Groups, languages and dendric shifts

Dominique Perrin

Université Paris Est, LIGM

Abstract. We present a survey of results obtained on symbolic dynamical systems called dendric shifts. We state and sketch the proofs (sometimes new ones) of the main results obtained on these shifts. This includes the Return Theorem and the Finite Index Basis Theorem which both put in evidence the central role played by free groups in these systems. We also present a series of applications of these results, including some on profinite semigroups and some on dimension groups.

Keywords: symbolic dynamical systems, free groups

1 Introduction

We present here a survey of results obtained in the last years on symbolic dynamical systems. These systems, also called shift spaces, are closely related to combinatorics on words. Familiar notions such as morphisms and words obtained by iterating morphisms (like the Fibonacci or the Thue-Morse word) play an important role in this subject. The aim of the presentation is to give a direct and, hopefully, easy access to some of the results obtained in the last years, as well as giving indications on simplifications in the proofs discovered since the publication of the results.

The main notion introduced is that of dendric shifts, defined by a condition on the possible right and left extensions of a word in the language of the shift. These extensions define a graph for every word in the language of the shift called its extension graph. A dendric shift is by definition such that the extension graph of every word in the language is a tree. This notion defines a natural class of shifts which contains several previously studied classes, such as Sturmian shifts or interval exchange shifts. The language of a dendric shift is called a tree set (this is the term used in the previous papers).

We will insist here on the surprisingly central role played in these systems by free groups. A first result, called the Return Theorem, states that in a dendric minimal shift, the group generated by the return words to every word in the language is the free group on the alphabet of the shift [7] (see below for a definition of return words). A second result, called the Finite Index Basis Theorem [8], implies that the language of a minimal dendric shift on the alphabet A contains a basis of any subgroup of finite index of the free group on A.

We include several kinds of applications of these results. In particular, the two last sections concern applications to substantially different perspectives, the profinite semigroups on the first hand and the dimension groups on the second one. In both cases, the presentation only sketches the notions introduced but can hopefully give an idea of the content.

2 Shift spaces

Let A be a finite alphabet. We denote by A^* the set of finite words on A. The empty word is denoted ε. We denote by $A^{\mathbb{Z}}$ the set of two-sided infinite words on A.

The *shift map* is the transformation σ on $A^{\mathbb{Z}}$ defined for $x = (x_n)_{n\in\mathbb{Z}}$ and $y = (y_n)_{n\in\mathbb{Z}}$ by $y = \sigma(x)$ if for every $n \in \mathbb{Z}$, one has $y_n = x_{n+1}$.

The set $A^{\mathbb{Z}}$ is considered as a topological space for the product topology induced by the discrete topology on A. This topology is also defined by the distance $d(x, y) = 2^{-r(x,y)}$ where

$$r(x, y) = \max\{k \geq 0 \mid x_{[-k,k]} = y_{[-k,k]}\}$$

with $d(x, y) = 0$ for $x = y$.

A *shift space* (or *shift*) on the alphabet A is a subset S of $A^{\mathbb{Z}}$ which is

1. invariant by the shift map, that is, such that $\sigma(S) = S$,
2. closed for the topology of $A^{\mathbb{Z}}$.

The set $A^{\mathbb{Z}}$ itself is a shift space called the *full shift*.

Example 1. The set of two-sided infinite words on the alphabet $A = \{a, b\}$ without factor bb is a shift space called the *golden mean shift*.

Let S be a shift space on the alphabet A. The *language* of S, denoted $L(S)$, is the set of words which appear as factors of the elements of S. It is factorial (that is, contains the factors of its elements) and extendable (that is, for any $w \in L(S)$ there are letters a, b such that $awb \in L(S)$). It is well known and easy to verify that $L(S)$ determines S.

In order to simplify the statements, we always assume that $L(S)$ contains A.

We denote by $L_n(S)$ (resp. $L_{\geq n}(S)$) the set of words of $L(S)$ of length n (resp. at least n).

Example 2. Let $A = \{a, b\}$ and let φ be the morphism from A^* into itself such that $\varphi : a \mapsto ab, b \mapsto a$. One has

$$\begin{aligned} \varphi(a) &= ab \\ \varphi^2(a) &= aba \\ \varphi^3(a) &= abaab \\ &\cdots \end{aligned}$$

The *Fibonacci shift* is the shift space S such that $L(S)$ is formed of the factors of the words $\varphi^n(a)$ for $n \geq 1$. One may easily verify that $bb \notin L(S)$. Thus S is contained in the golden mean shift.

For $w \in L(S)$, a *return word* to w is a word $u \in L(S)$ such that $wu \in L(S)$ and that wu has exactly two occurrences of w, one as a prefix and one as a suffix. We denote by $\mathcal{R}_S(w)$ the set of return words to w in S. Thus a return word indicates the word to be read before the next occurrence of w in a left to right scan. A symmetric notion of *left return word* is obtained replacing wu by uw in the definition.

Example 3. Let S be the golden shift (Example 1). Then $\mathcal{R}_S(a) = \{a, ba\}$ and $\mathcal{R}_S(b) = a^+b$.

A shift space S is said to be *minimal* if it does not contain properly any nonempty subshift. It is classical that S is minimal if and only if $L(S)$ is *uniformly recurrent*, which means that for any $w \in L(S)$ there is an integer $n \geq 1$ such that any word in $L_{\geq n}(S)$ contains w as a factor.

