PRODUIT DANS LE CÔNE RATIONNEL ENGENDRÉ PAR D_1^*

Michel LATTEUX

Université de Lille, 1, U.E.R. d'I.E.E.A., Service Informatique, B.P. 36, 59650 Villeneuve d'Ascq, France

Communiqué par Ronald Book
Reçu mars 1977

Résumé. Nous montrons que si le cône rationnel généré par D_1^*, le langage de Dyck sur une lettre, contient le produit de deux langages définis sur des alphabets disjoints, nécessairement l'un des deux langages est rationnel. Cette propriété n'est pas vraie pour le cône rationnel généré par $D_1'^*$, le langage de semi-Dyck sur une lettre.

0. Introduction

Les "Familles Agréables de Langages" (en anglais, "Full Abstract Families of Languages") ont été introduites par Ginsburg et Greibach [5]. Les transductions rationnelles caractérisées en terme de bimorphismes alphabétiques par Nivat [9], se sont révélées un outil essentiel pour l'étude de ces familles et des cônes rationnels qui sont justement les classes de langage closes par transduction rationnelle.

Considérons D_n^* (respectivement, $D_n'^*$) le langage de Dyck (respectivement, de semi-Dyck) sur n lettres, c'est-à-dire la classe du mot vide e dans la congruence engendrée par $a_i\bar{a}_i = \bar{a}_i a_i = e$ (respectivement, $a_i\bar{a}_i = e$) pour $i \in \{1, \ldots, n\}$. Le théorème de Chomsky et Schützenberger implique que, pour tout $n \geqslant 2$, D_n^* et $D_n'^*$ sont des générateurs du cône rationnel des langages algébriques (en anglais, "Context-free Languages"). Au contraire, Boasson a montré [1] que D_1^* et $D_1'^*$ sont rationnellement incomparables et ne sont pas générateurs du cône rationnel des langages algébriques. Dans [8], il était démontré que $D_1'^*$ ne pouvait pas être obtenu par des "FAL-opérations" à partir de langages commutatifs. Nous démontrons ici, pour $\mathscr{C}(D_1^*)$, le cône rationnel généré par D_1^*, une propriété qui n'est plus vérifiée pour le cône rationnel généré par $D_1'^*$: Si $\mathscr{C}(D_1^*)$ contient le produit de deux langages définis sur des alphabets disjoints, l'un des deux langages est rationnel. Cette propriété a été démontrée par Greibach [7] pour la famille des langages algébriques linéaires et par Goldstine [6], pour les cônes rationnels générés par des langages 1-bornés (cf. [4]). Enfin, en utilisant la commutativité de D_1^*, nous en déduisons (Corollaire 2.9) une propriété de l'opération "chevron" définie et étudiée dans [2] et [3].

1. Définitions et notations

Soient $Y=\{b_1,\ldots,b_p\}$ un alphabet et L_1,L_2 deux langages inclus dans Y^*. L'opération "Shuffle" est définie par:

$$\mathrm{Shuf}(L_1,L_2)=\{u_1v_1\cdots u_kv_k : u_i, v_i\in Y^*, u_1\cdots u_k\in L_1, v_1\cdots v_k\in L_2\}.$$

Le *quotient* de L_1 par L_2, noté L_1/L_2 est égal à $\{u\in Y^*\colon \exists v\in L_2, uv\in L_1\}$. Le mot u est un *facteur itérant* de L_1, si u est différent du mot vide e et s'il existe $u_1,u_2\in Y^*$ tels que $u_1u^*u_2\subseteq L_1$. La fonction de Parikh, notée ψ_Y (ou ψ s'il n'y a pas de confusion possible) est l'homomorphisme canonique de Y^* dans $\mathbf{N}^p$ où $\mathbf{N}$ désigne l'ensemble des entiers naturels. Un ensemble S de $\mathbf{N}^p$ est un ensemble *linéaire* s'il existe $x_0,\ldots,x_k\in\mathbf{N}^p$ tels que

$$S=\left\{x_0+\sum_{i=1}^{k}\lambda_ix_i\colon \lambda_i\in\mathbf{N}\right\}.$$

Désignons par $\mathbf{Z}$, l'ensemble des entiers relatifs et posons $\mathbf{N}_+=\mathbf{N}\setminus\{0\}$. Si A et B sont inclus dans $\mathbf{Z}$, notons $A+B=\{a+b\colon a\in A, b\in B\}$, $AB=\{ab\colon a\in A, b\in B\}$ et $-A=\{-a\colon a\in A\}$. Enfin, pour tout $x\in\mathbf{Z}$, $|x|$ désignera la valeur absolue de x et nous utiliserons x à la place de $\{x\}$.

