

L. BOASSON

A. PETIT

Langages algébriques déterministes non générateurs

Informatique théorique et applications, tome 21, n° 1 (1987), p. 41-57

<http://www.numdam.org/item?id=ITA_1987__21_1_41_0>

© AFCET, 1987, tous droits réservés.

L'accès aux archives de la revue « Informatique théorique et applications » 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/>

LANGAGES ALGÈBRIQUES DÉTERMINISTES NON GÉNÉRATEURS (*)

par L. BOASSON ⁽¹⁾ et A. PETIT ⁽²⁾

Communiqué par J. BERSTEL

Résumé. — On désigne par *Nge* le plus grand sous-cône strictement inclus dans celui des langages algébriques. L'objet de cet article est d'établir qu'aucun langage déterministe ne peut être générateur de *Nge*. Ce résultat laisse cependant ouverte la question générale de la principalité de *Nge*. Cependant, il renforce la conjecture que *Nge* devrait être non principal.

Abstract. — We denote by *Nge* the largest full sub-AFL of the family of context-free languages. The aim of this paper is to prove that no deterministic context-free language can be a full generator of *Nge*. This result leaves open the general question of whether *Nge* is principal or not. However, it strengthens the conjecture that *Nge* should be non principal.

La famille des langages algébriques (ou « context-free ») a fait l'objet de nombreuses études. L'apparition il y a environ quinze ans du cadre formel des familles de langages a permis non seulement de résoudre plusieurs problèmes, mais aussi de dégager un point de vue nouveau sur divers résultats déjà connus. Ainsi, les propriétés de clôture de la famille des langages algébriques maintenant classiques, permettent-elles de dire que cette famille forme un cône rationnel et même un « full AFL ». Ainsi encore, le théorème de Chomsky-Schützenberger peut se réenoncer en disant que le cône rationnel Alg des langages algébriques est principal : tout langage de Dyck en est un générateur. Cette façon de présenter ce résultat classique conduit naturellement à se poser de nouvelles questions. Par exemple, il est naturel de chercher à savoir quels sont les langages algébriques qui, comme les langages de Dyck, permettent d'obtenir la famille Alg toute entière. Plus généralement, il est facile de voir que tout cône rationnel principal \mathcal{C} contient un sous-cône strict

(*) Reçu novembre 1985, révisé juin 1986.

⁽¹⁾ U.E.R. de Mathématiques et Informatique, Université Paris-VII, 2, place Jussieu, 75005 Paris, France.

⁽²⁾ U.E.R. de Sciences, Département de Mathématiques et Informatique, Université d'Orléans, B.P. n° 67-59, 45067 Orléans Cedex 2, France.

maximal; il est formé de tous les langages de \mathcal{C} qui ne sont pas générateurs de \mathcal{C} . En particulier Alg contient un plus grand sous-cône noté ici Nge (pour famille des langages Non générateurs). On sait par ailleurs que ce sous-cône est clos par substitution [3]. La question se pose alors de savoir si ce cône Nge est principal ou non. Cette question est à ce jour non résolue bien que de nombreuses tentatives aient été faites pour prouver que non [4, 5, 7 à 10]. On notera à ce propos que pour aucun cône rationnel principal (autre que celui des langages rationnels où le problème est trivial) on ne connaît le statut vis-à-vis de la principalité de son plus grand sous-cône strict. Revenant à Nge, on a longtemps tenté de donner une description constructive de ce cône rationnel à l'aide de familles plus petites et de l'opération de substitution. Ceci est en effet l'un des moyens les plus puissants connus pour former de grands cônes non principaux. Cependant, on sait depuis peu que cette méthode est, par nature, vouée à l'échec [6]. Ainsi, pour résoudre le problème de la (non) principalité de Nge, il nous faut de nouvelles approches.

C'est dans ce cadre que s'inscrit le présent article. Son but est d'établir un résultat partiel qui distingue fortement Nge de tous les sous-cônes classiques (principaux ou non) de Alg. Plus précisément, nous montrons qu'aucune famille de langages déterministes ne peut engendrer Nge. On notera qu'au contraire, la famille Alg n'admet elle que des générateurs déterministes, en ce sens que, si G est générateur de Alg, il existe toujours un langage rationnel K tel que $G \cap K$ soit générateur et déterministe [2]. On notera aussi qu'il est assez difficile de trouver des langages algébriques n'ayant aucun langage rationnellement équivalent qui soit déterministe [1]. Ainsi, le résultat proposé ici montre-t-il que Nge est sûrement d'une nature très complexe. Il est également intéressant de signaler que la méthode de preuve utilisée ici permet d'obtenir des résultats semblables pour les sous-cônes rationnels maximaux des familles Lin des langages linéaires et RoCl des langages à un compteur. On trouvera dans [11] les preuves complètes de ces deux résultats; nous nous limitons ici au cas de Nge. En tout état de cause, tout ceci laisse entrevoir que la définition d'un cône rationnel comme sous-cône maximal d'un autre fait apparaître des familles de langages plus complexes que celles définies par des procédés consécutifs (grammaires, automates, ...).

Cet article est divisé en quatre parties. La première est constituée de rappels et de quelques résultats plus ou moins classiques que nous utilisons. Elle contient essentiellement la preuve d'un résultat généralement admis dont aucune preuve complète n'est jamais parue. La deuxième partie est très courte; elle présente le principe de la preuve et ses conséquences. La troisième partie

est le cœur de l'article: elle donne la preuve des propriétés exigées dans la seconde partie pour obtenir le résultat. Enfin, nous terminons par une dernière partie présentant quelques extensions possibles de notre résultat.

