Enumeration Formulæ in Neutral Sets

Francesco Dolce, Dominique Perrin

Enumeration formulæ in neutral sets

Francesco Dolce[1] and Dominique Perrin[1]

Université Paris Est, LIGM

Abstract. We present several enumeration results holding in sets of words called neutral and which satisfy restrictive conditions on the set of possible extensions of nonempty words. These formulae concern return words and bifix codes. They generalize formulae previously known for Sturmian sets or more generally for tree sets. We also give a geometric example of this class of sets, namely the natural coding of some interval exchange transformations.

Keywords: Neutral Sets, Bifix Codes, Interval Exchanges.

1 Introduction

Sets of words of linear complexity play an important role in combinatorics on words and symbolic dynamics. This family of sets includes Sturmian sets, interval exchange sets and primitive morphic sets, that is, sets of factors of fixed points of primitive morphisms.

We study here a family of sets of linear complexity, called neutral sets. They are defined by a property of a graph $E(x)$ associated to each word x, called its extension graph and which expresses the possible extensions of x on both sides by a letter of the alphabet A. A set S is neutral if the Euler characteristic of the graph of any nonempty word is equal to 1, as for a tree. The Euler characteristic of the graph $E(\varepsilon)$ is called the characteristic of S and is denoted $\chi(S)$. These sets were first considered in [1] and in [5]. The factor complexity of a neutral set S on k letters is for $n \neq 1$

$$p_n = n(k - \chi(S)) + \chi(S). \tag{1}$$

We prove here several results concerning neutral sets. The first one (Theorem 2) is a formula giving the cardinality of a finite S-maximal bifix code X of S-degree n in a recurrent neutral set S on k letters as

$$\mathrm{Card}(X) = n(k - \chi(S)) + \chi(S). \tag{2}$$

The remarkable feature is that, for fixed S, the cardinality of X depends only on its S-degree. In the particular case where X is the set of all words of S of length n, we recover Equation (1). Formula (2) generalizes the formula proved in [2] for Sturmian sets and in [7] for neutral sets of characteristic 1.

The second one concerns return words. The set of right first return words to a word x in a factorial set S, denoted $\mathcal{R}_S(x)$, is an important notion. It is the

set of words u such that xu is in S and ends with x for the first time. In several families of sets of linear complexity, the set of first return words to x is known to be of fixed cardinality independent of x. This was proved for Sturmian words in [13], for interval exchange sets in [16] (see also [11]) and for neutral sets of characteristic zero in [1].

We first prove here (Theorem 3) that the set $\mathcal{CR}_S(X)$ of complete first return words to a bifix code X in a uniformly recurrent neutral set S on k letters satisfies $\mathrm{Card}(\mathcal{CR}_S(X)) = \mathrm{Card}(X) + k - \chi(S)$. The remarkable feature here is that, for fixed S, the cardinality of $\mathcal{CR}_S(X)$ depends only on $\mathrm{Card}(X)$. When X is reduced to one element x, we have $\mathcal{CR}_S(x) = x\mathcal{R}_S(x)$ and we recover the result of [1]. When $X = S \cap A^n$, then $\mathcal{CR}_S(X) = S \cap A^{n+1}$. This implies $p_{n+1} = p_n + k - \chi(S)$ and also gives Equation (1) by induction on n. The proofs of these formulæ use a probability distribution naturally defined on a neutral set.

A third result concerns the decoding of a neutral set by a bifix code. We prove that the decoding of any recurrent neutral set S by an S-maximal bifix code is a neutral set. This property is proved for uniformly recurrent tree sets in [8].

We finally prove a result which allows one to obtain a large family of neutral sets of geometric origin, namely using interval exchange transformations. More precisely, we prove that the natural coding of an interval exchange transformation without connections of length ≥ 1 is a neutral set. This extends a result in [6] concerning interval exchange without connections as well as a result of [9] concerning linear involutions without connection.

2 Extension graphs

Let A be a finite alphabet. We denote by A^* the set of all words on A. We denote by ε or 1 the empty word. A *factor* of a word x is a word v such that $x = uvw$. If both u and w are nonempty, we say that x is an *internal factor*. A set of words on the alphabet A is said to be *factorial* if it contains the factors of its elements as well as the alphabet A.

Let S be a factorial set on the alphabet A. For $w \in S$, we denote $L_S(w) = \{a \in A \mid aw \in S\}$, $R_S(w) = \{a \in A \mid wa \in S\}$, $E_S(w) = \{(a,b) \in A \times A \mid awb \in S\}$, and further $\ell_S(w) = \mathrm{Card}(L_S(w))$, $r_S(w) = \mathrm{Card}(R_S(w))$, $e_S(w) = \mathrm{Card}(E_S(w))$.

