Revue française d'automatique informatique recherche opérationnelle. Informatique théorique

Bruno Courcelle

Une forme canonique pour les grammaires simples déterministes

Revue française d'automatique informatique recherche opérationnelle. Informatique théorique, tome 8, n° R1 (1974), p. 19-36

<http://www.numdam.org/item?id=ITA_1974__8_1_19_0>

R.A.I.R.O.
(8e année, R-1, 1974, p. 19-36)

UNE FORME CANONIQUE POUR LES GRAMMAIRES SIMPLES DETERMINISTES

par Bruno COURCELLE

Communiqué par M. NIVAT

Résumé. — *Cet article reprend l'étude des langages simples déterministes, introduits par Hopcroft et Korenjak. Voici les résultats essentiels : tout langage simple, défini par une grammaire, peut être effectivement décomposé en un produit de langages simples atomiques (c'est-à-dire de langages non produit de deux langages simples) ; tout langage simple est engendré par une grammaire simple particulière, qui ne dépend que du langage et que l'on peut construire. C'est l'analogue de l'automate minimal d'un langage rationnel.*

INTRODUCTION

Hopcroft et Korenjak ont introduit une classe particulière de grammaires de Chomsky, les grammaires simples dont l'équivalence est décidable [2], [1]. Nous reprenons l'étude des grammaires simples équivalentes : si N est l'ensemble des non-terminaux d'une grammaire simple G, nous étudions la congruence sur N^* définie par $m \equiv m'$ ssi $L(G, m) = L(G, m')$. Nous en déduisons : 1) que le monoïde des langages simples est librement engendré par l'ensemble des langages simples atomiques (i.e. ceux qui ne sont pas le produit de deux langages simples), 2) une nouvelle présentation de l'algorithme de Hopcroft et Korenjak, 3) que deux grammaires simples équivalentes dérivent d'une unique grammaire simple, dite canonique, par une transformation d'une certaine classe $\mathcal{T}$ de transformations des grammaires de Chomsky qui conservent le langage engendré. Toutes les propriétés des grammaires et des langages simples que nous énonçons sont décidables. La grammaire canonique d'un langage simple peut être effectivement construite. Les notations sont pour la plupart empruntées à [3] où l'on trouvera aussi la démonstration des théorèmes que nous rappelons au début.

IRIA, Rocquencourt.

1. LE MONOIDE DES LANGAGES SIMPLES

Nous désignerons par Σ^* le monoïde libre engendré par Σ, par 1 son unité le mot vide, par Σ^+ l'ensemble des mots non vides de Σ^*. $|x|$ désignera la longueur d'un mot x de Σ^*.

Un langage sur Σ est une partie de Σ^*. Définissons sur l'ensemble des langages $\mathcal{L}(\Sigma)$ l'opération binaire produit : $LL' = \{ xy/x \in L \text{ et } y \in L \}$ qui fait de $\mathcal{L}(\Sigma)$ un monoïde dont l'unité est $\{ 1 \}$. La réunion de deux langages L et L' sera notée $L + L'$.

Soit Σ un alphabet fini et $N = \{ S_1, ..., S_n \}$ un ensemble de *variables* (représentant des langages sur Σ). Une *grammaire algébrique* $G(N, \Sigma)$ est un système d'équations algébriques i.e. de la forme

$$\begin{cases} S_1 = m_{1,1} + ... + m_{1,r_1} \\ \vdots \\ S_n = m_{n,1} + ... + m_{n,r_n} \end{cases}$$

où $m_{i,j}$ sont des monômes en les variables non commutatives de N et à coefficients dans Σ, soit encore : $m_{i,j} \in (\Sigma \cup N)^*$.

Une *solution* d'une grammaire algébrique est un n-uplet de langages sur Σ vérifiant les équations du système.

A la grammaire algébrique $G(N, \Sigma)$, associons les règles de réécriture $S_i \to m_{i,j}$ pour $1 \leqslant j \leqslant r_i$ qui en font une grammaire de Chomsky habituelle. Nous désignerons par $\overset{*}{\underset{G}{\to}}$ la *dérivation*, clôture réflexive et transitive de la relation $m \underset{G}{\to} m'$ ssi $m = uS_iv$ et $m' = um_{i,j}v$ et par $L(G, m)$ pour $m \in (\Sigma \cup N)^*$ le langage $\{ w \in \Sigma^*/m \overset{*}{\underset{G}{\to}} w \}$ qui est dit *engendré par m.*

Une grammaire algébrique est dite *propre* si $r_i \geqslant 1$ pour tout i et si pour tout j, $m_{i,j} \notin \{ 1, S_1, ..., S_n \}$.

Le théorème suivant, dû à Schützenberger [3], établit la relation entre les langages engendrés et les langages-solution d'une grammaire algébrique :

Théorème : *Une grammaire algébrique propre,* $G(N, \Sigma)$ *admet une solution et une seule. C'est le n-uplet* $\langle L(G, S_1), ..., L(G, S_n) \rangle$.

Nous écrivons $m \underset{G}{\equiv} m'$ pour $m, m' \in (\Sigma \cup N)^*$ si $L(G, m) = L(G, m')$.

Corollaire : *Soient* $G(N, \Sigma)$ et $G'(N, \Sigma)$, *deux grammaires algébriques propres telles que pour tout i et j(j') il existe j'(j) tel que* $m_{i,j} \equiv {}_G m'_{i,j'}$. *Alors pour tout* $S \in N$, $L(G, S) = L(G', S)$.

