INFORMATIQUE THÉORIQUE ET APPLICATIONS

O. CARTON

# A hierarchy of cyclic languages

*Informatique théorique et applications*, tome 31, nº 4 (1997), p. 355-369

<http://www.numdam.org/item?id=ITA_1997__31_4_355_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 31, n° 4, 1997, pp. 355-369)

# A HIERARCHY OF CYCLIC LANGUAGES (*)

by O. CARTON ([1])

Abstract. – *We introduce a hierarchy of cyclic languages. The* k*th level of this hierarchy consists of all cyclic languages which are equal to some boolean combination of size* k *of strongly cyclic languages. We then show how this hierarchy can be characterized by chains of idempotents in monoids. Finally, we give a method to compute an optimal (in the number of terms) decomposition of a cyclic language into strongly cyclic languages.*

Résumé. – *Nous introduisons une hiérarchie des langages cycliques. Le* k*-ième niveau de cette hiérarchie comprend les langages cycliques qui sont égaux à une combinaison booléenne de taille* k *de langages fortement cycliques. Nous montrons ensuite comment cette hiérarchie peut être caractérisée par des chaînes d'idempotents dans des monoïdes. Finalement, nous donnons une méthode pour calculer une décomposition optimale (en nombre de termes) d'un langage cyclique en des langages fortement cycliques.*

## 1. INTRODUCTION

Cyclic languages and strongly cyclic languages are two classes of languages of finite words over a finite alphabet. A cyclic language is conjugation-closed and for any two words having a power in common, if one of them is in the language, then so is the other. A strongly cyclic languages is the set of words stabilizing a subset of the set of states of a finite deterministic automaton, the stabilized subset depending on the word stabilizing it. Every strongly cyclic language is rational.

It has been proved in [BCR96] that any rational cyclic language is a boolean combination of strongly cyclic languages. This result allows us to extend the computation of the zeta functions of strongly cyclic languages described in [Béa95] to rational cyclic languages. The connections of cyclic languages with algebraic geometry and symbolic dynamics are also discussed in [BR90]. We introduce in this paper a hierarchy among cyclic languages.

(*) Received December 1995.

([1]) Institut Gaspard Monge, Université de Marne-la-Vallée, F-93166 Noisy-le-Grand Cedex. E-mail: Olivier.Carton@univ-mlv.fr, http://www-igm.univ-mlv.fr/~carton/

This hierarchy measures the number of strongly cyclic languages needed to express a given cyclic language as a boolean combination of strongly cyclic languages. We prove that this hierarchy can be characterized by chains of idempotents in monoids. The level of the hierarchy to which a given cyclic language belongs can be computed in a monoid recognizing this language. In particular, it can be done in the syntactic monoid of the language.

In section 6, we prove that for any cyclic language $L$, there is a smallest strongly cyclic language containing $L$ which is called the *closure* of $L$. We show that the closure can be computed in the syntactic monoid of $L$. This result is used in section 7 to give a procedure to decompose $L$ as a boolean combination of strongly cyclic languages which uses less strongly cyclic languages than any other boolean combination of strongly cyclic languages equal to $L$.

We assume that the reader is familiar with the basic notions of automata and monoid theory. For example notions like syntactic monoid, Green relations, regular $\mathcal{D}$-classes are supposed to be known. We refer to [Lal79] and [Pin86] for a presentation of this subject.

The paper is organized as follows. Section 2 and 3 give the basic properties of cyclic languages and strongly cyclic languages. The chains of strongly cyclic languages and the hierarchy of cyclic languages are introduced in section 4. In section 5, we define chains of idempotents in monoids which characterize the classes of the hierarchy. In section 6, we define the closure of a cyclic language. This notion gives a method to decompose a cyclic language into strongly cyclic languages. This method is described in section 7.

## 2. CYCLIC LANGUAGES

In this section, we introduce cyclic languages and give some basic properties. In the following, we denote by $A$ a finite alphabet. In a finite monoid $M$, every element $s$ of $M$ has a power which is an idempotent. We denote by $s^\omega$ this idempotent.

DEFINITION 1: *A language $L$ of $A^*$ is said to be cyclic if it satisfies*

$$\begin{array}{ll} \forall u \in A^*, \forall n > 0 & u \in L \Leftrightarrow u^n \in L \\ \forall u, v \in A^* & uv \in L \Leftrightarrow vu \in L. \end{array}$$

A language is cyclic if it is closed under conjugation, power and root. By definition, the class of cyclic languages is closed under boolean operations.

EXAMPLE 1: *If* $A = \{a, b\}$, *the language* $L = A^*aA^* = A^* - b^*$ *is cyclic.*

Cyclic languages have the following straightforward characterization in terms of monoids.

PROPOSITION 1: *Let* $L \subset A^*$ *be a rational language. Let* $\varphi : A^* \twoheadrightarrow M$ *be a morphism from* $A^*$ *onto a monoid* $M$ *such that* $L = \varphi^{-1}(P)$ *for some* $P \subset M$. *The language* $L$ *is cyclic if and only if*

$$\begin{array}{ll} \forall s \in M, \forall n > 0 & s \in P \Leftrightarrow s^n \in P \\ \forall s, t \in M & st \in P \Leftrightarrow ts \in P. \end{array}$$

It is straightforward to verify that if those conditions are satisfied, the language $L$ is cyclic, and that there are necessary since the morphism is onto.

