

CHRISTIANE FROUGNY

Langages très simples générateurs

RAIRO. Informatique théorique, tome 13, n° 1 (1979), p. 69-86

<http://www.numdam.org/item?id=ITA_1979__13_1_69_0>

© AFCET, 1979, 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/>

LANGAGES TRÈS SIMPLES GÉNÉRATEURS (*)

par Christiane FROUGNY ⁽¹⁾

Communiqué par J. BERSTEL

Résumé. — En [9], S. Greibach a énoncé la conjecture suivante :

Un langage algébrique qui ne contient pas de sous-ensemble rationnel infini est soit quasi rationnel, soit générateur de la famille des langages algébriques.

Cette conjecture vient d'être infirmée en [2]. Elle reste cependant vérifiée dans un cas particulier, que nous prouvons ici, à savoir :

Un langage très simple qui ne contient pas de sous-ensemble rationnel infini est soit quasi rationnel, soit générateur de la famille des langages algébriques.

1. INTRODUCTION

Les langages algébriques (context-free) sont de bons modèles des langages de programmation. Pour cette raison, la famille des langages algébriques a été étudiée de manière approfondie. L'étude se poursuit actuellement, notamment par l'utilisation de la notion de transduction rationnelle pour comparer les complexités de structure de deux langages algébriques. On est ainsi amené à étudier des sous-familles des langages algébriques fermées par diverses opérations naturelles, comme les cônes rationnels qui sont fermés par transduction rationnelle, et les « full AFL » qui sont en plus fermées par union, produit et étoile.

La famille des langages algébriques, que nous notons Alg est un « full AFL » [6], et est de plus principal, à la fois en tant que cône [5] et en tant que « full AFL » [7], c'est-à-dire qu'il existe des langages algébriques L tels que le cône resp. le « full AFL » engendré par L soit Alg tout entier [3]. Ces langages, appelés générateurs, sont en quelque sorte les langages algébriques les plus « puissants » puisque tout autre langage algébrique peut en être déduit par des opérations simples. L'étude de la famille des générateurs, ou de manière équivalente de la famille Nge des langages algébriques qui ne sont pas générateurs, constitue l'un des problèmes les plus étudiés de la théorie, et de nombreuses questions restent encore sans réponse. Ainsi, l'on sait que Nge est un « full AFL » fermé par substitution [8], mais on ignore s'il est principal.

(*) Reçu novembre 1977, et dans sa version révisée octobre 1978.

(1) Université de Picardie, Amiens et L.I.T.P. Paris, L.A. 248 au C.N.R.S.

La famille Nge des non-générateurs contient la famille Lin des langages linéaires et également la famille Qrt des langages quasi-rationnels, qui est égale à la famille des langages non expansifs, et à celle des « derivation bounded languages ». D'autre part, Nge contient également la famille Ict des langages à compteur itéré.

S. Greibach [8] a formulé la conjecture que Nge est égale à la famille Gre définie comme la fermeture par substitution de $Ict \cup Qrt$. Si la conjecture avait été vraie, cela aurait impliqué que Nge est non principal. Or un contre-exemple à cette conjecture a été donné par Boasson [1]. La famille Gre a néanmoins des propriétés intéressantes, surtout si on la restreint à la classe des langages L définie par la condition suivante, appelée condition IRS : « Tout langage rationnel contenu dans L est fini ». Appelons Irs cette classe de langages. On obtient alors des résultats remarquables; ainsi S. Greibach [9] a montré que les familles Gre et Qrt coïncident sur les langages vérifiant la condition IRS, autrement dit : $Irs \cap Gre = Irs \cap Qrt$. Ainsi un langage L appartenant à Irs ou bien n'est pas dans la famille Gre, ou bien est quasi rationnel. S. Greibach [9] a formulé la conjecture plus forte, selon laquelle les familles Gre et Nge coïncident sur Irs, i. e. : $Irs \cap Gre = Irs \cap Nge$. En vue du résultat rappelé ci-dessus, la conjecture s'énonce aussi : un langage vérifiant la condition IRS est ou bien générateur, ou bien quasi rationnel. Un contre-exemple à cette conjecture a été donné récemment par L. Boasson [2]. Néanmoins, il existe des cas particuliers où la conjecture a été vérifiée. Ainsi un langage (multi-) parenthétique (qui est toujours dans Irs) est ou bien générateur, ou bien quasi rationnel [4].

Nous nous proposons de vérifier la conjecture dans un autre cas particulier, indépendant du premier, à savoir celui des langages très simples. Nous prouvons :

THÉORÈME : *Un langage très simple qui ne contient pas de sous-ensemble rationnel infini est ou bien quasi-rationnel, ou bien générateur de Alg.*

Pour prouver ce résultat, nous utilisons le fait que le langage E engendré par la grammaire $\{S \rightarrow aSbSc, S \rightarrow d\}$ est un générateur de la famille Alg [11]. Remarquons que E est très simple.

2. DÉFINITIONS, NOTATIONS

Une *grammaire algébrique* est un quadruplet $G=(X, V, P, A)$ où X est l'alphabet terminal, V l'alphabet non-terminal, A un élément de V appelé axiome, P un sous-ensemble de $V \times (X \cup V)^*$, l'ensemble des règles de G .

Si $f \in (X \cup V)^*$, on note

$$L(G, f) = \{g \in X^* / f \xrightarrow{*} g\} \quad \text{et} \quad \hat{L}(G, f) = \{g \in (X \cup V)^* / f \xrightarrow{*} g\}.$$

Le langage engendré par G est $L(G, A)$.

Une grammaire est *réduite* si, pour tout v dans V :

(i) $L(G, v) \neq \emptyset$;

(ii) il existe x et y dans X^* tels que $A \xrightarrow{*} xvy$ (i.e. tout non-terminal est accessible à partir de l'axiome);

(iii) toute règle est utile i.e. pour toute règle $v \rightarrow \alpha$ de P , $L(G, v) \neq L(G', v)$, où G' est la grammaire $G' = (X, V, P - \{v \rightarrow \alpha\})$.

Une grammaire algébrique est *simple* si $P \subset V \times XV^*$ et si, pour tout couple $(x, v) \in X \times V$, il existe au plus un m appartenant à V^* tel que $v \rightarrow xm$ soit une règle de P , i.e. $v \rightarrow xm, v' \rightarrow xm', v = v' \Rightarrow m = m'$.

L est un langage simple s'il existe une grammaire simple qui l'engendre.

Une grammaire algébrique est *très simple* si $P \subset V \times XV^*$ et si, pour tout x de X , il existe au plus une règle $v \rightarrow xm$, avec $v \in V, m \in V^*$, i.e. $v \rightarrow xm, v' \rightarrow xm' \Rightarrow v = v'$ et $m = m'$. Une telle grammaire est *a fortiori* simple.

L est un langage très simple s'il existe une grammaire très simple qui l'engendre.

Un langage L est *préfixe* si $f \in L, fg \in L \Rightarrow g = 1$.

On sait d'après [10], que les langages simples sont préfixes. Les langages très simples le sont donc aussi. Un langage L est *suffixe* si $g \in L, fg \in L \Rightarrow f = 1$.

Si α est un mot de $(X \cup V)^+$, on définira $\text{Int}(\alpha)$ comme la première lettre de α .

Nous ne considérerons que des grammaires très simples réduites.

Nous donnons d'abord quelques propriétés concernant les grammaires très simples. Ensuite, nous établirons un lemme concernant l'expansivité des grammaires très simples, et un autre lemme sur les langages très simples qui contiennent un sous-ensemble rationnel infini. Nous prouverons ensuite une proposition fondamentale dont le théorème annoncé sera corollaire immédiat.