2. Résultats

Nous utiliserons la propriété suivante qui est bien connue et s'obtient aisément à partir du théoréme de Kleene:

Lemme 2.1. *Si R est un langage rationnel, il existe des langages rationnels $R_1,\ldots,R_k$ tels que*
(i) $R=\bigcup_{i=1}^{k}R_i$,
(ii) *pour tout $i\in\{1,\ldots,k\}$, $\psi(R_i)$ est un ensemble linéaire.*

Dans toute la suite, T désignera l'alphabet $\{a_1,\bar a_1\}$. Considérons d, l'homomorphisme de T^* dans $\mathbf{Z}$ muni de l'addition, défini sur T par $d(a_1)=1$ et $d(\bar a_1)=-1$. Donnons quelques propriétés vérifiées par cet homomorphisme:

Lemme 2.2. *Pour tout $x\in\mathbf{Z}$ et tout $y\in\mathbf{N}_+$, $d^{-1}(x+\mathbf{Z}y)$ est un langage rationnel.*

Démonstration. On peut supposer, sans nuire à la généralité de la preuve, que $x\in\{0,\ldots,y-1\}$. Pour tout $i\in\{0,\ldots,y-1\}$, posons $R_i=\mathrm{Shuf}(a_1^t(a_1^y)^*,\bar a_1^i(\bar a_1^y)^*)$ où t est le reste dans la division entière de $x+i$ par y. Il est facile de vérifier que $d^{-1}(x+\mathbf{Z}y)$ est égal á $\bigcup_{i=0}^{y-1}R_i$ et, comme la famille des langages rationnels est close pour l'opération "shuffle", $d^{-1}(x+\mathbf{Z}y)$ est bien un langage rationnel. $\square$

Lemme 2.3. *Pour tout langage* $L \subseteq T^*$, D_1^*/L *est égal à* $d^{-1}(-d(L))$.

Démonstration. Prenons $u \in D_1^*/L$. Il existe $v \in L$ tel que $uv \in D_1^*$. Donc $d(u)+d(v)=0$, $d(u) \in -d(L)$ et $u \in d^{-1}(-d(L))$. Réciproquement si $u \in d^{-1}(-d(L))$, il existe $v \in L$ tel que $d(uv)=d(u)+d(v)=0$. Donc $uv \in D_1^* = d^{-1}(0)$ et $u \in D_1^*/L$. □

Lemme 2.4. *Soit R un langage rationnel inclus dans* T^*. *Si* $\psi(R)$ *est un ensemble linéaire et si R possède deux facteurs itérants u et v tels que* $d(u)\ d(v)<0$, *alors* D_1^*/R *est un langage rationnel.*

Démonstration. Comme $\psi(R)$ est un ensemble linéaire, il existe $x_0,\dots,x_k \in \mathbf{Z}$ tels que $d(R)=\{x_0+\sum_{i=1}^{k}\lambda_i x_i : \lambda_i \in \mathbf{N}\}$. Comme R possède un facteur itérant appartenant à $d^{-1}(\mathbf{N}_+)$, il existe $i \in \{1,\dots,k\}$ tel que $x_i \in \mathbf{N}_+$. De même, il existe $j \in \{1,\dots,k\}$ tel que x_j soit négatif. Nous pouvons en déduire que $d(R)$ est égal à $x_0+\mathbf{Z}y$ où y est le plus grand commun diviseur de $|x_1|,\dots,|x_k|$. Et d'après les Lemmes 2.2 et 2.3, $D_1^*/R = d^{-1}(-x_0+\mathbf{Z}y)$ est un langage rationnel. □

Lemme 2.5. *Soit R un langage rationnel inclus dans* T^*, *Si tout facteur itérant de R appartient à* $d^{-1}(\mathbf{N})$, *alors, pour tout* $k \in \mathbf{Z}$, *les langages* $R_1=\{x \in R: d(x)=k\}$, $R_2=\{x \in R: d(x)>k\}$ *et* $R_3=\{x \in R: d(x)<k\}$ *sont rationnels.*

