RAIRO-Theor. Inf. Appl. 59 (2025) 18
https://doi.org/10.1051/ita/2025016

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

A SYMMETRY PROPERTY OF CHRISTOFFEL WORDS

Yan Lanciault and Christophe Reutenauer*

Abstract. Motivated by the theory of trapezoidal words, whose sequences of cardinality of factors by length is symmetric, we introduce a bivariate variant of this symmetry, called Symmetry. We show that this property characterizes Christoffel words, and prove other related results.

Mathematics Subject Classification. 05A05, 68R15.

Received January 30, 2025. Accepted September 25, 2025.

1. Introduction

Trapezoidal words were in essence considered by Aldo de Luca in [1]; for such a word, w say, of length n, the graph of the discrete function $\{0, 1, \ldots, n\} \to \mathbb{N}$, giving the number of distinct factors of length k of w is an iscoceles trapezoid, with successive values $1, 2, \ldots, J, \ldots, J, \ldots, 2, 1$. He showed that Sturmian words are trapezoidal. The terminology "trapezoidal" was introduced by Flavio d'Alessandro in [2], who studied these words, giving in particular a condition for which a trapezoidal word is Sturmian. In [3], Michelangelo Bucci, Alessandro De Luca and Gabriele Fici gave many equivalent conditions for a word to be trapezoidal; one of them is that the number of factors of length k is at most $k + 1$

A remarkable property of trapezoidal words is that the sequence of the lengths of the factors of these words, from length 0 to length n, is symmetric.

In the present work, we present a generalization of this symmetry property. Let w be a word on the alphabet a, b, with p times the letter a and q times the letter b; in other words, the Parikh image (or commutative image) of w is (p, q). We say that w is *Symmetric* if for any i, j, w has as many factors with Parikh image (i, j) as factors of Parikh image $(p - i, q - j)$.

There are many equivalent definitions of Christoffel words. One of them is given in Section 2. It is well-known that Christoffel words are primitive finite Sturmian words.

This article is about the link between Christoffel words and Symmetric words. The main results are the three following ones. Theorem 3.1 asserts that each Christoffel word is Symmetric. Theorem 3.10 asserts, conversely, that each primitive, Sturmian and Symmetric word is a Christoffel word. Theorem 4.1 asserts that, for $k \geq 2$, w^k is Symmetric if and only w is a Christoffel word.

A one observes, the primitive and nonprimitive cases are rather different, and for the converse (Thm. 3.10) of Theorem 3.1, we need the Sturmian hypothesis. These observations are illustrated by the fact that *aabb* is primitive and Symmetric, but not Sturmian, and its square *aabbaabb* is not Symmetric.

Keywords and phrases: Combinatorics on words, Christoffel words, Sturmian words, trapezoidal, attractor, symmetry.

Département de mathématiques, Université du Québec à Montréal, Canada.

* Corresponding author: reutenauer.christophe@uqam.ca

An open problem is the characterization of primitive non Sturmian Symmetric words. We discuss this open problem in Section 5.

The above main results characterize Christoffel words. In the literature, one finds two other characterizations of Christoffel words: a Sturmian word is a Christoffel word, if and only if it is unbordered [4] (see also [5]), if and only if it is a Lyndon word for one of the two total orders on the alphabet [6]. There are also many characterizations of conjugates of Christoffel words, [7–12]), which do not distinguish between Christoffel words and their conjugates.

We end the article with several byproducts. Concerning the Symmetry of a Christoffel word w, we give an explicit bijection between the factors of w of Parikh image (i, j) onto those of Parikh image $(p - i, q - j)$ (Prop. 5.2); it rests on the notion of *attractor* [13, 14], and *circular attractor* [10]. Moreover, the support of the function δ_w, which maps $(i, j) \in \mathbb{N}^2$ onto the number of factors of w of Parikh image (i, j), which is a subset of the discrete plane, appears to be the set of integer points on the two paths defined by w and its reversal $\tilde{w}$, and between them (Prop. 5.1): see for example Figure 1.

2. Christoffel words, trapezoidal words, and Symmetry

Christoffel words are elements of A^*, the free monoid on the two-letter alphabet $A = \{a, b\}$. The original definition of Elwin Bruno Christoffel uses an algorithm based on modular arithmetic [15] (see also [11], Thm. 15.1.1); we use another definition.

Recall that a *palindrome* is a word h such that $h = \tilde{h}$, where $\tilde{h}$ is the *reversal* of h.

