Algebraic Characterization of Dendricity

France Gheeraert[a]   Herman Goulet-Ouellet[b]   Julien Leroy[c]
Pierre Stas[c]

Abstract

Dendric shift spaces simultaneously generalize codings of regular interval exchanges and episturmian shift spaces, themselves both generalizations of Sturmian words. One of the key properties enforced by dendricity is the Return Theorem. In this paper, we prove its converse, providing the following natural algebraic perspective on dendricity: *A minimal shift space is dendric if and only if every set of return words is a basis of the free group over the alphabet.*

Mathematics Subject Classifications: 37B10, 68R15

1   Introduction

Sturmian words, introduced in [23], are binary sequences well-known for having numerous properties and characterizations (see the surveys in [21] and [13]). Combinatorially, they are the simplest aperiodic words; dynamically, they encode irrational rotations on the circle; geometrically, they approximate lines of irrational slope; and number-theoretically, they are strongly related to continued fractions.

Among the many generalizations of Sturmian words, two are particularly studied: the episturmian shift spaces [16] with a combinatorial definition, and the codings of regular interval exchanges [12] originating in the dynamical approach. Both are examples of the main protagonists of this paper: dendric shift spaces. They are combinatorially defined by restricting the context surrounding finite patterns.

The study of dendric shift spaces begins in 2015 when a group of seven researchers (including the third author of this paper) publish a series of papers in which they consider so-called *tree sets* [3, 4, 5], which are precisely the languages of dendric shift spaces. Dendricity generalizes some of the remarkable combinatorial and dynamical properties of Sturmian words, such as affine factor complexity [3], constant number of return words [3],

---

[a]LAMFA, Université de Picardie Jules Verne, Amiens, France (`france.gheeraert@u-picardie.fr`).

[b]Czech Technical University in Prague, Prague, Czechia (`herman.goulet.ouellet@fit.cvut.cz`).

[c]Mathematics Department, University of Liège, Liège, Belgium (`j.leroy@uliege.be`, `pierre.stas@uliege.be`).

at most $d/2$ ergodic measures where d is the size of the alphabet [10], and the existence of an S-adic characterization [14].

In this paper, we focus on the algebraic aspects of dendric shift spaces, and more specifically on the following property known as the Return Theorem: *In a minimal dendric shift space, every return set forms a basis of the free group* [3]. This result is crucial to understand various aspects of dendric shift spaces, such as their dimension groups [7], S-adic representations [15, 14], Schützenberger groups [1], and stabilizers [6]. Our contribution is a proof that the converse of the Return Theorem also holds, leading to the following characterization.

Theorem 1. *Let X be a minimal shift space over $\mathcal{A}$. The following assertions are equivalent:*

1. *X is dendric;*

2. *for each word in the language of X, its set of return words is a tame basis of the free group $F_{\mathcal{A}}$;*

3. *for each word in the language of X, its set of return words is a basis of the free group $F_{\mathcal{A}}$.*

The statement of Theorem 1 is reminiscent of Vuillon's characterization of Sturmian words stating that an infinite binary word is Sturmian if and only if every factor has *exactly* two return words [24]. Observe that our condition is stronger than Vuillon's on the number of return words. This is unavoidable as cardinality alone is not sufficient to infer dendricity on alphabets with more than 2 letters [2].

Theorem 1 provides a natural algebraic interpretation of dendricity, which should be relevant to establish closure properties of the family of dendric shift spaces. Moreover, Theorem 1 and its proof emphasize once again the interactions between symbolic dynamics, combinatorics on words, and algebra, opening the door to further research at the intersection of these topics.

2   Preliminaries

We briefly recall the main notions used in this paper. For more details, we refer to the following monographs: on combinatorics on word [20]; on the free group [22]; on symbolic dynamics [19].