Démonstration. Soient $M=(Q,T,f,q_0,F)$ un automate d'état fini déterministe qui reconnait R et n le nombre d'états de M. Prenons $Q'=Q\times\{-n,\dots,k+n\}$, $q_0'=(q_0,0)$ et définissons f' sur $Q'\times T$ par:

$$f'((q,i),b)=\begin{cases}(f(q,b),i+d(b)) & \text{si } -n<i<k+n,\\ (f(q,b),i & \text{sinon.}\end{cases}$$

Pour $j=1,2,3$, posons $M_j'=(Q',T,f',q_0',F_j')$ où $F_1'=F\times\{k\}$, $F_2'=F\times\{k+1,\dots,k+n\}$ et $F_3'=F\times\{-n+1,\dots,k-1\}$. Comme tout facteur itérant de R appartient à $d^{-1}(\mathbf{N})$, pour toute factorisation $W_1W_2W_3$ d'un mot de R, $d(W_2)$ est supérieur à $-n$. Nous pouvons en déduire que pour $j=1,2,3$, le langage reconnu par M_j' est égal á R_j. □

Lemme 2.6. *Soient R et R' deux langages rationnels inclus dans* T^*. *Si* $\psi(R)$ *et* $\psi(R')$ *sont des ensembles linéaires, alors l'un des deux langages* $R\cap(D_1^*/R')$ *et* $R'\cap(D_1^*/R)$ *est rationnel.*

Démonstration. Distinguons plusieurs cas:

(i) Si R (resp. R') vérifie l'hypothèse du Lemme 2.4, D_1^*/R (resp. D_1^*/R') est un langage rationnel ainsi que $R'\cap(D_1^*/R)$ (resp. $R\cap(D_1^*/R')$).

(ii) Si tous les facteurs itérants de R et tous ceux de R' appartiennet à $d^{-1}(\mathbf{N})$, il existe, alors $x_0 \in \mathbf{Z}$, $x_1, \ldots, x_k \in \mathbf{N}$ tels que $d(R') = \{x_0 + \sum_{i=1}^{k} \lambda_i x_i : \lambda_i \in \mathbf{N}\}$. On peut donc trouver un ensemble fini $F \subseteq \mathbf{Z}$, $z \in \mathbf{Z}$ tels que $d(R') = F \cup (z + \mathbf{N}y)$ où y est le plus grand commun diviseur de $x_1, \ldots, x_k$. D'après le Lemme 2.3, D_1^*/R' est égal à $d^{-1}(-d(R')) = d^{-1}(F') \cup d^{-1}(z' + \mathbf{N}y')$ où $F' = -F$, $z' = -z$ et $y' = -y$. Le langage $R \cap d^{-1}(F') = \bigcup_{t \in F'} \{x \in R : d(x) = t\}$ est rationnel d'aprés le Lemme 2.5. Quant à $R \cap d^{-1}(z' + \mathbf{N}y')$, il est égal à $\{x \in R : d(x) < z' + 1\} \cap d^{-1}(z' + \mathbf{Z}y)$ qui est un langage rationnel d'après les Lemmes 2.2 et 2.5. Donc $R \cap (D_1^*/R') = (R \cap d^{-1}(F')) \cup (R \cap d^{-1}(z' + \mathbf{N}y'))$ est un langage rationnel.

(iii) Si tous les facteurs itérants de R appartiennent à $d^{-1}(\mathbf{N})$ et tous ceux de R' appartiennent à $d^{-1}(-\mathbf{N})$, on montre de la même façon que $d(R') = F \cup (z + \mathbf{N}y)$ où F est un ensemble fini de $\mathbf{Z}$, $z = -z' \in \mathbf{Z}$ et $y' = -y \in \mathbf{N}$. La seule différence avec le (ii) est que y est $\leqslant 0$. Donc $R \cap d^{-1}(z' + \mathbf{N}y')$ est égal à $\{x \in R : d(x) > z' - 1\} \cap d^{-1}(z' + \mathbf{Z}y')$. Mais, d'après les Lemmes 2.2 et 2.5, c'est encore un langage rationnel.