We omit the subscript S when it is clear from the context. A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *biextendable* if $e(w) > 0$. A factorial set S is called *right-extendable* (resp. *left-extendable*, resp. *biextendable*) if every word in S is right-extendable (resp. left-extendable, resp. biextendable).

A word w is called *right-special* if $r(w) \geq 2$. It is called *left-special* if $\ell(w) \geq 2$. It is called *bispecial* if it is both left-special and right-special. For $w \in S$, we denote
$$m_S(w) = e_S(w) - \ell_S(w) - r_S(w) + 1.$$

2

A word w is called *neutral* if $m_S(w) = 0$. We say that a set S is *neutral* if it is factorial and every nonempty word $w \in S$ is neutral. The *characteristic* of S is the integer $\chi(S) = 1 - m_S(\varepsilon)$.

Thus, a neutral set of characteristic 1 is such that all words (including the empty word) are neutral. This is what is called a neutral set in [5].

The following example of a neutral set is from [5].

Example 1. Let $A = \{a, b, c, d\}$ and let σ be the morphism from A^* into itself defined by $\sigma : a \mapsto ab, \ b \mapsto cda, \ c \mapsto cd, \ d \mapsto abc$. Let S be the set of factors of the infinite word $x = \sigma^\omega(a)$. One has $S \cap A^2 = \{ab, ac, bc, ca, cd, da\}$ and thus $m(\varepsilon) = -1$. It is shown in [5] that every nonempty word is neutral. Thus S is neutral of characteristic 2.

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and for any $u, w \in S$, there is a $v \in S$ such that $uvw \in S$. An infinite factorial set is said to be *uniformly recurrent* if for any word $u \in S$ there is an integer $n \geq 1$ such that u is a factor of any word of S of length n. A uniformly recurrent set is recurrent.

The *factor complexity* of a factorial set S of words on an alphabet A is the sequence $p_n = \mathrm{Card}(S \cap A^n)$. Let $s_n = p_{n+1} - p_n$ and $b_n = s_{n+1} - s_n$ be respectively the first and second order differences sequences of the sequence p_n.

The following result is [12, Proposition 3.5] (see also [10, Theorem 4.5.4]).

Proposition 1. *Let S be a factorial set on the alphabet A. One has $b_n = \sum_{w \in S \cap A^n} m(w)$ and $s_n = \sum_{w \in S \cap A^n} (r(w) - 1)$ for all $n \geq 0$.*

One deduces easily from Proposition 1 the following result which shows that a neutral set has linear complexity.

Proposition 2. *The factor complexity of a neutral set on k letters is given by $p_0 = 1$ and $p_n = n(k - \chi(S)) + \chi(S)$ for every $n \geq 1$.*

Let S be a biextendable set of words. For $w \in S$, we consider the set $E(w)$ as an undirected graph on the set of vertices which is the disjoint union of $L(w)$ and $R(w)$ with edges the pairs $(a, b) \in E(w)$. This graph is called the *extension graph* of w. We sometimes denote $1 \otimes L(w)$ and $R(w) \otimes 1$ the copies of $L(w)$ and $R(w)$ used to define the set of vertices of $E(w)$. We note that since $E(w)$ has $\ell(w) + r(w)$ vertices and $e(w)$ edges, the number $1 - m_S(w)$ is the Euler characteristic of the graph $E(w)$.

A biextendable set S is called a *tree set* of characteristic c if for any nonempty $w \in S$, the graph $E(w)$ is a tree and if $E(\varepsilon)$ is a union of c trees (the definition of tree set in [5] corresponds to a tree set of characteristic 1). Note that a tree set of characteristic c is a neutral set of characteristic c.

Example 2. Let S be the neutral set of Example 1. The graph $E(\varepsilon)$ is represented in Figure 1. It is acyclic with two connected components. It is shown in [5] that the extension graph of any nonempty word is a tree. Thus S is a tree set of characteristic 2.
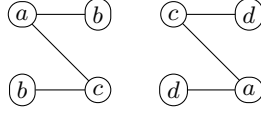
Fig. 1. The two trees forming the graph $E(\varepsilon)$. Vertices correspond to letters, while edges correspond to words of length 2 in S.