A word $w \in A^*$ is a *Christoffel word* if there exist palindromes $m, u, v \in A^*$ such that one has $w = amb$ or $w = bma$, and $w = vu$ (this characterization of Christoffel words is due to de Luca and Mignosi [16], Prop. 8; see also [11] Thm. 14.3.10). In this case, m is called the *central palindrome* of w, and vu is called the *palindromic factorization* of w. It follows from the definition that $\tilde{w}$ is also a Christoffel word. We denote by $\mathcal{C}$ the set of Christoffel words.

Exemple 2.1. Consider the Christoffel word $w = aabaabab$. Its central palindrome is $m = abaaba$, and its palindromic factorization is $w = aabaa \cdot bab$.

Recall that a word f is a *factor* of w if there exist words p, s such that $w = pfs$. Denote by $\mathcal{F}_w = \{f \in A^* \mid f \text{ is a factor of } w\}$ the set of factors of w.

Following D'Alessandro ([2], p. 11), we say that a word w is *trapezoidal* if for each i, w has at most one special factor of length i (that is, a factor which is the prefix of two distinct factors of w of length $i + 1$). Trapezoidal word have several characterizations. We give two of them. For this, consider for any word w the function $f_w : i \mapsto$ number of factors of length i of w, defined for $i = 0, 1, \ldots, |w|$.

Theorem 2.2. *([3] Prop. 2.8) The following conditions are equivalent for a word $w \in \{a, b\}^*$:*

(i) w is trapezoidal;

(ii) there exist $m_w \leq M_w$ such that $f_w(i) = i + 1$ for $i = 0, \ldots, m_w$, f_w is constant on $\{m_w, m_w + 1, \ldots, M_w\}$, and $f_w(i + 1) = f_w(i) - 1$ for $i = M_w, \ldots, |w| - 1$.

(iii) $f_w(i) \leq i + 1$ for any i.

Note that condition (iii) was called *stiff* by Alex Heinis in [17].

Remarks. The fact that (iii) implies (i) was stated in [3]; actually, the proof needs a refinement of a result of Aldo de Luca ([1] Prop. 4.2), which was given by Séébold and Levé ([18] Prop. 3.3) and also by Anisiu and Cassaigne ([19] Thm. 6).

There is another easy equivalent condition (the if implication seems to be new).

Corollary 2.3. *A word of length n is trapezoidal if and only if it has for $i = 0, 1, \ldots, n$ as many factors of length i as factors of length $n - i$.*

Lemma 2.4. *If a word w of length n has for $i = 0, 1, \ldots, n$ as many factors of length i as factors of length $n - i$, it has for each $i = 0, 1, \ldots, |w|$ at most $i + 1$ factors of length i.*

Proof. There are at most $i + 1$ occurrences of factors of length $|w| - i$; hence there are at most $i + 1$ factors of length $|w| - i$. Thus, by hypothesis, there at most $i + 1$ factors of length $|w| - (|w| - i) = i$. □

Proof of Corollary 2.3. Suppose that w has for $i = 0, 1, \ldots, n$ as many factors of length i as factors of length $n - i$. Then by Lemma 2.4 and Theorem 2.2 (iii), w is trapezoidal.

Conversely, if w is trapezoidal, it satisfies (ii) in Theorem 2.2; this expresses the fact that the graph of f_w is a discrete isosceles trapezoid; hence w has for $i = 0, 1, \ldots, n$ as many factors of length i as factors of length $n - i$. □

Define $\gamma : A^* \to \mathbb{N}^2$ by $\gamma(w) = (|w|_a, |w|_b)$, where $|w|_x$ is the number of occurrences of letter x in w; $\gamma(w)$ is called the *commutative image* of w, and γ is called the *Parikh mapping*. Note that γ is a monoid homomorphism. Next, for $w \in A^*$, we define $\delta_w : \mathbb{N}^2 \to \mathbb{N}$, by $\delta_w(i, j) = |\{f \in \mathcal{F}_w \mid \gamma(f) = (i, j)\}|$. Thus, $\delta_w(i, j)$ is the number of factors f of w such that $|f|_a = i$ and $|f|_b = j$.

Exemple 2.5. Let $w = aabaabab$. Then $\gamma(w) = (5, 3)$. Moreover, $\delta_w(3, 2) = 2$, where the two factors of w having three a's and two b's are $abaab$ and $aabab$.

Definition 2.6. Let $w \in A^*$ be such that $\gamma(w) = (p, q)$. We say that w is *Symmetric* if $\forall (i, j) \in [p] \times [q]$, one has $\delta_w(i, j) = \delta_w(p - i, q - j)$.

Here we have used the notation $[n] = \{0, 1, \ldots, n\}$. In other words, w is Symmetric if w has as many factors having i letters a and j letters b as factors having $p - i$ letters a and $q - j$ letters b.

3. Primitive case

Recall that Christoffel words are *primitive*, that is, not the power of another word.

