L. BOASSON

Non-générateurs algébriques et substitution

RAIRO – Informatique théorique, tome 19, n° 2 (1985), p. 125-136.
<http://www.numdam.org/item?id=ITA_1985__19_2_125_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 19, n° 2, 1985, p. 125 à 136)

NON-GÉNÉRATEURS ALGÉBRIQUES ET SUBSTITUTION (*)

par L. BOASSON ([1])

Résumé. — *Le but de cet article est d'établir le résultat suivant : si la clôture par substitution d'un cône rationnel $\mathscr{L}$ est égale à la famille Nge des langages algébriques non-générateurs, alors $\mathscr{L}$ était déjà égal à cette famille.*

Abstract. — *In this paper, we prove the following: given a rational cone $\mathscr{L}$ whose substitution closure is the largest sub-full AFL of the family of context-free languages, then $\mathscr{L}$ was already this subfamily.*

La théorie des langages formels, et donc aussi celle des langages algébriques se traite depuis une quinzaine d'années dans le cadre des cônes rationnels et des « full AFL's ». Nous renvoyons le lecteur à l'ouvrage de J. Berstel [2] pour la présentation de ces notions de même que pour les principales notations et les résultats classiques. Le présent article est une contribution négative au problème maintenant ancien concernant les non-générateurs de la famille des langages algébriques, c'est-à-dire des langages algébriques engendrant un cône rationnel strictement inclus dans Alg. On sait que Alg, famille de tous les langages algébriques, est un cône rationnel principal de générateur le langage E engendré par la grammaire $\langle S \rightarrow aSbSc+d \rangle$. Il en résulte que Alg contient un plus grand sous-cône rationnel Nge, celui formé des langages algébriques non générateurs. Le problème ouvert est le suivant :

Nge est-il un cône principal ou non?

La conjecture généralement admise est que non et, pour prouver ceci, de nombreuses tentatives ont été du type suivant : construire un sous-cône $\mathscr{L}$ de Alg, en construire la clôture par substitution, notée $\mathscr{T}^{\sigma}(\mathscr{L})$, et tenter de prouver que $\mathscr{T}^{\sigma}(\mathscr{L}) = \text{Nge}$. C'est ainsi que l'on a espéré un moment que Nge était égal à Gre, la famille des langages de Greibach, obtenue comme clôture

(*) Reçu en décembre 1983, révisé en mai 1984.
([1]) UER de Mathématiques et Informatique, Université Paris-VII, 2, place Jussieu, 75005 Paris.

0399-0540/85/02 125 12/$ 3,20/

par substitution des familles des langages linéaires et à un compteur. On sait que Nge contient strictement Gre [3]. Le but de cet article est de montrer que toutes ces tentatives ont échoué de par leur nature même, en établissant un résultat général :

Quel que soit le cône rationnel $\mathscr{L}$, si sa clôture par substitution $\mathscr{T}^{\sigma}(\mathscr{L})=\mathrm{Nge}$, alors $\mathscr{L}=\mathrm{Nge}$.

Cet article est divisé en quatre parties. La première est consacrée aux préliminaires, la seconde présente trois faits fondamentaux utilisés dans la troisième partie où nous prouvons le résultat annoncé ci-dessus. La quatrième et dernière partie présente quelques compléments et questions ouvertes.

I. PRÉLIMINAIRES

Sur l'alphabet $Z=\{a, \bar{a}\}$, on considère le langage de Dyck D_1^* qui est la classe du mot vide 1 dans la congruence engendrée par la relation $a\bar{a}\equiv 1$. On désigne par D_1 les mots premiers de D_1^*, i. e. ceux qui n'ont aucun facteur gauche propre dans D_1^*.

Si h est un mot de D_1 qui s'écrit $h=h_1 a h_2$, la hauteur de cette occurrence de la lettre a est définie par haut $(h_1, a, h_2)=k$ ssi $h_1 a\equiv a^k$. Si l'on considère une occurrence de la lettre $\bar{a}$, sa hauteur est celle de la lettre a associée. La hauteur du mot h est alors le maximum des hauteurs de ses diverses lettres.

Étant donné un alphabet X disjoint de Z, on désigne par π la projection de $(X\cup Z)^*$ sur Z^*. La notion de hauteur d'une lettre dans un mot est étendue aux mots de $\pi^{-1}(D_1)$ de la façon suivante : si g est un mot de $\pi^{-1}(D_1)$, une occurrence de lettre de Z a pour hauteur celle qu'elle a dans $\pi(g)$, soit la hauteur de cette occurrence dans le mot de Dyck sous-jacent; une occurrence de la lettre x de X a une hauteur définie par : ou bien x n'a aucune lettre de Z à sa gauche et sa hauteur est nulle, ou bien x a comme plus proche voisin à gauche dans Z une lettre a de hauteur i et x a pour hauteur i; ou bien x a comme plus proche voisin à gauche dans Z une lettre $\bar{a}$ de hauteur i et x a pour hauteur $(i-1)$. A nouveau, la hauteur de g est le maximum de la hauteur de ses lettres. Il est immédiat de vérifier que g et $\pi(g)$ ont la même hauteur.