Let S be a factorial set. For $x \in S$, we define

$$\rho_S(x) = e_S(x) - \ell_S(x), \quad \lambda_S(x) = e_S(x) - r_S(x).$$

Thus, when x is neutral, $\rho_S(x) = r_S(x) - 1$ and $\lambda_S(x) = \ell_S(x) - 1$. The following result shows that in a biextendable neutral set, ρ_S is a left probability distribution on S (and λ_S is a right probability), except for the value on ε which is $\rho(\varepsilon) = e(\varepsilon) - \ell(\varepsilon) = m(\varepsilon) + r(\varepsilon) - 1 = \mathrm{Card}(A) - \chi(S)$ and can be different of 1 (see [2] for the definition of a right or left probability distribution). We omit the subscript S when it is clear from the context.

Proposition 3. *Let S be a biextendable neutral set. Then for any $x \in S$, one has $\lambda_S(x), \rho_S(x) \geq 0$ and*

$$\sum_{a \in L(x)} \rho_S(ax) = \rho_S(x), \qquad \sum_{a \in R(x)} \lambda_S(xa) = \lambda_S(x).$$

Proof. Since S is biextendable, we have $\ell(x), r(x) \leq e(x)$. Thus $\lambda(x), \rho(x) \geq 0$. Next, $\sum_{a \in L(x)} \rho(ax) = \sum_{a \in L(x)} (r(ax) - 1) = e(x) - \ell(x) = \rho(x)$. The proof for λ is symmetric.

If $\rho(\varepsilon) = 0$, then $\rho(x) = 0$ for all $x \in S$. Otherwise, $\rho'(x) = \rho(x)/\rho(\varepsilon)$ is a left probability distribution. A symmetric result holds for λ.

3 Bifix codes

A prefix code is a set of nonempty words which does not contain any proper prefix of its elements. A suffix code is defined symmetrically. A *bifix code* is a set which is both a prefix code and a suffix code (see [3] for a more detailed introduction). Let S be a recurrent set. A prefix (resp. bifix) code $X \subset S$ is S-maximal if it is not properly contained in a prefix (resp. bifix) code $Y \subset S$. Since S is recurrent, a finite S-maximal bifix code is also an S-maximal prefix code (see [2, Theorem 4.2.2]). For example, for any $n \geq 1$, the set $X = S \cap A^n$ is an S-maximal bifix code.

Given a set X, we denote $\rho(X) = \sum_{x \in X} \rho(x)$. We prove the following result. It accounts for the fact that, in a Sturmian set S, any finite S-maximal suffix code contains exactly one right-special word [2, Proposition 5.1.5].

Proposition 4. *Let S be a neutral set and let X be a finite S-maximal suffix code. Then $\rho(X) = \mathrm{Card}(A) - \chi(S)$.*

Proof. If $\rho(\varepsilon) = 0$, then $\chi(S) = \mathrm{Card}(A)$ and thus the formula holds. Otherwise, ρ' is a left probability distribution (as seen at the end of Section 2), and the formula holds by a well-known property of suffix codes (see [2, Proposition 3.3.4]).

Example 3. Let S be the neutral set of characteristic 2 of Example 1. The set $X = \{a, ac, b, bc, d\}$ is an S-maximal suffix code (its reversal is the $\tilde{S}$-maximal prefix code $\tilde{X} = \{a, b, ca, cb, d\}$). The values of ρ on X are represented in Figure 2 on the left. One has $\rho(X) = \rho(a) + \rho(bc) = 2$, in agreement with Proposition 4.
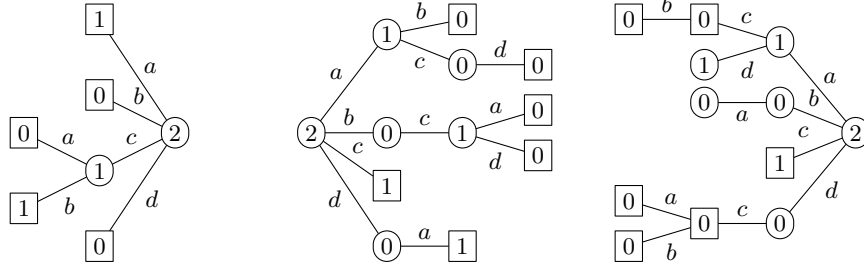

Fig. 2. An S-maximal suffix code (left) and an S-maximal bifix code represented as a prefix code (center) and as a suffix code (right).

Let X be a bifix code. Let Q be the set of words without any suffix in X and let P be the set of words without any prefix in X. A *parse* of a word w with respect to a bifix code X is a triple $(q, x, p) \in Q \times X^* \times P$ such that $w = qxp$. We denote by $d_X(w)$ the number of parses of a word w with respect to X. The S-degree of X, denoted $d_X(S)$ is the maximal number of parses with respect to X of a word of S. For example, the set $X = S \cap A^n$ has S-degree n.

