

FRANÇOIS RODRIGUEZ

Une double hiérarchie infinie de langages vérifiables

Revue française d'automatique, informatique, recherche opérationnelle. Informatique théorique, tome 9, n° R1 (1975), p. 5-19.

http://www.numdam.org/item?id=ITA_1975__9_1_5_0

© AFCET, 1975, tous droits réservés.

L'accès aux archives de la revue « Revue française d'automatique, informatique, recherche opérationnelle. Informatique théorique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). 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 DOUBLE HIERARCHIE INFINIE DE LANGAGES VERIFIABLES

par François RODRIGUEZ ⁽¹⁾

Communiqué par J.-F. PERROT

Résumé. — *En transposant aux accepteurs vérificateurs les notions d'automate à compteur et d'automates à pile à nombre fini de mouvements, il est possible de définir une double hiérarchie infinie croissante de langages vérifiables telle que toute sous-famille $C_{m,n}$ forme un cône rationnel principal.*

1. INTRODUCTION

Les accepteurs vérificateurs (« checking automata ») ont été introduits comme types particuliers d'accepteurs à mémoire unilatères (« one-way stack automata ») [3] [4]. La classe C des langages reconnus par ces accepteurs, ou « langages vérifiables » (« checking automata languages ») a été étudiée par S. Greibach [6]; Harrison et Schkolnick en ont donné une caractérisation en termes de grammaires indexées [7]. La sous-classe des langages vérifiables avec un nombre borné de mouvements sur la bande-mémoire (« finite-turn checking automata languages ») a fait également l'objet d'une caractérisation grammaticale par Siromoney [13].

La classe des langages vérifiables présente avec la classe des langages algébriques (« context-free »), reconnaissables par les automates à pile, certaines analogies structurelles : notamment, toutes deux forment des familles agréables de langages (en anglais « full AFL », cf. [1]) principales et fermées par substitution. Ces analogies nous ont conduit à transposer à la classe C la démarche de S. Greibach dans [5] pour les automates à pile, c'est-à-dire à étudier les sous-classes de langages vérifiables $C_{m,n}$ reconnues par les accepteurs

(1) E.N.S.E.E.I.H.T., Toulouse.

vérificateurs effectuant au plus un nombre fini m de mouvements en mémoire et dont tout mot-mémoire appartient à un langage borné de dimension n .

On montre que chaque sous-classe $C_{m,n}$ constitue un cône rationnel (cf. [1]) principal, i.e. qu'il existe un langage $L_{m,n}$ tel que $C_{m,n}$ soit exactement l'ensemble des images de $L_{m,n}$ par des transductions rationnelles. A ces cônes $C_{m,n}$ répondent, dans le domaine algébrique, les cônes des langages ultralinéaires de rang m et bornés de dimension n (cf. [2]), qui ont pour clôtures rationnelles les familles agréables de langage $\mathcal{F}_{m,n}$ étudiées en [5]. On trouvera en [11] une analyse détaillée de cette correspondance, dont les principaux résultats ont été résumés en [12].

Nous nous limitons, dans le présent article, à la démonstration que les cônes $C_{m,n}$ forment une hiérarchie croissante, i.e. que l'inclusion $C_{m,n} \subseteq C_{m',n'}$ a lieu si et seulement si $m \leq m'$ et $n \leq n'$ (théorème 3).

Une première partie est consacrée aux définitions (complexes mais nécessaires) relatives aux accepteurs vérificateurs qui permettent de définir précisément les familles $C_{m,n}$; nous donnons les langages $L_{m,n}$ générateurs des cônes $C_{m,n}$ sans démonstration. La seconde partie contient une preuve complète du théorème de hiérarchie, composée d'une suite de lemmes techniques dont le premier est un lemme d'itération avec positions marquées inspiré des méthodes d'Ogden [9].

Dans la suite, pour un ensemble quelconque E , E^* désignera le monoïde libre engendré par E , formé des suites finies d'éléments de E ou mots sur E ; la longueur d'un mot $w \in E^*$ sera notée $|w|$; le mot vide sera noté ε . \mathbb{N} désignera l'ensemble des entiers positifs ou nuls. En reprenant une notation de M. Nivat [8], on désignera par $\mathfrak{L} \cdot \Gamma$ le cône rationnel engendré par une famille de langage \mathfrak{L} .

Le symbole ■ indique une fin de démonstration.

2. LES ACCEPTEURS VERIFICATEURS

Définition 1.

Un accepteur vérificateur (abr. AV) est un 6-uplet $D = (K, X, G_D, \delta, q_0, F)$ tel que :

- 1) K, G_D sont des ensembles finis d'états et de symboles de mémoire. X est un alphabet d'entrée fini.
- 2) $q_0 \in K$ est l'état initial, $F \subseteq K$ est l'ensemble des états terminaux.
- 3) δ est une application de $K \times (X \cup \{ \varepsilon \}) \times (G_D \cup \{ \varepsilon \})$ dans $K \times (G_D \cup \{ 0, 1, -1, \varepsilon, E \})$.

4) Il existe un sous-ensemble K_E de K tel que $q_0 \in K_E$ et que, si :

$$(q', \gamma) \in \delta(q, a, Z) \text{ avec } q \in K_E, a \in X \cup \{ \varepsilon \},$$

$Z \in G_D \cup \{ \varepsilon \}, q' \in K - K_E, \gamma \in (G_D \cup \{ 0, 1, -1, \varepsilon \})$, alors :

$$\delta(q', a, Z) \cap K \times G_D = \emptyset, \quad \forall a \in X \cup \{ \varepsilon \}, \quad \forall Z \in G_D \cup \{ \varepsilon \}.$$

5) Si $(q', E) \in \delta(q, a, A)$ alors :

$$\delta(q', a, \varepsilon) = \emptyset, \quad \forall a \in X \cup \{ \varepsilon \}.$$

Définition 2.

