

MAURICE NIVAT

**Sur les ensembles de mots infinis engendrés par  
une grammaire algébrique**

*RAIRO. Informatique théorique*, tome 12, n° 3 (1978), p. 259-278

[http://www.numdam.org/item?id=ITA\\_1978\\_\\_12\\_3\\_259\\_0](http://www.numdam.org/item?id=ITA_1978__12_3_259_0)

© AFCET, 1978, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Informatique théorique » 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 ENSEMBLES DE MOTS INFINIS ENGENDRÉS PAR UNE GRAMMAIRE ALGÈBRIQUE (\*)

par Maurice NIVAT (<sup>1</sup>)

Communiqué par J. BERSTEL

---

Résumé. — Dans un article précédent paru ici même, nous avons défini l'ensemble des mots infinis engendrés par une grammaire algébrique. Nous étudions ici les  $\omega$ -langages algébriques, c'est-à-dire les langages algébriques constitués de mots finis et infinis.

### INTRODUCTION

Nous avons défini en [4], paru ici même, l'ensemble  $L^\omega(G, f)$  des mots infinis engendrés par une grammaire algébrique à partir d'un mot  $f$  (l'axiome).

Ci-dessous nous étudions les  $\omega$ -langages algébriques, c'est-à-dire les langages de la forme  $L^\omega(G, f) = L^\omega(G, f) \cup L(G, f)$  formés des mots finis et infinis engendrés par une grammaire algébrique. La notion essentielle est celle d'adhérence qui se relie immédiatement à celle de clôture [2] ou celle équivalente de limite [3]. Si la clôture d'une partie  $L$  de  $X^*$  est définie comme l'ensemble des mots infinis sur  $X$  qui ont une infinité de facteurs gauches dans  $L$ , l'adhérence de  $L$  est simplement la clôture de  $FG(L)$ , ensemble des facteurs gauches de mots de  $L$ .

Notre principal résultat (th. 6) caractérise les  $\omega$ -langages algébriques qui peuvent être engendrés par une grammaire de Greibach réduite comme ceux qui satisfont la condition

$$L \cap X^\omega = \text{Adh}(L \cap X^*)$$

La nécessité de cette condition rend claire la différence qui existe entre  $\omega$ -langages algébriques et langages algébriques ordinaires composés uniquement de mots finis : quand on met une grammaire algébrique sous forme de

---

(\*) Reçu décembre 1977 et en version définitive avril 1978.

(<sup>1</sup>) Laboratoire d'Informatique théorique et Programmation, Université Paris-VII et I.R.I.A.-Laboria.

Greibach on modifie l'ensemble des mots infinis qu'elle engendre. Et ceci explique certains phénomènes étranges étudiés par R. Cohen et A. Gold [1] ou M. Linna [3].

L'essentiel de notre article est la démonstration d'un théorème de substitution (th. 1) que nous donnons complète car elle montre bien où sont les difficultés que rencontre la théorie des  $\omega$ -langages algébriques : il ne suffit pas en effet de montrer l'existence d'une dérivation infinie il faut encore établir sa convergence c'est-à-dire montrer que l'ensemble des facteurs gauches terminaux des termes de la dérivation est infini.

Un corollaire de ce théorème est que tout mot infini produit par une dérivation quelconque l'est aussi par une dérivation gauche (en anglais "leftmost") et nos résultats se séparent là de ceux de [1].

L'algébricité de  $\hat{L}$  qui est le plus petit langage fermé par facteur gauche dont l'adhérence soit égale à celle d'un langage algébrique  $L$  (th. 4) ouvre la voie à une étude des parties de  $X^\omega$  qui sont adhérences de langages algébriques, étude qui fait l'objet de travaux de l'auteur, à paraître.

Nos définitions sont celles de [4], à quelques détails près : nous avons trouvé commode de noter  $\leq$  la relation « est un facteur gauche de » définie par

$$\forall f, g \in X^*, \quad f \leq g \Leftrightarrow \exists h \in X^* : g = fh.$$

Ainsi  $f < g$  signifie  $f \leq g \wedge f \neq g$  que l'on lit  $f$  est un facteur gauche propre de  $g$  ou encore

$$f < g \Leftrightarrow \exists h \in XX^* : g = fh.$$

Nous notons  $X^\infty = X^* \cup X^\omega$  et pour toute grammaire algébrique  $G : \xi_i = p_i, i \in [N]$  et  $f \in (X \cup \Xi)^*$ ,  $L^\infty(G, f) = L^\omega(G, f) \cup L(G, f)$ .

Nous appelons  $\omega$ -langage algébrique toute partie  $L$  de  $X^\infty$  telle qu'il existe une grammaire algébrique  $G$  et un mot  $f$  satisfaisant  $L = L^\infty(G, f)$ .

Si pour tout  $t \in (X \cup \Xi)^*$  nous notons  $\alpha(t)$  le plus long facteur gauche de  $t$  qui soit dans  $X^*$  (c'est-à-dire le plus long facteur gauche terminal) la dérivation  $\delta = \langle t_n \mid n \in \mathbf{N}_+ \rangle$  de  $G$  est dite gauche si et seulement si pour tout  $n \in \mathbf{N}_+$ , si  $t_n = \alpha(t_n) \xi_n \beta_n$  il existe  $m \in P_{i_n}$  tel que  $t_{n+1} = \alpha(t_n) m \beta_n$ .

La démonstration du théorème 1 nous oblige maintenant à donner de nombreuses définitions concernant les arbres de dérivations finis et infinis.

## I. ARBRES ET ARBRES DE DÉRIVATIONS

$\mathbf{N}_+$  dénote l'ensemble des entiers strictement positifs.

Un domaine d'arbre est une partie  $A$  de  $(\mathbf{N}_+)^*$  qui satisfait les deux

conditions :

$$1^\circ \forall u \in (\mathbf{N}_+)^*, n \in \mathbf{N}_+, un \in A \Rightarrow u \in A;$$

$$2^\circ \forall u \in (\mathbf{N}_+)^*, n, n' \in \mathbf{N}_+, un \in A \text{ et } n' < n \Rightarrow un' \in A.$$

L'ensemble des arbres sur  $X$ , où  $X$  est un alphabet fini, est l'ensemble des applications partielles de  $(\mathbf{N}_+)^*$  dans  $X \cup \{\varepsilon\}$  dont le domaine est un domaine d'arbre. On note  $X^{\natural}$  l'ensemble des arbres sur  $X$ .

Si  $\sigma$  est un arbre sur  $X$ ,  $A(\sigma)$  désigne son domaine. Les éléments de  $A(\sigma)$  sont appelés sommets ou nœuds de l'arbre  $\sigma$ . Les sommets de  $A(\sigma) \cap u\mathbf{N}_+$  sont dits fils du sommet  $u$  qui réciproquement est dit être leur père. Le sommet  $\varepsilon$  est appelé racine de l'arbre.

Si  $A$  est un domaine d'arbre, nous dirons que  $A'$  est un sous-domaine d'arbre de  $A$  si et seulement si :

- $A'$  est un domaine d'arbre;
- $\forall u \in A' u\mathbf{N}_+ \cap A' \neq \emptyset \Rightarrow u\mathbf{N}_+ \cap A' = u\mathbf{N}_+ \cap A$ .

L'arbre  $\sigma'$  est dit sous-arbre initial de  $\sigma$ , ce que l'on note  $\sigma' < \sigma$  si et seulement si :

- $A(\sigma')$  est un sous-domaine d'arbre de  $A(\sigma)$ ;
- $\sigma'$  est la restriction de  $\sigma$  à  $A(\sigma')$  ce que l'on note  $\sigma' = \sigma|_{A(\sigma')}$ .

La relation « est un sous-arbre initial de » est évidemment une relation d'ordre. Nous avons la :

**PROPRIÉTÉ 1 :** Tout sous-ensemble dirigé d'arbres, dirigé pour la relation « est un sous-arbre initial de », admet une borne supérieure.

*Démonstration :* Un sous-ensemble dirigé pour  $<$  est une partie  $\Delta$  de  $X^{\natural}$  telle que

$$\forall \sigma_1, \sigma_2 \in \Delta, \exists \sigma_3 \in \Delta \text{ tel que } \sigma_1 < \sigma_3 \text{ et } \sigma_2 < \sigma_3.$$

On note qu'une suite croissante  $\langle \sigma_n | n \in \mathbf{N} \rangle$  d'arbres est une partie dirigée de  $X^{\natural}$ .

Soit donc  $\Delta$  une partie dirigée de  $X^{\natural}$ .

