INFORMATIQUE THÉORIQUE ET APPLICATIONS

JEAN-MICHEL AUTEBERT
LUC BOASSON
MICHEL LATTEUX

Motifs et bases de langages

Informatique théorique et applications, tome 23, nº 4 (1989), p. 379-393

<http://www.numdam.org/item?id=ITA_1989__23_4_379_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 23, n° 4, 1989, p. 379 à 393)

MOTIFS ET BASES DE LANGAGES (*)

par Jean-Michel AUTEBERT (1), Luc BOASSON (1) et Michel LATTEUX (2)

Communiqué par J.-E. PIN

Résumé. – *Dans ce papier, nous introduisons les notions de prémotif, base et motif d'un langage. Un langage P est un prémotif de L si L est union de puissances de P. P est une base de L si c'est un prémotif inclus dans L; c'est un motif s'il n'admet aucun prémotif autre que lui-même. Nous démontrons que tout langage admet un motif (respectivement une base), et que tout langage rationnel admet un motif rationnel. En outre, nous donnons un algorithme qui permet de décider si un langage rationnel admet un motif fini, et de le construire (s'il existe). Restent ouvertes les questions de l'existence d'un motif minimal pour l'inclusion et de l'existence d'un motif (resp. d'une base) minimal rationnel pour les langages rationnels.*

Abstract. – *In this paper, we introduce the notions of "prémotif", "base" and "motif" of a language. A language P is a prémotif of L if L is a union of powers of P. P is a base of L if it is a prémotif included in L; it is a motif if it has no other prémotif than itself. We prove that every language has a motif (respectively a base), and that every rational language has a rational motif. Moreover, we give an algorithm to decide whether a rational language has a finite motif or not, and to construct this latter (if it exists). The questions of the existence of a minimal (with respect to the inclusion) motif and of the existence of a rational minimal motif (or base) for rational languages are still open.*

Les résultats présentés dans cet article ont pour origine la question suivante :

Étant donnés deux ensembles A et B tels que $AB=BA$, existe-t-il toujours un ensemble C tel que $A=\bigcup_{i\in I} C^i$ et $B=\bigcup_{j\in J} C^j$? On sait qu'il n'en est rien, même si A et B sont finis : soient $X=\{x, y\}$, $A=X^2+x^3$ et $B=X+X^2+x^3$; alors $AB=BA$ sans que A et B ne soient unions de puissances d'un même ensemble C. Par contre, si A et B sont des codes, la question reste posée [2].

(*) Reçu juin 1987, version définitive mars 1988.
(1) U.E.R. de Mathématiques et Informatique, Université Paris-VII, 4, place Jussieu, 75005 Paris.
(2) C.N.R.S.-U.A. n° 369, Université de Lille-Flandres-Artois, B.P. n° 36, 59650 Villeneuve-d'Ascq.

Quoiqu'il en soit, la question initiale conduit à se poser un problème différent intéressant : étant donné un ensemble de mots A, comment peut-on décomposer A en puissances d'un ensemble plus « simple » B? Plus précisément, étant donné A, peut-on construire des ensembles B tels que $A = \bigcup_{i \in I} B^i$?

Il est clair que l'on peut trouver des langages pour lesquels aucune vraie décomposition n'existe (langages primitifs). Nous montrons ici que, quelque soit A, il existe B primitif permettant de décomposer A. De plus, si A est rationnel, nous montrons que l'on peut effectivement construire B.

I. LANGAGES PRIMITIFS ET DECOMPOSITIONS D'UN LANGAGE

Nous supposons le lecteur familier avec les notions de base de théorie des langages telles qu'elles sont exposées dans [1] par exemple.

N désigne l'ensemble des entiers naturels. Si I et J sont deux parties de N, on note

$$I^J = \left\{ \sum_{t=1}^{j} i_t \,\middle|\, j \in J, \forall t \in [1, j] : i_t \in I \right\}.$$

Soit X un alphabet. X^* désigne le monoïde libre engendré par X, et on note ε le mot vide.

Pour tout langage $L \subset X^*$, on définit $L^1 = L$ et $L^{n+1} = L^n . L$, et pour toute partie I non vide de N, $L^I = \bigcup_{i \in I} L^i$.

Les implications suivantes sont des trivialités :

$\forall L, M \subset X^*, \forall I \subset N : L \subset M \Rightarrow L^I \subset M^I$

$\forall L \subset X^*, \forall I, J \subset N : I \subset J \Rightarrow L^I \subset L^J$

Enfin on a :

$\forall L \subset X^*, \forall I, J \subset N : (L^I)^J = L^{(I^J)}$.

DÉFINITION 1 : Un langage $L \subset X^*$ est dit *primitif* s'il vérifie : $L = M^I$ avec $M \subset X^*$ et $I \subset N \Rightarrow I = \{1\}$ (et donc $L = M$).

Exemple 1 : Le langage $E = a^* b$ est primitif

Preuve : Supposons que $E = M^I$ avec $I \neq \{1\}$

$$\varepsilon \notin E = M^I \quad \Rightarrow \quad 0 \notin I \quad et \quad \varepsilon \notin M$$

$$\left.\begin{array}{c} b \in E = M^I \\ \varepsilon \notin M \end{array}\right\} \quad \Rightarrow \quad 1 \in I \quad et \quad b \in M$$

$$\left.\begin{array}{r} I\neq\{1\}\wedge 0\notin I\wedge 1\in I \;\Rightarrow\; \exists i\geqq 2: i\in I \\ b\in M \end{array}\right\} \Rightarrow b^i\in M^I=E$$

ce qui n'est pas.

Remarques:

– Si L est entièrement constitué de mots primitifs (c'est le cas de $E=a^*b$), L est primitif, mais la réciproque est fausse: le lecteur se convaincra sans peine que le langage $A=\{aa, bb\}$ est primitif.