Soit $D = (K, X, G_D, \delta, q_0, F)$ un AV. Une *configuration* de D est un triplet soit de la forme (q, w, ε) soit de la forme $(q, w, (y, i))$ avec $q \in K, w \in X^*, y \in G_D^*$ et $1 \leq i \leq |y|$, où $|y|$ désigne la longueur du mot y .

Nous noterons \vdash la relation entre configurations, définie comme suit :

• Si (q, aw, ε) , avec $a \in X \cup \{ \varepsilon \}$ et $w \in X^*$, est une configuration et si $(q', \varepsilon) \in \delta(q, a, \varepsilon)$ alors :

$$(q, aw, \varepsilon) \vdash (q', w, \varepsilon),$$

• Si (q, aw, ε) est une configuration et si $(q', A) \in \delta(q, a, \varepsilon), A \in G_D$ alors :

$$(q, aw, \varepsilon) \vdash (q', w, (A, 1)),$$

• Si $(q, aw, (yA, 1))$ est une configuration, $y \in G_D^*, A \in G_D$, et si

$$(q', A') \in \delta(q, a, A), A' \in G_D \cup \{ \varepsilon \}$$

alors :

$$(q, aw, (yA, 1)) \vdash (q', w, (yAA', 1)),$$

• Si $(q, aw, (A_n \dots A_1, i))$ est une configuration, $A_i \in G_D, 1 \leq i \leq n$, et si $(q', j) \in \delta(q, a, A_i), j \in \{ 0, 1, -1 \}$ alors :

$$(q, aw, (A_n \dots A_1, i)) \vdash (q', w, (A_n \dots A_1, i + j)) \text{ si } 1 \leq i + j \leq n.$$

• Si $(q, aw, (y, i))$ est une configuration et si $(q', E) \in \delta(q, a, A_i)$ alors :

$$(q, aw, (y, i)) \vdash (q', w, \varepsilon).$$

Nous noterons \vdash^* la fermeture transitive de la relation \vdash .

Intuitivement un AV est un automate qui peut écrire un mot y dans sa mémoire mais (point 4) de la définition 1) qui, sitôt qu'il cesse d'écrire, n'a plus le droit de modifier le contenu de la mémoire.

Les instructions $+1, -1$ et 0 permettent d'examiner le contenu du mot mémoire.

L'instruction E provoque l'effacement du mot mémoire et bloque l'AV (point 5) de la définition 1).

Définition 3.

Nous appellerons langage accepté par un AV $D = (K, X, G_D, \delta, q_0, F)$ l'ensemble suivant, noté $L(D)$:

$$L(D) = \{ w \in X^* \mid \exists q_F \in F, (q_0, w, \varepsilon) \vdash^* (q_F, \varepsilon, \varepsilon) \}.$$

Tout langage L tel qu'il existe un AV D tel que $L = L(D)$ sera appelé *langage vérifiable*.

Nous noterons \mathcal{C} l'ensemble des langages vérifiables.

EXEMPLE : $L = \{ a^n b^n c^n \mid n \geq 1 \}$ est un langage vérifiable [6]. Par contre, on montre que $L = \{ aba^2 b^2 \dots a^n b^n \mid n \in \mathbb{N} \}$ n'est pas vérifiable [11].

Nous pouvons énoncer un résultat fondamental de S. Greibach :

Théorème 1 [6] : \mathcal{C} forme une famille agréable de langages principale, fermée par substitution.

Afin de construire la hiérarchie annoncée, définissons les évolutions possibles des AV dans leur mémoire.

Définition 4.

Soit $D = (K, X, G_D, \delta, q_0, F)$ un AV et soit la suite de configurations :

$$(q_0, w, \varepsilon) \vdash^* (q_i, w_i, (y, i)) \vdash^* (q_j, w_j, (y, j)) \vdash^* \\ (q_k, aw_k, (y, k)) \vdash (q, w_k, (y, l)) \vdash^* \dots$$

Si $k \geq j > i$ et $l < k$ nous dirons que D passe par un *pic* lors de l'analyse de « a », i.e. D passe de l'incréméntation à la décréméntation de son index de mémoire.

Si $k \leq j < i$ et $l > k$ nous dirons que D passe par un *col* lors de l'analyse de « a », i.e. D passe de la décréméntation à l'incréméntation de son index de mémoire.

Nous dirons que D fait un *mouvement* soit s'il passe par un pic soit si $(q, aw, (y, i)) \vdash (q, w, \varepsilon)$ pour $i > 1$.

Notations

Pour tout $n \in \mathbb{N}$ soit H_n l'ensemble :

- si $n = 2r$ alors $H_n = (a^* b^* \gamma)$
- si $n = 2 + 1$ alors $H_n = (a^* b^* \gamma a^*)$.

Soit enfin ω un nouveau symbole n'appartenant pas à \mathbb{N} : H_ω désignera l'ensemble $(a, b)^*$.

Définition 5.

Un AV $D = (K, X, G_D, \delta, q_0, F)$ sera dit strictement (m, n) borné si pour tout mot $w \in L(D)$, il existe une suite de configurations

$$(q_0, w, \varepsilon) \vdash^* (q', w', (y, i)) \vdash^* (q_F, \varepsilon, \varepsilon)$$

telle que :

- 1) D fait au plus m mouvements, $m \in \mathbf{N}$, sinon nous poserons $m = \omega$.
- 2) Les mots mémoires y appartiennent à H_n .

Pour tous m et n dans $\mathbf{N} \cup \{ \omega \}$, soit :

$$C_{m,n} = \{ L \mid \exists \text{ un AV } D, \text{ strictement } (m, n)\text{-borné, } L = L(D) \}.$$

Notons que $C_{0,n} = C_{m,0} = \text{Rat}(X^*)$, ensemble des langages rationnels sur X et que $C_{\omega,\omega} = \mathcal{C}$.

