

J. GABARRO

**Une application des notions de centre et index
rationnel à certains langages algébriques**

RAIRO. Informatique théorique, tome 16, n° 4 (1982), p. 317-330

http://www.numdam.org/item?id=ITA_1982__16_4_317_0

© AFCET, 1982, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Informatique théorique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

UNE APPLICATION DES NOTIONS DE CENTRE ET INDEX RATIONNEL A CERTAINS LANGAGES ALGÈBRIQUES (*)

by J. GABARRO ⁽¹⁾

Communicated by J. BERSTEL

Résumé. — *Dans cet article, nous prouvons la conjecture due à L. Boasson, B. Courcelle et M. Nivat : la famille des langages algébriques d'index au plus polynomiaux contient strictement la clôture par substitution de la famille des langages linéaires et de la famille des langages à compteur. Cette preuve passe par la notion de centre d'un langage.*

Abstract. — *We prove the following conjecture of L. Boasson, B. Courcelle and M. Nivat: the inclusion of the substitution closure of linear and one counter languages in the full-A.F.L. of polynomial languages is proper. We use the notion of center of a language.*

1. INTRODUCTION

La théorie des familles de langages, introduite par S. Ginsburg et S. Greibach [12], est le cadre désormais classique, dans lequel on étudie les familles de langages formels.

Une des idées directrices, dans ce cadre, est l'étude du comportement de familles de langages vis-à-vis d'opérations de fermeture, ce qui conduit naturellement à la définition de nouvelles familles de langages d'une part, et par abstraction à l'étude de familles de langages fermées pour un ensemble donné d'opérations de fermeture.

Dans la plupart des cas, les ensembles d'opérations considérés ont été ceux relatifs aux notions de cônes rationnels [14] (aussi appelé full-trio [11]) et full-A.F.L. [11]. Un cône rationnel est une famille fermée par transduction rationnelle, ou ce qui revient au même d'après le théorème de Nivat [14], par l'ensemble $C = \{ \text{morphisme, morphisme inverse, intersection avec un langage rationnel} \}$. Si on demande que la famille soit fermée pour les opérations de l'ensemble $A = C \cup \{ \cup, \cdot, * \}$ on obtient un full-A.F.L.

(*) Received July 1981.

(¹) L.I.T.P., Institut de Programmation 55-65, 4, place Jussieu, 75230 Paris Cedex 05.

Dans le cadre des familles de langages définies à l'aide des opérations de fermeture un des problèmes les plus étudiés est le problème de la *principalité*. Un cône rationnel (resp. full-A.F.L.) \mathcal{L} est dit principal si l'on peut trouver un langage L de \mathcal{L} tel que tout élément de \mathcal{L} s'obtient à partir de L à l'aide des opérations de cône rationnel (resp. full-A.F.L.); dans ce cas on dit que L est générateur de \mathcal{L} .

Il est bien connu que les langages algébriques ont le même ensemble, noté *gen* de générateurs, en tant que cône rationnel ou full A.F.L. On peut alors essayer de caractériser *Ng*, la famille des langages algébriques non *générateurs*.

Cette étude a conduit aux cinq familles suivantes [2]. Le cône des *langages linéaires* engendré par le langage symétrique à deux lettres; le full-A.F.L. des *langages quasi rationnels* qui est la fermeture par substitution des langages linéaires; le cône des *langages à compteur* engendré par le langage de Dyck à une lettre; le full-A.F.L. des langages à *compteur itéré* qui est la fermeture par substitution des langages à compteur; enfin la famille *Gre* des *langages de Greibach*, qui est la fermeture par substitution des langages linéaires et des langages à compteur. Cette famille est un A.F.L. non principal fermé par substitution. On sait que [3] :

$$Gre \not\subseteq Ng.$$

Bien d'autres critères de classification des langages formels existent. La notion d'*index rationnel* introduite par L. Boasson, B. Courcelle et M. Nivat [6], et qui est à la base de cet article, est une sorte de mesure de complexité d'un langage, en ce sens qu'elle indique la « distance » d'un langage aux langages rationnels. Contrairement à d'autres notions de ce type [15], l'index rationnel « passe bien » à travers diverses opérations de fermeture. On dispose ainsi d'un outil intéressant en soi, mais qui peut être utilisé pour attaquer avec des moyens nouveaux des vieilles conjectures concernant la théorie classique des langages formels. (Un exemple est donné dans [5].) L'index rationnel permet de définir les familles \overline{Pol} , des langages d'index au plus polynomial et \overline{Exp} , des langages d'index au moins exponentiel.

Par rapport aux familles précédentes on a :

$$Gre \subset \overline{Pol} \quad \text{et} \quad \overline{Gen} \not\subseteq \overline{Exp},$$

L. Boasson, B. Courcelle, M. Nivat ont conjecturé [6] que l'inclusion de *Gre* dans \overline{Pol} est stricte. Le but de cet article est de le démontrer. Il faut noter que la résolution se fait à l'aide du concept de *centre* d'un langage. L'opération de passage au centre permet d'associer à tout langage un autre langage appelé son

centre. Cette opération a été introduite par L. Boasson et M. Nivat dans ses études de mots infinis [7, 8], elle est intéressante parce que d'une part, elle ne peut être réduite à des suites d'opérations de cônes ou A.F.L. et d'autre part les langages algébriques sont fermés par passage au centre.

