INFORMATIQUE THÉORIQUE ET APPLICATIONS

J. DEVOLDER
E. TIMMERMAN

Finitary codes for biinfinite words

Informatique théorique et applications, tome 26, nº 4 (1992), p. 363-386.
<http://www.numdam.org/item?id=ITA_1992__26_4_363_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 26, n° 4, 1992, p. 363 à 386)

FINITARY CODES FOR BIINFINITE WORDS (*)

by J. DEVOLDER ([1]) and E. TIMMERMAN ([2])

Abstract. – *The aim of decoding, or factorizing in a single way, biinfinite words with a finitary language leads to define, according to the definition of factorizations, two distinct notions of finitary codes for biinfinite words we call "biω-codes" and "**Z**-codes". These codes are respectively close to the precircular codes and to the circular codes. The notion of biω-code is weaker but seems to be more suitable for biinfinite words. Indeed, biω-codes are characterized using coding morphisms, as are usual codes and codes for infinite words. **Z**-codes are rather codes for **Z**-words. The relationships between all these finitary codes are studied, and characteristic properties of biω-codes and **Z**-codes are given.*

Résumé. – *Le décodage ou la factorisation unique de mots biinfinis à l'aide d'un langage finitaire conduit à définir deux notions différentes de code finitaire pour les mots biinfinis selon la définition des factorisations. Nous appelons ces codes « biω-codes » et « **Z**-codes ». Ces codes sont respectivement proches des codes précirculaires et des codes circulaires. La notion de biω-code, moins restrictive, semble mieux appropriée aux mots biinfinis. On peut en effet la caractériser à l'aide de morphismes de codage comme le sont les codes usuels et les codes pour les mots infinis. Les **Z**-codes sont quant à eux mieux adaptés au codage des **Z**-mots. Les relations entre tous ces codes finitaires sont étudiées, et des propriétés caractéristiques des biω-codes et des **Z**-codes sont montrées.*

INTRODUCTION

The notion of "code language" means a single factorization, whenever it exists, of each word using words of the considered language. This notion has been very studied in the domain of finite words and finitary languages (sets of finite words).

If one deals with infinite or biinfinite words, two distinct approachs may be used: the first one is to consider finite factorizations (finite sequences of

(*) Received January 1991, final version November 1991.

This work is supported by the P.R.C. Mathématiques et Informatique, and by the EBRA project ALGEBRAIC and SYNTACTIC METHODS in COMPUTER SCIENCE.

([1]) C.N.R.S.-U.R.A. 369, L.I.F.L., Université de Lille-I, U.F.R. M.P.A., bât. M2, 59655 Villeneuve-d'Ascq Cedex, France.

([2]) C.N.R.S.-U.R.A. 369, L.I.F.L., Université de Lille-I, I.U.T.-A., Dépt. Informatique, B.P. n° 179, 59653 Villeneuve-d'Ascq Cedex, France.

0988-3754/92/04 363 24/$4.40/

words) using finite, infinite and biinfinite words (as done by Do Long Van *et al.*, [10]), and the second one is to consider infinite factorizations only using finite words. That leads to the notion of finitary code for infinite (resp. biinfinite) words. This second way has been used by Staiger for the study of the "ifl-codes" (codes for infinite words) that we call "ω-codes". In this paper, following this way, we are interested in the study of finitary codes for binfinite words.

Given a finitary language C, the first point is to define a factorization of a biinfinite words w by C (we say a C-factorization of w). For finite or infinite words, two equivalent definitions may be given:

1. A C-factorization of w is a sequence of words of C, the concatenation of them giving w.

2. A C-factorization of w is an increasing sequence of integers indicating letters positions where w is cut by words of C, the concatenation of them giving w.

The fact that a biinfinite word is an equivalent class of **Z**-words (mappings from **Z**, the set of relative integers, into a finite alphabet) implies that these two definitions are no more equivalent (since with the second definition one has to choose a representative of the biinfinite word). That leads to two distinct notions of finitary code for biinfinite words: the "biω-codes" (the weak notion, from definition 1) and the "**Z**-codes" (the strong notion, from definition 2). The simplest example is the language $C=\{aa\}$ which is a biω-code but not a **Z**-code: the only biinfinite word to be factorized: ${}^{\omega}a^{\omega}$ has the single C-factorization $(x_i=aa)_{i\in\mathbf{Z}}$ with definition 1, and the two non-equivalent C-factorizations $(2i)_{i\in\mathbf{Z}}$ and $(2i+1)_{i\in\mathbf{Z}}$ with definition 2. We call C-decompositions these C-factorizations with definition 2.

In the finitary case as well as for infinite words, the codes (and the ω-codes) may be defined or characterized with coding morphisms. A coding morphism is a bijection φ between an alphabet X and a language C, extended to X^* onto C^* (and to X^ω onto C^ω), and the language is a code (resp. an ω-code) iff φ (so extended) is injective. This mapping φ can also be extended to ${}^{\omega}X^{\omega}$ onto ${}^{\omega}C^{\omega}$, and then it comes that C is a biω-code iff φ is injective, which indicates that this notion of bi-ω codes seems to be the good one for biinfinite words.

The notion of **Z**-code (implicitely contained in the works of Beal, Beauquier, Blanchard and Hansel, Restivo,...) is more restrictive. It is in fact the suitable notion of finitary codes for **Z**-words.

These four classes of finitary codes: codes, ω-codes, biω-codes and **Z**-codes are defined in the first section. They form a strict decreasing hierarchy. The problem of defining finitary codes for words in ${}^{\infty}A^{\infty}$ (finite, left-infinite, right-infinite, biinfinite words on A) is then solved: the notion of biω-code is the good notion.

In the second part, we show how codes, precircular codes and circular codes deal with biinfinite words. This leads to see that the notion of biω-code does not coincide with some known code notion and that biω-codes (resp. **Z**-codes) are particular precircular (resp. circular) codes.

In the third part, we give a characterization of **Z**-codes: we show that a language is a **Z**-code iff it is a pure biω-code.

In the fourth section, we study biω-codes. Technical properties of biω-codes and conditions for precircular codes to be biω-codes are given.

The rational case is studied in the last section. In this case, the notion of circular code and the notion of **Z**-code coincide. The biω-codes are precircular codes, but the converse is false, even in the finite case. A technical characterization of biω-codes gives, as a consequence, the decidability for testing whether a rational language is a biω-code.

NOTATIONS AND BASIC DEFINITIONS

In the following A is a finite alphabet, A^* stands for the set of all (finite) words over A, A^+ denotes the language A^*-ε, where ε is the empty word. The length of the word u is denoted by $|u|$.

Two words x and x' are said to be *conjugate* if there exist u and v such that $x=uv$ and $x'=vu$. A word $z\in A^+$ is *primitive* if $z=u^n$ implies $n=1$. If $z=u^n$ with $n>1$, z is said to be imprimitive. For every word $x\in A^+$ there exists a unique primitive word $\sqrt{x}$ and an integer n for which $x=(\sqrt{x})^n$; $\sqrt{x}$ is referred to as the *primitive root* of x.

We shall also consider A^{ω} and ${}^{\omega}A$ which are the sets of right (resp. left)-infinite words over A.

A (right) infinite word $w=(w_n)_{n\in\mathbb{N}}$ is said to be *ultimately periodic* if

$\exists p\in\mathbb{N}\ \exists t>0$ such that $\forall k>0\ (\forall n\, 0\leq n<t)\ w_{p+n+kt}=w_{p+n}$. We denote this word by $w_0\ldots w_{p-1}(w_p\ldots w_{p+t-1})^{\omega}$. It is said to be *periodic* if p can chosen equal to 0; that is to say $w=(w_0\ldots w_{t-1})^{\omega}$.

Let $\mathbf{Z}$ denote the set of relative integers, $A^{\mathbf{Z}}$ stands for the set of $\mathbf{Z}$-sequences of elements of A. The elements of $A^{\mathbf{Z}}$ are called *$\mathbf{Z}$-words. Biinfinite words* [3, 18] over A are the equivalence classes for the shift relation over $A^{\mathbf{Z}}$:

$$(w_n)_{n \in \mathbf{Z}} \simeq (v_n)_{n \in \mathbf{Z}} \iff \exists p \in \mathbf{Z},\ \forall n \in \mathbf{Z},\ w_n = v_{n+p}.$$

The biinfinite word defined by $(w_n)_{n \in \mathbf{Z}}$ is written $\ldots w_0 w_1 \ldots$
The set of biinfinite words over A is denoted by ${}^{\omega}A^{\omega}$.

A $\mathbf{Z}$-word (resp. biinfinite word defined by) $(w_n)_{n \in \mathbf{Z}}$ is said to be *right-ultimately periodic* if $\exists p \in \mathbf{Z}\ \exists t > 0$ such that

$$\forall k > 0\ (\forall n\, 0 \leqq n < t)\ w_{p+n+kt} = w_{p+n}.$$

We denote such a biinfinite word by $\ldots w_{p-1} (w_p \ldots w_{p+t-1})^{\omega}$.