Theorem 3.1. *Christoffel words are Symmetric.*

We may visualize this symmetry by representating δ_w by a matrix, that we call the *matrix of* w, equal to $\mathcal{M}_w = (\delta_w(i, j))_{i \in [p],\ j \in [q]}$. Clearly, w is Symmetric if and only if the matrix of w is symmetric with respect to its center.

Exemple 3.2. Consider the Christoffel word $w = aababab$. Its matrix $\mathcal{M}_w$ is (we represent the rows $0, 1, \ldots,$ from bottom to top, and the columns from left to right):

$$\mathcal{M}_w = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

For the proof of the theorem, we need several notions on words. We call *circular factor* of a word w any factor of ww which is of length at most $|w|$; we denote by $\mathcal{F}_w^c$ the set of circular factors of w. A circular factor may also be seen as a factor of w, when the latter is written on circle. Denote also $\mathcal{P}_w := \{p \mid p \text{ is a prefix of } w\}$ and $\mathcal{S}_w := \{s \mid s \text{ is a suffix of } w\}$.

The next proposition will be needed in the proof; see [11], Theorem 15.3.1 and Corollary 15.3.5.

Proposition 3.3. *Let $w \in \mathcal{C}$ and $k < |w|$. Let $C_k = \{sp \mid s \in \mathcal{S}_w \text{ and } p \in \mathcal{P}_w,\ |sp| = k\}$. Then $|C_k| = k + 1$. Moreover the number of circular factors of w of length k is equal to $k + 1$.*

In other words, the words sp of length $< |w|$, with $s \in \mathcal{S}_w$ and $p \in \mathcal{P}_w$, are distinct and are all the circular factors of length $< |w|$.

Now two key notions for the proof are the notions of attractor [13, 14], and of circular attractor, [10]. If $w = a_1 \cdots a_n$ with $a_i \in \{a, b\}$, an *attractor* of w is a subset $K \subset \{1, 2, \ldots, n\}$ such that for any nonempty factor f of w, there exist $1 \leq i \leq j \leq n$ such that $f = a_i \cdots a_j$ and $\{i, \ldots, j\} \cap K \neq \emptyset$. Informally, an attractor is a set of marked letters in w such that any factor of w has an occurrence in w containing one of these marked letters.

Informally also, a circular attractor is a set of marked letters of w such that any circular factor has an occurrence containing one of these marked letters. Formally: a subset K of $\{1, \ldots, n\}$ is called a *circular attractor* of w if for any nonempty circular factor f of w, there exist i, j such that either: $1 \leq i \leq j \leq n$, $f = a_i \cdots a_j$, and K intersects the interval $\{i, i+1, \ldots, j\}$, or: $1 \leq i < j \leq n$, $f = a_j \cdots a_n a_1 \cdots a_i$, and K intersects $\{1, 2, \ldots, i\} \cup \{j, j+1, \ldots, n\}$.

The next result is a particular case of Theorem 34 in [10].

Corollary 3.4. *Let $w \in \mathcal{C}$ with its palindromic factorization $w = vu$. Let $k = |v|$. Then $\{k, k+1\}$ is a circular attractor of w.*

Proof. The reversal $\tilde{w}$ of w is equal to uv, since u, v are palindromes. Hence w and $\tilde{w}$, which are both Christoffel words, have the same circular factors. Applying Proposition 3.3 to $\tilde{w} = uv$, we see that each circular factor of w has an occurrence containing the last letter of v or the first letter of u. Hence we obtain the corollary. □

A *period* of a word $w = b_1 \cdots b_m$ ($b_i \in A$) is any $k \in \mathbb{N}$ such that for any $i = 1, \ldots, m-k$, one has $b_i = b_{i+k}$. It is known that if $w = a_1 \cdots a_n$ is a Christoffel word w with palindromic factorization $w = vu$, then the word $a_2 \cdots a_n$ has period $k = |v|$ (see for example [11], Thm. 12.1.8 and Lem. 12.2.5).

Corollary 3.5. *Let $w \in \mathcal{C}$ with $w = a_1 \cdots a_n$ ($a_i \in A$) and $w = vu$ its palindromic factorization. Let $k = |v|$. Then $K = \{k, k+1\}$ is an attractor of w. Moreover, for any nonempty factor f of w, there exists a unique occurrence of f which intersects K. That is,*

$$\forall f \in \mathcal{F}_w \setminus 1,\ \exists\, !(i,j),\ 1 \leq i \leq j \leq n \text{ such that } f = a_i \cdots a_j \text{ and } \{i, \ldots, j\} \cap K \neq \emptyset.$$

Proof. Uniqueness: We have $\tilde{w} = uv$ since u and v are palindromes. Thus, $\tilde{w}\tilde{w} = uvuv = uwv$. Applying Proposition 3.3 to $\tilde{w}$ (which is a Christoffel word), we obtain that, if $\{k, k+1\}$ is an attractor of w, then the factors intersect it uniquely.
Existence: Let $v = a_1 \cdots a_k$ and $u = a_{k+1} \cdots a_n$. Let $f \in \mathcal{F}_w$ with $|f| = \ell$. If the letter a_k or a_{k+1} is in f we are done. Otherwise, f is strictly at the left of a_k, or strictly at the right of a_{k+1}. We only treat the second case, the first being similar.

