RECURSIVITE ET CONES RATIONNELS FERMES PAR INTERSECTION

A. ARNOLD (1) - M. LATTEUX (1)

RESUME - Nous définissons les grammaires quasi-commutatives en modifiant légérement la definition des grammaires commutatives introduites par Crespi-Reghizzi et Mandrioli [5]. Nous obtenons, ainsi, une caractérisation grammaticale de $\mathcal{C}_\cap(D_1'^*)$, le plus petit cône rationnel clos par intersection et contenant $D_1'^*$, le langage de semi-Dyck sur une lettre. Nous en déduisons que tout langage de $\mathcal{C}_\cap(D_1'^*)$ est récursif. A l'aide d'une généralisation des « Vector Addition Systems », nous démontrons, ensuite, le même résultat pour les langages de $\mathcal{F}_\cap(\mathrm{Init}(D_1'^*))$, le plus petit cône rationnel clos par étoile et intersection, contenant le langage formé de tous les préfixes des mots de D_1'.

Introduction.

Dans bien des cas, pour une famille de languages $\mathcal{L}$, $\mathcal{C}_\cap(\mathcal{L})$, la clôture de $\mathcal{L}$ par transduction rationnelle et intersection est une famille beaucoup plus vaste que le cône rationnel engendré par $\mathcal{L}$, noté $\mathcal{C}(\mathcal{L})$, qui est la clôture de $\mathcal{L}$ par transduction rationnelle. Ainsi, Rat, la famille des languages rationnels est le seul cône rationnel clos par intersection, contenu dans la famille des langages algébriques (« context-free ») [13]. De même Baker et Book ont montré [3] que tout langage récursivement énumérable appartenait à $\mathcal{C}_\cap(\mathrm{Lin})$ où Lin désigne la famille des langages algébriques linéaires. Il en est de même pour $\mathcal{C}_\cap(C_1^*)$ où C_1 est le langage $\{a^n b^n / n \geq 0\}$ [9]. Par contre, dans [12], il était montré que $\mathcal{C}_\cap(C_1)$ était égal au cône rationnel engendré par la fermeture commutative des langages rationnels et était inclus dans la famille des langages à contexte lié. De plus, il était montré que $D'_1{}^*$ n'appartenait pas à $\mathcal{C}_\cap(D_1^*)=$

Reçu 2 Février 1977.
(1) Université de Lille I.U.E.R. I.E.E.A. Service Informatique Bat. M3 - BP 36, 59650 Villeneuve d'Ascq - France.

$= \mathcal{C}_\cap (C_1)$ où $D_1'^*$ (resp. D_1^*) désigne le langage de semi-Dyck (resp. de Dyck) sur une lettre, c'est-à-dire, la classe du mot vide ε dans la congruence engendrée par $a_1 \bar{a}_1 = \varepsilon$ (resp. $a_1 \bar{a}_1 = \bar{a}_1 a_1 = \varepsilon$). C'est la famille $\mathcal{C}_\cap (D_1'^*)$ qui nous intéresse en premier lieu.

En modifiant légérement la définition des grammaires commutatives de Crespi-Reghizzi et Mandrioli [5], nous introduisons la notion de grammaire quasi-commutative qui généralise aussi celle de « label-grammar » due à Höpner et Opp [10]. Nous obtenons, alors, une caractérisation grammaticale de la famille $\mathcal{C}_\cap (D_1'^*)$. En utilisant le résultat de Sacerdote et Tenney [15] sur la décidabilité du « reachability problem » pour les *VAS*, nous en déduisons que tout langage appartenant à $\mathcal{C}_\cap (D_1'^*)$ est récursif.

D'autre part, dans [8], Ginsburg et Goldstine se posaient la question de savoir si pour tout langage non rationnel L, $\mathcal{F}_\cap (L)$ la plus petite *FAL* contenant L et fermée par intersection contenait tous les langages récursivement énumérables. Ils répondaient négativement à cette question en donnant une condition suffisante pour qu'un langage non rationnel $L \subseteq a^*$ soit tel que $\mathcal{F}_\cap (L)$ ne contienne pas tous les langages récursivement énumérables. Clairement, les langages non rationnels inclus dans a^* ne sont pas algébriques. Nous allons, au contraire, considérer les langages algébriques non rationnels et montrer l'existence d'un tel langage L qui vérifie: tout langage de $\mathcal{F}_\cap (L)$ est récursif.

Pour obtenir ce résultat nous introduisons une généralisation des « Vector Addition System » (*VAS*) les *VAS* avec effacement (*VASE*), pour lesquels il existe un symbole spécial qui permet d'effacer une composante d'un vecteur. Pour les *VASE*, nou montrons que le « reachability problem » est indécidable. Par contre, nous montrons qu'un autre problème, dont la décidabilité avait été démontrée pour les *VAS* par Karp et Miller [11] reste décidable pour les *VASE*. C'est ce résultat qui nous permet de montrer que tout langage appartenant à $\mathcal{F}_\cap$ (Init $(D_1'^*)$) est récursif (Init $(D_1'^*)$ est formé des préfixes des mots de $D_1'^*$).

1. Préliminaires.

Soix X un ensemble fini, ou alphabet. Nous notons X^* le monoïde libre engendré par X et ε le mot vide de X^*.

Un omomorphisme h de X^* dans Y^* est dit *alphabétique* si pour tout $x \in X$, $h(x) \in Y \cup \{\varepsilon\}$. Une *transduction rationnelle* τ de X^* dans Y^* est une application de X^* dans $\mathcal{P}(Y^*)$ qui vérifie: il existe un alphabet Z, un langage rationnel $R \subseteq Z^*$ et deux homomorphismes h et g de Z^* dans respectivement X^* et Y^* tels que pour tout mot u de X^* $\tau(u) = g(h^{-1}(u) \cap R)$.

Une transduction rationnelle τ de X^* dans Y^* peut être étendue en une application, notée également τ, de $\mathcal{P}(X^*)$ dans $\mathcal{P}(Y^*)$ par $\tau(L) = \bigcup_{u \in L} \tau(u)$.

Une famille de langages est appelée *cône rationnel* si elle est fermée par transduction rationnelle.

Pour toute famille de langages $\mathcal{L}$, $\mathcal{C}(\mathcal{L})$ désigne le plus petit cône rationnel contenant $\mathcal{L}$. Si $\mathcal{L}=\{L\}$, on dira que le cône rationnel $\mathcal{C}(\mathcal{L})$ est *principal* et on le notera aussi $\mathcal{C}(L)$.