A *left-ultimately periodic* $\mathbf{Z}$-word (resp. biinfinite word) is defined in an analogous way.

A *bi-ultimately periodic* $\mathbf{Z}$-word (resp. biinfinite word) is a $\mathbf{Z}$-word (resp. biinfinite word) which is both left-ultimately periodic and right-ultimately periodic.

A $\mathbf{Z}$-word (resp. biinfinite word defined by) $(w_n)_{n \in \mathbf{Z}}$ is said to be *periodic* if $\exists t > 0$ such that $\forall k\ (\forall n\ 0 \leqq n < t)\ w_{n+kt} = w_n$. Such a biinfinite word is written ${}^{\omega}(w_1 \ldots w_t)^{\omega}$.

Given a language $C \subset A^+$ the submonoid generated by C is the language $C^* = \{ v_1 \ldots v_n \mid n \geqq 0\ v_i \in C \text{ for } 1 \leqq i \leqq n \}$, C^{ω} stands for the set of infinite words obtained by concatenation of an infinite sequence of words of C: $C^{\omega} = \{ v_0 v_1 v_2 \ldots \mid v_i \in C \text{ for } i \geqq 0 \}$. In a same way ${}^{\omega}C = \{ \ldots v_2 v_1 v_0 \mid v_i \in C \text{ for } i \geqq 0 \}$ and ${}^{\omega}C^{\omega} = \{ \ldots v_{-2} v_{-1} v_0 v_1 v_2 \ldots \mid v_i \in C \text{ for } i \in \mathbf{Z} \}$.

For $U,\ V \subset A^{\omega}$ and $U',\ V' \subset {}^{\omega}A$ we define: $UV^{-1} = \{ t \in A^* \mid \exists v \in V\ tv \in U \}$ and $V'^{-1} U' = \{ s \in A^* \mid \exists v \in V'\ vs \in U' \}$.

Given a language $C \subset A^+$ we shall often consider a bijection φ between an alphabet X and the language C. This mapping can be extended to X^* as a morphism $\varphi : X^* \to C^*$. This morphism is said to be a *coding morphism for C* (even if it is not injective). The mapping φ can also be extended to X^{ω} ($\varphi(z_0 z_1 \ldots)$ is the word $\varphi(z_0)\varphi(z_1)\ldots$), and also to ${}^{\omega}X$ and ${}^{\omega}X^{\omega}$ in an analogous manner. This extension is denoted also by φ.

1. FACTORIZATIONS AND CODES FOR FINITE, INFINITE AND BIINFINITE WORDS. DECOMPOSITIONS OF Z-WORDS. Z-CODES

In the sequel, **C** (the language C will be the "code") *is always a subset of* A^+. That means that we consider only notions of code for which the codes are constitued of finite words. Another notion of code for biinfinite words has been defined by Do Long Van [9, 10]; the codes used by Do Long Van have always infinite and biinfinite words as elements.

A. Codes for finite words: codes [4]

Definitions and recalls: For any word w in A^+, $(v_0, \ldots, v_{n-1})$ is a *C-factorization* of w if $n \geqq 1$, $w = v_0 \ldots v_{n-1}$ and $v_i \in C$ for $0 \leqq i < n$.

To have a factorization of w, it is equivalent to give a finite increasing sequence of integers $i_0, i_1, \ldots, i_n$ where $1 = i_0 < i_1 < \ldots < i_n = |w| + 1$. This gives a C-factorization if $v_j = w_{ij} \ldots w_{i(j+1)-1}$ (denoted in the sequel by $w[i_j, i_{j+1}[$) belongs to C for every j. Let $\varphi : X^* \to A^*$ by a coding morphism for C, the set of C-factorizations of $u \in A^*$ may be represented by $\varphi^{-1}(u)$.

A language $C \subset A^+$ is said to be a *code* if every word in C^+ has a *unique C-factorization*. In other terms: a language C is a *code* if and only if

(1) $$\forall u, v \in C \quad u C^* \cap v C^* \neq \varnothing \quad \Rightarrow \quad u = v.$$

In terms of coding morphisms, we can say: C is a *code* iff $\varphi : X^* \to A^*$ *is injective.*

B. Codes for infinite words: ω-codes [20]

Definitions: For any word w in A^ω, $(v_i)_{i \in \mathbb{N}}$ is a *C-factorization* of w if $w = v_0 v_1 v_2 \ldots$ and $v_i \in C$ for $i \geqq 0$.

To have a factorization of w, it is equivalent to give a strictly increasing infinite sequence of integers $i_0, i_1, \ldots, i_n, \ldots$ where $i_0 = 1$. This gives a C-factorization if $v_j = w[i_j, i_{j+1}[$ belongs to C for every $j \geqq 0$. The set of C-factorizations of a word $w \in A^\omega$ may be represented by $\varphi^{-1}(w)$, if $\varphi : X^\omega \to C^\omega$ denotes a coding morphism for C.

A language $C \subset A^+$ is said to be an *ω-code* if every word in C^ω has a *unique C-factorization*. In other terms: a language C is an *ω-code* if and only if

(1′) $$\forall u, v \in C \quad u C^\omega \cap v C^\omega \neq \varnothing \quad \Rightarrow \quad u = v.$$

This definition can be expressed in terms of morphisms: let φ be any *coding morphism for C*, C is an ω-*code* if and only if $\varphi : X^\omega \to C^\omega$ *is injective.*

C. Codes for biinfinite words: biω-codes [8]

We are going to propose a coding theory for biinfinite words. Of course, a code for biinfinite words (we say a biω-code) is a language C for which *every biinfinite word has at most one C-factorization.* A problem appears when we have to say what means "has at most one C-factorization". We suggest the following definition for a C-factorization, and according with this definition we obtain the notion of *biω-code.*

DEFINITIONS: We define *C-factorizations on* ${}^\omega A^\omega$ as the equivalence classes for the shift relation over $C^{\mathbf{Z}}$:

$$(c_n)_{n \in \mathbf{Z}} \simeq (d_n)_{n \in \mathbf{Z}} \quad \Leftrightarrow \quad \exists p \in \mathbf{Z}, \quad \forall n \in \mathbf{Z}, \quad c_n = d_{n+p}.$$

The sequence $(c_i)_{i \in \mathbf{Z}}$ represents a *C-factorization of* $w \in {}^\omega A^\omega$ if w is obtained by concatenation of the $c_n : w = \ldots c_{-1} c_0 c_1 \ldots$

A language $C \subset A^+$ is said to be a *biω-code* if every word in ${}^\omega C^\omega$ has a *unique C-factorization.*

This definition can be expressed in terms of morphisms: let φ be any *coding morphism for C.* There exists a bijection between the set of all the C-factorizations and ${}^\omega X^\omega$. In the sequel, when necessary, we shall denote a C-factorization by a word of ${}^\omega X^\omega$.

A language C is a *biω-code* if and only if $\varphi : {}^\omega X^\omega \to {}^\omega C^\omega$ *is injective.* The unique C-factorization of $w \in {}^\omega C^\omega$ is then represented by $\varphi^{-1}(w)$.

Clearly the notion of biω-code cannot be put on an analogous form to (1) and (1').

Remark: Let $u, v \in A^+$. If the primitive roots of u and v are conjugate, the words u and v cannot belong to a same biω-code: for example, the code $\{ab, ba\}$ is not a biω-code.

Examples 1.1:

- The singleton $\{u\}$ is a biω-code if $u \neq \varepsilon$.
- The languages $\{a^2\}$, $\{a, bab\}$, $\{ba^3, ba^2ba, a^2ba^2\}$ are finite biω-codes.
- For every n, the language $\{b^n\} \cup ab(a^2b)^*$ is an infinite biω-code.

DEFINITION: We call *C-factorization of a* $\mathbf{Z}$-*word u* every C-factorization of the class of u in ${}^\omega A^\omega$.

A C-factorization of a biinfinite word w can be expressed in terms of indices. This supposes that a representative of $w:(w_i)_{i \in \mathbf{Z}}$ has been chosen. To have a factorization of w (or equivalently of $(w_i)_{i \in \mathbf{Z}}$), one can give a strictly increasing biinfinite sequence of integers..., i_{-2}, i_{-1}, i_0, i_1,... This gives a C-factorization if $v_j = w[i_j, i_{j+1}[$ belongs to C for every j. To avoid the problems due to the shift on the indices of the sequence $(i_n)_{n \in \mathbf{Z}}$, one can impose $i_0 \geqq 0$ and $i_{-1} < 0$. But one can see (example 1.2) that uniqueness of the C-factorization of w does not correspond to uniqueness of such an increasing sequence.