REMARQUE 1

Étant donné un AV $D = (K, X, G_D, \delta, q_0, F)$ strictement (m, n) -borné, on peut construire sans difficulté un AV $D' = (K' \times \{ 0, 1, \dots, m \}, X, G_{D'}, \delta', (q_0, 0), F \times \{ 0, 1, \dots, m \})$ strictement (m, n) -borné tel que $L(D) = L(D')$ et que, si :

$$((q_0, 0), w', w'', \varepsilon) \vdash^* ((q, i), w'', (y, j))$$

alors D' a fait exactement i mouvements lors de l'analyse du préfixe w' .

Cette propriété sera utilisée dans la suite de ce travail.

Définition 6.

Soit $X = \{ a, b, \bar{a}, \bar{b} \}$ et ρ la congruence de X^* engendrée par les relations $a\bar{a} = \varepsilon$ et $b\bar{b} = \varepsilon$; on notera $\rho(w)$ la classe de $w \in X^*$. La classe $\rho(\varepsilon)$ est bien connue sous l'appellation de langage de Dyck restreint sur deux lettres, noté D_2^* : ce langage est générateur du cône rationnel des langages algébriques [1].

Désignons par v l'application de X^* dans \mathbf{N} définie par

$$v(w) = \max \{ k; w = w_1 x_1 \bar{x}_1 w_2 x_2 \bar{x}_2 \dots w_k x_k \bar{x}_k w_{k+1}, x_i \in \{ a, b \}, \\ 1 \leq i \leq k, w_j \in X^*, 1 \leq j \leq k + 1 \}.$$

Notons, pour $w \in X^*$, $\text{Init}(w)$ l'ensemble des préfixes de w , i.e. $\text{Init}(w) = \{ u \in X^*; w \in uX^* \}$, et appelons $L_{m,n}$, avec $m, n \in \mathbf{N} \cup \{ \omega \}$ le langage sur X ainsi défini :

$$L_{m,n} = \{ wy; w \in H_n, y \in D_2^*, v(y) \leq m, \rho(\text{Init}(y)) \subseteq \text{Init}(w) \}$$

On trouvera en [11] une démonstration du résultat suivant, fondamental pour notre étude :

Théorème 2 : Pour tous m et n dans $\mathbf{N} \cup \{ \omega \}$ la famille de langages $\mathcal{C}_{m,n}$ forme un cône rationnel principal engendré par $L_{m,n}$; i.e. on a $\mathcal{C}_{m,n} = \{ L_{m,n} \} \Gamma$.

Notation. Nous désignerons par $\mathcal{C}_{\infty,n}$, $\mathcal{C}_{m,\infty}$ et $\mathcal{C}_{\infty,\infty}$ les cônes rationnels suivants :

$$\mathcal{C}_{\infty,n} = \bigcup_{m \in \mathbf{N}} \mathcal{C}_{m,n} \quad ; \quad \mathcal{C}_{m,\infty} = \bigcup_{n \in \mathbf{N}} \mathcal{C}_{m,n} \quad ; \quad \mathcal{C}_{\infty,\infty} = \bigcup_{m,n \in \mathbf{N}} \mathcal{C}_{m,n}$$

Définition 7.

Soit \leq la relation d'ordre naturelle sur \mathbf{N} prolongée sur $\mathbf{N} \cup \{ \omega, \infty \}$ par $n < \infty < \omega, \forall n \in \mathbf{N}$.

Pour tous m, n, m', n' dans $\mathbf{N} \cup \{ \omega, \infty \}$ nous dirons que $(m, n) \leq (m', n')$ si et seulement si $m \leq m'$ et $n \leq n'$.

3. THEOREMES DE LA HIERARCHIE

Théorème 3 :

Pour tous m, n, m', n' dans $\mathbf{N} \cup \{ \omega, \infty \}$, nous avons $\mathcal{C}_{m,n} \subseteq \mathcal{C}_{m',n'}$ si et seulement si $(m, n) \leq (m', n')$.

Ce théorème découle de la proposition suivante :

Proposition 1 :

Pour tous m, n, m', n' dans $\mathbf{N} \cup \{ \omega \}$, le langage $L_{m,n}$ appartient à $\mathcal{C}_{m',n'}$ si et seulement si $(m, n) \leq (m', n')$.

Avant de démontrer cette proposition établissons quatre lemmes qui nécessitent eux-mêmes l'introduction de nouvelles définitions :

Définition 8 :

Soit $D = (K, X, G_D, \delta, q_0, F)$ un AV.

Soit $C_0, C_1, \dots, C_p, \dots, C_r$ une suite de configurations de D telle que :

$C_0 = (q_0, w_0, \varepsilon)$, $w_0 = w$, $C_i = (q_i, w_i, (y_i, 1))$ pour tout i , $1 \leq i \leq p$ avec $|y_i| > |y_{i-1}|$ et $y_p = y$.

$$C_j = (q_j, w_j, (y, i_j)) \quad , \quad p \leq j < r$$

$$C_r = (q_r, \varepsilon, \varepsilon) \quad , \quad q_r \in F,$$

et pour tout l , $0 \leq l \leq r$, $C_l \vdash C_{l+1}$.

On dira que $(C_0, C_1, \dots, C_p, \dots, C_r)$ est une *analyse* du mot w (par le mot y) de longueur (p, r) .

Intuitivement la suite de configurations C_0, \dots, C_p correspond à l'écriture par D d'un mot mémoire y de longueur p , tandis que C_{p+1}, \dots, C_r sont les configurations suivantes de l'analyse.

Une analyse $(C_0, C_1, \dots, C_p, \dots, C_r)$ de w par un mot y sera dite de *longueur minimale* si pour toute autre analyse de w , $(C'_0, \dots, C'_{p'}, \dots, C'_{r'})$ on a $p' \geq p$ et $r' \geq r$.