I. PRÉLIMINAIRES

Nous supposons le lecteur familier avec la théorie des langages algébriques et des familles de langages telle qu'elle est présentée dans [3]. Nous nous contenterons ici de revenir sur quelques notions particulières concernant les langages déterministes.

Un automate à pile (déterministe) est donnée par son alphabet d'entrée X , son alphabet de pile Z , son ensemble d'états Q , ses états terminaux Q_F , son état initial q^0 , son symbole initial de pile z^0 et sa fonction de transition δ soit

$$\mathcal{A} = \langle X, Z, Q, Q_F, q^0, z^0, \delta \rangle$$

Une configuration de l'automate est un triplet (f, q, m) de $X^* \times Q \times Z^*$. On note \vdash le passage direct d'une configuration à la suivante. Un calcul valide est une suite de configurations c_1, c_2, \dots, c_n telle que $c_i \vdash c_{i+1}$ ($0 < i < n$). Un calcul est maximal s'il se termine dans une configuration à laquelle aucune règle de \mathcal{A} ne soit applicable. Étant donné un mot f sur X , il admet une paire itérante automatique pour (q, z) relativement à \mathcal{A} si f se factorise en $aubvc$ de telle façon que

$$(a, q, z) \vdash^* (\varepsilon, s, mz')$$

$$(u, s, z') \vdash^* (\varepsilon, s, pz')$$

$$(b, s, z') \vdash^* (\varepsilon, t, \varepsilon)$$

$$(v, t, p) \vdash^* (\varepsilon, t, \varepsilon)$$

$$(c, t, m) \vdash^* (\varepsilon, q', m')$$

On sait que tout mot suffisamment long reconnu par \mathcal{A} contient au moins une paire itérante automatique (si le calcul de \mathcal{A} pour f vide la pile). On sait aussi que l'on peut normaliser un automate déterministe pour que sa fonction

de transition satisfasse :

- (1) les ε -règles sont toutes de type descendant :

$$\delta(\varepsilon, q, z) = (q', \varepsilon)$$

- (2) toutes les règles font varier la hauteur de pile de ± 1

$$\delta(x, q, z) = (q', \varepsilon) \quad \text{ou} \quad (q', z'z'')$$

Notons pour terminer que l'on sait que si D' est un langage déterministe, il existe toujours un langage déterministe préfixe D rationnellement équivalent tel que D' soit image de D dans un morphisme (alphabétique).

Nous allons établir la

PROPRIÉTÉ 1 : Soit $\mathcal{A} = \langle X, Z, Q, Q_F, q^0, z^0, \delta \rangle$ un automate à pile déterministe (sous forme normale), alors le langage

$$R = \{ w \in X^* \mid w \text{ est lu par } \mathcal{A} \text{ sans paire itérante automatique} \}$$

est rationnel.

Cette propriété est cruciale pour nos preuves. Bien qu'elle soit connue de tous, il n'en existe aucune preuve publiée; aussi, allons nous établir cette propriété.

Preuve de la propriété 1 : Introduisons les notations suivantes :

Pour y, \bar{y} dans Z et q, \bar{q} dans Q définissons les ensembles :

$$R(q, y, \bar{q}, \bar{y})^* = \{ w \in Z^* \mid \text{il existe un calcul valide } (c_1, \dots, c_n) \text{ avec } n \geq 2 \\ \text{allant de la configuration } c_1 = (w, q, y) \text{ à la configuration } c_n = (\varepsilon, \bar{q}, \bar{y}) \\ \text{et vérifiant : } w \text{ est lu par ce calcul sans paire itérante automatique} \}.$$

Remarque : Cette définition entraîne que la hauteur de pile des configurations c_i ($1 \leq i \leq n$) est au moins 1.

Définissons encore

$$RT(q, y, \bar{q}, \bar{y}) = \begin{cases} R(q, y, \bar{q}, \bar{y}) & \text{si } \delta(\varepsilon, \bar{q}, \bar{y}) = \emptyset \\ \emptyset & \text{sinon} \end{cases}$$

i. e. un ensemble $R(q, y, \bar{q}, \bar{y})$ est également un ensemble $RT(q, y, \bar{q}, \bar{y})$ si toutes les lectures des mots de $R(q, y, \bar{q}, \bar{y})$ sont maximales i. e. si on ne peut pas compléter ces lectures par des ε -transitions.

Les descentes élémentaires et les montées élémentaires sont définies par

$$\begin{aligned} \text{Del}(q, y, \bar{q}) &= \{x \in X \cup \{\varepsilon\} / (x, q, y) \vdash (\bar{q}, \varepsilon)\} \\ \text{Mel}(q, y, \bar{q}, \bar{y}) &= \{x \in X \cup \{\varepsilon\} / (x, q, y) \vdash (\bar{q}, z\bar{y})\}. \end{aligned}$$

Comme pour les $R(q, y, \bar{q}, \bar{y})$, on définit également :

$$\text{Mel } T(q, y, \bar{q}, \bar{y}) = \begin{cases} \text{Mel}(q, y, \bar{q}, \bar{y}) & \text{si } \delta(\varepsilon, \bar{q}, \bar{y}) = \emptyset \\ \emptyset & \text{sinon.} \end{cases}$$

Nous utiliserons plus loin la remarque suivante :

Remarque : Les ensembles $\text{Del}(q, y, \bar{q})$, $\text{Mel}(q, y, \bar{q}, \bar{y})$ et $\text{Mel } T(q, y, \bar{q}, \bar{y})$ sont évidemment finis. Les ensembles $R(q, y, \bar{q}, \bar{y})$ et $RT(q, y, \bar{q}, \bar{y})$ le sont également car d'après le lemme d'itération d'Ogden, tout mot suffisamment long a une paire itérante automatique.