3. QUELQUES PROPRIÉTÉS DE GRAMMAIRES TRÈS SIMPLES

Nous commençons par rappeler deux propriétés qui sont vraies dans le cas plus général des grammaires simples.

LEMME 1. — Soit $G = (X, V, P, A)$ une grammaire simple, m_0, m, m' dans V^* , u dans X^* , tels que $m_0 \xrightarrow{*} um, m_0 \xrightarrow{*} um'$. Alors $m = m'$.

Preuve : La démonstration se fait par récurrence sur la longueur de u .

Si $|u| = 0$, alors $m_0 = m$, et $m_0 = m'$. D'où $m = m'$. Si $u = au_1$, où $a \in X$ et $u_1 \in X^*$, alors il existe une règle $v \rightarrow ap$ où $v \in V, p \in V^*$. m_0 se factorise donc en $m_0 = vm'_0$, et pm'_0 engendre $u_1 m$ et $u_1 m'$. L'hypothèse de récurrence entraîne $m = m'$. \square

LEMME 2 : Soit $G=(X, V, P, A)$ une grammaire simple, m dans V^* , α dans $(X \cup V)^*$, f dans X^* tels que $m \xrightarrow{*} \alpha f$. Alors il existe \hat{f} dans V^* tel que $m \xrightarrow{*} \alpha \hat{f}$ et $\hat{f} \xrightarrow{*} f$.

Preuve : La démonstration se fait par récurrence sur la longueur de α .

Si $|\alpha|=0$, c'est trivial.

Supposons maintenant que $\alpha = a\alpha_1$, où $a \in X$. Alors il existe une règle $v \rightarrow ap$, où $v \in V$, $p \in V^*$. m se factorise donc en $m = vm'$, et pm' engendre $\alpha_1 f$. L'hypothèse de récurrence entraîne qu'il existe $\hat{f} \in V^*$ tel que $\hat{f} \xrightarrow{*} f$ et $pm' \xrightarrow{*} \alpha_1 \hat{f}$. D'où $m \rightarrow apm' \xrightarrow{*} a\alpha_1 \hat{f} = \alpha \hat{f}$. Supposons enfin que $\alpha = w\alpha_2$, où $w \in V$. Comme la grammaire est simple, m se factorise en $m = wm''$, et $m'' \xrightarrow{*} \alpha_2 f$. L'hypothèse de récurrence entraîne qu'il existe \hat{f} tel que $m'' \xrightarrow{*} \alpha_2 \hat{f}$ et $\hat{f} \xrightarrow{*} f$. Alors $m = wm'' \xrightarrow{*} w\alpha_2 \hat{f} = \alpha \hat{f}$. \square

Nous parlons maintenant de grammaires très simples.

LEMME 3 : Soit $G=(X, V, P, A)$ une grammaire très simple, m et p dans V^+ , f et g dans $(X \cup V)^+$, tels que $m \xrightarrow{*} f$, $p \xrightarrow{*} g$. Si f et g commencent par la même lettre, alors m et p commencent par la même lettre.

Preuve : Comme G est très simple, $\text{Init}(f)$ et $\text{Init}(g)$ sont éléments de X , i.e. $\text{Init}(f) = \text{Init}(g) = x$. Il existe une unique règle $v \rightarrow xw$, où $v \in V$, $w \in V^*$, qui fasse apparaître x . Donc m et p commencent tous les deux par v . \square

LEMME 4 : Soit $G=(X, V, P, A)$ une grammaire très simple, m dans V^+ , T dans V , u dans $(X \cup V)^+$, g dans $(X \cup V)^*$, tels que $m \xrightarrow{*} u$, $T \xrightarrow{*} ug$. Alors $m = T$ et $g = 1$.

Preuve : u et ug commencent par la même lettre, donc, en vertu du lemme 3, m commence par T , i.e. $m = Tm''$, où $m'' \in V^*$. Puisque $Tm'' \xrightarrow{*} u$, il existe une factorisation $u = u'u''$ avec $T \xrightarrow{*} u'$, $m'' \xrightarrow{*} u''$. Or $T \xrightarrow{*} ug$, et $\hat{L}(G, T)$ est préfixe; donc $g = u'' = 1$, et $m'' = 1$. \square

PROPOSITION 1 : Soit $G=(X, V, P, A)$ une grammaire très simple, m et p deux mots de V^+ . Alors $L(G, m) \cap L(G, p) \neq \emptyset$ si et seulement si $m = p$.

Preuve : La démonstration se fait par récurrence sur la somme des longueurs de m et de p . Soit $f \in L(G, m) \cap L(G, p)$. Supposons $|m| = |p| = 1$. Le lemme 3 appliqué à f entraîne que m et p commencent par la même lettre, donc sont égaux.

Supposons maintenant $|m| = 1$ et $p = w_1 \dots w_k$, où $w_i \in V$, pour $i = 1, \dots, k$. Le lemme 3 implique $m = w_1$, et le lemme 4 $w_2 \dots w_k = 1$, i.e. $m = p$.

Soit désormais $m = v_1 \dots v_n$, avec $v_i \in V$, et $p = w_1 \dots w_k$. Par le lemme 3, $v_1 = w_1$. Comme $L(G, v_1)$ est préfixe, il existe une factorisation $f = f_1 f_2$ telle que $v_1 = w_1 \xrightarrow{*} f_1, v_2 \dots v_n \xrightarrow{*} f_2, w_2 \dots w_k \xrightarrow{*} f_2$. Par récurrence $v_2 \dots v_n = w_2 \dots w_k$, d'où $m = p$.

La réciproque est triviale. \square

PROPOSITION 2 : Soit $G = (X, V, P, A)$ une grammaire très simple, m et p deux mots de V^+ , α un mot de $(X \cup V)^*$, u et h dans X^* . Si $m \xrightarrow{*} \alpha u h$ et $p \xrightarrow{*} u$, alors $m \xrightarrow{*} \alpha p h$.

Preuve : Puisque $m \xrightarrow{*} \alpha u h$, il existe m' dans V^* tel que $m \xrightarrow{*} \alpha m', m' \xrightarrow{*} u h$ (lemme 2). Il existe une factorisation $m' = m'_1 v m'_2$ où $v \in V$, telle que $m'_1 \xrightarrow{*} u_1, v \xrightarrow{*} u_2 h_1, m'_2 \xrightarrow{*} h_2$, avec $u_1 u_2 = u$ et $h_1 h_2 = h$.

1^{er} cas : $u_2 \neq 1$. Comme $p \xrightarrow{*} u_1 u_2$, il existe un mot $\Theta \in V^*$ tel que $p \xrightarrow{*} u_1 \Theta$, et $\Theta \xrightarrow{*} u_2$ (lemme 2). Le lemme 4 appliqué aux dérivations $v \xrightarrow{*} u_2 h_1$ et $\Theta \xrightarrow{*} u_2$ implique $\Theta = v, h_1 = 1$. D'où $m'_1 v \xrightarrow{*} u_1 u_2 = u$. Par la proposition 1, $m'_1 v = p$. Par ailleurs, $m'_2 \xrightarrow{*} h$. D'où $m \xrightarrow{*} \alpha p m'_2 \xrightarrow{*} \alpha p h$.

2^e cas : $u_2 = 1$. Alors $m'_1 \xrightarrow{*} u$. D'après la proposition 1, $m'_1 = p$. D'où $m \xrightarrow{*} \alpha p m'_2 \xrightarrow{*} \alpha p h$. \square

PROPOSITION 3 : Soit $G = (X, V, P, A)$ une grammaire très simple, T, T_1, \dots, T_n des éléments de V, f, g et u des mots de X^* , $u \neq 1$. Si $T \xrightarrow{*} u g$ et $T_1 \dots T_n \xrightarrow{*} f u$, alors $g = 1$.