THÉORÈME 1 [14]. *Soient L et L' deux langages inclus respectivement dans X^* et Y^*. Alors $L' \in \mathcal{C}(L)$ si il existe un alphabet Z, un langage rationnel $R \subseteq Z^*$ et deux homomorphismes alphabétiques h et g de Z^* dans respectivement X^* et Y^* tels que $L'=g(h^{-1}(L) \cap R)$.*

Une *famille agréable de langages* (FAL) est cône rationnel fermé par union, produit et étoile. Pour toute famille de languages $\mathcal{L}$, $\mathcal{F}(\mathcal{L})$ désigne la plus petite FAL contenant $\mathcal{L}$. Si $\mathcal{L}=\{L\}$, la FAL $\mathcal{F}(\mathcal{L})$, notée aussi $\mathcal{F}(L)$, est dite *principale.*

Pour tout langage $L \subseteq X^*$, Init (L) est le langage de X^* formé de tous les facteurs gauches de mots de L, c'est-à-dire $\{u \in X^* / \exists\, v \in X^* \text{ tq } uv \in L\}$.

Le « *shuffle* » de deux langages $L_1 \subseteq X^*$ et $L_2 \subseteq Y^*$ est le langage noté $L_1 \sqcup\!\sqcup L_2$ de $(X \cup Y)^*$ égal à $\{u_1 v_1 \ldots u_n v_n / u_i \in X^*,\ v_i \in Y^*,\ u_1 \ldots u_n \in L_1,\ v_1 \ldots v_n \in L_2\}$.

Il est clair que $\text{Init}(L_1 \sqcup\!\sqcup L_2) = \text{Init}(L_1) \sqcup\!\sqcup \text{Init}(L_2)$.

Pour tout langage $L_1 \subseteq X_1^*$, notons $\mathcal{C}_\cap(L_1)$ le plus petit cône rationnel fermé par intersection contenant L_1. Pour tout entier $i>1$ construisons par induction un alphabet X_i, et un langage $L_i \subseteq X_i^*$. Prenons Y_i un alphabet disjoint de X_{i-1} qui soit en bijection avec X_1 par un homomorphisme h_i; posons $X_i = X_{i-1} \cup Y_i$ et $L_i = L_{i-1} \sqcup\!\sqcup h_i(L_1)$. La famille $\mathcal{L}=\{L_i / i \geq 1\}$ vérifie:

PROPOSITION 1. [7]. $\mathcal{C}_\cap(L_1) = \mathcal{C}(\mathcal{L}) = \bigcup\limits_{i \geq 1} \mathcal{C}(L_i)$.

Prenons $X_1=\{a_1, \bar{a}_1\}$ et $D_1'^*$ le langage de semi-Dyck sur une lettre, c'est-à-dire la classe d'équivalence du mot vide ε dans la congruence sur X_1^* engendrée par $a_1 \bar{a}_1 = \varepsilon$. Pour tout $k>1$ posons $Y_k=\{a_k, \bar{a}_k\}$ et $X_k=\{a_j, \bar{a}_j / 1 \leq j \leq k\} = X_{k-1} \cup Y_k$; définissons l'homomorphisme bijectif h_k de X_1^* dans Y_k^* par $h_k(\bar{a}_1)=\bar{a}_k$, $h_k(a_1)=a_k$.

Alors d'après la proposition 1, $\mathcal{C}_\cap(D_1'^*) = \bigcup\limits_{k \geq 1} \mathcal{C}(O_k')$ où O_k' est le langage défini inductivement par $O_1' = D_1'^*$, $O'_{k+1} = O_k' \sqcup\!\sqcup h_{k+1}(O_1')$.

Il est facile de voir que pour tout $k \geq 1$ nous avons $O_k' = \{w \in X_k^* / \forall j \in \{1, \ldots, k\},\ l_{a_j}(w) = l_{\bar{a}_j}(w)$ et si $w=w'w''$, $l_{a_j}(w') \geq l_{\bar{a}_j}(w')\}$ où pour $b \in X_k$, $l_b(w)$est le nombre d'occurrences de la lettre b dans le mot w.

Prenons maintenant pour $k \geq 1$ $Y_k=\{a_k, \bar{a}_k, c_k\}$ et $X_k=\{a_i, \bar{a}_i, c_i / 1 \leq i \leq k\}$, d'où $X_{k+1} = X_k \cup Y_{k+1}$. Soit g_k l'homomorphisme bijectif de X_1^* dans Y_k^* défini par $g\ (a_1)=a_k$, $g_k(a_1)=a_k$, $\overline{g}_k(c_1)=c_k$. Soit I le langage $(\mathrm{Init}\,(D_1'^*)\,c_1)^*\,D_1'^*$. Alors d'aprés la proposition 1, $\mathcal{C}_\cap\ (I)$ est égal à $\bigcup_{k\geq 1} \mathcal{C}\,(I_k)$ où I_k est défini inductivement par:

$$I_1=I,\ I_{k+1}=I_k \sqcup\!\sqcup\ g_{k+1}(I_1).$$

Comme $D_1'^*=I\cap\{a_1,\overline{a}_k\}^*$, le langage $D_1'^*$ appartient à $\mathcal{C}\,(I)\subset\mathcal{C}_\cap\,(I)$, ainsi que le langage $h_c^{-1}(D_1'^*)$ où h_c est l'homomorphisme de $\{a_1,\overline{a}_1,c_1\}^*$ dans $\{a_1,\overline{a}_1\}^*$ défini par $h_c(a_1)=\overline{a}_1, h_c(\overline{a}_1)=\overline{a}_1, h_c(c_1)=\varepsilon$. Nous obtenons donc $(D_1'^*\,c_1)^*\,D_1'^*=I\cap h_c^{-1}(D_1'^*)\in\mathcal{C}_\cap\,(I)$.

Or

$$(D_1'^*\,c)^*\,D_1'^*\cap(a_1^*\,\overline{a}_1^*\,c)^*\,a_1^*\,\overline{a}_1^*=(\{a_1^n\,\overline{a}_1^n/n\geq 0\}\,c)^*\ \{a_1^n\,\overline{a}_1^n/n\geq 0\}$$

appartient encore à $\mathcal{C}_\cap\,(I)$.