L'idée intuitive de la démonstration de la propriété 1 est la suivante : considérons w lu par \mathcal{A} sans paire itérante automatique et w_1 le préfixe de w après lecture duquel la hauteur de pile est pour la dernière fois 1. w_1 appartient à un ensemble du type $R(q_0, z_0, q_1, z_1)$. Pour continuer la lecture de w ou bien on a une descente élémentaire $\text{Del}(q_1, z_1, q_2)$ ou bien une montée élémentaire $\text{Mel}(q_1, z_1, q_2, z_2)$ et l'on peut alors recommencer avec la hauteur 2 et ainsi la suite.

En suivant cette idée intuitive, nous allons construire un ensemble à partir des $R(q, y, q, y)$, $RT(q, y, q, y)$, $\text{Del}(q, y, q)$, $\text{Mel}(q, y, q, y)$, $\text{Mel } T(q, y, \bar{q}, \bar{y})$, qui sera l'ensemble des mots lus par l'automate \mathcal{A} sans paire itérante automatique.

Posons

$$\begin{aligned} T &= \{R(q, y, \bar{q}, \bar{y}), \text{Del}(q, y, \bar{q}), \text{Mel}(q, y, \bar{q}, \bar{y}) / q, \bar{q} \in Q, y, \bar{y} \in Z\}, \\ T' &= T \cup \{RT(q, y, \bar{q}, \bar{y}), \text{Mel } T(q, y, \bar{q}, \bar{y}) / q, \bar{q} \in Q, y, \bar{y} \in Z\} \end{aligned}$$

et définissons la substitution σ de T'^* dans X^* qui, à un élément de T' , associe le langage (fini) qu'il représente.

Définissons sur l'alphabet T' les ensembles suivants :

$$K_0 = \{R(q_0, z_0, q_1, z_1) \text{Del}(q_1, z_1, q_2) / q_1, q_2 \in Q, z_1 \in Z\}$$

K_0 sera l'ensemble des mots lus sans paire itérante automatique et dont le calcul mène à la hauteur zéro.

On construit un rationnel pseudo-local K_1 sur T^* : les transitions autorisées sont :

$$\begin{aligned}
 A &= \{ R(q, y, \bar{q}, \bar{y}) \text{ Mel}(\bar{q}, \bar{y}, q', y'), R(q, y, \bar{q}, \bar{y}) \text{ Mel} T(\bar{q}, \bar{y}, q', y') / \\
 &\hspace{25em} q, \bar{q}, q' \in Q, y, \bar{y}, y' \in Z \} \\
 B &= \{ \text{Mel}(q, y, \bar{q}, \bar{y}) \text{ Mel}(\bar{q}, \bar{y}, q', y'), \text{Mel}(q, y, \bar{q}, \bar{y}) \text{ Mel} T(\bar{q}, \bar{y}, q', y') / \\
 &\hspace{25em} q, \bar{q}, q' \in Q, y, \bar{y}, y' \in Z \} \\
 C &= \{ \text{Mel}(q, y, \bar{q}, \bar{y}) R(\bar{q}, \bar{y}, q', y'), \text{Mel}(q, y, \bar{q}, \bar{y}) RT(\bar{q}, \bar{y}, q', y') / \\
 &\hspace{25em} q, \bar{q}, q' \in Q, y, \bar{y}, y' \in Z \} \\
 V &= A \cup B \cup C.
 \end{aligned}$$

De plus les mots de K_1 doivent commencer par une lettre (de T) appartenant à

$$E = \{ R(q_0, z_0, q, y), \text{Mel}(q_0, z_0, q, y) / q \in Q, y \in Z \}$$

et finir par une lettre appartenant à :

$$F = \{ RT(q, y, \bar{q}, \bar{y}), \text{Mel} T(q, y, \bar{q}, \bar{y}) / q, \bar{q} \in Q, y, \bar{y} \in Z \}.$$

En outre on impose que seule la dernière lettre des mots de K_1 soit de la forme $RT(q, y, \bar{q}, \bar{y})$ ou $\text{Mel} T(q, y, \bar{q}, \bar{y})$. Ainsi on a

$$K_1 = (ET^*F \setminus T^*VT^*) \cup F$$

K_1 sera l'ensemble des mots sans paire itérante automatique et dont le calcul mène à une hauteur strictement positive.

Enfin on pose :