A shift S is *irreducible* if for any $u, v \in L(S)$ there is a $w \in L(S)$ such that $uwv \in L(S)$ (equivalently $L(S)$ is said to be *recurrent*). Any minimal shift is irreducible but the converse is not true.

Example 4. The Golden mean shift S is irreducible. It is not minimal since it contains properly the Fibonacci shift and accordingly $L(S)$ is not uniformly recurrent (since for example no word in a^* has a factor b).

Example 5. The Fibonacci shift is minimal. This can be verified easily checking by induction on n that every long enough word of $L(S)$ contains a factor equal to $\varphi^n(a)$ and thus all factors of $\varphi^n(a)$.

One may verify that an irreducible shift is minimal if and only if the set $\mathcal{R}_S(w)$ of return words on w is finite for every $w \in L(S)$.

Let S be a shift on the alphabet A. For $w \in L(S)$, denote by $L(w)$ (resp. $R(w)$) the set of letters a such that $aw \in L(S)$ (resp. $wa \in L(S)$) and by $E(w)$ the set of pairs $(a, b) \in A \times A$ such that $awb \in L(S)$. The *extension graph* of w is the undirected graph $\mathcal{E}(w)$ with vertices the disjoint union of $L(S)$ and $R(S)$ and edges the elements of $E(S)$.

A shift S is said to be *dendric* (or equivalently $L(S)$ is a *tree set*) if $\mathcal{E}(w)$ is a tree for every $w \in L(S)$.

The family of dendric shifts contains the Arnoux-Rauzy shifts as well as the interval exchange shifts (see [7]).

Example 6. The Fibonacci shift is dendric (see [7]). The graphs $\mathcal{E}(\varepsilon)$ and $\mathcal{E}(a)$ are shown in Figure 1.

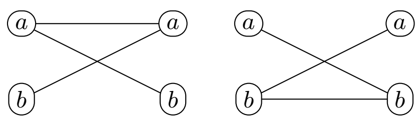

Fig. 1. The graphs $\mathcal{E}(\varepsilon)$ and $\mathcal{E}(a)$.

Let S be a shift space. For $w \in L(S)$, denote

$$\ell(w) = \mathrm{Card}(L(w)), \quad r(w) = \mathrm{Card}(R(w)), \quad e(w) = \mathrm{Card}(E(w))$$

and define the *multiplicity* of w as

$$m(w) = e(w) - \ell(w) - r(w) + 1.$$

The word w is called *neutral* if $m(w) = 0$ and the shift S is said to be neutral if all words in $L(S)$ are neutral. Clearly a dendric shift is neutral but the converse is false.

An important property of neutral shifts is that the numbers $\rho(w) = r(w) - 1$ are left additive, that is, satisfy

$$\rho(w) = \sum_{a \in L(w)} \rho(aw). \tag{1}$$

Thus ρ is almost a left probability, except for its value on the empty word which is $\mathrm{Card}(A) - 1$. This implies in particular the following very useful property. An *S-maximal suffix code* is a suffix code $X \subset L(S)$ which is not properly included in any suffix code $Y \subset L(S)$. Denoting $\rho(X) = \sum_{x \in X} \rho(x)$, we have for any neutral shift S and any S-maximal suffix code $X \subset L(S)$, the equality

$$\rho(X) = \rho(\varepsilon) \tag{2}$$

As a consequence, we have $\rho(L_n(S)) = \mathrm{Card}(A) - 1$ since both sides are equal to $\rho(\varepsilon)$. This implies in turn that for any neutral shift S, the numbers $p_n(S) = \mathrm{Card}(L_n(S))$ satisfy

$$p_n(S) = n(\mathrm{Card}(A) - 1) + 1. \tag{3}$$

Indeed, $p_n(S) = \sum_{w \in L_{n-1}(S)} r(w) = \mathrm{Card}(L_{n-1}(S)) + \rho(L_{n-1}(S))$ whence the result by induction on n. The function $n \mapsto p_n(S)$ is called the *complexity* of S. Thus Equation (3) expresses that a neutral shift has linear complexity.

3 Return Theorem

The following result is from [7]. We denote by $FG(A)$ the free group on A.

Theorem 1. *Let S be a dendric and minimal shift on the alphabet A. For any $w \in L(S)$, the set $\mathcal{R}_S(w)$ is a basis of the free group $FG(A)$.*

Example 7. Let S be the Fibonacci shift. Then $\mathcal{R}_S(aa) = \{baa, babaa\}$ is a basis of the free group on $\{a, b\}$. Set indeed $x = baa$ and $y = babaa$. Then $a = xy^{-1}x$ and $b = yx^{-1}(xy^{-1}x)^{-1}$.

It follows from Theorem 1 that when S is dendric and minimal, the set $\mathcal{R}_S(w)$ has $\mathrm{Card}(A)$ elements for every $w \in L(S)$. Actually, this result (originally proved in [4]) is not properly speaking a corollary because it is an element of the proof of Theorem 1, which can itself be sketched as follows.