– Si L ne contient qu'un mot u, L est primitif si et seulement si u est un mot primitif.

Dans toute la suite, nous excluerons les cas pathologiques où $L=\varnothing$ ou bien $L=\{\varepsilon\}$. De même, le cas des langages non propres, *i. e.* qui contiennent le mot vide est exclu: aucun langage contenant ε n'est primitif.

Remarque: Sur l'alphabet $X=\{a\}$, le seul langage primitif est X.

Exemple 2: Soit $F=\{b\}\cup X^4X^*$ où $X=\{a, b\}$

Si F s'écrit $F=M^I$, il vient sans peine: $b\in M$ et $1\in I$ et donc $2{,}3\notin I$.

Soit $M_1=b+X^4+X^5+\ldots+X^{15}+X^{16}$ et $I=\{4^i\,|\,i\in N\}$ on a alors $F=M_1^I$. Comme $I\neq\{1\}$, F n'est pas primitif. En outre, si $J\subset N$ est tel que $I\subset J\subset K$, avec $K=\{1\}\cup\{i\in N\,|\,i\geqq 4\}$ on a clairement:

$$F=M_1^I.$$

L'exposant vérifiant $L=M^I$ n'est donc pas unique pour L et M donnés. Par contre $i_0=\min(I)$ ne dépend que de L et de M, comme on s'en aperçoit par la considération des mots les plus courts de L. Parmi tous les ensembles d'exposants I vérifiant la relation $L=M^I$ il en existe un maximal pour l'inclusion (qui est leur réunion). En définissant pour tout couple de langages L et M, l'ensemble d'exposants $J_{\max}(L, M)=\{i\in N\,|\,M^i\subset L\}$, $J_{\max}(L, M)$ est cet ensemble maximal. Dans l'exemple, $J_{\max}(F, M_1)=K$. Par contre, il n'existe pas en général d'ensemble I d'exposants, parmi ceux qui vérifient $L=M^I$, qui soit minimal pour l'inclusion (le lecteur vérifiera que pour F et M_1, $I_1=\{1\}\cup\{2i\,|\,i\geqq 2\}$ et $I_2=\{1{,}4\}\cup\{2i+1\,i\geqq 2\}$ sont 2 tels ensembles d'exposants et que $I_1\cap I_2$ n'en est pas un).

Exemple 2, *suite*: Soit $M_k=b+X^4+X^5+\ldots+X^{15}+\ldots+X^{15+k}$. Il est facile de montrer que pour tout $k\geqq 0$, il existe $I_k\subset N$ tel que $F=M_k^{I_k}$. Le même ensemble d'exposants $J=\{K\}$ peut servir pour tous les M_k, mais clairement plus l'ensemble M_k est grand, et moins l'ensemble I_k nécessite d'éléments.

DÉFINITION 2 : On appelle *décomposition* d'un langage $L \subset X^*$ un couple $(M, I) \subset X^* \times N$ tel que $L = M^I$.

Un langage L est primitif si et seulement si il admet pour unique décomposition la décomposition $(L, \{1\})$.

Pour un langage L non primitif, on peut s'intéresser aux décompositions de L de diverses manières :

— on peut chercher à minimiser la première composante sans se soucier de la seconde,

— on peut chercher à minimiser la première composante en imposant que la seconde I ait pour élément minimal $1 = \min(I)$.

— etc, . . .

Dans notre travail, nous nous intéresserons essentiellement aux premières composantes de telles décompositions d'un langage L, que nous appellerons des prémotifs de L. Chaque fois que l'on en aura le choix, lorsque M est prémotif de L, on utilisera l'ensemble d'exposants $I = J_{\max}(L, M)$.

On étudiera le cas particulier où $1 \in \min(I)$ (ou, ce qui revient au même, où $M \subset L$). Dans ce cas particulier on a le lemme technique suivant, dont la preuve est laissée au lecteur :

LEMME TECHNIQUE : *Si $L = L^J$ avec $J = J_{\max}(L, L)$ (et donc $1 \in J$) alors :*

(*a*) $J^J = J$

(*b*) *dans tout produit de $j \in J$ mots de L donnant un mot de L, en remplaçant l'un de ces mots de L par un autre mot de L, on obtient encore un mot de L.*

PROPOSITION 1 : *Un langage $L \subset X^+$ est primitif si et seulement si on a l'implication : $L = M^I$ avec $M \subset X^*$ et $I \subset N \Rightarrow L = M$.*

Preuve : Il n'est à montrer que la condition suffisante, c'est-à-dire que :

$$[L = M^I \text{ avec } M \subset X^* \text{ et } I \subset N \Rightarrow L = M] \tag{1}$$

implique :

$$[L = M^I \text{ avec } M \subset X^* \text{ et } I \subset N \Rightarrow I = \{1\}] \tag{2}$$

Supposons donc (2) non vérifié :

$$L = M^I \quad \text{avec } M \subset X^* \qquad \text{et} \qquad I \subset N \quad \text{avec } I \neq \{1\}$$

et (1) vérifié simultanément :

$$\text{on a donc } L = M \quad \text{d'où} \quad L = L^I \quad \text{avec } I \neq \{1\}.$$

Nous allons montrer qu'alors il existe un $\hat{L} \neq L$ et un $J \subset N$ tels que $L = \hat{L}^J$, contredisant (1). Rappelons que nous supposons que L est un langage propre non vide. Soit $J = J_{\max}(L, L)$. Comme $1 \in J$, on a $J^J = J$. L étant propre, $0 \notin I$. Comme $I \neq \{1\}$, il existe donc un entier $k \in I$ tel que $k \geqq 2$. Posons $\hat{L} = L \setminus L^k$. Clairement $\hat{L}^J \subset L^J = L$. Supposons $\hat{L}^J \neq L$ et soit u un mot de longueur minimale de $L \setminus \hat{L}^J$. Il appartient à $L \setminus \hat{L} \subset L^k$ et s'écrit donc comme le produit de k mots plus courts de $L : u = u_1 u_2 \ldots u_k$, qui sont donc chacun dans $\hat{L}^J$. Leur produit u est donc dans $(\hat{L}^J)^k \subset (\hat{L}^J)^J = \hat{L}^{(J^J)} = \hat{L}^J$; contradiction. □

II. MOTIFS ET BASES D'UN LANGAGE

L'étude faite dans la partie précédente nous amène à poser la définition:

DÉFINITION 3 : Un langage $M \subset X^*$ est un *prémotif* de $L \subset X^*$ si et seulement s'il existe $I \subset N$ tel que $L = M^I$.