Example 4. Let S be the neutral set of characteristic 2 of Example 1. The set $X = \{ab, acd, bca, bcd, c, da\}$ is an S-maximal bifix code of S-degree 2 (see Figure 2 on the center and the right).

Let S be a recurrent set and let X be a finite bifix code. By [2, Theorem 4.2.8], X is S-maximal if and only if its S-degree is finite. Moreover, in this case, a word $w \in S$ is such that $d_X(w) < d_X(S)$ if and only if it is an internal factor of a word of X. The following is [2, Theorem 4.3.7].

Theorem 1. *Let S be a recurrent set and let X be a finite S-maximal bifix code of S-degree n. The set of nonempty proper prefixes of X is a disjoint union of $n - 1$ S-maximal suffix codes.*

Example 5. Let S and X be as in Example 4. The set of nonempty proper prefixes of X is the S-maximal suffix code represented on the left of Figure 2.

The following statement is closely related with a similar statement concerning the average length of a bifix code, but which requires an invariant probability distribution (see [2, Corollary 4.3.8]).

5

Proposition 5. *Let S be a recurrent neutral set and let X be a finite S-maximal bifix code of S-degree n. The set P of proper prefixes of X satisfies $\rho_S(P) = n(\mathrm{Card}(A) - \chi(S))$.*

Proof. By Theorem 1, we have $P \setminus \{\varepsilon\} = \cup_{i=1}^{n-1} Y_i$, where the Y_i are S-maximal suffix codes. By Proposition 4, we have $\rho(Y_i) = \mathrm{Card}(A) - \chi(S)$ and thus $\rho(P) = \rho(\varepsilon) + (n-1)(\mathrm{Card}(A) - \chi(S)) = n(\mathrm{Card}(A) - \chi(S))$.

4 Cardinality Theorem for bifix codes

The following theorem is a generalization of [7, Theorem 3.6] where it is proved for a neutral set of characteristic 1. We consider a recurrent set S, and we implicitly assume that all words of S are on the alphabet A.

Theorem 2. *Let S be a neutral recurrent set. For any finite S-maximal bifix code X of S-degree n, one has*

$$\mathrm{Card}(X) = n(\mathrm{Card}(A) - \chi(S)) + \chi(S).$$

Note that we recover, as a particular case of Theorem 2 applied to the set X of words of length n in S, the fact that for a set S satisfying the hypotheses of the theorem, the factor complexity is $p_0 = 1$ and $p_n = n(\mathrm{Card}(A) - \chi(S)) + \chi(S)$. Note that Theorem 2 has a converse (see [4]).

Proof (of Theorem 2). Since X is a finite S-maximal bifix code, it is an S-maximal prefix code (see Section 3). By a well-known property of trees, this implies that $\mathrm{Card}(X) = 1 + \sum_{p \in P}(r(p) - 1)$ where P is the set of proper prefixes of X. Since $\rho(p) = r(p) - 1$ for p non empty and $\rho(\varepsilon) = m(\varepsilon) + r(\varepsilon) - 1$, we have

$$\mathrm{Card}(X) = 1 + \sum_{p \in P}(r(p) - 1) = 1 + \sum_{p \in P}\rho(p) - m(\varepsilon)$$
$$= \rho(P) + \chi(S) = n(\mathrm{Card}(A) - \chi(S)) + \chi(S)$$

since $\rho(P) = n(\mathrm{Card}(A) - \chi(S))$ by Proposition 5.

Example 6. Let S be the neutral set of Example 1 and let X be the S-maximal bifix code of Example 4. We have $\mathrm{Card}(X) = 2(4 - 2) + 2 = 6$ according to Theorem 2.

5 Cardinality Theorem for return words

Let S be a factorial set of words. For a set $X \subset S$ of nonempty words, a *complete first return word* to X is a word of S which has a proper prefix in X, a proper suffix in X and no internal factor in X. We denote by $\mathcal{CR}_S(X)$ the set of complete first return words to X. The set $\mathcal{CR}_S(X)$ is a bifix code. If S is uniformly recurrent, $\mathcal{CR}_S(X)$ is finite for any finite set X. For $x \in S$, we denote $\mathcal{CR}_S(x)$ instead of $\mathcal{CR}_S(\{x\})$.

Theorem 3. *Let S be a uniformly recurrent neutral set. For any bifix code $X \subset S$, we have*

$$\mathrm{Card}(\mathcal{CR}_S(X)) = \mathrm{Card}(X) + \mathrm{Card}(A) - \chi(S).$$

Proof. Let P be the set of proper prefixes of $\mathcal{CR}_S(X)$. For $q \in P$, we denote $\alpha(q) = \mathrm{Card}\{a \in A \mid qa \in P \cup \mathcal{CR}_S(X)\} - 1$ and $\alpha(P) = \sum_{q \in P} \alpha(p)$.

