Revue française d'automatique informatique
recherche opérationnelle. Informatique théorique

J. M. Autebert
J. Beauquier

Une caractérisation des générateurs standard

Revue française d'automatique informatique recherche opérationnelle. Informatique théorique, tome 8, nº R1 (1974), p. 63-83
<http://www.numdam.org/item?id=ITA_1974__8_1_63_0>

R.A.I.R.O.
(8e année, R-1, 1974, p. 63-83)

UNE CARACTERISATION DES GENERATEURS STANDARD

par J. M. AUTEBERT et J. BEAUQUIER (1)
communiqué par J. BERSTEL

Résumé. — *Un langage algébrique L est générateur du cône rationnel des langages algébriques, si, pour tout élément L' de ce cône, il existe une transduction rationnelle* τ *telle que* $\tau(L) = L'$. *On caractérise dans cette article les générateurs standard (de la forme* $(D'^*_n \cap K)$, *où* D'^*_n *est le langage de Dyck sur un alphabet à 2n lettres et K un langage rationnel). Cette caractéristion porte sur le rationnel K, plus précisément sur une propriété des automates finis reconnaissant K. Elle permet de retrouver la plupart des générateurs connus.*

1. INTRODUCTION

On appelle générateur algébrique tout langage algébrique L tel que, pour tout langage algébrique L', il existe une transduction rationnelle de L sur L'. En d'autres termes le cône rationnel engendré par L, c'est-à-dire l'ensemble des images de L dans une transduction rationnelle, est la famille des langages algébriques.

Du théorème ancien de Chomsky et Schützenberger découle immédiatement que le langage de Dyck sur deux lettres, $D_2'^*$ est un tel générateur algébrique.

L'étude de ces générateurs a été développée par S. Ginsburg et S. Greibach [4] dans le cadre de leurs travaux sur les familles agréables de langages (AFL) et quelques-unes de leurs propriétés ont été démontrées par plusieurs auteurs (Boasson [2], Nivat [7], Schützenberger [8]). Mais leur caractérisation est un problème toujours ouvert, si bien que la démonstration qu'un langage déterminé est ou n'est pas générateur reste toujours chose pénible (cf. [8]).

---

(1) Institut de Programmation, Paris.

Ci-dessous, nous proposons une caractérisation des générateurs algébriques de la forme $(D_n'^* \cap K)$ (K est un langage rationnel) ou générateurs algébriques standard. Cette caractérisation permet de retrouver la plupart des générateurs connus. De plus, tout langage algébrique est image par transduction rationnelle d'un langage standard [3], et tout langage algébrique très simple est rationnellement équivalent à un langage standard [7].

De nombreux exemples d'application pourront être trouvés dans [1].

2. PRELIMINAIRES

Nous supposons connues du lecteur les notions classiques de Théorie des Langages : langage rationnel, algébrique, automate fini, transduction rationnelle... ([5], [6], [7]).

Signalons cependant que nous notons 1_{X^*} (ou plus simplement 1) l'élément neutre du monoïde libre X^* et énonçons quelques définitions plus particulières.

2.1. Langages de Dyck

Soient $X_n = \{ \sigma_1, \sigma_2, ..., \sigma_n \}$ et $\bar{X}_n = \{ \bar{\sigma}_1, \bar{\sigma}_2, ..., \bar{\sigma}_n \}$ deux alphabets disjoints et $\Sigma_n = X_n \cup \bar{X}_n$. Les lettres de X_n seront dites non barrées, celles de $\bar{X}_n$ barrées. Si x est une lettre de Σ_n, $\bar{x}$ désigne, soit $\bar{\sigma}_i$ si $x = \sigma_i$, soit σ_i si $x = \bar{\sigma}_i$.

On désigne par θ la congruence engendrée par les relations :

$$\sigma_i\bar{\sigma}_i \equiv 1 \text{ pour } i = 1, ..., n$$

et l'on appelle langage de Dyck $D_n'^*$ sur Σ_n la classe du mot vide modulo la congruence θ. L'ensemble des mots premiers de $D_n'^*$ est :

$$D_n' = \{ w \mid w \in D_n'^*/\{ 1 \} \text{ et } w = w_1w_2,\, w_1 \in D_n'^*/\{ 1 \} \text{ entraîne } w_2 = 1 \}$$

— $D_n'^*$ est le sous-monoïde libre engendré par D_n'.

— $D_n'^*$ est un langage algébrique.

— Si w est un mot de $D_n'^*$, considérons la suite de mots :

$$w = w_0, w_1, ..., w_k = 1$$

où :

$$\forall i \in [0, k-1], \exists j \in [1, n], \exists\, u_i, v_i \in \Sigma_n^* \text{ tels que } w_i = u_i\sigma_j\bar{\sigma}_jv_i \text{ et } w_{i+1} = u_iv_i$$

Une telle suite existe et permet d'établir une bijection unique entre les occurrences des lettres non barrées de w et les occurrences des lettres barrées. Nous appelons une telle bijection réduction de Dyck du mot w en le mot vide.

— Si $w = w_1\sigma_i w_2\bar{\sigma}_i w_3$ est un mot de $D_n'^*$ et si σ_i et $\bar{\sigma}_i$ se correspondent dans la réduction de Dyck de w, alors $w_2 \in D_n'^*$ et $w_1 w_3 \in D_n'^*$.

2.2. Lemmes combinatoires relatifs au langage de Dyck

Définitions

— Un mot h sera dit facteur strict d'un mot g si $g = \alpha h \beta$ avec $\alpha\beta \neq 1$.

— Un mot w de $D_n'^*$ est de puissance inférieure à k_0 (où $k_0 \in \mathbf{N}_+$) si et seulement si :

$$w = \alpha d_1 d_2 \dots d_k \beta \text{ et } d_i \in D_n' (\text{pour } i = 1, 2, \dots, k) \text{ entraîne } k \leqslant k_0$$

— Un mot h de $D_n'^*$ est de hauteur inférieure à k_0 (où $k_0 \in \mathbf{N}_+$) si et seulement si, pour toute suite $\mathrm{d}_0 = w, d_1, \dots, \mathrm{d}_k$ de mots de $D_n'^*/\{1\}$ tels que d_i est un facteur strict de d_{i-1}, on a $k \leqslant k_0$.

Lemme 1

Pour tout entier $N \geqslant 0$, l'ensemble :

$D_n'^* \cap \{ w \in \Sigma_n^* \mid w$ est de hauteur inférieure à N et de puissance inférieure à $N \}$ est fini.

