Theoretical Computer Science 581 (2015) 26–44

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Sturmian words and the Stern sequence

Aldo de Luca [a,1], Alessandro De Luca [b,*,2]

[a] Dipartimento di Matematica e Applicazioni, Università degli Studi di Napoli Federico II, via Cintia, Monte S. Angelo, I-80126 Napoli, Italy
[b] DIETI, Università degli Studi di Napoli Federico II, via Claudio 21, 80125 Napoli, Italy

ARTICLE INFO

Article history:
Received 20 October 2014
Received in revised form 11 February 2015
Accepted 25 February 2015
Available online 3 March 2015
Communicated by D. Perrin

Keywords:
Sturmian word
Central word
Standard word
Christoffel word
Stern sequence
Raney tree

ABSTRACT

Central, standard, and Christoffel words are three strongly interrelated classes of binary finite words which represent a finite counterpart to characteristic Sturmian words. A natural arithmetization of the theory is obtained by representing central and Christoffel words by irreducible fractions labeling, respectively, two binary trees, the Raney (or Calkin–Wilf) tree and the Stern–Brocot tree. The sequence of denominators of the fractions in the Raney tree is the famous Stern diatomic numerical sequence. An interpretation of the terms $s(n)$ of Stern's sequence as lengths of Christoffel words when n is odd, and as minimal periods of central words when n is even, allows one to interpret several results on Christoffel and central words in terms of Stern's sequence and, conversely, to obtain new insight into the combinatorics of Christoffel and central words. One of our main results is a non-commutative version of the "alternating bit sets theorem" by Calkin and Wilf. We also study the length distribution of Christoffel words corresponding to nodes of equal height in the tree, obtaining some interesting bounds and inequalities.

1. Introduction

Sturmian words are of great interest in combinatorics of infinite words because they are the simplest words that are not ultimately periodic. Starting with the seminal paper of 1940 by Marston Morse and Gustav A. Hedlund [31], there has arisen a large literature on this subject (see, for instance, [28, Chap. 2]). Sturmian words can be defined in many different ways of a combinatorial or geometric nature.

In the theory a key role is played by characteristic (or standard) Sturmian words, which can be generated in several different ways and, in particular, by a palindromization map ψ, introduced by the first author in [12], which injectively maps each finite word v, called the directive word, to a palindrome (cf. Section 2.1). The map ψ can be naturally extended to infinite words. In such a case, if v is any infinite binary word in which all letters occur infinitely often, one generates all characteristic Sturmian words. An infinite word is Sturmian if its set of finite factors equals the set of finite factors of a characteristic Sturmian word. The set of all $\psi(v)$, with v any finite word over the binary alphabet $\mathcal{A} = \{a, b\}$, coincides with the set of palindromic prefixes of all characteristic Sturmian words [12,17].

* Corresponding author.
E-mail addresses: aldo.deluca@unina.it (A. de Luca), alessandro.deluca@unina.it (A. De Luca).

1 Partially supported by GNSAGA, INdAM.
2 Partially supported by the Italian Ministry of Education (MIUR), under the PRIN project number 2010LYA9RH, titled "Automi e Linguaggi Formali: Aspetti Matematici e Applicativi".

http://dx.doi.org/10.1016/j.tcs.2015.02.043
0304-3975/

The words $\psi(v)$, called central words, may be also defined in a purely combinatorial way as the set of all words having two coprime periods p and q such that the length $|\psi(v)| = p + q - 2$. Central words $\psi(v)$ are strongly related [3,12] to proper finite standard words, which may be defined as $\psi(v)xy$ with $x, y \in \mathcal{A}$, and to proper Christoffel words $a\psi(v)b$.

Central, standard, and Christoffel words are considered in Section 3. They represent a finite counterpart to characteristic Sturmian words of great interest, since there exist several faithful representations of the preceding words by trees, binary matrices, and continued fractions [3]. These representations give a natural arithmetization of the theory. Some new results are proved at the end of the section.

As regards trees, we mainly refer in Section 4 to the Raney tree. This tree is a complete binary tree rooted at the fraction $\frac{1}{1}$ and any rational number represented in a node as the irreducible fraction $\frac{p}{q}$ has two children representing the numbers $\frac{p}{p+q}$ and $\frac{p+q}{q}$. Every positive rational number appears exactly once in the tree. This tree is usually called the Calkin–Wilf tree in the literature after Neil Calkin and Herbert Wilf, who considered it in their 2000 paper [7]. However, the tree was introduced earlier by Jean Berstel and the first author [3] under the name Raney tree, since they drew some ideas from a paper by George N. Raney [34].

