

# Coding Partitions: Regularity, Maximality and Global Ambiguity<sup>\*</sup>

Marie-Pierre Béal<sup>1</sup>, Fabio Burderi<sup>2</sup>, and Antonio Restivo<sup>2</sup>

<sup>1</sup> Institut Gaspard-Monge,  
Laboratoire d'informatique UMR 8049,  
Université de Marne-la-Vallée, 77454 Marne-la-Vallée Cedex 2, France  
[beal@univ-mlv.fr](mailto:beal@univ-mlv.fr)

<sup>2</sup> Dipartimento di Matematica ed Applicazioni,  
Università degli studi di Palermo,  
Via Archirafi 34, 90123 Palermo, Italy  
{Burderi, Restivo}@math.unipa.it

**Abstract.** The canonical coding partition of a set of words is the finest partition such that the words contained in at least two factorizations of a same sequence belong to a same class. In the case the set is not uniquely decipherable, it partitions the set into one unambiguous class and other parts that localize the ambiguities in the factorizations of finite sequences.

We firstly prove that the canonical coding partition of a regular set contains a finite number of regular classes. We give an algorithm for computing this partition. We then investigate maximality conditions in a coding partition and we prove, in the regular case, the equivalence between two different notions of maximality. As an application, we finally derive some new properties of maximal *UD* codes.

## 1 Introduction

In this paper, we call code a set of finite words. An important class of codes is the class of uniquely decipherable codes. This property allows the decoding of a sequence of concatenated codewords. Nevertheless, some classes of codes are used in information theory although they are not uniquely decipherable (see for instance [7], [9] and [10]). The condition of unique decipherability can also be weakened by considering that it applies only to codes with constraints (see [1]) or to codes with a constraint source (see [4], [6]). In [6], the classification of ambiguities of codes is investigated in the study of natural languages. From a combinatorial point of view, the study of ambiguities helps to understand the structure of a code.

To this purpose, the notions of coding partition and canonical coding partition of a code were introduced in [3] to study some decipherability conditions

---

<sup>\*</sup> Partially supported by Italian MURST Project of National Relevance “Linguaggi Formali e Automi: Metodi, Modelli e Applicazioni”

for codes weaker than the unique decipherability. The notion of coding partition generalizes that of *UD* code: indeed *UD* codes correspond to the extremal case in which each class contains exactly one element. In general, for codes that are not *UD*, the notion of coding partition allows to recover “unique decipherability” at the level of classes of the partition. In other words, such notion gives a tool to *localize* the ambiguities for a code that is not *UD*: indeed the ambiguities are bordered inside the individual classes of the partition and a sort of mutual unambiguity holds between the different classes.

By taking into account the natural ordering between the partitions of a set  $X$ , where finer is higher, we have that the coding partitions form a complete lattice. As a consequence, given a code  $X$ , we can define the finest coding partition  $P$  of  $X$ . It is called the *characteristic* partition of  $X$  and it is denoted by  $P(X)$ .

The structure of  $P(X)$  gives useful information about coding properties of  $X$ . In particular, an extremal case (each class of  $P(X)$  contains only one element) corresponds to *UD* codes. The opposite extremal case ( $P(X)$  contains only one class) gives rise to the definition of *globally ambiguous (GA)* code. Such considerations lead to define a *canonical decomposition* of a code in at most one unambiguous component and in a set (possibly empty) of *GA* components.

Remark that the notion of coding partition is closely related to some special cases of the notion of  $\mathfrak{F}$ -factorization, introduced in [8].

In [3] is given a Sardinas-Patterson like algorithm for computing the canonical coding partition of a finite code.

In this paper, we firstly prove that the canonical coding partition of a regular code has a finite number of classes, each one being regular. This result was conjectured in [3]. We give an exponential time algorithm for computing all classes of the partition which is based on automata constructions.

We then introduce the notion of maximality of coding partition with respect to a component, and we prove, in the regular case, that if a coding partition is maximal with respect to one component then it is maximal with respect to all the components. As an application, we prove, in the last section, that, if a regular *UD* code  $X$  is maximal, then any code containing strictly  $X$  is *GA*.

## 2 Partitions of a code

Let  $A$  be a finite alphabet. We denote by  $A^*$  the set of finite words over the alphabet  $A$ , and by  $A^+$  the set of nonempty finite words. A *code*  $X$  is here a subset of  $A^+$ . Its elements are called *code words*, the elements of  $X^*$  *messages*. Let  $X$  be a code and let

$$P = \{X_1, X_2, \dots\},$$

be a partition of  $X$  i.e. :  $\bigcup_{i \geq 1} X_i = X$  and  $X_i \cap X_j = \emptyset$ , for  $i \neq j$ .

A *P-factorization* of an element  $w \in X^+$  is a factorization  $w = z_1 z_2 \dots z_t$ , where

- $\forall i \ z_i \in X_k^+$ , for some  $k \geq 1$
- if  $t > 1$ ,  $z_i \in X_k^+ \Rightarrow z_{i+1} \notin X_k^+$ , for all  $1 \leq i \leq t - 1$ .

The partition  $P$  is called a *coding partition* if any element  $w \in X^+$  has a *unique  $P$ -factorization*, i.e. if

$$w = z_1 z_2 \cdots z_s = u_1 u_2 \cdots u_t,$$

where  $z_1 z_2 \cdots z_s, u_1 u_2 \cdots u_t$  are  $P$ -factorizations of  $w$ , then  $s = t$  and  $z_i = u_i$  for  $i = 1, \dots, s$ .

We say that a partition  $P$  is *concatenatively independent* if, for  $i \neq j$ ,

$$X_i^+ \cap X_j^+ = \emptyset.$$

Then a necessary condition for a partition  $P$  to be a coding partition, is that  $P$  is concatenatively independent.

Let  $X$  be a code and let  $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$  be two factorizations into code words of a message  $w \in X^+$ . We say that the relation  $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$  is *prime* if for all  $i < s$  and for all  $j < t$  one has  $x_1 x_2 \cdots x_i \neq y_1 y_2 \cdots y_j$ .