Since $\mathcal{CR}_S(X)$ is a finite nonempty prefix code, we have, by a well-known property of trees, $\mathrm{Card}(\mathcal{CR}_S(X)) = 1 + \alpha(P)$.

Let P' be the set of words in P which are proper prefixes of X and let $Y = P \setminus P'$. Since P' is the set of proper prefixes of X, we have $\alpha(P) = \mathrm{Card}(X) - 1$.

Since S is recurrent, any word of S with a prefix in X is comparable for the prefix order with a word of $\mathcal{CR}_S(X)$. This implies that for any $q \in Y$ and any $b \in R_S(q)$, one has $qb \in P \cup \mathcal{CR}_S(X)$. Consequently, we have $\alpha(q) = \rho_S(q)$ for any $q \in Y$. Thus we have shown that

$$\mathrm{Card}(\mathcal{CR}_S(X)) = 1 + \alpha(P') + \rho(Y) = \mathrm{Card}(X) + \rho(Y).$$

Let us show that Y is an S-maximal suffix code. This will imply our conclusion by Proposition 4. Suppose that $q, uq \in Y$ with u nonempty. Since q is in Y, it has a proper prefix in X. But this implies that uq has an internal factor in X, a contradiction. Thus Y is a suffix code. Consider $w \in S$. Since S is recurrent, there is some u and $x \in X$ such that $xuw \in S$. Let y be the shortest suffix of xuw which has a proper prefix in X. Then $y \in Y$. This shows that Y is an S-maximal suffix code. $\qquad\blacksquare$

Let S be a factorial set. A *right first return word* to x in S is a word w such that xw is a word of S which ends with x and has no internal factor equal to x (thus xw is a complete first return word to x). We denote by $\mathcal{R}_S(x)$ the set of right first return words to x in S. Since $\mathcal{CR}_S(x) = x\mathcal{R}_S(x)$, the sets $\mathcal{CR}_S(x)$ and $\mathcal{R}_S(x)$ have the same number of elements. Thus we have the following consequence of Theorem 3.

Corollary 1. *Let S be a uniformly recurrent neutral set. For any $x \in S$, the set $\mathcal{R}_S(x)$ has $\mathrm{Card}(A) - \chi(S) + 1$ elements.*

Example 7. Consider again the neutral set S of Example 1. We have $\mathcal{R}_S(a) = \{bca, bcda, cad\}$.

6   Bifix decoding

Let S be a factorial set and let X be a finite S-maximal bifix code. A *coding morphism* for X is a morphism $f : B^* \to A^*$ which maps bijectively an alphabet B onto X. The set $f^{-1}(S)$ is called a *maximal bifix decoding* of S.

Theorem 4. *Any maximal bifix decoding of a recurrent neutral set is a neutral set with the same characteristic.*

Let S be a factorial set. For two sets of words X, Y and a word $w \in S$, we denote $L_S^X(w) = \{x \in X \mid xw \in S\}$, $R_S^Y(w) = \{y \in Y \mid wy \in S\}$, $E_S^{X,Y}(w) = \{(x, y) \in X \times Y \mid xwy \in S\}$, and further

$$e_S^{X,Y}(w) = \operatorname{Card}(E_S^{X,Y}(w)), \ \ell_S^X(w) = \operatorname{Card}(L_S^X(w)), \ r_S^Y(w) = \operatorname{Card}(R_S^Y(w)).$$

Finally, for a word w, we denote $m_S^{X,Y}(w) = e_S^{X,Y}(w) - \ell_S^X(w) - r_S^Y(w) + 1$. Note that $E_S^{A,A}(w) = E_S(w)$, $m_S^{A,A}(w) = m_S(w)$, and so on.

Proposition 6. *Let S be a neutral set, let X be a finite S-maximal suffix code and let Y be a finite S-maximal prefix code. Then $m_S^{X,Y}(w) = m_S(w)$ for every $w \in S$.*

Proof. We use an induction on the sum of the lengths of the words in X and in Y.

If X, Y contain only words of length 1, since X (resp. Y) is an S-maximal suffix (resp. prefix) code, we have $X = Y = A$ and there is nothing to prove.

Assume next that one of them, say Y, contains words of length at least 2. Let p be a nonempty proper prefix of Y of maximal length. Set $Y' = (Y \setminus pA) \cup p$. If $wp \notin S$, then $m^{X,Y}(w) = m^{X,Y'}(w)$ and the conclusion follows by induction hypothesis. Thus we may assume that $wp \in S$. Then

$$m^{X,Y}(w) - m^{X,Y'}(w) = e^{X,A}(wp) - \ell^X(wp) - r^A(wp) + 1 = m^{X,A}(wp).$$

By induction hypothesis, we have $m^{X,Y'}(w) = m(w)$ and $m^{X,A}(wp) = 0$, whence the conclusion.