## 3. STRONGLY CYCLIC LANGUAGES

We now define the notion of a strongly language. The transitions of a deterministic automaton $\mathcal{A} = (Q, A, E)$ define a partial left action of $A^*$ on the set $Q$ of states. If $q \xrightarrow{w} q'$ is a path in the automaton labeled by a word $w$, we write $q' = q \cdot w$. For any state $q$, we have $q \cdot \varepsilon = q$ where $\varepsilon$ denotes the empty word. This action is extended to subsets by setting $P \cdot w = \{q \cdot w | q \in P\}$ for any subset $P$ of $Q$.

DEFINITION 2: *Let* $\mathcal{A} = (Q, A, E)$ *be a deterministic automaton where* $Q$ *is the set of states and* $E$ *the set of transitions. We say that a word* $w$ *stabilizes a nonempty subset* $P \subset Q$ *of states if we have* $P \cdot w = P$. *This means*

$$\begin{array}{lll} \forall p \in P & & p \cdot w \in P \\ \forall p' \in P & \exists p \in P & p \cdot w = p'. \end{array}$$

We denote by $\mathrm{Stab}(\mathcal{A})$ the set of the words $w$ such that $w$ stabilizes a nonempty subset $P$ of states in the automaton $\mathcal{A}$. It should be noticed that in this definition the subset $P$ of states stabilized by $w$ may depend on $w$ and that a word $w$ may stabilize several subset of $Q$. We say that a language $L$ is *strongly cyclic* if there is automaton $\mathcal{A}$ such that $L = \mathrm{Stab}(\mathcal{A})$. In this case, we say that the language $L$ *stabilizes the automaton* $\mathcal{A}$. The empty language $\varnothing$ is trongly cyclic since it stabilizes the empty automaton. The full language $A^*$ is also strongly cyclic since it stabilizes any complete automaton. Since

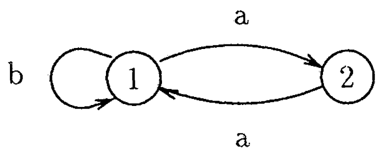


**Figure 1. – Automaton $\mathcal{A}_1$.**

the empty word stabilizes every subset of states, every nonempty strongly cyclic language contains the empty word.

EXAMPLE 2: *The language $(b + aa)^* + (ab^*a)^* + a^*$ is the strongly cyclic language associated with the automaton $\mathcal{A}_1$ of Figure 1. The subsets $\{1\}$, $\{2\}$ and $\{1, 2\}$ are respectively stabilized by the words of $(b + aa)^*$, $(ab^*a)^*$ and $a^*$.*

The following result gives a characterization of the words $w$ stabilizing a subset of states in an automaton. The proof of this proposition can be found in [BCR96].

PROPOSITION 2: *Let $\mathcal{A} = (Q, A, E)$ be a deterministic automaton. A word $w$ belongs to $\mathrm{Stab}(\mathcal{A})$ if and only if there is some state $q$ of $\mathcal{A}$ such that for every integer $n$, the transition $q \cdot w^n$ exists.*

Proposition 2 immediately implies that a strongly cyclic language is closed under power and root. It may also be directly verified that if a word $uv$ stabilizes a subset $P$ of states of an automaton, the word $vu$ stabilizes the set $P \cdot u$ of states. A strongly cyclic language is thus closed under conjugation. Putting together those two remarks, one obtains that a strongly cyclic language is cyclic and the terminology is justified.

Using Proposition 2, it may be easily verified that if $L_1$ and $L_2$ are two strongly cyclic languages stabilizing respectively automata $\mathcal{A}_1$ and $\mathcal{A}_2$, the union $L_1 + L_2$ stabilizes the disjoint union $\mathcal{A}_1 + \mathcal{A}_2$ and that the intersection $L_1 \cap L_2$ stabilizes the direct product $\mathcal{A}_1 \times \mathcal{A}_2$. Thus, the class of strongly cyclic languages is closed under union and intersection.

We now give some basic results. We first recall a characterization of strongly cyclic language and we state another characterization of these languages among rational cyclic languages. These results will be useful in the sequel.

The following theorem gives a characterization of the strongly cyclic languages. The proof of this theorem can be gound in [BCR96].

THEOREM 1: *Let $L$ be a rational language different from $A^*$. The following conditions are equivalent.*

*1. The language $L$ is strongly cyclic.*

*2. There is a morphism $\varphi$ from $A^*$ onto a monoid $M$ having a zero such that $L = \varphi^{-1}(\{s \in M | s^\omega \neq 0\})$.*

*3. The syntactic monoid $M(L)$ of $L$ has a zero and the image of $L$ in $M(L)$ is $\{s \in M(L) | s^\omega \neq 0\}$.*

Using Proposition 2, it can be shown that the transition monoid of an automaton stabilized by a strongly cyclic language $L$ has a zero which is the empty relation and that a relation $s$ of this monoid belongs to the image of $L$ if and only if $s^\omega \neq 0$. Conversely, the right representation of the syntactic monoid $L$ gives an automaton. The states are the elements of this monoids except 0 and the transitions are defined by the right action of the monoid on itself.