La preuve de l'inclusion stricte de *Gre* dans *Pol* utilise en plus la remarquable propriété [8] qui dit que *Gre* est fermée par passage au centre. Il faut noter que par contre, elle n'utilise aucun lemme d'itération.

Voici l'organisation de notre travail.

On introduit l'opération « dièze » qui associe à un langage L un autre noté $\#(L)$. Les langages $\#(L)$ ont la propriété de ne pas posséder de paires itérantes strictes (proposition 3). Ceci nous sert à construire un langage qui a des paires itérantes strictes jusqu'à un certain point, et complètement dégénérées après. On relie, finalement, les opérations d'index rationnel, centre et dièze, vis-à-vis d'un langage quelconque (théorème 6). L'application de ces trois opérations aux langages générateurs permet de trouver un langage polynomial qui n'est pas Greibach, ce qui résoud affirmativement la conjecture précédemment énoncée (corollaire 2).

2. PRÉLIMINAIRES

Rappelons quelques définitions et théorèmes qu'on utilisera dans la suite.

Nous utiliserons ici comme définition des transductions rationnelles leur caractérisation due à M. Nivat.

DÉFINITION 1 : Le langage $L' \subset Y^*$ est l'image par *transduction rationnelle* d'un langage $L \subset X^*$, noté $L' = \tau(L)$ ssi :

- (i) il existe un monoïde libre Z^* et un langage rationnel $K \subset Z^*$;
- (ii) il existe deux morphismes de monoïdes $\varphi : Z^* \rightarrow X^*$ et $\psi : Z^* \rightarrow Y^*$ tels que $L' = \tau(L) = \psi(\varphi^{-1}(L) \cap K)$.

DÉFINITION 2 : Une famille \mathcal{L} de langages est un *cône rationnel* ssi elle est fermée par transduction rationnelle.

On notera $\mathcal{C}(\mathcal{L})$ la fermeture par transduction rationnelle de \mathcal{L} (le plus petit cône contenant \mathcal{L}). Si $\mathcal{L} = \{L\}$ on dit que $\mathcal{C}(L)$ est le cône engendré par L .

THÉORÈME 1 [9] : La composition de deux transductions rationnelles est une transduction rationnelle.

Ce théorème permet la définition de :

$$\text{Gen} = \{L/\mathcal{C}(L) = \text{Alg}\} \quad \text{et} \quad \text{Ng} = \text{Alg} \setminus \text{Gen} = \{L/\mathcal{C}(L) \not\subseteq \text{Alg}\}.$$

Nous introduisons maintenant des familles de langages particuliers.

La famille des langages à compteur notée *Oct* est définie par : $Oct = \mathcal{C}(D_1^*)$ où D_1^* est le langage engendré par $S \rightarrow aS\bar{a} + SS + a\bar{a}$.

La famille des langages linéaires notée *Lin* est définie par : $Lin = \mathcal{C}(S_2)$ où S_2 est le langage engendré par $S \rightarrow aS\bar{a} + bS\bar{b} + a\bar{a} + b\bar{b}$.

Pour construire la famille des langages de Greibach, *Gre*, à l'aide de *Lin* et *Oct* on a besoin de l'opération de substitution définie par :

DÉFINITION 3 [2] : Une substitution σ est un morphisme (de monoïdes) de X^* dans $P(Y^*)$.

Une substitution est une \mathcal{L} -substitution ssi quelque soit $x \in X$, $\sigma(x) \in \mathcal{L}$.

Une famille \mathcal{L} est fermée par substitution ssi quelque soit $L \in \mathcal{L}$ et pour n'importe quelle \mathcal{L} -substitution on satisfait $\sigma(L) \in \mathcal{L}$. On appelle $F_s(\mathcal{L})$ la fermeture par substitution de la famille \mathcal{L} .

Avec l'opérateur de substitution on définit $Gre = F_s(Lin \cup Oct)$.

La famille *Gre* satisfait :

PROPOSITION 1 [3] : $Gre \not\subseteq Ng$.

La simple observation des lemmes d'itération des langages algébriques amène à la définition de paire itérante, afin d'étudier les contraintes locales des langages.

DÉFINITION 4 [4] : On dit que $\pi = (\alpha, u, \beta, v, \gamma)$ est une paire itérante de L ssi :

(i) $\forall n \geq 0, \alpha u^n \beta v^n \gamma \in L$;

(ii) $|uv| \geq 1$.

On appellera *exposants de la paire* l'ensemble :

$$\text{Exp } \pi = \{ (n, n') \mid \alpha u^n \beta v^{n'} \gamma \in L \}$$

La paire est *stricte* ssi $\text{Exp } \pi = \{ (n, n) \mid n \in N \}$.

Le théorème de transfert des paires [1] sera utilisé dans la forme faible suivante :

THÉORÈME 2 : Si $L \in Gen$ alors L possède une paire itérante stricte.

