Recent results on syntactic groups of codes

Jean Berstel[a], Clelia De Felice[b], Dominique Perrin[a], Christophe Reutenauer[c], Giuseppina Rindone[a]

[a] *Université Paris Est*
[b] *Università degli Studi di Salerno*
[c] *Université du Québec à Montréal*

Contents

1. Introduction

Finite permutation groups can be considered as building blocks of finite transformation monoids. Indeed, by the Krohn-Rhodes Theorem, any finite transformation monoid divides a semidirect product of permutation groups and three element transformation monoids.

Given a transformation monoid M on a finite set, defined by a finite set of generators, it is in general non-trivial to obtain a description of the groups contained in M. For example, there is no simple algorithm to check that all these groups are trivial (see [17]).

This paper, which is of expository nature, contains a new presentation of the techniques that allow to compute the groups contained in a transformation monoid. We use this technique to present a survey on some recent results that allow to describe the groups contained in transformation monoids. These transformation monoids arise as the transition monoid M of the minimal automaton recognizing the submonoid generated by a prefix code X. The maximal groups contained in M are called the syntactic groups of X. We treat two cases. In the first case, we discuss the problem of building bifix codes with the minimal possible number of elements having a given syntactic group. In the second one, we show that, for a particular class of prefix codes, all proper syntactic groups are cyclic.

The first case uses Sturmian words, which are a basic notion of combinatorics on words (see [12]). Using Sturmian words one may prove that any transitive permutation group G of degree d and rank k is a syntactic group of a bifix code with $(k-1)d+1$ elements [2]. We describe this construction here (Theorem 4.2). This result is an improvement of several previous ones due to Schützenberger [16] and to part of the authors of this paper [15].

The second case uses prefix codes such that no element of them is an internal factor of another one, called prefix codes with empty kernel. We describe here the result proved in [1]: if X is a prefix code with empty kernel, all proper syntactic groups of X are cyclic (Theorem 5.1).

The paper is organized as follows.

In Section 2, we give some basic definitions concerning automata, monoids and groups. A more detailed presentation can be found in [3].

In Section 3, we define the holonomy groups of a transformation monoid. This notion is due to Eilenberg [6]. We show how it is related to the classical notion of group in a monoid, as presented for example in [10] or [3]. We recall the notion of Schützenberger representation and show how it may be used to compute generators of the holonomy groups. We also give an alternative method to compute such a set of generators based upon the notion of fundamental group of a graph.

In Section 4, we present a result from [2] according to which any transitive permutation group of degree d which can be generated by k elements is a syntactic group of a bifix code with $(k-1)d+1$ elements (Theorem 4.2). The proof uses Sturmian words. We recall the necessary definitions of Sturmian and episturmian words. We illustrate the construction on several examples.

In Section 5, we define the kernel of a set of words. We present the main result of [1] which states that the proper syntactic groups of a prefix code with empty kernel are regular and cyclic (Theorem 5.1). We give several examples of this situation including ones using the well-known Černý automata.

2. Automata, monoids and groups

Let A be a finite set. We denote by A^* the free monoid on the set A. A deterministic automaton $\mathcal{A} = (Q, i, T)$ on the alphabet A is given by a set Q of states, an initial state $i \in Q$, a set $T \subset Q$ of terminal states and a partial map from $Q \times A$ into Q denoted $(q, a) \mapsto q \cdot a$. An *edge* from p to q labeled a is a triple $(p, a, q) \in Q \times A \times Q$ such that $p \cdot a = q$. A *path* is a sequence $c = (f_1, f_2, \ldots, f_n)$ of consecutive edges $f_j = (q_j, a_j, q_{j+1})$. The word $a_1 a_2 \cdots a_n$ is the *label* of the path.

For a state $p \in Q$ and a word $w \in A^*$, we denote $p \cdot w = q$ if there is a path labeled w from p to the state q and $p \cdot w = \emptyset$ otherwise.

The set *recognized* by the automaton is the set of words $w \in A^*$ such that $i \cdot w \in T$.

When the initial state and the terminal state need not be specified, we also denote an automaton $\mathcal{A} = (Q, E)$ with Q its set of states and E its set of edges.

Example 2.1 The automaton on the left of Figure 2.1 recognizes the words without any occurrence of bb. We use an incoming arrow to denote the initial state and an outgoing arrow to denote the terminal states. Thus, for the automata of Figure 2.1, state 1 is initial and terminal and state 2 is terminal.

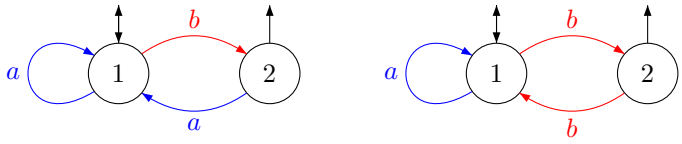

Figure 2.1: Two automata

Example 2.2 The automaton on the right of Figure 2.1 recognizes the words with an even number of b between two consecutive occurrences of a.

An automaton $\mathcal{A}$ is *trim* if for any $q \in Q$, there is a path from i to q and a path from q to some $t \in T$.

An automaton is called *simple* if it is trim and if it has a unique terminal state which coincides with the initial state.

For a set $X \subset A^*$, we denote by $\mathcal{A}(X)$ the *minimal automaton* of X. The states of $\mathcal{A}(X)$ are the nonempty sets $u^{-1}X = \{v \in A^* \mid uv \in X\}$ for $u \in A^*$. The initial state is the set X and the terminal states are the sets $u^{-1}X$ for $u \in X$. Its edges are the triples (p, a, q) such that $p = u^{-1}X$ and $q = (ua)^{-1}X$ for some $u \in A^*$. The minimal automaton $\mathcal{A}(X)$ is trim and recognizes the set X.

Example 2.3 Both automata of Figure 2.2 are simple and minimal. Note that the first automaton is identical with the first one of Figure 2.1 except for the choice of terminal states.

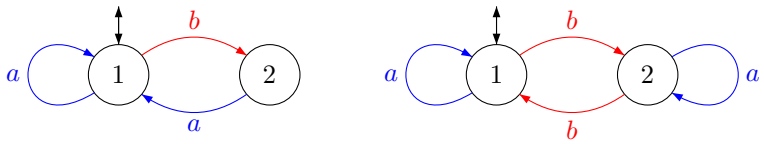

Figure 2.2: Two simple automata

A *prefix code* is a set of words which does not contain a proper prefix of one of its elements. The dual notion of a *suffix code* is defined symmetrically. A *bifix* code is a set of words which is simultaneously a prefix code and a suffix code.

Let $X \subset A^*$ be a prefix code. There is a simple automaton which recognizes the submonoid X^* generated by X, namely the minimal automaton of X^*. Conversely, any set is recognized by a simple automaton is a submonoid generated by a prefix code.

