A Characterization of Bispecial Sturmian Words (extended abstract)

Gabriele Fici

I3S, CNRS & Université Nice Sophia Antipolis, France
fici@i3s.unice.fr

Abstract. We show that bispecial Sturmian words are exactly the maximal internal factors of Christoffel words. This result is an extension of the known relation between central words and primitive Christoffel words. Our characterization allows us to give an enumerative formula for bispecial Sturmian words.

1 Introduction

Sturmian words are non-periodic infinite words of minimal factor complexity. They are characterized by the property of having exactly $n+1$ distinct factors of length n for every $n \geq 0$ (and therefore are binary words) [5]. The set *St* of finite factors of Sturmian words coincides with the set of binary *balanced* words, i.e., binary words having the property that any two factors of the same length have the same number of occurrences of each letter up to one. If one considers extendibility within the set *St*, one can define *left special Sturmian words* (resp. *right special Sturmian words*) [4] as those words w over the alphabet $\Sigma = \{a, b\}$ such that aw and bw (resp. wa and wb) are both Sturmian words. The Sturmian words that are both left special and right special are called *bispecial Sturmian words*. They are of two kinds: *strictly bispecial Sturmian words*, that are the words w such that awa, awb, bwa and bwb are all Sturmian words, or *weakly bispecial Sturmian words* otherwise. Strictly bispecial Sturmian words have been deeply studied (see for example [2, 4]) because they play a central role in the theory of Sturmian words. They are also called *central words*. Weakly bispecial Sturmian words, instead, received less attention.

One important field in which Sturmian words arise naturally is discrete geometry. Indeed, Sturmian words can be viewed as digital approximations of straight lines in the Euclidean plane. It is known that given a point (p, q) in the discrete plane $\mathbb{Z} \times \mathbb{Z}$, with $p, q > 0$, there exists a unique path that approximates from below (resp. from above) the segment joining the origin $(0, 0)$ to the point (p, q). This path, represented as a concatenation of horizontal and vertical unitary segments, is called the *lower (resp. upper) Christoffel word* associated to the pair (p, q). If one encodes horizontal and vertical unitary segments with the letters a and b respectively, a lower (resp. upper) Christoffel word is always a word of the form awb (resp. bwa), for some $w \in \Sigma^*$. If (and only if) p and q are coprime, the associated Christoffel word is primitive (that is, it is not the power of a shorter word). It is known that a word w is a strictly bispecial Sturmian word if and only if awb is a primitive lower Christoffel word (or, equivalently, if and only if bwa is a primitive upper Christoffel word). As a main result of this paper, we show that this correspondence holds in general between bispecial Sturmian words and Christoffel words. That is, we prove (in Theorem 2) that w is a bispecial Sturmian word if and only if there exist letters x, y in $\{a, b\}$ such that xwy is a Christoffel word. This characterization allows us to prove an enumerative formula for bispecial Sturmian words (Corollary 1).

2 Sturmian words and Christoffel words

A finite word w over $\Sigma = \{a, b\}$ is Sturmian if and only if for any $u, v \in Fact(w)$ such that $|u| = |v|$ one has $||u|_a - |v|_a| \leq 1$. We let St denote the set of finite Sturmian words.

Let w be a finite Sturmian word. The following definitions are in [4].

Definition 1. *A word $w \in \Sigma^*$ is a left special (resp. right special) Sturmian word if $aw, bw \in St$ (resp. if $wa, wb \in St$). A bispecial Sturmian word is a Sturmian word that is both left special and right special. Moreover, a bispecial Sturmian word is strictly bispecial if awa, awb, bwa and bwb are all Sturmian word; otherwise it is non-strictly bispecial.*

We let *LS*, *RS*, *BS*, *SBS* and *NBS* denote, respectively, the sets of left special, right special, bispecial, strictly bispecial and non-strictly bispecial Sturmian words.

The following lemma is a reformulation of a result of de Luca [3].

Lemma 1. *Let w be a word over Σ. Then $w \in LS$ (resp. $w \in RS$) if and only if w is a prefix (resp. a suffix) of a word in SBS.*

Given a bispecial Sturmian word, the simplest criterion to determine if it is strictly or non-strictly bispecial is provided by the following nice characterization [4]:

Proposition 1. *A bispecial Sturmian word is strictly bispecial if and only if it is a palindrome.*

Using the results in [4], one can derive the following classification of Sturmian words with respect to their extendibility.

Proposition 2. *Let w be a Sturmian word. Then:*

- $|\Sigma w \Sigma \cap St| = 4$ *if and only if w is strictly bispecial;*
- $|\Sigma w \Sigma \cap St| = 3$ *if and only if w is non-strictly bispecial;*
- $|\Sigma w \Sigma \cap St| = 2$ *if and only if w is left special or right special but not bispecial;*
- $|\Sigma w \Sigma \cap St| = 1$ *if and only if w is neither left special nor right special.*

We now recall the definition of central word [4].

Definition 2. *A word over Σ is central if it has two coprime periods p and q and length equal to $p+q-2$.*

We have the following remarkable result [4]:

Proposition 3. *A word over Σ is a strictly bispecial Sturmian word if and only if it is a central word.*

Another class of finite words, strictly related to the previous ones, is that of Christoffel words.

