A hierarchy of cyclic languages

Olivier Carton
Institut Gaspard Monge
Université de Marne-la-Vallée
93166 Noisy le Grand cedex

July 31, 1995

Abstract

We introduce a hierarchy of cyclic languages based on the boolean combinations of strongly cyclic languages. We then show how this hierarchy can be characterized by chains of idempotents in semigroups. Finally, we give a method to compute an optimal decomposition of a cyclic language into strongly cyclic languages.

Résumé

Nous introduisons une hiérarchie des langages cycliques basée sur les combinaisons booléennes de langages fortement cycliques. Nous montrons ensuite comment cette hiérarchie peut être caractérisée par des chaînes d'idempotents dans des semigroupes. Finalement, nous donnons une méthode pour calculer une décomposition optimale 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 language 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. A strongly cyclic language is rational.

It has been proved in [BCR] that any 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 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. This hierarchy measures the number of

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

We assume that the reader is familiar with the basic notions of automata and semigroup theory. For example notions like syntactic semigroup, 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 semigroups 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 semigroup S, every element s of S has a power which is an idempotent. We denote by s^ω this idempotent.

Definition 1 *A language L of A^* is said to be* cyclic *if it satisfies*

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

A language is cyclic if it is closed under conjugation, power and root.

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 finite semigroups.

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

$$\begin{array}{lll} \forall s \in S,\ \forall n > 0 & s \in P & \Leftrightarrow \quad s^n \in P \\ \forall s, t \in S & st \in P & \Leftrightarrow \quad ts \in P \end{array}$$

3 Strongly cyclic languages

We now define the notion of a strongly cyclic language.

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 subset $P \subset Q$ of states if we have $P.w = P$. This means*

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

We denote by $\mathrm{Stab}(\mathcal{A})$ the set of the words w such that w stabilizes a 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. 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* $\mathcal{A}$.

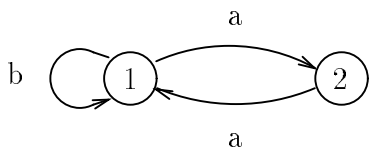

Figure 1: Automaton $\mathcal{A}_1$

Example 2 *The language $(b + aa)^* + (ab^*a)^* + a^*$ and b^* is the strongly cyclic language associated with the automaton $\mathcal{A}_1$ of Figure 1.*

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 [BCR].

Proposition 2 *Let $\mathcal{A} = (Q, A, E)$ be a deterministic automaton. A word w stabilizes a subset P of states in $\mathcal{A}$ if and only if there is some state q of $\mathcal{A}$ such that for any integer n, the transition $q.w^n$ exists.*

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 found in [BCR].

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 φ from A^* onto a semigroup S having a zero such that $L = \varphi^{-1}(\{s \in S \mid s^\omega \neq 0\})$.*
3. *The syntactic semigroup $S(L)$ of L has a zero and the image of L in $S(L)$ is $\{s \in S \mid s^\omega \neq 0\}$.*

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

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

Proof : We prove first that the Property (1) implies that the language L is strongly cyclic. Let J be the set of idempotents of S not belonging to the image P of L and let I be the ideal of S generated by J. We have $J = E(S) - P$ and $I = S^1 J S^1$ where $E(S)$ denotes the set of idempotents of the semigroup S. We prove first prove that $I \cap P = \emptyset$. Let $s \in S$ be a element of I. The element s can be written $s = xfy$ where f is an idempotent of J and $x, y \in S^1$. 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 $s \notin P$. We now prove that all the elements of I are equivalent for the Nerode congruence. For any $x, y \in S^1$ and $s \in I$, the element xsy also belongs to I. Since $I \cap P = \emptyset$, the contexts of s are $\emptyset \times \emptyset$. The language L is then recognized by the Rees quotient S/I. The language L is then recognized by a semigroup 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 $S(L)$ be the syntactic semigroup of L and φ the canonical morphism from A^* onto $S(L)$. Since the morphism ψ is onto, the syntactic semigroup $S(L)$ is a quotient of S : there is a morphism $\pi : S \twoheadrightarrow S(L)$ from S onto $S(L)$ such that $\pi \circ \psi = \varphi$. Let e, f be two idempotents of S satisfying $e \in P$ and $e \leq_{\mathcal{J}} f$. The images $\pi(e)$ and $\pi(f)$ are two idempotents of $S(L)$ 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 $S(L)$. We have then $\pi(f) \in \pi(P)$ by Theorem 1 and $f \in P$. This finishes the proof of the theorem. □

4 Chains of strongly cyclic languages

In this section, we introduce the notion of a chain of sets. We use this general notion 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 semigroups. 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 semigroup recognizing the language.

For X and Y two subsets of a set E, the complement of X is denoted by X^c. The union of X and Y is denoted by $X + Y$ and the intersection is denoted by XY. The difference set $X \setminus Y = X \cap Y^c$ is denoted by $X - Y$. The symmetric difference is denoted by $X \triangle Y = XY^c + X^cY$.