On vérifie d'abord que  $A = \bigcup_{\sigma \in \Delta} A(\sigma)$  est un domaine d'arbre. En effet

$un \in A \Rightarrow un \in A(\sigma)$  on en déduit  $u \in A$ . De même  $un \in A$  et  $n' < n$  impliquent  $un \in A(\sigma)$  pour quelque  $\sigma \in \Delta$  et  $n' < n$  d'où  $un' \in A(\sigma)$  et par suite  $un' \in A$ .

Maintenant si  $\Delta$  est dirigé,  $\sigma_1$  et  $\sigma_2$  appartiennent à  $\Delta$ ,  $u \in A(\sigma_1) \cap A(\sigma_2)$  on a  $\sigma_1(u) = \sigma_2(u)$ . En effet si  $\sigma_3$  est tel que  $\sigma_1 < \sigma_3$  et  $\sigma_2 < \sigma_3$ ,  $A(\sigma_1)$  et  $A(\sigma_2)$  sont des sous-domaines d'arbre de  $A(\sigma_3)$  ce qui implique que pour tout  $u \in A(\sigma_1) \cap A(\sigma_2)$ ,  $\sigma_3(u)$  est défini et égal à  $\sigma_1(u)$  et  $\sigma_2(u)$ .

D'où la borne supérieure  $\text{Sup } \Delta$  : son domaine est  $A(\text{Sup } \Delta) = \bigcup_{\sigma \in \Delta} A(\sigma)$  et pour tout  $u \in A(\text{Sup } \Delta)$ ,  $(\text{Sup } \Delta)(u)$  est égal à la valeur commune de tous les  $\sigma(u)$ , où  $\sigma \in \Delta$  est tel que  $u \in A(\sigma)$ .  $\square$

Entre les sommets d'un arbre  $\sigma$  nous avons les deux relations :

- $u \leq u'$  que l'on lit  $u$  est au-dessus de  $u'$  ou  $u'$  est en dessous de  $u$  si et seulement si  $u$  est facteur gauche de  $u'$ . On note  $u < u'$  la relation  $u \leq u'$  et  $u \neq u'$ ;
- $u <_g u'$  que l'on lit  $u$  « est à gauche » de  $u'$  si et seulement si  $\exists u_1, u_2, u'_2 \in (\mathbb{N}_+)^*$ ,  $n, n' \in \mathbb{N}_+$  tels que  $n < n'$ ,  $u = u_1 n u_2$  et  $u' = u_1 n' u'_2$ .

On remarquera que la relation  $(\leq \cup <_g)$  est une relation d'ordre : c'est la relation d'ordre lexicographique  $\leq_l$  sur  $(\mathbb{N}_+)^*$ .

Si  $\sigma$  est un arbre,  $u \in A(\sigma)$  un sommet, on vérifie immédiatement que  $A(\sigma : u) = \{v \mid uv \in A(\sigma)\}$  est un domaine d'arbre et l'on définit le sous-arbre issu de  $u$ , noté  $\sigma : u$  comme l'arbre de domaine  $A(\sigma : u)$  donné par  $(\sigma : u)(v) = \sigma(uv)$ .

Nous aurons besoin de définir encore une espèce de sous-arbres.

Si  $u \in A(\sigma)$  l'ensemble  $\{v \in A(\sigma) \mid v \leq u \text{ ou } v <_g u\}$  est un domaine d'arbre, dont on remarquera que ce n'est pas un sous-domaine d'arbre de  $A(\sigma)$ .

La restriction de  $\sigma$  à ce domaine est notée  $SG(\sigma, u)$  et appelée sous-arbre à gauche de  $u$ .

Il est facile d'établir que si  $\sigma'$  est un sous-arbre initial de  $\sigma$  et si  $u \in A(\sigma')$  alors  $SG(\sigma', u)$  est un sous-arbre initial de  $SG(\sigma, u)$ .

Nous appellerons branche d'un arbre  $\sigma$  toute suite de sommets tels que chacun soit fils de celui qui le précède et le premier soit  $\varepsilon$ .

Ainsi  $b = \langle u_1, \dots, u_n \rangle$  est une branche finie si et seulement si  $u_1 = \varepsilon$  et pour tout  $i \in [(n-1)]$ ,  $u_{i+1} \in u_i \mathbb{N}_+$ . Le sommet  $u_n$  est dit extrémité de la branche  $b$  et il est clair qu'il existe une et une seule branche d'extrémité  $u$  pour tout  $u \in A(\sigma)$ .

La branche  $b = \langle u_n \mid n \in \mathbb{N}_+ \rangle$  est une branche infinie si et seulement si  $n_1 = \varepsilon$  et pour tout  $i \in \mathbb{N}_+$ ,  $u_{i+1} \in u_i \mathbb{N}_+$ .

Définissons maintenant la frontière et le feuillage d'un arbre  $\sigma$ .

La frontière de  $\sigma$  notée  $fr(\sigma)$  est l'ensemble

$$fr(\sigma) = \{u \in A(\sigma) \mid u \mathbb{N}_+ \cap A(\sigma) = \emptyset\}.$$

Le feuillage de  $\sigma$  noté  $fg(\sigma)$  est un mot de  $X^*$  que l'on définit par récurrence pour tout  $\sigma$  de domaine fini de la façon suivante :

- si  $A(\sigma) = \emptyset$ ,  $fg(\sigma) = \varepsilon$ ;
- si  $A(\sigma) = \{\varepsilon\}$ ,  $fg(\sigma) = \sigma(\varepsilon)$ ;

– si  $A(\sigma) \cap \mathbf{N} = [p]$ ;

$$fg(\sigma) = fg(\sigma : 1) fg(\sigma : 2) \dots fg(\sigma : p).$$

Nous laisserons le soin au lecteur de vérifier les propriétés suivantes :

PROPRIÉTÉ 2 : Si  $u_1, u_2, \dots, u_p$  est la suite ordonnée dans l'ordre lexicographique  $\leq_l$  des sommets de  $fr(\sigma)$  où  $\sigma$  est un arbre fini alors :

$$fg(\sigma) = \sigma(u_1) \sigma(u_2) \dots \sigma(u_p).$$

PROPRIÉTÉ 3 : Pour tout arbre de domaine fini  $\sigma$ ,  $\alpha, \beta \in X^*$ ,  $x \in X$ , la condition  $\alpha x \beta = fg(\sigma)$  implique l'existence d'une et une seule branche  $b = \langle u_1, \dots, u_n \rangle$  de  $\sigma$  telle que

$$fg(SG(\sigma, u_n)) = \alpha x.$$

Par contre nous allons démontrer les deux propriétés suivantes :

PROPRIÉTÉ 4 : Si  $\sigma$  vérifie la condition

$$\forall u \in A(\sigma) : \text{card}(u \mathbf{N}_+ \cap A(\sigma)) < \infty$$

on dit alors que  $\sigma$  est de type fini. Si  $\sigma$  est infini et de type fini,  $\sigma$  possède une branche infinie.

PROPRIÉTÉ 5 : Disons que la branche infinie  $\langle v_n \mid n \in \mathbf{N} \rangle$  est à gauche de la branche infinie  $\langle u_n \mid n \in \mathbf{N} \rangle$  si et seulement si  $\forall n \in \mathbf{N} v_n \leq u_n$  ou  $v_n <_g u_n$ .

Alors si  $\sigma$  est un arbre infini et de type fini il existe une branche infinie plus à gauche que toutes les autres dite la branche infinie la plus à gauche de  $\sigma$ .

La propriété 4 est une forme du lemme de Koenig que nous énonçons sous la forme suivante, qui nous a déjà servi en [4] et nous servira plus loin.

LEMME DE KOENIG : Soit  $E$  un ensemble infini,  $R$  une relation sur  $E$ , une suite  $A_1, A_2, \dots, A_n, \dots$  de parties finies non vides de  $E$  telles que :

–  $\bigcup_{i \leq 1} A_i$  est de cardinalité infinie;

–  $\forall n \geq 1, y \in A_{n+1}, \exists x \in A_n : x R y$

alors il existe une suite infinie d'éléments  $a_1, a_2, \dots, a_n, \dots$  telle que :

–  $\forall n \in \mathbf{N}_+, a_n \in A_n$  et  $a_n R a_{n+1}$ .

Démonstration : Si  $\sigma$  est un arbre infini de type fini, définissons la suite des ensembles  $A_n$  par

$$A_n = A(\sigma) \cap \left( \bigcup_{k=0}^n (\mathbf{N}_+)^k \right),$$

qui est l'ensemble des sommets dits de hauteur inférieure ou égale à  $k$ . Chacun de ces ensembles est fini et non vide.

En effet  $A_0$  est fini. Et si  $A_n$  est fini,

$$A_{n+1} = A_n \cup \left( \bigcup_{n \in A_n} (u \mathbf{N}_+ \cap A(\sigma)) \right)$$

est fini puisque  $\sigma$  est de type fini. D'où l'on déduit par récurrence que  $A_n$  est fini pour tout  $n \in \mathbf{N}$ .