Thus f is strictly at the right of a_{k+1}: $f = a_t \cdots a_{t+\ell-1}$, $t > k+1$; we then have $\ell < n-k$. We know that the word $a_2 \cdots a_n$ has the period k. Hence $f = a_{t-\alpha k} \cdots a_{t+\ell-1-\alpha k}$ for any $\alpha \in \mathbb{N}$ such that $t - \alpha k > 1$. There exists an integer α such that $j = t - \alpha k \in \{2, \ldots, k+1\}$. Since $k+1 < t$, $\alpha \geq 1$. Thus f has an occurrence whose first letter is in $a_2 \cdots a_{k+1}$. Since the length of f is ℓ, if $\ell \geq k-1$, a_k or a_{k+1} is in this occurrence of f and we obtain the result. Otherwise, we have $\ell < k-1$; we also have $\ell < n-k$. By Corollary 3.4, $\{k, k+1\}$ is a circular attractor of w. Hence f has a circular occurrence which intersects $a_k a_{k+1}$. The two previous inequalities imply that this occurrence is a factor of w. Thus, in each case, there exists an occurrence of f which intersects $a_k a_{k+1}$.
In conclusion, $\{k, k+1\}$ is an attractor of w. □

In order to prove Theorem 3.1, it is enough to show that, for each Christoffel word w, the polynomial equal to the sum of all commutative monomials $x^{|f|_a} y^{|f|_b}$, for all factors f of w, is reciprocal (formal definition below); this sum is an element of the ring of polynomials $\mathbb{Z}[x, y]$. For this, it is convenient to consider first the noncommutative version of this polynomial, that is, the sum of all factors f of w, which is an element of the ring of noncommutative polynomials in the noncommuting variables a, b; then, to apply the natural ring homomorphism from the latter ring to the former, mapping a onto x and b onto y.

To be more formal, we denote by $\mathbb{Z}\langle a, b\rangle$ the $\mathbb{Z}$-algebra of polynomials in the noncommuting variables a et b. An element of this algebra is a $\mathbb{Z}$-linear combination of words. The reader may have a look at [20], Chapter 1, for more details.

We obtain the following corollary.

Corollary 3.6. *One has in* $\mathbb{Z}\langle a, b\rangle$

$$\sum_{f\in\mathcal{F}_w} f = \left(\sum_{\sigma\in\mathcal{S}_v} \sigma\right)\left(\sum_{\pi\in\mathcal{P}_u} \pi\right).$$

The reader who does not like noncommutative polynomials may translate this result in language theory, as follows: the language $\mathcal{F}_w$ is the unambiguous product of the languages $\mathcal{S}_v$ and $\mathcal{P}_u$.

Proof. Each word of the form $\sigma\pi$, for some $\sigma \in \mathcal{S}_v$ and $\pi \in \mathcal{S}_u$, is clearly a factor of w. Conversely, by Corollary 3.5, each nonempty factor of w has an occurrence containing the last letter de v or the first letter de u; hence each factor f of w is of the form $f = \sigma\pi$ for some $\sigma \in \mathcal{S}_v$ and $\pi \in \mathcal{S}_u$. Moreover, the same corollary implies that the words $\sigma\pi$ are all distinct, hence the coefficients in the product are all equal to 1. Thus, the two sides of the equality have the same support, and the words appearing have coefficient 1. Hence the equality holds. □

We say that a polynomial $P \in \mathbb{Z}[x, y]$ (commuting variables!) is *reciprocal* if there exists $(p, q) \in \mathbb{N}^2$ such that

$$P = \sum_{\substack{0\le i\le p\\ 0\le j\le q}} \alpha_{i,j} x^i y^j,$$

and that $\forall (i, j) \in [p] \times [q]$, one has $\alpha_{i,j} = \alpha_{p-i,q-j}$.

Exemple 3.7. The polynomial $P(a, b) = 1 + 3a^2b + 5a^3b^2 + 3a^3b^4 + 5a^2b^3 + a^5b^5$ is reciprocal, with $(p, q) = (5, 5)$.

