On Factorially Balanced Sets of Words

Gwenaël Richomme, Patrice Séébold

On factorially balanced sets of words

Gwénaël Richomme, Patrice Séébold

LIRMM (CNRS, Univ. Montpellier 2)
UMR 5506 - CC 477,
161 rue Ada,
34095 Montpellier Cedex 5, France

and

Université Paul-Valéry Montpellier 3,
UFR IV, Dpt MIAp,
Route de Mende,
34199 Montpellier Cedex 5, France

gwenael.richomme@lirmm.fr, patrice.seebold@lirmm.fr

August 27, 2010

Abstract

A set of words is factorially balanced if the set of all the factors of its words is balanced. We prove that if all words of a factorially balanced set have a finite index then this set is a subset of the set of factors of one Sturmian word. Moreover, characterizing the set of factors of a given length n of a Sturmian word by the left special factor of length $n - 1$ of this Sturmian word, we provide an enumeration formula for the number of sets of words that correspond to some set of factors of length n of a Sturmian word.

1   Introduction

Since the works of Morse and Hedlund [11], due to their numerous properties and applications, balanced words were given a lot of attention, especially the case of binary aperiodic balanced words, called *Sturmian words* (see for instance surveys and studies in [1, 4, 6, 9, 12, 15]). In this note, balanced sets of binary words rather than balanced words themselves are considered. In particular, we deal with sets of words such that the set of all the factors of all the words is balanced. Such sets, that we call *factorially balanced*, are introduced in Section 2. In Section 3, we prove that a finite set of binary words is a subset of the set of factors of a Sturmian word if and only if it is factorially balanced. The "only if" part of this result is very well known as a fundamental property of Sturmian words. As far as we know, the "if" part has never been stated except in the case of sets of cardinality one (a finite word is balanced if and only if it is the factor of a Sturmian word – see for instance [8, 14]).

In Section 4, we focus on *uniform* sets of words, that is on finite sets of words whose elements have all the same length. We provide an enumeration formula of uniform factorially balanced sets of binary words. For this we first prove that the set of factors of a given length n of a Sturmian word is characterized by the left special factor of length $n - 1$ of this Sturmian word.

In Section 5, the infinite case is considered and a characterization of factorially balanced sets of words is given using the additional notion of finite index. Results of Sections 3 and 5 are unified in the conclusion.

2   Factorially balanced sets of words

We assume that readers are familiar with combinatorics on words, quoting that for basic (possibly omitted) definitions we follow [9].

In this paper, we are interested on sets of finite words over the binary alphabet $A = \{a, b\}$, and, for any word u over A and letter α, $|u|$ and $|u|_\alpha$ denote respectively the length of u and the number of occurrences of α in u. Let us recall that a set of words X is *balanced* if for all words u and v over A, $|u| = |v|$ implies that $||u|_a - |v|_a| \leq 1$. A finite or infinite word is *balanced* if its set of factors is balanced, and it is well known that Sturmian words correspond to infinite aperiodic balanced words (see [9, Theorem 2.1.5]). A classical alternative definition is that Sturmian words are infinite words having exactly $n + 1$ factors of length n for each integer $n \geq 0$.

Note that in the case of Sturmian words, the sets usually considered are *factorial* sets, i.e., sets which contain all the factors of all their words. However, in the general case the notion of balanced set of words given above is not well adapted because the sets are not necessarily factorial. For instance any infinite set of words of different lengths, or, any uniform finite sets of words with all words containing the same number of occurrences of a given letter are balanced.

So we consider here a restriction of the notion of balanced sets of words. A set of words X is *factorially balanced* if its set of factors is balanced.

The following remark will be used several times (explicitely or not) in the rest of this paper.

Remark 2.1. *Every set of factors of a factorially balanced set of words is also a factorially balanced set of words.*

Of course a balanced factorial set of words is factorially balanced, thus all the results on Sturmian words applied with this notion. In particular we can reformulate the following useful Propositions 2.1.2 and 2.1.3 of [9] in term of factorially balanced set of words.