Let $\mathcal{A} = (Q, i, T)$ be an automaton. For $w \in A^*$, we denote $\varphi_{\mathcal{A}}(w)$ the partial map from Q to Q defined by $p\varphi_{\mathcal{A}}(w) = q$ if $p \cdot w = q$. The *transition monoid* of $\mathcal{A}$ is the monoid of partial maps from Q to Q of the form $\varphi_{\mathcal{A}}(w)$ for $w \in A^*$. It is denoted $M(\mathcal{A})$.

Example 2.4 The transition monoid of the automaton on the left of Figure 2.2 has six elements, images of $1, a, b, ab, ba$ and bb respectively.

Example 2.5 The transition monoid of the automaton on the right of Figure 2.2 is the regular representation of the cyclic group $\mathbb{Z}/2\mathbb{Z}$.

The *degree* of a permutation group G on a set R is the cardinality of R. The permutation group G is *transitive* if for any $r, s \in R$ there is some $g \in G$ such that $rg = s$.

A *group automaton* is a simple automaton $\mathcal{A} = (Q, 1, 1)$ on the alphabet A such that the maps $\varphi_{\mathcal{A}}(a)$, for all letters $a \in A$, are permutations of Q. The transition monoid of a group automaton is a permutation group on Q. Its degree is $\mathrm{Card}(Q)$. Since the automaton is trim, the group is transitive.

The following result gives equivalent definition of submonoids recognized by group automata (see [2]).

Proposition 2.6 *The following conditions are equivalent for a submonoid M of A^*.*

(i) *M is recognized by a group automaton with d states.*
(ii) *$M = \varphi^{-1}(H)$, where H is a subgroup of index d of a group G and φ is a surjective morphism from A^* onto G.*
(iii) *$M = H \cap A^*$, where H is a subgroup of index d of the free group on A.*

Let Z be the minimal generating set of a submonoid satisfying one of the conditions of Proposition 2.6. It is a bifix code called a *group code*. The integer d is its *degree*. The transition monoid of the minimal automaton of Z^* is the *group* of Z, denoted $G(Z)$. Note that, since a group automaton is minimal, the group automaton recognizing Z^* is unique (up to the names of the states). Note also that the degree of Z is equal to the degree of the permutation group $G(Z)$ and also to the index of the subgroup generated by Z.

Example 2.7 The set A^d is a group code by condition (ii). It has degree d. The submonoid generated by A^d, is composed of the words with length a multiple of d. The corresponding group automaton for $d = 2$ is represented on the left of Figure 2.3.

Example 2.8 The set $X = a \cup ba^*b$ is a group code of degree 2. The submonoid X^* is formed of the words with an even number of b. It is recognized by the group automaton on the right of Figure 2.3.

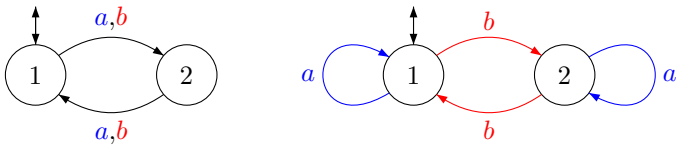

Figure 2.3: Two group automata

3. Holonomy groups

Let Q be a finite set. Recall that a *transformation monoid* on Q is a monoid of partial maps from Q into itself containing the identity on Q. Let M be a transformation monoid on Q. For $I \subset Q$, let

$$\mathrm{Stab}(I) = \{m \in M \mid Im = I\}$$

be the *stabilizer* of I. The restriction of $\mathrm{Stab}(I)$ to I, denoted $\mathrm{Group}(I)$ is a permutation group called the *holonomy group* of M relative to I [6].

Eilenberg uses the term holonomy group in a sligthly different sense. The definition of the holonomy group in [6], is the following (see also [8]). Let $\mathcal{J}$ be the set of subsets of Q which are of the form Qm for $m \in M$ or have cardinality at most one. Let $I \subset Q$ and let $\mathcal{J}(I)$ be the set of $J \in \mathcal{J}$ such that $J \subset I$, $J \neq I$ and which are maximal with respect to this property. Any element of $\mathrm{Stab}(I)$ defines a permutation of the set $\mathcal{J}(I)$. Indeed, for $J \in \mathcal{J}(I)$ and $m \in \mathrm{Stab}(I)$, the set Jm is in $\mathcal{J}(I)$ (it is maximal because m defines a permutation of the elements of $\mathcal{J}$ contained in I). The holonomy group, as defined in [6], is the permutation group on $\mathcal{J}(I)$ obtained in this way. Thus our definition differs in defining the action of the elements of $\mathrm{Stab}(I)$ on I itself instead of the set $\mathcal{J}(I)$.

Groups in monoids.. Let M be a monoid. A *group in* M is a subsemigroup of M which is isomorphic to a group. Note that the neutral element of a group contained in M needs not be equal to the neutral element of M.

A group in M is *maximal* if it not included in another group in M. For any idempotent e in M, there is a unique maximal group contained in M and containing e. It is denoted by $G(e)$. The following property is well-known (see [2]).

Proposition 3.1 *Let G be a group in a transformation monoid M. All elements of G have the same image I. The restriction of the elements of G to I is a faithful representation of G as a permutation group on I.*

Let G be a group in a transformation monoid on Q as above. The *canonical representation* of G is the group of permutations which is formed of the restrictions of the maps in G to their common image. We denote by G_e the canonical representation of the group $G(e)$. By Proposition 3.1, this representation is faithful.

For any set $I \subset Q$, there exists idempotents e of M such that $I \subset Qe$. Indeed, the neutral element of M is such an idempotent. We say that the idempotent

e *covers exactly* the set I if $I \subset Qe$ and Qe is minimal for this property. Note that any idempotent e covers exactly the set Qe.

The following statement shows that for any $I \subset Q$, the group $\mathrm{Group}(I)$ is obtained as the restriction to I of a group in M.

Proposition 3.2 *Let M be a transformation monoid on a finite set Q. For any $I \subset Q$, let e be an idempotent which covers exactly I. Then* $\mathrm{Group}(I)$ *is the restriction to I of the group $G(e) \cap \mathrm{Stab}(I)$.*

Proof. Let $H = G(e) \cap \mathrm{Stab}(I)$. The restriction to I of an element of H is in $\mathrm{Group}(I)$. Conversely, let $m \in \mathrm{Stab}(I)$ and let $g = eme$. Since M is finite, there is an integer n such that $h = g^n$ is idempotent. Since $I \subset Qe$, we have $e \in \mathrm{Stab}(I)$ and thus $g \in \mathrm{Stab}(I)$, which implies that $h \in \mathrm{Stab}(I)$ and $I \subset Qh$. On the other hand, since $h \in Me$, we have $Qh \subset Qe$. By the minimality of Qe, we obtain $Qh = Qe$. This implies $h = e$. Thus the submonoid generated by g is a cyclic group containing e, which implies that g belongs to $G(e)$ and thus to H. Since m and g have the same restriction to I, this shows that any element of $\mathrm{Group}(I)$ is obtained as the restriction to I of an element of H. ∎

The following result shows in particular that the holonomy groups of a monoid M relative to the image of an idempotent are isomorphic with maximal groups contained in M.