Remarquons que l'on pourrait donner une définition équivalente d'un prémotif: M est un prémotif si et seulement si $M^{J_{\max}(L, M)} = L$.

Nous allons étudier, pour L dans X^*, la famille des langages qui sont des prémotifs de L, notée PRÉMOTIFS(L).

Cette famille n'est jamais vide puisque L est lui-même un prémotif de L (si L est primitif, c'est le seul).

Il découle de la définition que si M est un prémotif de L et si P est un prémotif de M, alors P est un prémotif de L:

$$L = M^I \quad \text{et} \quad M = P^J \quad \Rightarrow \quad L = P^K \quad \text{où } K = J^I$$

Exemples:

- Dans l'exemple 2, tous les M_k décrits sont des prémotifs de F
- X et X^2 sont des prémotifs de $(X^2)^+$
- a + a* ba* est un prémotif de $a + a^2 + a^*ba^* + a^*ba^*ba^*$
- pour tout mot f, $X \cup \{f\}$ est un prémotif de X^+.

Si L est un langage non primitif, il existe un langage $M \neq L$ qui est un prémotif de L. Si M n'est pas lui-même primitif, il admet un langage $P \neq M$ comme prémotif, prémotif de M également, et ainsi de suite. Ceci amène à poser la définition:

DÉFINITION 4 : Un langage $M \subset X^*$ est un *motif* de $L \subset X^*$ si c'est un prémotif de L qui est un langage primitif.

Nous allons démontrer dans ce qui suit le théorème suivant:

THÉORÈME : *Tout langage $L \subset X^+$ admet un motif.*

Pour cela, à partir d'un langage L, on définit par une construction (c) un langage noté motif$^{(c)}(L)$, dont on montre ensuite (proposition 2) qu'il est bien un motif de L.

On pourra donc parler de la famille des langages qui sont des motifs de L, notée MOTIFS (L).

Ordonnons arbitrairement les lettres de $X=\{x_1, x_2, \ldots, x_n\}$, par exemple par: $x_i [x_j \Leftrightarrow i<j$. X^* est alors totalement ordonné par l'ordre hiérarchique

$$u[v \quad \Leftrightarrow \quad \begin{array}{c} |u|<|v| \\ \text{ou} \\ u=wx_i u', v=wx_j v', |u'|=|v'| \text{ et } i<j \end{array}$$

On désigne par u_i le ième mot de X^+ dans l'énumération des mots de X^+ dans l'ordre hiérarchique, et par $\min(L)$ le plus petit mot de L dans cet ordre.

Pour tout mot $f \in X^+$, on note $\operatorname{Inf}(f)=\{g \in X^+ \mid g[f\}$

Construction (c)

Soit $u_{k_0}=\min(\{\min(P) \mid P \in \text{PREMOTIFS}(L)\})$.

(C'est le mot le plus petit appartenant à un prémotif de L).

On pose $M_0=\{u_{k_0}\}$ et pour tout entier $n>0$

$$M_n=\begin{cases} M_{n-1} \text{ si il existe } P \in \text{PREMOTIFS}(L) \text{ tel que } P \cap \operatorname{Inf}(u_{k_0+n})=M_{n-1} \\ M_{n-1} \cup \{u_{k_0+n}\} \quad \text{sinon.} \end{cases}$$

Clairement les M_n sont croissants.

Posons pour tout entier $n \geqq 0$ $I_n=J_{\max}(L, M_n)$. Les I_n sont donc décroissants.

Enfin, on pose $M=\bigcup_{n \geqq 0} M_n$ et $I=\bigcap_{n \geqq 0} I_n$.

C'est ce langage M que l'on appelle le langage donné par la construction (c) et que l'on note motif$^{(c)}(L)$.

Avant de prouver que motif$^{(c)}(L) \in$ MOTIFS (L), explicitons la définition de M:

Au départ, on a vu que l'on prenait le plut petit mot qui appartienne à un prémotif de L dans l'ordre de l'énumération de X^+ choisi. On examine ensuite successivement tous les mots de X^+ à partir de celui-ci, à la lueur de l'ensemble de tous les prémotifs de L qui coïncident jusque là avec la partie retenue: ou bien, il existe un tel prémotif qui évite le mot considéré, et on ne retient pas ce mot (ceci a pour effet de réduire l'ensemble des prémotifs de L considérés), ou bien tous ces prémotifs contiennent le mot considéré, qui est alors ajouté à la partie que l'on construit.

Cette construction fait apparaître une suite infinie (possiblement stationnaire) de prémotifs de L:

– P_0 est un prémotif contenant u_{k_0}

– P_n $\begin{cases} \text{est 1 prémotif tel que } P_n \cap \operatorname{Inf}(u_{k_0+n}) = M_{n-1} \text{ s'il existe} \\ \text{est égal à } P_{n-1} \text{ sinon.} \end{cases}$

Remarquons que dans tous les cas $M_n = P_n \cap \operatorname{Inf}(u_{k_0+n})$.