Nous en déduisons que $\mathcal{C}_\cap\,(\{a^n\,b^n/n\geq 0\}^*)$ est inclus dans $\mathcal{C}_\cap\,(I)$ et comme $\mathcal{C}_\cap(\{a^n\,b^n/n\geq 0\}^*)=\mathcal{F}_\cap\,(\{a^n\,b^n/n\geq 0\})$ [7] et que tout langage récursivement énumerable appartient à $\mathcal{F}_\cap\,(\{a^n\,b^n/n\geq 0\})$ [9], nous avons:

LEMME 1. *Tout langage récursivement énumérable appartient à* $\mathcal{C}_\cap\,(I)$.

De plus nous savons [7] que $\mathcal{F}_\cap(\mathrm{Init}\,(D_1'^*))=\mathcal{C}_\cap((\mathrm{Init}\,(D_1'^*)\,c)^*=\mathcal{C}_\cap(J)$ où $J=(\mathrm{Init}\,(D_1'^*)\,c)^*\,\mathrm{Init}\,(D_1'^*)=\mathrm{Init}\,(I)$. D'après la proposition 1, ceci est encore égal a $\bigcup_{k\geq 1}\mathcal{C}\,(J_k)$, où, puisque les opérations « shuffle » et Init commutent, pour tout $k\geq 1$, $J_k=\mathrm{Init}\,(I_k)$.

Enfin, il est bien connu que pour les cônes rationnels certains problèmes de décidabilité sont équivalents.

THÉORÈME 2. *Soit* $\mathcal{L}$ *une famille de langages.*

Les propriétés suivantes sont équivalentes:

i) *tout langage de* $\mathcal{C}\,(\mathcal{L})$ *est récursif*

ii) *il est décidable de savoir si un langage de* $\mathcal{C}\,(\mathcal{L})$ *est vide*

iii) *il est décidable de savoir si l'intersection d'un langage de* $\mathcal{L}$ *et d'un langage rationnel est vide.*

DÉMONSTRATION. Puisque $x\in L$ ssi $\{x\}\cap L\neq\varnothing$ et que $\{x\}\cap L\in\mathcal{C}\,(\mathcal{L})$ quand $L\in\mathcal{C}\,(\mathcal{L})$, on a bien (ii) $\Rightarrow$ (i). Si maintenant h est l'homomorphisme qui à tout mot associe le mot vide a $L\neq\varnothing$ ssi $\varepsilon\in h\,(L)$ et comme $h\,(L)\in\mathcal{C}\,(\mathcal{L})$ quand $L\in\mathcal{C}\,(\mathcal{L})$ (i) $\Rightarrow$ (ii). Si $L\in\mathcal{L}$ et R est un langage rationnel $L\cap R\in\mathcal{C}\,(\mathcal{L})$ d'où (ii) $\Rightarrow$ (iii).

Si $L \in \mathcal{C}(\mathcal{L})$ d'après le théorème 1 il existe un langage rationnel R et deux homomorphismes h et g tels que $L = g(h^{-1}(L') \cap R)$. Il est clair que $L \neq \emptyset$ ssi $h^{-1}(L') \cap R \neq \emptyset$ ssi $L' \cap h(R) \neq \emptyset$. Comme $h(R)$ est encore un langage rationnel, (iii) $\Rightarrow$ (ii). $\square$

II. La famille $\mathcal{C}_\cap(D_1'^*)$.

DÉFINITIONS. Soit V un alphabet fini. Nous appellerons *multiplicité* sur V toute application de V dans N. A toute partie U de V nous associons la multiplicité notée également U définie par $\forall A \in V,\ U(A) = \begin{cases} 1 & \text{si } A \in U \\ 0 & \text{sinon.} \end{cases}$

Une *grammaire quasi-commutative* est un quadruplet $G = \langle V, X, P, S \rangle$ où V est un ensemble fini, l'alphabet non terminal, X est un ensemble fini, l'alphabet terminal, S est un élément de V, l'axiome et P est un ensemble fini de règles de la forme $u \to \langle w, v \rangle$ où u et v sont des multiplicités sur V et w un mot de X^*.

Si w_1 et w_2 sont des mots de X^* et u_1 et u_2 sont des multiplicités sur V nous dirons que $\langle w_1, u_1 \rangle$ se dérive immédiatement en $\langle w_2, u_2 \rangle$, ce qui sera noté $\langle w_1, u_1 \rangle \Rightarrow \langle w_2, u_2 \rangle$ ssi il existe une règle $u \to \langle w, u \rangle$ telle que

- $w_2 = w_1 \cdot w$
- $\forall A \in V,\ u_1(A) \geq u(A)$

et

$$u_2(A) = u_1(A) - u(A) + v(A).$$

Le langage $L(G)$ engendré par la grammaire quasi-commutative G est $\{ w \in X^* / \langle \Lambda, \{S\} \rangle \overset{*}{\Rightarrow} \langle w, \emptyset \rangle \}$. Nous noterons $\mathcal{L}_{QC}$ la famille des langages engendrés par des grammaires quasi-commutatives.

Si dans la définition des grammaires quasi-commutatives on remplace la condition que w est un mot de X^* par « w est une multiplicité sur X » on retrouve la définition des grammaires commutatives donnée par Crespi-Reghizzi et Mandrioli [5]. En notant $\mathcal{L}_{CG}$ la famille des ensembles de multiplicités engendrés par une grammaire commutative [5], on a alors clairement $\psi(\mathcal{L}_{CG}) = \psi(\mathcal{L}_{QC})$, ou ψ est la fonction de Parikh, qu'on peut définir aussi sur les multiplicités [5].

Par ailleurs si on ne considère que les grammaires quasi-commutatives dont les règles sont de la forme $\{A\} \to \langle w, v \rangle$ avec $w \in X \cup \{\varepsilon\}$ et $A \in V$, on obtient les « labeled grammars » de [10] et donc tout langage engendré par une « labeled grammar » est dans $\mathcal{L}_{QC}$.

Nous montrons que $\mathcal{L}_{QC}$ est égale à $\mathcal{C}_\cap(D_1'^*)$, ce qui fournit une caractérisation grammaticale de cette famille de langages.

THÉORÈME 3.

$$\mathcal{L}_{QC}=\mathcal{C}_\cap(D_1'^*).$$

DÉMONSTRATION.

A Soit $L=L(G)\in\mathcal{L}_{QC}$, où $G=\langle V, X, P, S\rangle$.