Proposition 3.3 *Let M be a transformation monoid on a finite set Q. Let e be an idempotent in M and set $I = Qe$. Then* $\mathrm{Group}(I) = G_e$.

Proof. By Proposition 3.2, and since $G(e) \subset \mathrm{Stab}(I)$, $\mathrm{Group}(I)$ is the restriction to I of the group $G(e)$, which is by definition G_e. ∎

Holonomy groups of automata. The *holonomy groups* of an automaton are the holonomy groups of its transition monoid.

Let $\mathcal{A} = (Q, E)$ be an automaton and let $I \subset Q$. Let $w \in A^*$ be such that $\varphi_{\mathcal{A}}(w) \in \mathrm{Stab}(I)$. The restriction of $\varphi_{\mathcal{A}}(w)$ to I is a permutation which belongs to $\mathrm{Group}(I)$. It is called the permutation of I *defined* by w.

Example 3.4 Let $\mathcal{A}$ be the automaton represented on Figure 3.1. The element

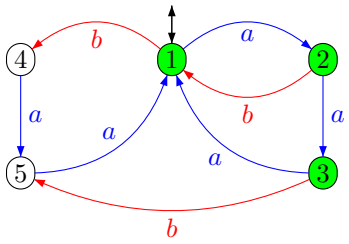

Figure 3.1: An automaton with holonomy group S_3.

$\varphi_{\mathcal{A}}(a^3)$ is idempotent since it is the identity on its image $I = \{1, 2, 3\}$. The action of the letters on the subsets with three elements, represented on Figure 3.2, shows that $\varphi_{\mathcal{A}}(a), \varphi_{\mathcal{A}}(baa) \in \mathrm{Stab}(I)$. Thus, the holonomy group relative to I

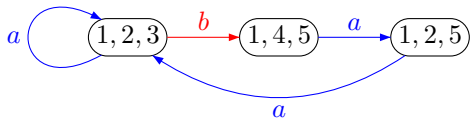

Figure 3.2: Action on the sets with 3 elements

contains the permutations (123) defined by a and (23) defined by baa. Thus it is the symmetric group S_3.

To compute a set of generators of a holonomy group, one may use two different methods. The first one is to compute the appropriate Schützenberger representation.

The Schützenberger representation. We recall here a description of the Schützenberger representation of a transformation monoid. For a more detailed description, see [10] or [3] (where the more general case of a monoid of unambiguous relations is treated).

Let M be a transformation monoid on a finite set Q. Let e be an idempotent of M and let $I = Qe$. Let $\mathcal{I}$ be the set of subsets J of Q such that $Im = J$ and $Jn = I$ for some $m, n \in M$. For each $J \in \mathcal{I}$ let $m_J, m'_J \in M$ be such that

$$Im_J = J, \quad Jm'_J = I$$

with the normalization conditions $m_I = m'_I = e$ and

$$em_Jm'_J = e \tag{3.1}$$

for all $J \in \mathcal{I}$. It is always possible to choose the m_J, m'_J in such a way that condition (3.1) is satisfied since $em_Jm'_J$ belongs to the group in M containing e.

Note that for any $J \in \mathcal{I}$, the restriction of m'_Jm_J to J is the identity on J. Indeed, let $j \in J$. Since $J = Im_J$ there is an $i \in I$ such that $j = im_J$. Then, using (3.1), we have

$$jm'_Jm_J = im_Jm'_Jm_J = iem_Jm'_Jm_J = iem_J = im_J = j.$$

For $J, K \in \mathcal{I}$ and $m \in M$ such that $Jm = K$, we define an element (J, m, K) of $\mathrm{Group}(I)$ as the restriction to I of $m_Jmm'_K$. It is straightforward to verify that if $Jm = K$ and $Kn = L$, then $(J, mn, L) = (J, m, K)(K, n, L)$. Indeed, for any $i \in I$, we have

$$i(J, m, K)(K, n, L) = im_Jmm'_Km_Knm'_L = im_Jmnm'_L = i(J, mn, L)$$

since $im_Jm \in K$ and m'_Km_K is the identity on K.

Thus, we have the following result (see Proposition 9.2.1 in [3]).

Proposition 3.5 *Let S be a set of generators of M. The permutations (J, m, K) for $m \in S$ and $J, K \in \mathcal{I}$ with $Jm = K$ form a set of generators of the group* Group(I).

The *Schützenberger representation* of M relative to the idempotent e is the map which assigns to $m \in M$ the $\mathcal{I} \times \mathcal{I}$ matrix with elements in Group$(I) \cup 0$ defined by

$$\mu(m)_{J,K} = \begin{cases} (J, m, K) & \text{if } Jm = K \\ 0 & \text{otherwise} \end{cases}$$

When M is the monoid of transitions of an automaton $\mathcal{A}$, the Schützenberger representation is defined by a transducer as in the following example.

Example 3.6 Let $\mathcal{A}$ be the automaton of Example 3.4. Set $\varphi = \varphi_{\mathcal{A}}$ and $M = \varphi(A^*)$. Let e be the idempotent $\varphi(a^3)$. The set $\mathcal{I}$ is composed of $I = \{1, 2, 3\}$, $J = \{1, 4, 5\}$ and $K = \{1, 2, 5\}$. We choose $m_J = \varphi(b)$, $m'_J = \varphi(a^2ba^2)$, $m_K = \varphi(ba)$, $m'_K = \varphi(aba^2)$.

The Schützenberger representation of M relative to e is the transducer of Figure 3.3.

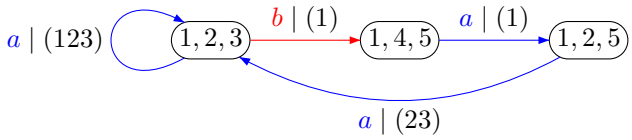

Figure 3.3: The Schützenberger representation

An alternative method. An alternative method can be used to compute a set of generators of the holonomy groups of an automaton.

Let $\mathcal{A} = (Q, E)$ be an automaton on a finite set Q of states. Let M be the transition monoid of $\mathcal{A}$. Let e be an idempotent of M and let $I = Qe$. Let $\mathcal{I}$ be, as above, the set of subsets J of Q such that $Im = J$ and $Jn = I$ for some $m, n \in M$. Let $\mathcal{G}$ be the graph with $\mathcal{I}$ as set of vertices and the triples $(J, a, K) \in \mathcal{I} \times A \times \mathcal{I}$ such that $J \cdot a = K$ as set of edges. We consider (J, a, K) as an edge from the vertex J to the vertex K with label a.

We also consider in the graph $\mathcal{G}$ inverses of the edges. If $f = (J, a, K)$, then f^{-1} is an edge from K to J labeled a^{-1}. A *generalized path* in $\mathcal{G}$ is a sequence of consecutive elements from the set of edges and their inverses. Its label is the element of the free group on A obtained by concatenating the labels of its elements. The iterated simplification of consecutive edges which are mutually inverse generates an equivalence on the set of generalized paths. The set of classes of generalized paths from I to I forms a group denoted by H. The group H is called the *fundamental group* at the point I (see [13] for example). It is well-known that H is a free group and that a basis of H can be obtained as follows. Let $T \subset E$ be a spanning tree with root I of the graph $\mathcal{G}$. Since $\mathcal{G}$ is strongly connected, there is for every J in $\mathcal{I}$ a unique path p_J in T from I to J.