The following theorem characterizes strongly cyclic languages among cyclic languages. The proof of this theorem is based on the former one.

THEOREM 2: *Let $L$ be a rational cyclic language. Let $\psi : A^* \twoheadrightarrow M$ be a morphism from $A^*$ onto a finite monoid $M$ and such that $L = \psi^{-1}(P)$ for some $P \subset M$. The language $L$ is strongly cyclic if and only if for any idempotents $e$ and $f$ of $M$,*

$$\left.\begin{array}{l} e \in P \\ e \leq_{\mathcal{J}} f \end{array}\right\} \quad \Rightarrow \quad f \in P \tag{1}$$

*Proof:* We prove first that the Property (1) implies that the language $L$ is strongly cyclic. Let $J$ be the set of idempotents of $M$ not belonging to the image $P$ of $L$ and let $I$ be the ideal of $M$ generated by $J$. We have $J = E(M) - P$ and $I = MJM$ where $E(M)$ denotes the set of idempotents of the monoid $M$. We first prove that $I \cap P = \emptyset$. Let $s \in M$ be a element of $I$. The element $s$ can be written $s = xfy$ where $f$ is an idempotent of $J$ and $x$, $y \in M$. The idempotent $e = s^\omega$ satisfies $e \leq_{\mathcal{J}} f$. Since $f \notin P$, we have $e \notin P$ by Property (1). Since the language $L$ is cyclic, we also have $s \notin P$. Since $I \cap P = \emptyset$, the language $L$ is then recognized by the Rees quotient $M/I$. The language $L$ is then recognized by a monoid having a zero and this zero is the only idempotent not belonging to the image of $L$. By Theorem 1, the language $L$ is strongly cyclic.

Suppose now that the language $L$ is strongly cyclic. Let $M(L)$ be the syntactic monoid of $L$ and $\varphi$ the canonical morphism from $A^*$ onto $M(L)$.

Since the morphism $\psi$ is onto, the syntactic monoid $M(L)$ is a quotient of $M$: there is a morphism $\pi : M \twoheadrightarrow M(L)$ from $M$ onto $M(L)$ such that $\pi \circ \psi = \varphi$. Let $e$, $f$ be two idempotents of $M$ satisfying $e \in P$ and $e \leq_{\mathcal{J}} f$. The images $\pi(e)$ and $\pi(f)$ are two idempotents of $M(L)$ satisfying $\pi(e) \in \pi(P)$ and $\pi(e) \leq_{\mathcal{J}} \pi(f)$ because $\pi \circ \psi = \varphi$. Both idempotents $\pi(e)$ and $\pi(f)$ are then different from the zero of $M(L)$. We have then $\pi(f) \in \pi(P)$ by Theorem 1 and $f \in P$. This finishes the proof of the theorem. $\square$

## 4. CHAINS OF STRONGLY CYCLIC LANGUAGES

In this section, we introduce the notion of a chain of sets. We first define this notion in a general framework and we use it to define a hierarchy among cyclic languages. This hierarchy is based on the fact that every cyclic language can be decomposed as a chain of strongly cyclic languages. We show then that this hierarchy can be characterized by chains of idempotents in monoids. Indeed, the level of the hierarchy to which a given cyclic language belongs is completely determined by the length of chains of idempotents in a monoid recognizing the language.

### 4.1. Sum of differences and chains

For two subsets $X$ and $Y$ of a set $E$, the union and the difference of $X$ and $Y$ are respectively denoted by $X + Y$ and $X - Y$. The symmetric difference is denoted by $X \Delta Y = (X - Y) + (Y - X)$.

Let $\mathcal{F}$ be a family of sets closed under union and intersection but not necessarily under complement. Every set $X$ of the boolean closure of $\mathcal{F}$ is equal to a finite union of differences of sets of $\mathcal{F}$.

A *sum of differences* of length $m$ is an expression

$$X = (X_1 - X_2) + (X_3 - X_4) + \ldots + (X_{m-1} - X_m) \quad \text{if } m \text{ is even}$$
$$X = (X_1 - X_2) + (X_3 - X_4) + \ldots + X_m \quad \text{if } m \text{ is odd.}$$

Every set $X$ of the boolean closure of $\mathcal{F}$ is then equal to some sum of differences $X = (X_1 - X_2) + (X_3 - X_4) + \cdots$ where the sets $X_i$ belong to $\mathcal{F}$.

*A chain of differences* (or simply a *chain*) is a sum of differences where the sequence of subsets $X_1, \ldots, X_m$ satisfies the additional condition $X_1 \supset \ldots \supset X_m$. In this case, we write

$$X = X_1 - X_2 + X_3 - \ldots \pm X_m$$

where the sign $\pm$ in front of $X_m$ depends on the parity of $m$.