Step 1. If S is neutral and minimal, the set $\mathcal{R}_S(w)$ has $\mathrm{Card}(A)$ elements for every $w \in L(S)$. The proof of this claim consists in considering the set Y formed of the proper prefixes of the set $w\mathcal{R}_S(w)$ which are not proper prefixes of w and showing that Y is an S-maximal suffix code. By a well known formula on trees, since $w\mathcal{R}_S(w)$ is a prefix code, we have

$$\mathrm{Card}(w\mathcal{R}_S(w)) = 1 + \rho(Y)$$

On the other hand, since Y is an S-maximal suffix code, we have by Equation (2), $\rho(Y) = \rho(\varepsilon) = \mathrm{Card}(A) - 1$. Thus

$$\begin{aligned}\mathrm{Card}(\mathcal{R}_S(w)) &= \mathrm{Card}(w\mathcal{R}_S(w)) \\ &= 1 + \rho(Y) = \mathrm{Card}(A).\end{aligned}$$

Example 8. Let S be the Fibonacci shift and let $w = aa$. We have $aa\mathcal{R}_S(w) = \{aabaa, aababaa\}$ and $Y = \{aa, aab, aaba, aabab, aababa\}$. The suffix code Y is represented in Figure 2 with the value of $\rho(y)$ indicated on the node y.

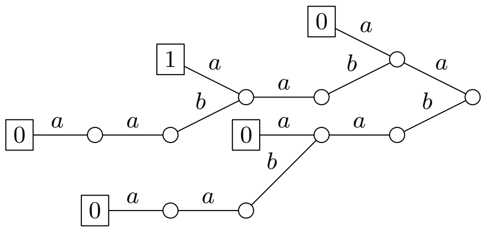

Fig. 2. The set Y.

The next step uses the notion of Rauzy graphs of a shift space. The *Rauzy graph* of order $n \geq 1$, denoted $\Gamma_n(S)$ is the following graph. Its set of vertices is $L_{n-1}(S)$ and its edges are the pairs (x, y) of vertices such that $xa = by$ for some letters a, b satisfying $xa = by \in L_n(S)$. Such an edge is labeled a, allowing to consider $\Gamma_n(S)$ as a deterministic automaton.

Example 9. Let S be the Fibonacci shift. The Rauzy graphs $\Gamma_n(S)$ for $1 \leq n \leq 3$ are shown in Figure 3.

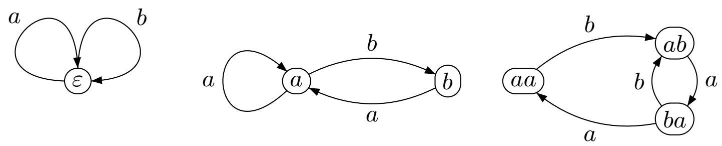

Fig. 3. The Rauzy graphs of order 1, 2, 3.

Step 2. If S is dendric and irreducible, the subgroup recognized by the Rauzy graph $\Gamma_n(S)$ with respect to some vertex considered as initial and terminal state is, for every $n \geq 1$, the free group on A [7, Theorem 4.1]. The proof of this claim uses Stallings foldings to reduce $\Gamma_n(S)$ to $\Gamma_{n-1}(S)$ while recognizing the same subgroup (see Example 10).

Example 10. Consider the Rauzy graph $\Gamma_3(S)$ represented in Figure 3 on the right as an automaton with initial and terminal state aa. Vertices aa and ba are the origin of edges with label b ending in ab. Thus there is a Stallings folding which merges these two vertices. The result is of course $\Gamma_2(S)$.

Step 3. Let n be the maximal length of the words in $w\mathcal{R}_S(w)$ and let $x \in L(S)$ be a word of length n ending with w. The submonoid S recognized by $\Gamma_{n+1}(S)$ with x as initial and terminal state is contained in $\mathcal{R}_S(w)^*$. This implies that the group generated by $\mathcal{R}_S(w)$ contains the group recognized by $\Gamma_{n+1}(S)$, which by Step 2 is the whole free group on A. Since any generating set of the free group on A having Card(A) elements is a basis, and since $\mathcal{R}_S(w)$ has Card(A) elements by Step 1, this implies our conclusion.

4 Finite Index Basis Theorem

A bifix code is a set of words which is both a prefix code and a suffix code. Let S be a shift space. An S-maximal prefix code (resp. bifix code) is a prefix code (resp. bifix code) $X \subset L(S)$ which is not properly contained in any prefix code (resp. bifix code) $Y \subset L(S)$.

For example, for every $n \geq 1$, the set $X = L_n(S)$ is an S-maximal bifix code.

A *parse* of a word w with respect to a bifix code X is a triple (p, x, s) such that $w = pxs$ with p without any suffix in X, $x \in X^*$ and s without any prefix in X. We denote by $d_X(w)$ the number of parses of w. The S-*degree* of a bifix code X, denoted $d_X(S)$ is the maximal value of $d_X(w)$ over the words w in $L(S)$. When X is a finite S-maximal bifix code, the S-degree of X is finite.

For example, if $X = L_n(S)$, then $d_X(S) = n$. More generally, let φ be a morphism from A^* onto a finite group G and let H be a subgroup of index n in G. Then $\varphi^{-1}(H)$ is a submonoid generated by a bifix code Y and $Y \cap L(S)$ is an S-maximal bifix code of S-degree n.

The following result is from [8].