In [3] is proved that  $P$  is a coding partition of a code  $X$  iff for every prime relation  $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$  these code words belong to the same component of the partition.

Recall that there is a natural order between the partitions of a set  $X$ : if  $P_1$  and  $P_2$  are two partitions of  $X$ ,  $P_1 \leq P_2$  if the elements of  $P_1$  are unions of elements of  $P_2$ . In [3] is proved the next theorem.

**Theorem 1.** *The set of the coding partitions of a code  $X$  is a complete lattice.*

As a consequence of previous theorem we can give the next definition.

Given a code  $X$ , the finest coding partition  $P$  of  $X$  is called the *characteristic partition* of  $X$  and it is denoted by  $P(X)$ .

A code  $X$  is called *ambiguous* if it is not  $UD$ . It is called *globally ambiguous (GA)* if  $|X| > 1$  and  $P(X)$  is the trivial partition.

So  $UD$  codes and  $GA$  codes correspond to the two extremal cases: a code is  $UD$  if  $|P(X)| = |X|$  and a code is  $GA$  if  $|P(X)| = 1$ .

Let  $X$  be a code and let  $P(X)$  be the characteristic partition of  $X$ . Let  $X_0$  be the union of all classes of  $P(X)$  having only one element, i.e. of all classes  $Z \in P(X)$  such that  $|Z| = 1$ . The code  $X_0$  is a  $UD$  code and is called the *unambiguous component* of  $X$ . From  $P(X)$  one then derives another partition of  $X$

$$P_C(X) = \{X_0, X_1, \dots\},$$

where  $|X_i| > 1$ , for  $i \geq 1$ . The sets  $X_i$ , with  $i \geq 1$ , are (see[3])  $GA$ . They are called the *GA components* of  $X$ . The partition  $P_C(X)$  is called the *canonical partition* of  $X$ : it defines a *canonical decomposition* of a code  $X$  in at most one unambiguous component and a (possibly empty) set of  $GA$  components. Roughly speaking, if a code  $X$  is not  $UD$ , then its canonical decomposition, on one hand separates the unambiguous component of the code (if any), and, on the other, localizes the ambiguities inside the  $GA$  components of the code. If,

on the contrary,  $X$  is  $UD$ , then its canonical decomposition contains only the unambiguous component  $X_0$ . Moreover if  $X$  is  $UD$  then every partition of  $X$  is a coding partition.

In [3] is given a Sardinas-Patterson like algorithm for computing the canonical coding partition of a finite code  $X$  and is also proved the next result.

**Theorem 2.** *Given a partition  $P = \{X_1, X_2, \dots, X_n\}$  such that  $X_i$ , for  $i = 1, 2, \dots, n$ , is a regular set, then it is decidable whether  $P$  is a coding partition.*

In the same paper it was also formulated the following conjecture.

**CONJECTURE :** *If  $X$  is regular, the number of classes of  $P_C(X)$  is finite and each class of  $P_C(X)$  is a regular set.*

The conjecture will be proved in the next section so the restrictive conditions considered in the last theorem are not actually a restriction for regular codes.

### 3 Coding partition of a regular code

In this section, we consider a regular code  $X$ .

We say that a coding partition of a code is *finite* if it has a finite number of components. We say that a coding partition of a code is *regular* if all the components of the partitions are regular. The following theorem gives a positive answer to previous conjecture.

**Theorem 3.** *The canonical partition of a regular code is finite and regular.*

*Remark 1.* Given a coding partition  $P = \{X_1, X_2, \dots\}$  of a code  $X \subseteq A^+$ , the condition that every word  $w \in X^+$  admits a unique  $P$ -factorization has the following algebraic interpretation: the submonoid  $X^*$  is isomorphic to the *free product* of the submonoids  $X_i^*$ . We say that a submonoid  $M \subseteq A^*$  is *indecomposable* if  $M$  is not factorizable in the free product of others submonoids. Then the previous theorem can be restated in the following algebraic setting.

**Theorem 4.** *Any regular monoid admits a canonical decomposition into a free product of at most one regular free monoid and a finite number (possibly zero) of regular indecomposable monoids.*

In order to prove Theorem 3, we give an algorithm for computing the finite automata accepting the components of the partition from a finite automaton accepting the code  $X$ .

A finite *automaton*  $\mathcal{A} = (Q, I, E, T)$  is made of a finite set of states  $Q$ , a set of edges  $E$  labelled on an alphabet  $A$ , a set of initial states  $I$  and a set of final states  $T$ . We shall also consider automata labelled in  $A^*$ . A *successful path* is a path going from a state of  $I$  to a state of  $T$ . The set of labels of successful paths is the *language accepted* by the automaton.

An automaton is *unambiguous* if for any word  $z$ , any states  $p, q$ , there is at most one path going from  $p$  to  $q$  and labelled by  $z$ .

Let  $\mathcal{A} = (Q, I, E, T)$  be a finite automaton. We define the automaton  $\mathcal{A} \times \mathcal{A} = (Q', I', E', T')$  called the *square* of  $\mathcal{A}$ , where  $Q' = Q \times Q$ ,  $E' = \{(p, q) \xrightarrow{a} (p', q') \mid p \xrightarrow{a} p' \text{ and } q \xrightarrow{a} q' \in E\}$ . The set of initial states  $I'$  and the set of final states  $T'$  will be specified later. A state  $(p, q)$  will be also denoted by  $\begin{bmatrix} p \\ q \end{bmatrix}$ .

*Proof (Proof of Theorem 3).* Let  $\mathcal{A} = (Q, I, E, T)$  be a finite unambiguous automaton accepting the code  $X$  such that  $I = \{i\}$ ,  $T = \{t\}$ , and which has no edge coming in  $i$  and no edge going out of  $t$ . Such an automaton, called a normalized automaton, can be obtained by standard constructions (see for instance [2]). By merging  $i$  and  $t$  into a single state denoted by 0, we get an automaton  $\mathcal{B} = (Q, 0, E, 0)$  accepting the set  $X^*$ . Note that  $\mathcal{B}$  is no more unambiguous unless  $X$  is UD.