Preuve : La démonstration se fait par récurrence sur n . Si $n = 1, T_1 \xrightarrow{*} f u$. Alors $T_1 \xrightarrow{*} f \hat{u}$, où $\hat{u} \in V^*$, et $\hat{u} \xrightarrow{*} u$. Mais $T \xrightarrow{*} u g$. D'après le lemme 4, $T = \hat{u}$ et $g = 1$.

Supposons maintenant le résultat vrai pour $n - 1$. Deux cas se présentent :

1^{er} cas : $T_n \xrightarrow{*} f' u$, où $f = f' f''$ (f'' peut être éventuellement vide). Comme $T \xrightarrow{*} u g$, l'hypothèse de récurrence entraîne $g = 1$.

2^e cas : $T_n \xrightarrow{*} u''$, où $u = u' u''$. Alors $T_1 \dots T_{n-1} \xrightarrow{*} f u'$. Mais $T \xrightarrow{*} u g = u' (u'' g)$. L'hypothèse de récurrence entraîne $u'' g = 1$, d'où $g = 1$.

4. DEUX LEMMES

On introduit sur V un préordre de la manière suivante : On dira que v domine v' (noté $v \geq v'$) si et seulement si il existe x et y dans X^* tels que $v \xrightarrow{*} x v' y$. Ce préordre induit une relation d'équivalence \simeq sur V définie par $v \simeq v'$ si et seulement si $v \geq v'$ et $v' \geq v$.

Le degré d'expansivité d'une règle $r = v \rightarrow am$ de G est défini ainsi : $d(r)$ est le nombre d'occurrences de lettres non terminales dans m qui sont équivalentes à v . Le degré d'expansivité $d(v)$ d'une variable v de V est alors le maximum des nombres $d(r)$ pour toutes les règles r qui ont v comme membre gauche.

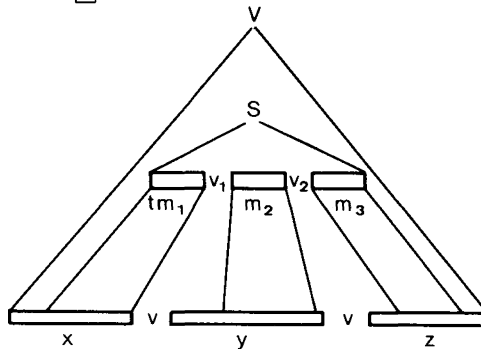
Rappelons qu'une grammaire est *expansive* (not derivation bounded) si elle contient une dérivation de la forme $v \xrightarrow{*} xvyvz$, avec v dans V , x , y et z dans X^* . Un langage est expansif (« not derivation bounded ») si toute grammaire qui l'engendre est expansive. Il est clair que si un langage très simple est expansif, alors toute grammaire très simple qui l'engendre est expansive.

On a alors le :

LEMME 5 : Si $G = (X, V, P, A)$ est une grammaire très simple, expansive, alors il existe un non-terminal dont le degré d'expansivité est supérieur ou égal à deux.

Preuve : Puisque G est expansive, il existe v dans V , x , y et z dans X^* tels que $v \xrightarrow{*} xvyvz$.

Considérons l'arbre de dérivation associé à cette dérivation. Et considérons le plus petit sous-arbre contenant les feuilles étiquetées par les deux occurrences de v dans le mot $xvyvz$. Soit S l'étiquette de la racine de ce sous-arbre. En vertu de la minimalité, on a la règle $S \rightarrow tm_1 v_1 m_2 v_2 m_3$, où $t \in X$, $m_1, m_2, m_3 \in V^*$, $v_1 v_2 \in V$, et où v_1 est la racine d'un sous-arbre contenant la première occurrence de v , v_2 la racine d'un sous-arbre contenant la seconde. (Ces deux sous-arbres sont distincts, à cause de la minimalité sur le sous-arbre de racine S .) Alors S est de degré au moins deux. \square



Soit G une grammaire très simple. On dira que G admet un facteur itérant si, et seulement si, il existe dans G une dérivation de la forme $T \xrightarrow{*} uT$, avec T dans V , u dans $(X \cup V)^*$.

On peut énoncer le :

LEMME 6 : Soit $G = (X, V, P, A)$ une grammaire très simple, $L = L(G, A)$. L contient un ensemble rationnel infini si et seulement si G admet un facteur itérant.

Preuve : Supposons qu'il existe α, u, β dans X^* tels que $\alpha u^* \beta \in L$. Alors il existe deux dérivations $A \xrightarrow{*} \alpha u \beta$ et $A \xrightarrow{*} \alpha u u \beta$. Par le lemme 2, il existe m, m' dans V^* tels que $A \xrightarrow{*} \alpha m, m \xrightarrow{*} u \beta, A \xrightarrow{*} \alpha m', m' \xrightarrow{*} u u \beta$. Par le lemme 1, on a $m = m'$. Par la proposition 2, $m \xrightarrow{*} u m$. Posons $m = v_1 \dots v_k$, où $v_i \in V$ pour $i = 1, \dots, k$. Donc il existe des entiers, i et j et une factorisation $u = u' u''$ tels que $v_1 \dots v_{i-1} \xrightarrow{*} u', v_i \xrightarrow{*} u'' v_1 \dots v_j, v_{i+1} \dots v_k \xrightarrow{*} v_{j+1} \dots v_k$. Comme la grammaire est très simple, la dernière de ces dérivations implique $i = j$. On a donc $v_i \xrightarrow{*} u'' v_1 \dots v_{i-1} v_i \xrightarrow{*} u'' u' v_i$, ce qui montre l'existence d'un facteur itérant dans G .

La réciproque est évidente, parce que la grammaire est réduite. \square

LEMME 7 : Soit $G = (X, V, P, A)$ une grammaire très simple. G est sans facteur itérant, si et seulement si, pour tout $m \in V^+, L(G, m)$ est suffixe (donc biffixe) [de même pour $\hat{L}(G, m)$].

Preuve : Supposons que u et $f \in \hat{L}(G, m)$ et que $m \xrightarrow{*} u, m \xrightarrow{*} u f$. Par la proposition 2 on a $m \xrightarrow{*} u m$, et une démonstration similaire à celle du lemme 6 montre que G possède alors un facteur itérant.

Réciproquement si G a un facteur itérant, alors il existe v tel que $v \xrightarrow{*} \alpha v$ avec α dans $(X \in V)^+$. Comme G est réduite, il existe f tel que $v \xrightarrow{*} f$. Alors f et αf sont tous deux dans $\hat{L}(G, v)$ qui n'est donc pas suffixe. \square

COROLLAIRE : Soit $G = (X, V, P, A)$ une grammaire très simple, sans facteur itérant, m_1 et $m_2 \in V^+, \alpha$ et $\beta \in (X \cup V)^*$, tels que $m_1 m_2 \xrightarrow{*} \alpha \beta$ et $m_2 \xrightarrow{*} \beta$. Alors $m_1 \xrightarrow{*} \alpha$.

Preuve : Supposons que $m_1 \xrightarrow{*} \alpha', m_2 \xrightarrow{*} \alpha' \beta$, avec $\alpha = \alpha' \alpha''$. En vertu de la suffixité, et de la dérivation $m_2 \xrightarrow{*} \beta, \alpha'' = 1$. D'où $m_1 \xrightarrow{*} \alpha$.

Supposons maintenant que $m_1 \xrightarrow{*} \alpha \beta', m_2 \xrightarrow{*} \beta''$, avec $\beta = \beta' \beta''$. Alors $m_2 \xrightarrow{*} \beta' \beta'', m_2 \xrightarrow{*} \beta''$ et la suffixité de $\hat{L}(G, m_2)$ entraînent $\beta' = 1$, d'où $m_1 \xrightarrow{*} \alpha$. Donc dans tous les cas $m_1 \xrightarrow{*} \alpha$. \square

5. LA PROPOSITION FONDAMENTALE

PROPOSITION 4 : Soit $G = (X, V, P, A)$ une grammaire très simple, expansive, sans facteur itérant. Alors il existe un non-terminal S tel que $L(G, S)$ soit générateur de la famille des langages algébriques.