Then the set $X = \{p_J f p_K^{-1} \mid f = (J, a, K) \in E \setminus T\}$ is a basis of H called the *Schreier basis* relative to T (see [13]).

Let $S \subset E$ be a set of edges such that S^{-1} is a spanning tree with root I of the reversal of the graph $\mathcal{G}$. In this way, for each vertex J of $\mathcal{G}$, there is a unique path q_J from J to I using edges in S.

The following result shows that each choice of the pair T, S gives a set of generators of the holonomy group Group(I).

Proposition 3.7 *The set $Y = \{p_J f q_K \mid f = (J, a, K) \in E \setminus T\}$ is a basis of H. The restriction to I of the labels of the elements of Y generates* Group(I).

Proof. We first observe that for any $f = (J, a, K) \in E$, we have $p_J f q_K \in Y$. It is true by definition if $f \notin T$. Otherwise, we use an induction on the length of q_K. Since T is a tree, we cannot have $K = I$ and thus q_K is not empty. Let $g = (K, b, L)$ be the first edge of q_K. Since the set of paths in S contains the suffixes of its elements, we have $q_K = g q_L$. Similarly, since T is a tree, $p_K = p_J f$. We have $p_K g q_L \in Y$ by induction hypothesis. Since $p_J f q_K = p_K g q_L$, we have proved the claim.

Since $Y \subset H = \langle X \rangle$, it is enough to prove that $X \subset \langle Y \rangle$. We prove that, for $f = (J, a, K) \in E \setminus T$, we have $p_J f p_K^{-1} \in \langle Y \rangle$. It is true if q_K is empty since then $p_J f$ is in $X \cap Y$. Otherwise, let $g = (K, b, L) \in E$ be the first edge of q_K. Since the set of paths in S contains the suffixes of its elements, we have $q_K = g q_L$. Then $p_J f q_K = (p_J f p_K^{-1})(p_K g q_L)$ and thus $p_J f p_K^{-1} = p_J f q_K (p_K g q_L)^{-1}$. We have $p_J f q_K \in Y$ by definition and $p_K g q_L \in Y$ by the preliminary remark. Thus $p_J f p_K^{-1} \in \langle Y \rangle$.

Finally, the labels of the elements of Y form the submonoid Stab(I) and thus their restrictions to I form the group Group(I). ∎

Note that, in the definition of Y, every element appears only once. Indeed, assume that $p_J f q_K = p_{J'} f' q_{K'}$ for two edges $f = (J, a, K)$ and $f' = (J', a', K')$ in $E \setminus T$. Suppose that p_J is a prefix of $p_{J'}$. If it is a proper prefix, then f is an edge of the path $p_{J'}$ and thus is in T, a contradiction. Thus $p_J = p_{J'}$ and $f = f'$.

Example 3.8 Consider again the automaton of Example 3.4 and the idempotent $e = \varphi(a^3)$ as in Example 3.6. Set also $I = \{1, 2, 3\}$. The group H is generated by the loop (I, a, I) and the cycle (I, baa, I). Thus Group(I) is generated by $\alpha(a) = (123)$ and $\alpha(baa) = (23)$ where $\alpha(w)$ denotes for $w \in A^*$ the restriction to I of $\varphi(w)$.

Example 3.9 Let $\mathcal{A}$ be the automaton represented in Table 3.1. Set $\varphi = \varphi_{\mathcal{A}}$ and let $e = \varphi(a^5)$. The action of A on the sets with 5 elements is represented in Figure 3.4. Set $I = \{1, 2, 3, 4, 5\}$ and $J = \{1, 2, 6, 7, 8\}$. Set $f = (I, a, I)$, $g = (I, b, J)$, $h = (J, b, J)$ and $k = (J, a, I)$. Let us choose the tree T reduced to the edge g and the set S reduced to the edge k.

	1	2	3	4	5	6	7	8
a	2	3	4	5	1	4	1	5
b	2	6	8	7	1	7	8	1

Table 3.1: The transitions of the automaton $\mathcal{A}$

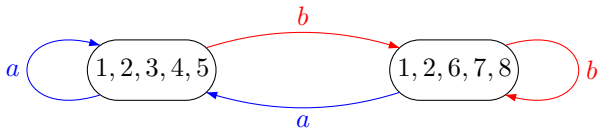

Figure 3.4: The action on 5 element subsets

The group H is generated by the finite set $Y = \{f, gk, ghk\}$. Thus the holonomy group relative to I is generated by

$$\alpha(a) = (12345), \quad \alpha(ba) = (13524), \quad \alpha(bba) = (14532).$$

Thus it is included in the alternating group A_5. Actually, we have $\alpha(ba) = \alpha(a)^2$ and the group generated by $\alpha(a)$ and $\alpha(bba)$ is equal to A_5 (this example appears in [3] as Example 11.7.5). The Schützenberger representation (with the choice of $m_J = \varphi(b)$ and $m'_J = \varphi(a(ba)^4)$) is given on Figure 3.5.

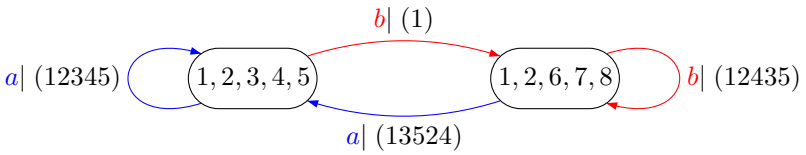

Figure 3.5: The Schützenberger representation

4. The degree of syntactic groups

Let X be a prefix code and let $\mathcal{A}$ be the minimal automaton of X^*. A *syntactic group* of X is a holonomy group of $\mathcal{A}$ relative to the image of an idempotent of $M(\mathcal{A})$. Thus, by Proposition 3.3, a syntactic group of X has the form $\mathrm{Group}(I)$ for a set $I = Qe$, which is the image of an idempotent e in the monoid $M(\mathcal{A})$.

Recall some terminology concerning permutation groups. Let G be a permutation group G on a set S. The *order* of G is its cardinality and its *degree* is the cardinality of the set S. The group G is said to be *transitive* if for any $r, s \in S$ there is an element g of G such that $rg = s$.

Let us call *minimal rank* of a group G the minimal cardinality of a generating set for G. We will discuss the following conjecture [14].

Conjecture 1 (Rank Conjecture) *Let X be a finite bifix code and let G be a transitive permutation group of degree d and minimal rank k. If G is a syntactic group of X, then* $\mathrm{Card}(X) \geq (k-1)d + 1$.

The inequality is related to Schreier's Formula. Indeed, if X is a basis of a subgroup of index d in a free group on k generators, then by Schreier's Formula, $\mathrm{Card}(X) = (k-1)d + 1$.