We build the square automaton  $\mathcal{B} \times \mathcal{B}$  and replace the state  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  by two states  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  and  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$  such that the edges going out of  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  go out of  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  and the edges coming in  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  come in  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$ . Note that  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  has no incoming edges and  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$  has no outgoing edges. We only keep in  $\mathcal{B} \times \mathcal{B}$  the states belonging to paths from  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  to  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$  and going at least one time through a state  $\begin{bmatrix} p \\ q \end{bmatrix}$  with  $p = 0, q \neq 0$  or  $p \neq 0, q = 0$ . By using the state-elimination technique (see for instance [11]), we remove the states  $\begin{bmatrix} p \\ q \end{bmatrix}$  with  $p$  and  $q$  distinct from 0 and get an automaton  $\mathcal{C}$  labelled in regular subsets of  $A^*$  whose states are  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$ ,  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$ , and  $\begin{bmatrix} p \\ q \end{bmatrix}$  with  $p = 0, q \neq 0$  or  $p \neq 0, q = 0$ . There is at most one edge between two states and each label is a regular non-empty subset of  $A^*$ .

States  $\begin{bmatrix} p \\ q \end{bmatrix}$  with  $p = 0$  are called upper-zero states while states  $\begin{bmatrix} p \\ q \end{bmatrix}$  with  $q = 0$  are called lower-zero states. Hence  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  and  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$  are both upper and lower-zero states.

We denote by  $E\begin{bmatrix} p \\ q \end{bmatrix}\begin{bmatrix} p' \\ q' \end{bmatrix}$  the regular set related to the edge  $\begin{bmatrix} p \\ q \end{bmatrix} \rightarrow \begin{bmatrix} p' \\ q' \end{bmatrix}$ . With a slight abuse of language, we sometimes say that there is an edge labelled by a word  $w$  from a state  $\begin{bmatrix} p \\ q \end{bmatrix}$  to state  $\begin{bmatrix} p' \\ q' \end{bmatrix}$  whenever  $w \in E\begin{bmatrix} p \\ q \end{bmatrix}\begin{bmatrix} p' \\ q' \end{bmatrix}$ .

Let  $p_i, q_i, p_j, q_j$  be states in  $Q$  with  $q_i$  and  $q_j$  distinct from 0. Let  $e, f$  be the edges

$$e = \begin{bmatrix} 0 \\ p_i \end{bmatrix} \rightarrow \begin{bmatrix} q_i \\ 0 \end{bmatrix} \text{ and } f = \begin{bmatrix} q_j \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ p_j \end{bmatrix}$$

(*i.e.* respectively an edge from an upper-zero state to a lower-zero state and an edge from lower-zero state to an upper-zero state).

We denote by

- $L\begin{bmatrix} q_i \\ 0 \end{bmatrix}\begin{bmatrix} q_j \\ 0 \end{bmatrix}$  the regular set of labels of paths from  $\begin{bmatrix} q_i \\ 0 \end{bmatrix}$  to  $\begin{bmatrix} q_j \\ 0 \end{bmatrix}$  with *all its states being lower-zero states*.
- $S\begin{bmatrix} q_i \\ 0 \end{bmatrix}\begin{bmatrix} q_j \\ 0 \end{bmatrix}$  the union of the labels of all edges contained in a path from  $\begin{bmatrix} q_i \\ 0 \end{bmatrix}$  to  $\begin{bmatrix} q_j \\ 0 \end{bmatrix}$  with *all its states being lower-zero states*.

Note that we may have  $q_i = q_j$ . In this case,  $L\begin{bmatrix} q_i \\ 0 \end{bmatrix}\begin{bmatrix} q_j \\ 0 \end{bmatrix}$  contains the empty word and  $S\begin{bmatrix} q_i \\ 0 \end{bmatrix}\begin{bmatrix} q_j \\ 0 \end{bmatrix}$  may be the empty set.

We define the regular sets

$$Y = E \begin{bmatrix} 0 \\ p_i \end{bmatrix} \begin{bmatrix} q_i \\ 0 \end{bmatrix} \cdot L \begin{bmatrix} q_j \\ 0 \end{bmatrix} \begin{bmatrix} q_j \\ 0 \end{bmatrix} \cdot E \begin{bmatrix} q_j \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ p_j \end{bmatrix} + S \begin{bmatrix} q_i \\ 0 \end{bmatrix} \begin{bmatrix} q_j \\ 0 \end{bmatrix},$$

$$S_{ef} = \begin{cases} Y & \text{if } p_i \neq 0, p_j \neq 0, \\ Y + E \begin{bmatrix} 0 \\ 0 \end{bmatrix}_s \begin{bmatrix} q_i \\ 0 \end{bmatrix} & \text{if } p_i = 0, p_j \neq 0, \\ Y + E \begin{bmatrix} q_j \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_t & \text{if } p_i \neq 0, p_j = 0, \\ Y + E \begin{bmatrix} 0 \\ 0 \end{bmatrix}_s \begin{bmatrix} q_i \\ 0 \end{bmatrix} + E \begin{bmatrix} q_j \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_t & \text{if } p_i = p_j = 0, \end{cases}$$

where the symbol  $+$  is the union symbol and the dot symbol is the concatenation symbol.