Pour montrer cette proposition, nous allons construire un langage rationnel K de X^* et un morphisme λ tels que E , le langage engendré par la grammaire de règles $\{S \rightarrow a S b S c, S \rightarrow d\}$ soit égal à $\lambda^{-1}(L(G, S) \cap K)$. Comme E est générateur de la famille Alg, $L(G, S)$ le sera également.

Prouver que E est inclus dans $\lambda^{-1}(L(G, S) \cap K)$ est facile, puisque l'on construit K dans ce but.

La réciproque se fera de la manière suivante : soit un mot f de $\lambda^{-1}(L(G, S) \cap K)$ et $g = \lambda f$. Nous montrerons que g s'écrit $\alpha g_1 \beta g_2 \gamma$ où α, β, γ sont des éléments de X^* , et g_1 et g_2 sont encore dans $L(G, S) \cap K$. Comme les longueurs de g_1 et de g_2 sont strictement inférieures à celle de g , on en déduira le résultat.

Preuve : D'après le lemme 5, il existe S dans V et une règle $r : S \rightarrow xm$, où $x \in X$ et $m \in V^*$ telle que m contient au moins deux non-terminaux dominant S . Soit U le non-terminal le plus à gauche dominant S , T le non-terminal le plus à droite dominant S . m s'écrit $m = m_1 U m_2 T m_3$, avec m_1, m_2, m_3 dans V^* . D'où $r : S \rightarrow x m_1 U m_2 T m_3$. Soit δ un mot de $L(G, S)$ tel qu'il existe une dérivation $S \xrightarrow{*} \delta$ dont la longueur est minimale parmi toutes les dérivations $S \xrightarrow{*} \delta'$ ($\delta' \in X^*$). Comme G est très simple, il est équivalent de dire que δ est un mot de longueur minimale dérivé de S .

Soient f_1 un mot de longueur minimale dans $L(G, m_1)$, f_3 un mot de longueur minimale dans $L(G, m_3)$.

Puisque U domine S , il existe u_1 dans X^* , \hat{u}_2 dans V^* tels que $U \xrightarrow{*} u_1 S \hat{u}_2$, et que u_1 soit de longueur minimale. On a donc :

$$S \xrightarrow{*} x f_1 u_1 S \hat{u}_2 m_2 T f_3 \quad (1)$$

Puisque T domine S :

$$\mathcal{F} = \{(t_1, \hat{t}_2) \in X^* \times V^*/T \xrightarrow{*} t_1 S \hat{t}_2\},$$

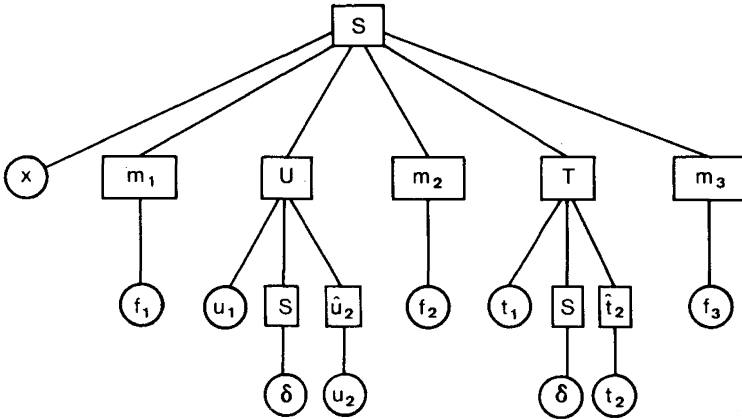
n'est pas vide. Deux cas se présentent alors :

1^{er} cas : Il existe dans \mathcal{F} des couples (t_1, \hat{t}_2) tels que \hat{t}_2 puisse se dériver en un mot t_2 de X^* sans réutiliser T . Alors on choisira parmi ces couples (t_1, \hat{t}_2) celui pour lequel les longueurs de t_1 et de t_2 sont minimales. Chaque non-terminal différent de S et de T figurant dans (1) est dérivé en un mot de X^* dont la longueur est minimale; S est dérivé en δ , T en $t_1 S \hat{t}_2$ puis $t_1 \delta t_2$.

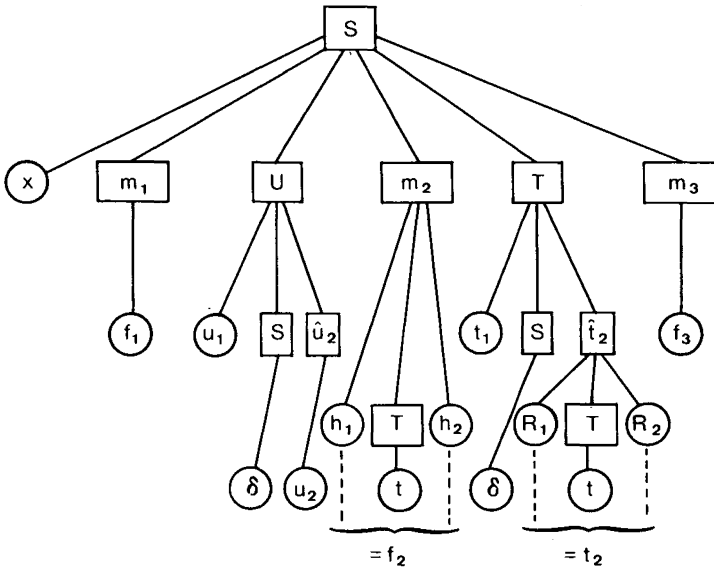
2^e cas : Il n'existe pas dans \mathcal{F} de (t_1, \hat{t}_2) tels que \hat{t}_2 puisse se dériver en un mot terminal sans réutiliser T . Alors il existe t dans X^* tel que $T \xrightarrow{*} t$ sans réutiliser S ; sinon $L(G, \hat{t}_2)$ serait vide, et la grammaire ne serait pas réduite. On dérivera donc dans (1) T en $t_1 S \hat{t}_2$, puis toute occurrence de T figurant dans $\hat{u}_2 m_2$ et \hat{t}_2 en t ; S est dérivé en δ , et tout non-terminal différent de S et de T comme dans le premier cas. On appelle t_2 le mot de X^* dérivé de \hat{t}_2 .

Appelons u_2 et f_2 les mots de X^* ainsi dérivés respectivement de \hat{u}_2 et m_2 . On a

$$S^* \rightarrow x f_1 u_1 S u_2 f_2 t_1 S t_2 f_3.$$



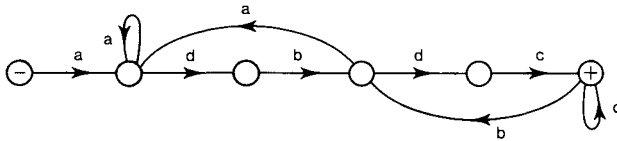
Cas I.



Cas II.

Posons $\alpha = x f_1 u_1$, $\beta = u_2 f_2 t_1$, $\gamma = t_2 f_3$. Soit $Y = \{a, b, c, d\}$ un nouvel alphabet distinct de X et de V , et E le langage engendré par la grammaire $(Y, \{S\}, P, \{S\})$, où $P = \{S \rightarrow a S b S c, S \rightarrow d\}$. On rappelle que E engendre la

famille Alg. Soit $R = R_1 \cup \{d\}$ où R_1 est le langage rationnel reconnu par l'automate fini schématisé de la manière suivante :



On définit un morphisme λ de Y^* dans X^* par $\lambda(a) = \alpha$, $\lambda(b) = \beta$, $\lambda(c) = \gamma$, $\lambda(d) = \delta$. On pose $K = \lambda(R)$. K est un langage rationnel de X^* . Il est clair que $E \subseteq \lambda^{-1}(L(G, S) \cap K)$.