La démonstration s'achève en remarquant que tous les autres cas possibles sont symétriques de l'un des cas étudiés. □

Nons pouvons maintenant énoncer:

Proposition 2.7. *Soient L_1 et L_2 deux langages sur des alphabets disjoints. Si L_1L_2 appartient à $\mathscr{C}(D_1^*)$, le plus petit cône rationnel contenant le Dyck sur une lettre, alors l'un des deux langages L_1, L_2 est rationnel.*

Démonstration. Soient T_1 et T_2, deux alphabets disjoints tels que $L_1 \subseteq T_1^*$ et $L_2 \subseteq T_2^*$. Si L_1L_2 appartient à $\mathscr{C}(D_1^*)$, il existe un alphabet Y, un langage rationnel $R \subseteq Y^*$ et deux homomorphismes alphabétiques h et g ($h(Y) \subseteq T \cup \{e\}$, $g(Y) \subseteq T_1 \cup T_2 \cup \{e\}$) tels que $L_1L_2 = g(h^{-1}(D_1^*) \cap R)$ (cf. [9]). Pour $i = 1, 2$, posons $Y_i = \{a \in Y : g(a) \in T_i \cup \{e\}\}$. Comme $L_1L_2 \subseteq T_1^*T_2^*$, $L_1L_2 = g(h^{-1}(D_1^*) \cap R')$ où $R' = R \cap Y_1^*Y_2^*$. Il est facile de vérifier que R' peut s'écrire $\bigcup_{j=1}^{k} R'_jR''_j$ où $\forall j \in \{1, \ldots, k\}$, R'_j, R''_j sont des langages rationnels vérifiant $R'_j \subseteq Y_1^*$ et $R''_j \subseteq Y_2^*$. On supposera, de plus, que $\psi(R'_j)$ et $\psi(R''_j)$ sont des ensembles linéaires (cf. Lemme 2.1). Donc $L_1L_2 = \bigcup_{j=1}^{k} V_j$ avec $V_j = g(h^{-1}(D_1^*) \cap R'_jR''_j)$.

Pour $i = 1, 2$, l'homomorphisme h_i de $(T_1 \cup T_2)^*$ dans T_i^* est défini sur $T_1 \cup T_2$ par

$$h_i(a) = \begin{cases} a & \text{si } a \in T_i, \\ e & \text{sinon.} \end{cases}$$

Montrons que pour tout $j \in \{1, \ldots, k\}$, l'un des deux langages $h_1(V_j)$, $h_2(V_j)$ est rationnel. Comme $g(R'_j)$ et $g(R''_j)$ sont inclus respectivement dans T_1^* et T_2^*, $h_1(V_j) = g(R'_j \cap (h^{-1}(D_1^*)/R''_j))$. De même, étant donné que D_1^* et $h^{-1}(D_1^*)$ sont des langages commutatifs, $h_2(V_j) = g(R''_j \cap (h^{-1}(D_1^*)/R'_j))$. En utilisant l'égalité $h^{-1}(D_1^*)/R_j'' = h^{-1}(D_1^*/h(R_j''))$, on obtient $h_1(V_j) = g(R_j' \cap h^{-1}(h(R_j') \cap (D_1^*/h(R_j''))))$. De la même façon, on peut montrer que $h_2(V_j) = g(R_j'' \cap h^{-1}(h(R_j'') \cap (D_1^*/h(R_j'))))$. Le Lemme 2.6 implique, alors, que l'un des langages $h_1(V_j)$, $h_2(V_j)$ est rationnel. Pour $i = 1,2$, posons $K_i = \{j \in \{1,\dots,k\}$: $h_i(V_j)$ et rationnel$\}$. D'aprés la propriété ci-dessus, $K_1 \cup K_2 = \{1,\dots,k\}$ et comme, pour $i = 1,2$, $h_i(L_1L_2) = L_i$, $h_i(V_j) \subseteq L_i$, $\forall j \in \{1,\dots,k\}$. Nous obtenons alors $L_1L_2 = R_1L_2' \cup L_1'R_2$ où pour $i = 1,2$, $R_i = \bigcup_{j \in K_i} h_i(V_j)$ est un langage rationnel et L_i' est égal à $\bigcup_{j \in K_{i'}} h_i(V_j)$ avec $i' = 3 - i$. Il est, alors, clair que L_1 où L_2 doit être un langage rationnel. □

Remarquons que la proposition précédente n'est plus vraie si on remplace le Dyck sur une lettre par $D_1'^*$, le semi-Dyck sur une lettre. En effet, montrons:

$$L_1L_2 \in \mathscr{C}(D_1'^*) \quad \text{avec } L_1 = \{a^nb^p\colon n \geq p \geq 0\} \text{ et } L_2 = \{c^sd^t\colon t \geq s \geq 0\}.$$

Considérons l'alphabet $Y = \{b_1, b_2, \dots, b_6\}$ et définissons l'homomorphisme h de Y^* dans T^* par $h(b_1) = h(b_6) = e$, $h(b_2) = h(b_4) = a_1$ et $h(b_3) = h(b_5) = \bar{a}_1$. De même l'homomorphisme g de Y^* dans $\{a,b,c,d\}^*$ est défini sur Y par $g(b_1) = g(b_2) = a$, $g(b_5) = g(b_6) = d$, $g(b_3) = b$ et $g(b_4) = c$. Il est, alors, facile de vérifier l'égalité entre L_1L_2 et $g(h^{-1}(D_1'^*) \cap b_1^*b_2^*b_3^*b_4^*b_5^*b_6^*)$.