DÉFINITION 5 [8] : On associe à tout langage $L \subset X^*$ son centre $C(L) \subset X^*$ par :

$$C(L) = \{ m \mid m \in X^* \text{ et } \text{Card}(mX^* \cap L) = \infty \}.$$

On sait que l'opération de passage au centre n'est pas une transduction rationnelle. Cependant la famille *Gre* a un comportement remarquable vis-à-vis de cette opération :

THÉORÈME 3 [8] : Si $L \in Gre$ alors $C(L) \in Gre$.

Donnons maintenant la définition de l'index rationnel qui sert à mesurer la complexité des langages et permet d'introduire des nouvelles familles.

DÉFINITION 6 [6] : A tout langage $L \subset X^*$ on associe une fonction $\rho_L : N \rightarrow N$: son index rationnel par :

$$\rho_L(n) = \lambda n \cdot \max \{ \delta_{K,L} \mid K \in \text{Rat}_n(X^*) \},$$

où $\delta_{K,L} = \min \{ |w| \mid K \cap L \neq \emptyset \text{ et } w \in K \cap L \}$ et $\text{Rat}_n(X^*)$ est la famille des langages reconnus par un automate fini non déterministe ayant au plus n états. Le comportement de ρ_L vis-à-vis des transductions rationnelles s'exprime par :

THÉORÈME 4 [6] : Si L est un langage et τ une transduction rationnelle, il existe un entier p tel que :

$$\rho_{\tau(L)} \leq \lambda n \cdot (p \cdot n + 1) \rho_L(p \cdot n).$$

Ce théorème permet de définir les familles Pol et \overline{Exp} par :

$$\underline{Pol} = \{ L \mid \rho_L(n) \leq c_1 n^{c_2}, c_1 \in R_+ \text{ et } c_2 \in N_+ \},$$

$$\overline{Exp} = \{ L \mid \rho_L(n) \geq c_1 2^{c_2 n}, c_1, c_2 \in R_+ \}.$$

PROPOSITION 2 [6] : $\underline{Gre} \subset \underline{Pol}$ et $\underline{Gen} \subset \overline{Exp}$.

Dans ce papier nous associons à tout langage $L \subset X^*$ un langage $\#(L)$ défini par :

DÉFINITION 7 [10] : Soit X un alphabet fini et $\#$ un marqueur, on note K le langage $K = \{ h \#^r \mid h \in X^* \text{ et } |h| \geq r \geq 0 \}$. On associe à tout langage $L \subset X^*$ le langage $\#(L)$ défini par $\#(L) = L \#^* \cup K$.

3. PAIRES ITÉRANTES DU LANGAGE $\#(L)$

On veut montrer que quel que soit $L \subset X^*$, le langage $\#(L)$ n'a aucune paire itérante stricte. Pour ce faire on montre (proposition 3) que quel que soit la paire π , il existe un entier k_π tel que :

$$\{ (n, n') \mid n \geq n' \geq k_\pi \} \subset \text{Exp } \pi.$$

La preuve se fait par lemmes. Le lemme 1 classe les trois types possibles de paires, les trois lemmes suivants servent à construire k_π pour chaque type.

Le théorème 5 montre que $\#(L) \in Ng$. La preuve se fait par application de la forme faible du théorème de transfert des paires rappelée ci-dessus.

LEMME 1 (types de paires) : Soit une paire itérante $\pi = (\alpha, u, \beta, v, \gamma)$ de $\#(L)$, alors π doit appartenir à un des trois types suivants :

type 1 : $u \in X^*$ et $v \in X^*$, on factorise la paire en, $\pi = (\alpha, u, \beta, v, \gamma' \#^r)$;

type 2 : $u \in X^*$ et $v \in \#^*$, on factorise la paire en, $\pi = (\alpha, u, \beta \#^r, \#^p, \#^s)$;

type 3 : $u \in \#^*$ et $v \in \#^*$, on factorise la paire en, $\pi = (\alpha \#^r, \#^p, \#^s, \#^q, \#^t)$.

Preuve : Raisonnons par l'absurde. Supposons que $u \in X^+ \#^+$ ou $v \in X^+ \#^+$, dans chacun des cas $\alpha u \beta v \gamma \notin X^* \#^*$ ce qui contredit que π est une paire de $\#(L)$. On s'aperçoit que les éléments itérants u et v doivent appartenir soit à X^* soit à $\#^*$ et il ne peut pas y avoir des chevauchements. Évidemment le cas $u \in \#^+$ et $v \in X^+$ est interdit parce que $\alpha u \beta v \gamma \in X^* \#^*$.

LEMME 2 : Soit une paire itérante $\pi = (\alpha, u, \beta, v, \gamma \#^r)$ où $\alpha u \beta v \gamma \in X^*$. Il existe $k_\pi \in N$ tel que :

$$\{n \geq k_\pi\} \times \{n' \geq k_\pi\} \subset \text{Exp } \pi.$$

Preuve : Si la factorisation de π vérifie $\gamma \leq |\alpha\beta\gamma|$ les mots $\alpha u^n \beta v^{n'} \gamma \#^r \in K$ quelques soient n, n' , on a alors $N \times N = \text{Exp } \pi$. On s'aperçoit qu'il suffit de prendre $k_\pi = 0$.