Proof (of Theorem 4). Let S be a recurrent neutral set and let $f : B^* \to A^*$ be a coding morphism for a finite S-maximal bifix code X. Set $U = f^{-1}(S)$. Let $v \in U \setminus \{\varepsilon\}$ and let $w = f(v)$. Then $m_U(v) = m_S^{X,X}(w)$. Since S is recurrent, X is an S-maximal suffix code and prefix code. Thus, by Proposition 6, $m_U(v) = m_S(w)$, which implies our conclusion.

The following example shows that the maximal decoding of a uniformly recurrent neutral set need not be recurrent.

Example 8. Let S be the set of factors of the infinite word $(ab)^\omega$. The set $X = \{ab, ba\}$ is a bifix code of S-degree 2. Let $f : u \mapsto ab$, $v \mapsto ba$. The set $f^{-1}(S)$ is the set of factors of $u^\omega \cup v^\omega$ and it is not recurrent.

7 Neutral sets and interval exchanges

Let $I =]\ell, r[$ be a nonempty open interval of the real line and A a finite ordered alphabet. For two intervals Δ, Γ, we denote $\Delta < \Gamma$ if $x < y$ for any $x \in \Delta$ and $y \in \Gamma$. A partition $(I_a)_{a \in A}$ of I (minus $\operatorname{Card}(A) - 1$ points) in open intervals is *ordered* if $a < b$ implies $I_a < I_b$.

We consider now two total orders $<_1$ and $<_2$ on A and two partitions $(I_a)_{a \in A}$ and $(J_a)_{a \in A}$ of I in open intervals ordered respectively by $<_1$ and $<_2$ and such that for every a, I_a and J_a have the same length λ_a. Let $\gamma_a = \sum_{b <_1 a} \lambda_b$ and $\delta_a = \sum_{b <_2 a} \lambda_a$.

An *interval exchange transformation* (with flips) relative to $(I_a)_{a \in A}$ and $(J_a)_{a \in A}$ is a map $T : I \to I$ such that for every $a \in A$, its restriction to I_a is either a translation or a symmetry from I_a to J_a (see for example [6] and [15] for interval exchanges with flips).

Observe that γ_a is the left boundary of I_a and that δ_a is the left boundary of J_a. If $\mathrm{Card}(A) = s$, we say that T is an s-interval exchange transformation.

Example 9. Let $A = \{a, b, c\}$. Consider the rotation of angle α with α irrational as a 3-transformation relative to the partition $(I_a)_{a \in A}$ of the interval $]0, 1[$, where $I_a =]0, 1 - 2\alpha[$, $I_b =]1 - 2\alpha, 1 - \alpha[$ and $I_c =]1 - \alpha, 1[$, while $J_c =]0, \alpha[$, $J_a =]\alpha, 1 - \alpha[$ and $J_b =]1 - \alpha, 1[$ (see Figure 3). Then, for each letter a, the restriction to I_a is a translation to J_a. Note that one has $a <_1 b <_1 c$ and $c <_2 a <_2 b$.
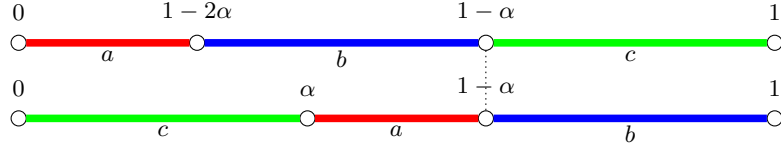

Fig. 3. A 3-interval exchange transformation.

For a word $w = b_0 b_1 \cdots b_m$ let I_w be the set

$$I_w = I_{b_0} \cap T^{-1}(I_{b_1}) \cap \cdots \cap T^{-m}(I_{b_m}).$$

Set $J_w = T^{|w|}(I_w)$. We set by convention $I_\varepsilon = J_\varepsilon =]\ell, r[$. Note that each I_w is an open interval and so is each J_w (see [6]).