Let $\mathcal{A}$ be a finite set that we call an *alphabet*, and let $\mathcal{A}^*$ and $F_{\mathcal{A}}$ respectively denote the free monoid (whose elements are called *words*) and the free group generated by $\mathcal{A}$. We let ε denote the neutral element of $\mathcal{A}^*$, i.e., the empty word, and we set $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. We naturally view $\mathcal{A}^*$ as a subset of $F_{\mathcal{A}}$, and thus we may consider the subgroup of $F_{\mathcal{A}}$ generated by a given set of words $W \subseteq \mathcal{A}^*$, which is then denoted $\langle W \rangle$.

We let $\mathcal{A}^{\mathbb{Z}}$ denote the set of two-sided infinite words equipped with the product topology of the discrete topology over $\mathcal{A}$. A *shift space* is a closed subset $X \subseteq \mathcal{A}^{\mathbb{Z}}$ invariant

under the shift map $(x_n)_{n \in \mathbb{Z}} \mapsto (x_{n+1})_{n \in \mathbb{Z}}$. The *language* of a shift space X is the set

$$\mathcal{L}(X) = \{x_i \cdots x_j \mid x \in X, \ i \leqslant j\} \cup \{\varepsilon\}.$$

A shift space is *minimal* (with respect to the inclusion of shift spaces) precisely when $\mathcal{L}(X)$ is *uniformly recurrent*, meaning that for every $u \in \mathcal{L}(X)$, there exists n such that u is a factor (in the free monoid $\mathcal{A}^*$) of every $w \in \mathcal{L}(X) \cap \mathcal{A}^n$.

Given a shift space X and a word $w \in \mathcal{L}(X)$, the set of *(left) return words* to w is

$$\mathcal{R}_X(w) = \{r \in \mathcal{A}^+ \mid rw \in \mathcal{L}(X) \cap w\mathcal{A}^* \setminus \mathcal{A}^+ w\mathcal{A}^+\}.$$

Example 2. Let $\mathcal{A} = \{a, b, c\}$. The *Tribonacci shift space* is the shift space X generated by the monoid morphism $\sigma \colon a \mapsto ab, b \mapsto ac, c \mapsto a$, i.e., its language is the set of factors of $\sigma^n(a)$, $n \geqslant 0$. The return words to aba are the words separating consecutive occurrences of aba in the elements of X. This can be seen in the following prefix of $\sigma^5(a)$:

$$abac|aba|abac|ab|abac.$$

Hence the words ab, aba, and $abac$ are return words to aba. As a matter of fact, they are the only return words (see [18, Corollary 4.5] for example). Moreover, they form a basis of the free group $F_\mathcal{A}$, since

$$a = (ab)^{-1}(aba), \quad b = (aba)^{-1}(ab)^2, \quad c = (aba)^{-1}(abac).$$

For a shift space X and a word $w \in \mathcal{L}(X)$, we consider the sets of *extensions*

$$\begin{aligned}
\mathrm{E}_X^-(w) &= \{a \in \mathcal{A} \mid aw \in \mathcal{L}(X)\}, \\
\mathrm{E}_X^+(w) &= \{b \in \mathcal{A} \mid wb \in \mathcal{L}(X)\}, \\
\mathrm{E}_X(w) &= \{(a, b) \in \mathcal{A} \times \mathcal{A} \mid awb \in \mathcal{L}(X)\}.
\end{aligned}$$

The *extension graph* of $w \in \mathcal{L}(X)$ is the bipartite graph $\mathcal{E}_X(w)$ where the vertex set is the disjoint union of $\mathrm{E}_X^-(w)$ and $\mathrm{E}_X^+(w)$, and there is an edge between $a \in \mathrm{E}_X^-(w)$ and $b \in \mathrm{E}_X^+(w)$ if $(a, b) \in \mathrm{E}_X(w)$.

A word $w \in \mathcal{L}(X)$ is said to be *dendric* if $\mathcal{E}_X(w)$ is a tree. We naturally extend this terminology to a shift space X when it is true for all $w \in \mathcal{L}(X)$.

Example 3. Continuing with the example of the Tribonacci shift space X, the extension graph $\mathcal{E}_X(a)$ is given in below.
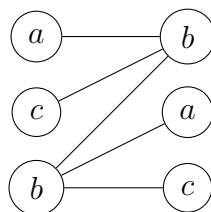
Again, the factors aab, baa, bab, bac, and cab can be seen in $abacabaabacababac$. Observe that $\mathcal{E}_X(a)$ is a tree, so a is dendric. More generally, the Tribonacci shift space is episturmian, and therefore, dendric [3].