Par contre, si la paire π vérifie $|\alpha\beta\gamma| > r$, démontrons que les mots $m = \alpha u^n \beta v^{n'} \gamma \#^r$ appartiennent à K lorsque n et n' sont plus grands que :

$$k_\pi = \left\lceil \frac{r - |\alpha\beta\gamma|}{|uv|} \right\rceil.$$

En effet $n = n_1 + k_\pi$ ou $n_1 \geq 0$ et $n' = n_2 + k_\pi$ ou $n_2 \geq 0$, alors :

$$|\alpha u^{n_1 + k_\pi} \beta v^{n_2 + k_\pi} \gamma| \geq |\alpha\beta\gamma| + k_\pi(|u| + |v|)$$

$$= |\alpha\beta\gamma| + \left\lceil \frac{r - |\alpha\beta\gamma|}{|uv|} \right\rceil |uv| \geq |\alpha\beta\gamma| + \frac{r - |\alpha\beta\gamma|}{|uv|} |uv| = r,$$

c'est-à-dire $m \in K$.

On s'aperçoit que k_π est le plus petit entier x que vérifie l'inéquation $|\alpha u^x \beta v^x \gamma| \geq r$.

Rappelons qu'on dit que $F = (\alpha, u, \beta)$ est un *facteur itérant* de L ssi $\alpha u^n \beta \in L$ quelque soit n et $|u| \geq 1$.

LEMME 3 : Soit une paire itérante $\pi = (\alpha, u, \beta \#^r, \#^p, \#^s)$ de $\#(L)$ où $\alpha u \beta \in X^*$, on vérifie:

- (i) Si $|u| > p$, existe k_π tel que $\{(n, n') \mid n \geq n' \geq k_\pi\} \subset \text{Exp } \pi$.
- (ii) Si $|u| < p$, existe k_π tel que $\{n \geq k_\pi\} \times N \subset \text{Exp } \pi$.
- (iii) Si $|u| = p$ alors :

Si $|\alpha\beta| < r + s$ on a $N \times N = \text{Exp } \pi$.

Si $|\alpha\beta| \geq r + s$ on a $\{(n, n') \mid n \geq n' \geq 0\} \subset \text{Exp } \pi$.

Preuve : Étudions chaque cas séparément.

(i) Dans ce premier cas il y a deux possibilités.

Si la factorisation de la paire π vérifie $|\alpha\beta| \geq r + s$, le mot $m = \alpha u^n \beta \#^{r+n'} \#^{p+s} \in K$ lorsque $n \geq n' \geq k_\pi = 0$ car :

$$\begin{aligned} |\alpha u^n \beta| &= |\alpha| + n|u| + |\beta| \\ &= |\alpha\beta| + n|u| \geq r + s + n|u| \geq r + s + n'|u| > r + s + n'p. \end{aligned}$$

Ces inégalités expriment la condition d'appartenance à K .

Par contre, si la paire vérifie $|\alpha\beta| < r + s$ démontrons que : $m = \alpha u^n \beta \#^{r+n'} \#^{p+s} \in K$ lorsque :

$$n \geq n' \geq k_\pi = \left\lceil \frac{r + s - |\alpha\beta|}{|u| - p} \right\rceil.$$

Il suffit de vérifier que $|\alpha u^n \beta| \geq r + n'p + s$ c'est-à-dire que :

$$|\alpha u^n \beta| - r + n'p + s \geq 0.$$

Calculons explicitement cette différence. Comme $n \geq n' \geq k_\pi$ il existe n_1 et n_2 tels que $n = n_1 + n_2 + k_\pi$ et $n' = n_2 + k_\pi$. Alors :

$$\begin{aligned} |\alpha u^n \beta| - (r + n'p + s) &= |\alpha\beta| + n|u| - (r + n'p + s) \\ &= |\alpha\beta| + (n_1 + n_2 + k_\pi)|n| - (r + (n_2 + k_\pi))p - s \\ &= |\alpha\beta| + n_1|n| + n_2(|n| - p) + \left\lceil \frac{r + s - |\alpha\beta|}{|u| - p} \right\rceil (|u| - p) - r - s \\ &\geq |\alpha\beta| + n_1|u| + n_2(|u| - p) + \frac{r + s - |\alpha\beta|}{|u| - p} (|u| - p) - r - s \\ &= n_1|n| + n_2(|u| - p) \geq 0. \end{aligned}$$

(ii) Dans le deuxième cas il y a aussi deux possibilités.

Dans un premier cas la factorisation de la paire vérifie $|\alpha\beta| < r + s$ et les mots $m = \alpha u^n \beta \#^{r+np+s} \in \#(L) \setminus K \subset L \#^*$ car $|\alpha u^n \beta| < r + np + s$, ceci entraîne que $m' = \alpha u^n \beta \in L$ quel que soit n . On a que $F = (\alpha, u, \beta)$ est un facteur itérant de L . Les mots $m'' = \alpha u^n \beta \#^{r+n'p+s} \in L \#^* \forall n, n' \in N$. Il suffit de prendre $k_\pi = 0$ et $\text{Exp } \pi = N \times N$.