Example 1.2: Let $C = \{aa\}$, the word ${}^{\omega}a^{\omega}$ has a single $\{aa\}$-factorization: ${}^{\omega}a^{\omega} = {}^{\omega}(aa)^{\omega}$.

But one can see that:

$$\ldots a_1\, a_2\, a_3\, a_4\, a_5\, a_6 \ldots = \ldots a_1\, a_2 \,.\, a_3\, a_4 \,.\, a_5\, a_6 \ldots a_1 \,.\, a_2\, a_3 \,.\, a_4\, a_5 \,.\, a_6 \ldots$$

(where $a_i = a$ for every i). The single $\{aa\}$-factorization of ${}^{\omega}a^{\omega}$ gives two ways to factorize its representative.

So, for a given factorization of a $\mathbf{Z}$-word (resp. a biinfinite word) w, one can sometimes consider several ways to factorize w (resp. any of its representatives). We shall give another notion of code for biinfinite words, the notion of *$\mathbf{Z}$-code*, stronger than the previous notion of biω-code, this other notion requiring a single way for factorizing $\mathbf{Z}$-words.

D. Decompositions of Z-words. Notion of Z-code

The notion of *decomposition over* $A^{\mathbf{Z}}$ is defined to count the different ways to factorize a $\mathbf{Z}$-word by a given C-factorization of this word.

DEFINITIONS: A *decomposition over* $A^{\mathbf{Z}}$ is a strictly increasing sequence of relative integers: $(d_n)_{n \in \mathbf{Z}}$ such that $d_0 \geqq 0$ and $d_{-1} < 0$. Let us denote by D the set of decompositons over $A^{\mathbf{Z}}$. A decomposition $d = (d_n)_{n \in \mathbf{Z}}$ is a *C-decomposition of* the $\mathbf{Z}$-word $\boldsymbol{u} = (u_i)_{i \in \mathbf{Z}}$ if for every $k \in \mathbf{Z}$ $u[d_k, d_{k+1}[\in C$.

A language $C \subset A^+$ is said to be a *$\mathbf{Z}$-code* if *every $\mathbf{Z}$-word has at most one C-decomposition* (*over* $A^{\mathbf{Z}}$).

This definition cannot be expressed immediately in terms of coding morphisms. This will be done in the third section (theorem 3.5).

Remark: Any element u of a $\mathbf{Z}$-code is primitive, otherwise $u^{\mathbf{Z}}$ has several C-decompositions.

Examples 1.3:

– The singleton $\{u\}$ is a **Z**-code for any primitive word $u \neq \varepsilon$.

– The codes $\{a, bab\}$, $\{ba^3, ba^2 ba, a^2 ba^2\}$ are not **Z**-codes (consider representatives of the words ${}^{\omega}(ab)^{\omega}$, ${}^{\omega}(ba^3 ba^2)^{\omega}$).

– The language $\{b^n\} \cup ab(a^2 b)^*$ is a **Z**-code if and only if $n=1$ (Every word except $b^{\mathbf{Z}}$ has at most one C-decomposition).

Let us now study how **Z**-codes deal with biinfinite words.

LEMMA 1.1: *Let u, $v \in A^{\mathbf{Z}}$ be representative of the same biinfinite word w, and consider C a subset of A^+. The set of C-decompositions of u and the set of C-decompositions of v have the same cardinality.*

Proof: Since $u \simeq v$ there exists p satisfying $v_i = u_{i+p}$ for all $i \in \mathbf{Z}$ (infinitely many p in the case of periodic words, a single p in the alternative case). A bijection f between the set of C-decompositions of u and the set of C-decompositions of v can be defined next way. Let us fix p such that $v_i = u_{i+p}$ for all $i \in \mathbf{Z}$. Consider a C-decomposition d of u. There exists a single k such that $d_{k-1} < p \leqq d_k$. Let $d'_n = d_{n+k} - p$ for all $n \in \mathbf{Z}$. We set $f(d) = d'$. The sequence d' is a C-decomposition of v, and f is a bijection.

COROLLARY 1.2: *A language $C \subset A^+$ is a **Z**-code if and only if for every biinfinite word w, one of the representative of w has at most one C-decomposition.*

Let us consider a **Z**-word w. To a C-decomposition d of w one can naturally associate a C-factorization of w: it suffices to consider the class of $(w[d_n, d_{n+1}[)_{n \in \mathbf{Z}}$ for the shift relation on n. In the following, we denote by $\tau_w(d)$ this C-factorization. It is clear that each C-factorization of w can be obtained using this way and that two distinct C-factorizations of w come from two distinct C-decompositions of w (whereas the converse does not hold: see example 1.2). So we have:

PROPOSITION 1.3: *Any **Z**-code is a biω-code.*

E. Relations between these codes

PROPOSITION 1.4: *For a language $C \subset A^+$, consider the properties: (a): the language C is a **Z**-code, (b): the language C is a biω-code, (c): the language C is an ω-code, (d): the language C is a code. One has: (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) and the converse implications are false.*

Proof: (a) $\Rightarrow$ (b): Already seen.

$(b) \Rightarrow (c)$: Let C be a biω-code.

The result is clear if C is a singleton.

If C has two elements, the language C is a two-element code and then an ω-code [12].

If C is not an ω-code but has at least three elements, there exist $u_0, u_1, u_2, \ldots, v_0, v_1, v_2, \ldots \in C$ such that $u_0 \neq v_0$ and $u_0 u_1 u_2 \ldots = v_0 v_1 v_2 \ldots$; Consider $w \in C - \{u_0, v_0\}$. The word ${}^{\omega}w u_0 u_1 u_2 \ldots (= {}^{\omega}w v_0 v_1 v_2 \ldots)$ has two distinct C-factorizations.

$(c) \Rightarrow (d)$: Let C be an ω-code and $u, v \in C$ such that $u C^* \cap v C^* \neq \varnothing$. Then $u C^{\omega} \cap v C^{\omega} \neq \varnothing$ and $u = v$.

Examples 1.4:

– The singleton $\{aa\}$ is a biω-code but not a **Z**-code.

– The language $\{ab, ba\}$ is an ω-code but not a biω-code

– The language $\{a, ab, b^2\}$ is a code but not an ω-code.

2. CHARACTERIZATIONS OF SOME SPECIAL CODES RELATED TO BIINFINITE WORDS

In this part we precise how some special codes (codes, precircular codes, circular codes) deal with biinfinite words and **Z**-words. The results are used in the sequel.

A. Codes and biinfinite words

Let $\varphi : X^* \to C^*$ be a coding morphism for C. A C-factorization denoted by $x \in {}^{\omega}X^{\omega}$ is said to be *right (resp. left)-ultimately periodic* if x is right (resp. left)-ultimately periodic. It is *periodic* if x is periodic. Necessary and sufficient conditions for languages to be codes was already given by Devolder [6]:

PROPOSITION 2.1 [6]: *Let C be a language $\subset A^+$. The following conditions are equivalent:*

– *The language C is a code,*

– *Every C-factorization of a periodic biinfinite word is periodic.*

– *Every C-factorization of a right-ultimately periodic biinfinite word is right-ultimately periodic.*

Let $\varphi : X^* \to C^*$ be a coding morphism for a code C.

If $xy^{\omega} \in {}^{\omega}X^{\omega}$ is a C-factorization of a biinfinite word uv^{ω} (where v is assumed to be a primitive word), the word $\varphi(y)$ is a power of a conjugate of v.

(Of course, in the previous proposition, one can substitute left for right, and $\mathbf{Z}$-words for biinfinite words.)

COROLLARY 2.2: *Let C be a code. A C-factorization is periodic if and only if it factorizes a periodic word.*

The result is false when C is not a code. A nonperiodic factorization can factorize a periodic word. For example, ${}^{\omega}a^{\omega}$ has a nonperiodic factorization if $C=\{a,a^2\}$.

B. Precircular codes and biinfinite words

The notion of precircular codes has been defined to study the factorizations of periodic biinfinite words [6]:

DEFINITION: A language $C \subset A^+$ is said to be *precircular* if $\forall n, p \geqq 1$, $\forall u_0, \ldots, u_{n-1}, v_0, \ldots, v_{p-1} \in C$, $\forall t, s \in A^*$ *such that* $v_0 = ts$ *we have*:

$$u_0 \ldots u_{n-1} = s v_1 \ldots v_{p-1} t \quad \Rightarrow \quad n=p \quad \text{and} \quad \exists h, \forall i \; u_i = v_{(i+h) \bmod p}.$$

In fact the precircular languages are codes [6]. Let us here recall some other useful properties of precircular codes:

LEMMA 2.3: *Let C be a language $\subset A^+$ and $\varphi : X^* \to C^*$ a coding morphism for C. The language C is a precircular code if and only if*

$$\forall x, y \in X^* \; \varphi(x) \text{ and } \varphi(y) \text{ conjugate} \quad \Rightarrow \quad x \text{ and } y \text{ conjugate.}$$

PROPOSITION 2.4 [6]: *A language $C \subset A^+$ is a precircular code if and only if every periodic word of ${}^{\omega}A^{\omega}$ has at most one C-factorization.*

COROLLARY 2.5: *Any biω-code is a precircular code.*

Example 2.1: Unfortunately a precircular code is not necessarily a biω-code: The language $\{a^2, ab, b^2\}$ is a precircular code but the word ${}^{\omega}ab^{\omega}$ has two C-factorizations.