Let T be an interval exchange transformation on $I =]\ell, r[$. For a given $z \in I$, the *natural coding* of T relative to z is the infinite word $\Sigma_T(z) = a_0 a_1 \cdots$ on the alphabet A defined by $a_n = a$ if $T^n(z) \in I_a$. We denote by $\mathcal{L}(T)$ the set of factors of the natural codings of T. We also say that $\mathcal{L}(T)$ is the *natural coding* of T. Note that, for every $w \in \mathcal{L}(T)$, the interval I_w is the set of points z such that $\Sigma_T(z)$ starts with w, while the interval J_w is the set of points z such that $\Sigma_T \left(T^{-|w|}(z) \right)$ starts with w. Moreover, it is easy to prove that a word u is in $\mathcal{L}(T)$ if and only if $I_u \neq \emptyset$ (and thus if and only if $J_u \neq \emptyset$).

Example 10. Let T be the interval exchange transformation of Example 9. The first element of $\mathcal{L}(T)$ are represented in Figure 4 (right-special words are colored).

A *connection* of an interval exchange transformation T is a triple (x, y, n) where x is a singularity of T^{-1}, y is a singularity of T, $n \geq 0$ and $T^n(x) = y$. We also say that (x, y, n) is a connection of length n ending in y. When $n = 0$, we say that $x = y$ is a connection.
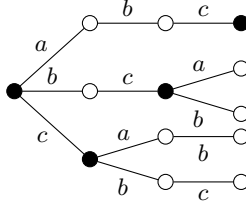
Fig. 4. The words of length ≤ 3 of $\mathcal{L}(T)$.

Interval exchange transformations without connections, also called *regular* interval exchange transformations, are well studied (see, for example, [14] and [6]). The natural coding of a linear involutions without connection (see [9]) is essentially the coding of an interval exchange transformation with exactly one connection of length 0 ending in the midpoint of the interval.

Example 11. Let T be the transformation of Example 9. The point γ_c is a connection of length 0. This connection is represented with a dotted line in Figure 3.

Let T be an interval exchange transformation with exactly c connections all of length 0. Denote $\gamma_{k_0} = \ell$ and $\gamma_{k_1}, \ldots, \gamma_{k_c}$ the c connections of T. For every $0 \leq i < c$ the interval $]\gamma_{k_i}, \gamma_{k_{i+1}}[$ is called a *component* of I.

Example 12. Consider again the transformation T of Example 9. The two components of $]0, 1[$ are the two intervals $]0, 1 - \alpha[$ and $]1 - \alpha, 1[$.

In the next statement we generalize a result of [5] and show that the natural coding of an interval exchange is acyclic.

Theorem 5. *Let T be an interval exchange transformation with exactly c connections, all of length 0. Then $\mathcal{L}(T)$ is neutral of characteristic $c + 1$.*

Lemma 1. *Let T be an interval exchange transformation. For every nonempty word w and letter $a \in A$, one has*

(i) $a \in L(w) \Longleftrightarrow I_w \cap J_a \neq \emptyset$,
(ii) $a \in R(w) \Longleftrightarrow I_a \cap J_w \neq \emptyset$.

Proof. A letter a is in the set $L(w)$ if and only if $aw \in \mathcal{L}(T)$. As we have seen before, this is equivalent to $J_{aw} \neq \emptyset$. One has $J_{aw} = T(I_{aw}) = T(I_a) \cap I_w = J_a \cap I_w$, whence point (i). Point (ii) is proved symmetrically.

We say that a path in a graph is *reduced* if it does not use twice consecutively the same edge.

Lemma 2. *Let T be an interval exchange transformation over I without connection of length ≥ 1. Let $w \in \mathcal{L}(T)$ and $a, b \in L(w)$ (resp. $a, b \in R(w)$). Then $1 \otimes a, 1 \otimes b$ (resp. $a \otimes 1, b \otimes 1$) are in the same connected component of $E(w)$ if and only if J_a, J_b (resp. I_a, I_b) are in the same component of I.*

Proof. Let $a \in L(w)$. Since the set $\mathcal{L}(T)$ is biextendable, there exists a letter c such that $(1 \otimes a, c \otimes 1) \in E(w)$. Using the same reasoning as that in Lemma 1, one has $J_a \cap I_{wc} \neq \emptyset$. Since $I_{wc} \subset I_w$, one has in particular $J_a \cap I_w \neq \emptyset$. This proves that J_a and I_w belong to the same component of I for every $a \in L(w)$.

Conversely, suppose that $a, b \in L(w)$ are such that J_a and J_b belong to the same component of I. We may assume that $a <_2 b$. Then, there is a reduced path $(1 \otimes a_1, b_1 \otimes 1, \ldots, b_{n-1} \otimes 1, 1 \otimes a_n)$ in $E(w)$ (see Figure 5) with $a = a_1$, $b = a_n$, $a_1 <_2 \cdots <_2 a_n$ and $wb_1 <_1 \cdots <_1 wb_{n_1}$. Indeed, by hypothesis, we have no connection of length ≥ 1. Thus, for every $1 \leq i < n$, one has $J_{a_i} \cap I_{wb_i} \neq \emptyset$ and $J_{a_{i+1}} \cap I_{wb_i} \neq \emptyset$. Therefore, a and b are in the same connected component of $E(w)$.