Mais ceci entraîne que  $A_n$  aussi est non vide, pour tout  $n$ , puisque  $A_n = \emptyset \Rightarrow A_{n+1} = \emptyset$  et si  $A_n$  était vide pour quelque  $n$  l'arbre serait fini.

Considérons alors la relation  $R$  définie par

$$(u, v) \in R \Leftrightarrow v \in u \mathbf{N}_+.$$

Pour tout  $v \in A_{n+1}$  il existe  $u \in A_n$  tel que  $(u, v) \in R$ , il suffit de prendre  $u$  égal au facteur gauche de longueur  $n$  de  $v$ .

Le lemme de Koenig fournit alors le résultat.  $\square$

Pour démontrer la propriété 5 on construit la branche infinie la plus à gauche de la manière suivante : disons que la branche  $b = \langle u_n \mid n \in \mathbf{N} \rangle$  passe par le sommet  $u$  si et seulement si il existe  $n \in \mathbf{N}$  tel que  $u = u_n$ .

Appelons alors  $v_n$  le sommet le plus à gauche de  $A_n$  par lequel passe une branche infinie de  $\sigma$  supposé infini et de type fini. La suite  $\langle v_n \mid n \in \mathbf{N} \rangle$  est une branche infinie, c'est-à-dire satisfait

$$\forall n \in \mathbf{N}, v_{n+1} \in v_n \mathbf{N}_+.$$

En effet  $v_{n+1} \in v' \mathbf{N}_+$  avec  $v' <_g v_n$  implique que  $v_n$  n'est pas le sommet le plus à gauche de  $A_n$  par lequel passe une branche infinie et  $v_{n+1} \in v' \mathbf{N}_+$  avec  $v_n <_g v'$  contredit l'hypothèse que  $v_{n+1}$  est le sommet le plus à gauche de  $A_{n+1}$  par lequel passe une branche infinie.  $\square$

De la construction précédente découle facilement la :

PROPRIÉTÉ 6 : Si  $\sigma$  est un arbre infini de type fini, sa branche infinie la plus à gauche est égale à la suite ordonnée par  $\leq$  des sommets de

$$B = \{ u \in A(\sigma) \mid A(\sigma : u) \text{ est infini et } \forall u', u' <_g u \Rightarrow A(\sigma : u') \text{ est fini} \}.$$

*Démonstration* : Si  $b$  est la branche infinie la plus à gauche de  $\sigma$ ,  $b = \langle u_n \mid n \in \mathbf{N}_+ \rangle$  il est clair que pour tout  $n$ ,  $A(\sigma : u_n)$  est infini et pour tout  $u'$  tel que  $u' <_g u_n$   $A(\sigma : u')$  est fini. Et si au contraire  $u \in A(\sigma)$  est un sommet tel qu'il existe  $u' <_g u$  avec  $A(\sigma : u')$  infini alors  $u$  n'appartient pas à la branche infinie la plus à gauche.  $\square$

On retiendra que le sous-arbre issu de tout sommet à gauche de la branche infinie la plus à gauche est fini.

**PROPRIÉTÉ 7 :** Pour toute branche  $b$  de l'arbre  $\sigma \in X^{\mathbb{N}}$  il existe un arbre minimal dans l'ensemble des sous-arbres initiaux de  $\sigma$  qui contiennent la branche  $b$ .

*Démonstration :* Soit  $b$  une branche infinie  $b = \langle u_n \mid n \in \mathbb{N} \rangle$ .

Appelons  $b_n$  la branche finie  $\langle u_1, \dots, u_n \rangle$ .

Nous construisons par récurrence  $\sigma_1, \dots, \sigma_n, \dots$  sous-arbres initiaux minimaux de  $\sigma$  tels que  $b_n$  soit une branche de  $\sigma_n$  et  $\sigma_1 < \sigma_2 < \dots < \sigma_n < \dots$ .

Pour  $\sigma_1$  on prend  $A(\sigma_1) = \varepsilon$  et  $\sigma_1(\varepsilon) = \sigma(\varepsilon)$ .

Pour  $n+1$ , en supposant construit  $\sigma_n$ , on prend

$$A(\sigma_{n+1}) = A(\sigma_n) \cup (u_n \mathbf{N}_+ \cap A(\sigma)) \quad \text{et} \quad \sigma_{n+1} = \sigma \upharpoonright A(\sigma_{n+1}).$$

Il est clair que  $\sigma_n < \sigma_{n+1}$  et  $\sigma_{n+1}$  est minimal puisque si  $\tau$  contient la branche  $b_{n+1}$  il contient la branche  $b_n$  donc, par récurrence,  $\sigma_n < \tau$ . Et si  $A(\tau)$  contient  $u_{n+1} = u_n p$  il doit aussi contenir  $u_n \mathbf{N}_+ \cap A(\sigma)$  donc  $A(\sigma_{n+1})$ .

A la limite  $\text{Sup}\{\sigma_n\}$  est minimal dans l'ensemble des sous-arbres qui contiennent la branche  $b$ .  $\square$

### Arbres de dérivation d'une grammaire algébrique $G$

Posons pour tout arbre  $\sigma \in X^{\mathbb{N}}$  et  $u \in A(\sigma)$  :

$$\pi_\sigma(u) = \sigma(u_1) \sigma(u_2) \dots \sigma(u_p) \quad \text{si} \quad [p] = A(\sigma) \cap u \mathbf{N}_+.$$

$\pi_\sigma(u)$  sera appelé descendante directe de  $u$  dans  $\sigma$ .

Soit  $G : \xi_i = P_i, i \in [N]$  une grammaire algébrique.

Un arbre  $\sigma$  sur  $(X \cup \Xi)$  est dit arbre de dérivation de  $G$  si et seulement si :

- $A(\sigma) \cap u \mathbf{N}_+ \neq \emptyset \Rightarrow \sigma(n) \in \Xi$ ;
- $\sigma(u) = \xi_i \Rightarrow \pi_\sigma(u) = \varepsilon$  ou  $\pi_\sigma(u) \in P_i$ .

On notera que tout arbre de dérivation de  $G$  est de type fini. Nous dirons qu'un arbre de dérivation  $\sigma$  de  $G$  contient une dérivation  $\delta = \langle t_1, \dots, t_n \rangle$  ou  $\delta = \langle t_n \mid n \in \mathbb{N} \rangle$  de  $G$  si et seulement si il existe une suite croissante de sous-arbres initiaux  $\sigma_1 < \sigma_2 < \dots < \sigma_n < \dots$  de  $\sigma$  telle que :

- $\forall i \in [n]$  (ou  $\forall i \in \mathbb{N}_+$ ) :  $fg(\sigma_i) = t_i$ ;

Nous avons les propriétés suivantes :

**PROPRIÉTÉ 8 :** Pour toute dérivation finie ou infinie  $\delta$  de  $G$ , d'origine  $\xi_i \in \Xi$  il existe un arbre de dérivation  $\sigma$  de  $G$  qui contient  $\delta$ .



*Démonstration* : On construit  $\sigma$  de la façon suivante : soit

$$\delta = \langle t_1, \dots, t_n, \dots \rangle, \quad \delta_n = \langle t_1, \dots, t_n \rangle.$$

On construit  $\sigma_n$  qui contient  $\delta_n$  de façon que  $\sigma_n$  soit un sous-arbre initial de  $\sigma_{n+1}$ .

Pour  $\sigma_1$  on prend l'arbre de domaine  $\{\varepsilon\}$  tel que  $\sigma_1(\varepsilon) = \xi_i$ .

Si l'on suppose construit  $\sigma_n$  et l'on suppose

$$t_n = \alpha \xi_j \beta, \quad t_{n+1} = \alpha m \beta \text{ pour quelque } m \in P_j,$$

on sait par la propriété 3 qu'il existe une et une seule branche  $b_n = \langle u_n^1, \dots, u_n^p \rangle$  telle que

$$fg(SG(\sigma_n, u_n^p)) = \alpha \quad \text{et} \quad \sigma(u_n^p) = \xi_j,$$

On construit alors  $\sigma_{n+1}$  de domaine

$$A(\sigma_n) \cup u_n^p[|m|]$$

avec

$$\sigma_{n+1}|A(\sigma_n) = \sigma_n \quad \text{et} \quad \sigma_{n+1}(u_n^p h) = \lambda_h m \text{ pour tout } h \in [|m|].$$

La notation  $\lambda_h m$  désigne la  $h$ ème lettre de  $m$ .