C. Circular codes and biinfinite words

Circular codes are particular cases of precircular codes. We recall here how these codes deal with biinfinite words or $\mathbf{Z}$-words.

DEFINITIONS: A *circular* code [13] is a language $C \subset A^+$ such that:

$\forall n, p \geqq 0$, $\forall u_0, \ldots, u_{n-1}, v_0, \ldots, v_{p-1} \in C$, $\forall t \in A^*$, $\forall s \in A^+$ such that $v_0 = ts$

we have: ($u_0 \ldots u_{n-1} = sv_1 \ldots v_{p-1}\, t \Rightarrow n=p$ $t=\varepsilon$ and $\forall\, i\, u_i = v_i$).

A code C is said to be pure if $\forall\, u \in A^+\ \forall\, n \geqq 1$ ($u^n \in C^* \Rightarrow u \in C^*$).

Let us recall some results used in the sequel.

LEMMA 2.6 [6]: *A language $C \subset A^+$ is a circular code if and only if C is a pure precircular code.*

LEMMA 2.7 [6]: *Let C be a precircular code $\subset A^+$ and $\varphi : X^* \to C^*$ a coding morphism for C. If ${}^{\omega}yxz^{\omega}$ (y, z primitive words) is a C-factorization of a word ${}^{\omega}uwv^{\omega}$ (u, v primitive words), then every C-factorization of a word belonging to ${}^{\omega}u\, A^*\, v^{\omega}$ belongs to ${}^{\omega}y\, X^*\, z^{\omega}$.*

LEMMA 2.8: *Let $\varphi : X^* \to C^*$ be a coding morphism for a code C. The code C is a pure code if and only if*

$$\forall\, y \in X^+, \quad y \text{ primitive} \quad \Rightarrow \quad \varphi(y) \text{ primitive}.$$

Proof: If C is a pure code, let us consider $y \in X^+$ such that $\varphi(y)$ is imprimitive. One has $\varphi(y) = u^n$ for some u and $n > 1$. Since $u^n \in C^+$, $u \in C^+$ and $u = \varphi(z)$ for some $z \in X^+$. Since $\varphi(z^n) = u^n$ and the language C is a code, $y = z^n$. Thus y is imprimitive.

Conversely, if $u^n \in C^+$ ($n > 1$), $u^n = \varphi(y)$ for some $y \in X^+$. The word $\varphi(y)$ is imprimitive so $y = z^p$ for some $p > 1$ and z primitive. The word $\varphi(z)$ is primitive and $\varphi(z)^p = u^n$, so u is a power of $\varphi(z)$ and $u \in C^+$ since $\varphi(z) \in C^+$.

In order to obtain a comparison between $\mathbf{Z}$-codes and circular codes, we prove that the notion of circular codes can be characterized by uniqueness of the decomposition of bi-ultimately periodic $\mathbf{Z}$-words. Notice the analogy between the next proposition and proposition 2.4.

PROPOSITION 2.9: *Let C be a language $\subset A^+$. The next properties are equivalent*:

1. *The language C is a circular code.*
2. *Every periodic $\mathbf{Z}$-word has at most one C-decomposition.*
3. *Every bi-ultimately periodic $\mathbf{Z}$-word has at most one C-decomposition.*

Proof: $3 \Rightarrow 2$: Obvious.

$2 \Rightarrow 1$: If C is not a circular code, there exists $u \in C^+$ such that the circular condition is not satisfied. The representatives of the word ${}^{\omega}u^{\omega}$ has two distinct C-decompositions.

$1 \Rightarrow 3$: Consider two C-decompositions of a representative of the word ${}^{\omega}uwv^{\omega}$. Since C is a code, the associated C-factorizations are bi-ultimately periodic. Let $\varphi : X \to C$ be a coding morphism for C, and ${}^{\omega}xyz^{\omega}$ (x, z primitive) be one of these factorizations. Since C is precircular, the other factorization can be written ${}^{\omega}xy_1 z^{\omega}$ (lemma 2.7). Since C is a pure code, $\varphi(x)$ is a primitive word and then cannot overlap; the same holds for $\varphi(z)$. So the C-decompositions coincide on the left and on the right. Then y and y_1 can be chosen such that $\varphi(y)=\varphi(y_1)$, and $y=y_1$ since C is a code. The two factorizations and moreover the two decompositions are in fact the same one.

COROLLARY 2.10: *Any* **Z**-*code is a circular code.*

Unfortunately this condition is not sufficient to have a **Z**-code (*see* example 2.2). However the converse of the corollary 2.10 will be proved in the rational case (see fifth section).

Example 2.2: The language $C=\{ab\} \cup \{ab^n ab^{n+1} \mid n>0\}$ is a circular code but the word ${}^{\omega}(ab)\, ab^2\, ab^3\, ab^4 \ldots$ has two C-factorizations and thus its representatives have two C-decompositions.

3. CHARACTERIZATIONS OF Z-CODES

First we give a simple characterization of **Z**-codes, which has an immediate consequence in the rational case.

PROPOSITION 3.1: *Let C be a language $\subset A^+$, C is a* **Z**-*code if and only if*

$$(({}^{\omega}C)^{-1}({}^{\omega}C) \cap A^+).((C^{\omega})(C^{\omega})^{-1} \cap A^+) \cap C = \varnothing.$$

COROLLARY 3.2: *One can decide whether a rational language $C \subset A^+$ is a* **Z**-*code.*

Proof: Since if A and B are rational infinitary languages, AB^{-1} is a rational finitary language, one has only to check whether the rational language:

$$(({}^{\omega}C)^{-1}({}^{\omega}C) \cap A^+).((C^{\omega})(C^{\omega})^{-1} \cap A^+) \cap C \text{ is empty.}$$

In the sequel we precise the relation between the notions of biω-code and **Z**-code. In this aim, let us consider for $w \in A^{\mathbf{Z}}$ the *surjective mapping* τ_w (already considered in proposition 1.3) which associates to a C-decomposition d of w the C-factorization represented by $(w[d_n, d_{n+1}[)_{n \in \mathbf{Z}}$. Of course, *a bi$\omega$-code is a* **Z**-*code if and only if for every $w \in A^{\mathbf{Z}}$, τ_w is an injective mapping.*

DEFINITION: A C-factorization c of $w \in A^{\mathbf{Z}}$ is said *to overlap* (or to be an *overlapping factorization*) if $\tau_w^{-1}(c)$ has at least two elements.

Of course, C is a $\mathbf{Z}$-code if and only if C is a biω-code such that no C-factorization overlaps.

Example 3.1: For the single $\{aa\}$-factorization c of $w=a^{\mathbf{Z}}$, $\tau_w^{-1}(c)$ has two elements: $(2n)_{n\in\mathbf{Z}}$ and $(2n+1)_{n\in\mathbf{Z}}$.

LEMMA 3.3: *An overlapping factorization factorizes a periodic biinfinite word.*

Proof: Let d, d' be two distinct C-decomposition of a $\mathbf{Z}$-word w, associated to the same C-factorization $c:\tau_w(d)=\tau_w(d')=c$. Let $(c_k)_{k\in\mathbf{Z}}$ be a representative of c; there exist p, q such that for every k

$$c_k=w\,[d_{p+k}, d_{p+k+1}[=w\,[d'_{q+k}, d'_{q+k+1}[.$$

It is clear that $d_p=d'_q$ implies $d=d'$.

Then

$$c_0\,c_1\ldots=w_{d(p)}\,w_{d(p)+1}\ldots=w_{d'(q)}\,w_{d'(q)+1}\ldots$$

and

$$\ldots c_{-2}\,c_{-1}=\ldots w_{d(p)-2}\,w_{d(p)-1}=\ldots w_{d'(q)-2}\,w_{d'(q)-1},$$

with $d_p\neq d'_q$. So the word w has the period $w\,[d_p, d'_q[$ if $d_p<d'_q$ and $w\,[d'_q, d_p[$ if $d_p>d'_q$.

Examples 3.2: The factorizations ${}^{\omega}(ab)^{\omega}$ and ${}^{\omega}(ab, a)^{\omega}$ are periodic factorizations and do not overlap.

The factorizations ${}^{\omega}(aa)^{\omega}$ and ${}^{\omega}(aba, b)^{\omega}$ are periodic factorizations and overlap.

