appl.*, **36**, 2002, 1–3.

[8] Morse, M., Hedlund, G. A.: Symbolic Dynamics, *Amer. J. Math.*, **60**, 1938, 815–866.

[9] Morse, M., Hedlund, G. A.: Symbolic Dynamics II. Sturmian Trajectories, *Amer. J. Math.*, **62**, 1940, 1–42.

[10] Perrin, D., Restivo, A.: A note on Sturmian words, *Theor. Comput. Sci.*, (in press), 2012.

[11] Rauzy, G.: Une généralisation du développement en fraction continue, in: *Séminaire Delange-Pisot-Poitou, 18e année: 1976/77, Théorie des nombres, Fasc. 1*, Secrétariat Math., Paris, 1977, Exp. No. 15, 16.

[12] Rauzy, G.: Échanges d'intervalles et transformations induites, *Acta Arith.*, **34**, 1979, 315–328.