Let  $p_i, q_i, p_j, q_j, p_k, q_k$  be states in  $Q$  with  $q_i, q_j, p_j, p_k$  distinct from 0. Let  $e, f, g$  be the edges

$$e = \begin{bmatrix} 0 \\ p_i \end{bmatrix} \rightarrow \begin{bmatrix} q_i \\ 0 \end{bmatrix}, \quad f = \begin{bmatrix} q_j \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ p_j \end{bmatrix} \quad \text{and} \quad g = \begin{bmatrix} 0 \\ p_k \end{bmatrix} \rightarrow \begin{bmatrix} q_k \\ 0 \end{bmatrix}.$$

We define the regular set

$$S_{efg} = E \begin{bmatrix} 0 \\ p_i \end{bmatrix} \begin{bmatrix} q_i \\ 0 \end{bmatrix} \cdot L \begin{bmatrix} q_j \\ 0 \end{bmatrix} \begin{bmatrix} q_j \\ 0 \end{bmatrix} \cdot E \begin{bmatrix} q_j \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ p_j \end{bmatrix} + E \begin{bmatrix} q_j \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ p_j \end{bmatrix} \cdot L \begin{bmatrix} 0 \\ p_j \end{bmatrix} \begin{bmatrix} 0 \\ p_k \end{bmatrix} \cdot E \begin{bmatrix} 0 \\ p_k \end{bmatrix} \begin{bmatrix} q_k \\ 0 \end{bmatrix}.$$

We define similar sets  $S_{ef}$  and  $S_{efg}$  when  $e, g$  are edges from a lower-zero state to an upper-zero state and  $f$  is an edge from an upper-zero state to a lower-zero state, by exchanging the roles played by the upper and lower states.

We get a finite number of regular subsets of  $X$ . Some of these states may have a nonempty intersection. We replace two parts having a non-empty intersection by their union. After a finite number of steps we get a finite number of regular subsets of  $X$  whose two by two intersections are empty. We denote these sets by  $X_1, X_2, \dots, X_r$ . We define the set  $X_0 = X - \bigcup_{i=1}^r X_i$ . We claim that  $(X_i)_{0 \leq i \leq r}$  is the canonical coding partition of  $X$ , which proves the proposition.

To prove our claim, we show that any two code words which belong to a same prime relation belong to a same component  $X_i$ . Let  $z = x_1 x_2 \dots x_n = y_1 y_2 \dots y_m$  be a prime relation where  $x_i, y_j$  are codewords. The existence of such a factorization is equivalent to the existence of a path in  $\mathcal{C}$ :

$$\begin{aligned} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_s &\xrightarrow{(e_1)} \begin{bmatrix} q_{01} \\ 0 \end{bmatrix} \dots \begin{bmatrix} q_{0j_0} \\ 0 \end{bmatrix} \xrightarrow{(e_2)} \begin{bmatrix} 0 \\ p_{11} \end{bmatrix} \dots \begin{bmatrix} 0 \\ p_{1i_1} \end{bmatrix} \xrightarrow{(e_3)} \begin{bmatrix} q_{11} \\ 0 \end{bmatrix} \dots \begin{bmatrix} q_{1j_1} \\ 0 \end{bmatrix} \xrightarrow{(e_4)} \begin{bmatrix} 0 \\ p_{21} \end{bmatrix} \dots \\ &\xrightarrow{(e_{k-2})} \begin{bmatrix} 0 \\ p_{r1} \end{bmatrix} \dots \begin{bmatrix} 0 \\ p_{ri_r} \end{bmatrix} \xrightarrow{(e_{k-1})} \begin{bmatrix} q_{r1} \\ 0 \end{bmatrix} \dots \begin{bmatrix} q_{rj_r} \\ 0 \end{bmatrix} \xrightarrow{(e_k)} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_t. \end{aligned}$$

In this path, we denote by  $e_i$  the edges going from an upper-zero state to a lower-zero one or the converse. Note that this path encodes two paths in the automaton  $\mathcal{A}$ . One is read on the upper track, the other one on the lower track. The label of any path read on the upper (or lower track) going from 0 to 0 without going through 0 in between belongs to  $X$ . Hence

$$\begin{aligned} i_1 + \dots + i_r + 1 &= n \\ j_0 + j_1 + \dots + j_r + 1 &= m. \end{aligned}$$

By renumbering the lower coefficients  $p_{ij}$  of the upper-zero states of this path  $p_1$  to  $p_n$ , and the upper coefficients  $q_{ij}$  of the lower-zero states of this path  $q_1$  to  $q_m$ , the label of each part of this path going from a state  $\begin{bmatrix} 0 \\ p_{i-1} \end{bmatrix}$  to a state  $\begin{bmatrix} 0 \\ p_i \end{bmatrix}$  is labelled by  $x_i$ . The label of each part of this path going from a state  $\begin{bmatrix} q_{j-1} \\ 0 \end{bmatrix}$  to a state  $\begin{bmatrix} q_j \\ 0 \end{bmatrix}$  is labelled by  $y_j$ .

By the definition of the sets  $S_{e_i e_{i+1}}$  and the sets  $S_{e_i e_{i+1} e_{i+2}}$ , we get that all  $x_i$  and all  $y_j$  belong to a same part of the canonical coding partition.

Conversely, we prove that if two words  $x$  and  $y$  belong to a same component of the partition, then there is a finite chain of words  $x = w_0, w_1, \dots, w_n = y$  such that  $w_i$  and  $w_{i+1}$  belong to a same prime relation for  $0 \leq i < n$ .

Let  $q_1, q_2$  be two non null states in  $Q$ . We first show that if two words  $y, y' \in S\begin{bmatrix} q_1 \\ 0 \end{bmatrix}\begin{bmatrix} q_2 \\ 0 \end{bmatrix}$ , then there is a finite chain of words  $y = w_0, w_1, \dots, w_n = y'$  such that  $w_i$  and  $w_{i+1}$  belong to a same prime relation for  $0 \leq i < n$ .

Since  $y, y' \in S\begin{bmatrix} q_1 \\ 0 \end{bmatrix}\begin{bmatrix} q_2 \\ 0 \end{bmatrix}$ , there are in  $\mathcal{C}$  two paths labelled  $xyz$  and  $x'y'z'$ , with  $x, x', z, z' \in A^*$ , containing respectively an edge labelled by  $y$  and an edge labelled by  $y'$ , with the following form:

$$\begin{aligned} \begin{bmatrix} q_1 \\ 0 \end{bmatrix} &\xrightarrow{x} \begin{bmatrix} q_{11} \\ 0 \end{bmatrix} \xrightarrow{y} \begin{bmatrix} q_{12} \\ 0 \end{bmatrix} \xrightarrow{z} \begin{bmatrix} q_2 \\ 0 \end{bmatrix}, \\ \begin{bmatrix} q_1 \\ 0 \end{bmatrix} &\xrightarrow{x'} \begin{bmatrix} q'_{11} \\ 0 \end{bmatrix} \xrightarrow{y'} \begin{bmatrix} q'_{12} \\ 0 \end{bmatrix} \xrightarrow{z'} \begin{bmatrix} q_2 \\ 0 \end{bmatrix}. \end{aligned}$$