Étant donné un mot g dans $\pi^{-1}(D_1)$, on désigne par num (g) le mot obtenu en indiçant les diverses lettres de g par leurs hauteurs.

Exemple : $X=\{x, y\}$.

$$g=axaxay\bar{a}xaxx\bar{a}\,\bar{a}xaxaxx\bar{a}xay\bar{a}\,\bar{a}\,\bar{a},$$

satisfait :

$$\pi(g)=aaa\bar{a}a\bar{a}\,\bar{a}aa\bar{a}a\bar{a}\,\bar{a}\,\bar{a}\in D_1, \qquad \text{haut}(g)=3.$$

$$\text{num}(g)=a_1\,x_1\,a_2\,x_2\,a_3\,y_3\,\bar{a}_3\,x_2\,a_3\,x_3\,x_3\,\bar{a}_3\,\bar{a}_2\,x_1$$
$$\times\, a_2\,x_2\,a_3\,x_3\,x_3\,\bar{a}_3\,x_2\,a_3\,y_3\,\bar{a}_3\,\bar{a}_2\,\bar{a}_1.$$

Il est clair que la fonction num n'est pas une transduction rationnelle. Cependant, si l'on borne à l'avance les hauteurs maximales (et donc les indices) que l'on peut obtenir, on peut réaliser la numérotation d'un mot à l'aide d'une machine séquentielle. Plus précisément, on désigne par H_k les mots de D_1 de hauteur au plus k. On sait que H_k est un langage rationnel donné par : $H_1=a\bar{a}$ et $H_{k+1}=a\,H_k^*\,\bar{a}$. La restriction de la fonction num aux mots premiers de hauteur au plus k est donnée par :

– si $g\notin\pi^{-1}(H_k)$, alors $\text{num}_k(g)=\varnothing$;

– si $g\in\pi^{-1}(H_k)$, alors $\text{num}_k(g)=\text{num}(g)$.

Nous laissons au lecteur le soin de vérifier que num_k est bien une transduction rationnelle réalisable par une machine séquentielle.

Étant donnés deux alphabets disjoints X et Y, une substitution syntaxique partielle de X^* dans $(X\cup Y)^*$ est définie par :

– une partition de X en $X'\cup X''$;

– un langage B sur Y.

La substitution syntaxique partielle $\sigma=\langle X', X'', B\rangle$ est alors donnée par :

$$\sigma(x)=x \quad \text{si } x\in X' \qquad \text{et} \qquad \sigma(x)=x\,B \quad \text{si } x\in X''.$$

Dans le cas particulier où $X'=\varnothing$ (et donc $X''=X$), on retrouve la substitution syntaxique classique, notée $\uparrow B$. Nous utiliserons souvent la substitution syntaxique partielle $\sigma=\langle Z, X, B\rangle$ où $Z=\{a, \bar{a}\}$ est disjoint de X. Elle sera notée $\Uparrow B$. $\Uparrow B$ consiste donc à substituer aux lettres $x\notin Z$ le langage $x\,B$ en laissant inchangées les deux lettres de Z.

Ainsi, par exemple, si $L=\{axx\bar{a}\}$ et $B=\{ayy\bar{a}\}$, $L\Uparrow B$ dénote-t-il $\{axayy\bar{a}xayy\bar{a}\,\bar{a}\}$. La notation $\Uparrow B$ sera aussi utilisée pour les diverses copies de $(X\cup Z)$.

Il est facile de prouver :

LEMME 1 : *Étant donnés deux langages rationnellement équivalents B et B', et deux substitutions syntaxiques partielles :*

$$\sigma=\langle X', X'', B\rangle \qquad \textit{et} \qquad \sigma'=\langle X', X'', B'\rangle,$$

quel que soit le langage A sur $X=X'\cup X''$, $\sigma(A)$ et $\sigma'(A)$ sont rationnellement équivalents.