Par contre si la paire vérifie $|\alpha\beta| \geq r + s$ démontrons que le mot $\alpha u^n \beta \#^{r+np+s}$ appartient à $\#(L) \setminus K \subset L \#^*$ lorsque :

$$n \geq \left\lfloor \frac{|\alpha\beta| - r - s}{p - |u|} + 1 \right\rfloor.$$

Pour ce faire vérifions que $r + np + s - |\alpha u^n \beta| > 0$ si $n = n_1 + k_\pi$ où $n_1 \in N$.

En effet :

$$\begin{aligned} r + s - |\alpha\beta| + \left(n_1 + \left\lfloor \frac{|\alpha\beta| - r - s}{p - |u|} + 1 \right\rfloor \right) (p - |u|) \\ > r + s - |\alpha\beta| + \left(n_1 + \frac{|\alpha\beta| - r - s}{p - |u|} \right) (p - |u|) = n_1 (p - |u|) \geq 0. \end{aligned}$$

Cela entraîne que $m' = \alpha u^n \beta \in L \forall n \geq k_\pi$ et en conséquence $\{n \geq k_\pi\} \times N \in \text{Exp } \pi$.

On s'aperçoit que $(\alpha u^{k_\pi}, u, \beta)$ est un facteur itérant de L .

Remarquons que k_π est l'entier x plus petit vérifiant $|\alpha u^x \beta| < r + xp + s$.

(iii) Étudions le cas $|u| = p$.

Si $|\alpha\beta| < r + s$, (α, u, β) est un facteur itérant de L parce que parce que $|\alpha\beta| + n|u| < r + s + np$.

Comme conséquence $\alpha u^n \beta \#^{r+np+s} \in \#(L) \setminus K \subset L \#^*$ quel que soit n . Cela entraîne que $N \times N = \text{Exp } \pi$.

Si par contre $|\alpha\beta| \geq r + s$, les mots $\alpha u^n \beta \#^{r+n'p+s}$ appartiennent à K si $n \geq n' \geq 0$.

LEMME 4 : Soit une paire itérante $\pi = (h \#^{k_1}, \#^p, \#^{k_2}, \#^q, \#^{k_3})$ alors $h \in L$ et $N \times N = \text{Exp } \pi$.

Preuve : Le mot $h^{k_1+k_n p+k_2+k_n q+k_3} \in L \#^*$ lorsque :

$$k_\pi = \left\lfloor \frac{|h| - k_1 + k_2 + k_3}{p + q} + 1 \right\rfloor.$$

Cela entraîne que $h \in L$. Alors les mots $h \# k_1 + np + h_2 + n'q + k_3 \in L \#^*$ quels que soient n et n' et $\text{Exp } \pi = N \times N$.

PROPOSITION 3 : Soit π une paire itérante de $\#(L)$; on vérifie :

- (i) existe $k_\pi \in N$ tel que $\{(n, n') \mid n \geq n' \geq k_\pi\} \subset \text{Exp } \pi$;
- (ii) le langage $\#(L)$ n'a aucune paire itérante stricte.

Preuve : (i) En groupant les lemmes 2, 3, 4, on obtient le résultat :

type 1 $\pi = (\alpha, u, \beta, v, \gamma \#^r)$		
$k_\pi = Si \mid \alpha\beta\gamma \mid r > 0$ alors $\left\lfloor \frac{r - \mid \alpha\beta\gamma \mid}{\mid uv \mid} \right\rfloor$ sinon 0		
type 2 $\pi = (\alpha, u, \beta \#^r, \#^p, \#^q)$		
$\mid u \mid > p$	$\mid u \mid = p$	$\mid u \mid < p$
$k_\pi = Si \mid r + s - \mid \alpha\beta \mid > 0$	$k_\pi = 0$	$k_\pi = Si \mid \alpha\beta \mid - (r + s) > 0$
alors $\left\lfloor \frac{r + s - \mid \alpha\beta \mid}{\mid u \mid - p} \right\rfloor$ sinon 0		alors $\left\lfloor \frac{\mid \alpha\beta \mid - r - s}{p - \mid u \mid} \right\rfloor$ sinon 0
type 3 $\pi = (h \#^{k_1}, \#^p, \#^{k_2}, \#^q, \#^{k_3})^3$		
$k_\pi = 0$		

(ii) il n'y a pas de paires itérantes strictes parce que :

$$\{(n, n) \mid n \geq k_\pi\} \not\subseteq \{(n, n') \mid n \geq n' \geq k_\pi\}.$$

COMMENTAIRE : L'opération $\#$ peut avoir des effets « curieux » sur les paires itérantes des langages algébriques.

Considérons la paire $\pi = (a, a, b, b, \epsilon)$ du langage $S = \{a^n b^n \mid n \geq 1\}$.