Définition 9 :

Soit $D = (K, X, G_D, \delta, q_0, F)$ un AV et $((C_0, C_1, \dots, C_p, \dots, C_r)$ une analyse d'un mot w de $L(D)$ par un mot $y = A_s \dots A_1, A_i \in G_D, 1 \leq i \leq s$ et

$$\begin{aligned} C_j &= (q_j, w_j, (y_j, 1)) \quad , \quad 1 \leq j \leq p. \\ C_l &= (q_l, w_l, (y, i_l)) \quad , \quad p \leq l \leq r. \end{aligned}$$

Pour tout $k, 1 \leq k \leq s$, nous appellerons *séquence de coupe de niveau k* et de longueur l , le l -uplet $\alpha_k = (q_{k_1}, \dots, q_{k_l}), 1 \leq k_1 < k_2 < \dots < k_l \leq r$, formé de la séquence des états q_j tels que $i_j = k$ et soit $i_j - 1 = k - 1$, soit $i_j + 1 = k - 1$. α_k est donc l'ensemble des états successifs de D lorsqu'il examine le symbole A_k dans sa mémoire.

Nous allons établir un lemme d'itération pour les familles $C_{m,\omega}$, utilisant les notions de positions distinguées développées par Ogden [9] :

Lemme 1 :

Pour tout AV D , strictement (m, ω) -borné, il existe un entier k_D tel que pour tout mot $w \in L(D)$, tel que w contient k_D ou plus positions distinguées, il existe un entier $r, 1 \leq r \leq 2m + 1$, des mots $x_1, x_2, \dots, x_r, y_0, y_1, \dots, y_r$ de X^* tels que :

- 1) $w = y_0 x_1 y_1 x_2 y_2 \dots x_r y_r$.
- 2) Pour tout $k \geq 1$

$$w_k = y_0 x_1^k y_1 x_2^k y_2 \dots x_r^k y_r \in L(D).$$

- 3) Si $w_0 = y_0 y_1 \dots y_r$ n'est pas le mot vide, alors $w_0 \in L(D)$.
- 4) Si $r > 1$, l'un des x_i au moins contient des positions distinguées.

5) Lors de l'analyse de tout $x_i, 1 \leq i \leq r, D$ ne passe ni par un col, ni par un pic. Lors de l'analyse de tout $y_i, 0 < i < r, D$ passe soit par un col, soit par un pic.

Démonstration :

Soit $D = (K, X, G_D, \delta, q_0, F)$ un AV.

Soient $\mathfrak{g} = \text{card} \{ G_D \}$ et $\mathfrak{K} = \text{card} \{ K \}$.

Soit $k_D = (\mathfrak{g} + 1)(\mathfrak{K} + 1)^{(\mathfrak{K}+1)}$.

Soit w un mot de $L(D)$ tel que w contienne k_D ou plus positions distinguées. Par définition il existe une analyse de w par un mot y , analyse de longueur minimale.

Dans le cas où $L(D)$ est rationnel ($m = 0$) on sait qu'il existe une factorisation de w , $w = y_0 x_1 y_1$ telle que $y_0 x_1^* y_1 L(D)$.

Considérons le cas non trivial où $m \neq 0$.

A) Supposons tout d'abord qu'il existe deux entiers $j \neq j'$ tels que $q_j = q_{j'}$ et $i_j = i_{j'}$.

On a donc une suite de configurations :

$$(q_0, w, \varepsilon) \stackrel{*}{\vdash} (q_j, w_j, (y, i_j)) \stackrel{*}{\vdash} (q_{j'}, w_{j'}, (y, i_{j'})) \stackrel{*}{\vdash} (q_F, \varepsilon, \varepsilon)$$

$$q_F \in F$$

telle que si

$$w_j = x_1 w_{j'} = x_1 y_1$$

et

$$w = y_0 w_j$$

alors, pour tout $n \geq 1$

$$(q_0, y_0 x_1^n y_1, \varepsilon) \stackrel{*}{\vdash} (q_j, x_1^n y_1, (y, i_j)) \stackrel{*}{\vdash} (q_{j'}, x_1^{n-1} y_1, (y, i_{j'}))$$

$$\stackrel{*}{\vdash} (q_j, y_1, (y, i_j)) \stackrel{*}{\vdash} (q_F, \varepsilon, \varepsilon)$$

i.e. : D accepte le mot $w_n = y_0 x_1^n y_1$ pour tout $n \geq 1$.

De plus $q_j = q_{j'}$ implique que D ne passe pas par un pic lors de l'analyse de x_1 . (Remarque 1.)

Enfin si D passait par au moins un col lors de l'analyse de x_1 , D passerait par au moins n cols donc $n - 1$ pics lors de l'analyse du mot $x_n \in L(D)$ et ceci pour tout $n \in \mathbb{N}$. D ne serait donc pas (m, ω) -borné.

Dans ce cas les parties 2, 4 et 5 du lemme sont donc vérifiées. Le point 3 résulte immédiatement du fait qu'il existe une analyse de $w_0 = y_0 y_1$ par le mot $y = Z_n \dots Z_k Z_{k-1} \dots Z_1$.

B) Nous pouvons donc supposer maintenant que les longueurs des séquences de coupe sont inférieures ou égales à \mathfrak{K} . Il existe donc au plus $\mathfrak{K}^{\mathfrak{K}}$ séquences de coupe distinctes. Si w contient au moins k_D positions distinguées, nous avons $|w| \geq k_D$ et donc $r \geq k_D$. D doit donc examiner au moins k_D / \mathfrak{K} positions dans la mémoire.