Étant donné un alphabet X, on désigne par X_i la copie (disjointe) de X obtenue en indiçant par i les lettres de X. La réunion des alphabets $X_1, X_2, \ldots, X_k$ est notée $X^{(k)}$. Si L est un langage sur X, L_i désigne sa copie sur X_i. On définit alors sur $X^{(k)}$ le langage :

$$L^{(k)}=L_1 \uparrow L_2 \uparrow L_3 \ldots \uparrow L_k.$$

Rappelons que $A \uparrow B \uparrow C$ dénote $A \uparrow (B \uparrow C)$ qui est différent de $(A \uparrow B) \uparrow C$. On sait que, soit L et $L^{(2)}$ sont équivalents, auquel cas $\mathcal{T}(L)=\mathcal{T}^{\sigma}(L)$, soit $L^{(2)}$ domine strictement L, auquel cas $\mathcal{T}^{\sigma}(L)$ est un cône rationnel non principal de base génératrice $\{L^{(i)} \mid i \in N\}$. Ces résultats sont aisément prouvés à l'aide du lemme syntaxique de S. Greibach [4] que nous utiliserons plus loin dans une version améliorée par M. Latteux [5] :

LEMME SYNTAXIQUE : *Étant donnés deux cônes rationnels $\mathcal{L}$ et $\mathcal{M}$ et deux langages A et B tels que $A \uparrow B \in \mathcal{L} \square \mathcal{M}$, on a soit $A \in \mathcal{L}$, soit $(Bc)^* \in \mathcal{M}$.*

Le résultat est vrai aussi si $\mathcal{L}$ est la famille des langages finis.

($\mathcal{L} \square \mathcal{M}$ désigne le cône obtenu en substituant des langages de $\mathcal{M}$ dans ceux de $\mathcal{L}$.)

A l'alphabet X de L, nous ajoutons l'alphabet de deux parenthèses $Z=\{a, \bar{a}\}$ pour obtenir $Y=X \cup Z$. On définit alors la version complètement parenthésée de $L^{(k)}$ de la façon suivante :

$$L'=a L \bar{a} \cup \{\varepsilon\};$$

$$C\,\mathrm{Par}\,(L)=L'_1 \qquad (\text{soit } a_1 L_1 \bar{a}_1 \cup \{\varepsilon\});$$

$$C\,\mathrm{Par}\,(L^{(k)})=L'_1 \Uparrow L'_2 \ldots \Uparrow L'_k.$$

Exemple : Si $L=\{xx\}$, $L'=\{axx\bar{a}\} \cup \{\varepsilon\}$ et C Par $(L^{(2)})=L'_1 \Uparrow L'_2$ soit :

$$\{\varepsilon,\ a_1 x_1 x_1 \bar{a}_1,\ a_1 x_1 x_1 a_2 x_2 x_2 \bar{a}_2 \bar{a}_1,\ a_1 x_1 a_2 x_2 x_2 \bar{a}_2 x_1 \bar{a}_1,\ a_1 x_1 a_2 x_2 x_2 \bar{a}_2 x_1 a_2 x_2 x_2 \bar{a}_2 \bar{a}_1\}.$$

On peut tout d'abord faire quelques simples remarques :

Rem. 1 : Les mots de $C\,\mathrm{Par}\,(L^{(k)})$ sont bien parenthésés sur $Z^{(k)}$.

Rem. 2 : Les lettres des mots de $C\,\mathrm{Par}\,(L^{(k)})$ portent en indices leurs hauteurs.

Rem. 3 : Les mots de $C\,\mathrm{Par}\,(L^{(k)})$ sont tous de hauteur au plus k.

Rem. 4 : Un mot de $C\,\mathrm{Par}\,(L^{(k)})$ de hauteur $i<k$ est aussi dans $C\,\mathrm{Par}\,(L^{(i)})$.

Enfin, on établit le :

LEMME 2 : *Les langages* $L^{(k)}$ *et* $C\,\mathrm{Par}(L^{(k)})$ *sont rationnellement équivalents.*

Preuve : Pour $k=1$, le résultat est évident puisque :

$$C\,\mathrm{Par}\,(L)=a_1\,L_1\,\bar{a}_1 \cup \{\varepsilon\} \equiv L_1 = L^{(1)}.$$

Supposons le résultat établi pour k; on sait que $L^{(k)}$ et $L_2 \uparrow L_3 \ldots \uparrow L_{k+1}$ sont deux copies du même langage et sont donc équivalents. De même :

$$C\,\mathrm{Par}(L^{(k)}) \qquad \text{et} \qquad C\,\mathrm{Par}\,(L_2) \Uparrow C\,\mathrm{Par}\,(L_3) \ldots \Uparrow C\,\mathrm{Par}\,(L_{k+1})$$

sont équivalents.