Proposition 2.2. [9, Prop. 2.1.2] *For any factorially balanced set X of words over A and for any integer $n \geq 0$, $\mathrm{Card}(X \cap A^n) \leq n + 1$.*

Proposition 2.3. [9, Prop. 2.1.3] *For any set X of words over A, the set X is not factorially balanced if and only if there exists a palindrome word w such that awa and bwb are factors of X.*

For any word (finite or infinite) w, $\mathrm{F}(w)$ denotes the set of all the finite factors of w. This notion extends to sets of words: if X is a set of words, $\mathrm{F}(X)$ denotes the set of all the finite factors of words of X.

Let us recall that a finite word u is a *left special factor* of a (finite or infinite) word w over A if both au and bu are factors of w. A remarkable property of Sturmian words is that every Sturmian word $\mathbf{s}$ contains exactly one left special factor of each length (see for instance [5, Prop. 3.1]). The next property is rather straightforward and it is perhaps already known (it can be seen as a slight improvement of [9, Lemma 2.2.35]), but to the best of our knowledge it is new.

Lemma 2.4. *Let $u \in A^*$ be a word such that there exists a Sturmian word $\mathbf{s}$ with $\{aub, bua\} \subset \mathrm{F}(\mathbf{s})$. Then one (and only one) of the two words aua, bub is a factor of $\mathbf{s}$.*

Proof. Since $\mathbf{s}$ is Sturmian, it is balanced thus aua and bub cannot be both factors of $\mathbf{s}$.

The word u is the left special factor of $\mathbf{s}$ of length $|u|$, thus the left special factor of $\mathbf{s}$ of length $|u| + 1$, say v, must have u as a prefix, so $v = ua$ or $v = ub$. If $v = ua$ then $av = aua$ is a factor of $\mathbf{s}$, otherwise $v = ub$ and bub is a factor of $\mathbf{s}$. $\qquad\square$

3 Characterization of factorially balanced finite sets of words

In this section we characterize finite sets of words that are subsets of the set of factors of a Sturmian word: they are factorially balanced sets of words. The infinite case will be studied in Section 5.

Theorem 3.1. *Let S be a finite set of binary finite words. There exists a Sturmian word* **s** *such that $S \subset \mathrm{F}(\mathbf{s})$ if and only if S is factorially balanced.*

Before proving this result, let us recall a few examples of Sturmian words. The most famous one is the *Fibonacci word*, denoted **F**, which is the fixed point of the morphism φ defined by $\varphi(a) = ab$ and $\varphi(b) = a$. Two other morphisms play an important role in the theory of Sturmian words, namely E and $\tilde{\varphi}$ defined by $E(a) = b$, $E(b) = a$, $\tilde{\varphi}(a) = ba$, $\tilde{\varphi}(b) = a$. Indeed, a morphism preserves Sturmian words (the image of any Sturmian word by such a morphism is still Sturmian) if and only if the morphism is obtained by composition of the morphisms φ, E, $\tilde{\varphi}$ (this was originally proved in [10], see also[9, Th. 2.3.7]). Hence for any integer $k \geq 0$, words $(\varphi \circ E)^k(\mathbf{F})$ and $\varphi \circ (\varphi \circ E)^k(\mathbf{F})$ are examples of Sturmian words, that contain respectively as a factor words a^k and $(ab)^k$.

Proof of Theorem 3.1. As mentioned in the introduction, the "only if" part of this theorem is well known since Morse and Hedlund's works [11], hence we only prove the "if" part.

Let S be a factorially balanced finite set of words. We prove by induction on $||S|| := \sum_{w \in S} |w|$ that there exists a Sturmian word **s** such that $S \subset \mathrm{F}(\mathbf{s})$. As mentioned in the introduction, this result is already known when $\mathrm{Card}(S) = 1$, therefore, we assume that $\mathrm{Card}(S) \geq 2$.

At least one of the two words aa and bb does not belong to $\mathrm{F}(S)$ because S is factorially balanced. When it is the case for both words aa and bb, there exists an integer k such that $S \subset \mathrm{F}((ab)^k)$, and the theorem is verified with $\mathbf{s} = \varphi \circ (\varphi \circ E)^k(\mathbf{F})$. Now, assume the result holds when $aa \in \mathrm{F}(S)$ and consider a factorially balanced finite set S' with $bb \in \mathrm{F}(S')$. Then $S := E(S')$ is factorially balanced with $aa \in \mathrm{F}(S)$, so that, by our previous assumption, there exists a Sturmian word **s** such that $S \subset \mathrm{F}(\mathbf{s})$. Consequently $S' = E(S) \subset E(\mathbf{s})$ which is a Sturmian word since E preserves Sturmian words: the theorem is thus verified for S'. Consequently, from now on we assume that $aa \in \mathrm{F}(S)$.