Since  $\begin{bmatrix} q_1 \\ 0 \end{bmatrix}$  is accessible from  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  and  $\begin{bmatrix} q_2 \\ 0 \end{bmatrix}$  is co-accessible from  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$ , these paths can be extended in  $\mathcal{C}$  by a shortest path from  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  to  $\begin{bmatrix} q_1 \\ 0 \end{bmatrix}$  labelled by a word  $u$ , and by a shortest path from  $\begin{bmatrix} q_2 \\ 0 \end{bmatrix}$  to  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$  labelled by a word  $w$ . The resulting paths are

$$\begin{aligned} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_s &\xrightarrow{u} \begin{bmatrix} q_1 \\ 0 \end{bmatrix} \xrightarrow{x} \begin{bmatrix} q_{11} \\ 0 \end{bmatrix} \xrightarrow{y} \begin{bmatrix} q_{12} \\ 0 \end{bmatrix} \xrightarrow{z} \begin{bmatrix} q_2 \\ 0 \end{bmatrix} \xrightarrow{v} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_t, \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix}_s &\xrightarrow{u} \begin{bmatrix} q_1 \\ 0 \end{bmatrix} \xrightarrow{x'} \begin{bmatrix} q'_{11} \\ 0 \end{bmatrix} \xrightarrow{y'} \begin{bmatrix} q'_{12} \\ 0 \end{bmatrix} \xrightarrow{z'} \begin{bmatrix} q_2 \\ 0 \end{bmatrix} \xrightarrow{v} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_t. \end{aligned}$$

Let for instance  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s \xrightarrow{u_1} \begin{bmatrix} q \\ 0 \end{bmatrix}$  be the first edge of the path  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s \xrightarrow{u} \begin{bmatrix} q_1 \\ 0 \end{bmatrix}$ . Hence  $u_1$  and  $y$  belong to a same prime relation, and  $u_1$  and  $y'$  belong to a same prime relation.

Let now  $x$  and  $y$  be two words in  $S_{ef}$ , where

$$e = \begin{bmatrix} 0 \\ p_i \end{bmatrix} \rightarrow \begin{bmatrix} q_i \\ 0 \end{bmatrix} \text{ and } f = \begin{bmatrix} q_j \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ p_j \end{bmatrix}.$$

Let us consider the first case in the definition of  $S_{ef}$ . For instance, one can assume that

$$\begin{aligned} x &\in E\begin{bmatrix} 0 \\ p_1 \end{bmatrix}\begin{bmatrix} q_1 \\ 0 \end{bmatrix} \cdot L\begin{bmatrix} q_1 \\ 0 \end{bmatrix}\begin{bmatrix} q_2 \\ 0 \end{bmatrix} \cdot E\begin{bmatrix} q_2 \\ 0 \end{bmatrix}\begin{bmatrix} 0 \\ p_2 \end{bmatrix}, \\ y &\in S\begin{bmatrix} q_1 \\ 0 \end{bmatrix}\begin{bmatrix} q_2 \\ 0 \end{bmatrix}. \end{aligned}$$

It follows that there is in  $\mathcal{C}$  a path labelled by  $x$  containing an edge labelled by  $y' \in S\begin{bmatrix} q_1 \\ 0 \end{bmatrix}\begin{bmatrix} q_2 \\ 0 \end{bmatrix}$  which has the following form:

$$\begin{bmatrix} 0 \\ p_1 \end{bmatrix} \rightarrow \begin{bmatrix} q_1 \\ 0 \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} q_{1i} \\ 0 \end{bmatrix} \xrightarrow{y'} \begin{bmatrix} q_{1(i+1)} \\ 0 \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} q_2 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ p_2 \end{bmatrix}.$$

Since  $\begin{bmatrix} 0 \\ p_1 \end{bmatrix}$  is accessible from  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  and  $\begin{bmatrix} 0 \\ p_2 \end{bmatrix}$  is co-accessible from  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$ , this path can be extended in  $\mathcal{C}$  by a shortest path from  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s$  to  $\begin{bmatrix} 0 \\ p_1 \end{bmatrix}$  labelled by a word  $u$  and, by a path from  $\begin{bmatrix} 0 \\ p_2 \end{bmatrix}$  to  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}_t$  labelled by a word  $w$ . The resulting path is

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}_s \xrightarrow{u} \begin{bmatrix} 0 \\ p_1 \end{bmatrix} \rightarrow \begin{bmatrix} q_1 \\ 0 \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} q_{1i} \\ 0 \end{bmatrix} \xrightarrow{y'} \begin{bmatrix} q_{1(i+1)} \\ 0 \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} q_2 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ p_2 \end{bmatrix} \xrightarrow{v} \begin{bmatrix} 0 \\ 0 \end{bmatrix}_t.$$

This defines a prime relation containing the words  $x$  and  $y'$ . Furthermore, we know that there is a word  $w$  such that  $y$  and  $w$  belong to a same prime relation, and  $y'$  and  $w$  belong to a same prime relation.

We consider similarly all cases in the definitions of  $S_{ef}$  and  $S_{efg}$  to conclude that for any two words  $x$  and  $y$  in a such a set, there is a finite chain of words  $w_0 = x, w_1, \dots, w_n = y$  such that  $w_i$  and  $w_{i+1}$  belong to a same prime relation for  $0 \leq i < n$ .

Note that, since the definition of the part  $X_0$  is  $X_0 = X - \bigcup_{i=1}^r X_i$ , the computation of the canonical coding partition cannot be achieved in a polynomial time. The computation of the sets  $S_{ef}$  and  $S_{efg}$  can be performed in polynomial time. Since it is necessary to compute some intersections to get the automata accepting  $X_i$ , the computation of the components  $X_i$  for  $i \neq 0$  also is exponential.