Il en résulte que :

$$L_2 \uparrow L_3 \ldots \uparrow L_{k+1} \qquad \text{et} \qquad C\,\mathrm{Par}\,(L_2) \Uparrow \ldots \Uparrow C\,\mathrm{Par}\,(L_{k+1})$$

sont rationnellement équivalents. On sait alors, en vertu du lemme 1, que :

$$C\,\mathrm{Par}(L_1)) \Uparrow (C\,\mathrm{Par}(L_2)) \ldots \Uparrow C\,\mathrm{Par}(L_{k+1})) = C\,\mathrm{Par}\,(L^{(k+1)})$$

et :

$$C\,\mathrm{Par}(L_1) \Uparrow (L_2 \uparrow \ldots \uparrow L_{k+1})$$

sont équivalents. Or ce dernier langage s'écrit :

$$\{\varepsilon\} \cup a_1\,L_1\,\bar{a}_1 \Uparrow (L_2 \uparrow L_3 \ldots \uparrow L_{k+1})$$
$$= \{\varepsilon\} \cup a_1\,(L_1 \uparrow (L_2 \uparrow \ldots \uparrow L_{k+1}))\,\bar{a}_1 = \{\varepsilon\} \cup a_1\,L^{(k+1)}\,\bar{a}_1,$$

qui est visiblement équivalent à $L^{(k+1)}$. ■

La version parenthésée de $L^{(k)}$, notée $\mathrm{Par}\,(L^{(k)})$, s'obtient en effaçant les indices des lettres dans les mots de $C\,\mathrm{Par}\,(L^{(k)})$. Ainsi, $\mathrm{Par}\,(L^{(k)})$ est-il défini sur $Y=X \cup Z$. Par construction, on sait que $C\,\mathrm{Par}\,(L^{(k)})$ domine rationnellement

Par $(L^{(k)})$. La réciproque est également vraie : en vertu de la remarque 2, on sait que :

$$C\,\mathrm{Par}\,(L^{(k)}) = \mathrm{num}_k\,\mathrm{Par}\,(L^{(k)}).$$

Ainsi :

LEMME 3 : *Les langages* $L^{(k)}$ *et* Par $(L^{(k)})$ *sont rationnellement équivalents.*

II. LE LANGAGE $L^{\dagger}$

Sur l'alphabet $Y = X \cup Z$, on considère le langage :

$$\hat{L} = \bigcup_{1}^{\infty} \mathrm{Par}\,(L^{(k)}).$$

Nous établissons d'abord le :

LEMME 4 : $\hat{L}$ *est algébrique avec* L.

Preuve : Soit $G = \langle Y, V, P \rangle$ une grammaire algébrique d'axiome S engendrant $a L \bar{a} \cup \{\varepsilon\}$. On désigne par G_i la grammaire $\langle Y_i, V_i, P_i \rangle$ engendrant $a_i L_i \bar{a}_i \cup \{\varepsilon\}$; cette grammaire est obtenue en indiçant les lettres terminales et non terminales de G par i. Si l'on désigne par $\bar{X}^{(k)}$ une copie de $X^{(k)}$, on sait construire une grammaire $G^{(k)}$ engendrant C Par $(L^{(k)})$:

$$G^{(k)} = \langle Y^{(k)}, V^{(k)} \cup \bar{X}^{(k)}, P^{(k)} \rangle,$$

où S_1 est l'axiome et où $P^{(k)}$ contient toutes les règles :

$$v_i \to \bar{m} \quad \text{si } (v_i \to m) \in P_i, \qquad 1 \leqq i \leqq k.$$

($\bar{m}$ est obtenu en remplaçant dans m les lettres de $X^{(k)}$ par leur copie dans $\bar{X}^{(k)}$.) :

$$\text{les règles : } \bar{x}_i \to x_i + x_i\, a_{i+1}\, S_{i+1}\, \bar{a}_{i+1}, \qquad 1 \leqq i \leqq k-1;$$
$$\text{et les règles : } \bar{x}_k \to x_k.$$

La réunion infinie des grammaires $G^{(k)}$ donne naissance à une pseudo-grammaire engendrant $\bigcup_{1}^{\infty} C\,\mathrm{Par}\,(L^{(k)})$. L'image par le morphisme effaçant les indices des lettres terminales donne donc une pseudo-grammaire engendrant $\bigcup_{1}^{\infty}$ Par $(L^{(k)})$. Il suffit de vérifier qu'en effaçant aussi les indices des variables, on trouve une grammaire engendrant encore $\hat{L}$:

$$\hat{G}=\langle Y,\ V\cup\bar{X},\ \hat{P}\rangle$$

avec :

$$(v\to\bar{m})\in\hat{P}\qquad\text{ssi}\quad(v\to m)\in P;$$

$$(\bar{x}\to x+xa\,S\,\bar{a})\in\hat{P},\qquad\forall x\in X.\quad\blacksquare$$

On sait que $\pi(\hat{L})\subset D_1\cup\{\varepsilon\}$. On pose alors :

$$Z^*\setminus D_1=\neg D_1\qquad\text{et}\qquad L^\dagger=\hat{L}\cup\pi^{-1}(\neg D_1).$$

Du lemme 4, il résulte immédiatement :

Fait 1 : Si L est algébrique, $L^\dagger$ l'est aussi.