Réciproquement, soit $f \in \lambda^{-1}(L(G, S) \cap K)$ et $g = \lambda f$. Nous allons montrer que soit $g = \delta$, soit $g = \alpha g_1 \beta g_2 \gamma$ où g_1 et g_2 appartiennent à $L(G, S) \cap K$.

On a $g \in L(G, S) \cap K$. Si $g = \delta$, le résultat annoncé est immédiat. Supposons donc que g est différent de δ :

(a) g s'écrit αh . Il existe, en vertu du lemme 2, \hat{h} dans V^* tel que $S \xrightarrow{*} \alpha \hat{h}$, $\hat{h} \xrightarrow{*} h$. Or on sait que $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 T m_3$. Le lemme 1 assure que $\hat{h} = S \hat{u}_2 m_2 T m_3$, d'où $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 T m_3 \xrightarrow{*} g$;

(b) faisons tout d'abord une remarque. Notons $X_v = \{x \in X / m \in V^* \text{ et } (v \rightarrow xm) \in P\}$. Comme la grammaire est très simple, il est clair que $(X_v)_{v \in V}$ forme une partition de X .

1. Supposons que l'on soit dans le 1^{er} cas, c'est-à-dire que toute occurrence de T ait été dérivée en $t_1 S \hat{t}_2$.

Notons $z = \text{Init}(t_1)$.

Deux cas se présentent à nouveau :

(i) S ne réutilise pas T dans la dérivation engendrant δ , et on n'a pas utilisé T dans la dérivation de T en $t_1 S \hat{t}_2$ pour faire apparaître t_1 .

Alors, si $z' \in X_T - \{z\}$, z' ne figure ni dans α , ni dans β , ni dans γ , ni dans δ .

En effet, supposons que z soit une lettre de α , β , ou γ . Le seul non-terminal faisant apparaître z' est T . Or T , par construction de K , est toujours dérivé en $t_1 S \hat{t}_2$, puis $t_1 \delta t_2$. On sait alors que \hat{t}_2 a été dérivé en t_2 sans utiliser T . Donc z' n'est pas une lettre de t_2 . Puisque z est distinct de z' , t_1 ne contient pas z' .

Reste le cas de δ . Si z' est une lettre de δ , c'est qu'on a utilisé T dans la dérivation de δ , contrairement à l'hypothèse. D'où si h est un mot de K , $|h|_{X_T} = |h|_z$, i. e. la seule lettre de X_T figurant dans h est z .

De plus, toute occurrence de z dans $u_2 f_2, f_3, \alpha$ est le début d'une occurrence de $t_1 \delta t_2$: en effet, z ne provient que de T , qui a toujours été dérivé en $z t'_1 S \hat{t}_2$, où $z t'_1 = t_1$.

On sait que $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 T m_3 \xrightarrow{*} g$. g se termine par $\gamma = t_2 f_3$ car $g \in K$. Or $m_3 \xrightarrow{*} f_3$. En vertu du corollaire du lemme 7 on a $S \hat{u}_2 m_2 T \xrightarrow{*} k t_2$, où $k \in (X \cup V)^*$.

Alors T ne peut engendrer un facteur droit de t_2 , car $T \xrightarrow{*} t_1 \delta t_2$, et dans ce cas $L(G, T)$ ne serait pas suffixe, contrairement au lemme 7. Donc $T \xrightarrow{*} k_2 t_2$, où k_2 est facteur droit strict de k . $k_2 t_2$ est facteur d'un mot de k , donc, d'après ci-dessus, $|k_2 t_2|_{X_T} = |k_2 t_2|_z$. Or on sait que z est toujours suivi de t'_1 . Donc $T \xrightarrow{*} z t'_1 t_2 = k_2 t_2$ [il n'est pas possible que t_1 et t_2 aient un facteur commun, i.e. qu'un facteur droit de t_1 soit facteur gauche de t_2 , car alors $t_1 \delta t_2$ ne serait pas de longueur minimale dans $L(G, T)$].

Mais on sait que $T \xrightarrow{*} t_1 S \hat{t}_2$. Or il existe, en vertu du lemme 2, $\Theta \in V^*$ tel que $T \xrightarrow{*} t_1 \Theta$, $\Theta \xrightarrow{*} l t_2$. Par le lemme 1, il vient $\Theta = S \hat{t}_2$. Donc $S \xrightarrow{*} l$, $T \xrightarrow{*} t_1 S \hat{t}_2 \xrightarrow{*} t_1 l t_2$, d'où $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 t_1 S \gamma \xrightarrow{*} g$.

(ii) S doit utiliser T pour engendrer δ , ou T a été utilisé dans la dérivation de T en $t_1 S \hat{t}_2$ pour faire apparaître t_1 . Alors il existe $t \in L(G, T)$ tel que $T \xrightarrow{*} t$ sans réutiliser S (sinon G ne serait pas réduite, et t_1 ne serait pas de longueur minimale). Un tel t peut être choisi de longueur minimale. Soit $j = \text{Init}(t)$. Alors, si $j' \in X_T - \{z, j\}$, j' ne figure ni dans α , ni dans β , ni dans γ , ni dans δ . En effet, si j' est une lettre d'un de ces mots, c'est qu'on a utilisé T , qui est toujours dérivé en faisant apparaître z ou j .

De plus, si h est un mot de K qui se factorise en $h_1 z h_2$ alors z est suivi de t'_1 tel que $t_1 = z t'_1$ [comme dans le cas (i)]. Si h est un mot de K qui se factorise en $h'_1 j h'_2$, alors j est suivi de t' où $t = j t'$ (car j ne peut provenir de T dans la dérivation de t).

On a $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 T m_3 \xrightarrow{*} g$, qui se termine par $\gamma = t_2 f_3$. Comme dans (i), $S \hat{u}_2 m_2 T \xrightarrow{*} k t_2$, où $k \in (X \cup V)^*$, et t n'est pas facteur droit de t_2 . Alors $T \xrightarrow{*} k_2 t_2$, où $k_1 k_2 = k$, $k_2 \neq 1$. $\text{Init}(k_2 t_2)$ est donc égal à z ou j :

– $\text{Init}(k_2 t_2) = z$; le même raisonnement que dans (i) montre que l'on a forcément $T \xrightarrow{*} t_1 S \hat{t}_2$ dans la dérivation de S en g , i.e. $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 t_1 S \gamma \xrightarrow{*} g$.

– $\text{Init}(k_2 t_2) = j$. Alors comme j est suivi de t' , $k_2 t_2 = j t' h = t h$, avec $h \in X^*$. Mais $T \xrightarrow{*} k_2 t_2 = t h$ et $T \xrightarrow{*} t$ entraînent, en vertu de la préfixité, $h = 1$.

D'où $k_2 t_2 = t$, et $S \hat{u}_2 m_2 T \xrightarrow{*} k_1 (k_2 t_2) = k_1 t$, avec $T \xrightarrow{*} t$.

g se termine par γ . Devant γ , on peut avoir δ ou γ :

1° On a δ . Deux cas se produisent :

– k_2 est facteur droit de δ , i.e. il existe $h \in X^*$ tel que $\delta = h k_2$. Alors $S \xrightarrow{*} h k_2$, $T \xrightarrow{*} t = k_2 t_2$, $k_2 \neq 1$ et la proposition 3 entraînent $t_2 = 1$. D'où $\hat{t}_2 = 1$, et $T \xrightarrow{*} k_2$.