Chains of differences and sums of differences are related by the following result due to F. Hausdorff [Hau57, p. 92].

PROPOSITION 3: *If the family $\mathcal{F}$ is closed under union and intersection, every sum of differences is equal to a chain of differences of the same length.*

The proof of this result is based of the following property of chains. If the subsets $X$ and $Y$ are respectively equal to chains length $m$ and $n$, the sets $X + Y$ and $X \cap Y$ are equal to chains of length at most $m + n$. For a new proof of this result, *see* [Car93].

## 4.2. The hierarchy of cyclic languages

We can now define the hierarchy of cyclic languages over an alphabet $A$. Let $\mathcal{S}$ be the class of strongly cyclic languages. The boolean closure of $\mathcal{S}$ is the class $\mathcal{C}$ of cyclic languages. We define the class $\mathcal{C}_m$ of cyclic languages in the following way. For $m = 0$, we set $\mathcal{C}_0 = \{\varnothing\}$ and for $m \geq 1$, we denote by $\mathcal{C}_m$ the class of cyclic languages $X$ that are equal to a chain of length at most $m$ of strongly cyclic languages, *i.e.*,

$$X = X_1 - X_2 + X_3 - \ldots \pm X_m \quad \text{where} \quad X_i \in \mathcal{S}.$$

For $m = 1$, the class $\mathcal{C}_1 = \mathcal{S}$ is the class of strongly cyclic languages. For $m' \leq m$, we have $\mathcal{C}_{m'} \subset \mathcal{C}_m$. Since every cyclic language can be written as a boolean combination of strongly cyclic languages, we have the equality $\mathcal{C} = \bigcup_{m \geq 0} \mathcal{C}_m$.

This hierarchy classifies the cyclic languages according to their complexity. The strongly cyclic languages are simple languages. The level of the hierarchy to which a cyclic language belongs is the minimal number of strongly cyclic languages needed to express it as a boolean combination.

The results about chains of subsets (*see* [Hau57, Car93]) imply the following properties of the hierarchy introduced above.

PROPOSITION 4: *If $X \in \mathcal{C}_m$ and $Y \in \mathcal{C}_n$, we have then*

$$X \cap Y \in \begin{cases} \mathcal{C}_{m+n-2} & if\ m\ and\ n\ even \\ \mathcal{C}_{m+n-1} & otherwise \end{cases}$$

$$X + Y \in \begin{cases} \mathcal{C}_{m+n-1} & if\ m\ and\ n\ odd \\ \mathcal{C}_{m+n} & otherwise \end{cases}$$

## 5. CHAINS OF IDEMPOTENTS

In this section we define the chains of idempotents. This notion allows to characterize the classes of cyclic languages introduced above.

DEFINITION 3: *Let $M$ be a monoid and $P$ a subset of $M$. A* chain of idempotents *of length $m$ is a sequence $e_0, \dots e_m$ of idempotents of $M$ satisfying the following two conditions:*

(i) $e_0 \leq_{\mathcal{J}} e_1 \leq_{\mathcal{J}} \dots \leq_{\mathcal{J}} e_m$.

(ii) $e_0 \in P$ *and* $e_i \in P \Leftrightarrow e_{i+1} \notin P$.

The first condition means that the sequence $e_0, \dots, e_m$ is a increasing sequence for the $\mathcal{J}$-order. The second one means that the idempotents $e_i$ are alternately in $P$ and out of $P$ and that the first idempotent $e_0$ of the sequence is in $P$.

We denote by $m(S, P)$ the maximal length of a chains. We set $m(S, P) = +\infty$ if the length of the chains is not bounded.

The following theorem states that the maximal length of the chains is a syntactic invariant. The integer $m(S, P)$ does not depend of the monoid considered, it just depends on the language recognized.

THEOREM 3: *Let $L$ be a rational language. Let $\varphi : A^* \twoheadrightarrow M$ and $\psi : A^* \twoheadrightarrow N$ be two morphisms from $A^*$ onto finite monoids $M$ and $N$ such that $L = \varphi^{-1}(P)$ and $L = \psi^{-1}(Q)$. We have then $m(M, P) = m(N, Q)$.*

*Proof:* It is sufficient to prove the result when $M$ is the syntactic monoid $M(L)$ of $L$. We suppose then that $M$ is the syntactic monoid of $L$ and that $\varphi$ is the canonical morphism from $A^*$ onto $M$. Since the morphism $\psi$ is onto, the monoid $M$ is a quotient of $N$: there is a morphism $\pi : N \twoheadrightarrow M$ from $N$ to $M$ such that $\pi \circ \psi = \varphi$. Since $\pi \circ \psi = \varphi$, we have $\pi(Q) = P$. We show that we can associate to any chain of idempotents of length $m$ in $N$, a chain of idempotents of the same length in $M$ and conversely.