The nonperiodic factorization of ${}^{\omega}a^{\omega}:(a^{|n|+1})_n\in\mathbf{Z}$ overlaps.

As a consequence of lemma 3.3 and proposition 2.1, an overlapping C-factorization is periodic whenever C is a code. The following lemma states precisely the structure of an overlapping periodic factorization.

LEMMA 3.4: *Consider a language C, a coding morphism for C $\varphi:X^*\to C^*$ a primitive word y of X^* and the periodic C-factorization $c={}^{\omega}y^{\omega}\in{}^{\omega}X^{\omega}$. The factorization c overlaps if and only if $\varphi(y)$ is imprimitive.*

Proof: Let us consider w a representative of ${}^{\omega}\varphi(y)^{\omega}$. Since c overlaps, w has two distinct C-decompositions d and d'. We keep the notations of the proof of lemma 3.3. Since c is periodic there exists h such that $d_p\leqq d'_{q+h}<d_p+|\varphi(y)|$ and moreover $d'_{q+h}\neq d_p$ (otherwise $d=d'$). Then w,

which has $\varphi(y)$ as a period, has also the period $w[d_p, d'_{q+h}[$ of length strictly smaller than $|\varphi(y)|$. So $\varphi(y)$ is imprimitive.

Conversely, let us assume that $\varphi(y)=u^n$ for some $n \geqq 1$ and $u \in A^+$. Then c factorizes any representative w of ${}^{\omega}u^{\omega}$. We consider a C-decomposition of $w: d$, associated to c. From the strictly increasing sequences: $(d_i + p|u|)_{i \in \mathbf{Z}}$ $(p=0, \ldots, n-1)$, by shifting the indices, one can consider n C-decompositions of w, associated to c. If c does not overlap, these decompositons coincide. Then $u \in C^+$ and there exists $z \in X^+$ such that $\varphi(z)=u$ et $y=z^n$. Thus $n=1$ since y is primitive.

Now we give the main result of this part. This characterization of $\mathbf{Z}$-code is easier to use than the previous ones, since a likewise geometric condition is replaced by algebraic ones. One can appreciate this fact for instance in the proof of proposition 4.2. In fact the theorem 3.5 can be expressed in terms of coding morphism properties (use lemma 2.8 and the definition of a biω-code).

THEOREM 3.5: *Let C be a language $\subset A^+$, C is a $\mathbf{Z}$-code if and only if C is a pure biω-code.*

Proof: A $\mathbf{Z}$-code is a circular biω-code and thus a pure biω-code.

Conversely, assuming C is a pure biω-code, one has to show that no C-factorization overlaps. From lemma 3.3, it suffices to consider the periodic biinfinite words. The result is then an immediate consequence of corollary 2.5, lemma 2.6 and proposition 2.9 (indeed, C is a pure precircular code and then a circular code).

As a consequence of theorem 3.5, we shall see in proposition 4.2 that the composition of $\mathbf{Z}$-codes gives $\mathbf{Z}$-codes.

4. PROPERTIES OF BIω-CODES

In order to have a lot of examples of biω-codes, we study simple biω-codes and composition of biω-codes.

Barbin and Le Rest have studied the two-element codes [1]. As an application, Devolder shows that $\{u, v\}$ is a precircular code if and only if the primitive roots of u and v are not conjugate [6]. It is now easy to show the same result for biω-codes: $\{u, v\}$ is a biω-code whenever it is a precircular code:

PROPOSITION 4.1: *Let $C=\{u,v\}$ be a two-word subset of A^+. Let $u_0=\sqrt{u}$ and $v_0=\sqrt{v}$. The language C is a biω-code if and only if u_0 and v_0 are not conjugate.*

Proof: The necessary condition is obvious.

Assume that u_0 and v_0 are not conjugate. Then C is a precircular code [6]. Let $C_0=\{u_0, v_0\}$. In ${}^{\omega}C_0^{\omega}$ there is at most one word whose representatives have two distinct C_0-decompositions [15, 1, 6]. It is the word ${}^{\omega}w^{\omega}$ where w is the single imprimitive word in $u_0 v_0^+ \cup u_0^+ v_0$ (if exists). Since C is a precircular code, the periodic word ${}^{\omega}w^{\omega}$ has a single C-factorization. So every word in ${}^{\omega}C^{\omega}$ has a single C-factorization.

PROPOSITION 4.2: *Let C be a language $\subset X^+$ and $\varphi : X^* \to A^*$ be a coding morphism for a language $D=\varphi(X) \subset A^+$.*

*If C and D are biω-codes (resp.: **Z**-codes), $\varphi(C)$ is a biω-code (resp.: **Z**-code). If $\varphi(C)$ is a biω-code (resp.: **Z**-code) and φ bijective from C to $\varphi(C)$, C is a biω-code (resp.; **Z**-code).*

Proof: One can consider a coding morphism $\Psi : Y^* \to X^*$ for$C$. As φ is bijective from C to $\varphi(C)$ in both cases, $\varphi \circ \Psi$ is a coding morphism for $\varphi(C)$. In the case of biω-codes, it remains to use the injective property of $\varphi : X^{\omega} \to A^{\omega}$ and $\Psi : Y^{\omega} \to X^{\omega}$, or that of $\varphi \circ \Psi : Y^{\omega} \to A^{\omega}$. In the case of **Z**-codes, it remains to prove the properties for pure codes. The result is easily obtained from the characterization 2.8 of pure codes.

Remarks: If C or D is not a biω-code, $\varphi(C)$ may or may not be a biω-code.

– It is possible to have a biω-code $\varphi(C)$ even if C is not a biω-code when $\varphi : C \to \varphi(C)$ is not a bijection (In this case, D is not a code of course). For example let us consider:

$$C=\{ab, ba\}, \qquad D=\{c, c^2\}, \qquad \varphi(a)=c,$$
$$\varphi(b)=c^2, \qquad \varphi(C)=\{c^3\}.$$

LEMMA 4.3: *Let Y be an alphabet (finite or not) and β be a mapping from Y to the set of natural numbers* N. *The language $Y_\beta=\{z^{\beta(z)} \mid z \in Y\ \beta(z) \neq 0\}$ is a biω-code.*

Proof: Two different words of Y_β cannot overlap. And the periodic words ${}^{\omega}(z^{\beta(z)})^{\omega}$ have only one C-factorization.

LEMMA 4.4: *Let C be a biω-code and β be a mapping from C to* N.

$$C_\beta = \{ u^{\beta(u)} \mid u \in C\,\beta(u) \neq 0 \} \text{ is a bi}\omega\text{-code.}$$

Proof. — The code C_β is the result of the composition of Y_β and C.

Example 4.2: The set C_β may be a biω-code even if C is not a biω-code:

Since ${}^\omega ab^\omega$ has two C-factorizations, the code $C = \{ a^2, ab, b^2 \}$ is not a biω-code. Nevertheless $\{ a^2, (ab)^2, b^2 \}$ is a biω-code.

In the sequel we give some characterizations for biω-codes. Some properties of biω-codes have already been found: a biω-code C is a precircular ω-code, and the mirror of C: $\tilde{C}$ is also an ω-code (we shall say that C is an $\tilde{\omega}$-code). Unfortunately, these conditions are not sufficient to have a biω-code, even if we consider a finite biprefix code:

Example 4.3: The language $C = \{ aa, ab, bb \}$ is such a code. The word ${}^\omega ab^\omega$ has two C-factorizations.

Examples 4.4: The sets of precircular codes, ω-codes and $\tilde{\omega}$-codes cannot be compared.

— The language $\{ ab, ba \}$ is a nonprecircular (biprefix) ω-code and is an $\tilde{\omega}$-code

— The language $\{ c, b^2, cb \}$ is a precircular $\tilde{\omega}$-code, but is not an ω-code.

Characterization 4.5: *Let C be a precircular code $\subset A^+$. The language C is a biω-code if and only if C satisfies:*

(2) $\forall s, t \in A^+$ such that $st \in C$ and $({}^\omega C \cap {}^\omega C s) . (C^\omega \cap t C^\omega) \neq \emptyset$,

there exists $u \in A^+$ such that $({}^\omega C \cap {}^\omega C s) . (C^\omega \cap t C^\omega) = {}^\omega u^\omega$.

Notation: For convenience, in the following, we shall denote by M_s the set ${}^\omega C \cap {}^\omega C s$ and by M'_s the set $C^\omega \cap s C^\omega$ (when $s \in A^+$).

Proof: Let C be a biω-code, then C is a precircular code.

Consider $v \in M_s$ and $w \in M'_t$. The representatives of the biinfinite word vw have two different C-decompositions. The corresponding factorizations are in fact the same one and the word is periodic. There exists u such that $v = {}^\omega u$ and $w = u^\omega$. For every $v' \in M_s$, $v' w$ is periodic and then $v' = {}^\omega u$. Thus $M_s = {}^\omega u$ and in a same way $M'_t = u^\omega$.