The degree of nonspecial groups. Let X be a prefix code and let $\mathcal{A} = \mathcal{A}(X^*)$. A syntactic group G of X is called *special* if $\varphi_{\mathcal{A}}^{-1}(G)$ is a cyclic submonoid. In particular a special syntactic group is cyclic.

The following result is from [15].

Theorem 4.1 *Let G be a permutation group of degree d. If G is a nonspecial syntactic group of a prefix code X, then* $\mathrm{Card}(X) \geq d + 1$.

This shows that the conjecture is true for groups of minimal rank 2. The theorem is clearly not true for special syntactic groups since $\mathbb{Z}/n\mathbb{Z}$ is a syntactic group of $X = a^n$ for any $n \geq 1$.

This theorem was proved before by Schützenberger [16] with a weaker bound ($\mathrm{Card}(X) \geq d$) but with a more general hypothesis (an arbitrary set X of words instead of a prefix code).

The general idea is that some parameters in the transition monoid of the minimal automaton of X^* (such as the degrees of the holonomy groups) can be bounded in terms of $\mathrm{Card}(X)$ only, instead of the sum of the lengths of the words of X.

The proof of Schützenberger uses the Critical Factorization Theorem (see [11]).

The following result (from [2]) shows that the bound in the Rank Conjecture can be reached for any transitive permutation group.

Theorem 4.2 *Any transitive permutation group of degree d which can be generated by k elements is a syntactic group of a bifix code with $(k-1)d+1$ elements.*

We first give two examples illustrating the result. We will see afterwards the general construction.

Example 4.3 Let $X = \{aaa, aaba, ab, baa\}$. The minimal automaton of X^* is the automaton of Figure 3.1.

The holonomy group relative to $\{1, 2, 3\}$ is the symmetric group S_3 generated by (123) defined by a and (23) defined by baa (see Example 3.4). Such a construction can be used to realize any group generated by a d-cycle α and another permutation β using $X = a^d \cup \{a^i b a^{d-(i+1)\beta} \mid 0 \leq i \leq d-1\}$.

Example 4.4 Let X be the bifix code with 5 elements represented below.

The action on the sets of states with four elements is shown in Figure 4.2 below.

The word ba defines the permutation (18)(24) and the word aba the permutation (14)(28). Thus the regular representation of $(\mathbb{Z}/2\mathbb{Z})^2$ is a syntactic group of this code.

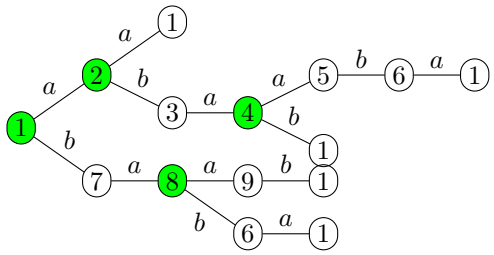

Figure 4.1: An automaton

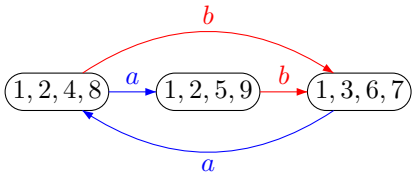

Figure 4.2: The action on four-element subsets

The proof of Theorem 4.2 relies on the notion of Sturmian and episturmian word that we describe now.

Episturmian words. Given a set F of words over an alphabet A, the right (resp. left) order of a word u in F is the number of letters a such that $ua \in F$ (resp. $au \in F$). A word u is *right-special* (resp. *left-special*) if its right order (resp. left order) is at least 2. A right-special (resp. left-special) word is *strict* if its right (resp. left) order is equal to $\mathrm{Card}(A)$. In the case of a 2-letter alphabet, all special words are strict.

For an infinite word x, we denote by $F(x)$ the set of factors of x. By definition, an infinite word x is *episturmian* if $F(x)$ is closed under reversal and if $F(x)$ contains, for each $n \geq 1$, at most one word of length n which is right-special.

Since $F(x)$ is closed under reversal, the reversal of a right-special factor of length n is left-special, and it is the only left-special factor of length n of x. A suffix of a right-special factor is again right-special. Symmetrically, a prefix of a left-special factor is again left-special.

As a particular case, a *strict* episturmian word is an episturmian word x with the two following properties: x has exactly one right-special factor of each length and moreover each right-special factor u of x is strict, that is satisfies the inclusion $uA \subset F(x)$ (see [5]).

A *Sturmian set* is the set of factors of a strict episturmian word.

It is easy to see that for a strict episturmian word x on an alphabet A with k letters, the set $F(x) \cap A^n$ has $(k-1)n+1$ elements for each n. Thus, for a binary alphabet, the strict episturmian words are just the Sturmian words, since a Sturmian word has one right-special factor for each length and its set of factors is closed under reversal.

An episturmian word s is called *standard* if all its left-special factors are prefixes of s. For any episturmian word s, there is a standard one t such that $F(s) = F(t)$. This is a rephrasing of Theorem 5 in [5].

Example 4.5 Set $A = \{a, b\}$. The *Fibonacci morphism* is the substitution $f : A^* \to A^*$ defined by $f(a) = ab$ and $f(b) = a$. The *Fibonacci word*

$$x = abaababaabaababaababaabaababaabaab\cdots$$

is the fixpoint $f^\omega(a)$ of f. It is a Sturmian word (see [12] Example 2.1.1). We call *Fibonacci set* the set of factors of the Fibonacci word.

Example 4.6 Consider the following generalization of the Fibonacci word to the ternary alphabet $A = \{a, b, c\}$. Consider the morphism $f : A^* \to A^*$ defined by $f(a) = ab$, $f(b) = ac$ and $f(c) = a$. The fixpoint

$$f^\omega(a) = abacabaabacababacabaabacabacabaabacab\cdots$$

is the *Tribonacci word*. It is a strict standard episturmian word (see [9]).

Let G be a transitive permutation group of degree d on a set Q which can be generated by k elements. Let A be a k-letter alphabet and let $\varphi : A^* \to G$ be a surjective morphism from A^* onto G. Let 1 be an element of Q. The map $q \mapsto q\varphi(a)$ defines an automaton $\mathcal{A} = (Q, 1, 1)$. It is trim since G is transitive and thus it is a group automaton. Furthermore, G is the transition monoid of $\mathcal{A}$. Let Z be the group code generating the submonoid recognized by $\mathcal{A}$. Then $G(Z) = G$ by definition. Thus Theorem 4.2 follows from the following result from [2].

Theorem 4.7 *Let A be an alphabet with k elements. Let $Z \subset A^*$ be a group code of degree d. Let F be a Sturmian set. The set $X = Z \cap F$ has $(k-1)d+1$ elements and $G(Z)$ is a syntactic group of X.*

We illustrate the construction on three examples. In each case, we start from a group code Z and we compute the minimal automaton of X^*, where $X = Z \cap F$. We exhibit in each case a set $I \subset Q$ such that Group(I) is equivalent to $G(Z)$.