Si on regarde maintenant la paire $\pi' = (a, a, b, b, \#^{2k+2})$ du langage $\#(S_2)$ on s'aperçoit que $\text{Exp } \pi' = \{(n, n) \mid n < k\} \cup \{(n, n') \mid n + n' \geq k\}$, c'est-à-dire la paire itérante π' est stricte jusqu'à k , et à partir de k , dégénérée. L'existence de langages algébriques admettant des paires itérantes avec de tels exposants n'était pas évidente.

THÉORÈME 5 : Quel que soit $L \subset X^*$, on a $\#(L) \in Ng$.

Preuve : Si $\#(L) \notin Ng$ alors $\#(L) \in Gen(Alg)$ et $\#(L)$ devrait admettre une paire itérante stricte d'après le théorème 2.

4. CENTRE DU LANGAGE $\#(L)$

Le but de cette partie est de calculer explicitement le centre du langage $\#(L)$. On s'aperçoit alors que L et le centre de $\#(L)$ engendrent le même cône rationnel.

PROPOSITION 4 : *Quel que soit $L \subset X^*$ on vérifie $C(\#(L)) = X^* \cup L \#^*$.*

Preuve : Rappelons que $FG(m)$ est l'ensemble des facteurs gauches du mot m . Sachant que l'opération du passage au centre est distributive par rapport à l'union on a :

$$C(\#(L)) = C(L \#^+ \cup K) = C(L \#^+) \cup C(K).$$

(i) Démontrons que $C(L \#^+) = FG(L \#^+) = FG(L) \cup L \#^+$.

On verra que n'importe quel facteur gauche de $m \#^r$, $m \in L$ et $r > 0$ peut être complété d'une infinité de façons.

Lorsque $w \in FG(m \#^r)$ soit $w \in FG(m)$, soit $w \in m \#^+$. Si $w \in FG(m)$ il existe w' tel que $ww' = m$ et quel que soit r' le mot $ww' \#^{r'} \in m \#^+$, cela entraîne que $w \in C(L \#^+)$. Sinon $w \in m \#^+$, w peut s'écrire $w = m \#^{r'}$, $r' \leq r$ et $\forall r'$, $m \#^{r'+r''} \in m \#^+$ cela entraîne que $w \in C(L \#^+)$.

(ii) Démontrons que $C(K) = X^*$. On va chercher parmi les facteurs gauches de K ceux qu'on peut compléter d'une infinité de façons. Lorsque $w \in FG(K)$, soit $w \in X^*$, soit $w \in X^* \#^+$. Si $w \in X^*$ quel que soit $w' \in X^*$ le mot $ww' \#^{|ww'|} \in K$, cela entraîne que $w \in C(K)$.

Au contraire si $w \in X^* \#^+$, $w \in C(K)$, en effet $w = w' \#^{n_1}$ où $w' \in X^*$ alors $ww'' \in K$ ssi $w'' = \#^{n_2}$ et $n_2 \leq |w'| - n_1$, on ne peut pas compléter w d'une infinité de façons.

REMARQUE : Ainsi qu'annoncé $\mathcal{C}(L) = \mathcal{C}(C(\#(L)))$ car $\#(L) \cap X^* \# = L \#$.

5. INDEX RATIONNELS DE $\#(L)$ ET $C(\#(L))$.

On travaille sur les index rationnels des langages $\#(L)$ et $C(\#(L))$. On montre que l'index de $\#(L)$ est au plus linéaire quel que soit celui de L , et que celui de $C(\#(L))$ est à peu près le même que l'index de L .

PROPOSITION 5 : *Soit \mathcal{L} une famille de langages, on vérifie que :*

$$\forall L \in \mathcal{L}, \quad \rho_L(n-1) + 1 \leq \rho_{C(\#(L))}(n) \leq \rho_L(n) + n.$$

Preuve : Comme $C(\#(L)) = L \#^+ \cup X^*$ l'index rationnel de $C(\#(L))$ vérifie :

$$\rho_{C(\#(L))}(n) \leq \max \{ \rho_{X^*}(n), \rho_{L \#^+}(n) \} = \rho_{L \#^+}(n) \leq \rho_L(n) + n.$$

Pour démontrer que $\rho C(\#(L))(n) \geq \rho_L(n-1) + 1$ considérons le langage $R \in \text{Rat}_{n-1}(X^*)$, où R est reconnu par l'automate $\mathcal{A} = \langle Q, Q_I, Q_F, P \rangle$ tel que $\rho_L(n-1) = \delta_r$, on construit, à partir de R le langage $\overline{R}, R \in \text{Rat}_n((X \cup \#)^*)$, reconnu par $\mathcal{A} = \langle Q \cup \{\bar{q}\}, Q_I, \bar{q}, P \cup \{(q, \#, \bar{q}) \mid q \in Q_F\} \rangle$. L'index rationnel vérifie :

$$\rho_L(n-1) + 1 = \delta_{\overline{R}, L^*} = \delta_{R, C(\#(L))} \leq \rho_{C(\#(L))}(n).$$

COROLLAIRE 1 : Si $\mathcal{L} \subset \underline{Pol}$ alors $C(\#(\mathcal{L})) \subset \underline{Pol}$.