Lemma 3.8. *1. If* $P(x, y) \in \mathbb{Z}[x, y]$*, then* P *is reciprocal if and only if there exists* $(p, q) \in \mathbb{N}^2$ *such that*

$$P(x, y) = x^p y^q P(x^{-1}, y^{-1}).$$

2. The product of two reciprocal polynomials is reciprocal.

The proof is straightforward and left to the reader.

Lemma 3.9. *If* $h \in A^*$ *is a palindrome, then the polynomials*

$$\sum_{\pi\in\mathcal{P}_h} x^{|\pi|_a} y^{|\pi|_b}, \quad \sum_{\sigma\in\mathcal{S}_h} x^{|\sigma|_a} y^{|\sigma|_b}$$

are reciprocal.

Proof. Let $\gamma(h) = (p, q)$ and $p + q = n$. Let π be a prefix of h such that $\gamma(\pi) = (i, j)$ and $|\pi| = i + j = k$. Let π' be the prefix of h of length $n - k$. We have $h = \pi\sigma = \pi'\sigma'$. We show that $\gamma(\pi') = (p - i, q - j)$. Since h is a palindrome, we have $\tilde{\pi} = \sigma'$. Thus, $\gamma(\pi) = \gamma(\tilde{\pi}) = \gamma(\sigma')$. Since γ is a monoid homomorphism, $\gamma(\pi') = \gamma(h) - \gamma(\sigma') = \gamma(h) - \gamma(\pi) = (p - i, q - j)$; hence the first polynomial in the lemma is reciprocal. For the second, the argument is similar. □

In order to prove the theorem, for $w \in \{a, b\}^*$ such that $\gamma(w) = (p, q)$, let

$$P_w(x, y) = \sum_{f\in\mathcal{F}_w} x^{|f|_a} y^{|f|_b} = \sum_{(i,j)\in[p]\times[q]} \delta_w(i, j) x^i y^j \in \mathbb{Z}[x, y].$$

Proof of Theorem 3.1. It is enough to show that for each $w \in \mathcal{C}$, P_w is reciprocal. By Corollary 3.6, applying the $\mathbb{Z}$-algebra homomorphism $\mathbb{Z}\langle a, b\rangle \to \mathbb{Z}[x, y]$, with $a \mapsto x$ et $b \mapsto y$, we obtain that

$$P_w = \left(\sum_{\sigma \in \mathcal{S}_v} x^{|\sigma|_a} y^{|\sigma|_b} \right) \left(\sum_{\pi \in \mathcal{P}_u} x^{|\pi|_a} y^{|\pi|_b} \right).$$

By Lemma 3.9, the two factors are reciprocal polynomial, since u, v are palindromes. Hence by Lemma 3.8, P_w is reciprocal. □

Strictly speaking, Theorem 3.1 has no converse. Indeed, *aabb* is not a Christoffel word (it has no central palindrome), but is Symmetric. indeed,

$$\mathcal{M}_{aabb} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

clearly has a central symmetry.

Recall that, among other equivalent definitions, a (finite) word is *Sturmian* if it is a factor of some Christoffel word.

Theorem 3.10. *Let w be a word which is Symmetric, primitive, and Sturmian. Then w is a Christoffel word.*

Note that *aabb* is not a Sturmian word.

Lemma 3.11. *Let $w \in A^*$ and $0 \leq \rho \leq |w|$. Then w has period ρ if and only if the set $D_\rho = \{(x, y) \mid x + y = \rho,\ \delta_w(x, y) > 0\}$ is a singleton.*

Proof. Suppose that w has the period ρ. Then all factors of length ρ have the same commutative image. Hence there is a unique pair (i, j) which is the commutative image of a factor of length ρ. Thus the set in the statement is a singleton.

Conversely, suppose that there is a unique pair (i, j) of sum ρ in the support of δ_w. Then all factors of w of length ρ have the same commutative image. Hence, if $xuy \in \mathcal{F}_w$, $x, y \in A$ and $u \in A^*$ such that $|xu| = \rho$, then $x = y$, since $|xu|$ and $|uy|$ have the same commutative image. Hence w has the period ρ. □

Lemma 3.12. *Let $w \in A^*$ with $|w| = n$ and $0 < \rho < n$. If w has the periods ρ and $n - \rho$, then $w = u^k$ for some word $u \in A^*$ of length $\gcd(n, \rho)$, and $k = \frac{n}{gcd(n,\rho)} > 1$.*

This lemma is certainly well known but we give a proof for sake of completeness.