Preuve

Nous allons prouver par récurrence sur h que les ensembles :

$\Delta(h,p) = D_n'^* \cap \{ w \in \Sigma_n^* \mid w$ est de hauteur inférieure à h et de puissance inférieure à $p \}$ vérifient :

$$w \in \Delta(h,p) \Rightarrow |w| \leqslant 2p \frac{(p^{h+1} - 1)}{p-1} \quad (|w| \text{ désignant la longueur du mot } w)$$

Si $h = 0$, alors $w \in \Delta(h,p)$ s'écrit :

$$w = x_1\bar{x}_1 x_2\bar{x}_2 \dots x_r\bar{x}_r \qquad (r \leqslant p)$$

et

$$|w| \leqslant 2p.$$

Si $h = 1$, alors :

$$w = w_1 w_2 \dots w_r \qquad (r \leqslant p) \qquad \text{où} \qquad w_i = \sigma_j w_i' \bar{\sigma}_j \text{ et } w_i' \in \Delta(0,p)$$

et

$$|w| \leqslant p[2p + 2] = 2p \frac{p^2 - 1}{p - 1}$$

Supposons démontré que :

$$w \in \Delta(h - 1,p) \Rightarrow |w| \leqslant 2p \frac{(p^h - 1)}{p - 1}$$

et soit w un mot de $\Delta(h, p)$. Alors :

$$w = w_1 w_2 \dots w_r (r \leqslant p) \quad \text{où} \quad w_i = \sigma_j w_i' \bar{\sigma}_j \text{ et } w_i' \in \Delta(h - 1,p)$$

Donc :

$$|w| \leqslant p[2p \frac{p^h - 1}{p - 1} + 2] = 2p \frac{(p^{h+1} - 1)}{p - 1}$$

Q.E.D.

Lemme 2

Soit w un mot de $D_n'^*$ qui s'écrit :

$$w = \alpha c \beta c \gamma c \delta, \qquad \alpha, \beta, \gamma, \delta \in \Sigma_n^* \qquad \beta, \gamma \notin D_n'^*$$
$$c \in \Sigma_n$$
$$\text{et } |\alpha\beta\gamma\delta|_c = 0 \qquad |\beta\gamma|_{\bar{c}} = 0$$

Soit u (resp. v) le plus petit facteur de w contenant β (resp. γ) et appartenant à $D_n'^*$:

$$w = fug = f'vg'$$

1° On n'a pas :

$$|f| = |f'| \qquad et \qquad |g| = |g'|$$

2° Si $|f'| \geqslant |f|$ *et* $|g'| \geqslant g$ alors :

$$f' = fh \qquad g' = kg$$

et le mot hk contient une occurrence de la lettre c.

3° Si $|f'| \leqslant |f|$ et $|g'| \leqslant |g|$ alors :

$$f = f'h' \qquad g = k'g'$$

et le mot $h'k'$ contient une occurrence de la lettre c.

La preuve se fait par cas.

2.3. Langage standard

Nous disons que le langage algébrique L est standard s'il existe un entier n et un rationnel K de Σ_n^* tels que :

$$L = D_n'^* \cap K$$

Il est d'ordre n si :

$$D_{n-1}'^* \cap K \subsetneq D_n'^* \cap K$$

3. CARACTERISATION DES GENERATEURS STANDARD

Nous allons maintenant démontrer une condition nécessaire et suffisante pour qu'un langage standard $L = D_n'^* \cap K$ soit un générateur du cône des langages algébriques. L'idée générale qui guide cette démonstration est la suivante : si le langage L est générateur, il est rationnellement équivalent au langage $D_2'^*$ que nous savons être générateur, et il contient alors $D_2'^*$ sous une forme « codée ».

La démonstration se fait en plusieurs étapes : chaque étape considère un mot de L d'une certaine complexité. L'étape suivante considère un mot d'une complexité plus grande et utilise les résultats démontrés à l'étape précédente. A chaque niveau, les résultats obtenus peuvent se représenter dans la mise en évidence d'un certain sous-automate d'un automate $A(K)$, reconnaissant le rationnel K.

Lemme I

Si le langage algébrique L de Σ_n^* s'envoie par transduction rationnelle sur le langage L' de Σ_2^*, il existe un homomorphisme de semi-groupe Λ :

$$\Lambda : \Sigma_2^+ \to \mathfrak{P}(\Sigma_n^*)$$

tel que :

$$\forall f \in \Sigma_2^+,\ f \in L' \Rightarrow \Lambda(f) \cap L \neq \varnothing$$

Preuve

Supposons :

$$L' = \psi[\varphi^{-1}(L) \cap R]$$

où $R \subset Z^*$ est un rationnel et $\varphi : Z^* \to \Sigma_n^*$ et $\psi : Z^* \to \Sigma_2^*$ sont deux homomorphismes alphabétiques :

Posons :

$$H_u = \{ z \in Z | \psi(z) = u \}, \lambda(u) = \varphi(H_u) \text{ et } W_1 = \varphi(\psi^{-1}(1))$$

Si le mot f de Σ_2^* s'écrit :

$$f = x_1 x_2 \dots x_k \qquad (x_i \in \Sigma_2)$$

soit :

$$\Lambda(f) = W_1 \lambda(x_1) W_1 \lambda(x_2) W_1 \dots W_1 \lambda(x_k) W_1$$

Λ est un homomorphisme et, de sa construction, il résulte que :

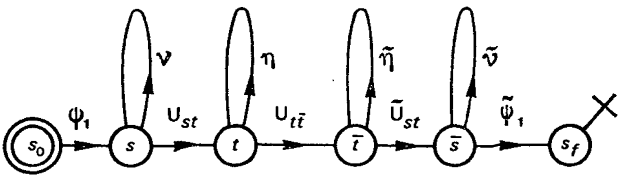
$$g \notin \Lambda(f) \Rightarrow \psi[\varphi^{-1}(g) \cap R] \neq f$$

Q.E.D.

Lemme II

Soit B la partie de $\Sigma_2^* : B = \{ \sigma_1^\alpha \sigma_2^\alpha \bar{\sigma}_2^\alpha \bar{\sigma}_1^\alpha \mid \alpha \in \mathbf{N}_+ \}$.

Si le langage standard $L = D_n'^* \cap K$ est tel qu'il existe une transduction rationnelle τ vérifiant : $B \subseteq \tau(L) \subseteq D_2'^*$ alors, il existe un rationnel $K' \subseteq K$, reconnu par un automate du type suivant :



avec : $\psi_1 \tilde{\psi}_1$, $U_{st} \tilde{U}_{st}$, $\nu \tilde{\nu}$, $\eta \tilde{\eta}$, $U_{t\bar{t}} \in D_n'^*$

Preuve :

Considérons les mots $M_\alpha = \sigma_1^\alpha \sigma_2^\alpha \bar{\sigma}_2^\alpha \bar{\sigma}_1^\alpha$ où $\alpha \in \mathbf{N}_+$.

Pour tout entier naturel α, M_α est un mot de $D_2'^*$.