(ii) Si $\mathcal{L} \subset \overline{Exp}$ alors $C(\#(\mathcal{L})) \subset \overline{Exp}$.

Preuve : Rappelons que $C(\#(\mathcal{L})) = \bigcup_{L \in \mathcal{L}} C(\#(L))$.

(i) Si $L \in \underline{Pol}$ son index vérifie $\rho_L(n) \leq c_1 n^{c_2}$ et l'index de $C(\#(L))$ d'après la proposition 5 satisfait $\rho_{C(\#(L))}(n) \leq c_1 n^{c_2} + n \leq (c_1 + 1) n^{c_2}$.

(ii) Si $L \in \overline{Exp}$ l'index vérifie $\rho_L(n) \geq c_1 2^{c_2 n}$ et l'index de $C(\#(L))$ d'après la proposition 5 satisfait $\rho_{C(\#(L))}(n) \geq c_1 2^{c_2(n-1)} + 1 \geq c_1 2^{-c_2} 2^{c_2 n}$.

PROPOSITION 6 : Soit \mathcal{L} une famille quelconque de langages et $X \cup \#)^*$ on a :

(i) $\forall L \in \mathcal{L}$ si $R \cap \#(L) \neq \emptyset$ existe $w \in R \cap \#(L)$ tel que $|w| \leq 3n$.

(ii) $\forall L \in \mathcal{L}$, $\rho_{\#(L)}(n) \leq 3n$ et $\#(\mathcal{L}) \subset \underline{Pol}$.

Preuve : (i) Raisonnons par l'absurde. Supposons qu'il existe $L \in \mathcal{L}$ tel que : quel que soit $m \in R \cap \#(L)$, m satisfait $|m| > 3n$. Dans ce cas $\delta_{R, \#(L)} = \min\{|m| \mid m \in R \cap \#(L)\} > 3n$.

Soit $w = h \#^r \in R \cap \#(L)$ un mot tel que $|w| = \delta_{R, \#(L)} > 3n$. On va arriver à une contradiction en trouvant un mot $w' \in R \cap \#(L)$, à partir de w tel que $|w'| \leq 3n$.

Construction de w' : Soit $w = h \#^r$, si $r > n$, le lemme d'Ogden appliqué à $R \in \text{Rat}_n(X^*)$ dit qu'il existe $r' \leq n$ tel que $h \#^{r'} \in R$. Étudions $|h|$; il y a trois possibilités :

(a) Si $|h|$ satisfait $0 < |h| < r'$, cela entraîne $|h| < r$, $h \in L$, alors $w' = h \#^{r'} \in L \#^{r'} \cap R \subset \#(L) \cap R$ et $|w'| < 2r' < 3n$.

(b) Si $|h|$ satisfait $r' \leq |h| \leq 2n$ alors $|h| > r'$, $h \#^{r'} \in K$,

$$w' = h \#^{r'} \in K \cap R \subset \#(L) \cap R \quad \text{et} \quad |w'| < 2n + r' \leq 3n.$$

(c) Si $|h|$ satisfait $|h| > 2n$, on factorise $h = h_1 h_2$ avec $|h_1| = r'$ et $|h_2| = |h| r' > n$. Le lemme d'Ogden permet de trouver $h'_2, h'_2| \leq n$ alors $|h_1 h'_2| \geq r'$ et $h_1 h'_2 \#'' \in K, w' = h_1 h'_2 \#'' \in K \cap R \subset \#(L) \cap R$ et $|w'| \leq 3n$.

(ii) Prenons $R \in \text{Rat}_n((X \cup \#)^*)$ tel que :

$$\rho_{\#(L)}(n) = \min \{ |w| \mid w \in \#(L) \cap R \} = \delta_{R, \#(L)}.$$

On sait par (i) que $\delta_{R, \#(L)} \leq 3n$, donc $\rho_{\#(L)}(n) \leq 3n$. Si $\#(\mathcal{L}) = \bigcup_{L \in \mathcal{L}} \#(L)$ on a $\#(\mathcal{L}) \subset \underline{\text{Pol}}$.

REMARQUE : Si on applique le théorème $\overline{\text{Gen}} \subset \overline{\text{Exp}}$ [6] on obtient, parce que $\rho_{\#(L)}(n) \leq 3n$, que $\#(L) \in \overline{\text{Ng}}$. Cette démonstration a l'inconvénient de ne rien dire sur la structure des paires itérantes de $\#(L)$ et ne permet pas d'obtenir la proposition 3. D'autre part l'existence de paires itérantes analogues à $\pi = (a, a, b, b, \#^{2k+2})$ en $\#(S_2)$ resterait inconnue. Le procédé de démonstration de la première partie donne donc des renseignements plus précis sur la structure des langages $\#(L)$.

6. APPLICATION

Nous résolvons, dans le corollaire 2, la conjecture $\overline{\text{Gre}} \not\subset \overline{\text{Pol}} \cap \overline{\text{Alg}}$ posée par L. Boasson, B. Courcelle et M. Nivat [6].

