

A. TERLUTTE

Sur les centres de DOL-langages

Informatique théorique et applications, tome 21, n° 2 (1987),
p. 137-145

http://www.numdam.org/item?id=ITA_1987__21_2_137_0

© AFCET, 1987, tous droits réservés.

L'accès aux archives de la revue « Informatique théorique et applications » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

SUR LES CENTRES DE DOL-LANGAGES (*)

par A. TERLUTTE ⁽¹⁾

Communiqué par J. BERSTEL

Résumé. – *Nous prouvons que le centre d'un DOL-langage est un langage co-algébrique.*

Abstract. – *We prove that the center of a DOL-language is a co-CFL.*

I. INTRODUCTION

Récemment, plusieurs conjectures énoncées dans [1] et [2] et qui concernaient différentes familles de langages algébriques ont reçu des réponses négatives [7]. Ces réponses sont des conséquences du résultat suivant : l'ensemble des facteurs gauches de la séquence de Thue-Morse est un langage co-algébrique [13]. Dans [3], Berstel a étendu ce résultat en démontrant le théorème suivant : si un morphisme itéré engendre un mot infini, alors l'ensemble des facteurs gauches de ce mot infini est un langage co-algébrique. Dans ce cas, ce mot est l'unique élément de l'adhérence du DOL-langage associé au morphisme itéré et le théorème de Berstel indique que le centre de ce DOL-langage, c'est-à-dire l'ensemble des facteurs gauches de son adhérence est un langage co-algébrique. Il est, alors, naturel de se poser la question de savoir si ce résultat reste valable pour tout DOL-langage [4]. Nous apportons une réponse à cette question en démontrant, dans ce papier, que le centre d'un DOL-langage est un langage co-algébrique.

(*) Reçu juin 1986, révisé octobre 1986.

(¹) U.E.R. d'J.E.E.A.-Informatique, Bât. M3, Université de Lille-I, 59655 Villeneuve-d'Ascq Cedex, France.

II. PRÉLIMINAIRES

\mathbb{N} désignera l'ensemble des entiers naturels et $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ l'ensemble des entiers strictement positifs.

Soit X un alphabet fini. X^* représente le monoïde libre engendré par X , c'est-à-dire l'ensemble des mots finis construits sur X , avec le mot vide noté ε . Si w est un mot de X^* , on note $|w|$ la longueur de w .

L'ensemble des langages algébriques est noté Alg. Le langage \bar{L} représente le complémentaire d'un langage L . Le langage L est co-algébrique si son complémentaire est algébrique.

Un mot infini α est une application α de \mathbb{N}_+ dans X . Pour $n \in \mathbb{N}_+$, $\alpha(n)$ désigne la n -ième lettre de α et $\alpha[n]$ représente le mot fini $\alpha(1)\alpha(2)\dots\alpha(n)$. On posera $\alpha[0] = \varepsilon$. L'ensemble des mots infinis sur X est noté X^ω et l'ensemble des mots finis et infinis sur X est noté $X^\infty = X^* \cup X^\omega$.

L'ensemble des facteurs gauches d'un mot infini α est noté $FG(\alpha)$ et représente l'ensemble $\{\alpha[n]/n \in \mathbb{N}\}$. Nous avons, pour tout $u \in X^\infty$,

$$FG(u) = \{w \in X^* / \exists v \in X^\omega, wv = u\}$$

et pour un langage $L \subset X^\infty$, $FG(L) = \{FG(u)/u \in L\}$.

A tout langage L inclus dans X^* on peut associer l'adhérence de L , ensemble de mots infinis, notée $Adh(L)$, et définie par

$$Adh(L) = \{\alpha \in X^\omega / \alpha[n] \in L \forall n \in \mathbb{N}_+\}.$$

On définit alors le centre d'un langage $L \subset X^*$, noté L^c , par

$$L^c = FG(Adh(L)) = \{w \in X^* / wX^* \cap L \text{ est infini}\}.$$

Un DOL-système est un triplet $\langle X, h, u \rangle$ où X est un alphabet fini, h est un morphisme de X^* dans X^* et l'axiome u appartient à X^* . Le DOL-langage $L(G)$ engendré par G contient tous les mots de la forme $h^n(u)$, avec $n \in \mathbb{N}$. On notera

$$L(G) = h^*(u) = \{h^n(u)/n \in \mathbb{N}\}.$$

Pour plus de détails, le lecteur consultera [12] et [14] en ce qui concerne les L -systèmes.