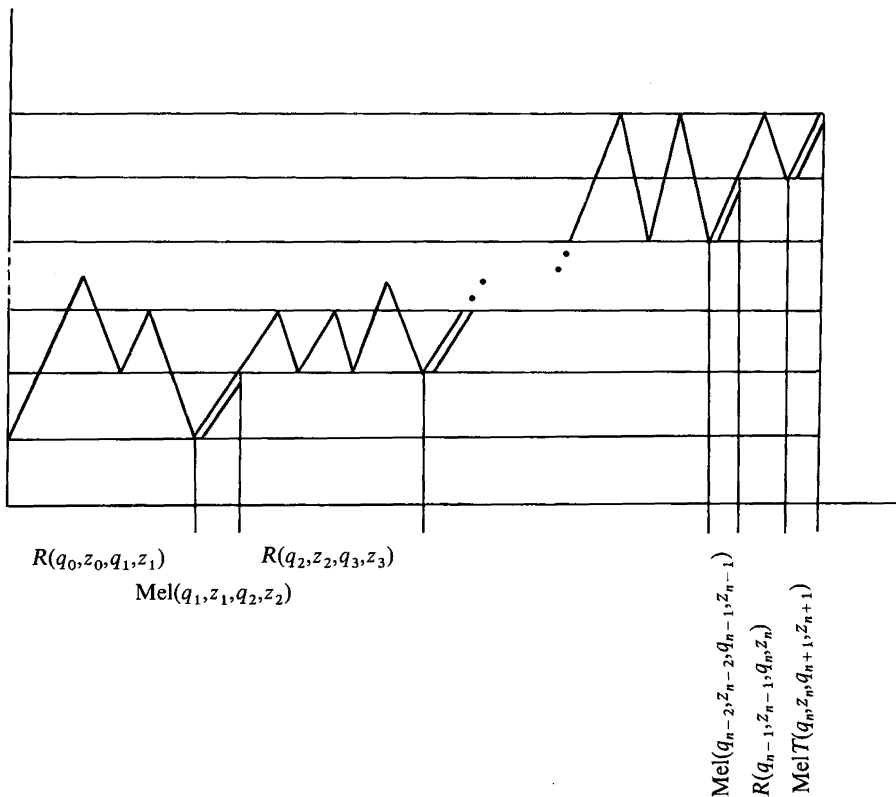
$$K = K_0 \cup K_1$$

K est clairement un ensemble rationnel sur T^* . Mais σ étant une substitution finie $\sigma(K)$ est un ensemble rationnel sur X^* .

Remarque : L'ensemble R est égal à l'ensemble $\sigma(K)$, ce qui montrera que R est un ensemble rationnel.

1^{re} étape: $R \subseteq \sigma(K)$

Soit $w \in R$, considérons le calcul maximal de w par \mathcal{A} :



L'ensemble K ayant été construit pour que les mots de $\sigma(K)$ soient lus justement sans paire itérante automatique, on démontre cette inclusion de façon immédiate par récurrence sur la hauteur de pile après lecture du mot w . Le raisonnement traduit exactement la construction ci-dessus.

2^e étape: $\sigma(K) \subseteq R$

Soit $w \in \sigma(K)$, considérons la décomposition de w sur T'^* [i.e. $\sigma^{-1}(w)$] et supposons que w n'appartienne pas à R i.e. que w soit lu avec une paire

itérante automatique $(\alpha, u, \beta, v, \gamma)$. On a alors deux possibilités :

- (i) ou bien $u\beta v$ « traverse » un $\text{Mel}(q, y, \bar{q}, \bar{y})$ [ou un $\text{Mel}T(q, y, \bar{q}, \bar{y})$];
- (ii) ou bien $u\beta v$ est inclus dans un $R(q, y, \bar{q}, \bar{y})$ [ou un $RT(q, y, \bar{q}, \bar{y})$].

Dans le premier cas, la hauteur de pile après lecture de v sera strictement plus grande que la hauteur de pile avant lecture de u contrairement à la définition d'une paire itérante automatique. Ce cas est donc exclu.

Dans le second cas $u\beta v$ sera facteur d'un mot de $R(q, y, \bar{q}, \bar{y})$ [ou $RT(q, y, \bar{q}, \bar{y})$] qui contiendra donc un mot ayant une paire itérante automatique contrairement à la définition même de $R(q, y, \bar{q}, \bar{y})$ [resp. $RT(q, y, \bar{q}, \bar{y})$]. Ce cas est également exclu.

Ainsi w ne peut avoir de paire itérante automatique et donc w appartient à R .

On a donc prouvé $R \subseteq \sigma(K)$ et $\sigma(K) \subseteq R$ ce qui entraîne $R = \sigma(K)$ et R est donc rationnel, ce qui prouve la propriété 1.

Remarque: La démonstration de la propriété 1 prouve que si \mathcal{A} est un automate à pile quelconque, $R = \{w \in X^* / \text{il existe une lecture de } w \text{ sans paire itérante automatique}\}$ est un ensemble rationnel. Par contre l'ensemble $S = \{w \in X^* / \text{pour toute lecture de } w, w \text{ est lu sans paire itérante automatique}\}$ n'est en général pas rationnel, même si l'on suppose l'automate non ambigu, comme le prouve le contre-exemple suivant. Les règles de l'automate \mathcal{A} sont :

$$\begin{array}{l}
 (a, q_0, z_0) \vdash (q_0, z_0 z_0) \\
 (b, q_0, z_0) \vdash (q_0, z_0 z_1) \quad (b, q_0, z_0) \vdash (q'_0, \varepsilon) \\
 (b, q_0, z_1) \vdash (q_0, z_1 z_1) \quad (b, q'_0, z_0) \vdash (q'_0, \varepsilon) \\
 (\#, q_0, z_1) \vdash (q_1, z_1) \quad (\#, q'_0, z_0) \vdash (q_3, z_0) \\
 (x, q_1, z_1) \vdash (q_1, \varepsilon) \quad (x, q_3, z_0) \vdash (q_3, z_0 z_0) \\
 (z, q_1, z_0) \vdash (q_1, \varepsilon) \quad (y, q_3, z_0) \vdash (q_4, \varepsilon) \\
 (\varepsilon, q_1, z_0) \vdash (q_2, \varepsilon) \quad (y, q_4, z_0) \vdash (q_4, \varepsilon) \\
 (\varepsilon, q_4, z_0) \vdash (q_2, \varepsilon)
 \end{array}$$

\mathcal{A} est un automate à pile non ambigu qui reconnaît par pile vide et état final q_2 le langage

$$L = \{a^n b^n \# x^p z^p\} \cup \{a^n b^p \# x^p y^n\}$$

et

$$S \cap a^* b^* \# = \{a^t b^s \#, t \neq s\} \notin \text{Rat.} \quad \blacksquare$$

R étant un ensemble rationnel, l'ensemble $RX \setminus R$ l'est également. On a ainsi de façon aisée la propriété suivante que nous utiliserons dans la suite :

PROPRIÉTÉ 2: Soit $\mathcal{A} = \langle X, Z, Q_0, Q_F, q_0, z_0, \delta \rangle$ un automate à pile déterministe sous forme normale alors

$\{w \in X^*/w = \alpha u \beta v \gamma \text{ et } (\alpha, u, \beta, v, \gamma) \text{ est la seule paire itérante}$
automatique de $w\}$ est un ensemble rationnel.