3 Proof of the main result

The proof of the converse of the Return Theorem uses techniques and results from symbolic dynamics, group theory, and combinatorics. We start with some results at the intersection between symbolic dynamics and combinatorics.

The first ingredient needed for the proof is the notion of *derived shift space*, which was introduced in [11] to characterize substitutive sequences. Let X be a minimal shift space and $w \in \mathcal{L}(X)$. Let $\mathcal{B}$ be an alphabet with a bijection $\theta_w \colon \mathcal{B} \to \mathcal{R}_X(w)$, which we naturally extend to maps $\theta_w \colon F_{\mathcal{B}} \to F_{\mathcal{A}}$ and $\theta_w \colon \mathcal{B}^{\mathbb{Z}} \to A^{\mathbb{Z}}$ whenever convenient. The *derived shift space* of X with respect to w is the shift space

$$D_w(X) = \theta_w^{-1}(X).$$

We call θ_w the *derivating substitution*. Note that the operation $X \mapsto D_w(X)$ only depends on w, up to a relabeling of the alphabet of $D_w(X)$. It is direct to check that $D_w(X)$ is also minimal. Moreover, the derived shift space of a dendric shift space is again dendric by [4, Theorem 5.13]. Using ideas from the proof of this result, we obtain the following technical lemma.

Lemma 4. *Let X be a minimal shift space and $w \in \mathcal{L}(X)$. Then $\mathcal{E}_X(w)$ is the image of $\mathcal{E}_{D_w(X)}(\varepsilon)$ under a graph morphism.*

Proof. Let $\theta_w \colon \mathcal{B} \to \mathcal{R}_X(w)$ be the derivating substitution used to define $D_w(X)$. We consider the set

$$\mathcal{R}'_X(w) = \{ r' \in \mathcal{A}^+ \mid wr' \in \mathcal{L}(X) \cap \mathcal{A}^* w \setminus \mathcal{A}^+ w \mathcal{A}^+ \}$$

of *right return words* to w and the morphism $\theta'_w \colon \mathcal{B}^* \to \mathcal{A}^*$ defined by $\theta'_w(u) = w^{-1} \theta_w(u) w$ (thus it restricts to a bijection between $\mathcal{B}$ and $\mathcal{R}'_X(w)$). Observe that $z \in \mathcal{L}(D_w(X))$ if and only if $\theta_w(z)w \in \mathcal{L}(X)$, if and only if $w\theta'_w(z) \in \mathcal{L}(X)$.

Consider the map Θ defined on the vertices of $\mathcal{E}_{D_w(X)}(\varepsilon)$ as follows:

- $r \in \mathrm{E}^-_{D_w(X)}(\varepsilon)$ is mapped to the last letter of $\theta_w(r)$;

- $s \in \mathrm{E}^+_{D_w(X)}(\varepsilon)$ is mapped to the first letter of $\theta'_w(s)$.

Assume that $r \in \mathrm{E}^-_{D_w(X)}(\varepsilon)$ and $s \in \mathrm{E}^+_{D_w(X)}(\varepsilon)$ are connected by an edge in $\mathcal{E}_{D_w(X)}(\varepsilon)$; in other words, $rs \in \mathcal{L}(D_w(X))$. It then follows that $\theta_w(rs)w = \theta_w(r)w\theta'_w(s) \in \mathcal{L}(X)$, and thus $\Theta(r)w\Theta(s) \in \mathcal{L}(X)$. This shows that Θ defines a graph morphism from $\mathcal{E}_{D_w(X)}(\varepsilon)$ to a subgraph of $\mathcal{E}_X(w)$.