Or $k_D/\mathcal{K} > (\mathcal{G} + 1)\mathcal{K}^{\mathcal{K}}$. Il existe donc nécessairement deux entiers k et \bar{k} tels que $A_k = A_{\bar{k}}$ et $\alpha_k = \alpha_{\bar{k}}$ soit, plus précisément (en utilisant la remarque 1) :

$$\alpha_k = ((\bar{q}_{k_1}, l_1), (\bar{q}_{k_2}, l_2), \dots, (\bar{q}_{k_p}, l_p)) = \alpha_{\bar{k}} = ((\bar{q}_{\bar{k}}, l_1), \dots, (\bar{q}_{\bar{k}}, l_p))$$

avec

$$q_{k_i} = (\bar{q}_{k_i}, l_i) = q_{k_i} = (\bar{q}_{\bar{k}}, l_i), \quad 1 \leq i \leq p$$

Il existe donc des mots $y_0, y_1, \dots, y_r, x_1, x_2, \dots, x_r$ de X^* tels que :

$w = y_0 x_1 y_1 \dots x_r y_r$ et des mots Z_1, Z_2, Z_3 dans G_D^* tels que

$y = Z_1 Z_2 Z_3, Z_1 = A_s \dots A_{\bar{k}}, Z_2 = A_{\bar{k}-1} \dots A_k$ et $Z_3 = A_{k-1} \dots A_1$

$(q_0, y_0 x_1 y_1 \dots x_r y_r, \varepsilon) \vdash^* (q_{\bar{k}_1}, x_1 y_1 \dots x_r y_r, (Z_1, 1)) \vdash^*$

$(q_{k_1}, y_1 \dots x_r y_r, (Z_1 Z_2, 1)) \vdash^* (q_{k_2}, x_2 \dots x_r y_r, (y, i_{k_2})) \vdash^*$

$(q_{\bar{k}_2}, y_2 \dots x_r y_r, (y, i_{\bar{k}_2})) \vdash^* \dots \vdash^* (q_{k'_r}, x_r y_r, (y, i_{k'_r}))$

$\vdash^* (q_{k''_r}, y_r, (y, i_{k''_r})) \vdash^* (q_F, \varepsilon, \varepsilon),$

avec si

$$k'_r = k_r \quad \text{alors} \quad k''_r = \bar{k}_r$$

et si

$$k'_r = \bar{k}_r \quad \text{alors} \quad k''_r = k_r$$

Remarquons que $q_{k_i} = q_{\bar{k}_i}$ implique que D ne passe pas par un pic lors de l'analyse de x_i sinon on aurait $q_{k_i} = (\bar{q}_{k_i}, l_i)$ et $q_{\bar{k}_i} = (\bar{q}_{k_i}, l_i + 1)$, ou bien $q_{\bar{k}_i} = (\bar{q}_{k_i}, l_i)$ et $q_{k_i} = (\bar{q}_{k_i}, l_i + 1)$. De la même façon il est facile de voir que si D passait par un col lors de l'analyse de x_i alors il passerait aussi, nécessairement par un pic. De plus, ayant exclus le cas A), D passe nécessairement par un col ou un pic lors de l'analyse de tout $y_i, 0 < i < r$ et par conséquent $1 \leq r \leq 2m + 1$. Il en résulte immédiatement qu'il existe un mot $y_k = Z_1 Z_2^k Z_3$ tel que :

$(q_0, y_0 x_1^k y_1 \dots x_r y_r, \varepsilon) \vdash^* (q_{\bar{k}_1}, x_1^k y_1 \dots x_r y_r, (Z_1, 1)) \vdash^*$

$(q_{k_1}, x_1^{k-1} y_1 \dots x_1^k y_r, (Z_1 Z_2, 1)) \vdash^* \dots (q_{k_1}, y_1 \dots x_r y_r, (Z_1 Z_2^k, 1)) \vdash^*$

$\vdash^* (q_{k_2}, x_2 y_2 \dots x_r y_r, (y_k, i_{k_2})) \vdash^* \dots$

$\vdash^* (q_{k_r}, x_r y_r, (y, i_{k_r})) \vdash^* (q_{k_r}, y_r, (y, i_{k_r}))$

$\vdash^* (q_F, \varepsilon, \varepsilon)$

Autrement dit le mot $w_k = y_0 x_1^k y_1 x_2^k y_2 \dots x_r y_r \in L(D)$ ce qui prouve les points 2 et 5 du lemme. Le 3) se démontre comme en A).

Il nous reste à prouver qu'un des x_i au moins contient des positions distinguées.

En effet, sinon, il existerait des y_i contenant la totalité des positions distinguées. Ce nombre étant supérieur ou égal à k_D il nous serait possible alors, par le raisonnement précédent, de trouver une nouvelle factorisation de w correspondant à deux nouveaux symboles A'_k et $A'_{\bar{k}}$, tels que $A'_k = A'_{\bar{k}}$ avec k et k' dans $\{1, k-1\} \cup \{\bar{k}+1, s\}$, factorisation qui soit telle qu'un des x_i ainsi obtenu contienne des positions distinguées.

Lemme 2 :

- 1) Si m est fini, alors $L_{m+1,1}$ n'est pas dans $C_{m,\omega}$.
- 2) Si m est fini, alors $L_{\omega,1}$ n'est pas dans $C_{m,\omega}$.

Démonstration :

1) Si $m = 0$ alors $C_{m,\omega} = \text{Rat}(X^*)$ et $L_{1,1} = \{a^r b^r c^r / q \geq r\}$ n'est pas algébrique.

Supposons donc $m \geq 1$. Considérons alors le langage

$$L = L_{m+1,1} = \left\{ A^r a^{p_1} b^{q_1} a^{p_2} b^{q_2} \dots a^{p_{m+1}} b^{q_{m+1}} \middle/ \sum_{i=1}^{m+1} p_i = \sum_{i=1}^{m+1} q_i \right. \\ \left. \sum_{i=1}^j p_i \geq \sum_{i=1}^j q_i \forall j, 1 \leq j \leq m+1, \sum_{j=0}^i (p_j - q_j) + p_{i+1} \leq r \forall i, 1 \leq i \leq m \right\}.$$

L appartient à $C_{m+1,\omega}$ de façon évidente.

Supposons que L appartienne à $C_{m,\omega}$. Alors il existe un AV D , strictement (m, ω) -borné tel que $L = L(D)$.