On peut aussi remarquer que $L^\dagger\cap\pi^{-1}(H_k)=\hat{L}\cap\pi^{-1}(H_k)$. Or, en vertu des remarques 3 et 4, on sait que :

$$\hat{L}\cap\pi^{-1}(H_k)=\operatorname{Par}(L^{(k)}).$$

Du lemme 3, on déduit alors :

Fait 2 : $L^\dagger$ domine rationnellement $L^{(k)}\,(k\geqq 1)$, soit $\mathscr{T}(L^\dagger)\supseteq\mathscr{T}^\sigma(L)$.

Étant donné un langage A sur l'alphabet X, on désigne par $\langle A\rangle$ le langage sur :

$$Y=X\cup Z,\qquad\langle A\rangle=\{a^n f\bar{a}^n\,|\,n\geqq 1\text{ et }f\in A\}.$$

Nous allons établir :

LEMME 5 : *Étant donnée une transduction rationnelle τ telle que $\langle B\rangle=\tau(A^\dagger)$, il existe un entier N tel que* :

$$\forall g\in B,\quad\forall n>N,\qquad\exists f\in A^\dagger\cap\pi^{-1}(H_N)$$

tel que $a^n g\bar{a}^n\in\tau(f)$.

Preuve : La transduction rationnelle τ étant donnée par $\langle\psi,K,\varphi\rangle$, on choisit $(N-1)$ comme le nombre d'états d'un automate reconnaissant K.

On choisit f, le mot le plus court de $A^\dagger$ tel que $a^n g\bar{a}^n\in\tau(f)$. Nous allons montrer que $f\in\pi^{-1}(H_N)$.

Par hypothèse, il existe un mot h dans K tel que :

$$\psi(h)=a^n g\bar{a}^n,\qquad\varphi(h)=f.$$

Comme n est supérieur à N, il existe une factorisation de h en $h_1 h_2 h_3$ satisfaisant :

(1) $\psi(h_2)=a^r$, $r\neq 0$;

(2) $h_1 h_2^* h_3' \subseteq K$;

(3) $\psi(h_3)=a^s g\bar{a}^{s'}$, $s\neq 0$.

On en déduit que $\lambda\neq 1 \Rightarrow \varphi(h_1 h_2^\lambda h_3)\notin A^\uparrow$, et donc :

$$|\varphi(h_2)|_a=|\varphi(h_2)|_{\bar{a}},$$

sinon, en itérant $\varphi(h_2)$, on se trouvera dans $\pi^{-1}(\neg D_1)$ et donc dans $A^\uparrow$ pour une infinité de valeurs de λ.

Dans ces conditions, on sait que $f\in A^\uparrow$ implique que $f\in\hat{A}$; sinon, à nouveau, en itérant $\varphi(h_2)$, on se trouvera dans $\pi^{-1}(\neg D_1)$ pour toutes les valeurs de λ sauf une au plus. Sachant que $f\in\hat{A}$, il reste à établir que la hauteur de f ne peut pas dépasser N. Si tel était le cas, on pourrait factoriser h en $h_1 h_2 h_3 = h = h' u h''$ avec :

(1) $h' u^* h'' \subseteq K$;

(2) $\pi\varphi(u)\equiv a^t$, $t\neq 0$.

On note alors que $\varphi(h' u^m h'')\in\eta^{-1}(\neg D_1)$ pour $m\neq 1$. Ainsi, $\psi(h' u^m h'')\in\langle B\rangle$ pour $m\neq 1$. Ceci implique que $\psi(u)$ ne contienne ni a ni $\bar{a}$. On sait par ailleurs que $\psi(u)\neq\varepsilon$ puisque f est de longueur minimale. Il en résulte que u est facteur de h_3 et donc :

$$h=h_1 h_2 h_3' u h''.$$

On sait maintenant que :

$$\psi(h_1 h_2^\lambda h_3' u^\mu h'')\notin\langle B\rangle,\qquad \lambda\neq 1,\quad \mu\in\mathbb{N};$$

et :

$$h_1 h_2^\lambda h_3' u^\mu h''\in K,\qquad \forall\lambda,\ \mu.$$

On doit donc avoir :

$$\varphi(h_1 h_2^\lambda h_3' u^\mu h'')\notin A^\uparrow,\qquad \lambda\neq 1,\quad \mu\in\mathbb{N}.$$

Or, si l'on fixe $\lambda=2$, pour toutes les valeurs de μ différentes de 1, on aura :

$$\pi\varphi(h_1 h_2^2 h_3' u^\mu h'')\in\neg D_1$$

et donc :

$$\varphi(h_1 h_2^2 h_3' u^\mu h'')\in A^\uparrow\qquad (\mu\neq 1),$$

ce qui est impossible.

Ainsi, la hauteur de $\varphi(h)$ ne peut-elle pas dépasser N et le lemme est établi.