Theorem 2. *Let S be a dendric minimal shift space on the alphabet A and let $d \geq 1$ be an integer. A finite bifix code X is S-maximal of S-degree d if and only if it is a basis of a subgroup of index d of the free group $FG(A)$.*

Note that Theorem 2 implies in particular that for any finite S-maximal bifix code X, one has

$$\mathrm{Card}(X) - 1 = d_X(S)(\mathrm{Card}(A) - 1) \tag{4}$$

Indeed, by Schreier's Formula, the rank $r(H)$ of a subgroup H of index d of a free group of rank k is $r(H) = d(k-1) + 1$. As for return words, this is not properly speaking a Corollary of Theorem 2 since it has to be proved directly and used in the proof.

Note that Equation (4) itself implies Equation (3) taking $X = L_n(S)$ since then $d_X(S) = n$ as we have seen above.

The proof that Formula (4) holds for any finite S-maximal bifix code in an irreducible neutral shift can be obtained in a conceptually simpler way than in [8] as follows.

Step 1. Use the well known equation relating the degrees of interior nodes to the number of leaves in a tree to write $\rho(P) = \mathrm{Card}(X) - 1$ where P is the set of proper prefixes of a finite S-maximal prefix code X viewed as a tree with X as set of leaves.

Step 2. Using the fact [5, Theorem 4.3.7] that if S is recurrent, the set of nonempty proper prefixes of a finite S-maximal bifix code X is a disjoint union of $d_X(S) - 1$ suffix codes, we can write, using Equation (2),

$$\begin{aligned}\mathrm{Card}(X) - 1 = \rho(P) &= (d_X(S) - 1)(\mathrm{Card}(A) - 1) + \rho(\varepsilon)\\ &= d_X(S)(\mathrm{Card}(A) - 1).\end{aligned}$$

A true corollary of Theorem 2 is the following [9, Theorem 5.10].

Corollary 1. *Let H be a subgroup of index n of the free group $FG(A)$. Then for any dendric minimal shift space S, the set $H \cap L(S)$ contains a basis of H.*

Actually, the set X of nonempty words in $H \cap L(S)$ which are not the product of two nonempty words in $H \cap L(S)$ is a bifix code which is a basis of H. Indeed, one may verify that the set X is a finite S-maximal bifix code of degree $m \leq n$. By Theorem 2, it is a basis of a subgroup of index m of $FG(A)$. Since $\langle X \rangle \subset H$, the integer m is a multiple of n. Since on the other hand $m \leq n$, this forces $m = n$ and $\langle X \rangle = H$.

In particular $L_n(S)$ is a basis of the subgroup of index n of the free group $FG(A)$ which is the kernel of the morphism from $FG(A)$ onto $\mathbb{Z}/n\mathbb{Z}$ sending every letter of A on 1.

Example 11. Let S be the Fibonacci shift. The set $L_3(S) = \{aab, aba, baa, bab\}$ is a basis of a subgroup of index 3, namely the kernel of the morphism from $FG(A)$ onto the additive group $\mathbb{Z}/3\mathbb{Z}$ which sends a, b to 1.

Another corollary of Theorem 2 concerns syntactic monoids. Let S be a shift space and let $X \subset L(S)$ be a finite bifix code. Denote by M the syntactic monoid of X^* and let $\varphi : A^* \to M$ be the canonical morphism. Let $J_X(S)$ be the minimum $\mathcal{J}$-class of M that has nonempty intersection with $\varphi(L(S))$. The structure group of $J_X(S)$ is called the *group of* X, denoted $G_X(S)$. This group is naturally represented as a permutation group of degree $d_X(S)$. Indeed, let Q be the set of states of the minimal automaton of X^*. The monoid M is represented by maps from Q into itself and the elements of $J_X(S)$ are precisely the maps with rank $d_X(S)$ (see [5]). Let G be a maximal subgroup of $J_X(S)$. The elements of G act by permutations on their common image, as maps from Q into itself. Thus G is naturally represented as a permutation group of degree $d_X(S)$.

Corollary 2. *Let S be a minimal dendric shift and let $X \subset S$ be an S-maximal bifix code. The group $G_X(S)$ is a transitive permutation group of degree $d_X(S)$ which is the representation of $FG(A)$ on the cosets of the subgroup generated by X.*

This follows indeed from Theorem 2 because it implies that a sequence of Stallings foldings reduces the minimal automaton of X^* to a group automaton with $d_X(S)$ states. We illustrate Corollary 2 by an example.

Example 12. Let S be the Fibonacci shift and let $X = L_2(S)$. The minimal automaton of X^* is represented in Figure 4 on the left. The minimum $\mathcal{J}$-class $J_X(S)$ of $M(X^*)$ is represented in the middle. The group automaton represented on the right is the representation of $FG(A)$ on the cosets of the group generated by X. Accordingly, the group $G_X(S)$ is the group $\mathbb{Z}/2\mathbb{Z}$.

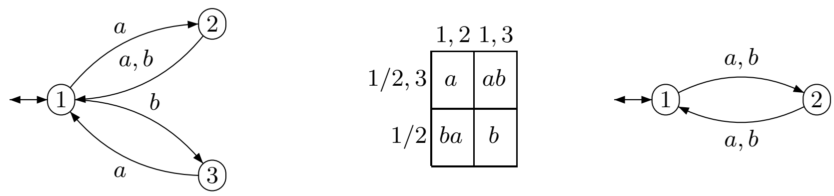

Fig. 4. The minimal automaton of X^*, the minimum $\mathcal{J}$-class $J_X(S)$ and the corresponding group automaton.