Let us denote S_2 the set $(S \cap aA^*) \cup a(S \cap bA^*)$ that is the set obtained from S keeping all words beginning with the letter a and adding an a in front of each word of S beginning with the letter b. Of course $S \subseteq \mathrm{F}(S_2)$.

Fact 3.2. *The set S_2 is factorially balanced.*

Proof. Assume by contradiction that S_2 is not factorially balanced. By Proposition 2.3, there exists a word x such that both words axa and bxb are factors of S_2. As S is factorially balanced, at least one of these two words is not a factor of S. By construction of S_2, we get $bxb \in \mathrm{F}(S)$, and so $axa \notin \mathrm{F}(S)$. Still by construction of S_2, this means that x is a non-empty word beginning with the letter b. This implies that both words aa and bb are factors of $\mathrm{F}(S)$ contradicting the fact that it is factorially balanced. $\square$

Now observe that all words in S_2 begin with the letter a and do not contain the factor bb, so that any word u in S_2 can be uniquely written either $u = \varphi(v)a$ when u ends with a, or $u = \varphi(v)$ (with v ending with a) when u ends with b. Therefore, there exist two (unique) sets of words S_3 and S_4 such that words of S_3 end with the letter a, and $S_2 = \varphi(S_3) \cup \varphi(S_4)a$.

Fact 3.3. *The set $S_3 \cup S_4$ is factorially balanced.*

Proof. Assume by contradiction that $S_3 \cup S_4$ is not factorially balanced. By Proposition 2.3, there exists a word x such that both words axa and bxb are factors of $S_3 \cup S_4$. Then words $ab\varphi(x)ab$ and $a\varphi(x)a$ are factors of S_2. More precisely since bxb ends with the letter b, either bxb is a factor of a word of S_3 without being a suffix of this word or bxb is a factor of a word in S_4. Observing that both $\varphi(a)$ and $\varphi(b)$ begin with the letter a, $a\varphi(x)aa$ is thus a factor of S_2, which contradicts Fact 3.2. □

Now, continuing the proof of Theorem 3.1, observe that if $a \in S$, since $aa \in F(S)$, we have $F(S \setminus \{a\}) = F(S)$ thus $F(S \setminus \{a\})$ is factorially balanced. In this case by inductive hypothesis there exists a Sturmian word $\mathbf{s}$ with $S \setminus \{a\} \subset F(\mathbf{s})$: since a is factor of all Sturmian words, we also have $S \subset F(\mathbf{s})$ and the theorem holds for set S. Therefore, from now on we assume that $a \notin S$.

Assume that $b \in S$. If there exists another word in S containing the letter b then, as previously seen, we can find a Sturmian word $\mathbf{s}$ such that $S \subset F(\mathbf{s})$, which shows that the theorem holds for set S. Otherwise there exists an integer k such that $F(S) \subseteq \{a^k, b\}$, and consequently $S \subset F((\varphi \circ E)^k(\mathbf{F}))$. From now on we also assume that $b \notin S$.

As neither word a nor word b belongs to S, we can observe that to each word u in S corresponds one and exactly one word v in $S_3 \cup S_4$ such that, for some word w, one of the following four cases holds:

- $u = awa$ and $u = \varphi(v)a$ (when $u \in S \cap aA^*$ and $v \in S_4$);

- $u = awb$ and $u = \varphi(v)$ (when $u \in S \cap aA^*$ and $v \in S_3$);

- $u = bwa$ and $au = \varphi(v)a$ (when $u \in S \cap bA^*$ and $v \in S_4$);

- $u = bwb$ and $au = \varphi(v)$ (when $u \in S \cap bA^*$ and $v \in S_3$).

In all cases $|v| < |u|$, which implies that $||S_3 \cup S_4|| < ||S||$. By Fact 3.3 and inductive hypothesis, there exists a Sturmian word $\mathbf{s}'$ such that $S_3 \cup S_4 \subset F(\mathbf{s}')$. Thus Theorem 3.1 holds since $\mathbf{s} := \varphi(\mathbf{s}')$ is a Sturmian word (let us recall that φ preserves Sturmian words) and $S \subseteq F(S_2) \subset F(\mathbf{s})$ (observe that $F(\varphi(S_4)a) \subset F(\mathbf{s})$ follows from the fact that both $\varphi(a)$ and $\varphi(b)$ begin with the letter a). □

This result admits the following interesting corollary.