Associons à chaque prémotif P_n l'ensemble d'exposants $J_n = J_{\max}(L, P_n)$ on a alors $J_n \subset I_n$.

LEMME 1 : *Le langage* $M = \text{motif}^{(c)}(L)$ *est un prémotif de L.*

Preuve:

– soit $u \in M^I$, u de rang $k_0 + n = k$

$$u \in M^I \cap \inf(u) \subset [M \cap \operatorname{Inf}(u)]^I = M_n^I \subset M_n^{I_n} \subset L$$

on a donc bien $M^I \subset L$;

– réciproquement, soit $u \in L$, u de rang $k_0 + n = k$, pour tout entier p, $u \in P_p^{J_p}$

$$u \in P_p^{J_p} \cap \operatorname{Inf}(u) \subset [P_p \cap \operatorname{Inf}(u)]^{J_p} = M_n^{J_p} \subset M_n^{I_p}$$

la même décomposition de u sert infiniment souvent: $u = u_1 u_2 \ldots u_l$ avec $\forall i : u_i \in M_n$ et $l \in I_p$. l étant dans une infinité de I_p il se trouve dans leur intersection I: donc $u \in M_n^I \subset M^I$.

LEMME 2 : *Aucun prémotif N de L contenant* u_{k_0} *ne peut être strictement inclus dans M.*

Preuve: En effet sinon soit $u = \min(M \setminus N)$. D'après la construction, lorsqu'on a examiné le mot u, N qui coïncidait jusque là avec la partie retenue (puisqu'il coïncide jusque là avec M) faisait partie des prémotifs de L envisagés, et comme N ne contient pas le mot u, celui-ci n'était pas retenu pour construire M, ce qui contredit le fait que u appartienne à M.

LEMME 3 : *Le langage* $M = \text{motif}^{(c)}(L)$ *est primitif.*

Preuve: Supposons que $M = N^J$. En vertu de la proposition 1, il suffit de montrer que cela entraîne que $M = N$.

L'égalité $M=N^J$ fait que $\min(M)=\min(N^J)$ et donc $\min(N)[\min(M)=u_{k_0}$. D'autre part, N étant un prémotif de M, lui-même un prémotif de L, N est aussi prémotif de L et donc $u_{k_0}[\min(N)$.

On a donc $u_{k_0}=\min(N)$, et $u_{k_0}\in N$.

Ceci entraîne que $1\in J$ et donc $N\subset M$.

D'après le lemme précédent $N=M$.

PROPOSITION 2 : *Le langage $M=Motif^{(c)}(L)$ est un motif de L.*

Ceci découle directement des lemmes 1 et 3.

Remarques: (1) Le langage motif$^{(c)}(L)$ dépend évidemment de la construction (c). *A priori* on peut faire une telle construction pour n'importe quelle énumération des mots de X^+, mais la preuve ci-dessus ne permet de conclure qu'on obtient bien un motif du langage considéré que dans le cas où l'ordre d'énumération est compatible avec l'ordre des longueurs des mots. On peut montrer que quelque soit une telle énumération, l'ensemble des mots les plus courts obtenus reste le même, mais nous ne savons pas si l'on construit le même motif.

(2) La construction proposée est en général non effective. En effet, considérons un langage L inclus dans XX^+ et un nouveau symbole $\sharp$. On pose $A=\sharp^2+X^+\sharp+\sharp X^++L$.

Fait 1: A admet un prémotif P contenant $\sharp$ ssi $L=XX^+$.

En effet si $L=XX^+$, alors $A=(\sharp+X^+)^2$. Réciproquement si A admet un prémotif P contenant $\sharp$, on voit immédiatement que $A=P^I$ avec $I=\{2\}$. Il s'en suit que X^+ est inclus dans P et $(X^+)^2$ inclus dans L. Ainsi $XX^+=(X^+)^2\subseteq L\subseteq XX^+$.

Toujours en utilisant le langage A, on voit que

Fait 2: A est un langage primitif ssi $L\neq XX^+$.

En effet si $L=XX^+$ alors $A=(\sharp+X^+)^2$ et A n'est pas primitif. Au contraire, si $L\neq XX^+$ alors tout prémotif de A contient $\sharp^2$ et $A=P^I$ implique $I=\{1\}$.

On voit donc que, par exemple, le problème de savoir si un langage algébrique est primitif ou non est indécidable. Nous verrons plus loin que si L, est rationnel, on peut décider si L est primitif ou non.

Revenant à la construction (c), le fait 1 montre qu'elle est ineffective en ce sens que pour le langage A, la première question posée pour construire MOTIF$^{(c)}(A)$ est: existe-t-il un prémotif de A contenant $\sharp$? Nous verrons plus loin que, dans le cas des langages rationnels, la construction (c) devient effective en ce sens que l'on peut répondre effectivement à la question posée par le test demandant si l'on ajoute ou non le mot u à M_n.

Appliquée à $R=a+a^2+a^*b+a^*ba+a^*ba^*b$, la construction ($c$) donnera alors motif$^{(c)}$ $(R)=a+a^*b$: ce motif est l'unique motif de R.

– Soit $R=(X^2)^+$ avec $X=\{a, b\}$, on a alors motif$^{(c)}(R)=X$.

Par ailleurs, $\forall i>1 : M_i=X^2+a^{2i}$ est un motif de R.

Tous ces motifs M_i contiennent X^2 qui est un prémotif de R. Par contre motif$^{(c)}(R)=X$ ne contient pas strictement de prémotif de R. Ceci amène à poser la définition suivante :

Définition 5 : Un prémotif M de L (respectivement : un motif) est *minimal* vis-à-vis de L s'il ne contient aucun autre prémotif de L.

La première question qui se pose est celle de l'existence de prémotifs (ou motifs) minimaux. Nous formulons les conjectures suivantes :

Conjecture 1 : Tout langage admet un motif minimal.