To show that it is onto, assume that $a \in \mathrm{E}^-_X(w)$ and $b \in \mathrm{E}^+_X(w)$ are connected by an edge in $\mathcal{E}_X(w)$; in other words $awb \in \mathcal{L}(X)$. By uniform recurrence, there exist words u and v such that uw starts with w and ends with aw, wv starts with wb and ends with w, and $uwv \in \mathcal{L}(X)$. Assuming that u and v are of minimal lengths, we get $u \in \mathcal{R}_X(w)$ and $v \in \mathcal{R}'_X(w)$, hence there exist $r, s \in \mathcal{B}$ such that $\theta_w(r) = u$ and $\theta'_w(s) = v$. Moreover, $\theta_w(r)w\theta'_w(s) = \theta_w(rs)w \in \mathcal{L}(X)$ so $rs \in \mathcal{L}(D_w(X))$. Considering r as an element of $\mathrm{E}^-_{D_w(X)}(\varepsilon)$ and s as an element of $\mathrm{E}^+_{D_w(X)}(\varepsilon)$, r and s are connected in $\mathcal{E}_{D_w(X)}(\varepsilon)$ and satisfy $\Theta(r) = a$ and $\Theta(s) = b$, which proves that Θ is onto. $\square$

We will also need the following result.

Lemma 5 ([11, Proposition 2.6]). *Let X be a minimal shift space. If θ_w is the derivating substitution for $w \in \mathcal{L}(X)$ used to define $D_w(X)$, then for every $u \in \mathcal{L}(D_w(X))$*

$$\theta_w(\mathcal{R}_{D_w(X)}(u)) = \mathcal{R}_X(\theta_w(u)w).$$

Proposition 6 below is the key to prove the converse of the Return Theorem. It relies on the previous lemmas and uses the notion of Rauzy graphs. The *order-m Rauzy graph* of a shift space X, denoted $\Gamma_m(X)$, is the directed graph whose vertices are the elements of $\mathcal{L}_m(X)$, and there is an edge from u to v if there are letters a and b such that $ub = av \in \mathcal{L}(X)$; this edge is labeled by a. The label of a path is the concatenation of the labels of the edges.

Proposition 6. *Let X be a minimal shift space and $w \in \mathcal{L}(X)$. If the set $\mathcal{R}_X(w)$ is a basis of $\langle \mathcal{R}_X(w) \rangle$, and if there exists $u \in \mathcal{R}_X(w)w$ for which $\langle \mathcal{R}_X(u) \rangle = \langle \mathcal{R}_X(w) \rangle$, then $\mathcal{E}_X(w)$ is connected.*

Proof. Let $\theta_w \colon \mathcal{B} \to \mathcal{R}_X(w)$ be a derivating substitution for w and let $a \in \mathcal{B}$ be such that $\theta_w(a)w = u$. By Lemma 5, we have $\langle \theta_w(\mathcal{R}_{D_w(X)}(a)) \rangle = \langle \mathcal{R}_X(u) \rangle = \langle \mathcal{R}_X(w) \rangle$. Since $\mathcal{R}_X(w)$ is a basis of $\langle \mathcal{R}_X(w) \rangle$, one can naturally extend θ_w into a group isomorphism from $F_{\mathcal{B}}$ to $\langle \mathcal{R}_X(w) \rangle$. It follows that

$$\langle \mathcal{R}_{D_w(X)}(a) \rangle = \theta_w^{-1} \langle \theta_w(\mathcal{R}_{D_w(X)}(a)) \rangle = F_{\mathcal{B}}.$$

Let us show that $\mathcal{E}_X(w)$ is connected. Let H be the subgroup of $F_{\mathcal{B}}$ generated by the labels of the loops based at the vertex a in $\Gamma_1(D_w(X))$ (this is sometimes called the *Rauzy group* with respect to a, detailed definitions may be found in [17, Section 4]). As return words are particular examples of labels of loops, $\langle \mathcal{R}_{D_w(X)}(a) \rangle$ is a subgroup of H. Hence, $H = F_{\mathcal{B}}$. By [17, Lemma 7.2], this implies that $\mathcal{E}_{D_w(X)}(\varepsilon)$ is connected. As, by Lemma 4, $\mathcal{E}_X(w)$ is the image of $\mathcal{E}_{D_w(X)}(\varepsilon)$ under a graph morphism, this shows that $\mathcal{E}_X(w)$ is connected. $\qquad\square$

To prove the main result, we need one last combinatorial ingredient: The *multiplicity* (or *bilateral order* [2, 9]) of a word $w \in \mathcal{L}(X)$ is the quantity

$$m_X(w) = \# \mathrm{E}_X(w) - \# \mathrm{E}_X^-(w) - \# \mathrm{E}_X^+(w) + 1.$$

A word is *strong* if $m_X(w) > 0$, *neutral* if $m_X(w) = 0$, and *weak* if $m_X(w) < 0$.