Proposition 3.4. *Let $u \in A^*$ be any word and let $X \subset A^*$ be a set of words of length at most $|u| + 1$. Then the following conditions are equivalent.*

a) There exists a Sturmian word $\mathbf{s}$ such that $(X \cup \{aub, bua\}) \subset F(\mathbf{s})$.

b) There exists a Sturmian word $\mathbf{s}'$ such that $(X \cup \{aub, bua, aua\}) \subset F(\mathbf{s}')$.

c) There exists a Sturmian word $\mathbf{s}''$ such that $(X \cup \{aub, bua, bub\}) \subset F(\mathbf{s}'')$.

Actually, Lemma 2.4 shows that one of the two words $\mathbf{s}'$ and $\mathbf{s}''$ could be taken equal to $\mathbf{s}$.

Proof. It is straightforward that *b)* and *c)* both imply *a)*. Assume that *a)* holds. By Theorem 3.1, $(X \cup \{aub, bua\})$ is a factorially balanced set of words. By Remark 2.1, the set $X \cup \{au, ua, bu, ub\}$ is factorially balanced. Let $\alpha \in \{a, b\}$. Observe that the set $Z_\alpha := \{aub, bua, \alpha u\alpha\}$ is a balanced set. Set $X_\alpha := X \cup Z_\alpha$. We have $F(X_\alpha) = F(Y) \cup Z_\alpha$, and this union is disjointed since elements of $F(Y)$ are of length at most $|u| + 1$, and those of Z_α of length $|u| + 2$. Since $F(Y)$ and Z_α are balanced, $F(X_\alpha)$ is also balanced, or equivalently X_α is factorially balanced. By Theorem 3.1, both *b)* and *c)* hold. □

4 Enumerating uniform factorially balanced sets of binary words

In this section we consider, for an integer $n \geq 0$, n-uniform sets of words, that is, sets that contain only words of the same length n. Such a set is necessarily finite. More precisely we are interested in sets of words that could correspond exactly to the set of words of length n of some Sturmian word. We set $\mathcal{S}_{sturm}(n) := \{F(\mathbf{s}) \cap A^n \mid \mathbf{s} \text{ Sturmian}\}$, that is, $\mathcal{S}_{sturm}(n)$ is the set of all the sets of factors of length n of each Sturmian word.

From Theorem 3.1, we know that any factorially balanced n-uniform set S is a set of factors of a Sturmian word. This does not mean that every factorially balanced n-uniform set of words belongs to some $\mathcal{S}_{sturm}(n)$, but each factorially balanced n-uniform set of words is a subset of some $F \in \mathcal{S}_{sturm}(n)$. Since any Sturmian word has exactly $n+1$ factors for each length n, we have the following characterization.

Fact 4.1. *For any integer $n \geq 0$, a set S of words belongs to $\mathcal{S}_{sturm}(n)$ if and only if S is a factorially balanced set of words of length n with cardinality $n+1$.*

The rest of this section is devoted to the proof of the next result, which is an answer to a question originally asked by Christophe Reutenauer [3]. Let ϕ denote the *Euler's totient function* ϕ that associates to each integer $n \geq 1$ the number of positive integers less than or equal to n and relatively prime to n.

Theorem 4.2. [1] *For every integer $n \geq 1$, $\mathrm{Card}(\mathcal{S}_{sturm}(n)) = \sum_{i=1}^{n} \phi(i)$.*

For any integer $n \geq 0$, set $\mathcal{S}_{lspec}(n) := \{u \in A^n \mid au \text{ and } bu \text{ balanced }\}$. This set denotes all potential left special factors of balanced words. To use this with Theorem 3.1 we would need the set $\{au, bu\}$ to be factorially balanced. Actually the two conditions are equivalent as proved below.