II. ÉNONCÉ

DÉFINITION : Une base d'un cône rationnel \mathcal{L} est un sous-ensemble B de \mathcal{L} vérifiant $\mathcal{T}(B) = \mathcal{L}$ [i. e. pour tout langage L de \mathcal{L} , il existe un langage E de B et une transduction rationnelle τ tels que $\tau(E) = L$].

Le but de cet article est d'établir le

THÉORÈME : *Nge n'a pas de base déterministe.*

Ceci implique, en particulier, que si Nge est principal, il n'a pas de générateur déterministe. La preuve de ce théorème résulte immédiatement de l'énoncé suivant :

Fait 1 : Soit \mathcal{L} un cône rationnel principal de générateur G et un langage L_0 de \mathcal{L} vérifiant :

Hypothèse 1 : L_0 n'est pas générateur de \mathcal{L} .

Hypothèse 2 : Pour tout morphisme alphabétique h , et pour tout langage déterministe préfixe D , l'égalité $h(D) = L_0$ entraîne que D domine G .

Alors, si A engendre le sous-cône maximal de \mathcal{L} , A n'est pas déterministe.

Preuve du Fait 1 : Si A engendre le sous-cône maximal de \mathcal{L} , A domine L_0 . Il existe donc deux morphismes alphabétiques f et g , et un langage rationnel K tels que $L = f(g^{-1}(A) \cap K)$. Posons $D' = g^{-1}(A) \cap K$; si A est déterministe, D' l'est également. On a donc l'égalité $f(D') = L_0$ avec D' langage déterministe. On sait alors qu'il existe un langage déterministe préfixe D rationnellement équivalent à D' et un morphisme alphabétique h tels que $h(D) = L_0$. Ceci implique que D domine G , et donc que D' et, *a fortiori*, A dominant G . Ainsi A est-il générateur de \mathcal{L} contrairement à l'hypothèse.

Du fait 1, on déduit alors le théorème : si le sous-cône maximal de \mathcal{L} avait une base déterministe, pour chaque langage de \mathcal{L} , il existerait un langage déterministe non générateur de \mathcal{L} le dominant et la même contradiction se produit.

III. PREUVE DU RÉSULTAT

Nous allons démontrer que Nge n'a pas de base déterministe. Commençons par exhiber le langage spécial A_0 de Alg qui vérifie les hypothèses du Fait 1.

I. Le langage A_0

Nous savons que le langage E engendré par la grammaire de règles $S \rightarrow aSbSc+d$ est un générateur des algébriques. E est également reconnu, par pile vide, par l'automate $\mathcal{A} = \langle \{a, b, c, d\}, \{S, b, c\}, S, t \rangle$ où t est défini par

$$t(d, S) = \varepsilon$$

$$t(a, S) = cSbS$$

$$t(b, b) = \varepsilon$$

$$t(c, c) = \varepsilon.$$

Rappelons sans démonstration, un lemme relatif aux facteurs gauches de E

$$\text{LEMME 1 : } FG(E) = (a + aEb)^*(\varepsilon + E).$$

Pour v préfixe d'un mot de E , on définit la *norme* de v , notée $\|v\|$, comme la longueur du plus petit mot v' de $\{a, b, c, d\}$ vérifiant vv' appartient à E . On démontre aisément que $\|v\|$ est égal à la hauteur de la pile de \mathcal{A} après lecture de v .

Ainsi $\|aad\| = 6$ car le plus petit mot v' tel que $aadv'$ appartienne à E est $bdc bdc$.

Pour tout mot f de $\{a, b, c, d\}^*$ définissons $e(f)$ comme le plus long préfixe de f qui soit un préfixe d'un mot de E . f s'écrit donc de manière unique $f = e(f)f'$ où f' est facteur droit de f , également déterminé de manière unique. On note ce facteur $q(f)$. Le mot f étant donné on trouve de façon déterministe $e(f)$ en essayant de reconnaître f par l'automate \mathcal{A} . La première lettre dont la lecture est impossible la première lettre de $q(f)$.

Avec ces notations on peut définir le langage particulier L_p , introduit par L. Boasson [5]

$$L_p = \{ f \in \{a, d, b, c\}^* / f = e(f)q(f) \text{ avec } \|e(f)\| \leq |q(f)| \leq 2\|e(f)\| \}.$$

On démontre que L_p est un langage algébrique non ambigu et non générateur vérifiant la condition IRS [10].

On définit le langage A_0 comme étant le miroir de L_p . Ainsi A_0 vérifie les propriétés de L_p qui se conservent par miroir. En particulier A_0 est un langage algébrique non ambigu et non générateur vérifiant la condition IRS. A_0 n'est donc pas générateur des algébriques ce qui démontre que A_0 vérifie l'hypothèse 1 du Fait 1.

2. A_0 vérifie l'hypothèse 2 du Fait 1

En vue de démontrer que A_0 vérifie l'hypothèse 2 (pour tout morphisme h et pour tout langage déterministe préfixe D l'égalité $h(D) = A_0$ entraîne que D est générateur de Alg) étudions tout d'abord certains mots de L_p .