Soit alors

$w = A^r a^{p_1} b^{q_1} \dots a^{p_{m+1}} b^{q_{m+1}}$ un mot de L tel que p_i et q_i soient supérieurs à k_D , $1 \leq i \leq m+1$, où k_D est l'entier associé à D par le lemme 1.

L n'est évidemment pas un langage rationnel.

Considérons alors les q_1 premiers « b » comme distingués.

Alors d'après le lemme 1 il existe un entier $r \geq 1$ et des mots

$$y_0, y_1, \dots, y_r, x_1, x_2, \dots, x_r$$

de X^* tels que $w = y_0 x_1 y_1 x_2 y_2 \dots x_r y_r$.

D'après le point 4) du lemme 1 il existe un entier i_0 tel que x_{i_0} contient des positions distinguées. Remarquons que x_{i_0} est donc nécessairement inclus dans b^* . Il en résulte que $i_0 \geq 3$ et donc que D est passé par un pic lors de l'analyse du préfixe $A^r a^{p_1} b^{q_1}$. En effet si $i_0 = 1$ alors $y_0 = A^r a^{p_1} b^s$, $s \geq 0$ et L contiendrait des mots de la forme $A^r a^{p_1} b^s (b^t)^k a^{p_2} \dots b^{q_{m+1}}$, $\forall k \in \mathbb{N}$, ce qui est absurde.

De même si $i_0 = 2$ alors soit $y_0 = A^r a^{p_1}$, $x_1 = a^{p'_1}$, $y_1 = a^{p''_1} b^s$ et $x_2 = b^t$ et donc L contiendrait des mots de la forme

$A^r a^{p_1} (a^{p'_1})^k a^{p''_1} b^s (b^t)^k \dots b^{q_{m+1}}$, avec $k p'_1 > r$ ce qui est impossible, soit

$$y_0 = A^{r'}, x_1 = A^{r''}, y_1 = A^{r'''} a^{p_1} b^s \text{ et } x_2 = b^t$$

auquel cas L contiendrait des mots de la forme

$$A^{r_1} a^{p_1} b^{q_1} \dots b^{q_{m+1}} \text{ avec } q'_1 > p_1.$$

Si l'on considère alors les q_2 « b » suivants comme distingués, il résulte du lemme 1 qu'il existe un entier $i_1 > i_0$ tel que x_{i_1} contienne des positions distinguées.

Supposons que $i_1 = i_0 + 1$. Alors $y_{i_0} \in a^*$ et donc D accepterait des mots de la forme

$$A' a^{p_1} b^{q_1} a^{p_2} b^{q_2} \dots b^{q_{m+1}}$$

avec $q'_2 + q'_1 > p'_1 + p_2$, ce qui est impossible.

Donc $i_1 \geq i_0 + 2$ et D est donc passé par au moins 2 pics pour analyser le préfixe $A' a^{p_1} b^{q_1} a^{p_2} b^{q_2}$.

En poursuivant ce raisonnement jusqu'à considérer les q_{m+1} derniers « b » comme distingués, on montre que D doit nécessairement faire $m + 1$ mouvements pour analyser w , i.e. D n'est pas strictement (m, ω) -borné.

Donc $L \notin \mathcal{C}_{m, \omega}$.

2) Supposons que $L_{\omega, 1}$ est dans $\mathcal{C}_{m, \omega}$ pour quelque m dans \mathbf{N} .

Alors on aurait

$$L_{m+1, 1} \in L_{\omega, 1} \cdot \Gamma = \mathcal{C}_{\omega, 1} \subset \mathcal{C}_{m, \omega}$$

ce qui contredit le point 1 du lemme. ■

Lemme 3 :

Pour tout AV D , il existe un AV D' tel que $L(D) = L(D')$ et si $w \in L(D) = L(D')$ alors il existe une analyse de w par D' par un mot y tel que $|y| \leq |w|$.

De plus si D est strictement (m, n) -borné alors D' l'est aussi.

Démonstration :

Soit $D = (K, X, G_D, \delta, q_0, F)$.

Si pour tout $w \in L(D)$, il existe une analyse de w par un mot y tel que $|y| \leq |w|$, alors $D = D'$.

Supposons alors qu'il existe des mots w tels que toute analyse se fasse par des mots y , $|y| > |w|$.

Remarquons tout d'abord que si cette analyse correspond à une suite de configurations :

$$\begin{aligned} (q_0, w, \varepsilon) \stackrel{*}{\vdash} (q_i, w_i, (y', 1)) \stackrel{*}{\vdash} (q_j, w_j, (y, 1)) \\ \stackrel{*}{\vdash} (q_k, w_k, (y, i_k)) \stackrel{*}{\vdash} (q_F, \varepsilon, \varepsilon) \end{aligned}$$

y' préfixe strict de $y = A_n \dots A_1$ telle que $\max_k (i_k) < n$; alors on peut construire

un AV $D' = (K', X, G_D, \delta', q_0, F')$ avec

$$K' = K \cup \{ q_Z, Z \in G_D \} \quad \text{et} \quad \delta'(q, a, Z) = \delta(q, a, Z)$$

et si $(q', Z') \in \delta(q, a, Z)$ alors $(q_{Z'}, \varepsilon) \in \delta'(q_{Z'}, a, X) \quad \forall X \in G_D$ et si $(q, Z') \in \delta(q_0, a, \varepsilon)$ alors $(q_{Z'}, \varepsilon) \in \delta'(q_0, a, \varepsilon)$.

Il est clair que $L(D) = L(D')$ et il existe une analyse de w (non déterministe) par D' , par un mot $y' = A_{\max(i_k)} \dots A_1$.