Example 4.8 Let Z be the group code corresponding to the automaton on the left of Figure 4.3. Let F be the Fibonacci set and let $X = Z \cap F$. The minimal automaton of X^* is represented on the right. The holonomy group relative to $\{1, 4, 5\}$ is the symmetric group S_3 generated by the permutations (45) defined by ab and (15) defined by aab.

Example 4.9 Let Z be group code defined by $Z^* = \varphi^{-1}(0, 0)$ where $\varphi : \{a, b\}^* \to (\mathbb{Z}/2\mathbb{Z})^2$ is defined by $\varphi(a) = (1, 0)$ and $\varphi(b) = (0, 1)$. The minimal automaton of Z^* is represented on the left of Figure 4.4. Let F be the Fibonacci set. The bifix code $X = Z \cap F$ is represented on the right of Figure 4.4.

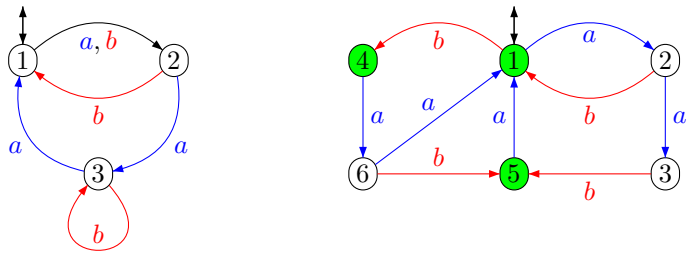

Figure 4.3: A code with syntactic group S_3

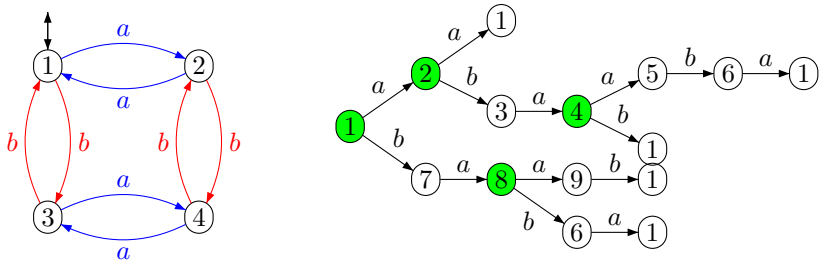

Figure 4.4: A code with syntactic group $(\mathbb{Z}/2\mathbb{Z})^2$

We have already seen (Example 4.4) that the holonomy group relative to $\{1, 2, 4, 8\}$ is the regular representation of $(\mathbb{Z}/2\mathbb{Z})^2$.

Example 4.10 Let $A = \{a, b, c\}$ and let $F \subset A^*$ be the set of factors of the Tribonacci word (see Example 4.6). Let $\varphi : A^* \to (\mathbb{Z}/2\mathbb{Z})^3$ be the morphism defined by $\varphi(a) = (1, 0, 0)$, $\varphi(b) = (0, 1, 0)$ and $\varphi(c) = (0, 0, 1)$. Let Z be the group code generating the submonoid $\varphi^{-1}(0, 0, 0)$ and let $X = Z \cap F$. The code X has 17 elements and the minimal automaton of X^* is given in Table 4.1. The word *abacaba* is an element of F of rank 8. Its image is the set $I = \{1, 2, 7, 9, 19, 21, 30, 32\}$. The action on the 8-element sets is shown on Figure 4.5.

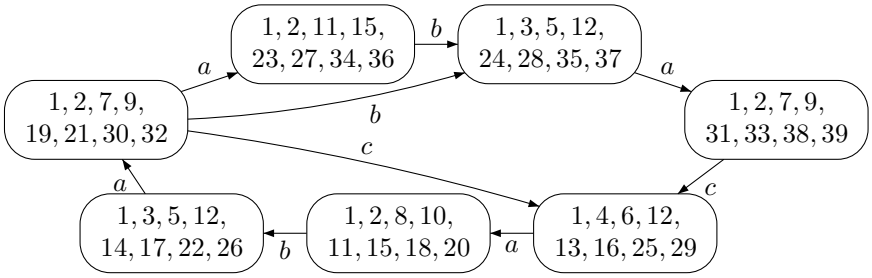

Figure 4.5: The action on 8-elements subsets

It shows that the holonomy group relative to I is generated by the three

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
a	2	1	7	8	9	10	11		15			1	18	19		
b	3	5					12	14	1	17	1				12	
c	4	6					13		16							
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
a	20	21		23		27	30		31	15	32		33	11	34	
b			22	24	26	28		28				24			35	
c				25		29									12	12
	32	33	34	35	36	37	38	39								
a	36			38		39										
b	37		37		35											
c	1	1					25	29								

Table 4.1: The transitions of the automaton $\mathcal{A}(X^*)$

permutations on the right below:

$$\begin{array}{ll} caba & (1,19)(2,21)(7,30)(9,32) \\ bacaba & (1,30)(2,32)(7,19)(9,21) \\ abacaba & (1,32)(2,30)(7,21)(9,19) \end{array}$$

defined respectively by the three words on the left. Thus this group is the regular representation of $(\mathbb{Z}/2\mathbb{Z})^3$.

5. Codes with empty kernel

A word s is an *internal* factor of w if $w = rst$ with r, t nonempty. The *kernel* of a set of words X is the set of words in X which are internal factors of the words of X. We will state in this section results on syntactic groups of prefix codes with empty kernel. There are two important particular cases of prefix codes with empty kernel.

The first case is that of *semaphore codes*. A semaphore code is a set of the form $X = A^*S \setminus A^*SA^+$ for some nonempty set $S \subset A^+$. Equivalently, a semaphore code is a maximal prefix code with empty kernel (see [3]).

The second particular case is that of *infix codes*. An infix code is a set of words which does not contain any proper factor of its elements. It is a bifix code with an empty kernel.

Let X be a prefix code and let F be the set of internal factors of X. Let $\mathcal{A} = \mathcal{A}(X^*)$ be the minimal automaton of X^*. A syntactic group G of X is *proper* if there is a word $w \notin F$ such that $\varphi_{\mathcal{A}}(w) \in G$. Note that if G is proper, then for any $g \in G$ there is a word $w \notin F$ such that $\varphi_{\mathcal{A}}(w) = g$.

Recall that a permutation group G is *regular* if no element of G distinct from the identity has a fixpoint.

The following result is from [1] where it is stated in the more general case of codes instead of prefix codes.

Theorem 5.1 *Let $X \subset A^+$ be a prefix code with empty kernel. Any proper syntactic group of X is a finite regular cyclic group.*

Example 5.2 Let $X = \{aa, aba, bab, bb\}$. The set X is an infix code. The minimal automaton $\mathcal{A}$ of X^* is represented in Figure 5.1. The transition monoid

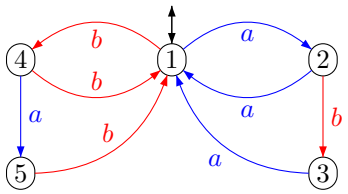

Figure 5.1: The automaton $\mathcal{A}(X^*)$

M of $\mathcal{A}$ contains groups which are cyclic of order 1, 2 or 3. For example, ab contains the cycle (134) while a contains the cycle (12).