For the converse, if a **Z**-word x has two distinct C-decompositions, there exist s and $t \in A^+$ such that $x \in M_s . M'_t$. Then x is periodic. Since C is a precircular code, by proposition 2.4, x has only one C-factorization.

Remark: In this characterization:

– one cannot omit the condition (2): *see* example 4.3.

– one cannot replace the condition "precircular code" by "almost circular code" (the definition of almost circular codes is given just after): *see* example 4.5.

Example 4.5: The language $C=\{ab, c, bca\}$ is an almost circular non-precircular code (so C is not a biω-code) and we have: $M_a={}^{\omega}(bca)$ and $M'_b=(bca)^{\omega}$, $M_{bc}={}^{\omega}(abc)$ and $M'_a=(abc)^{\omega}$ and $M_b={}^{\omega}(cab)$ and $M'_{ca}=(cab)^{\omega}$.

In order to weaken the condition (2) in the previous characterization we need the notion of almost circular code defined by Leconte.

DEFINITION 4.1 [14]: A language C is said to be *almost circular* if it is a code such that $\forall s, t \neq \varepsilon$ such that $st \in C$, $t C^* s \cap C^*$ is empty or is a monogeneous semigroup (*i.e.* of the form u^+ for some u).

LEMMA 4.6 [6]: *A precircular code is an almost circular code.*

PROPOSITION 4.7: *A code which satisfies*
(2') $\forall s, t \in A^+$, if $st \in C$ and $M_s . M'_t \neq \varnothing$ then M_s and M'_t are monogeneous (that is: there exist u, $v \in A^+$ such that $M_s={}^{\omega}u$ and $M'_t=v^{\omega}$) is an almost circular code and satisfies (2).

Proof: Let us consider w_0 and $w \in t C^* s \cap C^*$ (with s, $t \in A^+$ and $st \in C$). The words ${}^{\omega}w_0$ and ${}^{\omega}w$ belong to M_s. But $M_s={}^{\omega}u$, where u can be chosen primitive. Then w_0, w belong to u^*. In a same way, w_0, w belong to v^* if $M'_t=v^{\omega}$ and v primitive. So $u=v$ and $t C^* s \cap C^* \subset u^*$. As C is a code, $t C^* s \cap C^*$ is monogeneous [6]. So C is an almost circular code. Let us give a definition useful in the sequel of the proof:

Let z be a C-factorization of a word $uvw \in A^{\omega}$. We say that z define the cutting diagram $(v_1, \ldots, v_n)$ on the word v if $\exists p \geqq 0$ $\exists n \geqq 2$ such that $u=\varphi(z_1)\ldots\varphi(z_p)u'$ (with $u' \neq \varepsilon$),

$$\varphi(z_1)\ldots\varphi(z_{p+1})=uv_1, \qquad \varphi(z_{p+i})=v_i\,\varphi(z_{p+n})=v_n v',$$
$$uv<\varphi(z_1)\ldots\varphi(z_{p+n}) \qquad \text{and} \qquad v=v_1\ldots v_n.$$

Note that

$$v_1=\varepsilon \qquad \text{iff} \quad \varphi(z_1)\ldots\varphi(z_{p+1})=u$$

and

$$v_n=\varepsilon \qquad \text{iff} \quad \varphi(z_1)\ldots\varphi(z_{p+n-1})=uv.$$

Consider now s, t, u, $v \in A^+$ such that $st \in C$, $M_s = {}^{\omega}u$, $M'_t = v^{\omega}$, u and v primitive. It remains to prove that $u = v$.

Let $\varphi : X^* \to C^*$ be a coding morphism for C. Consider z (resp. z') $\in X^{\omega}$ a C-factorisation of v^{ω} (resp. $t^{-1} v^{\omega}$). Since C is a code, by proposition 2.1, z and z' are ultimately periodic ($z = z_1 z_2^{\omega}, z' = z'_1 z_2'^{\omega}$) and there exist v_1, v'_1 conjugate with v such that $\varphi(z_2) = v_1^n$, $\varphi(z'_2) = v_1'^{n'}$ for some n and n'. The conjugate words $v_1^{nn'}$ and $v_1'^{nn'}$ belong to C^*. We consider the successive occurrences of $v_1^{nn'}$ on which the cutting diagrams defined by z are the same one: $(\varepsilon, \ldots, \varepsilon)$ and on which the cutting diagrams defined by z' are also the same one: $(t', \ldots, s')$.

So we have two cases:

First case: s' and $t' \neq \varepsilon$. the factorizations z and z' define (in their periodic part) on the studied occurrences of $v_1^{nn'}$ cutting diagrams such that $s' t' \in C$, $v_1^{nn'} \in C^* \cap t' C^* s'$, ${}^{\omega}uv^{\omega} \in M_{s'} . M'_{t'}$. In the first part of the proof we proved that, in this case, $M_{s'} = {}^{\omega}v_1$, $M'_{t'} = v_1^{\omega}$. Then $u = v$. Second case: $t' = \varepsilon$ and then $s' = \varepsilon$. On all the studied occurrences of $v_1^{nn'}$, the cutting diagrams defined by z and z' are identical. Then we study what happens with a C-factorization y of ${}^{\omega}u$ and a C-factorization y' of ${}^{\omega}us^{-1}$, and we conclude if we are in the analogous first case. If we are in the analogous second case, we have for some z_3, z'_3, y_3, $y'_3 : z = z_3 z_2^{\omega}$, $z' = z'_3 z_2^{\omega}$ with $\varphi(z_3) = t\varphi(z'_3)$ and $y = {}^{\omega}y_2 y_3$, $y' = {}^{\omega}y_2 y'_3$ with $\varphi(y_3) = \varphi(y'_3) s$. As $st \in C$ and C is a code, we obtain a contradiction.

Remark: There exist almost circular codes which do not satisfy (2'). See example 4.3: $C = \{aa, ab, bb\}$, $M'_a = a^* b^{\omega} + a^{\omega}$.

– The conditions (2) and (2') are not sufficient to have a code: For example: $C = \{a, a^2\}$ is not a code and nevertheless $M_a = {}^{\omega}a$, $M'_a = a^{\omega}$.

As an immediate consequence of proposition 4.7 and characterization 4.5 we have:

THEOREM 4.8: *A precircular code $C \subset A^+$ is a biω-code if and only if C satisfies:*
(2') $\forall s, t \in A^+$ such that $st \in C$ and $M_g . M'_t \neq \emptyset$, the languages M_s and M'_t are monogeneous.

Remark: It is not difficult to see that if a code C satisfies the property (2'), it satisfies also the property:
(3) $\forall s, t, t' \in A^+$ such that $st \in C$, $st' \in C$, M_s, M'_t and $M'_t \neq \emptyset$, one has $t = t'$. And thus a biω-code satisfies the property (3).

But the condition (2′) in the theorem 4.8 can be replaced neither by (3) nor by

(4) $\forall s,\ t \in A^+$ such that $st \in C$ and $M_s . M'_t \neq \varnothing$ the languages M_s and M'_t are singletons.

Example 4.6: The language

$$C = \{ acab \} \cup \{ ac^{n+1} ac^n \mid n \geqq 1 \} \cup \{ ab^n ab^{n+1} \mid n \geqq 1 \}$$

is a circular ω-code but not a biω-code.

$\ldots ac^4 ac^3 ac^2 acabab^2 ab^3 ab^4 \ldots$ is the single word which has two C-factorizations. So, every (M_s, M'_t) such that $st \in C$, M_s and $M'_t \neq \varnothing$ is composed of two singletons. The code C satisfies also the property (3).

5. Z-CODES AND BIω-CODES IN THE RATIONAL CASE. DECIDABILITY

When a language C is rational, one can consider an automaton $\Omega_0 = (Q_0, q_0, q_F)$ with a finite set of states Q_0, a single initial state q_0 and a single final state q_F, which recognizes C and such that no edge comes to q_0 and no edge gœs from q_F. Ω_0 can be chosen trim (*i.e.* for every state q there exist a path from q_0 to q and a path from q to q_F) and unambiguous (*i.e.* the words of C have a single successful lecture). The automaton $\Omega = (Q, q_0, q_0)$ obtained by identification of q_0 and q_F recognizes C^*. If C is a code, the automaton Ω is unambiguous [4]. This automaton looked as a Büchi automaton recognizes ${}^{\omega}C^{\omega}$. Let $w \in A^{\mathbf{Z}}$. There is a bijection f between the set of successful lectures $1 = (q_{j(i)}, w_i, q_{j(i+1)})_{i \in \mathbf{Z}}$ of w on Ω and the set of C-decompositions of w: it suffices to define $f(1) = d$ where d satisfies: $d(\mathbf{Z}) = \{ i \in \mathbf{Z} \mid q_{j(i)} = q_0 \}$.