Nous pouvons supposer que $V=\{a_1, \dots, a_k\}$,

$$X=\{a_{k+1}, \dots, a_{k+n}\} \text{ et } S=a_1 .$$

A chaque multiplicité u sur V nous associons les mots $\overline{\text{Code}}(u)=\bar{a}_1^{n_1}\dots\bar{a}_k^{n_k}$ et $\text{Code}(u)=a_1^{n_1}\dots a_k^{n_k}$, où $n_i=u(a_i)$; et à chaque mot $w=a_{i_1}\dots a_{i_p}$ de X^* nous associons le mot $\text{Code}(w)=a_{i_1}\bar{a}_{i_1}\dots a_{i_p}\bar{a}_{i_p}$.

A la règle $r=u\to\langle w, v\rangle$ nous associons le mot $\text{Code}(r)=\overline{\text{Code}}(u)\ \text{Code}(v)\ \text{Code}(w)$.

Nous considérons enfin l'homomorphisme φ de $\{\bar{a}_i, a_i/1\leq i\leq k+n\}^*$ dans X^* défini par

$$\varphi(a_i)=\varphi(\bar{a}_i)=\varepsilon \quad \text{si} \quad i\leq k$$

$$\left.\begin{array}{l}\varphi(a_i)=a_i\\ \varphi(\bar{a}_i)=\varepsilon\end{array}\right\} \quad \text{si} \quad k<i\leq k+n .$$

Il découle des diverses définitions que $\langle\varepsilon,\{a_1\}\rangle \underset{r_1}{\Rightarrow} \langle w_1, u_1\rangle \dots \underset{r_p}{\Rightarrow} \langle w, \varnothing\rangle$ ssi $y=a_1\,\text{Code}(r_1)\cdot\text{Code}(r_2)\dots\text{Code}(r_p)\in O'_{k+n}$ et $w=\varphi(y)$.

Nous obtenons donc $L=\varphi(O'_{k+n}\cap a_1\{\text{Code}(r)/r\in P\}^*)$ et donc $L\in\mathcal{C}_\cap(D_1'^*)$.

B Soit $L\in\mathcal{C}_\cap(D_1'^*)$ et supposons que $L\subseteq D^*=\{d_1, d_2, \dots, d_n\}^*$. D'après la proposition 1 et le théorème 1 il existe un entier $k\geq 1$, un alphabet $B=\{b_1, \dots, b_p\}$, un langage rationnel $R\subseteq B^*$, un homomorphisme alphabétique $g\colon B^*\to D^*$ et un homomorphisme alphabétique $h\colon B^*\to\{a_i, \bar{a}_i/1\leq i\leq k\}^*$ tels que $L=g(h^{-1}(O'_k)\cap R)$.

Supposons que R soit reconnu par l'automate déterministe $\langle Q, B, \delta, q_0, F\rangle$. Nous construisons la grammaire quasi-commutative $G=\langle Q\cup\{a_i/1\leq i\leq k\}\rangle$, $D, P, q_0\}$ dont les règles sont

α) $\{q\}\to\langle\varepsilon, \varnothing\rangle$ si $q\in F$

β) pour tout $b_i \in B$, $q, q' \in Q$ tels que $\delta(q, b_i) = q'$

$$\cdot \{q\} \to \langle g(b_i), \{q', a_j\} \rangle \text{ si } h(b_i) = a_j$$

$$\cdot \{q\} \to \langle g(b_i), \{q'\} \rangle \text{ si } h(b_i) = \varepsilon$$

$$\cdot \{q, a_j\} \to \langle g(b_i), \{q'\} \rangle \text{ si } h(b_i) = \bar{a}_j .$$

Dans cette grammaire chaque dérivation terminale à partir de $\langle \varepsilon, \{q_0\} \rangle$ contient une et une seule règle de la forme α qui est la dernière; les autres règles sont entièrement caractérisées par le triplet $\langle q, b_i, q' \rangle$ tel que $q' = \delta(q, b_i)$. Il existe donc une bijection entre l'ensemble des dérivations terminales et l'ensemble des séquences $q_{j_1} b_{i_1} q_{j_2} b_{i_2} \dots q_{j_n} q_{i_n} q_{j_{n+1}}$ telles que

$$h(b_{i_1} \dots b_{i_n}) \in O_k' \text{ et } q_{j_1} = q_0, \; q_{i_{n+1}} \in F, \quad \forall l \leq n \; (q_{j_l}, b_{i_l}) = q_{j_{l+1}}$$

— et donc $b_{i_1} \dots b_{i_n} \in R$. Le mot engendré par ces dérivations est $g(b_{i_1} \dots b_{i_n})$ On a donc bien $L(G) = g(h^{-1}(O_k') \cap R) = L$. □

D'aprés le théorème 2 tout langage de $\mathcal{C}_\cap(D_1'^*)$ est récursif ssi la propriété $O_k' \cap R = \emptyset$ est décidable pour tout k et tout langage rationnel R. Nous avons montré dans [2] que cette propriété était équivalente à la décidabilité du « Reachability Problem » était équivalente à la décidabilité de $L \cap R = \emptyset$ quand Reghizzi et Mandrioli ont montré de leur côté [6] que la décidabilité du « Reachability Problem » était équivalente à la décidability de $L \cap R = \emptyset$ quand L est un langage de Szilard et R un langage rationnel. Or comme pour tout k le langage O_k'. **S** où **S** est un marqueur, est un langage de Szilard, on a $\mathcal{C}_\cap(D_1'^*) = \mathcal{C}(\{O_k'/k \geq 1\}) = \mathcal{C}(\{O_k' \cdot \mathbf{S}/k \geq 1\}) \subseteq \mathcal{C}(\mathcal{S}_z)$ où $\mathcal{S}_z$ est la famille des langages de Szilard. Par ailleurs tout langage de Szilard étant engendré par une « labeled grammar » [10] on a $\mathcal{S}_z \subseteq \mathcal{L}_{QC} = \mathcal{C}_\cap(D_1'^*)$ et donc $\mathcal{C}(\mathcal{S}_z) \subseteq \mathcal{C}_\cap(D_1'^*)$. Il en résulte que $\mathcal{C}(\mathcal{S}_z) = \mathcal{C}(\{O_k'/k \geq 1\})$. On voit alors que l'équivalence de la décidabilité des deux problèmes $L \cap R = \emptyset$ pour $L = O_k'$ ou $L \in \mathcal{S}_z$ est une conséquence immédiate du théorème 2.