III. RÉSULTATS

Nous introduisons la notion de DOL-langages et de DOL-systèmes unitaires, construits à partir de morphismes particuliers. Les propriétés des DOL-systèmes unitaires et la décomposition d'un DOL-système quelconque en DOL-systèmes unitaires nous permettront de démontrer que le centre d'un DOL-langage est co-algébrique.

Il a été montré dans [6] que l'adhérence d'une union finie de langages est égale à l'union des adhérences de ces langages. De plus, l'adhérence d'un langage est vide si et seulement si ce langage est fini. Nous en déduisons immédiatement une propriété que nous utiliserons par la suite.

PROPRIÉTÉ 1 : Soient $G = \langle X, h, u \rangle$ un DOL-système, n_0 un entier positif et n un entier strictement positif. Alors

$$\text{Adh}(L(G)) = \bigcup_{i=0}^{n-1} \text{Adh}(L(G_i))$$

où $\forall i \in [0, n-1]$, G_i désigne le DOL-système $\langle X, h^n, h^{n_0+i}(u) \rangle$.

Nous utiliserons le lemme suivant démontré par Head dans [11] :

LEMME 2 : Soient $G_1 = \langle X, h, u_1 \rangle$ et $G_2 = \langle X, h, u_1 u_2 \rangle$ deux DOL-systèmes. Alors, si $\text{Adh}(L(G_1))$ est non vide, $L(G_1)$ et $L(G_2)$ ont la même adhérence.

Nous distinguons dans la suite, pour un DOL-système $G = \langle X, h, u \rangle$, deux sous-ensembles disjoints de l'alphabet X : I , l'ensemble des lettres engendrant, par itérations successives du morphisme, des langages infinis et F , l'ensemble de celles engendrant des langages finis. Le lemme 2 nous permet d'affirmer que, parmi les lettres de I apparaissant dans l'axiome, seule celle située le plus à gauche détermine l'adhérence.

Nous définissons donc

$$F = \{x \in X/h^*(x) \text{ fini}\} \text{ et } I = \{x \in X/h^*(x) \text{ infini}\}.$$

Il est clair que

$$F = \{x \in X/\exists i \neq j \in \mathbb{N} \text{ tels que } h^i(x) = h^j(x)\},$$

$$I = \{x \in X/\forall i \neq j \in \mathbb{N}, h^i(x) \neq h^j(x)\} \quad \text{et} \quad X = F \cup I.$$

Remarquons aussi que $w \in F^*$ implique $h^n(w) \in F^*$, pour tout $n \in \mathbb{N}$. De plus, $w \in F^*$ implique l'existence de deux entiers $r_0 \in \mathbb{N}$ et $r \in \mathbb{N}_+$ vérifiant $h^{r_0}(w) = h^{r_0+r}(w)$. Enfin $x \in I$ implique que, pour tout entier $n \in \mathbb{N}$, $h^n(x)$ appartient à $X^* I X^* = F^* I X^*$.

DÉFINITION : Un DOL-système $G = \langle X, h, u \rangle$ est unitaire s'il existe une factorisation $u = u_1 u_2 u_3$ avec $u_1, u_2, u_3 \in X^*$, $h(u_1) = u_1$, $u_2 \in F^* \mid X^*$, $h(u_2) = w_1 u_2 w_3$ avec $w_3 \in X^*$, $w_1 \in F^*$, et $h(w_1) = w_1$.

Nous appellerons DOL-langage unitaire, un DOL-langage engendré par un DOL-système unitaire.

PROPRIÉTÉ 3 : Soit $G = \langle X, h, u \rangle$ un DOL-système unitaire engendrant le langage L . Alors, $\text{Adh}(L)$ contient exactement un élément et L^c est co-algébrique.