Let $e_0, \dots, e_m$ be a chain of idempotents in $N$. The sequence $\pi(e_0), \dots, \pi(e_m)$ is then a chain of idempotents in $M$. Obviously, the elements $\pi(e_i)$ are idempotents and these idempotents are ordered with respect to the $\mathcal{J}$-order. Since $\pi \circ \psi = \varphi$, we also have $e_i \in Q \Leftrightarrow \pi(e_i) \in P$. This implies $\pi(e_0) \in P$ and $\pi(e_i) \in P \Leftrightarrow \pi(e_{i+1}) \notin P$.

Let $f_0, \dots, f_m$ be a chain of idempotents in $M$. Since $f_0 \leq_{\mathcal{J}} \dots \leq_{\mathcal{J}} f_m$, there are $2m$ elements $y_i, y'_i$ of $M$ such that $y_i f_i y'_i = f_{i-1}$ for $1 \leq i \leq m$.

We choose elements $t_i$, $x_i$ and $x'_i$ of $N$ such that $\pi(t_i) = f_i$, $\pi(x_i) = y_i$ and $\pi(x'_i) = y'_i$. We define the idempotents $e_i$ of $N$ by

$$\begin{aligned}
e_m &= t_m^\omega \\
e_{m-1} &= (x_m e_m x'_m)^\omega \\
e_{m-2} &= (x_{m-1} e_{m-1} x'_{m-1})^\omega \\
&\vdots \\
e_0 &= (x_1 e_1 x'_1)^\omega
\end{aligned}$$

By definition, the sequence $e_0, \ldots, e_m$ is a sequence of idempotents ordered for the $\mathcal{J}$-order. Since $\pi(e_i) = f_i$, we have $e_0 \in Q$ and $e_i \in Q \Leftrightarrow e_{i+1} \in Q$ and the sequence $e_0, \ldots, e_m$ is a chain of idempotents.

Since for each chains in $M$ of length $m$ there exists a chain in $T$ of length $m$ and vice-versa, we have proved that $m(M, P) = m(N, Q)$. $\square$

Since the integer $m(M, P)$ only depends on the language recognized and not on the monoid considered, we can define $m(L)$ as $m(M, P)$ for any morphism $\varphi : A^* \twoheadrightarrow M$ from $A^*$ onto a finite monoid $M$ such that $L = \varphi^{-1}(P)$ for some $P \subset M$.

The definition of chains of idempotents is motivated by the following result.

THEOREM 4: *Let $L$ be a rational cyclic language. Let $\varphi : A^* \twoheadrightarrow M$ be a morphism from $A^*$ onto a finite monoid $M$ such that $L = \varphi^{-1}(P)$ for some $P \subset M$. We have then*

$$L \in C_m \quad \Leftrightarrow \quad m(M, P) \leq m - 1.$$

We first prove the following lemma which states that the function $m$ is "subadditive".

LEMMA 1: *Let $X$ and $Y$ be two rational languages. We have then*

$$m(X \Delta Y) \leq m(X) + m(Y) + 1.$$

*Proof:* We suppose that the languages $X$ and $Y$ are respectively recognized by the morphisms $\varphi : A^* \twoheadrightarrow M$ and $\psi : A^* \twoheadrightarrow N$ from $A^*$ onto the finite monoids $M$ and $N$. Let $P$ and $Q$ be the images of $X$ and $Y$ in $M$ and $N$. We have $X = \varphi^{-1}(P)$ and $Y = \psi^{-1}(Q)$. By definition, we have $m(X) = m(M, P)$ and $m(Y) = m(N, Q)$. The language $X \Delta Y$ is recognized by the morphism $\varphi \times \psi : A^* \to M \times N$ where $M \times N$ is the product of $M$ and $N$. The morphism $\varphi \times \psi$ may not be onto. Let $R$ be the submonoid of $M \times N$ defined by $R = \varphi \times \psi(A^*)$. The language $X\Delta Y$ is then recognized by the morphism $\varphi \times \psi : A^* \twoheadrightarrow R$ and the image of $X\Delta Y$ in $R$ is given by

$$\varphi \times \psi(X\Delta Y) = (P \times (N - Q) + (M - P) \times Q) \cap R.$$

We prove that if there is a chain in $R$ of length $m$, there are two integers $p$ and $q$ satisfying $p + q \geq m - 1$, a chain in $M$ of length $p$ and a chain in $N$ of length $q$. Let $(e_0, f_0), \ldots, (e_m, f_m)$ be a chain of idempotents in $R$. We consider the integers $i$ for which one of the idempotents $e_{i-1}, e_i$ belongs to $P$ and the other does not. We also consider the integers $j$ for which one of the idempotents $f_{j-1}, f_j$ belongs to $Q$ and the other does not. Formally, we define the sets of integers $I$ and $J$ by

$$I = \{1 \leq i \leq m | e_{i-1} \in P \Leftrightarrow e_i \notin P\}$$
$$J = \{1 \leq j \leq m | f_{j-1} \in Q \Leftrightarrow f_j \notin Q\}$$