THÉORÈME 6 : Si $\mathcal{L} \subset \overline{\text{Exp}}$ alors $\#(\mathcal{L}) \subset \overline{\text{Pol}} \setminus \overline{\text{Gre}}$.

Preuve : Démontrons que $\#(\mathcal{L}) \subset \overline{\text{Pol}}$ et que quel que soit $L \in \mathcal{L}$ on a $\#(L) \notin \overline{\text{Gre}}$. La proposition 6 montre que $\#(\mathcal{L}) \subset \overline{\text{Pol}}$. La preuve de $\#(\mathcal{L}) \not\subset \overline{\text{Gre}}$ se fait par l'absurde et passe par les notions d'index rationnel et de centre d'un langage. Si on suppose que $\#(L) \in \overline{\text{Gre}}$, la proposition 2 montre que $\#(L) \in \overline{\text{Pol}}$ et le corollaire 1 affirme que $C(\#(L)) \in \overline{\text{Pol}}$. D'autre part par hypothèse $L \in \overline{\text{Exp}}$, alors là le corollaire 1 permet de dire que $C(\#(L)) \in \overline{\text{Exp}}$. L'index $\rho_{C(\#(L))}(n)$ devrait satisfaire une inéquation :

$$C_1 2^{C_2 n} \leq \rho_{C(\#(L))}(n) \leq C'_1 n^{C_2},$$

ce qui est impossible.

Reste maintenant la question de construire effectivement des langages algébriques appartenant à $\underline{Pol} \setminus Gre$. La proposition 2 dit que $Gen \subset Exp$. Parmi les générateurs d'Alg les plus connus nous avons :

$$\forall n \geq 2. D_n^*, \text{ les langages engendrés par } S \rightarrow \sum_{i=1}^n a_i S \bar{a}_i + SS + \sum_{i=1}^n a_i \bar{a}_i,$$

$$1 E_n, \text{ les langages engendrés par } S \rightarrow \sum_{i=1}^n a_i S b S c_i + d.$$

On a alors le corollaire :

COROLLAIRE 2 : $\#(Gen) \subset \underline{Pol} \cap Alg \setminus Gre$.

En particulier :

(i) $\forall n \geq 2, \#(D_n^*) \in \underline{Pol} \cap Alg \setminus Gre$.

(ii) $\forall n \geq 1, \#(E_n) \in \underline{Pol} \cap Alg \setminus Gre$.

REMARQUE : On trouve ici des langages non générateurs et dont le centre est générateur des langages algébriques. En outre ceux-ci sont aussi des langages non générateurs qui ne sont pas dans la famille Gre.

BIBLIOGRAPHIE

1. J. BEAUQUIER, *Générateurs algébriques et systèmes de paires itérantes*, Théor. Comp. Sc., vol. 8, 1979, p. 293-323.
2. J. BERSTEL, *Transductions and Context-Free Languages*, Teubner Verlag, 1979.
3. L. BOASSON, *The Inclusion of the Substitution Closure of Linear and One Counter Languages in the Largest Sub-A.F.L. of the Family of C.F.L's is Proper*, Inform. Process. Let., vol. 2, 1973, p. 135-140.
4. L. BOASSON, *Langages algébriques, paires itérantes et transductions rationnelles*, Théor. Comp. Sc., vol. 2, 1976, p. 209-223.
5. L. BOASSON, *Un langage particulier*, R.A.I.R.O. informatique théorique, vol. 13, n° 3, 1979, p. 203-215.
6. L. BOASSON, B. COURCELLE et M. NIVAT, *The Rational Index, a Complexity Measure for Languages*, S.I.A.M. J. Compt., vol. 10, 1981, p. 284-296.
7. L. BOASSON et M. NIVAT, *Adherences of Languages*, J. Comput. and Syst. Sc., vol. 20, 1980, p. 285-309.
8. L. BOASSON et M. NIVAT, *Centers of Languages*, in Theoretical Computer Science, 5th Gi-Conference, Lec. Not. Comp. Sc., vol. 104, 1981, p. 245-251.
9. C. ELGOT et G. MEZEL, *On Relations Defined by Generalized Finite Automata*, I.B.M. J. of Res. and Dev., vol. 9, 1965, p. 88-101.
10. J. GABARRO, *Index rationnel, centre et langages algébriques*, Thèse, 3° Cycle, Paris-VI, 1981.

11. S. GINSBURG, *Algebraic and Automata-Theoretic Properties of Formal Languages*. North-Holland, 1975.
12. S. GINSBURG et S. GREIBACH, *Abstract Families of Languages*, in *Studies in Abstract Families of Languages*, Mem. Amer. Math. Soc., vol. 87, 1969, p. 1-32.
13. S. GREIBACH, *Chains of full-A.F.L.'s*, Math. Syst. Theory, vol. 4, 1970, p. 231-242.
14. M. NIVAT, *Transductions des langages de Chomsky*, Ann. Inst. Fourier, vol. 18, 1968, p. 339-456.
15. J. PAREDAENS et R. VYNCKE, *A Class of Measures on Formal Languages*, Acta Informatica, vol. 9, 1977, p. 73-86.