D'après le lemme I, il existe un mot w de L tel que :

$$w \in \Lambda(M_\alpha) \qquad (\Lambda \text{ étant l'homomorphisme associé à } \tau)$$

et l'on peut écrire w sous la forme :

$$w = \upsilon_0 x_1^1 \upsilon_1 x_1^2 \upsilon_2 \ldots x_1^\alpha \upsilon_\alpha x_2^1 \upsilon_{\alpha+1} \ldots x_2^\alpha \upsilon_{2\alpha} x_3^1 \ldots x_3^\alpha \upsilon_{3\alpha} x_4^1 \ldots x_4^\alpha \upsilon_{4\alpha}$$

avec $\upsilon_i \in W_1$ pour $i \in [0,4\alpha]$

$x_1^i \in \lambda(\sigma_1)$

$x_2^i \in \lambda(\sigma_2)$

$x_3^i \in \lambda(\bar{\sigma}_2)$

$x_4^i \in \lambda(\bar{\sigma}_1)$ pour $i \in [1, \alpha]$

$\lambda(\sigma_1)$ est une partie finie de $\Sigma_n \cup \{ 1 \}$. Donc, k étant un entier arbitraire, il existe un entier α et un élément x de $\lambda(\sigma_1)$, tels que l'élément x ait au moins k occurrences dans le mot :

$$x_1^1 x_1^2 \ldots x_1^\alpha$$

Le mot w correspondant à l'entier α choisi s'écrit, dans ces conditions :

$$w = \xi_0 x \xi_1 x \ldots x \xi_{k-1} x w' \qquad \xi_i \in \Sigma_n^*, \ i = 0, \ldots, (k-1)$$

en mettant en évidence les occurrences de l'élément x.

Un automate reconnaissant le rationnel K, $A(K) = \{ \Sigma_n, E, s_0, S', \delta \}$, est un automate fini (avec μ états). Considérons la suite de ses états lorsqu'il lit successivement les mots :

$$\xi_0,\ \xi_0 x \xi_1,\ \xi_0 x \xi_1 x \xi_2, \ldots, \xi_0 x \xi_1 x \xi_2 x \ldots x \xi_{k-1}$$

l étant un entier arbitraire, on peut choisir l'entier k tel qu'il existe un état S_u de E, qui apparaît au moins l fois dans cette suite. Nous pouvons alors écrire le mot w correspondant au choix fait, sous la forme :

$$w = \xi_0' x \xi_1' x \ldots x \xi_{l-1}' x w''$$

avec :

$$\delta(s_0, \xi_0') = \delta(s_0, \xi_0' x \xi_1') = \ldots = \delta(s_0, \xi_0' x \xi_1' x \ldots x \xi_{l-1}')$$

Montrons que l'on peut toujours faire un choix tel que les mots :

$$\xi_0',\ \xi_0' x \xi_1', \ldots, \xi_0' x \xi_1' x \ldots x \xi_{l-1}'$$

soient tous distincts. Cette justification concerne uniquement le cas où :

$$x = 1.$$

Supposons donc que, pour tout entier m :

$$1 \in \Lambda(\sigma_1^m)$$

$$\exists\, g_1, g_2, g_3 \in Z^* \; g = g_1 g_2 g_3 \in R,\ \varphi(g) \in L \text{ et } \varphi(g_2) = 1$$
$$\psi(g_2) = \sigma_1^m$$

En choisissant m suffisamment grand, on peut décomposer le facteur g_2 en :

$$g_2 = g_2' g_2'' g_2''' \qquad (g_2'' \neq 1)$$

de telle manière que

$$\forall p \in \mathbf{N} \qquad g_{(p)} = g_1 g_2' (g_2'')^p g_2''' g_3 \in R$$

Or : $\varphi(g_{(p)}) = \varphi(g) \in L$

et $\psi(g_{(p)}) \notin D_2'^*$ pour $p \geqslant 1$, d'où une contradiction.

Nous pouvons donc choisir tous les $x\xi_i'$ non vides. On définit des mots de Dyck χ_i de la façon suivante : à chaque mot $x\xi_i'$ on fait correspondre le plus petit mot de Dyck qui le contient et qui soit un facteur de w. Pour tout entier arbitraire l, on peut trouver α tel que l'on ait l mots χ_i de ce type.

Nous désignerons par Q_i (resp. F_i) l'état de $A(K)$ lorsque, lors de la lecture du mot w, l'automate est sur le point d'absorber χ_i (resp. vient d'absorber χ_i).

D'après le lemme 1 plusieurs cas sont à envisager :

Quel que soit l'entier m, on peut choisir l (et par conséquent α) de telle manière qu'il y ait m facteurs χ_{j_i} pour $i \in [1, m]$ qui sont :

— soit strictement emboîtés,

— soit non strictement emboîtés,

— soit consécutifs.

Dans chacun de ces cas, K et R étant deux rationnels, on peut choisir l'entier m (et par conséquent α) de telle manière qu'il y ait 2 facteurs, χ_p et χ_q satisfaisant :

— $Q_p = Q_q$

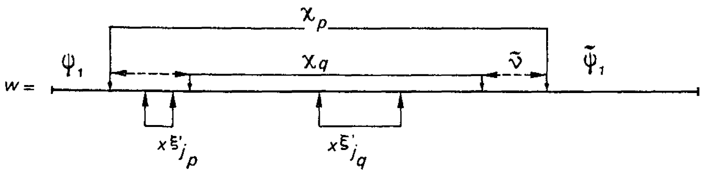
— $F_p = F_q$

— Il existe un mot $g \in R$ tel que $\varphi(g) = w$ qui s'écrit $g = g_1' c_1 g_1'' = g_2' c_2 g_2''$ avec $\varphi(c_1) = \chi_p$ et $\varphi(c_2) = \chi_q$,

tel que les états de l'automate reconnaissant R soient les mêmes après la lecture de chaque g_i' (resp. chaque $g_i' c_i$) $(i = 1, 2)$.

1er cas : χ_p et χ_q sont strictement emboîtés.

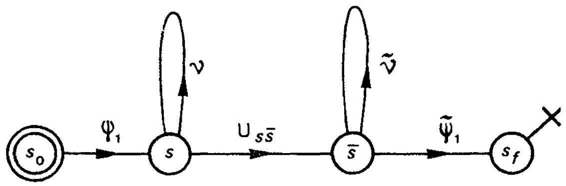
Cette hypothèse correspond à la configuration suivante :



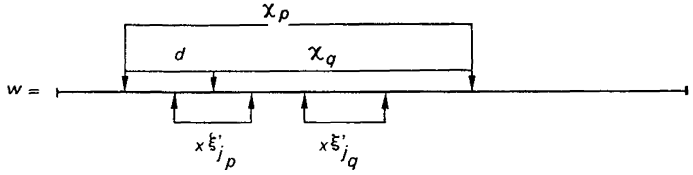
D'après le lemme 2 l'un au moins des mots ν ou $\tilde{\nu}$ contient un élément de $\Lambda(\sigma_1)$. Il est clair que l'on peut choisir les χ_i de telle sorte que ce mot ne contienne que des éléments de $\Lambda(\sigma_1)$ et $\Lambda(1)$.