$S \xrightarrow{*} hk_2$; par la proposition 2, on tire $S \xrightarrow{*} hT$. Or $T \xrightarrow{*} t_1 S \hat{t}_2 = t_1 S$, donc $S \xrightarrow{*} hT \xrightarrow{*} ht_1 S$, ce qui entraîne l'existence d'un facteur itérant, contrairement à l'hypothèse;

– δ est facteur droit de k_2 , i. e. $k_2 = k'_2 k''_2$ et $\delta = k''_2$. Alors $S \xrightarrow{*} k'_2$, $T \xrightarrow{*} k'_2 k''_2 t_2 = t$ et la proposition 2 entraînent $T \xrightarrow{*} k'_2 S t_2 \xrightarrow{*} t$. Or on a supposé que t était dérivé de T sans S . Donc c'est impossible.

2° On a γ . Nous avons encore deux cas :

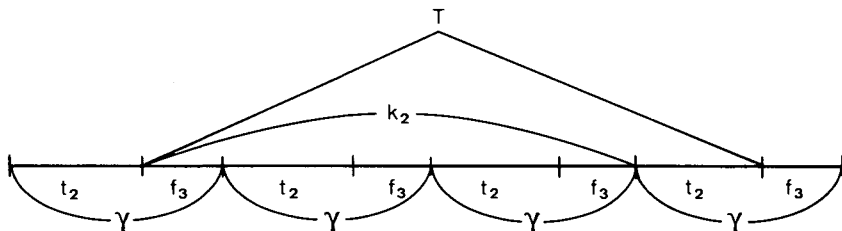
– k_2 est facteur droit de γ . Alors $\gamma = rk_2$, où $r \in X^*$. De $T \xrightarrow{*} k_2 t_2$, $\hat{t}_2 m_3 \xrightarrow{*} \gamma = rk_2$ et $k_2 \neq 1$, on déduit, en vertu de la proposition 3, $t_2 = 1$. Donc $T \xrightarrow{*} k_2$, $\hat{t}_2 = 1$ et $m_3 \xrightarrow{*} rk_2$. D'où $m_3 \xrightarrow{*} rT$ (prop. 2), et $T \xrightarrow{*} t_1 S$. On a donc :

$$S \xrightarrow{*} \alpha S \hat{u}_2 m_2 T m_3 \xrightarrow{*} \alpha S \hat{u}_2 m_2 T r T \xrightarrow{*} \alpha S \hat{u}_2 m_2 T r t_1 S,$$

et on a un facteur itérant contrairement à l'hypothèse;

– γ est facteur droit de k_2 .

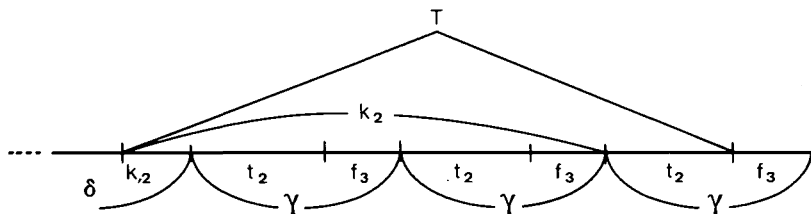
(a)



Il existe $n > 0$ tel que γ^n est facteur droit de k_2 , et γ^{n+1} est plus grand que k_2 . (Car dans le rationnel, devant γ ou encore γ ou δ .)

Alors $k_2 = k'_2 \gamma^n$ et le même raisonnement que plus haut appliqué à k'_2 au lieu de k_2 aboutit à une contradiction.

(b) $k_2 = k'_2 \gamma^n$, mais devant γ^n on a δ .



Si k'_2 est facteur droit de δ , on fait le même raisonnement que dans 1° (premier cas).

Si δ est facteur droit de k'_2 , c'est la même chose que dans 1° (deuxième cas).

Aussi, dans tous les cas de figure, on aboutit à une contradiction.

La seule possibilité est donc $\text{Init}(k_2 t_2) = z$ et $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 t_1 S \gamma \xrightarrow{*} g$.

2. On suppose maintenant que l'on se trouve dans le deuxième cas, c'est-à-dire que l'on a dérivé dans (1) T en $t_1 S \hat{t}_2$, puis toute occurrence de T pouvant survenir en t . Le même raisonnement que dans le cas 1 (ii) permet de conclure que $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 t_1 S \gamma \xrightarrow{*} g$.

(c) On est donc, dans tous les cas dans la situation :

$$S \xrightarrow{*} \alpha S \hat{u}_2 m_2 t_1 S \gamma \xrightarrow{*} g.$$

On rappelle que $x = \text{Init}(\alpha)$. Notons $y = \text{Init}(\delta)$. Par construction, toute occurrence de S dans la règle r a été dérivée en δ , ou αh avec $h \in (X \cup V)^*$. Si $x' \in X_S - \{x, y\}$, alors x' ne figure ni dans α ni dans β , ni dans γ , ni dans δ : en effet, supposons que $S \rightarrow x' m'$, et que x' soit une lettre de δ ; alors on a $S \xrightarrow{*} h_1 S h_2 \xrightarrow{*} h_1 x' m' h_2 \xrightarrow{*} \delta$, avec h_1 et h_2 dans $(X \cup V)^*$, $h_1 \neq 1$ et dans ce cas δ n'est pas le mot le plus court de $L(G, S)$.

x' ne peut être une lettre de α , β ou γ : supposons le contraire, alors on aurait utilisé S . Or S a toujours été dérivé en δ (qui ne contient pas x') et u_1 et t_1 étant de longueur minimale, x' ne peut y figurer.

Donc si un mot $h \in L(G, S) \cap K$, sa lettre initiale est x ou y . Si $\text{Init}(h) = y$, alors $h = \delta$, car S est toujours dérivé en δ après y ; si $\text{Init}(h) = x$, alors h commence par α , et on a utilisé la règle r .

Revenons à la dérivation $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 t_1 S \gamma \xrightarrow{*} g$. L'occurrence de S placée à droite dans $\alpha S \hat{u}_2 m_2 t_1 S \gamma$ ne peut donner que δ , ou un mot commençant par α (d'après le raisonnement supra) :

— si $S \xrightarrow{*} \delta$, alors ce δ est précédé de β , sinon g ne serait pas dans K . Or on sait que $S \xrightarrow{*} \alpha S \beta S \gamma$, et donc (lemme 2), qu'il existe $\Theta \in V^*$ tel que $S \xrightarrow{*} \alpha S \Theta$, $\Theta \xrightarrow{*} \beta S \gamma$. Or $S \xrightarrow{*} \alpha S \hat{u}_2 m_2 T m_3$. En vertu du lemme 1, on a

$$S \hat{u}_2 m_2 T m_3 = S \Theta \quad \text{i. e.} \quad \Theta = \hat{u}_2 m_2 T m_3.$$

Donc $\hat{u}_2 m_2 T m_3 \xrightarrow{*} \beta S \gamma$, et $S \xrightarrow{*} \alpha S \beta S \gamma \xrightarrow{*} g$;

— si S engendre un mot commençant par α , on a devant cet α , soit β soit α (sinon g ne serait pas dans K);

— si c'est β , alors $S \xrightarrow{*} \alpha S \beta S \gamma \xrightarrow{*} g$ comme ci-dessus;

– si c’est α , alors $\alpha = \alpha_1 t_1$. Or $\hat{u}_2 m_2 \xrightarrow{*} l \alpha_1$, où $l \in (X \cup V)^*$ et $S \xrightarrow{*} \alpha k = \alpha_1 (t_1 k)$. D’après la proposition 3, $t_1 k = 1$, ce qui est impossible.