Pour cela considérons une grammaire G de L_p et soit N_0 son entier d'Ogden. A tout mot h de E on associe l'entier $n(h)$ défini de la manière suivante : on considère pour une décomposition quelconque $h_1 h_2 h_3 h_4$ de h le mot $ah_1 h_2^2 h_3$ et on calcule $\|e(ah_1 h_2^2 h_3)\|$. On définit alors $A(h)$ comme étant le maximum des normes ainsi obtenues pour toutes les décompositions (en nombre fini) $h_1 h_2 h_3 h_4$ possibles de h . On pose alors $n(h) = \max \{ A(h), N_0 \}$. On définit enfin le mot $f(h)$ par

$$f(h) = ahba^{n(h)} c^{3n(h)+3}$$

Remarque :

$$e(f(h)) = ahba^{n(h)}, \quad q(f(h)) = c^{3n(h)+3}$$

d'où

$$\|e(f(h))\| = 3n(h) + 2$$

donc

$$\|e(f(h))\| \leq |q(f(h))| \leq 2\|e(f(h))\|$$

et le mot $f(h)$ appartient à L_p .

Remarque : Le choix de $n(h)$ garantira que

$$ah_1 h_2^2 h_3 ba^{n(h)+1} c^{3n(h)+3} \notin L_p$$

et

$$\|ah_1 h_2^2 h_3\| \leq n(h).$$

Établissons une proposition sur une paire itérante de $f(h)$:

PROPOSITION 1: Soit $h \in E$, G la grammaire de L_p considérée, alors $f(h)$ admet une paire itérante grammaticale $(ahba^{n_1}, a^{n_2}, \beta, v, c^{m_3})$ relativement à G , où v contient au moins un « c » i. e. :

$$(\beta = a^{n_3} c^{m_1}, v = c^{m_2} \text{ et } m_1, m_2 > 0)$$

ou

$$(\beta = a^{n_3}, v = a^{n_4} c^{m_1}, m_1 > 0).$$

Preuve : Soit h un mot de E , on pose $n = n(h)$. Marquons, dans $f(h)$, les N_0 « a » de gauche de a^n . On obtient, par le lemme d'itération d'Ogden, un facteur u d'une paire itérante grammaticale relativement à G . Cherchons où peut se trouver le second facteur v de la paire itérante :

(a) $v = a^l$ facteur de a^n . On doit donc avoir :

$$\forall k, f_k = ahba^{n+kl} c^{3n+3} \in L_p$$

ce qui est impossible car

$$\begin{aligned} \|e(f_k)\| &= 3n + 3kl + 2 \\ |q(f_k)| &= 3n + 3. \end{aligned}$$

Ce cas est donc exclu.

(b) $v = h_2$ facteur de h . On a $h = h_1 h_2 h_3$ et

$$\forall k, f_k = ah_1 h_2^k h_3 ba^{n_1} a^{kn_2} a^{n_3} c^{3n+3} \in L_p.$$

— S'il existe un itéré f_{k_0} tel que $ah_1 h_2^{k_0}$ ne soit pas un facteur gauche de E et donc que $e(ah_1 h_2^{k_0})$ soit un préfixe propre de $ah_1 h_2^{k_0}$ on a :

$$\forall k \geq k_0, e(ah_1 h_2^k) \text{ facteur gauche propre de } ah_1 h_2^{k_0}$$

d'où

$$\forall k \geq k_0, \|e(f_k)\| = \|e(f_{k_0})\|$$

alors que $|q(f_k)|$ croît avec k . Il existera donc un itéré f_k vérifiant $|q(f_k)| > 2\|e(f_k)\|$ donc n'appartenant pas à L_p .

— Sinon on a donc en particulier $e(ah_1 h_2^2) = ah_1 h_2^2$.

- Si $e(ah_1 h_2^2 h_3) = ah_1 h_2^2 h_3$ on a

$$e(f_2) = ah_1 h_2^2 h_3 ba^{n_1} a^{2n_2} a^{n_3}$$

d'où $\|e(f_2)\| \geq 1 + 3(n + n_2) \geq 3n + 4$ donc $\|e(f_2)\| > |q(f_2)| = 3n + 3$ ce qui contredit $f_2 \in L_p$.

- Si $e(ah_1 h_2^2 h_3) = ah_1 h_2^2 h'_3$ avec $h'_3 \neq h_3$, $h_3 = h'_3 h''_3$ on a

$$e(f_2) = ah_1 h_2^2 h'_3$$

et donc, par définition de $n(h)$, $\|e(f_2)\| \leq n$.

D'autre part $q(f_2) = h'_3 ba^n c^{3n+3}$ d'où $|q(f_2)| \geq 4n + 5$ donc $|q(f_2)| > 2\|e(f_2)\|$ ce qui contredit $f_2 \in L_p$.

Ce cas est donc également impossible.

(c) $v = ah_1$ avec $h_1 \in FG(hba^n)$. On a alors

– ou bien $\exists k_0$ tel que $(ah_1)^{k_0} \notin FG(E)$ et alors $\forall k \geq k_0 e(f_k) = e(f_{k_0})$ et lorsque k augmente $\|e(f_k)\|$ reste constant mais $|q(f_k)|$ croît donc il existe un itéré k vérifiant $f_k \notin L_p$.

– ou bien

$$\forall k \quad e(f_k) = (ah_1)^{k+1} h_2 a^{n+kn_2}$$

et lorsque k augmente $|q(f_k)|$ reste constant mais $\|e(f_k)\|$ croît donc il existe un itéré k vérifiant $f_k \notin L_p$.

Ce cas est donc impossible.

(d) $v = h_2 ba^{n_1}$ avec h_2 facteur droit de h . Posons $h = h_1 h_2$.