The sequence $(e_0, f_0), \ldots, (e_m, f_m)$ is a chain in $R$. Every integer $1 \leq k \leq m$ belongs to exactly one on the sets $I$ and $J$. Otherwise, both idempotents $(e_{k-1}, f_{k-1})$ and $(e_k, f_k)$ of $R$ are in the image of $X\Delta Y$ or out of the image of $X\Delta Y$. We set $I = \{i_1 < \ldots < i_p\}$ and $J = \{j_1 < \ldots < j_l)$ where $p$ and $l$ are the cardinals of $I$ and $J$. We have then $p + l \geq m$. Since the idempotent $(e_0, f_0)$ belongs to the image of $X\Delta Y$ in $R$, if $e_0$ belongs to $P$, $f_0$ does not belong to $Q$ and conversely. By symmetry, we suppose that $e_0$ belongs to $P$. The sequences $e_0, e_{i_1}, \ldots, e_{i_p}$ and $f_{j_1}, \ldots, f_{j_l}$ are respectively chains in $M$ and $N$ of length $p$ and $q = l - 1$. We then have $m \leq p + l \leq p + q + 1 \leq m(X) + m(Y) + 1$. □

We can now complete the proof of the theorem.

*Proof:* We suppose first that $L \in C_m$. The language $L$ can be written $L = X_1 - X_2 + \ldots \pm X_m$ or equivalently $L = X_1\Delta \ldots \Delta X_m$ with $X_i$ strongly cyclic language. By Theorem 2, we have $m(X_i) = 0$ and the preceding lemma implies that $m(L) \leq m - 1$.

We suppose now that $m(X) \leq m - 1$. For an idempotent $e$ of $M$, we denote by $m(e)$, the maximal length of a chain $e_0, \ldots, e_n$ such that $e_n = e$. We have of course the inequality $m(e) \leq m(M, P)$ for any idempotent $e$ of $M$. Let $J_k$ be the set of idempotents $J_k = \{e \in M | m(e) \geq k\}$. By construction, we have that $e \in J_k$ and $e \leq_{\mathcal{J}} f$ imply that $f \in J_k$. Let $P_k$ be the subset $P_k = \{s \in M | s^\omega \in J_k\}$. Since every idempotent $e$ satisfies

$e^\omega = e$, we have for any idempotents $e$ and $f$ of $M$

$$\left.\begin{array}{l} e \in P_k \\ e \leq_{\mathcal{J}} f \end{array}\right\} \quad \Rightarrow \quad f \in P_k.$$

By Theorem 2, the languages $X_k = \varphi^{-1}(P_k)$ are then strongly cyclic. Since the language $L$ is cyclic, a element $s$ of $M$ belongs to $P$ if and only if $s^\omega$ belongs to $P$. We have then $L = X_0 - X_1 \ldots \pm X_{m-1}$. □

The previous theorem can be used to give an another proof that any cyclic language is a boolean combination of strongly cyclic languages. To get this result, we must prove that any cyclic language belong to the class $C_m$ for some integer $m$. By the previous Theorem, it is sufficient to prove that the length of chains of idempotents in a monoid recognizing $L$ is bounded. We have the following proposition.

PROPOSITION 5: *Let $L$ be a rational cyclic language. Let $\varphi : A^* \twoheadrightarrow M$ be a morphism from $A^*$ onto a finite monoid $M$ such that $L = \varphi^{-1}(P)$ for some $P \subset M$. Let $n$ be the number of $\mathcal{D}$-classes of the monoid $M$. We have then the inequality*

$$m(M, P) \leq n.$$

*Proof:* Let $e_0, \ldots, e_m$ be a chain of idempotents in $M$. The idempotents $e_i$ satisfy $e_{k-1} \leq_{\mathcal{J}} e_k$ for $1 \leq k \leq m$. We will see that all these inequalities are strict. The idempotents $e_i$ satisfy in fact $e_{k-1} \leq_{\mathcal{J}} e_k$. Suppose that one of the inequality is not strict. Two idempotent $e_{k-1}$ and $e_k$ belongs to the same $\mathcal{D}$-class and are then conjugated. There are two elements $x$ and $y$ of $M$ such that $xy = e_{k-1}$ and $yx = e_k$. Since the language $L$ is cyclic, we have by Proposition 1, $e_{k-1} \in P \Leftrightarrow e_k \in P$ and this leads to a contradiction. The idempotents $e_i$ belong to different $\mathcal{D}$-classes and the length of the chain is bounded by the number of $\mathcal{D}$-classes of the monoid $M$. □

## 6. CLOSURE OF A CYCLIC LANGUAGE

In this section, we first prove that for any cyclic language $L$, there is a smallest strongly cyclic language containing $L$.

We first recall that the syntactic monoid of a rational cyclic language has a zero. It has been proved in [BCR96, Cor. 5].