Lemma 4.3. *For any word u over A, the words au and bu are balanced if and only if the set $\{au, bu\}$ is factorially balanced.*

Proof. The "if" part corresponds to the definition. Assume that the set $\{au, bu\}$ is not factorially balanced. By Proposition 2.3, there exists a word x such that words axa and bxb are factors of $\{au, bu\}$. As it is not possible that these two words simultaneously occur as prefixes of words au and bu, at least one of axa and bxb is a factor of u. This implies that both words axa and bxb are factors of au or of bu, showing that one of this word is not balanced. $\square$

We are now ready to state our cornerstone for the proof of Theorem 4.2.

Proposition 4.4. *For every integer $n \geq 1$, there exists a bijection from the set $\mathcal{S}_{sturm}(n)$ into the set $\mathcal{S}_{lspec}(n-1)$.*

In other terms, the set of factors of a given length n of a Sturmian word is characterized by only one word, its left special factor of length $n-1$. It is known that such a special factor exists and is unique (see for instance [5, Prop. 3.1]).

Before proving Proposition 4.4, we start by the following observation.

Fact 4.5. *For $S \in \mathcal{S}_{sturm}(n)$, there exists a unique word u in $\mathcal{S}_{lspec}(n-1)$ such that $\{au, bu\} \subseteq S$.*

Proof. Indeed by choice of S, there exists a Sturmian word $\mathbf{s}$ such that $F(\mathbf{s}) \cap A^n = S$. This Sturmian word has a unique left special factor u of length $n-1$. In particular u is the unique word such that $\{au, bu\} \subseteq S$, which implies that both words au and bu are balanced (see Lemma 4.3). $\square$

From now on we denote by LS_n the function from $\mathcal{S}_{sturm}(n)$ to $\mathcal{S}_{lspec}(n-1)$ that associates to each set S in $\mathcal{S}_{sturm}(n)$ the word u considered in the previous fact.

Fact 4.6. *For all integers $n \geq 1$, the function LS_n is injective.*

Proof. Let $S_1, S_2 \in \mathcal{S}_{sturm}(n)$ such that $LS_n(S_1) = LS_n(S_2)$: we denote by u this left special word of length $n-1$, and for all integers i between 1 and n, we denote by u_i the prefix of u of length $i - 1$. Since sets S_1 and S_2 are factorially balanced and since $\{au, bu\} \subseteq S_1 \cap S_2$, for all integers i between 1 and n, and for all words $v \in F(S_1 \cup S_2) \cap A^i$, we have $||v|_a - |au_i|_a| \leq 1$ and $||v|_a - |bu_i|_a| \leq 1$, which implies $|v|_a \in \{|u_i|_a, |u_i|_a + 1\}$. Consequently the set $F(S_1 \cup S_2)$ is balanced. By Proposition 2.2, $\mathrm{Card}(S_1 \cup S_2) = \mathrm{Card}(F(S_1 \cup S_2) \cap A^n) \leq n + 1$. As by Fact 4.1 $\mathrm{Card}(S_1) = \mathrm{Card}(S_2) = n + 1$, we deduce that $S_1 = S_2$. $\qquad\square$

Fact 4.7. *For all integers $n \geq 1$, the function LS_n is surjective.*

Proof. This is a direct consequence of a result proved by de Luca [6] (see Corollary 1). However, in order to be self-contained we give here a direct proof of this result.

For any word u such that au and bu are balanced, from Lemma 4.3 the set $\{au, bu\}$ is factorially balanced thus, by Theorem 3.1 there exists a Sturmian word $\mathbf{s}$ such that $\{au, bu\} \subset F(\mathbf{s})$. $\qquad\square$

Proof of Proposition 4.4. Facts 4.6 and 4.7 prove that the function LS_n is bijective. $\qquad\square$

Corollary 4.8. *For all integers $n \geq 1$, $\mathrm{Card}(\mathcal{S}_{sturm}(n)) = \mathrm{Card}(\mathcal{S}_{lspec}(n-1))$.*

Proof of Theorem 4.2. Actually Theorem 4.2 is a direct corollary of Corollary 4.8, and of a result by de Luca and Mignosi [7] stating that the number of words w of length n such that aw and bw are balanced is $\sum_{i=1}^{n+1} \phi(i)$. $\qquad\square$

5 The infinite case

In previous sections we have considered finite sets that can be sets of factors of some Sturmian words. A question remains after Section 3: can we characterize infinite sets of words that are sets of factors of some Sturmian words? The following examples show that the answer cannot be a simple extension of Theorem 3.1.