On a alors, comme pour le cas (c),

ou bien

$$\exists k_0 \quad \text{tel que } ah_1 (h_2 ba^{n_1})^{k_0} \notin FG(E)$$

ou bien

$$\forall k \quad e(f_k) = ah_1 (h_2 ba^{n_1})^{k+1} a^{n+kn_2}$$

dans les deux cas on montre, de façon identique à celle employée au (c), qu'il existe un itéré f_k n'appartenant pas à L_p .

Ce cas est également impossible.

Finalement le facteur itérant v ne peut être qu'à droite de u et comme v ne peut être contenu dans le « a », en vertu du cas (a), on a $v = a^{n_4} c^{m_2}$ avec $n_4 \geq 0$, $m_2 > 0$ ce qui démontre la proposition 1. ■

Prouvons maintenant que A_0 vérifie l'hypothèse 2 du fait 1 :

Pour tout langage déterministe préfixe D et pour tout morphisme alphabétique h , l'égalité $h(D) = A_0$ entraîne que D est générateur des algébriques.

Notons d'abord que l'on peut se restreindre au cas où h est strictement alphabétique puisque A_0 est IRS [3].

Soit D un langage déterministe préfixe sur un alphabet Z et h un morphisme strictement alphabétique tel que $h(D) = A_0$.

Soit \mathcal{A} un automate à pile déterministe reconnaissant par pile vide et états finals le langage déterministe préfixe D et G la grammaire associée à \mathcal{A} (par la construction des triplets). Soit B l'ensemble des mots w de Z'' qui vérifient :

- $w = \alpha u \beta v$ avec $(\alpha, u, \beta, v, \varepsilon)$ paire itérante automatique;
- $(\alpha, u, \beta, v, \varepsilon)$ est la seule paire itérante automatique de w .

On sait (propriété 2) que B est un ensemble rationnel.

Posons

$$R = B \cap h^{-1}(c)^+ h^{-1}(a)^+ \\ L = D \cap R h^{-1}(a)^* h^{-1}(b) Z^* h^{-1}(a).$$

LEMME 2 : $\forall v \in E \exists g \in L / h(g) = f(\tilde{v}) = c^{3n(v)+3} a^{n(v)} b \tilde{v} a$.

Preuve : En effet puisque $c^{3n(v)+3} a^{n(v)} b \tilde{v} a$ appartient à A_0 il existe g dans D tel que $h(g) = f(\tilde{v})$. D'après la proposition 1, $f(v)$ admet relativement à $h(G)$ (qui est une grammaire de A_0) une paire itérante grammaticale $(c^{m^3}, w, \beta, a^{n^2}, a^{n^1} b \tilde{v} a)$ où w contient au moins un « c ». Ceci entraîne l'existence dans g d'une paire itérante grammaticale relativement à G (autrement dit une paire itérante automatique relativement à \mathcal{A}) $(\alpha', u', \beta', v', v')$ avec $\alpha' u' \beta' v' \in h^{-1}(c)^+ h^{-1}(a)^*$. Il est alors clair que g a un préfixe dans R et donc que g appartient à L . Le lemme 2 est ainsi démontré.

LEMME 3 : $\forall g \in L, h(g) \in c^* a^* b \widetilde{FG}(E) a$.

Preuve : Soit $g \in L = D \cap R h^{-1}(a)^* h^{-1}(b) Z^* h^{-1}(a)$, on a

$$h(g) = c^n a^m b f a.$$

De plus \mathcal{A} étant déterministe, $g a$, par définition de R , une paire itérante dans A^0 de la forme (α, u, β, v, v) avec $h(\alpha u \beta v) \in c^+ a^+$. Ceci entraîne que $h(g)$ a une paire itérante, dans A_0 , de la forme $(\alpha_1, u_1, \beta_1, v_1, v_1)$ avec $\alpha_1 u_1 \beta_1 v_1 \in c^+ a^+$.

Il est clair que si $u_1 \beta_1 v_1$ est une puissance de a (ou de c) on obtiendra en itérant des mots qui ne seront plus dans A_0 . $(\alpha_1, u_1, \beta_1, v_1, v_1)$ est donc de

l'une des trois formes suivantes :

$$(c^{n_1}, c^{n_2}, c^{n_3}, a^{m_1}, a^{m_2}, a^{m_3} bfa)$$

ou

$$(c^{n_1}, c^{n_2}, c^{n_3}, c^{n_4} a^{m_1}, a^{m_2} bfa)$$

ou

$$(c^{n_1}, c^{n_2} a^{m_1}, a^{m_2}, a^{m_4} bfa).$$

Supposons que $\tilde{f} \notin FG(E)$, on obtient en itérant cette paire des mots w_k de L_p de la forme

$$w_k = a\tilde{f}ba^{m+kl}c^{n+kl}.$$

L'hypothèse $\tilde{f} \notin FG(E)$ entraîne pour tout k : $e(w_k)$ est facteur gauche de af ce qui est incompatible avec l'appartenance de w_k à L_p pour tout k .

Finalement on a $f \in \widetilde{FG}(E)$ et le lemme 3 est démontré.