5 Bifix decoding

Let X be a prefix code on the alphabet A. A *coding morphism* for X is a morphism from an alphabet B into A^* which defines a bijection from B onto $X = f(B)$.

Let S be a shift on the alphabet A and let X be a finite S-maximal bifix code. Let $f : B \to X$ be a coding morphism for X. We denote by $f^{-1}(S)$ the shift space such that $L(f^{-1}(S)) = f^{-1}(L(S))$, called the *maximal bifix decoding* of S with respect to f.

Example 13. Let S be the Fibonacci shift and let $X = L_2(S) = \{aa, ab, ba\}$. Set $B = \{u, v, w\}$ and $f : u \mapsto aa, v \mapsto ab, w \mapsto ba$. The decoding $T = f^{-1}(S)$ of S by f is the shift obtained by decoding S by non overlapping blocks of length 2.

The following result is from [9].

Theorem 3. *The family of minimal dendric shifts is closed under maximal bifix decoding.*

The proof uses a result interesting in its own concerning the derived shift of a minimal shift space. Let us first recall this notion, introduced in [13]. Let S be a shift space on the alphabet A and let $w \in L(S)$. Let $f : B \to \mathcal{R}_S(w)$ be a coding morphism for the prefix code $\mathcal{R}_S(w)$. Let $\Gamma_S(w) = \{x \in L(S) \mid wx \in L(S) \cap A^+w\} \cup \varepsilon$.

The derived shift of S with respect to f is the shift denoted $f^{-1}(S)$ such that $L(f^{-1}(S))) = f^{-1}(\Gamma_S(w))$. The following is [9, Theorem 5.13].

Theorem 4. *Any derived shift of a minimal dendric shift is a minimal dendric shift.*

This generalizes the well-known fact that the derived set of a Sturmian shift is Sturmian [17].

Example 14. Let S be the Tribonacci shift which is the shift generated by the substitution $\varphi : a \mapsto ab, b \mapsto ac, c \mapsto a$, that is such that $L(S)$ is formed of the factors of the $\varphi^n(a)$ for $n \geq 1$. The set of return words to a is $\mathcal{R}_S(a) = \{a, ba, ca\}$. Let f be the coding morphism $f : a \mapsto a, b \mapsto ba, c \mapsto ca$. The derived shift with respect to f is the image of S under the permutation (abc).

The proof of Theorem 3 itself uses the following steps (which represent a substantial simplification of the proof given in [9]).

Step 1 Prove that the maximal bifix decoding of an irreducible dendric shift is a dendric shift [7, Theorem 3.13].

Step 2 Prove that the maximal bifix decoding of a minimal dendric shift is irreducible [9, Lemma 6.6].

Step 3 Use the fact that any irreducible neutral shift is actually minimal [12, Corollary 5.3] (the direct proof that the maximal bifix decoding of a minimal dendric shift is minimal is one of the hairy points of the proof in [9]).

6 S-adic representation

The following definition is taken from [10, Definition 4.11.1]. Let $\mathcal{S}$ be a set of morphisms and $(\sigma_n)_{n\geq 0}$ be a sequence of morphisms in $\mathcal{S}$ with $\sigma_n : A_{n+1}^* \to A_n^*$ and $A_0 = A$. Let $\Delta = a_0a_1 \cdots$ with $a_i \in A_i$. Assume that

$$u = \lim_{n\to\infty} \sigma_0\sigma_1 \cdots \sigma_{n-1}(a_n)$$

exists and is an infinite word $u \in A^{\mathbb{N}}$. Let S be the shift such that $L(S)$ is the set of factors of u. We call the sequence (σ_n) an $\mathcal{S}$-*adic representation* of S with *directive sequence* Δ.

Thus $\mathcal{S}$-adic representations are a generalization of the construction of shifts by iterating morphisms as shown more precisely below.

Example 15. Let $\varphi : A^* \to A^*$ be a morphism such that $u = \lim \varphi^n(a)$ exists and is an infinite word for some $a \in A$. Then $(\varphi, \varphi, \ldots)$ is an $\mathcal{S}$-adic representation of the shift S generated by u (that is such that $L(S)$ is the set of factors of u) with directive word $aa \cdots$.

This notion is linked to an open question, called the *$\mathcal{S}$-adic conjecture* which aims at relating shifts with a finite $\mathcal{S}$-adic representation with shifts with at most linear complexity (see [15]).

A sequence of morphisms $(\sigma_n)_{n \geq 0}$ is said to be *primitive* if for all $r \geq 0$ there exists $s > r$ such that all letters of A_r occur in all images $\sigma_r \cdots \sigma_{s-1}(a)$, $a \in A_s$. When all morphisms σ_n are equal, one finds the usual notion of primitive morphism.