Conjecture 1 *bis* : Si L est un langage quelconque, motif$^{(c)}(L)$ est un motif minimal de L.

A l'inverse, un langage peut posséder plusieurs motifs minimaux :

Soit $F=a+a^2+\{a^i ba^j \mid 0 \leqq i, j \leqq 4;\ i+j \leqq 7\}+\{a^i ba^j ba^k \mid i, k \leqq 3;\ j \leqq 6\}$.

F admet les deux motifs minimaux ci-dessous

$$M_1=a+\{a^i ba^j \mid 0 \leqq i, j \leqq 3\} \setminus \{aba^2\}$$
$$M_2=a+\{a^i ba^j \mid 0 \leqq i, j \leqq 3\} \setminus \{a^2 ba\}.$$

Dans les deux cas, on utilise les mêmes exposants : $F=M_1+M_1^2=M_2+M_2^2$.

Une autre question est celle de savoir si des contraintes imposées à L impliquent l'existence de contraintes sur un motif de L. Nous allons montrer dans la partie suivante que si L est rationnel, on peut trouver un motif rationnel de L, bien que ses motifs ne soient pas tous rationnels. Nous ne savons pas s'il en est de même dans le cas algébrique.

Le lecteur peut noter que l'on utilise ici deux ordres sur les langages

– d'une part l'inclusion

– d'autre part l'ordre $\prec$ « être un prémotif de »

Ces deux ordres sont incomparables :

$\{a, b\} \subset \{a, b, aa\}$ et $\{a, b\} \not\prec \{a, b, aa\}$ car $\{a, b\}$ n'est pas un prémotif de $\{a, b, aa\}$. D'autre part, $\{a\} \prec \{a^2\}$ et $\{a\} \not\subset \{a^2\}$.

Cependant si M est un prémotif de $L\,(M \prec L)$ avec $1 \in J_{\max}(L, M)$, on a évidemment $M \subset L$.

DÉFINITION 6 : On appelle *base* de L, un langage M qui est un prémotif de L et qui est inclus dans L (ou, de façon équivalente, tel que $\exists I \subset N : L = M^I$ avec $1 \in I$).

La relation «être une base de» est une relation qui est plus fine que la relation («être un prémotif de») et également plus fine que la relation d'inclusion.

DÉFINITION 7 : Une base de L est *minimale* si elle ne contient aucune autre base de L.

Une base minimale est donc un langage n'admettant aucune autre base que lui-même.

Remarque: Si L est un sous-monoïde de X^* $(L = L^*)$, il admet une unique base minimale qui est le système générateur B minimal du sous-monoïde: $B = (L - \{\varepsilon\}) \setminus (L - \{\varepsilon\})^2$.

L étant une base de lui-même, la construction (c) peut être appliquée en se restreignant aux prémotifs de L qui sont des bases. En vertu du lemme 2, la base obtenue sera minimale. Par ailleurs, si l'ensemble de toutes les bases de L, noté BASES (L), n'est pas strictement inclus dans PREMOTIFS (L) (*i. e.* si ces deux ensembles sont égaux), motif$^{(c)}$ (L) est un motif minimal, ce qui conforte la conjecture 1 *bis*.

Comme pour la construction d'un motif de L, celle proposée ici d'une base est ineffective. Reprenant en effet l'exemple utilisé pour les motifs $A = \sharp^2 + x^+ \sharp + \sharp x^+ + L$, on considère maintenant $B = \sharp + A$. On vérifie facilement que si B_0 est une base de B, $\sharp \in B_0$ et soit $B = B_0$ soit $B_0 = B_0 + B_0^2$. Le second cas n'est possible que si $L = XX^+$. D'où il vient:

B admet une base ne contenant pas $\sharp^2$ ssi $L = XX^+$ et tout se passe comme pour les motifs.

III. LE CAS RATIONNEL

Dans le cas où $L \subset X^+$ est un langage rationnel, nous ne savons pas si motif$^{(c)}$ (L) et base$^{(c)}$ (L) sont aussi des langages rationnels. Par contre, nous allons montrer qu'il est toujours possible de construire pour un langage rationnel un motif rationnel.

Considérons un langage rationnel $R \subset X^+$ et notons S le monoïde syntaxique de R. Pour $x \in X^*$, $[x] \in S$ désigne la classe d'équivalence de x tandis que pour $L \subset X^*$, $[L] \subset S$ désigne $\{[x] \mid x \in L\}$. Enfin, si T est une partie de S,

$\bar{T}=\{x\in X^* \mid [x]\in T\}$. Les lemmes suivants sont évidents :

LEMME 4 : *Pour $L \subset X^*$, $i\in N$ et R langage rationnel, $L^i \subset R$ implique* $\bar{\mu}^i \subset R$ avec $\mu=[L]$.

LEMME 5 : *Pour $L \subset X^*$ et R langage rationnel, $J_{\max}(R, L)=\{i \mid L^i \subset R\}$ est ultimement périodique.*

LEMME 6 : *Si $R=L^I$ avec $L \subset X^*$ et $I \subset N$, alors il existe un langage rationnel K et un ensemble ultimement périodique J tels que $R=K^J$ avec $L \subset K$ et $I \subset J$.*

PROPOSITION 3 : *Pour tout langage rationnel $R \subset X^+$, on peut construire un motif rationnel de R.*