Récemment Sacerdote et Tenney ont montré que le « Reachability problem » était décidable [15], d'où

THÉORÈME 4. *Tout langage de* $\mathcal{C}_\cap(D_1'^*)$ *est récursif.*

III. La famille $\mathcal{F}_\cap(\text{Init}(D_1'^*))$.

De même que la famille $\mathcal{C}_\cap(D_1'^*)$ a certaines relations avec les systèmes d'additions de vecteurs [2, 6], la famille $\mathcal{F}_\cap(\text{Init}(D_1'^*)) = \mathcal{C}_\cap((\text{Init}(D_1'^*)\, c)^*)$ est à rapprocher de systèmes d'additions de vecteurs où l'on a introduit une possibilité de remise à zéro. Cette extension des systèmes d'additions de vecteurs que nous allons définir plus loin correspond grosso-modo au phénomène suivant: si l'on dispose d'un automate à pile reconnaissant $D_1'^*$ on peut le transformer en un automate qui reconnait $(\text{Init}\,(D_1')^*\,c)^*$ en ajoutant des mouvements qui réinitialisent la pile lorsque l'automate lit en entrée le symbole c.

Définition. Soit $\#$ un nouveau symbole. On pose $\mathbf{Z}_1=\mathbf{Z}\cup\{\#\}$. On étend l'addition sur $\mathbf{Z}$ en une addition partielle sur $\mathbf{Z}_1$ par

$$\forall x\in\mathbf{Z},\quad x+\#=0.$$

On remarquera que cette addition n'est plus associative puisque $(x+\#)+y=y$ et que $x+(\#+y)$ n'est pas défini.

Un *système d'addition de vecteurs avec effacement* (VASE), de dimension n, est un couple $\langle W, x_0\rangle$ où W est un ensemble fini de vecteur de $\mathbf{Z}_1^n$ et x_0 un élément de $\mathbf{N}^n$.

L'ensemble engendré par un VASE $\langle W, x_0\rangle$, noté $R\,(W, x_0)$ est défini par:

$y\in R\,(W, x_0)$ ssi il existe une séquence $y_0, y_1, \dots, y_k$ d'éléments de $\mathbf{N}^n$ telle que $y_0=x_0$, $y_k=y$ et $\forall i\in\{0,\dots,k-1\}$

$$\exists\, w\in W \text{ t. q. } y_{i+1}=y_i+w.$$

Si on impose à un VASE $\langle W, x_0\rangle$ la restriction que $W\subseteq\mathbf{Z}^n$, on retrouve alors les VAS et la définition habituelle de leur « Reachability Set » [11].

Nous allons montrer maintenant que pour tout VASE $\langle W, x_0\rangle$, de dimension n, et pour tout $y_0\in\mathbf{N}^n$, il est décidable de savoir s'il existe $y\in R\,(W, x_0)$ tel que $y\geq y_0$.

Dans le cas des VAS, Karp et Miller [11] démontrent le même résultat en examinant un arbre fini construit à partir de W et x_0 et ne dépendant pas de y. Cette méthode n'est plus applicable ici à cause de l'effacement. Cependant nous allons procéder de manière analogue en construisant encore un arbre fini, mais à partir de W et y.

Nous définissons d'abord l'application partielle, notée $\dot{-}$ de $\mathbf{N}\times\mathbf{Z}_1$ dans $\mathbf{N}$ par

$$y\dot{-}w=\begin{cases}\max\,(y-w, 0) \text{ si } w\neq\#\\ 0 \text{ si } w=\# \text{ et } y=0\\ \text{indéfini sinon.}\end{cases}$$

Cette opération partielle est étendue composante par composante en une application partielle, notée encore $\dot{-}$, de $\mathbf{N}^n\times\mathbf{Z}_1^n$, dans $\mathbf{N}^n$.

Lemme 2. *Soient* $x\in\mathbf{N}^n$, $y\in\mathbf{N}^n$ et $w\in\mathbf{Z}_1^n$.
Alors $y\leq x+w$ *ssi* $y\dot{-}w$ *défini et* $y\dot{-}w\leq x$.

DÉMONSTRATION. Comme l'ordre sur $\mathbf{N}$ est aussi étendu composante par composante à $\mathbf{N}^n$, il suffit de montrer le résultat quand $n=1$.

— si $w \neq \#$, $y \dot{-} w$ est toujours défini et vaut $\max(y-w, 0)$. On a donc

$\max(y-w, 0) \leq x$ ssi $y-w \leq x$ ssi $y \leq x+w$.

— si $w=\#$, $y \leq x+\#$ ssi $y=0$ ssi $y \dot{-} \#$ est défini. De plus si $y \dot{-} \#$ est défini il vaut 0 et donc $y \dot{-} \# \leq x$. □

LEMME 3. *Soient y et $y' \in \mathbf{N}^n$, $w \in \mathbf{Z}_1{}^n$.*
Si $y \leq y'$ et si $y' \dot{-} w$ est défini alors $y \dot{-} w$ est défini et $y \dot{-} w \leq y \dot{-} w$.

DÉMONSTRATION. Posons $x=y' \dot{-} w$. D'aprés le lemme 2, $y' \leq x+w$ et comme $y \leq y'$ on a $y \leq x+w$ et encore d'aprés le lemme 2, $y \dot{-} w$ est défini et $y \dot{-} w \leq x = y' \dot{-} w$. □

LEMME 4. *Soit $\langle W, x_0 \rangle$ un VASE de dimension n et soit $y_0 \in \mathbf{N}^n$. Il existe $y \in R(W, x_0)$ tel que $y \geq y_0$ ssi il existe une séquence $y_0, y_1, \dots, y_k$ d'éléments de $\mathbf{N}^n$ telle que*

1) $\forall i \leq k \; \exists \; w_i \in W$ tq $y_i = y_{i-1} \dot{-} w_i$

2) $y_k \leq x_0$

3) $\forall i, j, \; 0 \leq i < j \leq k \Rightarrow y_i \not\leq y_j$.

DÉMONSTRATION.

A Soit $y_0, \dots, y_k \in \mathbf{N}^n$ tel que $y_i = y_{i-1} \dot{-} w_i$ et $y_k \leq x_0$. Posons $y_k = x_0$ et pour tout $i \leq \{1, \dots, k\}$, $y'_{i-1} = y' + w_i$.