D'autre part si dans la suite de configurations

$$(q_0, w, \varepsilon) \stackrel{*}{\vdash} (q_i, w_i, (y', 1)) \stackrel{*}{\vdash} (q_j, w_j, (y, 1)) \stackrel{*}{\vdash} \dots \stackrel{*}{\vdash} (q_k, w_k, (y, i_k)) \stackrel{*}{\vdash} (q_{k'}, w_{k'}, (y, i_{k'})) \stackrel{*}{\vdash} (q_F, \varepsilon, \varepsilon),$$

il existe r et s , $r > s$, tels que pour tous k et k' vérifiant $i_k = r$ et $i_{k'} = s$ (resp. $i_k = s$ et $i_{k'} = r$) on ait $w_k = w_{k'}$, on peut modifier trivialement D en un AV D' , par le même procédé que ci-dessus, pour que $L(D) = L(D')$ et D' analyse w par un mot $y' = Z_n \dots Z_{r+1} Z_s \dots Z_1$.

Autrement dit pour tout Z_k , $1 \leq k \leq n$, il existe au moins un suffixe w_i de w et $a \in X$ tels que :

$$(q_0, w, \varepsilon) \stackrel{*}{\vdash} (q_{l-1}, aw_L, (y, k)) \stackrel{*}{\vdash} (q, w_b, (y, i_i)) \stackrel{*}{\vdash} (q_F, \varepsilon, \varepsilon)$$

Il existe donc pour tout $w \in L(D) = L(D')$ une analyse de w par D' , par un mot y , $|y| \leq |w|$.

Enfin de par la construction de D' il est clair que si D est strictement (m, n) -borné, alors D' l'est aussi. ■

Lemme 4 :

- 1) Si n est fini, alors $L_{1, n+1}$ n'est pas dans $C_{\omega, n}$.
- 2) Si n est fini, alors $L_{1, \omega}$ n'est pas dans $C_{\omega, n}$.

Démonstration :

- 1) Si $n = 0$ alors $C_{\omega, 0} = \text{Rat}(X^*)$ et $L_{1, 1}$ n'est pas rationnel.

Si $n \geq 1$, considérons le langage

$$L_{n+1} = \{ wc\tilde{w}/w \in a_1^* a_2^* \dots a_{n+1}^* \}$$

L_{n+1} est trivialement dans $C_{1, n+1}$. Il nous suffit donc de montrer que L_{n+1} n'est pas dans $C_{\omega, n}$.

Supposons qu'il existe un AV D , strictement (ω, n) -borné, qui accepte L_{n+1} . D'après le lemme 3 on peut supposer que l'analyse de tout mot $\tilde{w}cw \in L_{n+1}$ est faite par un mot y tel que $|y| \leq 3|w|$.

Si D a \mathcal{K} états distincts, après avoir lu un facteur w de longueur r inclus dans $a_1^* a_2^* \dots a_{n+1}^*$, D passe par au plus $\mathcal{K} \cdot f_n(3r)$ configurations distinctes avant d'analyser le facteur w , avec

$$f_n(x) = \text{card} \{ y \mid y \in a_1^* \dots a_n^*, |y| \leq |x| \}.$$

On montre [10] que $f_n(x)$ est un polynôme de degré n de la variable x et donc pour tous entiers k et k' il existe un entier $r(k, k')$ tel que

$$f_{n+1}(x) > k f_n(k'x) \quad \text{si} \quad x \geq r(k, k')$$

Or le nombre de facteurs différents inclus dans $a_1^* \dots a_{n+1}^*$ et de longueur r est exactement $f_{n+1}(r)$.

Il en résulte qu'il existe au moins un entier r tel que $f_{n+1}(r) > \mathcal{K} \cdot f_n(3r)$ et donc qu'il existe des mots w_1 et w_2 , $w_1 \neq w_2$ un état q , un mot mémoire y et deux états terminaux q_F et q'_F tels que :

$$(q_0, w_1 c \tilde{w}_1, \varepsilon) \stackrel{*}{\vdash} (q, c \tilde{w}_1, (y, 1)) \stackrel{*}{\vdash} (q_F, \varepsilon, \varepsilon)$$

et

$$(q_0, w_2 c \tilde{w}_2, \varepsilon) \stackrel{*}{\vdash} (q, c \tilde{w}_2, (y, 1)) \stackrel{*}{\vdash} (q'_F, \varepsilon, \varepsilon)$$

Ainsi l'AV D accepterait le mot $w_1 c \tilde{w}_2 \notin L_{n+1}$, donc $L_{1, n+1}$ n'est pas dans $\mathcal{C}_{\omega, n}$.

2) Si $L_{1, \omega}$ est dans $\mathcal{C}_{\omega, n}$ pour n fini alors

$$L_{1, n+1} \in \{ L_{1, n+1} \} \cdot \Gamma \subset \mathcal{C}_{1, \omega} = \{ L_{1, \omega} \} \cdot \Gamma \subset \mathcal{C}_{\omega, n}$$

ce qui contredit le point 1) du lemme.

Démonstration de la proposition 1 :

- . Si $(m, n) \leq (m', n')$ alors $\mathcal{C}_{m, n} \subseteq \mathcal{C}_{m', n'}$ par définition.
- . Supposons que $L_{m, n} \in \mathcal{C}_{m', n'}$. Alors soit $(m, n) \leq (m', n')$, soit l'on est dans l'un des quatre cas suivants :

1^{er} cas : Si $n' < n < \infty$ alors $n' \leq n - 1$ et par conséquent

$$\mathcal{C}_{m', n'} \subseteq \mathcal{C}_{m', n-1} \subseteq \mathcal{C}_{\omega, n-1}. \quad \text{Si } L_{m, n} \in \mathcal{C}_{m', n'}$$

alors $L_{1, n} \in \mathcal{C}_{1, n} \subseteq \mathcal{C}_{m', n'} \subseteq \mathcal{C}_{\omega, n-1}$ ce qui contredit le lemme 4.