Preuve: Considérons la famille finie de parties de monoïdes syntaxiques S de R, $F=\{\mu \subset S/\bar{\mu}$ est un prémotif de $R\}$. Prenons dans F un élément minimal pour l'inclusion μ_0 vérifiant de plus: $\forall\mu\in F$, $|\min(\bar{\mu}_0)|\leqq|\min(\bar{\mu})|$. Posons $J=\{i\in N/\bar{\mu}_0^i \subset \bar{\mu}_0\}$ et $I=J\setminus\{1\}$. Nous allons montrer que le langage rationnel $M=\bar{\mu}_0\setminus\bar{\mu}_0^I$ est un motif de R. Pour cela, prouvons d'abord que $\bar{\mu}_0=M^J$ ce qui implique que M est un prémotif de R. L'inclusion de M dans $\bar{\mu}_0$ entraîne $M^J \subset \bar{\mu}_0^J=\bar{\mu}_0$. Pour montrer l'inclusion inverse, supposons que $\bar{\mu}_0 \not\subset M^J$ et considérons $u=\min(\bar{\mu}_0\setminus M^J)$. Alors $u\notin M$ et $u\in\bar{\mu}_0^I$, donc $u=u_1 u_2 \ldots u_i$ avec $i\in I \subset J$ et $\forall s\in[1, i]$ $u_s\in\bar{\mu}_0$. D'après le choix de u, on peut déduire que $\forall s\in[1, i]$, $u_s\in M^J$ et $u\in(M^J)^J$. Mais par définition $J^J=J$ et on obtient $u\in M^J$, d'où la contradiction.

Il nous reste à établir que M est un langage primitif. Supposons que $M=N^K$. D'après le lemme 4, $M=\bar{\mu}^K$ avec $\mu=[N]$, donc $\bar{\mu}$ un prémotif de R et $\mu\in F$. Alors, le choix de $\bar{\mu}_0$ implique $\min(M)=\min(\bar{\mu}_0)\leqq\min(N)$ d'où $1\in K$, $N \subset M$ et $\mu=[N] \subset [M] \subset \mu_0$. De nouveau, le choix de μ_0 entraîne $\mu=\mu_0$, d'où $\bar{\mu}_0^K=M \subset \bar{\mu}_0$ et $K \subset J$. Supposons $N \subseteq M$ et posons $u=\min(M\setminus N)$. Alors $u\in N^K\setminus N$ et $u=u_1 u_2 \ldots u_k$ avec $k\in K\setminus\{1\} \subset I$ et $\forall s\in[1, k]$, $u_s\in N$ ce qui implique, d'après le choix de u, $u_s\in M$. On en déduit que $u\in M^I \subset \bar{\mu}_0^I$, d'où la contradiction, puisque $u\in M=\bar{\mu}_0\setminus\bar{\mu}_0^I$. Donc $M=N$ et d'après la proposition 1, M est un langage primitif. □

Remarques: Si nous prenons, par exemple, $R=a+a^2+a^* ba^*+a^* ba^* ba^*$, le monoïde syntaxique $S=\{[\varepsilon], [a], [a^2], [a^3], [b], [b^2], [b^3]\}$, l'ensemble F construit lors de la preuve précédente est égal à $\{[a+b], [a+a^2+b+b^2]\}$. Alors $\mu_0=[a+b]$ et $\bar{\mu}_0=a+a^* ba^*$. L'ensemble $J=\{i/\bar{\mu}_0^i \subset \bar{\mu}_0\}=\{1\}$ et $M=\bar{\mu}_0=a+a^* ba^*$ est le motif rationnel de R construit à l'aide de la preuve précédente. On peut remarquer que $M \not\supseteq \text{motif}^{(c)}(R)$: tout motif de R est inclus dans $M=a+a^*ba^*$; $M\setminus a^2ba$ est un motif de R et donc $a^2ba\notin \text{motif}^{(c)}(R)$.

CONJECTURE 2: Tout langage rationnel admet un motif minimal rationnel.

CONJECTURE 2 *bis*: Si R est un langage rationnel, motif$^{(c)}$ (R) est un motif minimal rationnel.

S'intéressant aux bases au lieu des motifs:

CONJECTURE 3: Tout langage rationnel admet une base minimale rationnelle.

Il est évident qu'un langage rationnel peut admettre des bases non rationnelles. Ainsi $\{a, b\}^+=M^+$ avec $M=\{a, b\}+\{a^n b^n \mid n\geqq 1\}$. De même, on peut trouver des bases non algébriques pour un langage algébrique: soit $L=(\varepsilon+X^*b).\ \{a^i ba^j b \mid i\neq j\}.\ (\varepsilon+X^*b)$ avec $X=\{a, b\}$; $L=L^+$ possède une unique base $B=L\setminus L^2$. Or B n'est pas algébrique car $B\cap ab^2(a^*b)^3=\{ab^2(a^n b)^3 \mid n\geqq 0\}$. Notons que $\min(L)=abb$ est un mot primitif; donc $\text{motif}^{(c)}(L)=\text{base}^{(c)}(L)=B$ n'est pas algébrique. Pourtant, ce même langage L admet un motif M algébrique; en effet $L=M^+$ avec $M=(a^*b+\varepsilon)\ \{a^n ba^n b \mid n\geqq 0\}^*\ \{a^n ba^p b \mid n\neq p\}\ \{a^n ba^n b \mid n\geqq 0\}^*$.

CONJECTURE 4: Tout langage algébrique admet un motif algébrique.

Revenant à la construction (c) dans le cas rationnel, nous laissons au lecteur le soin de vérifier qu'à chaque étape de la construction, la question posée « L admet-il un prémotif P ne contenant pas u et contenant un ensemble fini connu de mots (ceux de $\inf(u)$ déjà ajoutés)? » est décidable dans la mesure où le problème se pose seulement pour des prémotifs saturés par la congruence syntaxique de L. Notons cependant que si L est un rationnel sans motif fini, la construction (c) n'est pas un algorithme: elle ne s'arrête jamais. Ainsi, sait-on seulement répondre dans le cas rationnel à la question suivante: u est-il ou non dans motif$^{(c)}(L)$? Ce qui montre au moins que motif$^{(c)}(L)$ est récursif (*cf.* Conjecture 2 *bis*).