Il est clair que  $\sigma_n < \sigma_{n+1}$  et  $\sigma_{n+1}$  contient  $\delta_{n+1}$ .

La borne supérieure  $\text{Sup}\{\sigma_n\}$  contient la dérivation  $\delta$ .  $\square$

**PROPRIÉTÉ 9** : Si  $\sigma'$  est un sous-arbre initial fini de  $\sigma$ , où  $\sigma$  est un arbre de dérivation de  $G$ , il existe une dérivation  $\delta$  dans  $G$  de  $\sigma(\varepsilon)$  en  $fg(\sigma')$  contenue dans  $\sigma$ .

*Démonstration* : On fait une récurrence sur  $\text{card } A(\sigma')$ .

Si  $\text{card } A(\sigma') = 1$ ,  $A(\sigma') = \{\varepsilon\}$  c'est clair.

Sinon considérons  $A(\sigma')$ . Il existe dans  $\text{fr}(\sigma')$ , qui est fini, des mots de longueur maximale c'est-à-dire des mots  $u$  tels que  $v \in \text{fr}(\sigma') \Rightarrow |v| \leq |u|$ .

Prenons-en un, soit  $u = u' p$ ,  $u' \in (\mathbf{N}_+)^*$ ,  $p \in \mathbf{N}_+$ .

Il existe une dérivation de  $\xi_i$  en  $fg(\sigma'')$  où  $\sigma'' = \sigma|(A(\sigma') \setminus u' \mathbf{N}_+)$  contenue dans  $\sigma$  soit  $\delta = \langle t_1, \dots, t_{k+1} \rangle$ . Mais si  $\sigma(u') = \xi_j$  on a  $t_{k+1} = fg(\sigma'') = \alpha \xi_j \beta$  pour quelque  $\alpha, \beta$ , et si l'on pose  $t_{k+2} = fg(\sigma') = \alpha \pi_\sigma(u') \beta$  il est clair que  $\langle t_1, \dots, t_{k+2} \rangle$  est une dérivation de  $\xi_i$  en  $fg(\sigma')$  contenue dans  $\sigma$ .  $\square$

## II. THÉORÈME DE SUBSTITUTION

Nous devons d'abord définir la substitution.

Soit  $u \in (X \cup \Xi)^\circ$  et  $\vec{Q} = \langle Q_1, \dots, Q_N \rangle$  sur  $N$ -vecteur de parties de  $X^*$ .

Nous notons  $u[\vec{Q}/\vec{\xi}]$  la substitution de  $\vec{Q}$  à  $\vec{\xi}$  dans  $u$  définie par :

- $u[\vec{Q}/\vec{\xi}] \subset X^\omega$ ;
- $\forall v \in X^\omega, v \in u[\vec{Q}/\vec{\xi}]$  si et seulement si il existe une suite d'entiers  $p_n$  telle que :

$$p_n \xrightarrow[n \rightarrow \infty]{} \infty \quad \text{et} \quad \forall n, v[p_n] \in u[n][\vec{Q}/\vec{\xi}].$$

La notation  $u[n]$  désigne le facteur gauche de longueur  $n$  de  $u$ .

Une définition équivalente est donnée par le :

LEMME 1 : Soit  $u \in (X \cup \Xi)^\omega$  le mot

$$u = \sigma_1 \xi_{i_1} \alpha_2 \xi_{i_2} \dots \alpha_p \xi_{i_p} \alpha_{p+1} \dots,$$

où pour tout  $p \in \mathbb{N}_+$ ,  $\alpha_p \in X^*$  et  $\xi_{i_p} \in \Xi$ .

Alors le mot  $v \in X^\omega$  appartient à  $u[\vec{Q}/\vec{\xi}]$  si et seulement si il existe  $h_1, h_2, \dots, h_p, \dots$  tels que

$$v = \alpha_1 h_1 \alpha_2 h_2 \dots \alpha_p h_p \alpha_{p+1} \dots \quad \text{et} \quad \forall p \in \mathbb{N}_+, h_p \in Q_{i_p}.$$

Démonstration : Dans un sens c'est évident puisque

$$\alpha_1 h_1 \dots \alpha_p h_p \in \alpha_1 \xi_{i_1} \dots \alpha_p \xi_{i_p} [\vec{Q}/\vec{\xi}]$$

et que  $v = \alpha_1 h_1 \dots \alpha_p h_p \alpha_{p+1} \dots$  implique que

$$|\alpha_1 h_1 \dots \alpha_p h_p| \rightarrow \infty \quad \text{si} \quad p \rightarrow \infty.$$

Réciproquement, soit  $p_n$  la suite telle que

$$v[p_n] \in u[n][\vec{Q}/\vec{\xi}].$$

Considérons la suite des entiers  $n_1, \dots, n_k, \dots$  tels que

$$u[n_k] = \alpha_1 \xi_{i_1} \dots \alpha_k \alpha_{i_k}.$$

Pour tout  $k \in \mathbb{N}_+$  il existe  $h_1^{(k)}, \dots, h_k^{(k)}$  tels que  $v[p_{n_k}] = \alpha_1 h_1^{(k)} \dots \alpha_k h_k^{(k)}$ .

Une simple application du lemme de Koenig permet d'établir qu'il existe alors  $h_1, \dots, h_k, \dots$  tels que

$$v[p_{n_k}] = \alpha_1 h_1 \dots \alpha_k h_k \quad \text{pour tout } k. \quad \square$$

On remarquera que  $u[\vec{Q}/\vec{\xi}]$  ne contient pas les mots finis qui pourraient provenir de la substitution telle qu'elle est traditionnellement définie.

Nous pouvons alors énoncer le :

THÉORÈME 1 : Soit  $G$  la grammaire algébrique  $\xi_i = P_i, i \in [N]$ .

Soit  $G$  la grammaire algébrique en les variables  $\Xi = \{\xi_1, \dots, \xi_N\}$  sur l'alphabet

terminal  $X \cup \Xi$  définie par

$$\bar{\xi}_i = \bar{P}_i, \quad i \in [N]$$

où

$$\bar{P}_i = \{ \alpha \bar{\xi}_j \mid \alpha \in (X \cup \Xi)^*, \xi_j \in \Xi, \alpha \xi_j (X \cup \Xi)^* \cap P_i \neq \emptyset \}.$$

Alors pour tout  $i \in [N]$ ,

$$L^\omega(G, \xi_i) = L^\omega(\bar{G}, \bar{\xi}_i) [\overrightarrow{L(\bar{G})/\bar{\xi}_i}].$$

*Démonstration* : On démontre d'abord l'inclusion  $\subset$ .

Considérons une dérivation infinie terminale  $\delta = \langle t_n \mid n \in \mathbf{N}_+ \rangle$  à partir de  $t_1 = \xi_i$  dans  $G$ . Supposons qu'elle engendre le mot  $w \in X^\omega$  où  $w = \lim_{n \rightarrow \infty} \alpha(t_n)$ ,  $\alpha(t_n)$  étant le plus long facteur gauche terminal de  $t_n$ .

Soit  $\sigma$  un arbre de dérivation de  $G$  contenant  $\delta$ . Soit  $b = \langle u_n \mid n \in \mathbf{N}_+ \rangle$  sa branche infinie la plus à gauche. Nous notons  $\sigma_n$  le sous-arbre initial minimal de  $\sigma$  tel que  $u_n \in A(\sigma_n)$  et  $\sigma' = \text{Sup} \{ \sigma_n \}$ .

La démonstration réside essentiellement dans la considération des arbres  $\bar{\sigma}_n$  ainsi construits :

si  $\tau$  est un sous-arbre initial de  $\sigma$ , nous définissons :

$$\begin{aligned} A(\bar{\tau}) &= \{ v \in A(\tau) \mid \exists n \in \mathbf{N}_+ : v \leq u_n \text{ ou } v <_g u_n \}, \\ \bar{\tau}(v) &= \tau(v) \quad \text{pour tout } v \in b, \\ \bar{\tau}(v) &= \tau(v) \quad \text{pour tout } v \notin b. \end{aligned}$$

Intuitivement pour obtenir  $\bar{\tau}$  on supprime toute la partie de  $\tau$  qui est strictement à droite de la branche  $b$  et l'on barre les étiquettes des sommets qui sont sur  $b$ .

Nous pouvons établir plusieurs propriétés :

LEMME 2 : Pour tout  $n \in \mathbf{N}_+$  :

$$fg(SG(\sigma, u_n)) = w_n \sigma(u_n) \quad \text{où } w_n \in (X \cup \Xi)^* \text{ et } \sigma(u_n) \in \Xi$$

et

$$fg(SG(\bar{\sigma}_n, u_n)) = \bar{w}_n \overline{\sigma(u_n)} \quad \text{où } \bar{w}_n \in (X \cup \Xi)^*.$$

Enfin  $w_n \in \bar{w}_n [\overrightarrow{L(\bar{G})/\bar{\xi}_i}]$ .