Comme $y_{i-1} \dot{-} w_i = y_i$, d'après le lemme 2, $y_{i-1} \leq y_i + w_i$. Supposons que $y_i \leq y_i'$; il est clair que $y_i + w_i \leq y_i' + w_i$ et donc $y_{i-1} \leq y'_{i-1}$. Comme $y_k \leq x_0 = y_k'$, on en déduit que pour tout $i \in \{0, \dots, k\}$, $y_i \leq y_i'$. Puisque $y_i \in \mathbf{N}^n$ on a aussi $y_i' \in \mathbf{N}^n$ et donc, d'après la définition de $R(W, x_0)$, $y_0' \in R(W, x_0)$, et de plus $y_0 \leq y_0'$.

B Si $y \in R(W, x_0)$ il existe une séquence $y_k' = x_0, y_{k-1}, \dots, y_0' = y$ d'éléments de N^n telle que pour $i \in \{1, \dots, k\}$, $y'_{i-1} = y'_i + w_i$.

Soit $y_0 \in \mathbf{N}^n$ tel que $y_0 \leq y_0'$. D'après le lemme 2 si $z \leq y_i' = y_{i+1} + w_{i+1}$ alors $z \dot{-} w_{i+1}$ est défini et $z \dot{-} w_{i+1} \leq y'_{i+1}$.

Comme $y_0 \leq y_0'$, on peut donc construire une séquence $y_0, y_1, \dots, y_k$ d'éléments de $\mathbf{N}^n$ qui vérifie $y_{i+1} = y_i \dot{-} w_{i+1} \leq y'_{i+1}$ et donc en particulier $y_k \leq y_k' = x_0$ Les conditions 1 et 2 du lemme sont donc verifiées.

Supposons maintenant qu'il existe i et j tels que $0 \leq i < j \leq k$ et $y_i \leq y_j$

— si $j = k$ alors $y_i \leq y_k = x_0$ et la séquence $y_0, y_1, \dots, y_i$ vérifie encore les conditions 1 et 2

— si $j < k$, d'aprés le lemme 3 si $z \leq y_l$ alors $z \dot{-} w_{l+1}$ est défini et $z \dot{-} w_{l+1} \leq y_l \dot{-} w_{l+1} = y_{l+1}$; on peut donc construire la séquence $z_{i+1}, \dots, z_{i+k-j}$ qui vérifie

$$z_{i+1} = y_i \dot{-} w_{j+1} \leq y_{j+1}$$

$$z_{i+l} = z_{i+l-1} \dot{-} w_{j+l} \leq y_{j+l}$$

d'où en particulier $z_{i+k-j} \leq y_k$ et la séquence $y_0, y_1, \dots, y_i, z_{i+1}, \dots, z_{i+k-j}$ vérifie encore les conditions 1 et 2.

Dans les deux cas, tant que la condition 3 n'est pas remplie on peut trouver une séquence plus courte qui vérifie encore les conditions 1 et 2. Comme pour $k = 0$ la condition 3 est toujours trivialement remplie, on obtiendra toujours une séquence qui vérifie les conditions 1, 2 et 3. □

Soit maintenant T l'ensemble des séquences finies de $\mathbf{N}^n$ qui vérifient les conditions 1 et 3 du lemme 4.

LEMME 5. *L'ensemble T est fini.*

DÉMONSTRATION. Posons T_n l'ensemble des séquences de T longueur n.

Il est clair que $y_0, y_1, \dots, y_n$ est une séquence de T_n ssi $y_0, y_1, \dots, y_{n-1}$ est une séquence de T_{n-1} et $\exists\, w \in W$ tel que $y_n = y_{n-1} \dot{-} w$. Il en résulte d'une part que toute séquence de T_n admet une sous séquence initiale dans T_{n-1} et d'autre part que, puisque $T_1 = \{y_0\}$, chacun des ensembles T_i est fini.

Si $T = \bigcup_n T_n$ est infini, on peut appliquer le lemme de Koenig: il existe une séquence infinie $y_0, y_1, \dots, y_k, \dots$ telle que toute sous-séquence initiale finie est dans T.

Notons $(y_j)_l$ pour $l = 1, \dots, n$, la $l^{\text{ième}}$ composante de y_j.

Considérons l'ensemble $\{(y_j)_1 / j \in \mathbf{N}\}$. Si cet ensemble est fini, il existe un sous ensemble infini J_1 de $\mathbf{N}$ tel que $\forall j,\ j' \in J_1\ (y_j)_1 = (y_{j'})_1$. Si cet ensemble est infini il existe un sous-ensemble infini J_1 de N tel que $\forall j,\ j' \in J_1\ j < j' \Rightarrow (y_j)_1 < (y_{j'})_1$. Dans les deux cas on obtient un ensemble infini J_1 tel que $j, j' \in J_1,\ j < j' \Rightarrow (y_j)_1 \leq (y_{j'})_1$.

Supposons maintenant qu'il existe un ensemble infini $J_k \subseteq \mathbf{N}$ avec $k < n$, tel que pour $j, j \in J_k, j < j' \Rightarrow \forall i \leq k (y_j)_i \leq (y_{j'})_i$. Considérons l'ensemble $\{(y)_{k+1} / j \in J_k\}$. En procédant comme ci-dessus on obtient un ensemble infini $J_{k+1} \subset J_k$ tel que $j, j' \in J_{k+1},\ j < j' \Rightarrow \forall i \leq k+1,\ (y_j)_i \leq (y_{j'})_i$. On en déduit qu'il existe un ensemble infini J tel que $j, j' \in J$ $j < j' \Rightarrow y_j \leq y_{j'}$. Soit alors j_0 et j_1 les deux premiers éléments de J.

La sous-séquence initiale $y_0, \dots, y_{j_1}$ de $y_0, \dots, y_k, \dots$ est dans T_{j_1} par hypothèse, mais elle ne vérifie pas la condition 3 d'où une contradiction. L'ensemble T doit nécessairement être fini. □

THÉORÈME 5. *Soit* $\langle W, x_0 \rangle$ *un VASE de dimension* n *et soit* $y_0 \in \mathbf{N}^n$. *Il est décidable de savoir si il existe* $y \in R(W, x_0)$ *tel que* $y \geq y_0$.