2^e cas : Si $n = \omega$, alors si $L_{m, \omega} \in \mathcal{C}_{m', n'}$ on a

$$L_{1, \omega} \in \mathcal{C}_{1, \omega} \subseteq \mathcal{C}_{m, \omega} \subseteq \mathcal{C}_{m', n'} \subseteq \mathcal{C}_{\omega, n'}$$

ce qui contredit le lemme 4.

3^e cas : Si $m' < m < \infty$, on aboutit comme dans le 1^{er} cas à une contradiction avec le lemme 2.

4^e cas : Si $m = \omega$, on aboutit comme dans le 2^e cas à une contradiction avec le lemme 2. ■

Démonstration du théorème 3 :

La condition suffisante est évidente.

Réciproquement supposons $\mathcal{C}_{m,n} \subseteq \mathcal{C}_{m',n'}$.

Quatre cas sont à considérer :

1^{er} cas : Si $m \neq \infty \neq n$. Si $\mathcal{C}_{m,n} \subseteq \mathcal{C}_{m',n'}$ alors $L_{m,n} \in \mathcal{C}_{m',n'}$ et $(m, n) \leq (m', n')$ d'après la proposition 1.

2^e cas : Si $n = \infty$ et $m \neq \infty$. Alors pour tout n_1 fini

$$\mathcal{C}_{m,n_1} \subseteq \mathcal{C}_{m,\infty} \subseteq \mathcal{C}_{m',n'}$$

D'après la proposition 1 on a donc $(m, n_1) \leq (m', n')$ pour tout n_1 fini soit donc $n' \in \{\infty, \omega\}$ et dans tous les cas $(m, n) \leq (m', n')$.

3^e cas : Si $m = \infty$ et $n \neq \infty$. Par un raisonnement analogue à celui fait dans le 2^e cas, il vient $(m, n) \leq (m', n')$.

4^e cas : Si $n = m = \infty$, alors pour tous m_1 et n_1 dans \mathbb{N} on a

$$\mathcal{C}_{m_1,n_1} \subseteq \mathcal{C}_{\infty,\infty} \subseteq \mathcal{C}_{m',n'}$$

et donc m' et $n' \in \{\infty, \omega\}$ et $(m, n) \leq (m', n')$. ■

Corollaire 1 :

Soient m, n, m' et n' dans $\mathbb{N} \cup \{\infty, \omega\}$. Alors $\mathcal{C}_{m,n} = \mathcal{C}_{m',n'}$ si et seulement si $(m, n) = (m', n')$.

Corollaire 2 :

Soient m, n, m' et n' dans $\mathbb{N} \cup \{\infty, \omega\}$, tels que $(m, n) \leq (m', n')$ $(m, n) \neq (m', n')$. Alors $\mathcal{C}_{m',n'} - \mathcal{C}_{m,n}$ est infini.

Corollaire 3 :

Pour tous m et n dans $\mathbb{N} \cup \{\infty, \omega\}$, les cônes rationnels $\mathcal{C}_{m,\infty}$ et $\mathcal{C}_{\infty,n}$ sont non principaux.

Démonstration :

Raisonnons par exemple sur $\mathcal{C}_{m,\infty}$. Si $\mathcal{C}_{m,\infty}$ étant principal, il existerait un langage L et un entier n tels que $L \in \mathcal{C}_{m,n}$ et $\{L\} \Gamma = \mathcal{C}_{m,\infty}$.

On aurait donc $\mathcal{C}_{m,n'} = \mathcal{C}_{m,n}$ pour tout $n' > n$. Ce qui contredit le théorème 3.

REFERENCES

- [1] L. BOASSON et M. NIVAT, *Sur diverses familles de langages fermées par transduction rationnelle*, Acta Informatica 2 (1973), 180-188.
- [2] J.-P. CRESTIN, *Langages quasi-rationnels*, in Actes de l'École de Printemps sur la Théorie des langages algébriques, Bonascre, 1973, à paraître chez Gauthier-Villars.
- [3] S. GINSBURG, S. GREIBACH and M. HARRISON, *Stack Automata and Compiling*, J. Assoc. Comput. Mach., 14, (1967), p. 172-201.
- [4] S. GINSBURG, S. GREIBACH and M. HARRISON, *One-Way Stack Automata*, J. Assoc. Comput. Mach., 14, (1967), p. 389-418.
- [5] S. GREIBACH, *An Infinite Hierarchy of Context-free Languages*, J. Assoc. Comput. Mach., 16, (1969), p. 91-106.
- [6] S. GREIBACH, *Checking Automata and One-Way Stack Languages*, J. Comput. System Sci., 3, (1969), p. 196-217.
- [7] M. HARRISON and M. SCHKOLNICK, *A Grammatical Characterisation of One-Way Nondeterministic Stack Languages*, J. Assoc. Comput. Mach., 18, (1971), p. 148-172.
- [8] M. NIVAT, *Opérateurs sur des familles de langages*, in Actes de l'École de Printemps sur la théorie des langages algébriques, Bonascre, 1973, à paraître chez Gauthier-Villars.
- [9] W. F. OGDEN, *Intercalation Theorems for Pushdown Store and Stack Languages*, Ph. D. Thesis, Stanford, 1968.
- [10] J. RIORDAN, *An Introduction to Combinatorial Analysis*, Wiley, New York, 1958.
- [11] F. RODRIGUEZ, *Cônes d'Accepteurs-Application à l'Étude d'une Hiérarchie Infinie de Cônes Rationnels de Langues d'Accepteurs Vérificateurs*, Thèse Docteur-Ingénieur, Toulouse, 1973.
- [12] F. RODRIGUEZ, *Une Hiérarchie Infinie de Cônes d'Accepteurs Vérificateurs*, in 1. Fachtagung über Automatentheorie und Formale Sprachen, Lecture Notes in Computer Science, Springer-Verlag, 1973, p. 81-87.
- [13] R. SIROMONEY, *Finite-Turn Checking Automata*, J. Comput. System Sci., 5, (1971), p. 549-559.