Example 5.1. *The set $\{a^n b a^n \mid n \geq 0\}$ is factorially balanced, but there exists no (right) infinite balanced word having all words of this set as factors. However one can observe that this set is a subset of factors of the balanced biinfinite word $^\omega aba^\omega$.*

Remember that Sturmian words are words with $n + 1$ factors of length n for every integer n. Therefore, one may consider adding a condition on the number of factors of each length in the set.

Example 5.2. *The set $\{a^n, a^k b a^{n-k} \mid n \geq 0, 0 \leq k \leq n-1\}$ is a factorially balanced set (the set is factorial) containing exactly $n + 1$ factors of length n for all integers $n \geq 1$, but there exists no balanced infinite word (neither biinfinite word) that contains this set as a subset of factors.*

The previous example is not satisfactory since factors containing the letter b occur only at most once in each factor of the considered set. Let us recall that an infinite word $\mathbf{w}$ is *uniformly recurrent* if for every integer $n \geq 0$ there exists an integer N such that each factor of $\mathbf{w}$ of length at least N contains all factors of length n of $\mathbf{w}$. Sturmian words are uniformly recurrent (see [9]). We extend the definition of uniform recurrence to infinite sets of finite words: an infinite

set of finite words S is *uniformly recurrent* if for every integer $n \geq 0$ there exists an integer N such that each factor of S of length at least N contains all factors of length n of S. Next lemma provides a characterization of infinite sets of words that are subsets of the sets of factors of a Sturmian word.

Proposition 5.3. *Let S be an infinite set of binary finite words. There exists a Sturmian word $\mathbf{s}$ such that $S \subseteq \mathrm{F}(\mathbf{s})$ if and only if S is uniformly recurrent and, for all $n \geq 0$, $\mathrm{Card}(\mathrm{F}(S) \cap A^n) = n+1$.*

Proof. We have already mentioned that the "only if" part holds. Now if S is uniformly recurrent, one can construct an infinite sequence of words $(u_n)_{n \geq 0}$ such that u_0 is a letter, and for all $n \geq 1$, u_n is a prefix of u_{n+1} and u_n contains all factors of S of length at most n (this is possible thanks to uniform recurrence). This sequence of words tends to a unique infinite word $\mathbf{w}$ having all words u_n has prefixes. Consequently, any factor of $\mathbf{w}$ is a factor of one word of the sequence and so is a factor of S: $\mathrm{F}(S) = \mathrm{F}(\mathbf{w})$. The word $\mathbf{w}$ is Sturmian since it has exactly $n+1$ factors of length n for all $n \geq 0$. $\qquad \square$

Examples at the beginning of the section show that the hypothesis "S factorially balanced" cannot substitute any of the two conditions in the previous proposition. Of course a direct consequence of the previous lemma is that, for any infinite set S of binary finite words, there exists a Sturmian word $\mathbf{s}$ such that $S \subseteq \mathrm{F}(\mathbf{s})$ if and only if S is factorially balanced and uniformly recurrent, and, for all $n \geq 0$, $\mathrm{Card}(\mathrm{F}(S) \cap A^n) = n+1$, but the hypothesis "$S$ factorially balanced" is purely artificial here. In order to get a stronger characterization of those infinite factorially balanced sets of words that are subsets of factors of a Sturmian word, let us recall that a factor u of an infinite word $\mathbf{w}$ *has a finite index* in $\mathbf{w}$ if there exists an integer k such that u^k is not a factor of $\mathbf{w}$. This definition naturally extends to indexes of factors of an infinite set of words. A basic result on indexes of Sturmian word is:

Proposition 5.4. (See [2]) *Every nonempty factor of a Sturmian word has a finite index.*

We are ready to state our characterization on infinite sets.

Theorem 5.5. *Let S be an infinite set of binary finite words. There exists a Sturmian word $\mathbf{s}$ such that $S \subseteq \mathrm{F}(\mathbf{s})$ if and only if S is factorially balanced and each nonempty word of $\mathrm{F}(S)$ has a finite index in $\mathrm{F}(S)$.*

Let us denote $\mathrm{Pref}(S)$ the set of prefixes of words in a set S. Next lemma will be useful.