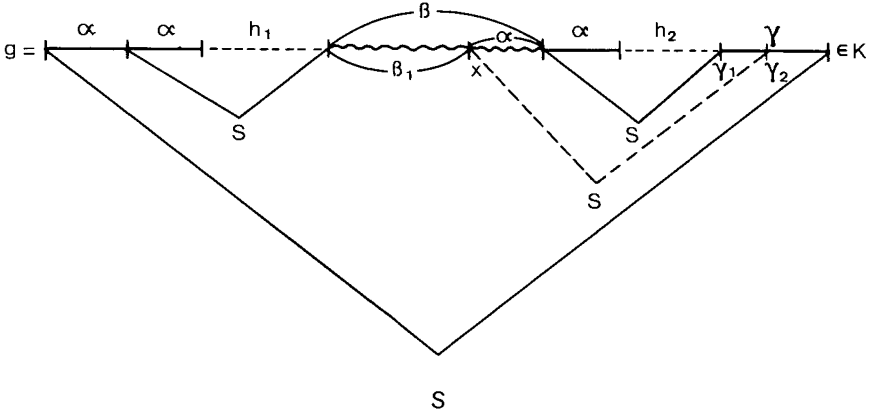
Dans tous les cas on a donc

$$S \xrightarrow{*} \alpha S \beta S \gamma \xrightarrow{*} g \in K.$$

g s’écrit alors $\alpha g_1 \beta g_2 \gamma$, avec g_1 et g_2 dans $L(G, S)$. Montrons que g_1 et g_2 sont aussi dans K .

Après α , on peut avoir α ou δ . Supposons que ce soit δ ; on a alors β . Après β , on a δ ou α . Supposons que ce soit δ ; alors on a forcément γ . D’où, on a $g_1 = \delta \in K$ et $g_2 = \delta \in K$. Considérons maintenant le cas moins trivial, i. e. $g_1 \neq \delta$ et $g_2 \neq \delta$.

Comme $S \xrightarrow{*} g_1$, c’est que $g_1 = \alpha h_1$, $h_1 \in X^*$. De même, $g_2 = \alpha h_2$, $h_2 \in X^*$.



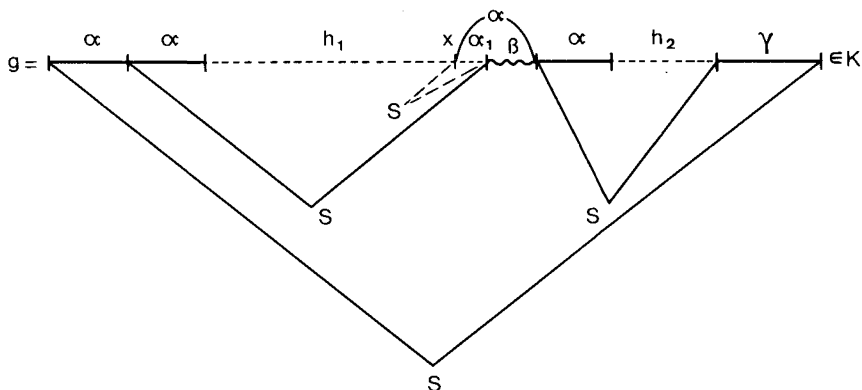
Devant l’occurrence de α la plus à droite, on peut avoir α ou β (sinon $g \notin K$).

Supposons que ce soit α , et que $|\alpha| < |\beta|$.

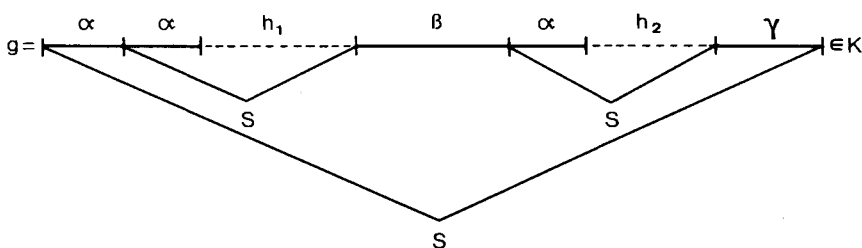
Comme la lettre initiale de α est x , et que seul S fait apparaître x , on a $S \xrightarrow{*} \alpha S \gamma_1$, où γ_1 est facteur gauche strict de γ ; on en déduit donc $S \xrightarrow{*} \alpha S \beta_1 S \gamma_2$, avec $\gamma_1 \gamma_2 = \gamma$ et $\beta_1 \alpha = \beta$.

Or $\alpha S \beta_1 S \gamma_2 \xrightarrow{*} g \in K$. On est donc ramené au problème initial, avec $|\beta_1| < |\beta|$ et $|\gamma_2| < |\gamma_1|$.

De proche en proche, on arrive au même cas, avec $|\alpha| > |\beta|$.



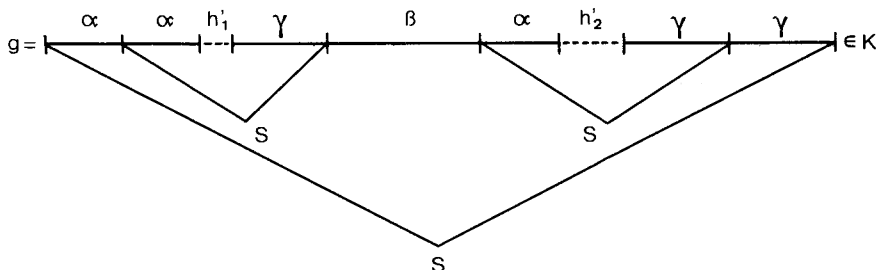
Cette fois-ci, comme $|\alpha| > |\beta|$, la lettre x se trouve dans h_1 . $\alpha = \alpha_1 \beta$, et $S \xrightarrow{*} \alpha_1$. Or $S \xrightarrow{*} \alpha S \beta S \gamma$. Ceci contredit la préfixité. Donc devant α on ne peut avoir que β .



Devant β , on peut avoir δ ou γ .

Si c'est δ , alors $S \xrightarrow{*} \alpha h'_1 \delta$ d'où $S \xrightarrow{*} \alpha h'_1 S$, d'où l'existence d'un facteur itérant, contrairement à l'hypothèse. On a donc γ devant β .

Devant le γ final, on peut avoir δ ou γ . Si c'est δ , le même raisonnement montre qu'il doit y avoir un facteur itérant. Donc c'est γ . On a donc $h_1 = h'_1 \gamma$, $h_2 = h'_2 \gamma$.



$g_1 = \alpha h'_1 \gamma$ est forcément dans K , sinon g ne serait pas dans K . De même pour $\alpha h'_2 \gamma = g_2$.

Le cas où l'un des g_i seulement est égal à δ se traite de la même façon, mais d'un seul côté seulement.

D'où $g = \alpha g_1 \beta g_2 \gamma \in L(G, S) \cap K$ entraîne g_1 et $g_2 \in L(G, S) \cap K$.

DÉFINITION : Soit h un mot de $K \setminus \{\delta\}$. La décomposition $h = h_1 \dots h_n$, où $h_i \in X^*$ pour $i = 1, \dots, n$ est une décomposition de h suivant K si :

- (i) $h_1 = \alpha, h_n = \gamma$;
- (ii) $\forall i = 2, \dots, n-1, h_i \in \{\alpha, \beta, \gamma, \delta\}$;
- (iii) $\forall i = 2, \dots, n-1, (h_i, h_{i+1})$ est une transition du rationnel K .

Nous devons maintenant prouver que tout mot de $K \setminus \{\delta\}$ admet une décomposition unique suivant K . En effet, sinon on aurait $g = \alpha g'_1 \beta g'_2 \gamma$, avec $g'_1 \neq g_1, g'_2 \neq g_2$. g'_1 et g'_2 seraient encore dans K , mais pas forcément dans $L(G, S)$. Or pour terminer la démonstration de cette proposition, nous utilisons le fait que g_1 et g_2 sont non seulement dans K , mais encore dans $L(G, S)$, et que leurs longueurs sont strictement inférieures à celle de g . De proche en proche, on arrive au plus petit mot de $L(G, S) \cap K$, à savoir δ . Ceci entraîne que g est dans $\lambda(E)$.