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 intersections of sets of $\mathcal{F}$ and complements of sets of $\mathcal{F}$.

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

$$\begin{array}{ll} X = (X_1 - X_2) + (X_3 - X_4) + \cdots + (X_{m-1} - X_m) & \text{if } m \text{ is even} \\ X = (X_1 - X_2) + (X_3 - X_4) + \cdots + X_m & \text{if } m \text{ is odd} \end{array}$$

Every set X of the boolean closure of $\mathcal{F}$ is then equal to 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 $X_1 \supset \cdots \supset X_m$. In this case, we write

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

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

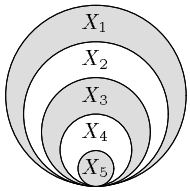

Figure 2: Chain of length 5

Example 3 *A chains of differences of length* 5 *is shown in Figure 2. The sets* X_i *are represented by circles of decreasing sizes. The set* $X = X_1 - X_2 + X_3 - X_4 + X_5$ *is marked in grey.*

The chains of differences and the sums of differences are related by the following result due to F. Hausdorff [Hau57].

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 XY are equal to chains of length at most $m+n$. For a new proof of this result, see [Car93].

We can now define the hierarchy of cyclic languages over an alphabet A. Let $\mathcal{S}$ the class of strongly cyclic languages. The boolean closure of $\mathcal{S}$ is the class $\mathcal{C}$ of cyclic languages. We define the class C_m of cyclic languages in the following way. For $m = 0$, we set $C_0 = \{A^*\}$ and for $m \geq 1$, we denote by 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.*,

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

For $m = 1$, the class $C_1 = \mathcal{S}$ is the class of strongly cyclic languages. For $m' \leq m$, we have $C_{m'} \subset C_m$. Since every cyclic language can be written as a boolean combination of strongly cyclic languages, we have the equality $\mathcal{C} = \bigcup_{m \geq 0} C_m$.

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]) implies the following properties of the hierarchy introduced above.

Proposition 4 *If $X \in C_m$ and $Y \in C_n$, we have then*

$$XY \in \begin{cases} C_{m+n-2} & \text{if } m \text{ and } n \text{ even} \\ C_{m+n-1} & \text{otherwise} \end{cases}$$

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

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 S be a semigroup and P a subset of P. A* chain of idempotents *of length m is a sequence $e_0, \ldots, e_m$ of idempotents of S satisfying the following two conditions:*

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

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

The first condition means that the sequence $e_0, \ldots, e_m$ 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 semigroup considered, it just depends on the language recognized.

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

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

Let $e_0, \ldots, e_m$ be a chain of idempotents in T. The sequence $\pi(e_0), \ldots, \pi(e_m)$ is then a chain of idempotents in S. Obviously, the elements $\pi(e_i)$ are idempotents and these idempotents are ordered for 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 $f_0, \ldots, f_m$ be a chain of idempotents in S. Since $f_0 \leq_{\mathcal{J}} \cdots \leq_{\mathcal{J}} f_m$, there are $2m$ elements y_i, y_i' of S^1 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 T 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 T 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 the chains of idempotents in S relative to P are in correspondence with the chains of idempotents in T relative to Q, we have proved that $m(S, P) = m(T, Q)$. □

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

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

Theorem 4 *Let L be a rational cyclic language. Let $\varphi : A^* \twoheadrightarrow S$ be a morphism from A^* onto a finite semigroup S such that $L = \varphi^{-1}(P)$. We have then*

$$L \in C_m \iff m(S,P) \leq m-1$$

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

Lemma 1 *Let X and Y two rational languages. We have then*

$$m(X \triangle Y) \leq m(X) + m(Y) + 1$$

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

$$\varphi \times \psi(X \triangle Y) = P \times (T - Q) + (S - P) \times Q$$

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 S of length p and a chain in T 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

$$\begin{aligned} I &= \{i \in [1;m] \mid e_{i-1} \in P \Leftrightarrow e_i \notin P\} \\ J &= \{j \in [1;m] \mid f_{j-1} \in Q \Leftrightarrow f_j \notin Q\} \end{aligned}$$

The sequence $(e_0, f_0), \ldots, (e_m, f_m)$ is a chain in R. every integer k (in $[1;m]$) 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 $X \triangle Y$ or out of the image of $X \triangle Y$. We set $I = \{i_1 < \cdots < i_p\}$ 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 $X \triangle Y$ 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 S and T of length p and $q = l - 1$. We have then $p + q \geq m - 1$. In particular, if the integer m is strictly greater than $m(X) + m(Y) + 1$, p is greater than $m(X)$ or q is greater than $m(Y)$ and this leads to a contradiction. □

We can now complete the proof of the theorem.