D'autre part de la proposition précédente, nous pouvons déduire immédiatement:

Corollaire 2.8. *Soit L un langage défini sur un alphabet ne contenant pas la lettre c. Alors, LcL appartient à $\mathscr{C}(D_1^*)$ si et seulement si L est un langage rationnel.*

L'opération "chevron" qui à tout langage L fait correspondre $\langle L \rangle = \{a^nxb^n\colon x \in L, n \geq 0\}$ où a et b sont de nouveaux symboles, a été définie et étudiée dans [2] et [3]. En utilisant la commutativité de D_1^*, nous pouvons déduire de la Proposition 2.7:

Corollaire 2.9. *Soient L un langage inclus dans T_1^*, a et b deux symboles n'appartenant pas à T_1. Alors $\langle L \rangle = \{a^nxb^n : x \in L,\ n \geq 0\}$ appartient à $\mathscr{C}(D_1^*)$ si et seulement si L est un langage rationnel.*

Démonstration. Si $\langle L \rangle \in \mathscr{C}(D_1^*)$, il existe [9], un alphabet Y, un langage rationnel $R \subseteq Y^*$ et deux homomorphismes alphabétiques h et g tels que $\langle L \rangle = g(h^{-1}(D_1^*) \cap R)$. Posons $A = \{x \in Y\colon g(x) \in \{a, e\}\}$, $B = \{x \in Y\colon g(x) \in \{b, e\}\}$ et $Y_1 = \{x \in Y\colon g(x) \in T_1 \cup \{e\}\}$. Comme $\langle L \rangle \subseteq a^*T_1^*b^*$, $\langle L \rangle = g(h^{-1}(D_1^*) \cap R')$ où $R' = R \cap A^*Y_1^*B^*$. Donc R' peut s'écrire $\bigcup_{i=1}^k R_i'R_i''R_i'''$ où R_i', R_i'', R_i''' sont des langages rationnels inclus respectivement dans A^*, Y_1^* et B^*. Comme $h^{-1}(D_1^*)$ est un langage commutatif, il est facile de vérifier que $L\{a^nb^n\colon n \geq 0\}$ est égal á $g(h^{-1}(D_1^*) \cap R'')$ où $R'' = \bigcup_{i=1}^k R_i''R_i'R_i'''$ et d'aprés la Proposition 2.1, L est rationnel.

Réciproquement si L est un langage rationnel, il est clair que $\langle L \rangle \in \mathscr{C}(D_1^*)$. □

Références

[1] L. Boasson, Two iteration theorems for some families of languages, *J. Comput. system Sci.* **7** (1973) 583–596.

[2] L. Boasson et M. Nivat, Sur Diverses familles de langages fermées par transduction rationnelle, *Acta Informat.* **2** (1973) 180–188.

[3] L. Boasson, J.P. Crestin et M. Nivat, Familles de langages translatables et fermées par crochet, *Acta Informat.* **2** (1973) 383–393.

[4] S. Ginsburg, *Algebraic and Automata-Theoretic Properties of Formal languages* (North-Holland, Amsterdam, 1975).

[5] S. Ginsburg et S. Greibach, Abstract families of languages, *Mem. Am. Math. Soc.* **87** (1969) 1–32.

[6] J. Goldstine, Abstract families of languages generated by bounded languages, Thesis, Université de Californie, Berkeley, CA (1970).

[7] S. Greibach, The unsolvability of the recognition of linear context-free languages, *J. Assoc. Comput. Mach.* **13** (1966) 582–587.

[8] M. Latteux, Cônes rationnels commutativement clos, *Rev. Française Automat. Informat. Recherche Opérationnelle* (à paraître).

[9] M. Nivat, Transductions des langages de Chomsky, *Ann. Inst. Fourier (Grenoble)* **18** (1968) 339–456.