Les lemmes 2 et 3 nous donnent la double inclusion :

$$\{c^{3n(v)+3}a^{n(v)}b\tilde{v}a/v \in E\} \subseteq h(L) \subseteq c^*a^*b\widetilde{FG}(E)a.$$

Si on applique la transduction τ qui efface les préfixes dans c^*a^*b et le dernier « a » on obtient :

$$\tilde{E} \subseteq \tau(h(L)) \subseteq \widetilde{FG}(E)$$

et donc

$$E \subseteq \tilde{\tau}(h(\tilde{L})) \subseteq FG(E).$$

Pour conclure il nous reste à démontrer le résultat suivant :

PROPOSITION 2: *Si un langage algébrique A vérifie $E \subseteq A \subseteq FG(E)$ il est générateur des algébriques.*

Preuve : Considérons le code : $\psi(a_1) = aa$, $\psi(a_2) = adba$, $\psi(b) = b$, $\psi(c_2) = cc$, $\psi(c_1) = cbdc$, $\psi(d) = d$ introduit par J. Berstel [3]. Ce code vérifie $\psi^{-1}(E) = E_2$. Écrivons $FG(E) = \varepsilon + E + (a + aEb)^*a(\varepsilon + E + EbE + Eb)$ et appliquons ψ^{-1} à la double inclusion $E \subseteq A \subseteq FG(E)$ on obtient :

$$E_2 \subseteq \psi^{-1}(A) \subseteq E_2 + (a_1 + a_2)^+(\varepsilon + E_2 + E_2bE_2 + E_2b)$$

d'où

$$a_1 E_1 b E_1 c_2 \subseteq \psi^{-1}(A) \cap a_2 \{a_1, d, c_1, b\}^* c_2 \subseteq a_2 E_1 b E_1 c_2$$

et de là $\psi^{-1}(A) \cap a_2 \{a_1, c_1, d, b\}^* c_2$ est un générateur de Alg donc A également ce qui démontre la proposition.

D'après cette proposition $\tau(h(L))$ est générateur de Alg donc $(h(L))$ est générateur de Alg et ainsi L puis D sont générateurs de Alg.

Nous avons par conséquent prouvé que A_0 vérifie l'hypothèse 2 du Fait 1. Il s'en suit

THÉORÈME : *Nge n'a pas de base déterministe.*

IV. CONCLUSIONS

Nous ne donnerons, en conclusion, qu'une liste (non exhaustive) d'extensions possibles.

Notre résultat est-il encore vrai lorsqu'on remplace le déterminisme par la non-ambiguïté? Plus précisément peut-on démontrer la propriété suivante :

Nge n'a pas de base non ambiguïté.

Si l'on veut appliquer la méthode utilisée dans ce travail, ce problème semble être le plus difficile pour au moins deux raisons. Il faudra exhiber des langages L_0 inhéremment ambigus, et même plus, des langages qui, si l'on reprend la terminologie de [1], sont intrinsèquement ambigus (very strong en anglais), i. e. tel que tous les langages rationnellement équivalents sont inhéremment ambigus. Or, en pratique, il est souvent très délicat de démontrer qu'un langage donné vérifie ces propriétés. La seconde raison est que nous utilisons, pour Nge, de façon fondamentale, le fait que l'ensemble des mots lus par un automate à pile déterministe sans paire itérante automatique est un langage rationnel. Et, comme nous l'avons rappelé, il n'existe pas de propriété analogue lorsque l'on remplace l'hypothèse « déterministe » par celle de « non ambigu ». Il semble donc nécessaire d'utiliser, si ce n'est une autre méthode, du moins un autre argument si l'on veut prouver des résultats sur la non-ambiguïté.

Le même principe de preuve que celui utilisé ici permet de démontrer que le sous-cône NLin des langages non générateurs de Lin (cône des langages linéaires) n'a pas de base déterminante [11]. La question se pose alors de savoir si le même résultat vaut pour :

- (a) Qrt(k) la famille des langages quasi-rationnels d'ordre k ;
- (b) Fcl(k) la famille des langages à compteur itéré d'ordre k ;

(c) Gre(k) la famille des langages de Greibach d'ordre k .

Nous pensons que la méthode utilisée dans ce travail devrait permettre de répondre à certaines de ces questions. Par exemple des langages de la forme $L_0 \uparrow S_2$ ou $S_2 \uparrow L_2$ permettent peut-être de démontrer que NQrt(2) n'a pas de base déterministe. La difficulté vient du fait qu'un langage qui domine $L_0 \uparrow S_2$ ne s'écrit pas toujours $L \uparrow L'$ et donc on ne peut utiliser facilement le résultat de NLin.

BIBLIOGRAPHIE

1. J. BEAUQUIER, *Strong Non Deterministic Context-Free Languages*, 4th. G.I. Conference. L.N. in C.S. N° 55, 1929, p. 4.
2. J. BEAUQUIER, *Générateurs algébriques et systèmes de paires itérantes*, T.C.S., vol. 8, 1979, p. 293-323.
3. J. BERSTEL, *Transductions and Context-Free Languages*, Teubner, 1979.
4. L. BOASSON, *The Inclusion of the Substitution Closure of Linear and One-Counter Languages in the Largest Sub AFL of the Family of CFL's is Proper*, I.P.L., vol. 2, 1973, p. 135-140.
5. L. BOASSON, *Un langage algébrique particulier*, R.A.I.R.O., Informatique théorique, vol. 13, 1979, p. 203-215.
6. L. BOASSON, *Non-générateurs algébriques et substitution*, R.A.I.R.O., Informatique théorique, vol. 19, 1985, p. 125--136.
7. L. BOASSON et M. NIVAT, *Parenthesis Generators*, 17th I.E.E.E. Symposium on Foundations of Computer Science, Houston, 1976, p. 253-257.
8. C. FROUGNY, *Langages très simples générateurs*, R.A.I.R.O., Informatique théorique, vol. 13, 1979, p. 68-86.
9. S. GREIBACH, *Chains of Full AFL's*, math. Syst. Theory, vol. 4, 1970, p. 231-242.
10. S. GREIBACH, *One Counter Languages and the IRS Condition*, J.C.S.S., vol. 10, 1975, p. 237-247.
11. A. PETIT, *Déterminisme et non-générateurs*, Thèse de 3^e cycle, Université Paris-VII, 1985.