Ces remarques conduisent naturellement à se demander si, dans le cas rationnel, on peut décider de l'existence ou non d'un motif fini. C'est à ce problème qu'est dévolue la section suivante. Elle vise à établir que, étant donné un langage rationnel R, on peut décider s'il admet un motif fini ou non.

Notons que, dès que L est algébrique, ce problème est indécidable. Considérons sur l'alphabet $X=\{a, b\}$ le langage $A_L=a^*b^*\cup X^*b.\ \{a^n b^p \mid n\neq p;\ n>0\}\cup Lb.\ \{a^n b^n \mid n>0\}$ pour $L\subset X^*$. On vérifie que A_L est primitif ssi $L\neq X^*$, auquel cas A_L n'admet pas de motif fini. Au contraire si $L=X^*$, A_L s'écrit X^* qui admet le motif fini X. Ainsi A_L admet un motif fini si et seulement si $L=X^*$ et dès que L est algébrique (A_L l'est aussi), la question devient indécidable.

Revenons maintenant au cas rationnel.

LEMME 7 : *Soit* $R=A^*B\cup C$ *avec* $A\subset X^+$ *et* $B.C\subset X^*$. *Alors, il existe des langages finis* $A_0\subset A$, $B_0\subset B$ *et* $C_0\subset C$ *tels que* $R=A_0^*B_0\cup C_0$ *si et seulement si les* 3 *conditions suivantes sont vérifiées*:

(1) $C\setminus A^*B$ *est fini,*

(2) $B\setminus A^+B$ *est fini,*

(3) *il existe un langage fini* $A_1\subset A$ *tel que* $A_1^*B=A^*B$.

Preuve: Si les 3 conditions sont vérifiées, nous pouvons écrire

$$R=A^*B\cup C_0 \qquad \text{avec} \quad C_0=C\setminus A^*B\subset C.$$

Alors, $A^*B=A_1^*B=A_1^*(B\setminus A_1^+B)$. Comme $B\setminus A_1^+B=B\setminus A^+B$ est fini, on obtient $R=A_0^*B_0\cup C_0$, en posant $B_0=B\setminus A^+B$ et $A_0=A_1$.

Réciproquement, si $R=A_0^*B_0\cup C_0$, nous pouvons en déduire que

$$C\setminus A^*B=R\setminus A^*B=(A_0^*B_0\cup C_0)\setminus A^*B=C_0\setminus A^*B$$

puisque $A_0^*B_0\subset A^*B$. Donc $C\setminus A^*B\subset C_0$ est fini. De même,

$$B\setminus A^+B\subset R\setminus A^+B=(A_0^+B_0\cup B_0\cup C_0)\setminus A^+B=(B_0\cup C_0)\setminus A^+B$$

puisque $A_0^+B_0\subset A^+B$. Ceci implique que la propriété 2 est vérifiée. Enfin

$$\begin{aligned}A^*B=A^*B\cap R&=A^*B\cap(A_0^*B_0\cup C_0)\\&=A_0^*B_0\cup(A^*B\cap C_0)\subset A_0^*B\cup(A^*B\cap C_0).\end{aligned}$$

Comme $A^*B\cap C_0$ est fini, il existe une partie finie $A_2\subset A$ telle que $A^*B\cap C_0\subset A_2^*B$. On obtient alors

$$A^*B\subset A_0^*B\cup A_2^*B\subset(A_0\cup A_2)^*B\subset A^*B$$

et donc $A^*B=A_1^*B$ avec $A_1=A_0\cup A_2$. □

Si A, B et C sont des langages rationnels, les conditions 1 et 2 du lemme précédent sont évidemment décidables. Montrons, maintenant, que la condition 3 est aussi décidable. Pour cela, posons $K=A^*B$ et construisons le langage rationnel :

$$A_B=\{u\in A/\exists\, v\in K \text{ tel que } uv\notin B \text{ et } u\notin A\,(Kv^{-1}\setminus\{\varepsilon\})\}$$

où $Kv^{-1}=\{z\in X^*/zv\in K\}$.

LEMME 8 : *Soient* $A \subset X^+$, $B \subset X^*$ *des langages rationnels et* $K = A^* B$. *Alors,* A_B *est un langage rationnel effectivement constructible vérifiant* $A_B^* B = K$. *De plus, il existe un langage fini* $A_1 \subset A$ *tel que* $A_1^* B = K$ *si et seulement si* A_B *est fini.*

Preuve: Comme B et K sont des langages rationnels, les familles $\{Bv^{-1}/v \in X^*\}$ et $\{Kv^{-1}/v \in X^*\}$ sont des familles finies effectivement constructibles de langages rationnels. Il en est de même pour la famille $\{Bv^{-1} \cup A(Kv^{-1} \setminus \{\varepsilon\}) \mid v \in K\}$ que nous noterons $\{B_1, B_2, \ldots, B_n\}$. Montrons que $A_B = A \setminus \left(\bigcap_{i=1}^{n} B_i\right)$.

Prenons $u \in A_B$. Alors $u \in A$ et il existe $v \in K$ tel que $uv \notin B$ et $u \notin A(Kv^{-1} \setminus \{\varepsilon\})$. Donc il existe $i \in [1, n]$ tel que $u \notin B_i$, ce qui implique $u \notin \left(\bigcap_{i=1}^{n} B_i\right)$ et $u \in A \setminus \left(\bigcap_{i=1}^{n} B_i\right)$.

Réciproquement, si $u \in A \setminus \left(\bigcap_{i=1}^{n} B_i\right)$, il existe $i \in [1, n]$ tel que $u \notin B_i$ et donc il existe $v \in K$ tel que $u \notin Bv^{-1} \cup A(Kv^{-1} \setminus \{\varepsilon\})$, ce qui entraîne $u \in A_B$.