Definition 3. *Let $n > 1$ and $p, q > 0$ be integers such that $p+q = n$. The lower Christoffel word $w_{p,q}$ is the word defined for $1 \leq i \leq n$ by*

$$w_{p,q}[i] = \begin{cases} a \text{ if } iq \ mod(n) > (i-1)q \ mod(n), \\ b \text{ if } iq \ mod(n) < (i-1)q \ mod(n). \end{cases}$$

If one draws a word in the discrete grid $\mathbb{Z} \times \mathbb{Z}$ by encoding each a with a horizontal unitary segment and each b with a vertical unitary segment, the lower Christoffel word $w_{p,q}$ is in fact the best grid approximation from below of the segment joining $(0, 0)$ to (p, q), and has slope q/p, that is, $|w|_a = p$ and $|w|_b = q$.

Analogously, one can define the upper Christoffel word $w'_{p,q}$ by

$$w'_{p,q}[i] = \begin{cases} a \text{ if } ip \bmod(n) < (i-1)p \bmod(n), \\ b \text{ if } ip \bmod(n) > (i-1)p \bmod(n). \end{cases}$$

Of course, the upper Christoffel word $w'_{p,q}$ is the best grid approximation from above of the segment joining $(0, 0)$ to (p, q).

The next result follows from elementary geometrical considerations.

Lemma 2. *For every pair (p, q) the word $w'_{p,q}$ is the reversal of the word $w_{p,q}$.*

If (and only if) p and q are coprime, the Christoffel word $w_{p,q}$ intersects the segment joining $(0, 0)$ to (p, q) only at the end points, and is a primitive word. Moreover, one can prove that $w_{p,q} = aub$ and $w'_{p,q} = bua$ for a palindrome u. Since u is a bispecial Sturmian word and it is a palindrome, u is a strictly bispecial Sturmian word (by Proposition 1). Conversely, given a strictly bispecial Sturmian word u, u is a central word (by Proposition 3), and therefore has two coprime periods p, q and length equal to $p+q-2$. Indeed, it can be proved that $aub = w_{p,q}$ and $bua = w'_{p,q}$. The previous properties can be summarized in the following theorem (cf. [1]):

Theorem 1. *$SBS = \{w \mid xwy$ is a primitive Christoffel word, $x, y \in \Sigma\}$.*

If instead p and q are not coprime, then there exist coprime integers p', q' such that $p = rp'$, $q = rq'$, for an integer $r > 1$. In this case, we have $w_{p,q} = (w_{p',q'})^r$, that is, $w_{p,q}$ is a power of a primitive Christoffel word. Hence, there exists a central Sturmian word u such that $w_{p,q} = (aub)^r$ and $w'_{p,q} = (bua)^r$. So, we have:

Lemma 3. *The word xwy, $x \neq y \in \Sigma$, is a Christoffel word if and only if $w = (uyx)^n u$, for an integer $n \geq 0$ and a central word u. Moreover, xwy is a primitive Christoffel word if and only if $n = 0$.*

Recall from [3] that the right (resp. left) palindromic closure of a word w is the (unique) shortest palindrome $w^{(+)}$ (resp. $w^{(-)}$) such that w is a prefix of $w^{(+)}$ (resp. a suffix of $w^{(-)}$).

Lemma 4. *Let xwy be a Christoffel word, $x, y \in \Sigma$. Then $w^{(+)}$ and $w^{(-)}$ are central words.*

Theorem 2. $BS = \{w \mid xwy$ *is a Christoffel word,* $x, y \in \Sigma\}$.

Proof. (Sketch) Let xwy be a Christoffel word, $x, y \in \Sigma$. Then, by Lemma 3, w is of the form $w = (uyx)^n u$, $n \geq 0$, for a central word u. By Lemma 4, w is a prefix of the central word $w^{(+)}$ and a suffix of the central word $w^{(-)}$, and therefore, by Lemma 1 and Proposition 3, w is a bispecial Sturmian word.

Conversely, let w be a bispecial Sturmian word. If w is strictly bispecial, then w is a central word by Proposition 3, and xwy is a (primitive) Christoffel word by Theorem 1. So suppose $w \in NBS$. By Lemma 3, it is enough to prove that w is of the form $w = (uyx)^n u$, $n \geq 1$, for a central word u and letters $x \neq y$. This can be proven by contradiction using the property of balanceness of Sturmian words. □

3 Enumeration of bispecial Sturmian words

It is known [4] that $SBS(n) = \phi(n+2)$. Therefore, in order to find an enumerative formula for bispecial Sturmian words, we only have to enumerate the non-strictly bispecial Sturmian words. We do this in the next proposition.

Proposition 4. *For every $n > 1$, one has*

$$NBS(n) = 2\,(n + 1 - \phi(n + 2))$$

Corollary 1. *For every $n \geq 0$, there are $2(n+1) - \phi(n+2)$ bispecial Sturmian words of length n.*

References

1. J. Berstel and A. de Luca. Sturmian Words, Lyndon Words and Trees. *Theoret. Comput. Sci.*, 178(1-2):171–203, 1997.
2. A. Carpi and A. de Luca. Central Sturmian Words: Recent Developments. In *Developments in Language Theory, 9th International Conference, DLT 2005*, volume 3572 of *Lecture Notes in Comput. Sci.*, pages 36–56. Springer, 2005.
3. A. de Luca. Sturmian Words: Structure, Combinatorics, and Their Arithmetics. *Theoret. Comput. Sci.*, 183(1):45–82, 1997.
4. A. de Luca and F. Mignosi. Some combinatorial properties of Sturmian words. *Theoret. Comput. Sci.*, 136(2):361–385, 1994.
5. M. Morse and G. A. Hedlund. Symbolic dynamics II: Sturmian Trajectories. *Amer. J. Math.*, 62:1–42, 1940.