Dans cette écriture  $\hat{L}(G, \xi_i) = \{ f \in (X \cup \Xi)^* \mid \xi_i \xrightarrow{*} f \}$ .

*Démonstration* : Si  $n = 1$  il n'y a rien à démontrer.

Nous faisons une récurrence sur  $n$ .

Supposons  $u_{n+1} = u_n p$  pour  $p \in \mathbf{N}_+$ . Nous pouvons écrire :

$$\begin{aligned} A(SG(\sigma, u_{n+1})) &= \{v \in A(\sigma) \mid v \leq u_{n+1} \text{ ou } v <_g u_{n+1}\} \\ &= A(SG(\sigma, u_n)) \cup \{v \in A(\sigma) \mid u_n < v \text{ et } v <_g u_{n+1}\} \cup \{u_{n+1}\} \\ &= A(SG(\sigma, u_n)) \cup A(\sigma : u_n 1) \cup \dots \cup A(\sigma : u_n (p-1)) \cup \{u_{n+1}\}. \end{aligned}$$

En remarquant que les termes de cette union sont disjoints et que les éléments de chacun d'eux précèdent dans l'ordre lexicographique les éléments du terme suivant, nous déduisons

$$fg(SG(\sigma, u_{n+1})) = w_{n+1} \sigma(u_{n+1})$$

et

$$w_{n+1} = w_n fg(\sigma : u_n 1) \dots fg(\sigma : u_n (p-1)).$$

De façon similaire, en se reportant à la définition de  $\sigma_n$  et  $\bar{\sigma}_n$  :

$$fg(SG(\bar{\sigma}_{n+1}, u_{n+1})) = \bar{w}_{n+1} \overline{\sigma(u_{n+1})}$$

et

$$\bar{w}_{n+1} = \bar{w}_n \sigma(u_n 1) \dots \sigma(u_n (p-1)).$$

Les deux suites  $w_n$  et  $\bar{w}_n$  sont ainsi des suites croissantes pour l'ordre  $\leq$ . La propriété de  $b$  d'être la branche infinie la plus à gauche entraîne que pour  $i=1, \dots, p-1$   $A(\sigma : u_n i)$  est fini puisque  $u_n i <_g u_{n+1}$ . D'où  $fg(\sigma : u_n i) \in \hat{L}(G, \sigma(u_n i))$  puisque  $fg(\sigma : u_n i)$  est un mot de  $(X \cup \Xi)^*$ .

Par récurrence nous obtenons bien  $w_n \in \bar{w}_n [\overrightarrow{L(G)}/\vec{\xi}]$ .  $\square$

Une autre remarque essentielle est :

LEMME 3 : Si la suite  $\bar{w}_n$  est stationnaire il en est de même de la suite  $w_n$ .

Démonstration : En effet la suite  $\bar{w}_n$  est stationnaire si et seulement si à partir d'un certain  $n_0$ ,  $u_{n+1} = u_n 1$  pour tout  $n \geq n_0$ .

L'identité ci-dessus montre qu'alors  $w_n$  est aussi stationnaire.  $\square$

La démonstration s'achève par une suite de remarques.

La suite  $w_n$  n'est pas stationnaire : en effet la dérivation  $\delta$  est terminale ce qui implique que pour tout  $n$ .

$$w[n] \leq fg(\tau_n) \text{ pour quelque sous-arbre initial } \tau_n \text{ de } \sigma.$$

Or si  $u_p$  est le sommet où la branche infinie  $b$  rencontre la frontière de  $\tau_n$  nous avons

$$\alpha(fg(\tau_n)) \leq fg(\sigma, u_p).$$

Il s'ensuit que les mots  $w_n$  contiennent des facteurs gauches terminaux de longueur aussi grande que l'on veut. Comme ils forment une suite croissante

pour  $\leq$  cela implique qu'ils sont tous dans  $X^*$ .

On peut alors remplacer la relation

$$w_n \in \bar{w}_n [\bar{L}(\bar{G})/\bar{\xi}]$$

pour la relation

$$w_n \in \bar{w}_n [\bar{L}(\bar{G})/\bar{\xi}].$$

De fait que cette relation est satisfaite pour tout  $n$  et que la suite des  $\bar{w}_n$  n'est pas stationnaire on déduit que :

1° la suite des  $\bar{w}_n$  a une limite quand  $n \rightarrow \infty$ , notons la  $\bar{w}$ , et  $\bar{w} \in (X \cup \Xi)^\circ$ ;

2°  $w \in \bar{w} [L(\bar{G})/\bar{\xi}]$ .

Or le mot  $\bar{w}$  est engendré par une dérivation infinie terminale dans  $\bar{G}$ , précisément la dérivation

$$\langle fg(SG(\bar{\sigma}_n, u_n)), n \in \mathbf{N}_+ \rangle.$$

Que cette suite soit une dérivation est immédiat par la relation établie plus haut :

$$fg(SG(\bar{\sigma}_{n+1}, u_{n+1})) = \bar{w}_n \sigma(u_n 1) \dots \sigma(u_n (p-1)) \bar{\sigma}(u_n p),$$

où

$$fg(SG(\bar{\sigma}_n, u_n)) = \bar{w}_n \overline{\sigma(u_n)}.$$

Qu'elle soit terminale et engendre  $\bar{w} = \lim_{n \rightarrow \infty} \bar{w}_n$  ne l'est pas moins par la remarque que le plus long facteur de  $fg(SG(\bar{\sigma}_n, u_n))$  qui soit dans  $(X \cup \Xi)^*$  est précisément  $\bar{w}_n$  et le fait que la suite  $w_n$  n'est pas stationnaire.

Nous avons établi la première inclusion.  $\square$

Établissons l'inclusion inverse :

– soit  $w \in L^\circ(\bar{G}, \bar{\xi}_i) [L(\bar{G})/\bar{\xi}]$ ;

– soit  $\bar{w} \in L^\circ(\bar{G}, \bar{\xi}_i)$  un mot tel que  $w \in \bar{w} [L(\bar{G})/\bar{\xi}]$ .

Nous pouvons écrire

$$\bar{w} = \alpha_0 \xi_{i_1} \alpha_1 \xi_{i_2} \dots \alpha_{n-1} \xi_{i_n} \alpha_n \dots$$

et

$$w = \alpha_0 h_1 \alpha_1 h_2 \dots \alpha_{n-1} h_n \alpha_n \dots \quad \text{avec } h_n \in L(\bar{G}, \xi_{i_n}).$$

Considérons une dérivation infinie terminale de  $\bar{\xi}_i$  en  $\bar{w}$  dans  $\bar{G}$ , soit  $t_1 = \bar{\xi}_1$ ,  $t_2, \dots, t_p, \dots$ . La grammaire  $\bar{G}$  étant linéaire à droite on peut écrire pour tout  $p \in \mathbf{N}_+$  :

$$t_p = u'_p \bar{\xi}_i, \quad \text{avec } u'_p \in (X \cup \Xi)^* \quad \text{et } u'_p \leq \bar{w}.$$

Appelons  $u_p$  le facteur gauche de  $w$  obtenu en substituant aux occurrences de non terminaux  $\xi_{i_1}, \dots, \xi_{i_k}$  dans  $u'_p$  les mots  $h_1, \dots, h_k$  correspondants.

Supposons connue une dérivation gauche  $\delta_p$  de  $\xi_i$  en un mot  $u_p \xi_{i_p} v_p$  : nous construisons  $\delta_{p+1}$  prolongeant  $\delta_p$ , dérivation gauche de  $\xi_i$  en  $u_{p+1} \xi_{i_{p+1}} v_{p+1}$ .

Nous avons puisque  $\langle t_n \rangle$  est une dérivation de  $\bar{G}$ ,

$$u'_{p+1} = u'_p m'_{p+1} \quad \text{avec} \quad m'_{p+1} \bar{\xi}_{i_{p+1}} \in \bar{p}_{i_p}.$$

Il existe ainsi  $n'_{p+1}$  tel que  $m'_{p+1} \bar{\xi}_{i_{p+1}} n'_{p+1} \in P_{i_p}$  d'où

$$u_p \xi_{i_p} v_p \xrightarrow{G} u_p m'_{p+1} \bar{\xi}_{i_{p+1}} n'_{p+1} v_p.$$

Or  $u_{p+1} = u_p m_{p+1}$  où  $m_{p+1}$  est obtenu en substituant aux occurrences de non-terminaux  $\xi_{i_{k+1}}, \dots, \xi_{i_{p+1}}$  dans  $m'_{p+1}$  les  $h_{i_{k+1}}, \dots, h_{i_{p+1}}$  correspondants.