Prouvons, maintenant, l'égalité $K = A_B^* B$. Soit $w \in K \setminus B$. Alors $w \in AK$. Considérons la factorisation $w = uv$ avec $u \in A$, $v \in K$ et u de longueur minimum. Montrons que $u \in A_B$. Raisonnons par l'absurde et supposons que $u \notin A_B$. Alors, par définition de A_B et puisque $v \in K$ et $uv \notin B$, on en déduit que $u \in A(Kv^{-1} \setminus \{\varepsilon\})$. Donc $u = u_1 u_2$ avec $u_1 \in A$, $u_2 \neq \varepsilon$ et $u_2 v \in K$. Nous obtenons une factorisation $w = u_1 u_2 v$ avec $u_1 \in A$, $u_2 v \in K$ et $|u_1| < |u|$, ce qui contredit le choix de u. Donc $u \in A_B$ et $K \setminus B \subset A_B K$. Nous en déduisons $K \subset A_B K \cup B$. Comme $B \subset A^* B = K$ et $A_B \subset A$ donc $A_B K \subset AA^* B \subset A^* B = K$, nous obtenons $K = A_B K + B$. Comme $A_B \subset A \subset X^+$, nous en déduisons que K et $A_B^* B$ sont égaux puisque solution de l'équation $Z = A_B Z + B$.

Prenons, maintenant, $D \subset A$ tel que $D^* B = K$ et montrons que $A_B \subset FG(D)$, l'ensemble des facteurs gauches de D. Prenons $u \in A_B$. Il existe $v \in K$ tel que $uv \notin B$ et $u \notin A(Kv^{-1} \setminus \{\varepsilon\})$. Donc $uv \in A_B K \subset K = D^* B$; ce qui implique $uv \in D^+ B$ puisque $uv \notin B$, d'où $uv \in DK$ et $uv = u'v'$ avec $u' \in D \subset A$ et $v' \in K$. Supposons que $|u'| < |u|$. Alors $u = u'u''$ avec $u'' \neq \varepsilon$ et $u''v = v' \in K$, donc $u'' \in Kv^{-1} \setminus \{\varepsilon\}$ et $u = u'u'' \in A(Kv^{-1} \setminus \{\varepsilon\})$, d'où la contradiction. On en déduit que $|u| \leqq |u'|$, ce qui entraîne $u \in FG(u') \subset FG(D)$ et $A_B \subset FG(D)$. En particulier, si $D = A_1 \subset A$ avec A_1 fini, $FG(D)$ est fini ainsi que A_B. □

Des deux lemmes précédents, nous pouvons déduire un premier résultat de décidabilité.

LEMME 9 : *Soient $M \subset X^*$ un langage rationnel et $I \subset N$ un ensemble ultimement périodique. Alors, on peut décider s'il existe un ensemble fini $F \subset M$ tel que $F^I = M^I$.*

Preuve: Comme I est ultimement périodique, il existe $k \in N$ et deux ensembles finis I_0 et $I_1 \subset N$ tels que $M^I = (M^k)^* M^{I_1} + M^{I_0}$. Montrons l'équivalence entre les deux propriétés :

(i) il existe un ensemble fini $F \subset M$ tel que $F^I = M^I$,

(ii) il existe des ensembles finis $F_0 \subset M^{I_0}, F_1 \subset M^{I_1}$ et $F_2 \subset M^k$ tels que :

$$M^I = F_2^* F_1 + F_0.$$

Si la propriété (i) est vérifiée, $M^I = (F^k)^* F^{I_1} + F^{I_0}$ et la propriété (ii) est vérifiée en prenant $F_0 = F^{I_0}$, $F_1 = F^{I_1}$ et $F_2 = F^k$.

Réciproquement, posons $t = \sup\{|u| / u \in F_1 \cup F_2 \cup F_3\}$ et $F = \{u \in M / |u| \leqq t\}$. Alors $F_0 \subset F^{I_0}$, $F_1 \subset F^{I_1}$ et $F_2 \subset F^k$. Donc

$$M^I = (F_2)^* F_1 + F_0 \subset (F^k)^* F^{I_1} \cup F^{I_0} \subset (M^k)^* M^{I_1} + M^{I_0} = M^I,$$

d'où l'équivalence entre les deux propriétés. D'après les deux lemmes précédents, la propriété (ii) est décidable, d'où le résultat. □

PROPOSITION 4 : *On peut décider si un langage rationnel $R \subset X^*$ possède un motif fini (resp. une base finie).*

Preuve: Il est clair que R possède un motif fini si et seulement si R possède un prémotif fini, c'est-à-dire si et seulement si il existe $T \subset S$, où S est le monoïde syntaxique de R, vérifiant $\bar{T}^I = R$ avec $I = J_{\max}(R, \bar{T}) = \{i / \bar{T}^i \subset R\}$ et il existe un ensemble fini $F \subset \bar{T}$ tel que $F^I = \bar{T}^I$. Comme S est fini, ceci est décidable d'après le lemme précédent. En effet, si F est un motif fini de R, $R = F^I$ avec $I = J_{\max}(R, F)$ et il existe $T = [F]$ vérifiant $R = \bar{T}^I$ avec $F \subset \bar{T}$.

De même R possède une base finie si et seulement si il existe $T \subset [R]$, vérifiant $\bar{T}^I = R$ avec $I = J_{\max}(R, \bar{T})$ et il existe un ensemble fini $F \subset \bar{T}$ tel que $F^I = \bar{T}^I$. □

BIBLIOGRAPHIE

1. J.-M. AUTEBERT, *Langages algébriques*, Masson, Paris, 1987.

2. B. RATOANDROMANANA, *Codes et motifs*, R.A.I.R.O. Informatique Théorique et Applications, Vol. 23, n° 4, p. 425-444, 1989. Rapport interne du L.I.F. de Lille, n° IT 112-87.