The symmetrical statement is proved similarly.

We can now prove the main result of this section.

Proof (of Theorem 5). Let us first prove that for any $w \in \mathcal{L}(T)$, the graph $E(w)$ is acyclic. Assume that $(1 \otimes a_1, b_1 \otimes 1, \ldots, 1 \otimes a_n, b_n \otimes 1)$ is a reduced path in $E(w)$ with $a_1, \ldots, a_n \in L(w)$ and $b_1, \ldots, b_n \in R(w)$. Suppose that $n \geq 2$ and that $a_1 <_2 a_2$. Then one has $a_1 <_2 \cdots <_2 a_n$ and $wb_1 <_1 \cdots <_1 wb_n$ (see Figure 5). Thus one cannot have an edge (a_1, b_n) in the graph $E(w)$.
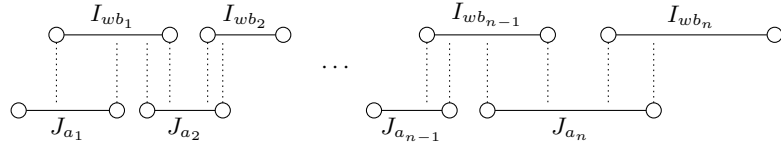

Fig. 5. A path from a_1 to a_n in $E(w)$.

Let us now prove that the extension graph of the empty word is a union of $c + 1$ trees. Let $a, b \in A$. If J_a and J_b are in the same component of I, then $1 \otimes a, 1 \otimes b$ are in the same connected component of $E(\varepsilon)$ by Lemma 2. Thus $E(\varepsilon)$ is a union of $c + 1$ trees.

Finally, if $w \in \mathcal{L}(T)$ is a nonempty word and $a, b \in L(w)$, then J_a and J_b are in the same component of I, by Lemma 1, and thus a and b are in the same connected component of $E(w)$ by Lemma 2. Thus $E(w)$ is a tree.

The previous proof shows actually a stronger result: the set $\mathcal{L}(T)$ is a tree set of characteristic $c + 1$. This result generalizes the corresponding result for regular interval exchange in [5].

Example 13. Let T be the interval exchange transformation of Example 9. In Figure 6 are represented the extension graphs of the empty word (left) and of the letters a (center) and b (right).
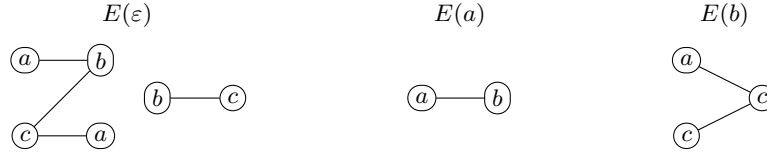
Fig. 6. Some extension graphs.

References

1. Ľubomíra Balková, Edita Pelantová, and Wolfgang Steiner. Sequences with constant number of return words. *Monatsh. Math.*, 155(3-4):251–263, 2008.
2. Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.
3. Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009.
4. Valérie Berthé, Clelia De Felice, Vincent Delecroix, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Specular sets. 2015. In preparation (http://arxiv.org/abs/1505.00707).
5. 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.
6. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and interval exchanges. *J. Pure Appl. Algebra*, 219(7):2781–2798, 2015.
7. 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.
8. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Math.*, 2015.
9. Valérie Berthé, Vincent Delecroix, Francesco Dolce, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Return words of linear involutions and fundamental groups. 2015. Submitted (http://arxiv.org/abs/1405.3529).
10. Valérie Berthé and Michel Rigo. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.* Cambridge Univ. Press, Cambridge, 2010.
11. A. Blondin Massé, S. Brlek, S. Labbé, and L. Vuillon. Palindromic complexity of codings of rotations. *Theoret. Comput. Sci.*, 412(46):6455–6463, 2011.
12. Julien Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1):67–88, 1997. Journées Montoises (Mons, 1994).
13. Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000.
14. Michael Keane. Interval exchange transformations. *Math. Z.*, 141:25–31, 1975.
15. Arnaldo Nogueira, Benito Pires, and Serge Troubetzkoy. Orbit structure of interval exchange transformations with flip. *Nonlinearity*, 26(2):525–537, 2013.
16. Laurent Vuillon. On the number of return words in infinite words constructed by interval exchange transformations. *Pure Math. Appl. (PU.M.A.)*, 18(3-4):345–355, 2007.