Proof. By the lemma of Fine and Wilf (*e.g.* [11] Cor. 12.2.9), since the length of $w \geq \rho + (n - \rho) - 1$, it has the period $\gcd(\rho, n - \rho) = \gcd(\rho, n)$; this latter number is $< n$ and divides n. Hence w is the power of some word u of length $\gcd(n, \rho)$, with exponent $\frac{n}{\gcd(n,\rho)} > 1$. □

Corollary 3.13. *If $w \in A^*$ is Symmetric and has the period ρ with $0 < \rho < |w|$, then w is not primitive.*

Proof. By Lemma 3.11, D_ρ is singleton. Since w is Symmetric, $D_{n-\rho}$ is a singleton too, and by the same lemma, $n - \rho$ is a period of w. By Lemma 3.12, w is not primitive. □

Proof of Theorem 3.10. The theorem follows from Corollary 3.13, and from a theorem of Chuan ([4]; see also Thm. 13.4.2 in [11]): each Sturmian word, which has no nontrivial period, is a Christoffel word. □

4. Nonprimitive case

The nonprimitive case is quite different from the primitive case, since one no more needs the Sturmian hypothesis: one has a complete characterization of Christoffel words.

Theorem 4.1. *Let $w \in \{a,b\}^*$, primitive, and $k \geq 2$. Then w^k is Symmetric if and only if w is a Christoffel word.*

The theorem implies that that the word *aabbaabb* is not Symmetric; contrarily to *aabb*, which is Symmetric as we have seen.

Lemma 4.2. *Let $w \in \mathcal{C}$, and $k \geq 2$. There is a bijection ζ from $\mathcal{F}_{w^k}$ into itself such for any factor f of w^k, $\gamma(f) + \gamma(\zeta(f)) = \gamma(w^k)$ $(*)$.*

Proof. Let $|w| = n$. It is enough to construct a bijection $\zeta : \mathcal{F}_{w^k} \cap A^\ell \to \mathcal{F}_{w^k} \cap A^{kn-\ell}$, satisfying $(*)$, for each $\ell = 0, 1, \ldots, kn$.

1. If $\ell = 0$ or kn, it is evident.

2. Suppose $\ell = 1, \ldots, n-1$. Since $\ell < n$ and $k \geq 2$, the set of factors of length ℓ of w^k is equal to the set of circular factors of length ℓ of w. Each circular factor f of w of length ℓ has by Proposition 3.3 a unique factorization $f = sp$, $s \in \mathcal{S}_w, p \in \mathcal{P}_w$, $|sp| = \ell$; in this case, one has $w = pys$ and we define $\zeta(f) = (ysp)^{k-1}y$. This mapping is well defined and clearly, $(*)$ holds. Concerning injectivity, we note that if $w = pys = p'y's'$, with $|sp| = |s'p'|$, and if $\zeta(sp) = \zeta(s'p')$, then $(ysp)^{k-1}y = (y's'p')^{k-1}y'$, so that $y = y'$ (since they have the same length and are both suffix of the same word), $ysp = y's'p'$, and then $sp = s'p'$. Concerning subjectivity, let g be a factor of w^k of length $kn - \ell$; we have $w^k = pgs$, with $|p| + |s| = \ell$, and since $0 < \ell < n$, $w = pys$; thus $w^k = (pys)^k = p(ysp)^{k-1}ys$, and we deduce that $g = (ysp)^{k-1}y = \zeta(sp)$.

3. Suppose that ℓ is a multiple of n, $\ell = in$, $i = 1, \ldots, k-1$. If f is a factor of w^k of length $\ell = in$, then $f = u^i$, u a conjugate of w, and we define $\zeta(f) = u^{k-i}$. We obtain clearly a bijection from the set of factors of w^k of length $\ell = in$ onto the set of factors of w of length $kn - ki = kn - \ell$. Since $\gamma(u) = \gamma(w)$, $(*)$ holds.

4. Suppose that ℓ is not a multiple of n and that $n < \ell < (k-1)n$. Then $\ell = qn + r$, $1 \leq q \leq n-2$, $1 \leq r \leq n-1$. Note that the factors of w of length ℓ have a unique expression $u^q p$, u a conjugate of w, p a nontrivial proper prefix of u; likewise they have a unique expressions su^q, u a conjugate of w, s a nontrivial proper suffix of u. Define $\zeta(u^q p) = su^{n-1-q}$, with $u = ps$. Then ζ is clearly a bijection and $(*)$ holds.

5. If $n < \ell < kn$, we take the inverse bijection of 2. □

Recall that a *border* of a word w is a nonempty word $\tau \neq w$ such that $\tau \in \mathcal{P}_w \cap \mathcal{S}_w$. The next result is well-known.

Proposition 4.3. *Let $w \in A^*$ and t, $1 \leq t \leq |w|$. Then w has a border of length t if and only if w has the period $|w| - t$.*

We also need the next result.