De plus, l'hypothèse où $\tilde{\nu}$ contient un élément de $\Lambda(\sigma_1)$ contredit le fait que la transduction rationnelle τ envoie L dans $D_2'^*$.

On a donc mis ainsi en évidence un rationnel K' contenu dans K et reconnu par un automate du type suivant :



2e cas : χ_p et χ_q sont non strictement emboîtés. On a donc la configuration suivante :



(Ou la configuration symétrique.)

Il existe un mot g de R tel que

$$g = g_1 h g_2 \qquad \text{et} \qquad g_1 h^* g_2 \text{ est inclus dans } R$$

avec :

$$\varphi(h) = d, d \in \Lambda(\sigma_1^r) \quad (r > 0)$$

comme d appartient à $D_n'^*$ et que l'on est dans le même état de $A(K)$ avant et après avoir lu ce facteur de w, si $w_1 d w_2 = w$ le mot $w' = w_1 d^2 w_2$ appartient à $D_n'^* \cap K$ et il existe $g' \in R$, $g' = g_1 h^2 g_2$ tel que $\varphi(g') = w'$.

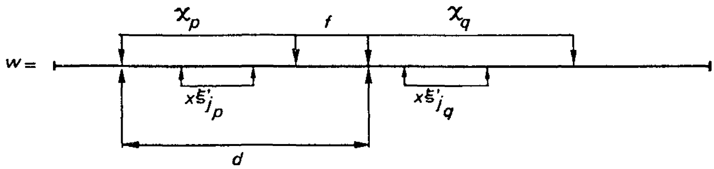
Or on a alors $\psi(g') \notin D_2'^*$ car $\psi(h) = \sigma_1^r \quad (r > 0)$.

D'où une contradiction qui exclut cette hypothèse.

3e cas :

χ_p et χ_q sont consécutifs.

Cette dernière hypothèse correspond à la configuration suivante :



χ_p contient au moins un élément de $\Lambda(\sigma_1)$ donc χ_p appartient à $\Lambda(\sigma_1^i)$ avec $i > 0$ (comme dans le premier cas). En posant $\chi_p f = d$, on se retrouve dans les conditions du second cas. Seule la première hypothèse est donc valable.

De la même façon, on peut faire cette construction en considérant la lettre σ_2 au lieu de la lettre σ_1.

Il en résulte qu'il existe des mots ψ_1', ψ_2', η, $\tilde{\eta}$, $U_{t\bar{t}}$ de Σ_n^* et des états t et $\bar{t}$ de $A(K)$ vérifiant :

$$\eta\tilde{\eta} \in D_n'^*, \; U_{t\bar{t}} \in D_n'^*$$

$$\delta(s_0, \psi_1') = \delta(s_0, \psi_1'\eta) = t$$

$$\delta(s_0, \psi_2') = \delta(s_0, \psi_2'\tilde{\eta}) = \bar{t}$$

$$\delta(t, U_{t\bar{t}}) = \bar{t}$$

C'est-à-dire que nous mettons ainsi en évidence l'existence d'un rationnel inclus dans K, reconnu par un automate analogue à celui représenté ci-dessus.

La conjonction de ces deux résultats prouve le lemme II.

Posons :

$$(\forall\, \alpha \in \mathbf{N}_+) \quad M_{\alpha,1} = \sigma_1^\alpha \sigma_2^\alpha \bar{\sigma}_2^\alpha \bar{\sigma}_1^\alpha$$

et

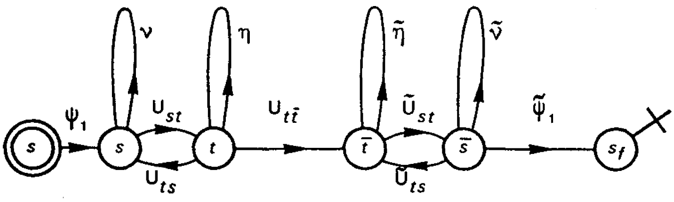
$$(\forall\, i \in \mathbf{N}_+) \quad M_{\alpha,i+1} = \sigma_1^\alpha \sigma_2^\alpha M_{\alpha,i} \bar{\sigma}_2^\alpha \bar{\sigma}_1^\alpha$$

enfin, soit C la partie de $\Sigma_2^* : C = \{ M_{\alpha,i} \mid \alpha, i \in \mathbf{N}_+ \}$.

Nous allons démontrer le :

Lemme III

Si le langage algébrique standard $L = (D_n'^* \cap K)$ est tel qu'il existe une transduction rationnelle τ vérifiant $C \subseteq \tau(L) \subseteq D_2'^*$ alors il existe un langage rationnel K', inclus dans K, reconnu par un automate fini du type suivant :



avec :

$$\psi_1\,\tilde{\psi}_1,\ U_{st}\tilde{U}_{st},\ U_{ts}\tilde{U}_{ts},\ \nu\tilde{\nu},\ \eta\tilde{\eta},\ U_{t\bar{t}} \in D_n'^*.$$

Preuve :

Si l'on refait la construction du Lemme II, le passage du mot $M_{\alpha,i}$ au mot $M_{\alpha,i+1}$, introduit deux paires d'états. Il suffit de remarquer que, en choisissant i suffisamment grand, on peut toujours trouver deux telles paires identiques. La mise en évidence d'un automate du type ci-dessus est alors aisée.

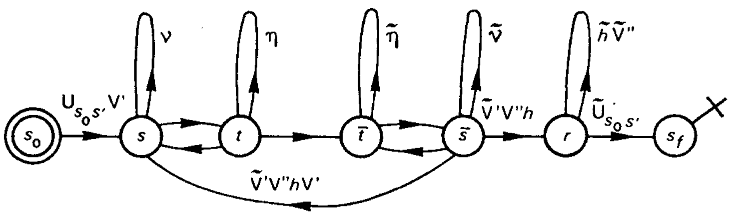
Posons :

$$(\forall\, m \in \mathbf{N}_+) T_{\alpha,p,m} = [M_{\alpha,p}]^m \text{ et } E = \{\, T_{\alpha,p,m} \mid \alpha, p, m \in \mathbf{N}_+ \,\}$$

Démontrons le :

Lemme IV

Si le langage algébrique standard $L = (D_n'^* \cap K)$ est tel qu'il existe une transduction rationnelle τ vérifiant $E \subseteq \tau(L) \subseteq D_2'^*$, alors il existe un langage rationnel K', inclus dans K, reconnu par un automate fini du type suivant :



avec :

$$U_{s_0 s}, \tilde{U}_{s_0 s'}, V'\tilde{V}', V''\tilde{V}'', h\tilde{h}, \nu\tilde{\nu}, \eta\tilde{\eta} \in D_n'^{*}.$$