THEOREM 5: *Let $L$ be a rational cyclic language and $\varphi : A^* \twoheadrightarrow M$ the canonical morphism from $A^*$ onto the syntactic monoid $M$ of $L$. There* *is then a smallest strongly cyclic language containing L. This language is* $\overline{L} = \varphi^{-1}(\overline{P})$ *where* $\overline{P} = \{s | s^{\omega} \neq 0\}$ *if the zero of* $M$ *does not belong to the image of* $L$ *in* $M$ *and is* $A^*$ *otherwise.*

We point out that the result is false if the monoid considered is not the syntactic monoid. Let us consider the strongly cyclic language $L = b^*$ over the alphabet $A = \{a, b\}$. The syntactic monoid of $L$ is the monoid $\{b = 1, a = 0\}$. The language $L$ is also recognized by the idempotent monoid $M = \{1, a, b, ab = ba = 0\}$ with the canonical morphism from $A^*$ onto this monoid. The image of $L$ in $M$ is $P = \{1, b\}$ but the subset $\overline{P}$ is $\{1, a, b\}$. The language $\overline{L}$ is then $a^* + b^*$ which is not the smallest strongly cyclic language containing $L$.

*Proof:* We first consider the case in which the zero of $M$ does not belong to the image of $L$ in $M$. The language $\overline{L} = \varphi^{-1}(\overline{P})$ where $\overline{P} = \{s | s^{\omega} \neq 0\}$ is strongly cyclic by Theorem 1 and contains the language $L$. Let prove now that this language is the smallest one.

Let $X$ be a strongly cyclic language containing $L$ and $w$ be a word of $\overline{L}$. Let $\mathcal{A} = (Q, A, E)$ be a deterministic automaton such that $X = \mathrm{Stab}(\mathcal{A})$. By definition, we have $\varphi(w) = s$ where $s^{\omega} \neq 0$. For every integer $n$, the element $\varphi(s)^n$ is different from the zero of $M$. There are two words $x_n$ and $y_n$ such that $x_n w^n y_n$ belongs to $L$. By Proposition 2, there is a state $q_n$ of $\mathcal{A}$ such that the transition $q_n \cdot x_n w^n y_n$ is defined. The transition $(q_n \cdot x_n) \cdot w^n$ is then defined and the word $w$ belongs to $X$. We have proved that $\overline{L} \subset X$. The language $\overline{L}$ is then the smallest strongly cyclic language containing $L$.

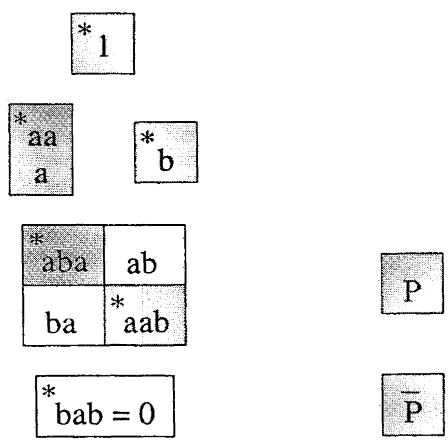


**Figure 2. – Structure of the syntactic monoid of** $L$.

Let us now consider the case in which the zero of $M$ does belong to the image of $L$ in $M$. In this case, the languages $L$ intersects every ideal $I$ of $A^*$, *i.e.*, $L \cap I \neq \emptyset$. Let $X$ be a strongly cyclic language different from $A^*$. By Theorem 1, the syntactic monoid of $X$ has a zero which does not belong to the image of $X$. The language $X$ does not intersect the ideal equal to the inverse image of 0 and cannot contain the language $L$. The only strongly cyclic language containing $L$ is then $A^*$. □

EXAMPLE 3: *Let $L$ be the language $(b + aa)^* + (ab^*a)^* + a^* - b^*$. The structure of the syntactic monoid of $L$ is given in Figure 2. The image $P$ of $L$ in $M(L)$ is equal to $P = \{a, aa, aba, aab\}$.*

*The subset $\overline{P}$ defined in the proof is equal to $\overline{P} = \{1, a, aa, b, aba, aab\}$ and the language $\overline{L}$ is $(b + aa)^* + (ab^*a)^* + a^*$.*

## 7. APPROXIMATIONS BY CHAINS

In this section, we will see how the existence of a smallest strongly cyclic language $\overline{L}$ containing a cyclic language $L$ can be used to compute a chain of strongly cyclic languages equal to the language $L$.

We remark that if the language $L$ is equal to the chain $L = X_1 - \ldots \pm X_m$, the languages $L_k$ for $1 \leq k \leq m$ defined by $L_k = X_1 - \ldots \pm X_k$ satisfy

$$\begin{aligned} L_k &\supset L \quad \text{if } k \text{ is odd} \\ L_k &\subset L \quad \text{if } k \text{ is even.} \end{aligned}$$

Suppose now that the language $L$ is equal to the chain $L = X_1 - \ldots \pm X_m$ where the languages $X_i$ are strongly cyclic. We set $L_k = X_1 - \ldots \pm X_k$ for $1 \leq k \leq m$. We introduce two other sequences of languages $Y_i$ and $M_i$ defined by $Y_1 = M_1 = \overline{L}$ and

$$\begin{aligned} Y_k &= \overline{M - M_{k-1}} \quad \text{and} \quad M_k = M_{k-1} + Y_k \quad \text{if } k \text{ is odd} \\ Y_k &= \overline{M_{k-1} - M} \quad \text{and} \quad M_k = M_{k-1} - Y_k \quad \text{if } k \text{ is even.} \end{aligned}$$

In particular, we have $Y_2 = \overline{\overline{L} - L}$ and $M_2 = \overline{L} - \overline{\overline{L} - L}$.