Preuve : Deux cas peuvent se présenter.

1. $w_1 = \varepsilon$.

Nous remarquons alors une certaine analogie entre le DOL-système G et ceux décrits par J. Berstel dans [3].

Soit g un homomorphisme défini sur $X \cup \{\theta\}$, où θ est une nouvelle lettre, par $g(\theta) = \theta w_3$ et $g(x) = x$, pour tout $x \in X$. Alors, le DOL-système $G_\theta = \langle X \cup \{\theta\}, g, \theta \rangle$ engendre le mot infini

$$\theta\beta = \theta w_3 g(w_3) g^2(w_3) \dots g^i(w_3) \dots$$

$= \theta w_3 h(w_3) \dots h^i(w_3) \dots$ et, d'après Berstel, $FG(\theta\beta)$ est co-algébrique.

Il s'agit de mettre en relation G_θ et G .

D'après l'hypothèse, u se factorise en $u_1 u_2 u_3$ avec u_2 engendrant un langage infini. Donc $G = \langle X, h, u_1 u_2 u_3 \rangle$ et $G' = \langle X, h, u_1 u_2 \rangle$ engendrent deux langages qui ont même adhérence (lemme 2).

Nous pouvons définir un homomorphisme non effaçant f tel que $f(L(G_\theta)) = L(G')$. Il suffit de poser $f(\theta) = u_1 u_2$ et $f(x) = x$, pour tout $x \in X$. D'après [6], $\text{Adh}(L(G')) = \text{Adh}(f(L(G_\theta))) = f(\text{Adh}(L(G_\theta))) = \{u_1 u_2 \beta\}$ et donc $\text{Adh}(L)$ contient exactement un élément.

La décomposition sur $X^* = FG(u_1 u_2) \cup u_1 u_2 X^* \cup \overline{FG(u_1 u_2 X^*)}$ nous permet d'écrire

$$\overline{L^c} = \overline{FG(u_1 u_2 \beta)} = u_1 u_2 \overline{FG(\beta)} \cup \overline{FG(u_1 u_2 X^*)}.$$

Le préfixe $u_1 u_2$ appartenant à X^* , $FG(u_1 u_2 X^*)$ est rationnel. De plus, nous avons

$$\overline{FG(\beta)} = X^* - FG(\beta) = \{z \in X^* / \theta z \in \overline{FG(\theta\beta)}\}$$

qui est algébrique car $\overline{FG(\theta\beta)}$ est algébrique. Nous en déduisons que

$$\overline{L^c} = u_1 u_2 \overline{FG(\beta)} \cup \overline{FG(u_1 u_2 X^*)}$$

est algébrique.

2. $w_1 \neq \varepsilon$.

Alors $\text{Adh}(L) = \{u_1 w_1^{n_0}\}$ et $L^c = FG(u_1 w_1^*)$ est rationnel.

Donc $\overline{L^c}$ est rationnel. \square

Montrons, maintenant, que tout DOL-langage infini est égal à l'union d'un langage fini et d'une union finie de DOL-langages unitaires.

PROPOSITION 4: Soit $G = \langle X, h, u \rangle$ un DOL-système. Si $L = L(G)$ est infini, alors il existe deux entiers $n \in \mathbb{N}_+$ et $n_0 \in \mathbb{N}$ tels que pour tout $i \in \mathbb{N}$, $G_i = \langle X, h^n, h^{n_0+i}(u) \rangle$ est unitaire.

Preuve. — Dans un premier temps, on établit le résultat suivant : il existe $x \in I$, $t_0 \in \mathbb{N}$ et $t \in \mathbb{N}_+$ tels que $h^{t_0}(u) = v_1 x v_3$ avec $v_1 \in F^*$, $v_3 \in X^*$, $h^t(x) = w'_1 x w'_3$, $w'_1 \in F^*$ et $w'_3 \in X^*$.