Preuve :

Écrivons : $T_{\alpha,p,m} = M^{(1)}_{\alpha,p} M^{(2)}_{\alpha,p} \dots M^{(m)}_{\alpha,p}$. Pour chaque facteur $M^{(i)}_{\alpha,p}$ ($i = 1, \dots, m$), la construction du lemme III permet d'obtenir m quadruplets d'états. Le mot w choisi pour cette construction s'écrit : $w = w_1 w_2 \dots w_m$. Soit : $\omega = \omega_1 \omega_2 \dots \omega_m$, le sous-mot de w obtenu en ne conservant que les occurrences x telles que : $\psi[\varphi^{-1}(x)] \neq 1$. Pour tout entier k, nous pouvons choisir m de telle façon que la suite : $\omega_1, \omega_2, \dots, \omega_m$ contienne au moins k mots identiques. Pour tout entier l, nous pouvons choisir k de telle façon que, parmi les k mots identiques déterminés ci-dessus, il y en ait l : $\Omega_1, \Omega_2, \dots, \Omega_l$ qui, en posant :

$$w = \xi_1 V_1 \xi_2 V_2 \dots V_l \xi_{l+1} \qquad (V_i \text{ étant le facteur correspondant à } \Omega_i)$$

satisfont la relation :

$$(\forall\, i = 1, \dots, l),$$

$$[(V_i = \alpha_i x_i \beta_i) \text{ et } (x_i \text{ est la } j^{\text{e}} \text{ occurrence de } \Omega_i)] \Rightarrow [\delta(s_0, \xi_1 V_1 \dots \xi_i \alpha_i x_i)] = q_j$$

Il suffit en effet de choisir : $k > l\mu^{(2p\alpha+1)}$. l étant un entier arbitraire nous pouvons prendre les entiers α, p et m et le mot w de $\Lambda[T(\alpha, n, m)]$ de telle façon que, en particulier, chaque facteur $V_i (i = 1, \dots, l)$, puisse se mettre sous la forme :

$$V_i = \rho_{i_1} \nu_i \rho_{i_2} \eta_i \rho_{i_3} \tilde{\eta}_i \rho_{i_4} \tilde{\nu}_i \rho_{i_5}$$

avec :

$$\nu_i \tilde{\nu}_i, \eta_i \tilde{\eta}_i \in D_n'^{*}$$

$$\rho_{i_3}, \rho_{i_2} \rho_{i_4} \in D_n'^{*}$$

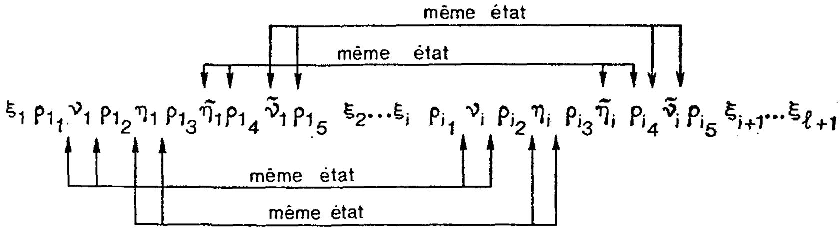
$$\delta(s_0, \xi_1 V_1 \xi_2 V_2 \dots V_{i-1} \xi_i \rho_{i_1}) = \delta(s_0, \xi_1 V_1 \dots V_{i-1} \xi_i \rho_{i_1} \nu_i)$$

$$\delta(s_0, \xi_1 V_1 \dots V_{i-1} \xi_i \rho_{i_1} \nu_i \rho_{i_2} \eta_i \rho_{i_3} \tilde{\eta}_i \rho_{i_4}) = \delta(s_0, \xi_1 V_1 \dots \rho_{i_4} \tilde{\nu}_i)$$

$$\delta(s_0, \xi_1 V_1 \dots V_{i-1} \xi_i \rho_{i_1} \nu_i \rho_{i_2}) = \delta(s_0, \xi_1 V_1 \dots \rho_{i_2} \eta_i)$$

$$\delta(s_0, \xi_1 V_1 \dots V_{i-1} \xi_i \rho_{i_1} \nu_i \rho_{i_2} \eta_i \rho_{i_3}) = \delta(s_0, \xi_1 V_1 \dots \rho_{i_3} \tilde{\eta}_i)$$

Ce que nous pouvons représenter schématiquement sous la forme :



Pour simplifier les notations, posons :

$$\begin{aligned} h_1 &= \xi_1 \rho_{1_1} \\ &\vdots \\ h_i &= \rho_{(i-1)_5} \xi_i \rho_{i_1} \qquad i = 2, \ldots, l \\ &\vdots \\ h_{l+1} &= \rho_{l_5} \xi_{l+1} \end{aligned}$$

Deux cas sont à envisager :

1er cas :

Il existe un entier $j (2 \leqslant j \leqslant l)$ tel que le mot h_j soit un mot de $D_n'^*$.
On pose alors :

$$\begin{cases} \delta(s_0, \xi_1 V_1 \ldots V_{(j-2)} \xi_{(j-1)} \rho_{(j-1)1}) = s \\ \delta(s_0, \xi_1 V_1 \ldots \rho_{(j-1)1} \nu_j \rho_{(j-1)2} \eta_j \rho_{(j-1)3} \tilde{\eta}_j \rho_{(j-1)4}) = \bar{s} \\ h_j = U_{\bar{s}s} \end{cases}$$

et l'on a :

$$\delta(\bar{s}, U_{\bar{s}s}) = s \quad \text{avec} \quad U_{\bar{s}s} \in D_n'^*$$

C'est-à-dire que l'on a mis en évidence l'existence d'un rationnel K' inclus dans K reconnu par un automate du type annoncé dans le lemme III.

2e cas :

Quel que soit l'entier $j (2 \leqslant j \leqslant l)$, le mot h_j n'est pas un mot de $D_n'^*$.

Or, d'après le lemme 1, pour tout entier k_0, il existe un entier l tel que le mot : $h_1 h_2 \ldots h_l$ soit ou bien de puissance supérieure à k_0, ou bien de hauteur supérieure à k_0.