Il existe une dérivation gauche de  $\xi_{i_n}$  en  $h_n$  dans  $G$  pour tout  $n$  donc aussi une dérivation gauche de  $m'_{p+1}$  en  $m_{p+1}$ , donc enfin une de  $u_p \xi_{i_p} v_p$  en  $u_{p+1} \xi_{i_{p+1}} n'_{p+1} v_p$ .

La dérivation  $\delta_{p+1}$  est la dérivation  $\delta_p$  prolongée de cette dérivation gauche de  $u_p \xi_{i_p} v_p$  en  $u_{p+1} \xi_{i_{p+1}} v_{p+1}$ .

La suite croissante de dérivations  $\delta_1, \delta_2, \dots, \delta_p, \dots$  a une limite  $\delta$  qui est une dérivation infinie gauche et terminale de  $\xi_i$  en  $w$  dans  $G$ . Ce qui établit le théorème.  $\square$

Nous avons aussi établi la :

PROPRIÉTÉ 10 : Pour toute grammaire algébrique  $G$ , tout non terminal  $\xi_i$  et tout mot infini  $w \in L^\omega(G, \xi_i)$  il existe une dérivation infinie terminale gauche de  $\xi_i$  en  $w$  dans  $G$ .

### III. ADHÉRENCES DE LANGAGES ALGÈBRIQUES

Si  $A$  est une partie de  $X^*$  on définit  $A^\omega$  de la façon suivante :

—  $A^\omega \subset X^\omega$  et pour tout  $u \in X^\omega$  on a  $u \in A^\omega \Leftrightarrow$  il existe une suite croissante d'entiers  $p_1, p_2, \dots, p_n, \dots$  telle que  $p_n \xrightarrow[n \rightarrow \infty]{} \infty$  et  $u[p_n] \in A^n$ .

On écrit un peu improprement  $A^\omega = \{a_1 a_2 \dots a_n \dots \mid a_n \in A\}$  en sous-entendant que  $a_n$  est non vide pour une infinité de  $n$ .

Le théorème de substitution a de nombreuses conséquences.

La première est le :

THÉORÈME 2 : *Tout  $\omega$ -langage algébrique,  $L = L^\omega(G, \xi_i)$  peut se mettre sous la forme*

$$L = \bigcup_{k=1}^p L_k (L'_k)^\omega,$$

où les  $L_k$  et  $L'_k$  sont des langages algébriques.

La démonstration repose essentiellement sur le :

LEMME 4 : *Si  $G$  est une grammaire algébrique linéaire à droite satisfaisant pour tout  $i \in [N]$ ,  $P_i \subset X^* \Xi$  alors pour tout  $i \in [N]$ ,  $L^\omega(G, \xi_i)$  est contenu dans  $X^\omega$  et se met sous la forme*

$$L^\omega(G, \xi_i) = \bigcup_{k=1}^p R_k (R'_k)^\omega.$$

*Démonstration* : Celle que nous donnons n'est qu'une esquisse de preuve, une démonstration rigoureuse se trouve dans l'ouvrage d'Eilenberg [2]. Le résultat est originellement dû à MacNaughton.

Considérons d'abord la simple grammaire

$$\xi = A \xi.$$

Il est clair qu'elle engendre  $A^\omega = \{ a_1 a_2 \dots a_n \dots \mid a_n \in A \}$ .

Considérons la grammaire

$$G \begin{cases} \xi_1 = A_{11} \xi_1 + A_{12} \xi_2, \\ \xi_2 = A_{21} \xi_1 + A_{22} \xi_2. \end{cases}$$

Encore clairement on a

$$L^\omega(G, \xi_1) = (A_{11})^* A_{12} L(G, \xi_2) + (A_{11})^\omega,$$

en séparant dans l'ensemble des dérivations infinies à partir de  $\xi_1$  celles qui ne contiennent jamais  $\xi_2$  et les autres.

On en tire

$$L(G, \xi_2) = A_{21} A_{11}^* A_{12} L(G, \xi_2) + A_{22} L(G, \xi_2) + A_{21} (A_{11})^\omega$$

qui se résout en

$$L(G, \xi_2) = (A_{22} A_{11}^* A_{12} + A_{22})^* A_{21} (A_{11})^\omega + (A_{21} A_{11}^* A_{12} + A_{22})^\omega,$$

formule que l'on peut aussi établir *a priori* en séparant les dérivations infinies qui contiennent  $\xi_2$  une infinité de fois et les autres.

Ce raisonnement peut être appliqué, par récurrence, à toute grammaire linéaire à droite, c'est-à-dire de la forme

$$\xi_i = \sum_{j=1}^N A_{ij} \xi_j, \quad i \in [N]$$

et conduit au résultat cherché.  $\square$

Le théorème 2 découle immédiatement du théorème 1 et de la propriété puisque

$$L^\omega(G, \xi_i) = L^\omega(\bar{G}, \bar{\xi}_i) [\overrightarrow{L(\bar{G})/\bar{\xi}}],$$

où  $\bar{G}$  est une grammaire linéaire à droite.

Ainsi

$$L^\omega(\bar{G}, \bar{\xi}_i) = \bigcup_{k=1}^p R_k (R'_k)^\omega$$

et

$$L^\omega(G, \xi_i) = \left( \bigcup_{k=1}^p R_k (R'_k)^\omega \right) [\overrightarrow{L(\bar{G})/\bar{\xi}}] = \bigcup_{k=1}^p R_k [L(\bar{G})/\bar{\xi}] (R'_k [\overrightarrow{L(\bar{G})/\bar{\xi}}])^\omega.$$

La substitution algébrique  $[L(\bar{G})/\bar{\xi}]$  applique les rationnels  $R_k$  et  $R'_k$  sur des langages algébriques  $L_k$  et  $L'_k$ .  $\square$

*Exemple :* On prend  $G : \xi = a \xi \xi + b$ , grammaire qui engendre comme langage de mots finis le langage de Lukasiewicz :

$$\mathcal{L} = \{ f \in \{a, b\}^* \mid |f|_a = |f|_b - 1 \text{ et } f' < f \Rightarrow |f'|_a \geq |f'|_b \}.$$

Le langage de mots infinis engendré par  $G$  s'obtient en considérant

$$\bar{G} : \bar{\xi} = a \bar{\xi} + a \xi \bar{\xi},$$

qui engendre  $L^\omega(\bar{G}, \bar{\xi}) = (a + a \xi)^\omega$ .

D'où

$$\begin{aligned} L^\omega(G, \xi) &= (a + a \xi)^\omega \mathcal{L} / \xi \\ &= (a/a\mathcal{L})^\omega. \end{aligned}$$

On montrera plus loin que ce langage est aussi

$$\{ u \in X^\omega \mid u' < u \Rightarrow |u'|_a \geq |u'|_b \}.$$

REMARQUE : Avec les définitions qui sont les nôtres la réciproque du lemme 4 n'est pas vérifiée. Ainsi  $a^* b^\omega$  n'est engendré par aucune grammaire linéaire à droite telle que  $P_i \subset X^* \Xi$  pour tout  $i \in [N]$ .



Supposons en effet qu'il en soit ainsi et considérons une dérivation infinie de  $\xi_i$  en  $a^n b^\omega$ , soit  $t_1 = \xi_i, t_2, \dots, t_n \dots$ . On peut écrire pour tout  $n : t_n = u_n \xi_{i_n}$  et pour tout  $n < p$  on a  $u_n \leq u_p$  et si  $u_p = u_n v, \xi_{i_n} \xrightarrow{*} v \xi_{i_p}$ . Si l'on prend  $n$  assez grand on trouvera certainement  $n$  et  $p$  tels que  $n < p, u_n < u_p \leq a^n$  et  $\xi_{i_n} = \xi_{i_p}$ . D'où une dérivation  $\xi_{i_n} \xrightarrow{*} a^l \xi_{i_n}$ . Mais si une telle dérivation existe, il existe une dérivation infinie de  $\xi_i$  en  $a^\omega$  dans  $G$ .  $\square$

Nous définissons maintenant l'adhérence d'un langage.

DÉFINITION 1. — Pour toute partie  $L \subset X^\omega$  nous appelons adhérence de  $L$  et notons  $\text{Adh}(L)$  l'ensemble des mots infinis  $u$  de  $X^\omega$  tels que tous leurs facteurs gauches sont facteurs gauches d'un mot de  $L$ .

DÉFINITION 2. — La partie  $L$  de  $X^\omega$  est dite amène si et seulement si le sous-ensemble  $L^{\text{inf}} = L \cap X^\omega$  des mots infinis de  $L$  est l'adhérence de l'ensemble  $L^{\text{fin}} = L \cap X^*$  des mots finis de  $L$ .