When the code  $X$  is not regular, even when context-free, the canonical coding partition may have an infinite number of classes, as shows the following example.

*Example 1.* Let

$$X = \bigcup_{n \geq 1} (a^n b + a^n b c^n + c^n a^n b).$$

The code  $X$  is context free and its canonical coding partition is  $(X_i)_{i \geq 1}$  with  $X_i = a^i b + a^i b c^i + c^i a^i b$  for  $i \geq 1$  and  $X_0 = \emptyset$ .

It is also possible to get a finite canonical coding partition with non regular classes.

*Example 2.* Let  $X$  be a code, for instance a uniquely decipherable code. Let  $Y$  be the code

$$Y = \{ax, xb \mid x \in X\} + \{a, b\},$$

where  $a, b$  are two symbols which do not appear in the words of  $X$ . The canonical coding partition of  $Y$  is made of a unique class since  $axb = ax \cdot b = a \cdot xb$ . Such code is  $GA$ .

## 4 Maximality

In this section we introduce the notion of maximality of a coding partition. Actually two different notions of maximality can be introduced: maximality with respect to one component (Definition1) and maximality with respect to all the components (Definition2). The main result of this section states that the two notions coincide for regular codes.



**Definition 1.** Let  $P = \{X_1, X_2, \dots\}$  be a non-trivial coding partition of a code  $X \in A^+$ . We say that  $P$  is maximal with respect to the component  $X_i$  if  $\forall w \in A^+$ , the partition  $P' = \{X_1, \dots, X_i \cup \{w\}, \dots\}$  is a coding partition of  $X \cup \{w\}$  iff  $w \in X_i^+$ .

**Definition 2.** A non-trivial coding partition  $P$  is said to be maximal if it is maximal with respect to every component of  $P$ .

*Remark 2.* It is straightforward that if  $P$  is a maximal coding partition of a code  $X$  and  $P' > P$  then also  $P'$  is a maximal coding partition of  $X$ .

**Theorem 5.** Let  $X$  be a code and let  $P = \{X_1, X_2, \dots\}$  be a non-trivial coding partition of  $X$ . If  $P$  is maximal with respect to at least one component, then  $X$  is complete.

*Proof.* Let  $X$  be a code over the alphabet  $A$ , with  $\text{card}(A) \geq 2$  (the case  $\text{card}(A) < 2$  is trivial). We will first prove that, if  $X$  is not complete, then there exists a word  $w \in A^* \setminus X$  such that the partition  $P_1 = \{\{w\}, X_1, X_2, \dots\}$  is a coding partition of  $X \cup \{w\}$ . Indeed, if  $X$  is not complete, there exists a word  $v \in A^*$  such that  $v$  does not belong to  $F(X^*)$ . Let  $a$  be the first letter of  $v$  and let  $b \in A \setminus \{a\}$ . Consider the word  $w = vb^{|v|-1}$ . By construction,  $w$  is unbordered, i.e. no proper prefix of  $w$  is a suffix of  $w$ . Since  $v$  does not belong to  $F(X^*)$ , we have that also  $w$  does not belong to  $F(X^*)$ .

Let us first remark that  $X^+ \cap \{w\}^+ = \emptyset$ . We now prove that every word  $t \in (X \cup \{w\})^*$  admits a unique  $P_1$ -factorization. Indeed, since  $w$  is unbordered, we can uniquely distinguish all occurrences of  $w$  in  $t$ , i.e.  $t$  has a unique factorization of the form

$$t = u_1 w u_2 w \cdots w u_n,$$

with  $n \geq 1$  and  $u_i \in X^*$ , for  $i = 1, \dots, n$ . From this factorization, since  $P$  is a coding partition, we obtain a unique  $P_1$ -factorization of  $t$  and therefore, by definition,  $P_1$  is a coding partition. From this is trivial that  $\forall i P' = \{X_1, \dots, X_i \cup \{w\}, \dots\}$  is still a coding partition and so  $X_i$  is not maximal. This concludes the proof.

The next lemma and its proof is just a little variation of a lemma due to Schutzenberger (see Theorem 7.4 in [5]).

**Lemma 1.** Let  $X \subseteq A^+$  be a regular and complete code and let  $x_1, x_2 \in X^*$ . Then, there exist a word  $v_1 \in X^+$  and a positive integer  $m$  such that for any word  $w \in A^*$ ,  $(v_1 w v_1)^m \in X^+$  where  $v = x_1 v_1 x_2$ .

*Proof.* Since  $X$  is a regular set,  $X^+$  is a regular set too. Let

$$\mathcal{A} = (A, Q, \delta, i, F)$$

be a finite state automaton recognizing  $X^+$ . For any set of states  $S \subseteq Q$  and for any word  $u \in A^+$ , denote by  $Su$  the set  $\{\delta(q, u); q \in S\}$  of states reached

by paths having label  $u$  and starting at any state of  $S$ . Let  $n = \min\{\text{card}(Qu)\}$  with  $u$  ranging over  $A^+$ , and choose  $u$  such that  $n = \text{card}(Qu)$ . Since  $X$  is complete, we have  $xuy \in X^+$  for some  $x, y \in A^*$  and so  $v' := x_1xuyx_2 \in X^+$ . Since  $\text{card}(Qx_1xuyx_2) \leq \text{card}(Qx_1xu)$  and  $Qx_1xu \subseteq Qu$ , it follows that  $\text{card}(Qv') \leq \text{card}(Qu)$ . Thus, by minimality,  $\text{card}(Qv') = n$ . Let  $P = Qv'$ . Since  $Pv' = Qv'v' \subseteq Qv' = P$ , it follows from the minimality of  $n$  that  $Qv'v' = Qv'$  and  $Pv' = P$ ; thus  $v'$  defines a permutation of  $P$ . Thus, put  $v$  a suitable power of  $v'$  and wrote  $v = x_1v_1x_2$  for a certain  $v_1 \in X^+$ , we may assume that  $pv = p$  for all  $p \in P$  and  $Qv = Qv' = P$ . Consider now a word  $w \in A^*$  and let  $z = vvw$ . Again we have  $Pz = Qvwwv \subseteq Qv = P$  and thus  $Pz = P$ . Then for  $m = n!$  we have  $pz^m = p$  for all  $p \in P$ . To prove that

$$z^m = (vww)^m \in X^+,$$

it suffices to show that  $qz^m = qv$  for all  $q \in Q$ . Since  $Qv = P$  and  $pv = p$  for all  $p \in P$ , then  $qvv = qv$ . It follows that  $qz = qvww = qvwwv = qvz$  and therefore that  $qz^m = qvz^m$ . Since  $pz^m = p$  for all  $p \in P$ , we have that  $qvz^m = qv$ . Thus  $qz^m = qv$  as required. This completes the proof.