Séparons en deux sous-cas :

1er sous-cas

S'il existe un choix de l'entier k_0, supérieur à $(\mu + 1)$, tel que le mot correspondant : $h_1 h_2 \dots h_l = h$, soit de puissance supérieure à k_0, alors, ce mot peut s'écrire : $h = h_1 h_2 \dots h_{(i-1)}(h_i' d_1 \tilde{d}_1 h_j'')h_{j+1} \dots h_l$ avec :

$$h_i = h_i' h_i'',\ h_j = h_j' h_j'',\ h_i'' = d_1,\ h_{(i+1)} \dots h_{(j-1)} h_j' = \tilde{d}_1,\ d_1 \tilde{d}_1 \in D_n'^*;$$

et de plus :

$$\delta(s_0, \xi_1 V_1 \dots \xi_i \rho_{i_1} \nu_i \dots \tilde{\nu}_i h_i') = \delta(s_0, \xi_1 V_1 \dots \xi_j \rho_{j_1} \nu_j \dots \tilde{\nu}_j h_j').$$

En effet, d'après l'hypothèse faite, le mot h peut s'écrire :

$$h = f_1 f_2 \dots f_{(\mu+1)} \quad \text{où} \quad f_i \in D_n'^*;\ i = 1, \dots, (\mu + 1)$$

Désignons par $f_i^{(u)}$ la dernière lettre du mot f_i et considérons la suite des états de l'automate fini $A(K)$, lisant le mot w :

$$\delta(s_0, \dots f_1^{(u)}),\ \delta(s_0, \dots f_2^{(u)}),\ \dots,\ \delta(s_0, \dots f_{(\mu+1)}^{(u)})$$

Deux éléments de cette suite sont identiques : $\delta(S_0, \dots f_i^{(u)}) = \delta(S_0, \dots f_j^{(u)})$ d'où la forme annoncée du mot h.

Posons maintenant :

$$\xi_1 V_1 \dots \xi_i \rho_i = U_{s_0 s'} d_1,\ \rho_{j5} \xi_{(j+1)} V_{(j+1)} \dots \xi_1 V_1 \xi_{(l+1)} = U_{rs_f}$$

$$\delta(s_0, U_{s_0 s'} d_1) = s,\ \delta(s, \nu_i \dots \tilde{\nu}_i) = \bar{s},\ \delta(\bar{s}, \tilde{d}_1) = r.$$

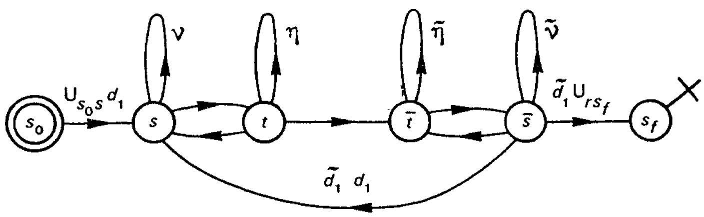
Nous avons ainsi mis en évidence l'existence de mots vérifiant :

$$\delta(s_0, U_{s_0 s'} d_1) = s,\ \delta(\bar{s}, \tilde{d}_1 d_1) = s \text{ (avec } (d_1 \tilde{d}_1) \in D_n'^*),$$

et

$$(\exists\, s_f \in F), \delta(\bar{s}, \tilde{d}_1 U_{rs_f}) = s_f \text{ (avec } (U_{s_0 s} U_{rs_f}) \in D_n'^*).$$

c'est-à-dire, en fait d'un rationnel K', inclus dans K, reconnu par l'automate fini représenté schématiquement de la manière suivante :



2^e sous-cas

Si l'hypothèse ci-dessus n'est pas vérifiée, alors, en choisissant $k_0 = (\mu + 1)^2 + 1$, on peut trouver un entier l tel que le mot h soit de puissance inférieure à k_0, mais de hauteur supérieure à k_0.

Dans ces conditions, h peut s'écrire :

$$h = h_1 h_2 \dots h_{(i-1)}(h_i' U_3 U_1 U_2 d U_4 h_j'') h_{(j+1)} \dots h_l$$

où : $h_i = h_i' U_3$, $(U_3 U_1 U_2 U_4) \in D_n'^*$, $d \in D_n'^*$, U_3 et U_2 sont équivalents (au sens de la réduction de Dyck) à des mots uniquement constitués d'occurrences de lettres non barrées, $\delta(s_0, \xi_1 V_1 \dots V_{(i-1)} h_i') = \delta(s_0, \xi_1 V_1 \dots V_{(i-1)} h_i' U_3 U_1 U_2)$.

En effet, d'après l'hypothèse faite, h s'écrit : $h = f_{(\mu+1)^2+1}$

où :

$$f_{(\mu+1)^2+1} \in D_n'^*$$

et,

$$\begin{array}{llll} f_{(\mu+1)^2+1} & = \alpha_{(\mu+1)^2} f_{(\mu+1)^2} \beta_{(\mu+1)^2} & & f_{(\mu+1)^2} \in D_n'^*, \ (\alpha_{(\mu+1)^2}\beta_{(\mu+1)^2}) \neq 1 \\ & \vdots & & \\ f_{i+1} & = \alpha_i \quad f_i \quad \beta_i & & f_i \in D_n'^*, \ \alpha_i\beta_i \neq 1 \\ & \vdots & & \\ f_2 & = \alpha_1 \quad f_1 \quad \beta_1 & & f_1 \in D_n'^*, \alpha_1\beta_1 \neq 1 \end{array}$$

Considérons la suite des états de l'automate fini $A(K)$, lisant w :

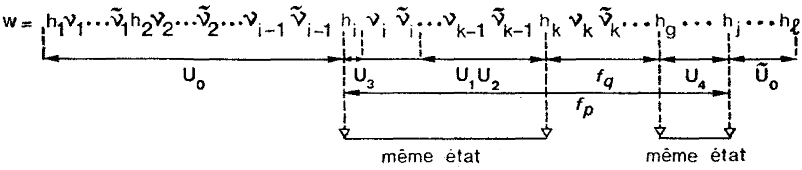
$$\delta(s_0, \dots f^{(u)}_{(\mu+1)^2+1}), \delta(s_0, \dots f^{(u)}_{(\mu+1)^2}), \dots, \delta(s_0, \dots f_i^{(u)}), \dots, \delta(s_0, \dots f_1{}^{(u)})$$

Il y a dans cette suite au moins $(\mu + 1)$ états identiques, correspondant aux indices : $k_1, k_2, \dots, k_{(\mu+1)}$.

Désignons par f_i^1 la première occurrence du mot f_i et considérons la suite des états de $A(K)$ lisant w :

$$\delta(s_0, \dots f^{(1)}_{k1}), \delta(s_0, \dots f^{(1)}_{k2}), \dots, \delta(s_0, \dots f^{(1)}_{k(\mu+1)})$$

Deux au moins de ces états sont identiques (correspondant aux indices p et q), d'où la forme annoncée de h.



Nous avons ainsi mis en évidence l'existence de mots, vérifiant :

$$\delta(s_0, U_0) = s', \delta(s', U_3) = s, \delta(\bar{s}, U_1U_2U_3) = s, \delta(\bar{s}, U_1U_2f_q) = r, \delta(r, U_4) = r$$

et :

$$\exists (s_f \in F)\, \delta(r, \tilde{U}_0) = s_f$$

avec :

$$(U_3U_1U_2U_4) \in D_n'^*, f_q \in D_n'^*, (U_0\tilde{U}_0) \in D_n'^*.$$

Finalement, et de manière générale, nous avons ainsi mis en évidence un rationnel K', inclus dans K, et reconnu par un automate fini du type annoncé dans le lemme IV.