REMARQUE : La clôture  $\bar{A}$  d'un ensemble  $A$  de  $X^*$  étant définie comme dans l'ouvrage d'Eilenberg [2] par

$$\bar{A} = \{ u \in X^\omega \mid \text{card} \{ n \mid u[n] \in A \} = \infty \},$$

on a immédiatement pour  $L \subset X^*$

$$\text{Adh}(L) = \overline{FG(L)}$$

On remarquera que  $\overline{A^*} = \{ u \in X^\omega \mid \text{card} \{ n \mid u[n] \in A^* \} = \infty \}$  n'est pas égal à  $A^\omega$  comme le montre l'exemple suivant :

Exemple :

$$A = ab^*, \quad A^\omega = \{ ab^{n_1} ab^{n_2} \dots ab^{n_p} \dots \mid n_1, n_2, \dots, n_p, \dots \in N \}$$

et

$$\overline{A^*} = A^\omega \cup \{ ab^{n_1} ab^{n_2} \dots ab^{n_p} ab^\omega \mid n_1, \dots, n_p \in N \}.$$

D'une manière générale on a

$$\overline{A^*} = A^* \bar{A} \cup A^\omega.$$

Nous pouvons établir tout de suite la :

PROPRIÉTÉ 11 : Soit  $L \subset X^*$  et  $\mathring{L}$  le langage

$$\mathring{L} = \{ u \in X^* \mid \text{card} \{ v \in X^* \mid uv \in L \} = \infty \}.$$

Alors :

$$\mathring{L} = FG(\text{Adh}(L)) \text{ et } \text{Adh}(\mathring{L}) = \text{Adh}(L).$$

Démonstration : Soit  $u \in \text{Adh}(L)$  et  $u' = u[n]$  un de ses facteurs gauches. Pour tout  $p \in \mathbf{N}_+$   $u' \leq u[n+p]$ , il existe  $v_{n+p}$  tel que  $u[n+p]v_{n+p} \in L$ .

D'où  $u(n+1) \dots u(n+p)v_{n+p} \in \{v' \mid u'v' \in L\}$  et ce dernier ensemble est infini.

Réciproquement si  $\text{card}\{v \mid uv \in L\} = \infty$  on peut construire un mot  $\bar{w} \in \text{Adh}(L)$  tel que  $u < \bar{w}$ . L'alphabet  $X$  étant comme toujours supposé fini  $\{v \mid uv \in L\} \cap X^*$  est infini pour un moins un  $x \in X$  soit  $x_{i_1}$ . Cela s'écrit  $\text{card}\{v \mid ux_{i_1}v \in L\} = \infty$  et permet de trouver  $x_{i_2}$  tel que  $\{v \mid ux_{i_1}x_{i_2}v \in L\} \cap X^*$  soit infini ou encore  $\text{card}\{v \mid ux_{i_1}x_{i_2}v \in L\} = \infty$ .

On peut ainsi construire une suite croissante  $u, ux_{i_1}, ux_{i_1}x_{i_2}, \dots$  dont la limite est un mot infini  $w$  de l'adhérence de  $L$  qui a  $u$  comme facteur gauche.

L'égalité  $\text{Adh}(L) = \text{Adh}(\overset{\circ}{L})$  se vérifie immédiatement.  $\square$

Les facteurs gauches  $u$  de mots de  $L$  qui ont la propriété que  $\text{card}\{v \mid uv \in L\} = \infty$  sont dits infiniment complétables. Le langage  $\overset{\circ}{L}$  sera dit intérieur de  $L$ .

Dans une étude des langages de mots infinis qui sont adhérence d'un langage algébrique, on est amené à utiliser constamment l'intérieur d'un langage algébrique.

Nous allons établir le :

**THÉORÈME 4 :** *L'intérieur de tout langage algébrique est un langage algébrique.*

Mais auparavant nous devons établir une importante propriété des langages de mots infinis engendrés par des grammaires de Greibach.

**THÉORÈME 5 :** *Si la grammaire algébrique  $G : \xi_i = P_i, i \in [N]$  est de Greibach et réduite alors pour tout  $i \in [N]$ , le langage  $L^\infty(G, \xi_i)$  est amène c'est-à-dire*

$$L^\infty(G, \xi_i) = \text{Adh}(L(G, \xi_i)).$$

*Démonstration :* Rappelons que  $G$  est de Greibach ssi pour tout  $i \in [N]$ ,  $P_i \subset X(X \cup \Xi)^*$  et que  $G$  est réduite ssi pour tout  $i$ ,  $L(G, \xi_i) \neq \emptyset$ .

Disons que la dérivation gauche  $\delta = \langle t_n \mid n \in [k] \rangle :$

- dépasse le mot  $g$  de  $X^*$  ssi  $g \leq \alpha(t_k)$ ;
- dépasse juste  $g$  ssi  $\alpha(t_{k-1}) < g \leq \alpha(t_k)$ .

Il est clair que si la dérivation  $\delta = \langle t_n \mid n \in [k] \rangle$  dépasse  $g$ , il existe  $k' \leq k$  tel que  $\delta' = \langle t_n \mid n \in [k'] \rangle$  dépasse juste  $g$ .

Si la grammaire  $G$  est de Greibach, l'ensemble des dérivations gauches qui dépassent juste  $g$  est fini pour tout  $g$ . En effet, si  $\delta = \langle t_n \mid n \in [k] \rangle$  est gauche et dépasse juste  $g$  on a  $\alpha(t_{k-1}) < g$  donc  $|\alpha(t_{k-1})| < |g|$ . Or  $\delta$  étant gauche et  $G$  de Greibach  $|\alpha(t_{k-1})| > k-1$  d'où  $k \leq |g|$ .

Montrons alors que  $L^\omega(G, \xi_i) \subset \text{Adh}(L(G, \xi_i))$  si  $G$  est réduite : il suffit de montrer que si  $\delta = \langle t_n \mid n \in \mathbf{N}_+ \rangle$  est une dérivation infinie terminale d'origine  $\xi_i$  et  $u \leq \alpha(t_n)$  il existe  $v$  tel que  $uv \in L(G, \xi_i)$ . Or  $G$  étant réduite pour tout mot  $t \in (X \cup \Xi)^*$ ,  $L(G, t) \neq \emptyset$  donc il existe  $w \in L(G, t_n)$ . On a bien  $w \in L(G, \xi_i)$  et  $u \leq w$ . L'inclusion inverse  $\text{Adh}(L(G, \xi_i)) \subset L^\omega(G, \xi_i)$  se démontre en utilisant le lemme de Koenig.

Soit  $u \in \text{Adh}(L(G, \xi_i))$  et pour tout  $n$ ,  $\Delta_n$  l'ensemble des dérivations gauches qui dépassent juste  $u[n]$ . Nous avons vu que  $\Delta_n$  est fini et que pour tout  $\delta \in \Delta_{n+1}$  il existe  $\delta'$  dans  $\Delta_n$  tel que  $\delta$  prolonge  $\delta'$ .  $\Delta_n$  est aussi non vide puisque si  $v_n$  est tel que  $u[n]v_n \in L(G, \xi_i)$  il existe une dérivation gauche de  $\xi_i$  en  $u[n]v_n$  qui dépasse  $u[n]$ . Appliquant le lemme de Koenig on construit alors une suite infinie de dérivations  $\delta_n$  telles que  $\delta_n \in \Delta_n$  et  $\delta_{n+1}$  prolonge  $\delta_n$  pour tout  $n$ . La limite  $\delta = \lim_{n \rightarrow \infty} \delta_n$  est une dérivation infinie gauche terminale de  $\xi_i$  en  $u$  dans  $G$ .  $\square$

Le théorème 5 admet une réciproque :

**THÉORÈME 6 :** *Pour tout  $\omega$ -langage algébrique amène  $L$  il existe une grammaire de Greibach réduite  $G'$  telle que*

$$L^\omega(G, \xi_i) = L \setminus \{ \varepsilon \}.$$

*Démonstration :* Supposons  $L = L^\omega(G, \xi_i)$  pour quelque grammaire algébrique  $G$  et  $L$  amène qui se traduit par

$$L^\omega(G, \xi_i) = \text{Adh}(L(G, \xi_i)).$$

Il est bien connu qu'on peut construire une grammaire de Greibach réduite  $G'$  telle que

$$L(G', \xi_i) = L(G, \xi_i) \setminus \{ \varepsilon \}.$$

Par le théorème 5 on a alors :

$$L^\omega(G', \xi_i) = \text{Adh}(L(G', \xi_i)),$$

donc aussi

$$L^\omega(G', \xi_i) = \text{Adh}(L(G, \xi_i)),$$

puisque pour tout  $L$ ,  $\text{Adh}(L) = \text{Adh}(L \setminus \{ \varepsilon \})$  et  $L^\omega(G', \xi_i) = L^\omega(G, \xi_i)$ .