DÉMONSTRATION. D'après le lemme 4, il suffit de trouver dans T une séquence $y_0, \dots, y_k$ tel que $y_k \leq x_0$. Or comme T est fini il existe n_0 tel que $n \geq n_0 \Rightarrow T_n = \varnothing$.

D'autre part puisque $T_{n+1} \neq \varnothing \Rightarrow T_n \neq \varnothing$, n_0 est le plus petit entier n tel que $T_n = \varnothing$.

La construction de T est donc effective: il est facile de construire successivement $T_1, T_2, \dots$ jusqu'à ce qu'on obtienne un T_n qui soit vide. □

Nous allons maintenant faire apparaitre le lien entre les langages de $\mathcal{F}$ (Init $(D_1')^*$) et les VASE.

Pour tout entier $k \geq 1$, posons $\Delta_k = \{a_i, \bar{a}_i, C_i / 1 \leq i \leq k\}$. Rappelons que $J_k =$ Init $(I_k) \subset \Delta_k^*$ est le « shuffle » de k copies disjointes de (Init $(D_1'^*)$ $C)^*$ Init $(D_1'^*)$. Pour tout mot y de J_k nous posons $t(y)$ égal à $a_1^{i_1} \dots a_k^{i_k} \dots a_k^{i_k}$ avec $\forall j \in \{1, \dots, k\}$ $i_j = l_{a_j}(y_j') - l_{\bar{a}_j}(y_j')$ où y_j' est l'unique facteur droit de $C_j y$ appartenant à $C_j (\Delta_k \{C_j\})^*$. En d'autres termes y s'écrit $y_1 \sqcup\!\sqcup y_2 \sqcup\!\sqcup \dots \sqcup\!\sqcup y_k$, y_j étant une copie sur $\{a_j, a_j, C_j\}$ d'un mot de (Init $(D_1'^*)$ $C)^*$ Init $(D_1'^*)$. Ce mot y_j peut donc s'écrire de façon unique $z z'$ avec z' ne contenant aucun C_j, et donc z' est une copie d'un mot de Init $(D_1'^*)$. Par définition y_j' est alors $C_j \cdot z'$.

Soit $M = \langle Q, \Delta_k, \delta, q_1, F \rangle$ un automate d'états fini avec $Q = \{q_1, \dots, q_n\}$. Définissons l'homomorphisme φ de $(Q \cup \{a_1, \dots, a_k\})^*$ dans $\mathbf{N}^{n+k}$ par $\varphi(y) = \langle l_{q_1}(y), l_{q_2}(y), \dots, l_{q_n}(y), l_{a_1}(y), \dots, l_{a_k}(y) \rangle$ pour tout y.

LEMME 6. *Si l'automate* M *vérifie la propriété*

$$(*) \qquad \forall q \in Q, \ \forall y \in \Delta_k, \ \delta(q, y) \neq q$$

alors il existe un VASE $\langle W, x_0 \rangle$ *tel que*

$$R(W, x_0) = \{\varphi(\delta(q_1, y) \cdot t(y)) / y \in J_k\}.$$

DÉMONSTRATION. Prenons $\bar{Q} = \{\bar{q}_i / 1 \leq i \leq n\}$ et étendons φ en un homomorphisme, noté encore φ de $(Q \cup \bar{Q} \cup \Delta_k)^*$ dans $(\mathbf{Z} \cup \{\#\})^{n+k}$ en posant

· pour $i \in \{1, \dots, n\}$ $\varphi(\quad) = -\varphi(q_i)$

· pour $i \in \{1, \ldots, k\}$ $\varphi(\bar{a}_i) = -\varphi(a_i)$

· pour $i \in \{1, \ldots, k\}$ $\varphi(c_i) = v_j$

où v_j est l'élément de $\mathbf{Z}_1^{n+k}$ dont toutes les composantes sont nulles, sauf la $n+j^{\text{ème}}$ qui est égale à $\#$. Considérons alors le VASE $\langle W, x_0 \rangle$ avec $W = \{\varphi(\bar{q}\, z\, q')/q,\ q' \in Q,\ z \in \Delta_k,\ q' = \delta(q, z)\}$ et $x_0 = \varphi(q_1)$.

Le résultat se démontre alors aisément par induction soit sur la longueur de y dans un sens, soit sur le nombre d'additions dans le VASE dans l'autre. □

THÉORÈME 6. *Tout langage de* $\mathcal{F}_\cap$ (Init $(D_1'^*)$) *est récursif.*

DÉMONSTRATION. D'aprés le théorème 2 et l'égalité $\mathcal{F}_\cap(\text{Init}(D_1'^*)) = \mathcal{C}(\{J_k/k \geq 1\})$ il suffit de montrer que pour tout $k \geq 1$ et tout langage rationnel $R \subseteq \Delta_k^*$, $J_k \cap R = \varnothing$ est décidable. Considérons un automate $M = \langle Q, \Delta_k, \delta, q_1, F \rangle$ qui reconnait le langage R; il est toujours possible de modifier cet automate de façon à ce que la condition $(*)$ soit vérifiée (par dédoublement des états par exemple); Considérons aussi le VASE $\langle W, x_0 \rangle$ construit dans le lemme 6. Alors $J_k \cap R \neq \varnothing$ ssi il existe $y \in J_k$ tel que $\delta(q_1, y) \in F$ et donc ssi il existe $q_i \in F$ tel que $\varphi(\delta(q_1, y)\, t(y)) \geq \varphi(q_i)$ ou encore d'aprés le lemme 6 ssi $\exists\, q_i \in F, x \in R(W, x_0)$ tel que $x \geq \varphi(q_i)$, ce qui est décidable d'après le lemme 5. □

Remarquons d'autre part que pour tout $k \geq 1$, $y \in I_k$ si et seulement si $y \in J_k$ et $t(y) = \varepsilon$. En reprenant les notations du lemme 6, et en notant R le langage rationnel reconnu par l'automate M, nous obtenons $I_k \cap R \neq \varnothing$ si et seulement si il existe $q_i \in F$ tel que $\varphi(q_i) = \varphi(\delta(q_1, y)\, t(y)) \in R(W, x_0)$.

Comme $\mathcal{C}_\cap(I) = \mathcal{C}(\{I_k/k \geq 0\})$ contient tous les langages récursivement énumérables, d'aprés le théorème 2 $I_k \cap R = \varnothing$ ne peut être décidable. Ce qui implique que $R(W, x_0)$ n'est pas un ensemble récursif et donc que le « Reachability Problem » pour les VASE est indécidable, résultat analogue au théorème 5 de Araki et Kasami [1].