Lemma 5.6. *(König's lemma – see for instance [9, Prop. 1.2.3]) If S is an infinite set of words, there exists an infinite word $\mathbf{w}$ having all its prefixes in $\mathrm{Pref}(S)$.*

Proof of Theorem 5.5. We have already mentioned that the set $\mathrm{F}(\mathbf{s})$ of factors of a Sturmian word is factorially balanced, and so is any of its subsets (see Remark 2.1). Since each factor of $\mathbf{s}$ has finite index in $\mathbf{s}$, if $S \subseteq \mathrm{F}(\mathbf{s})$ then any factor of S is also of finite index in $\mathrm{F}(S)$.

Assume from now on that S is an infinite factorially balanced set of finite words, all of its factors having a finite index in $\mathrm{F}(S)$. Using König's lemma, we get an infinite word $\mathbf{s}$ such that all its prefixes are prefixes of words of S. As any factor of $\mathbf{s}$ is a factor of a prefix of $\mathbf{s}$ and so a factor of $\mathbf{s}$, $\mathbf{s}$ is balanced. If it is not aperiodic, then there exist words x (possibly empty) and y (not empty) such that $\mathbf{s} = xy^\omega$, and consequently we have a contradiction with the fact that y should be of finite index in S. Thus $\mathbf{s}$ is an aperiodic balanced word, that is a Sturmian word. Moreover $\mathrm{F}(\mathbf{s}) \subseteq \mathrm{F}(S)$. As by Proposition 2.2, for all integers $n \geq 0$, $\mathrm{Card}(\mathrm{F}(S) \cap A^n) \leq n+1 = \mathrm{Card}(\mathrm{F}(\mathbf{s}) \cap A^n)$, we get $\mathrm{F}(\mathbf{s}) = \mathrm{F}(S)$ and so $S \subseteq \mathrm{F}(\mathbf{s})$. $\qquad \square$

6 Conclusion

In our study of factorially balanced sets, we have considered separately the finite case (Section 3) and the infinite case (Section 5) because, in the infinite case, there exist factorially balanced sets of words which are not subsets of the set of factors of some sturmian word.

However, in every finite set of finite words each factor is trivially of finite index. Thus we can add, in the statement of Theorem 3.1, the (unnecessary) condition of finite index in order to unify Theorems 3.1 and 5.5 in the following general result claimed in the abstract.

Theorem 6.1. *Let S be a set of binary finite words. There exists a Sturmian word $\mathbf{s}$ such that $S \subseteq F(\mathbf{s})$ if and only if S is factorially balanced and each nonempty factor of S has a finite index in S.*

References

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

[2] J. Berstel. On the Index of Sturmian Words. In *Jewels are forever*, pages 287–294. Springer, Berlin, 1999.

[3] J. Berstel. Private communication, 2010.

[4] J. Berstel, A. Lauve, C. Reutenauer, and F. Saliola. *Combinatorics on Words: Christoffel Words and Repetitions in Words*, volume 27 of *CRM Monograph Series.* American Mathematical Society, 2008.

[5] J. Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc.*, 4:67–88, 1997.

[6] A. de Luca. Sturmian words: structure, combinatorics, and their arithmetics. *Theoret. Comput. Sci.*, 183:45–82, 1997.

[7] A. de Luca and F. Mignosi. On some combinatorial properties of Sturmian words. *Theoret. Comput. Sci.*, 136:361–385, 1994.

[8] S. Dulucq and D. Gouyou-Beauchamps. Sur les facteurs des suites de Sturm. *Theoret. Comput. Sci.*, 71:381–400, 1990.

[9] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, 2002.

[10] F. Mignosi and P. Séébold. Morphismes sturmiens et règles de Rauzy. *J. Théor. Nombres Bordeaux*, 5:221–233, 1993.

[11] M. Morse and G. A. Hedlund. Symbolic dynamics II: Sturmian sequences. *Amer. J. Math.*, 61:1–42, 1940.

[12] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics.* Springer, 2002. (V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel, editors).

[13] C. Reutenauer. Private communication, 2010.

[14] G. Richomme. Test-words for Sturmian morphisms. *Bull. Belg. Math. Soc.*, 6:481–489, 1999.

[15] L. Vuillon. Balanced words. *Bull. Belg. Math. Soc.*, 10(5):787–805, 2003.