D'où l'égalité  $L^\omega(G', \xi_i) = L \setminus \{ \varepsilon \}$ .  $\square$

Revenons alors à la démonstration du théorème 4.

Soit  $L$  un langage algébrique. Il existe une grammaire  $G$  de Greibach réduite telle que  $L(G, \xi_i) = L \setminus \{ \varepsilon \}$ .

Il est même loisible de supposer  $G$  en forme normale c'est-à-dire satisfaisant la condition pour tout  $i \in [N]$ ,  $P_i \subset X \Xi^*$ .

Par le théorème 5,  $\text{Adh}(L) = \text{Adh}(L \setminus \{\varepsilon\}) = L^\circ(G, \xi_i)$  et par le théorème 3  $L(G, \xi_i) = L^\circ(\bar{G}, \xi_i) [\overrightarrow{L(\bar{G})} \setminus \bar{\xi}]$  où  $\bar{G}$  est la grammaire construite plus haut.

Ainsi  $\dot{L} = FG(L^\circ(\bar{G}, \xi_i) [L(\bar{G}) \setminus \bar{\xi}])$ .

Cherchons donc à quelles conditions un mot  $u$  de  $X^*$  est facteur gauche d'un mot de  $L^\circ(G, \xi_i) [\overrightarrow{L(\bar{G})} \setminus \bar{\xi}]$ . Il est clair que si  $w = \alpha_1 h_1 \alpha_2 h_2 \dots \alpha_n h_n \dots$  est un mot de  $L^\circ(G, \xi_i) [\overrightarrow{L(\bar{G})} \setminus \bar{\xi}]$ , où  $\bar{w} = \alpha_1 \xi_{i_1} \alpha_2 \xi_{i_2} \dots \alpha_n \xi_{i_n} \dots \in L^\circ(\bar{G}, \xi_i)$  et pour tout  $n$ ,  $h_n \in L(G, \xi_{i_n})$  on a  $u \in FG(w)$  si et seulement si il existe  $k$  tel que soit

$$u = \alpha_1 h_1 \dots \alpha_{k-1} h_{k-1} \alpha_k h'_k \quad \text{avec} \quad h'_k \leq h_k.$$

soit

$$u = \alpha_1 h_1 \dots \alpha_{k-1} h_{k-1} \alpha'_k \quad \text{avec} \quad \alpha'_k \leq \alpha_k.$$

Ceci s'écrit encore  $u \in FG(w) \Leftrightarrow$  il existe  $\bar{u} \in FG(\bar{w})$  tel que soit  $\bar{u} \xi_{i_k} \leq \bar{w}$  et  $u = u_1 h'_k$  avec  $u_1 \in \bar{u} [\overrightarrow{L(\bar{G})} \setminus \bar{\xi}]$  et  $h'_k \in FG(L(G, \xi_{i_k}))$  soit  $u \in \bar{u} [\overrightarrow{L(\bar{G})} \setminus \bar{\xi}]$ .

Pour construire la grammaire  $\bar{G}$  on se munit d'un alphabet  $\Theta = \{\theta_1, \dots, \theta_N\}$  et l'on écrit les équations

$$\bar{\xi}_i = \bar{P}_i + \Sigma \{ \alpha \theta_j \mid \alpha \xi_j \in \bar{P}_i \} + \Sigma \{ \alpha \theta_j \mid \exists \beta, \alpha \xi_j \beta \in \bar{P}_i \} + \Sigma \{ x \in X \mid x \Xi^* \Xi \cap \bar{P}_i \neq \emptyset \}.$$

On a la propriété suivante :

$$L(\bar{G}, \bar{\xi}_i) = \{ \bar{u} \theta_j \mid \bar{u} \xi_j \text{ est facteur gauche d'un mot de } L^\circ(\bar{G}, \bar{\xi}_i) \} \\ \cup \{ \bar{u} \mid \bar{u} \in (X \cup \Xi)^* X \text{ et } \bar{u} \text{ est facteur gauche d'un mot de } L^\circ(\bar{G}, \bar{\xi}_i) \}.$$

Cette propriété n'est pas difficile à démontrer du fait que  $\bar{G}$  est comme  $\bar{G}$  linéaire à droite et que par suite une seule règle terminale peut être utilisée dans une dérivation et ce tout à la fin.

Le langage  $\dot{L}$  est alors égal à la substitution dans  $L(\bar{G}, \bar{\xi}_i)$  de  $L(G, \xi_i)$  à  $\xi_i$  et  $FG(L(G, \xi_i)) \setminus \{\varepsilon\}$  à  $\theta_i$  pour tout  $i \in [N]$ . C'est-à-dire que  $L = L(\bar{G}, \bar{\xi}_i)$  où l'on obtient  $\bar{G}$  en rajoutant à  $\bar{G}$  toutes les équations de  $G$  et pour chaque  $\theta_i$  les équations d'une grammaire engendrant  $FG(L(G, \xi_i)) \setminus \{\varepsilon\}$  (grammaire qui peut donc être prise de Greibach).

Exemple :

$$G : \xi = a \xi \xi + b$$

est une grammaire qui engendre le langage de Lukasiewicz :

$$\dot{L} = \{ f \{ a, b \}^* \mid |f|_a = |f|_b - 1 \text{ et } \forall f' < f : |f'|_a \geq |f'|_b \}.$$

La grammaire  $\bar{G}$  est la grammaire  $\bar{G} : \bar{\xi} = a\bar{\xi} + a\xi\bar{\xi}$  qui engendre  $L(\bar{G}, \bar{\xi}) = (a + a\xi)^\omega$ .

Ainsi  $\text{Adh}(L) = (a + a\xi)^\omega [\mathcal{L}/\xi] = (a + a\mathcal{L})^\omega$  et d'autre part

$$\text{Adh}(\mathcal{L}) = \{ u \in X^\omega \mid \forall f < u, |f|_a \geq |f|_b \}$$

par la définition de l'adhérence.

Pour construire une grammaire engendrant  $\overset{\circ}{L}$  on commence par construire  $\bar{\bar{G}}$  :

$$\bar{\bar{G}} : \bar{\bar{\xi}} = a\bar{\bar{\xi}} + a\xi\bar{\bar{\xi}} + a\theta + a\xi\theta + a$$

qui engendre  $(a + a\xi)^*(a + a\theta + a\xi\theta)$ .

Il nous faut encore une grammaire pour  $FG(\mathcal{L}) \setminus \{ \varepsilon \}$ .

On peut prendre  $\theta = a\xi\theta + a\theta + a + b + a\xi$ .

Et finalement l'on obtient  $\bar{\bar{\bar{G}}}$  :

$$\bar{\bar{\bar{\xi}}} = a\bar{\bar{\bar{\xi}}} + b,$$

$$\bar{\bar{\bar{\xi}}} = a\bar{\bar{\bar{\xi}}} + a\xi\bar{\bar{\bar{\xi}}} + a\theta + a\xi\theta + a,$$

$$\bar{\bar{\bar{\theta}}} = a\xi\bar{\bar{\bar{\theta}}} + a\bar{\bar{\bar{\theta}}} + a + b + a\xi.$$

On vérifie que  $L(\bar{\bar{\bar{G}}}, \bar{\bar{\bar{\xi}}})$  n'est autre que l'ensemble des facteurs gauches propres et non vides de mots de  $\mathcal{L}$ , comme on s'en aperçoit facilement.

#### REMERCIEMENTS

Nous tenons à remercier ici L. Boasson avec lequel nous avons eu d'innombrables et fructueuses discussions sur le sujet, B. Courcelle qui nous a fourni une première version du théorème 6 et attiré notre attention sur la notion d'« aménité », enfin A. Arnold qui a bien voulu lire ce texte avec soin et nous a fourni d'innombrables remarques de détail qui ont permis de l'améliorer grandement.

#### BIBLIOGRAPHIE

1. R. COHEN et A. GOLD, *Theory of  $\omega$ -languages, Part I et II*, J. Comp. Syst. Sc., vol. 15, 1977, p. 169-208.
2. S. EILENBERG, *Automata, Languages and Machines*, Vol. A, Academic Press, 1974.
3. M. LINNA, *On  $\omega$ -words and  $\omega$ -computations*, Annales Universitatis Turkuensis, série A, 1975, p.
4. M. NIVAT, *Mots infinis engendrés par une grammaire algébrique*, R.A.I.R.O. Inform. Theor., vol. 11, 1977, p. 311-327.