Proposition 4.4. *([7, 21]; see also [11] Lem. 15.3.2 and Thm. 15.3.1) Let $w \in \{a,b\}^*$ be primitive.*

(i) The word w has at least $i+1$ circular factors of length i for each $i = 0, 1, \ldots, n-1$.

(ii) The word w is the conjugate of some Christoffel word if and only if it has $i+1$ factors of length i for any $i = 0, 1, \ldots, |w| - 1$.

Proof of theorem 4.1. Suppose that w is a Christoffel word. It follows from Lemma 4.2 that w^k is Symmetric.

Conversely, let $n = |w|$ and suppose that w^k is Symmetric. Then by Lemma 2.4, w^k has for each $i = 0, 1, \ldots, n-1$, at most $i+1$ factors of length i. Since each circular factor of w is a factor of w^k, we deduce that w has at most $i+1$ circular factors of length i. Since w is primitive, it follows by Proposition 4.4 (i) that w has exactly $i+1$ factors of length i. Thus, by (ii) in the same proposition, w is the conjugate of some Christoffel word. Hence w is Sturmian. If w is not a Christoffel word, then it has a border by Chuan's theorem, and w^k has the same border, of length $t, 1 \leq t \leq n-1$. Thus, by Proposition 4.3, w^k has the period $kn - t$. Hence, by

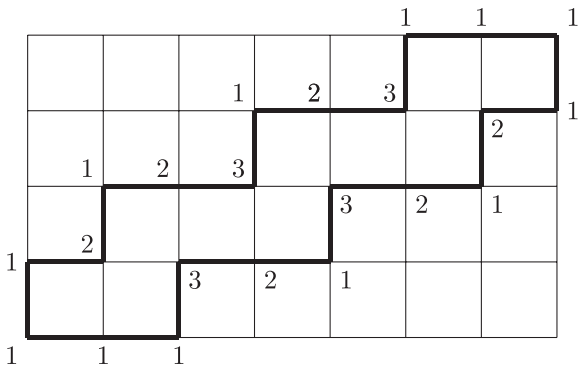

FIGURE 1. Paths of lower and upper Christoffel words $w = aabaabaabab$ and $\tilde{w}$ and the function δ_w.

Lemma 3.11 and Symmetry, w^k has the period t. Thus by Lemma 3.12, w^k is the power of some word of length $\gcd(t, kn)$; since this number is smaller that n, we contradict the primitivity of w. Thus w must be a Christoffel word. □

A particular case of Theorem 4.1 is the following characterization of Christoffel words.

Corollary 4.5. *Let $w \in A^*$. Then w is a Christoffel word if and only if w^2 is Symmetric.*

5. Byproducts

Open question: in the light of Theorems 3.1 and 4.1, it is natural to ask which primitive trapezoidal words, which are non Sturmian, are Symmetric. The work of Flavio D'Alessandro may help, since he characterizes non-Sturmian trapezoidal words (Thm. 5 in [2]). Also, the work of Fici may help, since he shows that each non Sturmian trapezoidal word is *open*, see [22].

In the following, to each word on the alphabet $\{a, b\}$, we associate the path in the discrete plane starting from the origin, where a represents an horizontal step towards East, and b a vertical step towards North.

Proposition 5.1. *Let w be a lower Christoffel word, $\tilde{w}$ the corresponding upper Christoffel word, and E_w the set of integer points on the paths corresponding to w and $\tilde{w}$. Then E_w is the support of the function δ_w.*

See for example Figure 1.

Proof. We have $w = amb$ with m palindrome. Let $d = 1, \ldots, n = |w|$. Denote by π and σ the prefix and the suffix of m of length $d-1$. Since m is a palindrome, $\gamma(\pi)$ and $\gamma(\sigma)$ are equal. Let $(i, j) = \gamma(\pi) = \gamma(\sigma)$. Note that $i + j = d - 1$. By construction $a\pi$, σb are factors of w such that $\gamma(a\pi) = (i+1, j)$ and $\gamma(\sigma b) = (i, j+1)$. Since $w \in \mathcal{C}$, it is balanced; thus, each factor f of w of length $|f| = d$ must satisfy $\gamma(f) = (i+1, j)$ or $\gamma(f) = (i, j+1)$. Since $a\pi$ is prefix of w and $b\tilde{\sigma}$ is a prefix of $\tilde{w}$, the points $\gamma(a\pi)$ and $\gamma(b\tilde{\sigma})$ are on the paths corresponding to w and $\tilde{w}$. All points on the paths are obtained in this way and the result follows. □

The proof of our last result, which is a bijective version of Theorem 3.1, is left to the reader; it is obtained by going carefully through the proof of this theorem.