Définissons maintenant l'application Φ de $C \times \Sigma_2^*$ dans Σ_2^* par :

$$(\forall\, w \in \Sigma_2^*)(\forall\, \alpha \in \mathbf{N}_+)\, \Phi(M_{\alpha,1}, w) = \sigma_1^\alpha\sigma_2^\alpha w\bar{\sigma}_2^\alpha\bar{\sigma}_1^\alpha$$

$$(\forall\, w \in \Sigma_2^*)(\forall\, \alpha \in \mathbf{N}_+)(\forall\, i \in \mathbf{N}_+)\, \Phi(M_{\alpha,(i+1)}, w) = \sigma_1^\alpha\sigma_2^\alpha\Phi(M_{\alpha,i}, w)\bar{\sigma}_2^\alpha\bar{\sigma}_1^\alpha$$

et posons :

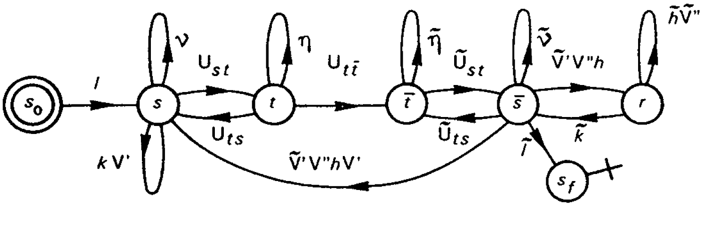
$$P_{\alpha,p,m,0} = T_{\alpha,p,m}$$

$$(\forall\, i \in \mathbf{N}_+) P_{\alpha,p,m,i} = [\Phi(M_{\alpha,p}, P_{\alpha,p,m,(i-1)})]^m$$

enfin, soit F la partie de $\Sigma_2^* : F = \{ P_{\alpha,p,m,i} | \alpha, p, m, i \in \mathbf{N}_+ \}$.

Lemme V

Si le langage algébrique standard $L = (D_n'^* \cap K)$ est tel qu'il existe une transduction rationnelle τ vérifiant $F \subseteq \tau(L) \subseteq D_2'^*$ alors il existe un langage rationnel K', inclus dans K, reconnu par un automate fini du type suivant :



avec :

$$l\tilde{l}, k\tilde{k}, V'\tilde{V}', V''\tilde{V}'', h\tilde{h}, U_{st}\tilde{U}_{st}, U_{ts}\tilde{U}_{ts}, U_{t\bar{t}} \in D_n'^*.$$

Preuve :

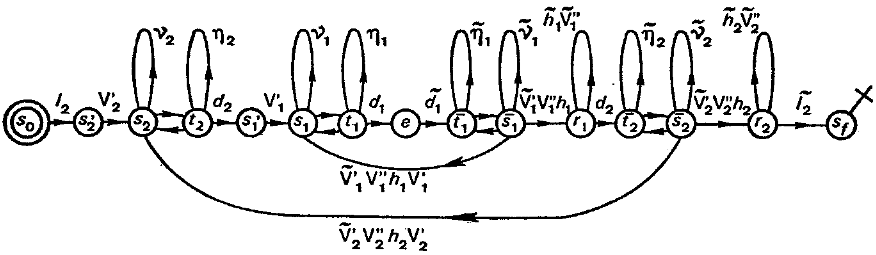
Considérons un mot w de $\Lambda(P_{\alpha,p,m,1})$:

$$w = w_1 w_1' w_1'' w_2 w_2' w_2'' \dots w_i w_i' w_i'' \dots w_m w_m' w_m'' \text{ où } \begin{cases} w_i' \in \Lambda\,(T_{\alpha,p,m}) \\ w_i w_i'' \in \Lambda\,(M_{\alpha,p}) \end{cases}$$

La construction du lemme IV appliquée aux mots :

$$w_1', w_2', \dots, w_m' \quad \text{et} \quad w_1 w_1'' w_2 w_2'' \dots w_m w_m''$$

permet la mise en évidence d'un rationnel K'', inclus dans K, reconnu par un automate fini du type suivant :

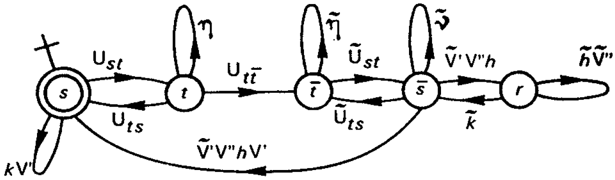


Il s'agit d'un automate obtenu en « emboîtant » deux automates du type de ceux définis au lemme IV. En considérant le mot $P_{\alpha,p,m,i}$, on met, de la même façon, en évidence, un rationnel $K^{(i+1)}$, inclus dans K, reconnu par un automate obtenu en « emboîtant « $(i+1)$ automates du type de ceux définis au lemme IV. Il suffit de remarquer que, en choisissant i suffisamment grand, on peut trouver deux tels automates correspondants aux mêmes quadruplets d'états $(s, \bar{s}, t, \bar{t})$, pour établir le lemme V.

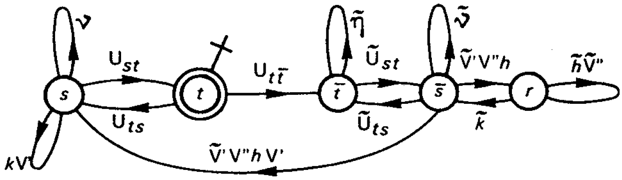
La construction faite en considérant la partie F s'applique à $D_2'^*$ et nous pouvons affirmer que tout langage algébrique standard qui s'envoie par transduction rationnelle sur $D_2'^*$, vérifie une propriété analogue à celle énoncée au lemme V. Dès lors, énonçons le :

Lemme VI

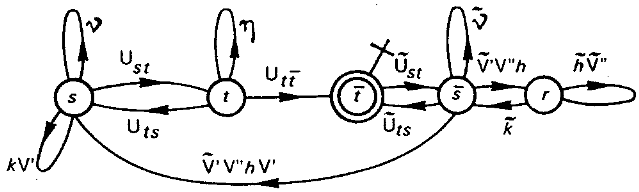
Si $K_1, K_2, \dots, K_8$ désignent les langages rationnels reconnus par les automates finis suivants, sous-automates de l'automate fini du lemme V (voir page suivante), A(K_1)



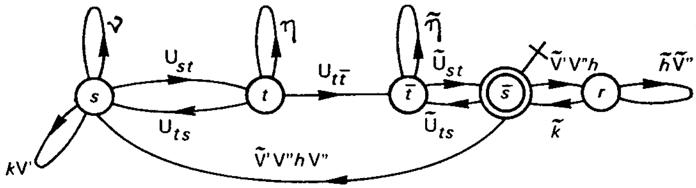
A(K_2)