Proof : We suppose first that $L \in C_m$. The language L can be written $L = X_1 - X_2 + \cdots \pm X_m$ or equivalently $L = X_1 \triangle \cdots \triangle X_m$ 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 S, 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 $m(e) \leq m(S, P)$ for any idempotent e of S. Let J_k be the set of idempotents $J_k = \{e \in S \mid m(e) \geq k\}$. By construction, we have that $e \in J_k$ and $e \leq_{\mathcal{J}} f$ imply that $f \in J_k$. The language $X_k = \varphi^{-1}(P_k)$ where $P_k = \{s \in S \mid s^\omega \in J_k\}$ is then strongly cyclic language by Theorem 2 and we have $L = X_0 - X_1 \cdots \pm X_{m-1}$. □

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 semigroup recognizing L is bounded. We have have the following proposition.

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

$$m(S, P) \leq n$$

Proof : Let $e_0, \ldots, e_m$ be a chain of idempotents in S. 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 equalities are strict. The idempotents e_i satisfy in fact $e_{k-1} <_{\mathcal{J}} e_k$. Suppose that one of the equality is not strict. Two idempotent e_{k-1} and e_k belong to the same $\mathcal{D}$-class and are then conjugated. There are two elements x and y of S 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 semigroup S. □

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.

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

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

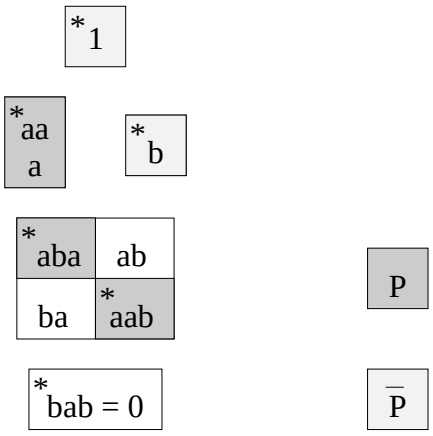

Figure 3: Structure of the syntactic semigroup of L.

Proof : We first consider the case in which the zero of S does not belong to the image of L in S. The language $\overline{L} = \varphi^{-1}(\overline{P})$ where $\overline{P} = \{s \mid s^\omega \neq 0\}$ 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 S. There 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 the transition $q_n.x_n w^n y_n$. The transition $(q_n.x_n).w^n$ 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.

Let now consider the case in which the zero of S does belong to the image of L in S. In this case, the languages L intersects every ideal I of A^*, *i.e.*, $L \cap I \neq \emptyset$. Let X a strongly cyclic language different from A^*. By Theorem 1, the syntactic semigroup 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 can not contain the language L. The only strongly cyclic language containing L is then A^*. □

Example 4 *Let L be the language $(b+aa)^* + (ab^*a)^* + a^* - b^*$. The structure of the syntactic semigroup of L is given on figure 3. The image P of L in $S(L)$ 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 $L = X_1 - \cdots \pm X_m$, the languages L_k for $1 \leq k \leq m$ defined by $L_k = X_1 - \cdots \pm X_k$ satisfy

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

Suppose now that the language L is equal to the chain $L = X_1 - \cdots \pm X_m$ where the languages X_i are strongly cyclic. We set $L_k = X_1 - \cdots \pm X_k$ 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

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

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 *The languages Y_i and M_i satisfy the following inclusions*

1. $Y_1 \supset \cdots \supset Y_m$.
2. *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 $L = Y_1 - \cdots \pm Y_m$ 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:

$$\begin{aligned} f(X) &= X + \overline{L - X} \\ g(X) &= X - \overline{X - L} \end{aligned}$$

The key property is expressed in the following lemma

Lemma 2 *The functions f and g verify the following properties:*

$$\begin{aligned} X \subset Y \subset L &\implies f(X) \supset f(Y) \supset f(L) = L \\ X \supset Y \supset L &\implies g(X) \subset g(Y) \subset g(L) = L \end{aligned}$$

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

$$\begin{aligned} f(L) &= X + \overline{L - L} = X + \overline{\emptyset} = X \\ g(L) &= X - \overline{L - L} = X - \overline{\emptyset} = X \end{aligned}$$

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

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

we can easily complete the proof of the theorem. □

References

[BCR] Marie-Pierre Béal, Olivier Carton, and Christophe Reutenauer. Cyclic languages and strongly cyclic languages. Submitted to STACS'96.

[Béa95] Marie-Pierre Béal. Puissance extérieure d'un automate déterministe, application au calcul de la fonction zêta d'un système sofique. *R.A.I.R.O.-Informatique Théorique et Applications*, 29(2):85–103, 1995.

[BR90] Jean Berstel and Christophe Reutenauer. Zeta functions of formal languages. *Trans. Amer. Math. Soc.*, 321:533–546, 1990.

[Car93] Olivier Carton. *Mots infinis, ω-semigroupes et topologie.* Thèse, Université Paris 7, 1993. Rapport LITP-TH 93-08.

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

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

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