Multiplicity is related to the *factor complexity* $p_X(n) = \#(\mathcal{L}(X) \cap \mathcal{A}^n)$ as follows (see [9, Proposition 3.5]): If $s_X(n) = p_X(n+1) - p_X(n)$ and $b_X(n) = s_X(n+1) - s_X(n)$, then

$$b_X(n) = \sum_{w \in \mathcal{L}(X) \cap \mathcal{A}^n} m_X(w). \tag{1}$$

Multiplicity is also related to dendricity. Indeed, if we say that a word $w \in \mathcal{L}(X)$ is *connected* whenever the graph $\mathcal{E}_X(w)$ is connected, then we have the following simple observations.

Lemma 7. *Let X be a shift space and $w \in \mathcal{L}(X)$.*

1. If w is connected, then w is strong or neutral.

2. If w is connected and neutral, then w is dendric.

Proof. This directly follows from the observation that $m_X(w) - 1$ is the difference between the number of edges and the number of vertices in $\mathcal{E}_X(w)$. By a classical graph theory result (see [8, Exercise 2.1.5 and Corollary 2.4.2] for example), if w is connected then $m_X(w) \geqslant 0$, with equality if and only if $\mathcal{E}_X(w)$ is a tree. $\square$

We can now prove the converse of the Return Theorem.

Theorem 8. *Let X be a minimal shift space over $\mathcal{A}$ such that, for all $w \in \mathcal{L}(X)$, $\mathcal{R}_X(w)$ is a basis of the free group $F_\mathcal{A}$. Then X is dendric.*

Proof. We first observe that any $w \in \mathcal{L}(X)$ is connected by Proposition 6. In particular, $\mathcal{L}(X)$ has no weak factor by Lemma 7. Moreover, all the return sets have the same cardinality $\#\mathcal{A}$ by assumption, therefore, the factor complexity of X is given by $p_X(n) = (\#\mathcal{A} - 1)n + 1$ for all n by [2, Theorem 4.5]. It follows that $s_X(n) = \#\mathcal{A} - 1$ and $b_X(n) = 0$. Using again the fact that X has no weak factor, this implies, by Equation (1), that $m_X(w) = 0$ for all $w \in \mathcal{L}(X)$, and so X is dendric by Lemma 7. $\square$

Using [4, Theorem 5.19], we can strengthen the Return Theorem by observing that, for any factor w in a dendric shift space X, $\mathcal{R}_X(w)$ is a tame basis of $F_\mathcal{A}$. A basis is said to be *tame* if it is obtained by applying to $\mathcal{A}$ a composition of permutations and morphisms of the form $\alpha_{a,b}$ and $\tilde{\alpha}_{a,b}$ where, for $a \neq b$ in $\mathcal{A}$,

$$\alpha_{a,b}(c) = \begin{cases} ab, & \text{if } c = a; \\ c, & \text{otherwise} \end{cases} \quad \text{and} \quad \tilde{\alpha}_{a,b}(c) = \begin{cases} ba, & \text{if } c = a; \\ c, & \text{otherwise.} \end{cases}$$

Putting Theorem 8 together with [4, Theorem 5.19] and the Return Theorem, we obtain the claimed algebraic characterization of dendricity (Theorem 1).

References

[1] J. Almeida and A. Costa. A geometric interpretation of the Schützenberger group of a minimal subshift. *Ark. Math.*, 54(2): 243–275, 2016. doi:10.1007/s11512-016-0233-7.