Dans [4], Berstel a étudié la structure d'ordre des cônes rationnels engendrés par les langages algébriques bornés sur un alphabet à deux lettres; il montre en particulier que les langages $L_> = \{a^n b^p/n \geq p \geq 0\}$ et $L_< = \{a^n b^p/0 \leq n \leq p\}$ sont rationnellement incomparables, c'est-à-dire que $\mathcal{C}(L_>)$ et $\mathcal{C}(L_<)$ sont des familles incomparables pour l'inclusion. Nous pouvons renforcer ce résultat en montrant que les familles $\mathcal{F}_\cap(L_>)$ et $\mathcal{F}_\cap(L_<)$ sont aussi incomparables pour l'inclusion.

COROLLAIRE 1. *Les familles* $\mathcal{F}_\cap(L_>)$ *et* $\mathcal{F}_\cap(L_<)$ *sont incomparables pour l'inclusion.*

DÉMONSTRATION. Supposons que $L_< \in \mathcal{F}_\cap(L_>)$. Alors $\{a^n b^n/n \geq 0\} = L_> \cap L_<$ appartient encore à $\mathcal{F}_\cap(L_>)$ et donc $\mathcal{F}_\cap(L_>)$ contient tous les langage récursivement énumérables. Or $L_> \in \mathcal{C}$ (Init($D_1'^*$)) et donc à fortiori $L_> \in \mathcal{F}_\cap$(Init($D_1'^*$)) et donc $\mathcal{F}_\cap(L_>) \subset \mathcal{F}_\cap$ (Init ($D_1'^*$)), ce qui contredit le théorème 5. Comme $L_<$ est l'image miroir de $L_>$ et réciproquement, il est clair que $L_> \in \mathcal{F}_\cap$ ($L_<$) implique $L_< \in \mathcal{F}_\cap$ ($L_>$) et donc nous avons aussi $L_> \notin \mathcal{F}_\cap$ ($L_<$). □

Le théoreme 5 permet aussi de montrer que Init ($D_1'^*$) est rationnellement incomparable avec sa clôture commutative qui est isomorphe à $C(L_>) = \psi^{-1} \circ \psi(L_>)$. De façon plus précise nous avons

COROLLAIRE 2. *Le langage Init* ($D_1'^*$) *n'appartient pas à* $\mathcal{C}_\cap$ ($C(L_>)$) *et le langage* $C(L_>)$ *n'appartient pas à* $\mathcal{F}_\cap$ (*Init* ($D_1'^*$)).

DÉMONSTRATION. Comme $L_< \in \mathcal{C}_\cap(C(L_>))$ on a

$$\mathcal{C}_\cap(C(L_>)) = \mathcal{C}_\cap(\{a^n b^n / n \geq 0\}) = \mathcal{C}_\cap(D_1'^*)$$

(cf. [12]), où $D_1^* = C(D_1'^*)$.

Si Init ($D_1'^*$) appartenait à $\mathcal{C}_\cap(D_1^*)$ on aurait $D_1'^* = D_1^* \cap$ Init (D_1') $\in \mathcal{C}_\cap(D_1^*)$ ce qui contredit un résultat de [12]. Supposons maintenant que $C(L_>) \in \mathcal{F}_\cap$ (Init ($D_1'^*$)). On aurait alors $\mathcal{F}_\cap(C(L_>)) \subseteq \mathcal{F}_\cap$ (Init ($D_1'^*$)). Or $\mathcal{F}_\cap(C(L_>)) = \mathcal{F}_\cap(\{a^n b^n / n \geq 0\})$ contient tous les langages récursivement énumérables, ce qui contredit le théorème 5. □

BIBLIOGRAPHIE

[1] T. ARAKI, T. KASAMI, *Some decision problems related to the reachability problem for Petri Nets*, Theoret. Comput. Sci. **3** (1977), 85-104.

[2] A. ARNOLD, M. LATTEUX, *Vector addition systems and semi-Dyck languages*, Publication du Laboratoire de Calcul de Lille 1, n⁰ 78 (1977).

[3] B. S. BAKER, R. V. BOOK, *Reserval-Bounded Multipushdown Machines*, J. Comput. System. Sci. **8** (1974), 315-332.

[4] J. BERSTEL, *Une hiérarchie des parties rationnelles de N^2*, Math. Systems Theory **7** (1973), 114-137.

[5] S. CRESPI-REGHIZZI, D. MANDRIOLI, *Commutative grammars*, Calcolo **13** (1976), 173-189.

[6] S. CRESPI-REGHIZZI, D. MANDRIOLI, *Petri Nets and Szilard languages*, Information and Control **33** (1977), 177-192.

[7] S. GINSBURG, *Algebraïc and Automata-theoretic Properties of Formal Languages*, (1975) North-Holland Publishing Company.

[8] S. GINSBURG, J. GOLDSTINE, *Intersection-closed Full AFL and the recursively enumerable languages*, Information and Control **22** (1973). 201-231.

[9] J. HARTMANIS, J. E. HOPCROFT, *What makes some languages theory problems undecidable*, J. Comput. System. Sci. **4** (1970), 368-376.

[10] M. HÖPNER, M. OPP, *Renaming and Erasing in Szilard languages*, in « Automata, Languages and Programming; Fourth Colloquium, Turku » (A. Salomaa and M. Steinby, Eds.), pp. 244-257, Lectures Notes in Computer Science **52**, Springer-Verlag, Berlin, 1977.

[11] R. M. KARP, R. E. MILLER, *Parallel Program Schemata*, J. Comput. System. Sci. **3** (1969), 147-195.

[12] M. LATTEUX, *Cônes rationnels commutativement clos*, RAIRO Informatique Théorique **11** (1977), 29-51.

[13] M. LATTEUX, *Cônes rationnels commutatifs*, Publication du Laboratoire de Calcul de Lille 1, n⁰ 86 (1977).

[14] M. NIVAT, *Transductions des langages de Chomsky*, Ann. Inst. Fourier (Grenoble) **18** (1968), 339-455.

[15] G. S. SACERDOTE, R. L. TENNEY, *The decidability of the Reachability Problem for Vector Addition Systems*, 9 th annual Symposium on Theory of Computing, 1977.