Let us note an interesting feature of this example. For any word w in $(a \cup babb^*aba)^*$, we have $\varphi_{\mathcal{A}}(w) \in \mathrm{Stab}(\{1, 2\})$ and w defines a cyclic group of order 2 (see Figure 5.2). But the set of these words is a submonoid which is not cyclic and not even finitely generated.

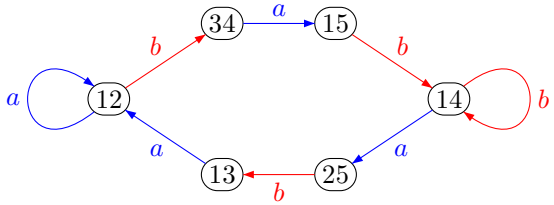

Figure 5.2: The action of A^* on the 2-element subsets reachable from $\{1, 2\}$

Example 5.3 Let $X = \{aaa, aab, abaa, abab, baba, babb, bba, bbb\}$. The set X is an infix code. The minimal automaton $\mathcal{A}$ of X^* is represented in Figure 5.3. The transition monoid $M = \varphi(A^*)$ of $\mathcal{A}$ contains cyclic groups of degree 1, 2, 3 and 4. For example the word a contains the cycle (123). In turn, ba contains the permutation (16)(23).

This example shows another interesting feature. Consider the group G containing $\varphi(ba)$. The neutral element of G is $e = \varphi(baba)$ and its set of fixpoints is $\{1, 2, 3, 6\}$. The group G is of degree 4. It is composed of the permutation (16)(23) and the identity. It is thus of order 2 (note that this gives an example of a syntactic group which is not transitive). Actually, M does not contain any cyclic group of order 4.

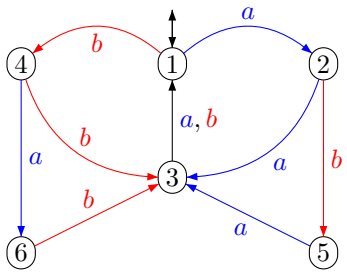

Figure 5.3: An infix code with a group of degree 4 and order 2.

From groups to monoids. Let G be a transitive permutation group on $R = \{1, 2, \ldots, n\}$ and let $\varphi : A^* \to G$ be a surjective morphism. Let H be the subgroup of G which fixes 1. Let Z be the bifix code generating the submonoid $\varphi^{-1}(H)$. Let X be the set of elements of Z which have no proper factor in Z. Let M be the transition monoid of the automaton $\mathcal{A}(X^*)$.

The following result appears in [1].

Proposition 5.4 *The set X is a finite infix code. The syntactic groups of X, except possibly the group reduced to the neutral element of M, are cyclic and regular of degree at most n.*

Note that the group reduced to the neutral element of M is of degree equal to the number of states of $\mathcal{A}(X^*)$ which may be larger than n.

Example 5.5 Let G be the symmetric group on the set $R = \{1, 2, 3\}$ and let $A = \{a, b\}$. Let $\varphi : A^* \to G$ be the morphism defined by $\varphi(a) = (12)$, $\varphi(b) = (13)$. The bifix code Z is the infinite set represented on the left of Figure 5.4 and the code X is the finite set represented on the right. It is the code of Example 5.2.

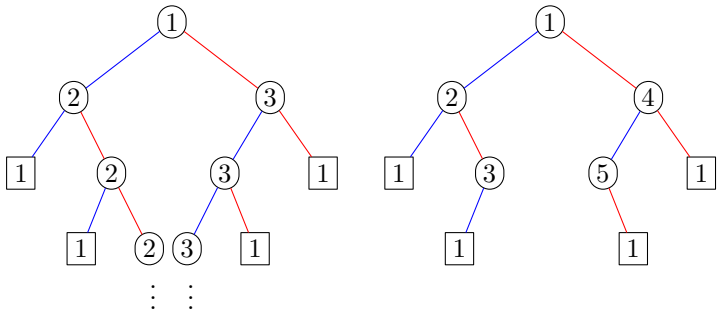

Figure 5.4: The infinite group code Z and the finite code X

Example 5.6 The code X of Example 5.3 corresponds to the above construction with G being the alternating group on the set $\{1, 2, 3, 4\}$ and $\varphi(a) = (123)$, $\varphi(b) = (143)$.

Example 5.7 Let φ be the morphism onto the alternating group A_5 defined by $\varphi(a) = (123)$, $\varphi(b) = (145)$. We obtain

$$X = \{aaa, aaba, aabba, abaa, ababa, abbaa, baabb, babab, babb, bbaab, bbab, bbb\}.$$

The transitions of the minimal automaton $\mathcal{A}$ of X^* are represented on Table 5.1. The transition monoid of $\mathcal{A}$ has 14351 elements (result obtained using the soft-

	1	2	3	4	5	6	7	8	9	10	11	12	13
a	2	3	1	8	9	7	1	13	10	–	1	11	–
b	4	6	7	5	1	12	11	9	1	1	–	–	10

Table 5.1: The transitions of $\mathcal{A}(X^*)$

ware Semigroupes [7]). Moreover

- ab defines the cycle (1 6 11 4 9).
- $a^2ba^2b^2$ defines the permutation (1 12)(5 11) of degree 4 but there is no cyclic group of order 4 in $M(\mathcal{A})$.
- a and b define cycles of degree 3.
- a^4ba^5 defines the identity on $\{1, 2\}$ and $a^2bab^2a^2ba^4$ defines the transposition (1 2).

Černý automata. Let $\mathcal{A}$ be a complete deterministic automaton on a set Q of states. A word w is called *synchronizing* for $\mathcal{A}$ if the set $Q \cdot w$ of states reached after reading w has only one element. The automaton $\mathcal{A}$ is called *synchronized* if there exists a synchronizing word for $\mathcal{A}$. Recall that Černý's conjecture [4] expresses that a complete deterministic synchronized automaton with n states has a synchronizing word of length at most $(n-1)^2$.

The Černý automaton of order n, denoted $\mathcal{C}_n$ is the following automaton on the alphabet $A = \{a, b\}$. Its set of states is the set $\{0, 1, \ldots, n-1\}$ and the transitions are defined by

$$i \cdot a = i + 1 \bmod n, \quad i \cdot b = \begin{cases} 1 & \text{if } i = 0 \\ i & \text{otherwise} \end{cases}$$

The automata $\mathcal{C}_n$ for $n = 4, 5, 6$ are represented in Figure 5.5.

For each $n \geq 1$, the automaton $\mathcal{C}_n$ is synchronized. Indeed, it is easy to verify that the word $(ba^{n-1})^{n-2}b$ is a synchronizing word of length $(n-1)^2$. It is known that there is no synchronizing word of length less than $(n-1)^2$ [4]. We use Theorem 5.1 to prove the following result. The proof uses semaphore codes (see the beginning of Section 5).