En effet, L infini implique que, quelque soit l'itération i , $h^i(u)$ appartient à $F^* I X^*$. Soit k le nombre d'éléments de I . Pour $i \in [0, k]$, $h^i(u)$ représente $k+1$ itérations. Donc un élément x de I a été utilisé deux fois en tant que première lettre de I apparaissant dans $h^i(u)$. Il existe donc deux entiers t et t_0 , avec $0 \leq t_0 < t_0 + t \leq k$ pour lesquels x est facteur de $h^{t_0}(u)$ et de $h^t(x)$ et, dans les deux cas, n'est précédé que de lettres de F .

Nous avons $v_1 \in F^*$ et $w'_1 \in F^*$; Il existe donc $p_0, q_0 \in \mathbb{N}$ et $p, q \in \mathbb{N}_+$ tels que

$$h^{p_0}(v_1) = h^{p_0+p}(v_1) \quad \text{et} \quad h^{q_0}(w'_1) = h^{q_0+q}(w'_1).$$

Posons $n_0 = p_0 + q_0 + t_0$ et $n = ptq$.

Pour tout $i \in \mathbb{N}$, nous avons

$$h^{n_0+i}(u) = h^{p_0+q_0+i}(h^{t_0}(u)) = h^{p_0+q_0+i}(v_1 x v_3).$$

Posons

$$u_1 = h^{p_0+q_0+i}(v_1), \quad u_2 = h^{p_0+q_0+i}(x) \quad \text{et} \quad u_3 = h^{p_0+q_0+i}(v_3).$$

Nous vérifions que :

$$(a) \quad h^n(u_1) = h^{ptq}(u_1) = h^{ptq+p_0+q_0+i}(v_1) \\ = h^{q_0+i}(h^{p_0+ptq}(v_1)) = h^{q_0+i}(h^{p_0}(v_1)) = u_1;$$

(b) puisque $x \in I$, nous avons $u_2 \in F^* I X^*$;

$$(c) \quad h^n(u_2) = h^{ptq}(h^{p_0+q_0+i}(x)) = h^{p_0+q_0+i+(pq-1)t}(w'_1 x w'_3)$$

$$= h^{p_0+q_0+i+(pq-1)t}(w'_1) h^{p_0+q_0+i+(pq-2)t}(w'_1) \dots h^{p_0+q_0+i}(w'_1) \\ h^{p_0+q_0+i}(x) h^{p_0+q_0+i}(w'_3) \dots h^{p_0+q_0+i+(pq-1)t}(w'_3) = w_1 u_2 w_3$$

en posant

$$w_1 = h^{p_0+q_0+i+(pq-1)t}(w'_1) \dots h^{p_0+q_0+i}(w'_1) \\ w_3 = h^{p_0+q_0+i}(w'_3) \dots h^{p_0+q_0+i+(pq-1)t}(w'_3).$$

(d) puisque $w'_1 \in F^*$, $h^{p_0+q_0+i+(pq-j)t}(w'_1) \in F^*$, pour tout $j \in [1, pq]$.

De plus, pour tout $j \in [1, pq]$, on a :

$$h^{pq}(h^{p_0+q_0+i+(pq-j)t}(w'_1)) = h^{p_0+i+(pq-j)t}(h^{q_0+pq}(w'_1)) \\ = h^{p_0+i+(pq-j)t}(h^{q_0}(w'_1)) \\ = h^{p_0+q_0+i+(pq-j)t}(w'_1).$$

Donc

$$h^n(w_1) = h^{pq}(w_1) = h^{pq}(h^{p_0+q_0+i+(pq-1)t}(w'_1) \dots h^{p_0+q_0+i}(w'_1)) = w_1.$$

Pour tout $i \in \mathbb{N}$, $G_i = \langle X, h^n, h^{n_0+i}(u) \rangle$ est donc unitaire. \square

Des propositions 3 et 4, nous pouvons déduire que l'adhérence d'un DOL-langage est toujours fini, résultat qui avait été annoncé dans [9].