LEMME 6 : *Étant donnés deux langages A et B, $A^{\uparrow}$ domine $\langle B \rangle$ implique qu'il existe un entier k tel que $A^{(k)}$ domine B.*

Preuve : Ceci est une conséquence du lemme 5. En effet, si $A^{\uparrow}$ domine $\langle B \rangle$, on sait que $A^{\uparrow} \cap \pi^{-1}(H_N)$ domine $\{a^n g \bar{a}^n \mid n \geqq N, g \in B\}$. Or, nous savons que $A^{\uparrow} \cap \pi^{-1}(H_N)$ et $A^{(N)}$ sont équivalents. Ainsi, $A^{(N)}$ domine-t-il $\{a^n g \bar{a}^n \mid n \geqq N, g \in B\}$ et donc B.

LEMME 7 : *Étant donnés deux langages A et B, $A^{(k)}$ domine $B^{\uparrow}$ implique que $\mathscr{T}(A)$ contienne $\mathscr{T}^{\sigma}(B)$, et donc que A domine B.*

Preuve : Si $A^{(k)}$ domine $B^{\uparrow}$, on a $\mathscr{T}(A^{(k)}) \supseteq \mathscr{T}(B^{\uparrow})$. En vertu du fait 2, $\mathscr{T}(B^{\uparrow}) \supseteq \mathscr{T}^{\sigma}(B)$, et donc $\mathscr{T}(A^{(k)}) \supseteq \mathscr{T}^{\sigma}(B)$. Or, si $L \in \mathscr{T}^{\sigma}(B)$, on a $L^{(k)} \in \mathscr{T}^{\sigma}(B)$ et donc $A^{(k)}$ domine $L^{(k)}$. Alors, en vertu du lemme syntaxique, soit A domine L, soit $A^{(k-1)}$ domine $L^{(k-1)}$. Le lemme est alors immédiat :

$$L \in \mathscr{T}^{\sigma}(B) \quad \Rightarrow \quad L \in \mathscr{T}(A).$$

Ce lemme appelle la question :

Si $A^{(k)}$ domine $B^{\uparrow}$, A domine-t-il $B^{\uparrow}$?

Nous pouvons maintenant établir le :

Fait 3 : Le langage $L^{\uparrow}$ est générateur ssi L est générateur.

Preuve : Si $\mathscr{T}(L) = \mathrm{Alg}$, on sait que $\mathscr{T}^{\sigma}(L) = \mathrm{Alg}$. Par ailleurs, le fait 1 assure $\mathscr{T}(L^{\uparrow}) \subseteq \mathrm{Alg}$; le fait 2 assure :

$$\mathscr{T}(L^{\uparrow}) \supseteq \mathscr{T}^{\sigma}(L) = \mathrm{Alg} \qquad \text{et donc} \qquad \mathscr{T}(L^{\uparrow}) = \mathrm{Alg}.$$

Ainsi, $\mathscr{T}(L) = \mathrm{Alg}$ implique $\mathscr{T}(L^{\uparrow}) = \mathrm{Alg}$.

Réciproquement, si $\mathscr{T}(L^{\uparrow}) = \mathrm{Alg}$, on en déduit que $L^{\uparrow}$ domine $\langle E^{\uparrow} \rangle$.

Du lemme 6, on déduit qu'il existe un entier k tel que $L^{(k)}$ domine $E^{\uparrow}$.

Le lemme 7 affirme alors que L domine E et donc $\mathscr{T}(L) = \mathrm{Alg}$.

III. LE RÉSULTAT

Nous commençons par quelques corollaires immédiats de ce qui précède. Quel que soit le langage L engendrant un cône non clos par substitution, on a $\mathscr{T}^{\sigma}(L) \nsubseteq \mathscr{T}(L^{\uparrow}) \subseteq \mathrm{Nge}$. En particulier :

COROLLAIRE 1 : *Il existe un cône principal strictement inclus dans* Alg *contenant la famille des langages quasirationnels.*

Ceci répond à une question de [1].

On notera également que :

COROLLAIRE 2 : *Étant donné un langage L tel que* $\mathscr{T}^{\sigma}(L)=\mathrm{Nge}$, *alors* $\mathscr{T}(L)=\mathrm{Nge}$.

Preuve : On sait que $\mathscr{T}^{\sigma}(L)\subseteq\mathscr{T}(L^{\dagger})\subseteq\mathrm{Nge}$, et donc que ces trois cônes sont égaux.

Alors, $L^{\dagger}$ domine $\langle L^{\dagger}\rangle$ et donc il existe un entier k tel que $L^{(k)}$ domine $L^{\dagger}$.

Ainsi, en vertu du lemme 7, $\mathscr{T}(L)\supseteq\mathscr{T}^{\sigma}(L)$, soit $\mathscr{T}(L)=\mathrm{Nge}$.

Ce dernier corollaire n'est autre que le résultat annoncé dans le cas particulier d'un cône principal. Pour obtenir le résultat général, nous établissons d'abord :

PROPOSITION 1 : *Étant donné un cône rationnel* $\mathscr{L}$, *sa clôture par union* $\mathscr{L}_u$ *contient* Nge *si et seulement si* $\mathscr{L}$ *contient* Nge.