**Theorem 6.** *Let  $X$  be a regular code and let  $P = \{X_1, X_2, \dots\}$  be a non-trivial coding partition of  $X$ . If  $X$  is complete then  $P$  is maximal.*

*Proof.* Let  $w \in A^+$  and  $i \geq 1$  such that  $P' = \{X_1, \dots, X_i \cup \{w\}, \dots\}$  is a coding partition of  $X' = X \cup \{w\}$ . Since  $P$  is non-trivial,  $\exists x \in X_j \neq X_i$ . By previous lemma there exist  $v_1 \in X^+$  and a positive integer  $m$  such that  $z = (xv_1xwxv_1x)^m \in X^+$ . Since  $x \notin X_i$  the  $P'$  factorization of  $z$  is of the form:

$$z = z_1 \cdots z_{s_1} w z_{s_1+1} \cdots z_{s_m} w z_{s_m+1} \cdots z_t$$

where  $z_h$ ,  $1 \leq h \leq t$  are the blocks of the factorization. But  $z \in X^+$  so there exists a factorization without  $w$  that is again a  $P'$  factorization. By the uniqueness of the  $P'$  factorization the block corresponding to  $w$  must be the same and so  $\exists y_1, y_2, \dots, y_k \in X_i$  s.t.  $w = y_1y_2 \cdots y_k$ . This shows that  $P$  is maximal.

From Theorem 5 and Theorem 6 we get the following corollary.

**Corollary 1.** *Let  $P = \{X_1, X_2, \dots\}$  be a non-trivial coding partition of a regular code  $X$ . If  $P$  is maximal with respect to a component  $X_i$ , then  $P$  is maximal.*

## 5 UD codes versus GA codes

In this section we consider an application of previous results to maximal  $UD$  codes. By definition, a  $UD$  code  $X$  is maximal if any code  $Y$  containing strictly  $X$  is ambiguous. We here prove that, if a regular  $UD$  code  $X$  is maximal, then any code  $Y$  containing strictly  $X$  is globally ambiguous. Moreover, if  $X$  is a finite maximal  $UD$  code, we prove that for a given word  $v \in A^+$ , there exists a prime relation involving all the elements of  $X \cup \{v\}$ .

A generalization of this result to the case of non- $UD$  codes, is given at the end of the paper.

**Theorem 7.** *Let  $X \subseteq A^+$  be a maximal UD code. If  $X$  is regular then, for all  $v \in A^+$  such that  $v \notin X^+$ ,  $X \cup \{v\}$  is GA.*

The proof is an immediate consequence of the following proposition that has an independent interest.

**Proposition 1.** *Let  $X \subseteq A^+$  be a regular code and let  $P = \{X_1, X_2, \dots\}$  be a non-trivial coding partition of  $X$ . If  $P$  is maximal then, for all  $v \in A^+$  such that  $v \notin X^+$ ,  $X \cup \{v\}$  is GA.*

*Proof.* Let, by contradiction,  $X' := X \cup \{v\}$  be non-GA, let  $P' = \{X'_1, X'_2, \dots\}$  be a non-trivial coding partition of  $X'$  and suppose, without loss of generality,  $v \in X'_1$ . We claim that  $|X'_1| > 1$ . Indeed if  $X'_1 = \{v\}$  then  $v$  is in the unambiguous component of  $X'$  and so there are not relations between  $v$  and the others words of  $X$ . Then  $\forall i \geq 1 \{X_1, \dots, X_i \cup \{v\}, \dots\}$  is a coding partition of  $X \cup \{v\}$  against the hypothesis that  $P$  is maximal. Then  $P'' = \{X'_1 \setminus \{v\}, X'_2, \dots\}$  is a coding partition of  $X$ . From Theorem 5 and Theorem 6 one derives that if a coding partition of a code is maximal then any other non-trivial coding partition of that code is maximal too. Then  $P''$  is maximal so  $v \in (X'_1 \setminus \{v\})^+$  and we have a contradiction.

In the case the code  $X$  is *finite* we can derive stronger results. Recall that a code  $X$  is called a *base* if  $X$  is a minimal set of generators of  $X^*$ .

**Theorem 8.** *Let  $C \subseteq A^+$  be a finite maximal prefix UD code. If  $C \neq A$  then there exists a word  $v \in A^+$  such that  $C' := C \cup \{v\}$  has the following properties:*

- $C'$  is a base
- $C'$  is GA
- there exists a prime relation involving all the elements of  $C'$ , i.e. a relation  $x_1 x_2 \cdots x_s = x_{s+1} x_{s+2} \cdots x_t$  such that  $\{x_1, x_2, \dots, x_t\} = C'$ .

*Proof.* Let  $C = \{c_1, c_2, \dots, c_n\}$  and let  $u := c_1 c_2 \cdots c_n$ . By hypothesis there is  $c_{i_0} \in C$  with  $|c_{i_0}| > 1$ , and let  $w$  be a prefix of  $c_{i_0}$  s.t.  $|w| = |c_{i_0}| - 1$ . Let us put  $v := uw$  and  $C' := C \cup \{v\}$ . We claim that  $C'$  is a base. Indeed since  $|v| > |c_i|$ ,  $1 \leq i \leq n$ , it is sufficient to show that  $v \notin C^+$ . If, by contradiction,  $v \in C^+$ , being  $C^+$  right unitary (see [2]), we have  $w \in C^+$  with  $w$  prefix of  $c_{i_0}$ , and this is a contradiction because  $C$  is a prefix code. Finally since  $C$  is maximal  $C'$  is not UD so there is a prime relation involving  $v$ . This relation by definition of  $v$ , being  $C$  a prefix code, must have the form  $v x_1 \cdots x_s = c_1 \cdots c_n y_1 \cdots y_t$  for some  $x_i, y_j \in C'$ .