A morphism $\alpha : A^* \to A^*$ is called *elementary* if it is a permutation of A or it is one of the morphisms $\alpha_{a,b}$, $\tilde{\alpha}_{a,b}$ defined for $a, b \in A$ with $a \neq b$ by

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

Thus $\alpha_{a,b}$ places a letter b after each a and $\tilde{\alpha}_{a,b}$ places a letter b before each a. We denote by $\mathcal{S}_e$ the family of elementary morphisms. The elementary morphisms are automorphisms of the free group which are *positive*, that is, preserving A^*.

A morphism is called *tame* if it belongs to the submonoid generated by the elementary morphisms. Thus the tame morphisms form a finitely generated submonoid of the monoid of positive automorphisms of $FG(A)$ which is not itself finitely generated [18].

The following is [9, Theorem 5.23].

Theorem 5. *Every minimal dendric shift has a primitive $\mathcal{S}_e$-adic representation.*

The converse of Theorem 5 is not true and it is an open problem to characterize minimal dendric shifts by their $\mathcal{S}$-adic representation. Such a characterization exists for Sturmian shifts as recalled below.

Example 16. Any Sturmian shift has an $\mathcal{S}$-adic representation using the particular tame morphisms ψ_a for $a \in A$ which are defined by $\psi_a(a) = a$ and $\psi_a(b) = ab$ for $b \neq a$ (see [3]). Such a representation characterizes Sturmian shifts in the sense that a shift is Sturmian if and only if it has an $\mathcal{S}$-adic representation of the form (ψ_{a_n}) where every letter in A appears infinitely often in the directive word $\Delta = a_0 a_1 \cdots$. For example, the Fibonacci shift has directive word $(ab)^\omega$. Indeed, $\psi_a \psi_b(a) = aba$ and $\psi_a \psi_b(b) = ab$. Thus $\psi_a \psi_b = \varphi^2$ where φ is the Fibonacci morphism.

The proof of Theorem 5 uses the following remarkable consequence of Theorem 2. A positive basis X of the free group $FG(A)$ is tame if there is a tame automorphism α of $FG(A)$ such that $X = \alpha(A)$.

Theorem 6. *Let S be a minimal dendric shift. Any basis of the free group included in $L(S)$ is tame.*

Let indeed X be a basis of $FG(A)$ included in $L(S)$. If X is bifix, by Theorem 2, it is S-maximal of S-degree 1 and thus it is equal to the alphabet A. Otherwise, assuming that X is not suffix, there are words in X of the form uv, v. Then $Y = (X \setminus uv) \cup u$ is again a basis contained in $L(S)$ and an induction shows that Y is tame. Thus $X = \beta(Y)$ where β is an elementary automorphism and thus X is tame.

7 Profinite semigroups

We give below a short introduction to profinite algebra and its links with symbolic dynamics (see [1] for a more detailed presentation).

The *profinite distance* on the free monoid A^* is defined for distinct elements x, y of A^* by $d(x, y) = 2^{-r(x,y)}$ where $r(x, y)$ is the minimal cardinality of a monoid M for which there is a morphism $\varphi : A^* \to M$ with $\varphi(x) \neq \varphi(y)$. The *free profinite monoid* on A, denoted $\widehat{A^*}$, is the completion of the free monoid with respect to the profinite distance. It is a topological monoid which is compact. As in any compact monoid, the closure of the semigroup generated by an element x is compact and contains an idempotent denoted x^ω. Actually one has, in the free profinite monoid, $x^\omega = \lim x^{n!}$.

The elements of $\widehat{A^*}$ are called *pseudowords*, which can be finite words, elements of A^* or *infinite pseudowords*.

Similarly, the *free profinite group*, denoted $\widehat{FG(A)}$ is obtained as the completion of the free group under the profinite distance defined on the free group by replacing finite monoids finite by groups. It is classical that the natural projection $p : \hat{A}^* \to \widehat{FG(A)}$ is surjective. If x a pseudoword x, the inverse of $p(x)$ is $p(x^{\omega-1}) = p(\lim x^{n!-1})$.

It is shown in [1] that for any minimal shift space S, the set of infinite pseudowords in the closure of $L(S)$ is a regular $\mathcal{J}$-class of the monoid $\widehat{A^*}$. This $\mathcal{J}$-class is denoted $J(S)$ and its group is denoted $G(S)$.

The following result is from [2]

Theorem 7. *Let S be a minimal dendric shift. Then $G(S)$ is a free profinite group. More precisely, the restriction to any maximal subgroup of $J(S)$ of the natural projection $\widehat{A^*} \to \widehat{FG(A)}$ is an isomorphism.*

The proof of Theorem 7 uses the Return Theorem (Theorem 1).

There is a close connection between Theorem 7 and the Finite Index Basis Theorem (Theorem 2). Indeed, let $X \subset S$ be a finite maximal bifix code. By Theorem 2, the subgroup $\langle X \rangle$ of $FG(A)$ generated by X has index $d_X(S)$. Let G be a maximal subgroup of $J(S)$. By Theorem 7, the intersection of G with the closure H of $\langle X \rangle$ is a subgroup of index $d_X(S)$. Let φ be the natural morphism from A^* onto the syntactic monoid M of X^* and let $\hat{\varphi}$ be the extension of φ by continuity to a morphism from $\widehat{A^*}$ onto $M(X^*)$. Then $\hat{\varphi}$ maps G onto a maximal subgroup of $J_X(S)$ identified to $G_X(S)$.

$$\begin{array}{ccc} G & \longrightarrow & J(S) \\ \downarrow \hat{\varphi} & & \downarrow \hat{\varphi} \\ G_X(S) & \longrightarrow & J(X) \end{array} \tag{5}$$

The intersection of $G_X(S)$ with the image of X^* in $M(X^*)$ is thus of index $d_X(S)$, which is the content of Corollary 2.