Preuve : Clairement, si $\mathscr{L}$ contient Nge, il en va de même de sa clôture par union. Il suffit donc d'établir l'inverse. Pour ce faire, admettons d'abord que Rat et Fin désignant les familles des langages rationnels et des langages finis respectivement :

$$\text{(i)} \qquad \mathrm{Rat}\,\square\,\mathscr{L}\subseteq\mathrm{Fin}\,\square\,\mathscr{M} \text{ implique } \mathrm{Rat}\,\square\,\mathscr{L}\subseteq\mathscr{M}.$$

De (i), on déduit que, si $\mathscr{L}_u\supseteq\mathrm{Nge}$ on a $\mathrm{Fin}\,\square\,\mathscr{L}\supseteq\mathrm{Nge}$. Par ailleurs, $\mathrm{Rat}\,\square\,\mathrm{Nge}$ est égal à Nge et donc $\mathrm{Rat}\,\square\,\mathrm{Nge}\subseteq\mathrm{Fin}\,\square\,\mathscr{L}$. De (i), il résulte que $\mathrm{Nge}\subseteq\mathscr{L}$.

Le lemme sera donc prouvé si nous prouvons (i).

Soit donc $L\in\mathrm{Rat}\,\square\,\mathscr{L}$. Alors, $c^*\uparrow L\in\mathrm{Rat}\,\square\,\mathscr{L}$ et donc $c^*\uparrow L\in\mathrm{Fin}\,\square\,\mathscr{M}$. On déduit du lemme syntaxique que soit $c^*\in\mathrm{Fin}$, soit $(Lc)^*\in\mathscr{M}$. Ainsi, $L\in\mathscr{L}$ implique $(Lc)^*\in\mathscr{M}$ et donc $\mathrm{Rat}\,\square\,\mathscr{L}\subseteq\mathscr{M}$.

PROPOSITION 2 : *Soit* $\mathscr{L}$ *un cône rationnel, si* $\mathrm{Nge}\subseteq\mathscr{T}^{\sigma}(\mathscr{L})$, *alors* $\mathrm{Nge}\subseteq\mathscr{L}$.

Preuve : Commençons par supposer que $\mathscr{L}$ est clos par union. On pose alors $\mathscr{L}^{\dagger}=\{L^{\dagger}\mid L\in\mathscr{L}\}$. On établit d'abord :

$$\text{(i)} \qquad A\in\mathscr{T}^{\sigma}(\mathscr{L}) \quad\Rightarrow\quad \exists\, B\in\mathscr{L} \quad \text{tel que} \quad A\in\mathscr{T}(B^{\dagger}).$$

En effet, si $A\in\mathscr{T}^{\sigma}(\mathscr{L})$, il existe un nombre fini de langages $L_1, L_2, \ldots, L_k$ de $\mathscr{L}$ tels que $A\in\mathscr{T}^{\sigma}(L_1, L_2, \ldots, L_k)$. $\mathscr{L}$ étant clos par union, il existe un langage B dans $\mathscr{L}$ tel que $A\in\mathscr{T}^{\sigma}(B)$. Mais alors $\mathscr{T}^{\sigma}(B)\subseteq\mathscr{T}(B^{\dagger})$ assure $A\in\mathscr{T}(B^{\dagger})$.

(i) étant établi, on considère un langage L dans Nge. Alors, en vertu du fait 3, $L^{\dagger}$ est non générateur, et donc $\langle L^{\dagger}\rangle\in\mathrm{Nge}$. Ainsi, $\langle L^{\dagger}\rangle\in\mathscr{T}^{\sigma}(\mathscr{L})$. Il existe donc un langage B dans $\mathscr{L}$ tel que $\langle L^{\dagger}\rangle\in\mathscr{T}(B^{\dagger})$. Du lemme 6, il résulte qu'il existe un entier k tel que $B^{(k)}$ domine $L^{\dagger}$. Le lemme 7 garantit alors que B domine L. Ainsi, $L \in$ Nge implique $L \in \mathscr{T}(\mathscr{L})$, ce qui établit la proposition dans le cas où $\mathscr{L}$ est clos par union.

Si $\mathscr{L}$ n'est pas clos par union, on remarque que $\mathscr{T}^{\sigma}(\mathscr{L}) = \mathscr{T}^{\sigma}(\mathscr{L}_u) =$ Nge. Alors, $\mathscr{L}_u \supseteq$ Nge et, en vertu de la proposition 1, $\mathscr{L} \supseteq$ Nge.

Nous pouvons maintenant établir le résultat principal de cet article :

THÉORÈME : *Étant donné un cône rationnel $\mathscr{L}$, la clôture par substitution de $\mathscr{L}$ est égale à* Nge *si et seulement si $\mathscr{L} =$* Nge.