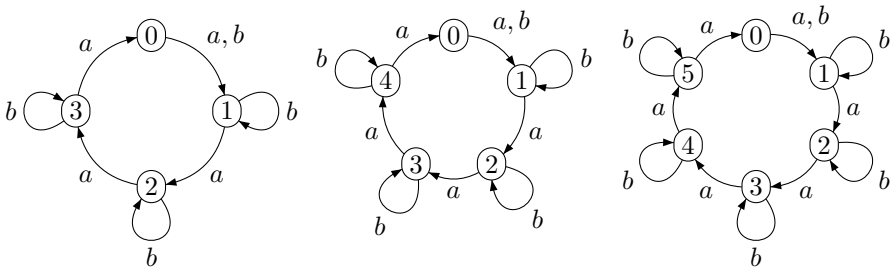

Figure 5.5: The automata $\mathcal{C}_4, \mathcal{C}_5, \mathcal{C}_6$

Proposition 5.8 *The holonomy groups of $\mathcal{C}_n$ are cyclic and regular.*

Proof. Let X_n be the prefix code generating the stabilizer of 0 in $\mathcal{C}_n$, which is the submonoid formed by the words $x \in A^*$ such that $0 \cdot x = 0$. We have

$$X_n = a(b^*a)^{n-1} \cup b(b^*a)^{n-1}$$

It is easy to verify that X_n is a maximal prefix code which has empty kernel. Thus it is a semaphore code. The automaton $\mathcal{A} = \mathcal{C}_n$, with 0 as initial and terminal state, is the minimal automaton of X_n^*. Let G be a syntactic group of X_n. Since $\varphi_{\mathcal{A}}(a^n)$ is the neutral element of $M(\mathcal{A})$, the set $\varphi_{\mathcal{A}}^{-1}(G)$ contains words which have an arbitrary high number of occurrences of a, and thus which are not factors of X. Thus G is proper and we conclude that it is cyclic and regular by Theorem 5.1. By Proposition 3.2, the holonomy groups of $\mathcal{A}$ are the restriction to I of the group $G(e) \cap \mathrm{Stab}(I)$ for some idempotent e in $M(\mathcal{A})$ with image J containing I. Since $\mathrm{Group}(J)$ is a syntactic group of X, it is cyclic and regular. Thus the same holds for the holonomy groups of $\mathcal{A}$. ∎

Note that the semaphore code X_n above is the product of the two semaphore codes A and $(b^*a)^{n-1}$ (the product of two semaphore codes is a semaphore code by Proposition 3.5.12 of [3]).

We do not know if there is a relation between the property of Černý automata expressed by Proposition 5.8 and the fact that they are an extremal exemple for the shortest length of synchronizing words. A connexion between Černý's conjecture and holonomy groups has already been established. Indeed, it is known that the conjecure holds for aperiodic automata, i.e. automata with trivial holonomy groups [18].

We give below examples showing the following features of the automata $\mathcal{C}_n$ for $n = 4, 5, 6$: *All holonomy groups are cyclic of order and degree k for all k with $1 \leq k \leq n$.*

We do not know if this property holds for all automata $\mathcal{C}_n$.

Example 5.9 The automaton $\mathcal{C}_4$ has cyclic holonomy groups of order and degree $1, 2, 3, 4$. Indeed, the table below indicates the permutations defined by

each word on the first row.

a	ab	a^2b	a^3b
(0123)	(123)	(13)	(1)

Example 5.10 The automaton $\mathcal{C}_5$ has cyclic holonomy groups of order and degree $1, 2, 3, 4, 5$. Indeed, the table below indicates the permutations defined by each word on the first row.

a	ab	a^3b	a^2b	a^4b
(01234)	(1234)	(142)	(13)	(1)

Example 5.11 The automaton $\mathcal{C}_6$ has cyclic holonomy groups of order and degree $1, 2, 3, 4, 5, 6$. Indeed, the table below indicates the permutations defined by each word on the first row.

a	ab	a^3bab	a^2b	a^3ba^5b	a^5b
(012345)	(12345)	(1532)	(135)	(13)	(1)

It is easy to see that, in general,

(i) a defines the n-cycle $(01\cdots n-1)$
(ii) ab defines the $n-1$ cycle $(12\cdots n-1)$.
(iii) $a^{n-1}b$ defines the cycle (1).

It does not seem easy to describe in general the shape of the words defining a cycle of length k for $1 < k < n-1$.

References

[1] Jean Berstel, Clelia De Felice, Dominique Perrin, and Giuseppina Rindone. On the groups of codes with empty kernel. *Semigroup Forum*, 80(3):351–374, 2010. 2, 16, 17

[2] Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. 2011. `http://arxiv.org/abs/1011.5369`. 2, 4, 5, 11, 13

[3] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009. 2, 7, 10, 15, 19

[4] Ján Černý. Poznamka k homogenym s konecnymi automati. *Mat.-fyz. cas. SAV.*, 14:208– 215, 1964. 18

[5] Xavier Droubay, Jacques Justin, and Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001. 12, 13

[6] Samuel Eilenberg. *Automata, Languages, and Machines. Vol. B.* Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. 2, 5

[7] Véronique Froidure and Jean-Eric Pin. Algorithms for computing finite semigroups. In *Foundations of computational mathematics (Rio de Janeiro, 1997)*, pages 112–126. Springer, Berlin, 1997. 18

[8] W. Michael L. Holcombe. *Algebraic Automata Theory*. Cambridge University Press, 1982. 5

[9] Jacques Justin and Laurent Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5):343–356, 2000. 13

[10] Gérard Lallement. *Semigroups and combinatorial applications*. John Wiley & Sons, New York-Chichester-Brisbane, 1979. Pure and Applied Mathematics, A Wiley-Interscience Publication. 2, 7

[11] M. Lothaire. *Combinatorics on Words*. Cambridge University Press, second edition, 1997. (First edition 1983). 11

[12] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002. 2, 13

[13] Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, 1977. 8, 9

[14] Dominique Perrin. Sur les groupes dans les monoïdes finis. In *Noncommutative Structures in Algebra and Geometric Combinatorics (Naples 1978)*, volume 109 of *Quaderni de "La Ricerca Scientifica"*, pages 27–36. CNR, 1981. 10

[15] Dominique Perrin and Giuseppina Rindone. On syntactic groups. *Bull. Belg. Math. Soc. Simon Stevin*, 10(suppl.):749–759, 2003. 2, 11

[16] Marcel-Paul Schützenberger. A property of finitely generated submonoids of free monoids. In *Algebraic theory of semigroups (Proc. Sixth Algebraic Conf., Szeged, 1976)*, volume 20 of *Colloq. Math. Soc. János Bolyai*, pages 545–576. North-Holland, Amsterdam, 1979. 2, 11

[17] Jacques Stern. Complexity of some problems from the theory of automata. *Inform. and Control*, 66(3):163–176, 1985. 1

[18] Avraham Trakhtman. The Černý conjecture for aperiodic automata. *Discrete Mathematics & Theoretical Computer Science*, 9(2), 2007. 19