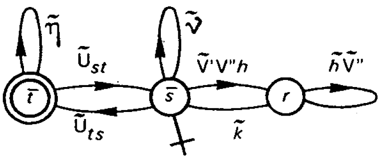
A(K_3)



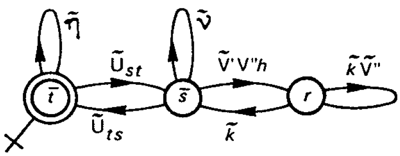
A(K_4)



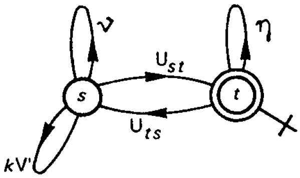
A(K_5)



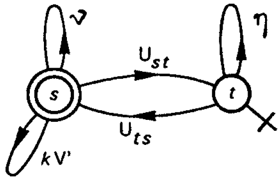
A(K_6)



A(K_7)



A(K_8)



alors, on a :

$$[[K_1 U_{st}\tilde{\eta}^* U_{t\bar{t}} K_5]\tilde{\nu}^+] \cap D_n'^* = \varnothing$$

$$[[K_2 U_{t\bar{t}} K_6]\tilde{\eta}^+] \cap D_n'^* = \varnothing$$

$$[\eta^+[K_7 U_{t\bar{t}} K_3]] \cap D_n'^* = \varnothing$$

$$[\nu^+[K_8 U_{t\bar{t}}\tilde{\eta}^* \tilde{U}_{st} K_5]] \cap D_n'^* = \varnothing$$

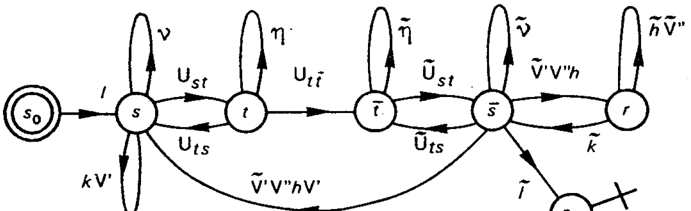
Preuve

Ce lemme résulte de la construction même effectuée aux lemmes II, III, IV, V, associée au fait qu'un mot de Dyck admet une réduction unique.

Les lemmes V et VI donnent donc une condition nécessaire pour qu'un langage algébrique standard soit générateur. Le fait qu'elle soit suffisante découle de propriétés bien connues des transductions rationnelles. Énonçons donc le :

Théorème I

Une condition nécessaire et suffisante pour que le langage algébrique standard $L = (D_n'^* \cap K)$ $(n > 1)$ soit générateur est que le langage K contienne un langage rationnel K', reconnu par un automate fini du type suivant :



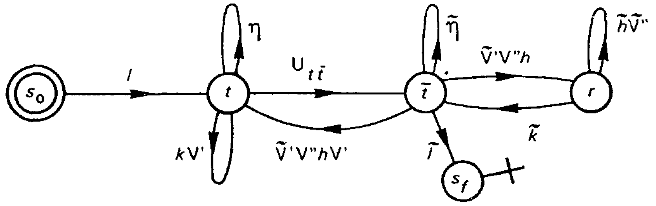
satisfaisant le lemme VI avec :

$$l\tilde{l},\ k\tilde{k},\ U_{st}\tilde{U}_{st},\ U_{ts}\tilde{U}_{ts},\ \nu\tilde{\nu},\ \eta\tilde{\eta},\ h\tilde{h},\ U_{t\bar{t}} \in D_n'^*$$

Si on remplace $D_2'^*$ par $D_1'^*$, une construction analogue fournit le :

Théorème II

Une condition nécessaire et suffisante pour que le langage algébrique standard $L = (D_n'^* \cap K)(n > 1)$ s'envoie par transduction rationnelle sur le langage $D_1'^*$ est que le langage rationnel K contienne un langage rationnel K', reconnu par un automate fini du type suivant :



satisfaisant le lemme VI avec :

$$l\tilde{l},\ k\tilde{k},\ h\tilde{h},\ V'\tilde{V}',\ V''\tilde{V}'',\ \eta\tilde{\eta}, U_{t\tilde{t}} \in D_n'^*$$

4. EXEMPLE D'APPLICATION

Proposition

Le langage algébrique L, engendré par la grammaire :

$$G \begin{cases} S \to aSSb \\ S \to d \end{cases}$$

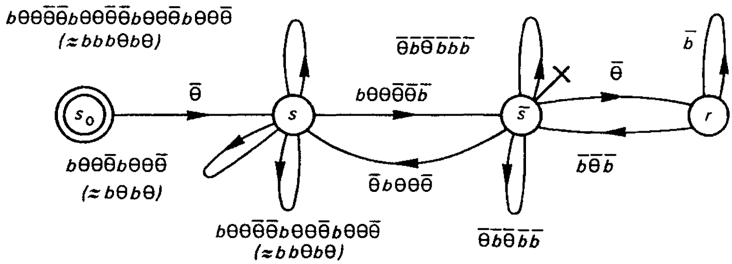
est un générateur du cône rationnel des langages algébriques.

Preuve :

Soit $\Sigma_2 = \{ \theta, b, \bar{\theta}, \bar{b} \}$, $D_2'^*$ le langage de Dyck construit sur Σ_2^* et K le langage rationnel : $K = \{ \bar{\theta} b\theta\theta, \bar{\theta}, \bar{b} \}^*$.

D'après une propriété connue [7], le langage L est rationnellement équivalent au langage $L' = (\bar{\theta} D_2'^* \cap K)$ (c'est-à-dire qu'il existe deux transductions rationnelles τ et τ' telles que : $\tau(L) = L'$ et $\tau'(L') = L$).

Considérons le rationnel K', reconnu par l'automate fini suivant :



Clairement, K' est inclus dans K et d'après le théorème I, L' est générateur. L l'est donc aussi.

BIBLIOGRAPHIE

[1] BEAUQUIER J., Thèse de 3e Cycle, Paris VII, 1973.

[2] BOASSON L., Thèse de 3e Cycle, Paris VII, 1971.

[3] CHOMSKY N. et SCHUTZENBERGER M. P., *The Algebraic Theory of Context-Free Languages*, Computer Programming and Formal Systems, North-Holland Publishing Co., Amsterdam, 1963.

[4] GINSBURG S. et GREIBACH S., *Principal A.F.L.*, Journal Comp. System Sciences, *4*, 1970.

[5] GINSBURG S., *The Mathematical Theory of Context-Free Languages*, Mac Graw Hill, 1966.

[6] GINZBURG A., *Algebraic Theory of Automata*, A.P. New York, London, 1968.

[7] NIVAT M., *Transductions de langages de Chomsky*, Ann. Inst. Fourier, Grenoble, 1968.

[8] SCHUTZENBERGER M. P., *Sur un langage équivalent au langage de Dyck* (non publié).