Preuve : On sait que si $\mathscr{L} =$ Nge, la clôture par substitution de $\mathscr{L}$ est aussi égale à Nge. Réciproquement, si Nge $= \mathscr{T}^{\sigma}(\mathscr{L})$, la proposition 2 garantit que $\mathscr{L} \supseteq$ Nge $= \mathscr{T}^{\sigma}(\mathscr{L}) \supseteq \mathscr{L}$; donc $\mathscr{L} =$ Nge.

IV. COMPLÉMENTS

Le passage de L à $\hat{L} = \bigcup_{1}^{\infty} \operatorname{Par}(L^{(k)})$ est une opération puissante.

En effet, si $L = \{xx\}$, $\hat{L}$ est engendré par la grammaire $\langle S \to ax\,Sx\,S\bar{a} + \varepsilon \rangle$ qui est générateur. De façon générale, on a :

Observation 1 : $L^{\dagger}$ n'est jamais rationnel.

Clairement, si f est un mot de $L^{\dagger}$ ne commençant pas par a, il est dans $\pi^{-1}(\neg D_1)$, langage non rationnel dominé par $L^{\dagger}$.

On peut aussi observer que si $\mathscr{T}(L) \nsupseteq \mathscr{T}^{\sigma}(L)$, il suffit que $\{a^n b^n \mid n \in N\}$ soit dans $\mathscr{T}(L)$ pour garantir que $\mathscr{T}(L^{\dagger}) \nsupseteq \mathscr{T}((L^{\dagger})^{\dagger})$. En effet, sinon on aurait $\langle L^{\dagger} \rangle \in \mathscr{T}(L^{\dagger})$ et, en vertu des lemmes 6 et 7, $\mathscr{T}^{\sigma}(L) \subseteq \mathscr{T}(L)$.

On aurait alors $\mathscr{T}(L) = \mathscr{T}^{\sigma}(L)$. Ceci conduit à deux questions :

Question 1 : Existe-t-il dans Nge un langage tel que L domine $L^{\dagger}$?

Question 2 : Peut-on construire un cône rationnel principal différent de Rat et de Alg qui soit clos par substitution?

Pour terminer, nous donnons une description directe de $L^{\dagger}$: étant donné un mot de $\pi^{-1}(D_1)$, on considère les sous-mots ainsi formés :

— choisir deux parenthèses associées;

— sélectionner les lettres de X qui ne sont entourées d'aucune autre parenthèse à l'intérieur du facteur choisi;

— conserver le sous-mot obtenu.

On note $D_1 \circ L$ le langage satisfaisant que tous ces sous-mots sont dans L.

Exemple : Si $f = axaxx\bar{a}axy\bar{a}xax\bar{a}\bar{a}$, en considérant les parenthèses extérieures, on construit $ax(axx\bar{a})(axy\bar{a})x(ax\bar{a})\bar{a}$ qui donne le sous-mot xx; les parenthèses intérieures donnent les sous-mots xx, xy et x. Ainsi $f \in D_1 \circ L$ ssi x, xx et xy sont dans L. $D_1 \circ L$ est voisin de $\hat{L}$: il n'en diffère que par les mots contenant aa ou $\bar{a}a$. Ainsi, on voit que $D_1 \circ L$ domine rationnellement $L^{\uparrow}$.

On vérifie alors sans peine que $D_1 \quad L \cup \pi^{-1}(\neg D_1) = L^{\downarrow}$ jouit des mêmes propriétés que $L^{\uparrow}$:

$L^{\downarrow}$ est algébrique avec L;

$L^{\downarrow}$ est non générateur avec L;

$L^{\downarrow}$ engendre un cône contenant $\mathscr{T}^{\sigma}(L)$.

Sur cette description, il est facile de voir que $L^{\downarrow}$ et $L^{\uparrow}$ sont déterministes avec L.

REMERCIEMENTS

Je tiens à remercier Michel Latteux qui est à l'origine de l'extension du résultat principal au cas des cônes non principaux, ses relectures attentives ayant permis également de simplifier bon nombre de preuves.

BIBLIOGRAPHIE

1. J.-M. AUTEBERT, J. BEAUQUIER, L. BOASSON, M. NIVAT, *Quelques problèmes ouverts en théorie des langages algébriques*, R.A.I.R.O. Informatique théorique, vol. 13, 1979, p. 363-379.
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-Full AFL of the Family of CFL's is Proper*, Information Proc. Letter, vol. 2, 1973, p. 135-140.
4. S. GREIBACH, *Chains of Full AFL'S*, Math. System Theory, vol. 4, 1970, p. 231-242.
5. M. LATTEUX, *A propos du lemme de substitution*, Theoretical Comp. Science, vol. 14, 1981, p. 119-123.