The fraction $\mathrm{Ra}(w)$ in the node of the Raney tree represented by the binary word w is equal to the ratio $\frac{p}{q}$ of the periods of the central word $\psi(w)$, where p (resp., q) is the minimal period of $\psi(w)$ if w terminates with the letter a (resp., b).

Another very important tree that can be considered as dual of Raney tree is the Stern–Brocot tree (see, for instance, [22, 29]). One can prove (cf. [3]) that the fraction $\mathrm{Sb}(w)$ in the node w of the Stern–Brocot tree is equal to the slope $\frac{|a\psi(w)b|_b}{|a\psi(w)b|_a}$ of the Christoffel word $a\psi(w)b$. The duality is due to the fact that

$$\mathrm{Sb}(w) = \mathrm{Ra}(w^{\sim}),$$

where $w^{\sim}$ is the reversal of the word w.

The sequence formed by the denominators of the fractions labeling the Raney tree is the famous diatomic sequence introduced in 1858 by Moritz A. Stern [35]. There exists a large literature on this sequence, that we shall simply refer to as Stern's sequence, since its terms admit interpretations in several parts of combinatorics and satisfy many surprising and beautiful properties (see, for instance, [1,7,8,11,18,26,32,36] and references therein).

In this paper we are mainly interested in the properties of Stern's sequence which are related to combinatorics of Christoffel and central words. In Section 5, using some properties of the Raney tree, we prove that there exists a basic correspondence (cf. Theorem 5.2) between the values of Stern's sequence on *odd* integers and the lengths of Christoffel words, as well as a correspondence between the values of the sequence on *even* integers and the minimal periods of central words. Thus there exists a strong relation between Sturmian words and Stern's sequence which strangely, with the only exception of [18], has not been observed in the literature.

As a consequence of the previous correspondence several results on Stern's sequence can be proved by using the theory of Sturmian words and, conversely, properties of Stern's sequence can give new insight into the combinatorics of Christoffel and central words.

In Section 6 we show that one can compute the terms of Stern's sequence by continuants in two different ways. The first uses a result concerning the length of a Christoffel word $a\psi(v)b$ and the minimal period of the central word $\psi(v)$, which can be expressed in terms of continuants operating on the integral representation of the directive word v. The second is of a more arithmetical nature and uses known results on Stern's sequence.

In Section 7 we consider a very interesting and unpublished theorem of Calkin and Wilf on Stern's sequence [8, Theorem 5]. The Calkin–Wilf theorem states that for each n, the term $s(n)$ of the sequence represents the number of "alternating bit sets" in n, i.e., the number of occurrences of subsequences (subwords) in the binary representation of n belonging to the set $b(ab)^*$. We provide a formula allowing to compute the length of the Christoffel word $a\psi(v)b$ as the number of occurrences of subwords $u \in b(ab)^*$ in bvb. Moreover, if v is not a power of a letter, the minimal period of $\psi(v)$ equals the number of occurrences $u \in b(ab)^*$ in bv_+b, where v_+ is the longest prefix of v immediately followed by a letter different from the last letter of v.

The main result of the section is the following quite surprising theorem (cf. Theorem 7.3). For any $w \in \mathcal{A}^*$, consider the reversed occurrences of words of the set $b(ab)^*$ as subwords in bwb, and then sort these in decreasing lexicographic order. An occurrence is said to be initial if it begins in the first position of bwb. Then, marking the reversed initial occurrences with a and the reversed non-initial ones with b, one obtains the standard word $\psi(w)ba$. This can be regarded as a non-commutative version of the Calkin–Wilf theorem, from which the original theorem can be easily derived. In particular, we obtain that $|a\psi(v)b|_a$ is equal to the number of initial occurrences of subwords $u \in b(ab)^*$ in bvb.

In Section 8 we prove a formula (cf. Theorem 8.2) that for each $w \in \mathcal{A}^*$ relates the length of the Christoffel word $a\psi(w)b$ with the occurrences in bwb of a certain kind of factors whose number is weighted by the lengths of Christoffel words associated with suitable directive words which are factors of w. The result is a consequence of an interesting theorem on Stern's sequence due to Michael Coons and Jeffrey Shallit [11].

In Section 9 we study the distribution of the lengths of Christoffel words $a\psi(v)b$ of order k, i.e., the directive word v has a fixed length k. Using a property of Stern's sequence we show that the average value of the length is $2 \cdot (3/2)^k$. Moreover, the maximal value given by F_{k+1}, where $(F_k)_{k \geq -1}$ is the well-known Fibonacci sequence, is attained if and only if v is alternating, i.e., any letter in v is immediately followed in v by its complementary letter.