Theoretical Computer Science 412 (2011) 5492–5497

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Note

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

ARTICLE INFO

Article history:
Received 2 August 2010
Received in revised form 9 June 2011
Accepted 22 June 2011
Communicated by D. Perrin

Keywords:
Balanced words
Sturmian words

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 a 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 [12], 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 [1,5,7,10,13,16]). 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 [9,15]).

In Section 4, we focus on *uniform* sets of words, that is, on finite sets of words all of whose elements have 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 the readers are familiar with combinatorics on words, quoting that for basic (possibly omitted) definitions we follow [10].

In this paper, we are interested in 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

* Corresponding author at: Université Paul-Valéry Montpellier 3, UFR IV, Dpt MIAp, Route de Mende, 34199 Montpellier Cedex 5, France. Tel.: +33 0 467142373.

E-mail addresses: gwenael.richomme@lirmm.fr, gwenael.richomme@univ-montp3.fr (G. Richomme), patrice.seebold@lirmm.fr, patrice.seebold@univ-montp3.fr (P. Séébold).

0304-3975/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2011.06.027

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 [10, 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, here we consider 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 (explicitly 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 [10] in terms of factorially balanced set of words.

Proposition 2.2. *[10, 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. *[10, 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 [6, Prop. 3.1]). The following, rather straightforward property was used informally by Arnoux and Rauzy in [2] to explain the evolution of word graphs in the Sturmian case. We prove it in order to self-contain our paper.

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 both cannot be 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 $\mathbf{s}$. $\square$

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 $\mathbf{s}$ 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 by $\mathbf{F}$, 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 [11], see also[10, 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 factors a^k and $(ab)^k$ respectively.

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 [12], 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 $\mathbf{s}$ 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 the 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 that 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 $\mathbf{s}$ 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 by 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 $S \subseteq \mathrm{F}(S_2)$.