We denote by $A_{q,q'}$ the set of words which can be read on Ω from q to q', and by $B_{q,q'}$ the set $A_{q,q_0} . A_{q_0,q'}$ when $(q, q') \neq (q_0, q_0)$.

The next technical proposition will be used thoroughly in the following.

Proposition 5.1: *Let C be a rational language $\subset A^+$ and Ω_0 an unambiguous automaton recognizing C such as considered before.*

Let D be the set of finite words which have at least two C-factorizations. Put $X_1 = {}^{\omega}CDC^{\omega}$ and $X_2 = \bigcup {}^{\omega}(B_{q',q'} \cap C^+) . (B_{q',q} \cap C^+) . (B_{q,q} \cap C^+)^{\omega}$ (where $(q, q') \in Q^2 - (q_0, q_0)$).
The language $Y = X_1 \cup X_2$ is the set of words whose representatives have several C-decompositions. Y is a rational language.

Proof: It is known that the language D is rational [4, 11, substraction lemma]. A word of X_1 has clearly two distinct C-factorizations. A representative w of a word of X_2 has two distinct successful lectures on Ω since $q' \neq q_0$ or $q \neq q_0$, then w have two distinct C-decompositions.

Conversely, if $w \in A^{\mathbf{Z}}$ has two distinct C-decompositions, w has two successful lectures on Ω $(q_{j(i)}, w_i, q_{j(i+1)})_{i \in \mathbf{Z}}$ and $(q_{k(i)}, w_i, q_{k(i+1)})_{i \in \mathbf{Z}}$. Consider a strictly increasing sequence $(i_p)_{p \in \mathbf{Z}}$ such that $q_{j(ip)} = q_0$ and define $1(p) = k(i_p)$. Since Q is finite and the second lecture successful, there exist $q, q' \in Q$ and a strictly increasing sequence $(p_n)_{n \in \mathbf{Z}}$ such that $q_{l(pn)} = q$ for $n > 0$, $q_{l(pn)} = q'$ for $n < 0$ and $\forall n \exists h$ $q_{k(h)} = q_0$ and $i(p_n) \leqq h < i(p_{n+1})$ (between the considered crossings through q (resp. q') there is a crossing through q_0). If $q \neq q_0$ or $q' \neq q_0$, $w \in X_2$. If $q = q' = q_0$, since the C-decompositions are distinct, $w \in X_1$.

As a consequence of the previous proposition, Beal's next result [2] is obtained:

THEOREM 5.2: *Let C be a rational language $\subset A^+$, C is a $\mathbf{Z}$-code if and only if C is a circular code.*

Proof: The condition $Y = \varnothing$ is equivalent with $X_1 = \varnothing$ (that is $D = \varnothing$) and $X_2 = \varnothing$ (that is $\forall q \neq q_0$ $B_{q,q} \cap C^+ = \varnothing$); the condition $Y = \varnothing$ is then equivalent with "C is a circular code".

When C is a nonrational language, this condition is not sufficient to have a biω-code. *See* example 2.2.

Corollary 3.2 gives a method to decide whether a rational language is a circular code.

In the case of infinite words, it suffices to study the rational infinite words to know whether a rational language is an ω-code [7]. In the case of $\mathbf{Z}$-words, from theorem 5.2 and proposition 2.9 we have an analogous result for $\mathbf{Z}$-codes. For biω-codes, this is also true:

PROPOSITION 5.3: *A rational language $C \subset A^+$ is a biω-code (resp: $\mathbf{Z}$-code) if and only if every bi-ultimately periodic biinfinite word (resp: $\mathbf{Z}$-word) has at most one C-factorization (resp: C-decomposition).*

Proof (for biω-codes): Let us consider the set Y defined at proposition 5.1. Let us denote by Y_1 the set of words having at least two distinct C-factorizations and Y_2 the set of periodic words having a single C-factorization and whose C-factorization overlaps. Clearly $Y = Y_1 \cup Y_2$. Since Y is rational, it is a finite union of sets ${}^{\omega}A_i B_i C_i^{\omega}$ [18]. If ${}^{\omega}A_i B_i C_i^{\omega}$ is not reduced to a single periodic word, there exists a rational nonperiodic word in Y, and thus a rational word in Y_1. Therefore, if every rational word has at most one C-factorization, $Y=\bigcup {}^{\omega}u_i^{\omega}$. By hypothesis, $Y=Y_2$ and Y_1 is empty.

It is then clear that the following corollary holds:

COROLLARY 5.4: *Let C be a rational biω-code $\subset A^+$, the set X of biinfinite words which representatives have several C-decompositions is finite.*

Another easy consequence of proposition 5.3 is:

COROLLARY 5.5: *Let C be a rational language $\subset A^+$, C is a biω-code (resp: $\mathbb{Z}$-code) if and only if all its finite subsets are biω-codes (resp: circular codes).*

These properties 5.3 and 5.5 are false for a nonrational language (*see* example 2.2).

The next lemma gives another property of biω-codes, that will be used in the rational case:

LEMMA 5.6: *A biω-code $C\subset A^+$ satisfies the property* (4):
(4) *$\forall s, t, s', t', s'', t''\in A^+$ such that st, $s't'$, st'' and $s''t'\in C$, $C^*\cap tC^*s$, $C^*\cap t''C^*s''$ and $C^*\cap t'C^*s'\neq\varnothing$, one has $t''=t$, $s''=s'$ and there exist $p>0$, a primitive word u, a conjugate v of u ($v=u''u'$, $u=u'u''$) such that $C^*\cap tC^*s=(u^p)^+$, $C^*\cap t''C^*s''\subset u^*u'$ and $C^*\cap t'C^*s'=(v^p)^+$.*

Proof: A biω-code C is an almost circular code and we have $C^*\cap tC^*s=(u^p)^+$ and $C^*\cap t'C^*s'=(v^n)^+$ for some primitive words u and v and some integers p and n. Let us consider $w\in C^*\cap t''C^*s''$. The representatives of ${}^{\omega}uwv^{\omega}$ have two distinct C-decompositions. Then ${}^{\omega}uwv^{\omega}$ has an overlapping C-factorization and ${}^{\omega}uwv^{\omega}$ is periodic, hence $v=u''u'$, $u=u'u''$ and $w\in u^*u'$. Since C is a precircular code, the proof of [6], lemma 5.2, gives $n=p$.

Since C is a precircular code, ${}^{\omega}u^{\omega}$ has a unique C-factorization. Let $t^{-1}ut=c_1\ldots c_p$ where $c_p=st$ and $c_i\in C$ for every i. The C-factorization of ${}^{\omega}u^{\omega}$ is ${}^{\omega}(c_1,\ldots,c_p)^{\omega}$ and belongs to ${}^{\omega}(c_1\ldots c_p)\,c_1\ldots c_{p-1}st''C^*s''t'C^{\omega}$, then $st''=c_p=st$ and $t=t''$. In a same way $s''=s'$.

A biω-code must factorize in a single manner the periodic biinfinite words, the right-infinite words, the left-infinite words and also the biultimately-periodic biinfinite words. In the rational case, it turns out that these conditions are shown to be sufficient:

THEOREM 5.7: *A rational language $C\subset A^+$ is a biω-code if and only if C satisfies the four next conditions*:

- *the language C is a precircular code*
- *the language C is an ω-code*

– *the language C is an* $\tilde{}\omega$-*code*

– *the language C satisfies the property* (4).

Proof: We have already seen the necessity of the four conditions (corollary 2.5, proposition 1.3, and lemma 5.6).

Conversely, it is sufficient to prove that every rational word has at most one C-factorization. Let us assume that a representative r of the rational word ${}^{\omega}uwv^{\omega}$ (u and v primitive) has two C-decompositions d_1 and d_2 and consider a coding morphism φ: $X^* \to C^*$ for C.

τ_r maps the two decompositions on two C-factorizations f_1 and f_2. Since C is a code, $f_1 = {}^{\omega}yxz^{\omega}$ for some primitive words y, $z \in X^+$. Since C is a precircular code, $f_2 = {}^{\omega}yx'z^{\omega}$ for some $x' \in X^*$ (lemma 2.7).

Look at the decompositions d_1 and d_2. Two cases can happen:

– the decompositions d_1 and d_2 cut the word r at (at least) a same point ($d_1(\mathbf{Z}) \cap d_2(\mathbf{Z}) \neq \varnothing$), then $d_1 = d_2$ since C is both an $\tilde{}\omega$-code and an ω-code.