En effet, si le DOL-langage est fini, son adhérence est vide. Dans le cas contraire, le DOL-langage est égal à l'union d'un langage fini, dont l'adhérence est vide, et de l'union, pour $i \in [0, n-1]$, des DOL-langages unitaires $L(G_i)$ définis par la proposition 4, dont l'adhérence contient un seul élément. Nous avons donc

PROPOSITION 5: *L'adhérence d'un DOL-langage est fini.*

Pour démontrer que le centre d'un DOL-langage est co-algébrique, nous utilisons le lemme suivant.

LEMME 6: *Soient $\alpha_1, \alpha_2, \dots, \alpha_n \in X^\omega$. Si, pour tout $i \in [1, n]$, $L_i = \overline{FG(\alpha_i)}$ est algébrique alors $L_1 \cap L_2 \cap \dots \cap L_n$ est encore algébrique.*

Preuve : On peut supposer, sans nuire à la généralité, que tous les α_i sont différents. Il existe alors un entier $k \in \mathbb{N}_+$ tel que les préfixes de longueur k des α_i sont différents. Posons $\alpha_i = \alpha_i[k] \beta_i$, pour tout $i \in [1, n]$. $i \neq j$ implique $\alpha_i[k] \neq \alpha_j[k]$.

X^* peut s'écrire

$$X^* = \left[\bigcup_{i=1}^n FG(\alpha_i[k]) \right] \cup \left[\bigcup_{i=1}^n \alpha_i[k] X^* \right] \cup \left[\bigcap_{i=1}^n \overline{FG(\alpha_i[k] X^*)} \right]$$

en fonction des différents préfixes $\alpha_i[k]$. Le langage $L = \bigcap_{i=1}^n L_i$ est examiné par rapport à cette décomposition.

$L \cap FG(\alpha_i[k])$ est inclus dans $L_i \cap FG(\alpha_i[k])$ qui est vide, et ceci pour tout $i \in [1, n]$.

$$L \cap \alpha_i[k] X^* = L_i \cap \alpha_i[k] X^* \text{ pour tout } i \in [1, n].$$

Enfin

$$L \cap \left[\bigcap_{i=1}^n \overline{FG(\alpha_i[k] X^*)} \right] = \bigcap_{i=1}^n \overline{FG(\alpha_i[k] X^*)}$$

puisqu'il y a inclusion.

Donc nous pouvons écrire

$$L = \left[\bigcup_{i=1}^n (L_i \cap \alpha_i[k] X^*) \right] \cup \left[\bigcap_{i=1}^n \overline{FG(\alpha_i[k] X^*)} \right].$$

L_i étant algébrique, $\alpha_i[k] X^*$ et $\overline{FG(\alpha_i[k] X^*)}$ étant rationnels, pour tout $i \in [1, n]$, nous déduisons que $L = \bigcap_{i=1}^n L_i$ est algébrique. \square

Nous avons vu que si l'adhérence n'était pas vide, elle pouvait être déterminée par une union finie de DOL-langages unitaires :

$$\text{Adh}(L(G)) = \bigcup_{i=0}^{n-1} \text{Adh}(L(G_i)) = \{\alpha_0, \alpha_1, \dots, \alpha_m\} \quad \text{avec } m \leq n-1$$

(propositions 1 et 4). Chaque α_j est l'adhérence d'un $L(G_i)$ et, G_i étant unitaire, $\overline{L(G_i)^c} = \overline{FG(\alpha_j)}$ est algébrique (proposition 3). Le complémentaire d'une union finie étant l'intersection des complémentaires,

$$\overline{L(G)^c} = \overline{\bigcup_{j=0}^m FG(\alpha_j)} = \bigcap_{j=0}^m \overline{FG(\alpha_j)},$$

le lemme 6 nous permet de conclure par la proposition suivante.

PROPOSITION 7: *Le centre d'un DOL-langage est co-algébrique.*

IV. REMARQUES

Il est intéressant de signaler que cette propriété d'algébricité ne reste pas vraie si l'on prend le complémentaire des facteurs de l'adhérence d'un DOL-langage. Il suffit de prendre le DOL-système suivant pour s'en convaincre : $G = \langle \{a, b\}, h, a \rangle$ avec $h(a) = aba$ et $h(b) = b^2$. On vérifiera aisément que $\overline{F(\text{Adh}(L(G)))} \cap ab^*a = \{ab^m a \mid n \in \mathbb{N} \text{ tel que } m = 2^n\}$ qui n'est pas algébrique.