The previous result can be extended to any finite maximal UD code.

**Theorem 9.** *Let  $C \subseteq A^+$  be a finite maximal UD code. If  $C \neq A$  then there exists a word  $v \in A^+$  such that  $C' := C \cup \{v\}$  has the following properties:*

- $C'$  is a base

- $C'$  is GA
- there exists a prime relation involving all the elements of  $C'$ , i.e. a relation  $x_1x_2 \cdots x_s = x_{s+1}x_{s+2} \cdots x_t$  such that  $\{x_1, x_2, \dots, x_t\} = C'$ .

*Proof.* If  $C = \{c_1, c_2, \dots, c_n\}$  is a prefix set then the thesis holds because of previous theorem. If  $C$  is not a prefix set then  $C$  is not right complete and let  $w \in A^+$  s.t.  $w$  is not right completable. Of course also  $w^m$  is not right completable and we can choose  $m \geq 2$  in such a way that  $|w^m| > |c_i| + |w| \forall c_i \in C$ . Now we put  $w_1 := w^m w'$  with  $w' \in A^*$  s.t.  $w_1$  is unbordered. Since  $C$  is maximal then put  $C' := C \cup \{w_1\}$  there exists a prime relation  $x_1x_2 \cdots x_s w_1 x_{s+1} \cdots x_l = y_1 \cdots y_k$ ,  $x_i, y_i \in C'$ ,  $s \geq 1, l \geq s, k \geq 2$ . Let  $p \geq 1$  the first index s.t.  $|y_1 \cdots y_p| > |x_1 \cdots x_s w|$ : by choice of  $m$  and  $w_1$ ,  $|y_1 \cdots y_p| < |x_1 \cdots x_s w^m|$  so  $y_1 \cdots y_p = x_1 \cdots x_s w^q u$ , with  $1 \leq q < m$  and, since  $w$  is not right completable,  $u \in A^+$ . Now we put  $v := w^q u c_1 \cdots c_n z w^{m-q-1} w'$  with  $z = u^{-1} w$ . We have the relation  $x_1 x_2 \cdots x_s v x_{s+1} \cdots x_l = y_1 \cdots y_p c_1 \cdots c_n y_{p+1} \cdots y_k$  that is clearly prime. Finally, by definition,  $v \notin C^+$  and, by a length argument, one has that  $C \cup \{v\}$  is a base and the proof is complete.

*Remark 3.* We observe as consequence of Theorem 7 that, if  $X$  is a base and it is not GA, then any regular set  $Y \subsetneq X$  is not a maximal UD code.

**Theorem 10.** *Let  $X \subseteq A^+$  be a non-GA finite code that is a base. If  $X$  is complete then there exists a word  $v \in A^+$  such that  $X' := X \cup \{v\}$  has the following properties:*

- $X'$  is a base
- $X'$  is GA
- there exists a prime relation involving all the elements of  $X'$ , i.e. a relation  $x_1x_2 \cdots x_s = x_{s+1}x_{s+2} \cdots x_t$  such that  $\{x_1, x_2, \dots, x_t\} = X'$ .

*Proof.* We recall that a code  $Y \subseteq A^+$  is right complete iff  $Y \setminus YA^+$  is a maximal prefix UD code (see [2]). Let  $P$  be a non-trivial coding partition of  $X$  then, by Theorem 6,  $P$  is maximal. Because of Theorem 8 we can suppose that  $X$  is not a prefix UD code and then  $\emptyset \neq X \setminus XA^+ \subsetneq X$ . Moreover, because of previous remark,  $X \setminus XA^+$  is not a maximal prefix UD code and so  $X$  is not right complete. Then there exists  $w \in A^+$  s.t.  $w$  is not right completable and we can proceed like in the previous theorem.

## References

1. M.-P. BÉAL AND D. PERRIN, *Codes, unambiguous automata and sofic systems*, Theoret. Comput. Sci., 356 (2006), pp. 6–13.
2. J. BERSTEL AND D. PERRIN, *Theory of codes*, vol. 117 of Pure and Applied Mathematics, Academic Press Inc., Orlando, FL, 1985. <http://www-igm.univ-mlv.fr/~berstel/LivreCodes/Codes.html>.
3. F. BURDERI AND A. RESTIVO, *Coding partitions*, Discret. Math. Theor. Comput. Sci. (to appear).

4. M. DALAI AND R. LEONARDI, *Non prefix-free codes for constrained sequences*, in International Symposium on Information Theory, 2005. ISIT 2005, IEEE, 2005, pp. 1534–1538.
5. S. EILENBERG, *Automata, Languages and Machines, Vol. A*, Academic Press, New York, 1974.
6. G. GÖNENÇ, *Unique decipherability of codes with constraints with application to syllabification of Turkish words*, in COLING 1973: Computational And Mathematical Linguistics: Proceedings of the International Conference on Computational Linguistics, vol. 1, 1973, pp. 183–193.
7. F. GUZMÁN, *Decipherability of codes*, J. Pure Appl. Algebra, 141 (1999), pp. 13–35.
8. W. P. J. KARHUMÄKI AND W. RYTTER, *Generalized factorizations of words and their algorithmic properties*, Theoret. Comput. Sci., 218 (1999), pp. 123–133.
9. A. LEMPEL, *On multiset decipherable codes*, IEEE Trans. Inform. Theory, 32 (1986), pp. 714–716.
10. A. RESTIVO, *A note on multiset decipherable codes*, IEEE Trans. Inform. Theory, 35 (1989), pp. 662–663.
11. J. SAKAROVITCH, *Éléments de théorie des automates*, Vuibert, Paris, 2003. English translation to appear, Cambridge University Press.