Montrons maintenant l'unicité de la décomposition suivant K : nous construisons un nouveau langage rationnel \bar{K} de la manière suivante : des dérivations $S \xrightarrow{*} \alpha S \beta S \gamma$ et $S \xrightarrow{*} \delta$, on tire $S \xrightarrow{*} \alpha^n S (\beta \delta \gamma)^n$, d'où $S \xrightarrow{*} \alpha^n S (\beta \delta \gamma)^p \beta S \gamma (\beta \delta \gamma)^q$, avec $n = p + q + 1$. Posons

$$\bar{\alpha} = \alpha^n, \quad \bar{\beta} = (\beta \delta \gamma)^p \beta, \quad \bar{\gamma} = \gamma (\beta \delta \gamma)^q, \quad \bar{\delta} = \delta.$$

On définit \bar{K} comme on l'a fait pour K (en surlignant $\alpha, \beta, \gamma, \delta$). En fait, nous allons montrer qu'un mot \bar{g} de $\bar{K} \setminus \{\bar{\delta}\}$ se décompose de manière unique suivant \bar{K} . Il suffit de regarder la définition de \bar{K} pour voir qu'il n'y a que trois cas possibles de non-unicité de la décomposition :

- 1° $\bar{g} = h'_1 \bar{\alpha} h'_1 = h'_1 \bar{\delta} \bar{\beta} h'_1$, avec unicité de la décomposition sur h'_1 ;
- 2° $\bar{g} = h'_2 \bar{\alpha} h'_2 = h'_2 \bar{\delta} \bar{\gamma} h'_2$, avec unicité de la décomposition sur h'_2 ;
- 3° $\bar{g} = h'_3 \bar{\beta} h'_3 = h'_3 \bar{\gamma} h'_3$, avec unicité de la décomposition sur h'_3 .

Traitons successivement ces trois cas :

$$1^\circ h'_1 \bar{\alpha} h'_1 = h'_1 \bar{\delta} \bar{\beta} h'_1.$$

On a $\bar{\alpha} = \alpha^n$ et $\bar{\delta} = \delta$. On peut toujours choisir n suffisamment grand pour que $|\bar{\alpha}| > |\bar{\delta}|$, d'où $\bar{\alpha} = \bar{\delta} \bar{\alpha}_1$.

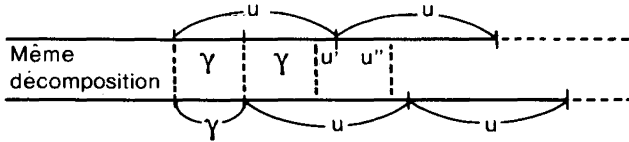
D'où $S \xrightarrow{*} \alpha \delta \beta \delta \gamma = \delta \alpha_1 \delta \beta \delta \gamma$; or $S \xrightarrow{*} \bar{\delta}$, ce qui contredit la préfixité de $L(G, S)$. Ce cas est donc impossible;

2° c'est le même que le 1^{er} cas;

3° $h'_3 \bar{\beta} h'_3 = h'_3 \bar{\gamma} h'_3$.

On a $\bar{\beta} = (\beta\delta\gamma)^p \beta$ et $\bar{\gamma} = \gamma(\beta\delta\gamma)^q$; posons $\beta\delta\gamma = u$. Il est clair que $|u| > |\gamma|$.

On peut faire le schéma suivant :



On obtient $u = \gamma^r u'$, avec $|u'| < |\gamma|$, $r \geq 1$, $\gamma = u' u''$. Alors $u' u'' = \gamma = u'' u'$. Il est bien connu qu'alors u' et u'' sont puissances d'un même mot. Comme $u = \gamma^r u'$ et que $\gamma = u' u''$, γ et u sont puissances d'un même mot, et donc γ et $\beta\delta$ aussi (car $u = \beta\delta\gamma$).

On en tire que γ et $\beta\delta$ commutent, d'où $\alpha\delta\beta\delta\gamma = \alpha\delta\gamma\beta\delta$. Donc $\alpha\delta\gamma\beta\delta$ appartient à $L(G, S)$, ainsi que δ . Ceci contredit la suffixité de $L(G, S)$ (lemme 7).

La décomposition suivant \bar{K} d'un mot de $\bar{K} \setminus \{\bar{\delta}\}$ est donc unique.

Il n'y a aucune perte de généralité à supposer que la décomposition d'un mot de $K \setminus \{\delta\}$ suivant K est unique.

Ceci nous permet de conclure que $E = \lambda^{-1}(L(G, S) \cap K)$, et donc que $L(G, S)$ est générateur. □

On sait qu'une grammaire est quasi-rationnelle si et seulement si elle est non expansive.

On a donc le :

THÉORÈME : *Tout langage très simple qui ne contient pas de sous-ensemble rationnel infini est soit quasi-rationnel, soit générateur de Alg.*

Preuve : Supposons L non quasi-rationnel. Soit $G = (X, V, P, A)$ une grammaire de L . G est expansive, et d'après le lemme 4, G est sans facteur itérant. En vertu de la proposition 4, il existe S dans V tel que $L(G, S)$ soit générateur. Mais $L = L(G, A)$.

G étant réduite, il existe c_1 et c_2 dans X^* tels que $A \xrightarrow{*} c_1 S c_2$. Il est clair que

$$L(G, A) \cap c_1 X^* c_2 = c_1 L(G, S) c_2,$$

et que

$$L(G, A) \cap c_1 K c_2 = c_1 (L(G, S) \cap K) c_2.$$

Donc L s'envoie par transduction rationnelle sur $L(G, S)$ ce qui achève la démonstration. \square

BIBLIOGRAPHIE

1. L. BOASSON, *The Inclusion of the Substitution Closure of Linear and One-Counter Languages in the Largest Full Sub-AFL of the Family of C.F.L. is Proper*, Information Processing Letters, vol. 2, n° 5, 1973, p. 135-140.
2. L. BOASSON, *Un langage algébrique particulier* [R.A.I.R.O., Informatique théorique (à paraître)].
3. L. BOASSON et M. NIVAT, *Sur diverses familles de langages fermées par transduction rationnelle*, Acta Informatica, vol. 2, 1973, p. 180-188.
4. L. BOASSON et M. NIVAT, *Parenthesis Generators*, 17th I.E.E.E. Symposium on foundations of computer science, Houston, 1976, p. 253-257.
5. S. EILENBERG, in Actes Congrès intern. Math., t. 3, 1970, p. 265-267.
6. S. GINSBURG et S. GREIBACH, *Studies in Abstracts Families of Languages*, Memoirs Amer. math. Soc., vol. 87, 1969, p. 1-32.
7. S. GINSBURG et S. GREIBACH, *Principal AFL*, J. Comput. System. Sc., vol. 4, 1970, p. 308-338.
8. S. GREIBACH, *Chains of Full AFL's*, Math. System. Theory, vol. 4, 1970, p. 231-242.
9. S. GREIBACH, *One-Counter Languages and the I.R.S. Condition*, J. Comput. System. Sc., vol. 10, 1975, p. 237-247.
10. HOPCROFT et KORENJAK, *Simple Deterministic Languages*, 7th I.E.E.E.E., Symposium on switching and automata theory, 1966, p. 36-46.
11. M. P. SCHUTZENBERGER, *Sur un langage équivalent au langage de Dyck*, in Logic, Methodology and Philosophy of Sciences, vol. 4, 1973, North Holland, p. 197-203.