Un autre mécanisme de génération de mots infinis est fourni par les Tag-systèmes [8]. Un Tag-système est un quintuplet $T = \langle X, h, x_0, Y, g \rangle$ où X, Y sont des alphabets finis, h est un morphisme sur X prolongeable en x_0 , c'est-à-dire $h(x_0) = x_0 u$ avec $u \in X^+$, g est un morphisme littéral (strictement alphabétique) de X sur Y . Un Tag-système est uniforme si toutes les lettres de X ont par h des images de même longueur. Il est montré dans [5] que certains mots infinis peuvent être engendrés par des Tag-systèmes uniformes sans pouvoir l'être par des DOL-systèmes. En s'inspirant de la démonstration de Berstel pour les morphismes prolongeables [3] on montre facilement que le complémentaire des facteurs gauches de la séquence engendrée par un Tag-système uniforme est algébrique. Pour les Tag-systèmes quelconques, cette méthode ne permet pas d'obtenir un résultat. Cependant les exemples de mots infinis dont le complémentaire des facteurs gauches n'est pas algébrique [10], nous incitent à émettre la conjecture suivante :

CONJECTURE : Le complémentaire des facteurs gauches d'une séquence engendrée par un Tag-système est algébrique.

REMERCIEMENTS

Je tiens à remercier le rapporteur pour ses remarques et, plus particulièrement, pour une simplification de la preuve de la Proposition n° 4.

BIBLIOGRAPHIE

1. J. M. AUTEBERT, J. BEAUQUIER, L. BOASSON et M. NIVAT, *Quelques problèmes ouverts en théorie des langages algébriques*, R.A.I.R.O. Informatique Théorique, vol. 13, 1979, p. 363-379.
2. J. M. AUTEBERT, J. BEAUQUIER, L. BOASSON et M. LATTEUX, *Very Small Families of Algebraic Nonrational Languages*, in Formal Language Theory, Perspectives and Open Problems, R. V. BOOK éd., Academic-Press, 1980, p. 89-107.
3. J. BERSTEL, *Every Iterated Morphism Yields a co-CFL*, Information Processing Letters, vol. 22, 1986, p. 7-9.
4. J. BERSTEL, *Exposé aux journées du groupe LANFOR*, Bayonne, 20-22 mai 1985.

5. J. BERSTEL, *Mots sans carré et morphismes itérés*, Discrete Math., vol. 29, 1980, p. 235-244.
6. L. BOASSON et M. NIVAT, *Adherences of Languages*, J. Comput. System Sci. vol. 20, 1980, p. 285-309.
7. W. BUCHER, D. HAUSSLER et M. G. MAIN, *Applications of an Infinite Square Free co-CFL*, 12^e I.C.A.L.P., Napflion, 15-19 juillet 1985.
8. A. COBHAN, *Uniform Tag sequences*, Math. Systems Theory, vol. 6, 1972, p. 164-192.
9. K. CULIK II et A. SALOMAA, *On Infinite Words Obtained by Iterating Morphisms*, Theoret. Comput. Sci., vol. 19, 1982, p. 29-38.
10. A. GRAZON, *Contribution à l'étude des petites familles de langages*, Thèse de 3^e cycle, Université de Paris-VII, 1985.
11. T. HEAD, *Adherences of DOL-Languages*, Theoret. Compt. Sci., vol. 31, 1984, p. 139-149.
12. G. T. HERMAN et G. ROZENBERG, *Developmental Systems and languages*, North-Holland, Amsterdam, 1975.
13. M. G. MAIN, *An Infinite Square Free co-CFL*, 1985, Informatique Processing Letters (à paraître).
14. G. ROZENBERG et A. SALOMAA, *The Mathematical Theory of L Systems*, Academic-Press, New York, 1980.