8 Dimension groups

We now turn to a somewhat different perspective involving a notion called the dimension group of a minimal shift space (see [11] for a detailed presentation).

Let S be a minimal shift space and consider the group $C(S,\mathbb{Z})$ of continuous maps from S into $\mathbb{Z}$. Such a map can actually be defined as a map on the words of $L(S)$. Indeed, let ϕ be a function from $L_n(S)$ into $\mathbb{Z}$. The *cylinder function* $\underline{\phi}$ associated to ϕ is defined by $\underline{\phi}(x) = \phi(x_0 \cdots x_{n-1})$. Any continuous map from S into $\mathbb{Z}$ is a composition of a cylinder function with some power of the shift.

For $f \in C(S,\mathbb{Z})$, the *coboundary* of f is the map $\partial f = f - f \circ \sigma$ where σ denotes the shift. The set $\partial C(S,\mathbb{Z})$ of coboundaries is a subgroup of $C(S,\mathbb{Z})$.

We denote by $H(S)$ the quotient $H(S) = C(S,\mathbb{Z})/\partial C(S,\mathbb{Z})$. It is an ordered abelian group with the order induced on the quotient by the order on $C(S,\mathbb{Z})$ defined by $f \leq g$ if $f - g \in C(S,\mathbb{N})$. Let $H^+(S)$ denote the nonnegative elements of $H(S)$, called the *positive cone*, and let $\mathbf{1}$ denote the class of the constant function equal to 1. The triple $(H(S), H^+(S), \mathbf{1})$ is called an *ordered group with unit*.

An ordered group with unit is formed in general of an ordered Abelian group G, its positive cone G^+ and an element $\mathbf{1}_G$ such that for any $g \in G^+$ there is an $n \geq 1$ such that $g < n\mathbf{1}_G$. A morphisms $i : (G, G^+, \mathbf{1}_G) \to (H, H^+, \mathbf{1}_H)$ between ordered groups with unit is a group morphism from G to H such that $i(G^+) \subset H^+$ and $i(\mathbf{1}_G) = \mathbf{1}_H$.

The *dimension group* of a minimal shift S is the ordered group with unit $K^0(S) = (H(S), H^+(S), \mathbf{1})$ where $H^+(S)$ is formed of the nonnegative elements of $H(S)$ and $\mathbf{1}$ is the class of the constant function equal to 1.

Dimension groups are of interest in the study of minimal shifts because they allow a classification of shifts up to an equivalence called *strong orbit equivalence* (see [16]).

Example 17. Let w be a primitive word of length n and let S be the shift formed of the n infinite words with period w. Then $C(S,\mathbb{Z})$ is isomorphic to $\mathbb{Z}^n$ and $H(S)$ is isomorphic to $\mathbb{Z}$. Indeed, the coboundaries are the vectors with a sum of components equal to 0. The order unit is n since it is the image in the quotient of the vector with n components equal to 1. Thus the dimension group is $(\mathbb{Z}, \mathbb{N}, n)$.

Dimension groups can be described using the notion of *direct limit* of ordered groups with unit. We recall briefly the definition. Let $(G_n)_{n\geq 0}$ be a sequence of ordered Abelian groups with unit, with morphisms $i_n : G_n \to G_{n+1}$. The direct limit of this sequence is the quotient of the disjoint union $\cup_{n\geq 0}(G_n, n)$ of the G_n by the equivalence generated by the pairs $((g, n), (i_n(g), n+1))$ for $g \in G_n$.

Example 18. Consider the sequence $\mathbb{Z} \to \mathbb{Z} \to \ldots$ where at each step the morphism is the multiplication by 2. The direct limit of this sequence can be identified via the map $(z, n) \mapsto z/2^n$ with the group of *dyadic rationals* of the form $x/2^n$.

By a theorem of Giordano, Putnam and Skau [16], for any minimal shift S, the dimension group of S is a direct limit of groups $\mathbb{Z}^{k_n}$ with the usual ordering.

We denote by $\mathcal{M}(S)$ the set of invariant probability measures on S. Such a measure μ is determined by its values $\mu([w])$ on the cylinders $[w] = \{x \in A^{\mathbb{Z}} \mid x_0 \cdots x_{n-1} = w\}$ for a word $w \in L_n(S)$. The functions $\pi(w) = \mu([w])$, called *weight functions*, satisfy the left and right additivity rule

$$\pi(w) = \sum_{a\in R(w)} \pi(wa) = \sum_{a\in L(w)} \pi(aw) \tag{6}$$

together with $\pi(\varepsilon) = 1$. For $\mu \in \mathcal{M}(S)$, we denote $f_\mu : A \to \mathbb{Z}$ the function defined by $f_\mu(a) = \mu([a])$. Thus f_μ is the restriction of π to the alphabet.