– the decompositions d_1 and d_2 coincide nowhere on the word r. Since $f_1 = {}^{\omega}yxz^{\omega}$ and $f_2 = {}^{\omega}yx'z^{\omega}$, one can obtain s, t, s', $t' \in A^+$ such that $st \in C$ and $\varphi(y) \in C^* \cap tC^*s$ ($s, t \neq \varepsilon$ since d_1 and d_2 coincide nowhere), $s't' \in C$ and $\varphi(z) \in C^* \cap t'C^*s'$. Since d_1 and d_2 decompose the same word, one can obtain s'', $t'' \in A^+$ such that $st'' \in C$, $t''s' \in C$ *and* $C^* \cap t''C^*s'' \neq \varnothing$ and the word ${}^{\omega}uwv^{\omega}$ belongs to ${}^{\omega}(C^* \cap tC^*s)\,(C^* \cap t''C^*s'')\,(C^* \cap t'C^*s')^{\omega}$. The fourth condition implies that the word ${}^{\omega}uwv^{\omega}$ is periodic. Therefore, since C is a precircular code, ${}^{\omega}uwv^{\omega}\,(= {}^{\omega}u^{\omega})$ has a single C-factorization. Note that this factorization is ${}^{\omega}y^{\omega}$.

The theorem 5.7 is false for a nonrational language.

No condition can be suppressed in the theorem 5.7, even in the finite case.

Examples 5.3:

– The language $C = \{abab\} \cup \{ab^n ab^{n+1} \mid n > 0\}$ is a precircular suffix ω-code which satisfies the property (4). Indeed for all $s, t, s', t', s'', t'' \in A^+$ such that st, $s't'$, st'' and $s''t' \in C$, and $C^* \cap tC^*s$, $C^* \cap t''C^*s''$ and $C^* \cap t'C^*s' \neq \varnothing$, one has

$$s = s' = s'' = t = t' = t'' = ab \qquad \text{and} \qquad C^* \cap tC^*s = (abab)^+.$$

But the word ${}^{\omega}(ab)\,ab^2\,ab^3\,ab^4\ldots$ has two distinct C-factorizations.

– The language $C = \{a^2, ab, b^2\}$ is a finite precircular biprefix code, but ${}^{\omega}ab^{\omega}$ has two C-factorizations.

– The language $C = \{a, ab, b^2\}$ is a finite precircular $\tilde{}\omega$-code which satisfies the fourth condition, but ${}^{\omega}ab^{\omega}$ has two C-factorizations.

– The language $C=\{ab, ba\}$ is a finite biprefix code which satisfies the fourth condition, but ${}^{\omega}(ab)^{\omega}$ has two C-factorizations.

Remark: In theorem 5.7, the property (4) cannot be replaced by the property (4'):

(4') $\forall s, t, s', t', s'', t'' \in A^+$ such that $st, s't', st'', t' \in C$, $C^* \cap t C^* s$, $C^* \cap t'' C^* s''$, $C^* \cap t' C^* s' \neq \varnothing$, one has $t''=t$ and $s''=s'$.

Example 5.4: The language $C=\{a, aba^2 b, a^2 bc, c^2\}$ is a precircular suffix ω-code which does not satisfy the property (4): ${}^{\omega}(a^2 b) c^{\omega}$ has two C-factorizations but C satisfies the property (4') since:

– $C^* \cap t C^* s \neq \varnothing$ and $st \in C$ implies $(s, t) \in \{(c, c), (ab, a^2 b), (aba, ab)\}$

– $st \in C$ and $s \in \{c, ab, aba\}$ implies $(s, t) \in \{(c, c), (ab, a^2 b), (aba, ab)\}$.

Using an automaton, in the rational case, theorem 5.7 and theorem 4.8 may be improved. In the following, Ω_0 is an unambiguous automaton such as considered in the beginning of the section 5. Let us denote by: $C_{q, q'}$ the set of words which can be read on Ω_0 from q to q', $T_{q, q'}$ the set $C^+ \cap C_{q, q_F} . C^* . C_{q_0, q'}$, M_q the set ${}^{\omega}C \cap {}^{\omega}C . C_{q_0, q}$, M'_q the set $C^{\omega} \cap C_{q, q_F} . C^{\omega}$.

In theorem 5.7 the condition (4) may be replaced by the next one: $\forall q, q' \in Q_0 - \{q_0, q_F\}$, $q \neq q'$, if $T_{q, q}$, $T_{q, q'}$ and $T_{q', q'}$ are not empty there exist $p, n > 0$, a primitive word u and a conjugate v of u ($v = u'' u'$, $u = u' u''$) such that $T_{q, q} = (u^p)^+$, $T_{q, q'} \subset u^* u'$ and $T_{q', q'} = (v^n)^+$.

Another characterization for rational biω-code can be obtained from theorem 4.8. This gives a decidability result.

THEOREM 5.8: *Let C be a rational precircular code $\subset A^+$ and Ω_0 an unambiguous automaton recognizing C such as considered before. The precircular code C is a biω-code if and only if for every $q \in Q_0 - \{q_0, q_F\}$ such that $M_q . M'_q \neq \varnothing$, M_q and M'_q are monogeneous.*

Proof: From the characterization 4.8, one can see that the condition is sufficient.

Conversely, consider s_0, $s \in C_{q_0, q}$, $t \in C_{q, q_F}$ such that M_{s0}, M_s and $M'_t \neq \varnothing$. From the characterization 4.5, $M_{s0} = {}^{\omega}u$, $M'_t = {}^{\omega}u$ and then $M_s = {}^{\omega}u$. Then $M_q = {}^{\omega}u$ and $M'_q = u^{\omega}$.

COROLLARY 5.9: *One can decide whether a rational language is a biω-code.*

Proof: One can decide whether a rational language is a precircular code [6] and whether a rational ω-language L is monogeneous (consider an ultimately periodic word $uv^{\omega} \in L$, verify if uv^{ω} is periodic and check the equality $L = uv^{\omega}$).

ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their interesting remarks and suggestions.

REFERENCES

1. E. Barbin-Le Rest and M. Le Rest, Sur la combinatoire des codes à deux mots, *Theor. Comp. Sc.*, 1985, *41*, pp. 61-80.
2. M. P. Beal, Codages, automates locaux et entropie, *Thèse*, Publications du L.I.T.P., 38, Université de Paris-VII, 1988.
3. D. Beauquier, Automates sur les mots biinfinis, *Thèse d'état*, Université Paris-VII, 1986.
4. J. Berstel and D. Perrin, Theory of Codes, Academic Press, Orlando, 1985.
5. F. Blanchard, Codes engendrant certains systèmes sofiques, *Theoret. Comput. Sci.*, 1989, *68*, pp. 253-265.
6. J. Devolder, Precircular Codes and Periodic Biinfinite Words, Publications du L.I.F.L., I.T. 175, 1989, *Inform. and Comput.* (to appear).
7. J. Devolder, M. Latteux, I. Litovsky and L. Staiger, Codes and Infinite Words, Publications du L.I.F.L., I.T. 220, 1991.
8. J. Devolder and E. Timmerman, Codes for Biinfinite Words, Publications du L.I.F.L., I.T. 194, 1990.
9. Do Long Van, Codes avec des mots infinis, *R.A.I.R.O. Inform. Theor. Appl.*, 1982, *16*, pp. 371-386.
10. Do Long Van, D. G. Thomas, K. G. Subramanian and R. Siromoney, Bi-Infinitary Codes, *RAIRO Inform. Theor. Appl.*, 1990, *24*, 1, pp. 67-87.
11. S. Eilenberg, Automata, Languages, and Machines, Academic Press, New York, 1974.
12. J. Karhumaki, On Three-Element Codes, *Theoret. Comput. Sci.*, 1985, *40*, pp. 3-11.
13. J. L. Lassez, Circular codes and synchronisation, *Internat. J. Comput. Syst. Sci.*, 1976, *5*, pp. 201-208.
14. M. Leconte, K-th Power-Free Codes, Automata on infinite words, *Lecture notes in Comput. Sci.*, 1984, *192*, pp. 172-187.
15. A. Lentin and M. P. Schutzenberger, A Combinatorial Problem in the Theory of Free Monoids, *Proc. Univ. of North Carolina*, 1967, pp. 128-144.
16. I. Litovsky, Générateurs des langages rationnels de mots infinis, *Thèse*, Université de Lille-I, 1988.
17. A. de Luca and A. Restivo, On Some Properties of Very Pure Codes, *Theoret. Comput. Sci.*, 1980, pp. 157-170.
18. M. Nivat and D. Perrin, Ensembles reconnaissables de mots biinfinis, *Proc. 14^e Symp. on Theory of Computing*, 1982, pp. 47-59.
19. A. Restivo, Codes and automata, *Lecture notes in Comput. Sci.*, 1989, *386*, pp. 186-198.
20. L. Staiger, On Infinitary Finite Length Codes, *R.A.I.R.O., Inform. Théor. Appl.*, 1986, *20*, 4, pp. 483-494.