[2] L. Balková, E. Pelantová, and W. Steiner. Sequences with constant number of return words. *Monatsh. Math.*, 155(3-4): 251–263, 2008. doi:10.1007/s00605-008-0001-2.

[3] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176(4): 521–550, 2015. doi:10.1007/s00605-014-0721-4.

[4] V. Berthé, C. D. Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Maximal bifix decoding. *Discrete Math.*, 338(5): 725–742, 2015. `doi:10.1016/j.disc.2014.12.010`.

[5] V. Berthé, C. D. Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. The finite index basis property. *J. Pure Appl. Algebra*, 219(7): 2521–2537, 2015. `doi:10.1016/j.jpaa.2014.09.014`.

[6] V. Berthé, F. Dolce, F. Durand, J. Leroy, and D. Perrin. Rigidity and substitutive dendric words. *Int. J. Found. Comput.*, 29(05): 705–720, 2018. `doi:10.1142/S0129054118420017`.

[7] V. Berthé, P. Cecchi Bernales, F. Durand, J. Leroy, D. Perrin, and S. Petite. On the dimension group of unimodular S-adic subshifts. *Monatsh. Math.*, 194(4): 687–717, 2021. `doi:10.1007/s00605-020-01488-3`.

[8] J. A. Bondy and U. S. R. Murty. *Graph Theory with Applications*. Macmillan London, 1976.

[9] J. Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1): 67–88, 1997. `doi:10.36045/bbms/1105730624`.

[10] M. Damron and J. Fickenscher. The number of ergodic measures for transitive subshifts under the regular bispecial condition. *Ergod. Theory Dyn. Syst.*, 42(1): 86–140, 2020. `doi:10.1017/etds.2020.134`.

[11] F. Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3): 89–101, 1998. `doi:10.1016/S0012-365X(97)00029-0`.

[12] S. Ferenczi. Combinatorial methods for interval exchange transformations. *Southeast Asian Bull. Math.*, 37(1): 47–66, 2013.

[13] N. P. Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*. Springer Berlin Heidelberg, 2002. `doi:10.1007/b13861`.

[14] F. Gheeraert and J. Leroy. $\mathcal{S}$-adic characterization of minimal dendric shifts. *DMTCS*, 22(7), 2024. `doi:10.46298/dmtcs.13130`.

[15] F. Gheeraert, M. Lejeune, and J. Leroy. $\mathcal{S}$-adic characterization of minimal ternary dendric shifts. *Ergod. Theory Dyn. Syst.*, 42(11): 3393–3432, 2022. `doi:10.1017/etds.2021.84`.

[16] A. Glen and J. Justin. Episturmian words: A survey. *RAIRO Theor. Inform. Appl.*, 43(3): 403–442, 2009. `doi:10.1051/ita/2009003`.

[17] H. Goulet-Ouellet. Suffix-connected languages. *Theoret. Comput. Sci.*, 923: 126–143, 2022. `doi:10.1016/j.tcs.2022.05.001`.

[18] J. Justin and L. Vuillon. Return words in Sturmian and episturmian words. *RAIRO Theor. Inform. Appl.*, 34(5): 343–356, 2000. `doi:10.1051/ita:2000121`.

[19] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995. `doi:10.1017/CBO9780511626302`.

[20] M. Lothaire. *Combinatorics on Words*. Cambridge University Press, 1997. `doi:10.1017/CBO9780511566097`.

[21] M. Lothaire. *Algebraic combinatorics on words.* Cambridge University Press, 2002. `doi:10.1017/CBO9781107326019`.

[22] R. C. Lyndon and P. E. Schupp. *Combinatorial Group Theory.* Springer Berlin Heidelberg, 2001. `doi:10.1007/978-3-642-61896-3`.

[23] M. Morse and G. A. Hedlund. Symbolic dynamics II: Sturmian trajectories. *Am. J. Math.*, 62(1): 1–42, 1940. `doi:10.2307/2371431`.

[24] L. Vuillon. A characterization of Sturmian words by return words. *Eur. J. Comb.*, 22(2): 263–275, 2001. `doi:10.1006/eujc.2000.0444`.