Example 19. Let S be the Fibonacci shift. There is a unique invariant probability measure on S. The corresponding weight function π is such that $\pi(a) = \lambda$ where $\lambda = (\sqrt{5} - 1)/2$. The other values can be computed using the rules of Equation (6). For example $\pi(ab) = \pi(a) - \pi(aa) = \pi(a) - \pi(a) + \pi(ba) = \pi(ba) = \pi(b) = 1 - \lambda$.

The set $\mathcal{M}(S)$ of invariant probability measures on S is a convex set. Its extremal points are called the *ergodic measures*. We denote by $\mathcal{E}(S)$ the set of ergodic measures on S. By a theorem of Boshernitzan, a neutral shift S on the alphabet A has at most $\mathrm{Card}(A) - 1$ distinct ergodic measures (see [10, Theorem 7.3.2]).

The following result is from [6].

Theorem 8. *Let S be a minimal dendric shift on the alphabet A. Then $H(S) = \mathbb{Z}^A$ with the order defined by*

$$H^+(S) = \{x \in \mathbb{Z}^A \mid \langle f_\mu, x\rangle > 0 \text{ for all } \mu \in \mathcal{E}(S)\} \cup 0$$

The proof of Theorem 8 uses the Return Theorem. Indeed, one can show that the dimension group is related to return words as follows. Fix some $x \in X$ and for $n \geq 1$, let W_n be the set of left return words to $x_0 \cdots x_{n-1}$. Let $G_n(S)$ be the group of maps from W_n into $\mathbb{Z}$ with $G_n^+(S)$ denoting the cone of nonnegative maps in $G_n(S)$ and the order unit being the function which associates to a word in $W_n(S)$ its length. Since $x_0 \cdots x_{n-1}$ is a prefix of $x_0 \cdots x_n$, we have $W_n \subset W_{n+1}^*$ and thus we have a morphism $i_n : G_n(S) \to G_{n+1}(S)$ defined for $\phi \in G_n(S)$ by

$$(i_n\phi)(u) = \sum_{i=1}^{k} \phi(w_i)$$

if $u = w_1 \cdots w_k$. It can then be shown that $H(S)$ is the direct limit of the sequence $(G_n(S))$ with the morphisms i_n (see [14]). When S is dendric, the sets W_n generate the free group and thus the morphisms i_n are isomorphisms. This shows that $H(S) = \mathbb{Z}^A$.

Example 20. Let S be the Fibonacci shift. Then $H(S)$ is isomophic to $\mathbb{Z}^2$. There is a unique invariant probability measure μ (see Example 19) and one has $\mu([a]) = (\sqrt{5} - 1)/2$. Thus, by Theorem 8, $H(S)$ is isomorphic as an ordered group to the group $\mathbb{Z}[(1 + \sqrt{5})/2]$ with the order induced by the reals.

Acknowledgement. The results presented in this paper have been obtained through a cooperation with a group including Valérie Berthé, Paulina Cecchi, Clelia De Felice, Vincent Delecroix, Francesco Dolce, Fabien Durand, Julien Leroy, Samuel Petite, Christophe Reutenauer and Giuseppina Rindone. They are gratefully acknowledged for their help in the preparation.

References

1. Jorge Almeida. Profinite groups associated with weakly primitive substitutions. *Fundam. Prikl. Mat.*, 11(3):13–48, 2005.
2. Jorge Almeida and Alfredo Costa. A geometric interpretation of the Schützenberger group of a minimal subshift. *Ark. Mat.*, 54(2):243–275, 2016.
3. Pierre Arnoux and Gérard Rauzy. Représentation géométrique de suites de complexité $2n + 1$. *Bull. Soc. Math. France*, 119(2):199–215, 1991.
4. Lubomíra Balková, Edita Pelantová, and Wolfgang Steiner. Sequences with constant number of return words. *Monatsh. Math.*, 155(3-4):251–263, 2008.
5. Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.
6. Valérie Berthé, Paulina Cecchi, Francesco Dolce, Fabien Durand, Julien Leroy, Dominique Perrin, and Samuel Petite. Dimension groups of dendric shifts. 2018. preprint.
7. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176(4):521–550, 2015.
8. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. *J. Pure Appl. Algebra*, 219:2521–2537, 2015.
9. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Mathematics*, 338(5):725–742, 2015.
10. Valérie Berthé and Michel Rigo, editors. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.
11. Kenneth R. Davidson. *C^*-algebras by Example*, volume 6 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 1996.
12. Francesco Dolce and Dominique Perrin. Neutral and tree sets of arbitrary characteristic. *Theor. Comput. Sci.*, 658:159–174, 2017.

13. Fabien Durand. A characterization of substitutive sequences using return words. *Discrete Math.*, 179(1-3):89–101, 1998.
14. Fabien Durand, Bernard Host, and Christian Skau. Substitutional dynamical systems, Bratteli diagrams and dimension groups. *Ergodic Theory Dynam. Systems*, 19(4):953–993, 1999.
15. N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
16. Thierry Giordano, Ian F. Putnam, and Christian F. Skau. Topological orbit equivalence and C^*-crossed products. *J. Reine Angew. Math.*, 469:51–111, 1995.
17. Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000.
18. Bo Tan, Zhi-Xiong Wen, and Yiping Zhang. The structure of invertible substitutions on a three-letter alphabet. *Adv. in Appl. Math.*, 32(4):736–753, 2004.