Proposition 5.2. *Let $w = vu$ be a Christoffel word of length n with its palindromic factorization. Let $k, 0 \leq k \leq n$. Consider all factors of length k of w which intersect the cut of the factorization, and order them from left to right: $f_1, f_2, \ldots, f_r$. Consider all factors of length $n-k$ of w which intersect this cut, and order them from right to left: $g_1, g_2, \ldots, g_s$. Then $r = s$, the words f_i are distinct, the words g_i are distinct, and the mapping $f_i \mapsto g_i$ is a bijection from the set of factors of length k of w to the set of factors of length $n-k$ of w, which complements the Parikh image $\gamma(w)$ of w, that is: $\gamma(f_i) + \gamma(g_i) = \gamma(w)$.*

An example: let $w = aababab$, $v \cdot u = aa \cdot babab$, $k = 4$, $f_1 = aaba, f_2 = abab, f_3 = baba$, $g_1 = bab, g_2 = aba, g_3 = aab$. The bijection is $aaba \mapsto bab, abab \mapsto aba, baba \mapsto aab$.

Acknowledgements

The authors thank Julien Cassaigne for useful mail exchange. Also the anonymous referee for many suggestions and corrections.

Funding

The second author is partially supported by a grant from NSERC, 2020-04685.

Conflicts of interest

There is no conflict of interest.

Data availability statement

No data has been used.

Author contribution statement

The two authors contributed equally to this article.

References

[1] A. de Luca, On the combinatorics of finite words. *Theor. Comput. Sci.* **218** (1999) 13–39.
[2] F. D'Alessandro, A combinatorial problem on trapezoidal words. *Theor. Comput. Sci.* **273** (2002) 11–33.
[3] M. Bucci, A. De Luca and G. Fici, Enumeration and structure of trapezoidal words. *Theor. Comput. Sci.* **468** (2013) 12–22.
[4] W.-F. Chuan, Unbordered factors of the characteristic sequences of irrational numbers. *Theor. Comput. Sci.* **205** (1998) 337–344.
[5] T. Harju and D. Nowotka, Minimal Duval extensions. *Int. J. Found. Comput. Sci.* **15** (2004) 349–354.
[6] J. Berstel and A. de Luca, Sturmian words, Lyndon words and trees. *Theor. Comput. Sci.* **178** (1997) 171–203.
[7] J.-P. Borel and C. Reutenauer, On Christoffel classes. *RAIRO Inform. Théor. Appl.* **40** (2006) 15–27.
[8] W.-F. Chuan, Sturmian morphisms and α-words. *Theor. Comput. Sci.* **225** (1999) 129–148.
[9] S. Mantaci, A. Restivo and M. Sciortino, Burrows–Wheeler transform and Sturmian words. *Inform. Process. Lett.* **86** (2001) 241–246.
[10] S. Mantaci, A. Restivo, G. Romana, G. Rosone and M. Sciortino, A combinatorial view on string attractors. *Theor. Comput. Sci.* **850** (2021) 236–248.
[11] C. Reutenauer, From Christoffel Words to Markoff Numbers. Oxford University Press (2019).
[12] C. Reutenauer, Christoffel words and weak Markoff theory. *Adv. Appl. Math.* **127** (2021) 102179.
[13] D. Kempa, and N. Prezza, At the roots of dictionary compression: string attractors, in *STOC 2018*. ACM (2018) 827–840.
[14] N. Prezza, String attractors. CoRR, arXiv:709.05314 [abs] (2017).
[15] E.B. Christoffel, Observatio arithmetica. *Ann. Mat. Pura Appl.* **6** (1875) 145–152.
[16] A. de Luca and F. Mignosi, Some combinatorial properties of Sturmian words. *Theor. Comput. Sci.* **136** (1994) 361–385.
[17] A. Heinis, On low-complexity bi-infinite words and their factors. *J. Théor. Nombres Bordeaux* **13** (2001) 421–442.
[18] F. Levé and P. Séébold, Proof of a conjecture on word complexity. *Bull. Belg. Math. Soc.* **8** (2001) 277–291.
[19] M.-C. Anisiu and J. Cassaigne, Properties of the complexity function for finite words. *Rev. Anal. Numér. Théor. Approx.* **33** (2004) 123–139.
[20] J. Berstel and C. Reutenauer, Noncommutative Series with Applications. Cambridge University Press (2011).
[21] J. Clough, and G. Myerson, Variety and multiplicity in diatonic systems. *J. Music Theory* **29** (1985) 249–270.
[22] G. Fici, A classification of trapezoidal words. *Proceedings of 8th International Conference Words* (2011).