By definition, the languages $Y_i$ are strongly cyclic. The following theorem states that the languages $Y_i$ form a chain and this chain is the best approximation of the language $L$.

THEOREM 6: *Let $L$ be a cyclic language equal to $L = X_1 - \ldots \pm X_m$ where the languages $X_i$ are strongly cyclic. Let $L_k = X_1 - \ldots \pm X_k$ for $1 \leq k \leq m$. Define the languages $Y_i$ and $M_i$ by $Y_1 = M_1 = \overline{L}$ and*

$$\begin{array}{lcll} Y_k = \overline{M - M_{k-1}} & \text{and} & M_k = M_{k-1} + Y_k & \text{if } k \text{ is odd} \\ Y_k = \overline{M_{k-1} - M} & \text{and} & M_k = M_{k-1} - Y_k & \text{if } k \text{ is even.} \end{array}$$

*The languages $Y_i$ and $M_i$ then satisfy*

1. *For any $1 \leq k \leq m$, $Y_k$ is strongly cyclic.*
2. $Y_1 \supset \ldots \supset Y_m$.
3. *For any $1 \leq k \leq m$,*

$$\begin{array}{ll} L_k \supset M_k \supset L & \text{if } k \text{ is odd} \\ L_k \subset M_k \subset L & \text{if } k \text{ is even.} \end{array}$$

The last inclusions mean that each set $M_i$ is closer to $L$ than the set $L_i$. In particular, if $L$ is equal to a chain of length $m$ of cyclic languages, the language $M_m$ computed by the previous procedure is equal to $L$. The chain $L = Y_1 - \ldots \pm Y_m$ computed is then the closest (in the sense of the inclusions) and shortest chain of strongly cyclic languages equal to $L$.

*Proof:* We introduce the functions $f$ and $g$ defined on $\mathcal{P}(A^*)$ by:

$$f(X) = X + \overline{L - X}$$
$$g(X) = X - \overline{X - L}$$

The key property of the functions $f$ and $g$ is expressed in the following lemma.

LEMMA 2: *The functions $f$ and $g$ satisfy the following properties:*

$$\begin{array}{lcl} X \subset Y \subset L & \Rightarrow & f(X) \supset f(Y) \supset f(L) = L \\ X \supset Y \supset L & \Rightarrow & g(X) \subset g(Y) \subset g(L) = L \end{array}$$

*Proof:* An easy calculation proves that $L$ is a fixed point of $f$ and $g$, *i.e.*, $f(L) = L$ and $g(L) = L$.

$$f(L) = L + \overline{L - L} = L + \overline{\varnothing} = L$$
$$g(L) = L - \overline{L - L} = L - \overline{\varnothing} = L$$

Suppose now that $X \subset Y \subset L$. The inclusion $L - X \supset Y - X$ implies $\overline{L - X} \supset Y - X$. We have $X + \overline{L - X} = Y + \overline{L - X} \supset Y + \overline{L - Y}$ since

$\overline{L - X} \supset \overline{L - Y}$. This ends the proof of property of $f$. The property of $g$ is handled in the same way. □

Since the languages $M_i$ can be defined by

$$M_k = f(M_{k-1}) \quad \text{if } k \text{ is odd}$$
$$M_k = g(M_{k-1}) \quad \text{if } k \text{ is even.}$$

we can easily complete the proof of the theorem. □

ACKNOWLEDGMENT

We would like to sincerely thank the anonymous referee for his pertinent remarks and his very detailed report on the first version of this paper.

## REFERENCES

[BCR96] M.-P. Béal, O. Carton and C. Reutenauer, Cyclic languages and strongly cyclic languages. In *STACS' 96*, vol. 1046 of *Lect. Notes in Comput. Sci.*, 1996, pp. 49-59.

[Béa95] M.-P. Béal, Puissance extérieure d'un automate déterministe, application au calcul de la fonction zêta d'un système sofique. *RAIRO-Informatique Théorique et Applications*, 1995, *29* (2), pp. 85-103.

[BR90] J. Berstel and C. Reutenauer, Zeta functions of formal languages. *Trans. Amer. Math. Soc.*, 1990, *321*, pp. 533-546.

[Car93] O. Carton, *Mots infinis, $\omega$-semigroupes et topologie.* Thèse, Université Paris 7. Rapport LITP-TH 93-08, 1993.

[Hau57] F. Hausdorff, *Set Theory*, Chelsea, New York, 1957.

[Lal79] G. Lallement, *Semigroups and combinatorial applications*, Wiley, 1979.

[Pin86] J.-E. Pin, *Varieties of formal languages*, North Oxford, London and Plenum, New York, 1986 (Traduction de *Variétés de langages formels*).