Démonstration : Puisque G' est propre, il suffit de vérifier que $\langle L(G, S) \rangle_{S \in N}$ satisfait les équations du systèmes G'. Or, si $1 \leqslant i \leqslant n$,
$L(G, S_i) = \bigcup_{j=1}^{ri} L(G, m_{i,j}) = \bigcup_{=1}^{r'i} L(G, m'_{i,j'})$ d'après l'hypothèse. Si

$m'_{i,j'} = \alpha_1 S_{i1} \alpha_2 \ldots S_{ik} \alpha_{k+1}$ alors $L(G, m'_{i,j'}) = \alpha_1 L(G, S_{i1})\, \alpha_2 \ldots L(G, S_{ik}) \alpha_{i_{k+1}}$

et donc $\langle L(G, S_j) \rangle_{1 \leqslant j \leqslant n}$ satisfait la i-ème équation du système G'. C.Q.F.D.

Le théorème de Schützenberger peut être renforcé ainsi :

Soit $G(N, \Sigma)$ un système algébrique propre, $\bar{G}$ un système algébrique à coefficients dans $\mathfrak{L}(\Sigma)$ obtenu en remplaçant dans les membres de droite des équations de G certaines occurrences des variables par le langage $L(G, S_i)$ correspondant. $\langle L(G, S_i) \rangle_{1 \leqslant i \leqslant n}$ est l'unique solution du système $\bar{G}$.

REMARQUE : Les langages solution des grammaires algébriques sont exactement les langages « context-free » ne contenant pas le mot vide. Nous appellerons ces langages *algébriques*. Nous appellerons encore *non-terminaux* les variables d'une grammaire algébrique.

Une grammaire algébrique est dite *réduite* si pour tout $S \in N, L(G, S) \neq \emptyset$. Toutes les grammaires que nous allons considérer seront propres et réduites, sans que nous devions le préciser.

Une grammaire algébrique $G(N, \Sigma)$ est dite en *forme normale de Greibach* si toute équation est de la forme $S_i = a_1 m_{i,1} + \ldots + a_{ri} m_{\,i,ri}$ où $m_{i,j} \in N^*$ et $a_j \in \Sigma$. On notera $M_{S_i,a}$ l'ensemble des $m_{i,j}$ tels que $a_j = a$.

Une grammaire en forme normale de Greibach est dite *simple* si pour tout $a \in \Sigma$ et $S \in N$, $M_{S,a}$ contient au plus un élément, que l'on notera alors $m_{S,a}$.

$L \subset \Sigma^*$ est un *langage simple* si il existe une grammaire simple $G(N, \Sigma)$ et $m \in N^+$ tel que $L(G, m) = L$.

Pour tout $L \in L(\Sigma)$ et $u \in \Sigma^*$, soit $L/u = \{ w \in \Sigma^* |\, uw \in L \}$.

Lemme 1 : *Si $G(N, \Sigma)$ est une grammaire simple, $uv \in \Sigma^*$ et $m_0 \in N^*$, $uv \in L(G, m_0)$ ssi il existe $m \in N^*$ tel que $m_0 \xrightarrow{*} um$ et $m \xrightarrow{*} v$. S'il existe un tel m, il est unique et $L(G, m_0) \,|u = L(G, m)$.*

Démonstration : Il est clair que la condition est suffisante. Montrons, par récurrence sur la longueur de u l'existence et l'unicité de m.

Si $|u| = 0$ c'est trivial.

Si $u = u_1 a$ et $a \in \Sigma$ alors par hypothèse il existe un et un seul $Sm_1 \in N^+$ tel que $m_0 \xrightarrow{*} u_1 S m_1$ et $Sm_1 \xrightarrow{*} av$; cette dernière condition n'est réalisée que si il existe m_2 tel que $S \to am_2$ et $m_2 m_1 \xrightarrow{*} v$. Donc

$$m_0 \xrightarrow{*} u_1 a m_2 m_1 \text{ et } m_2 m_1 \xrightarrow{*} v.$$

L'unicité résulte de la définition d'une grammaire simple.

Il est clair alors que $L(G, m_0) \mid u = L(G, m)$.

Il en résulte en particulier que les *langages simples sont préfixes* (i.e. si $u \in L$ et $|v| \neq 0, uv \notin L$).

Soit $\mathcal{M}(\Sigma)$ l'ensemble de langages simples sur l'alphabet Σ, augmenté du langage $\{1\}$ (qui est en fait $L(G, 1)$). La proposition suivante montre que $\mathcal{M}(\Sigma)$ est un *monoïde* dont l'unité est $\{1\}$.

Proposition 1 : *Le produit de deux langages simples est un langage simple.*

Démonstration : Soient L et L' deux langages simples, soient $G(N, \Sigma)$ et $G(N', \Sigma')$ deux grammaires simples, $m \in N^*$ et $m' \in N'^*$ tels que $N \cap N' = \emptyset$, $L = L(G, m)$ et $L' = L(G', m')$. La réunion de G et de G' est une grammaire simple $G''(N \cup N', \Sigma \cup \Sigma')$ et $LL' = L(G'', mm')$.

REMARQUE : Si u est facteur gauche d'un mot de L et si pour $m_1 \in N^*$, $m \xrightarrow{*} um_1$ alors $(LL') \mid u = L(G'', m_1 m') = L(G'', m_1)L(G''m') = (L \mid u)L'$.

On appelle *valuation* sur un monoïde M un homomorphisme $\tau : M \to \mathbf{N}$, le monoïde additif des entiers naturels, tel que si $m \neq 1$ alors $\tau(m) \neq 0$.

$\mathcal{M}(\Sigma)$ est un monoïde valué par $\tau(L) = \inf \{ |w| / w \in L \}$. Il est clair en effet que $\tau(LL') = \tau(L) + \tau(L')$ et que si L est un langage simple (différent de $\{1\}$) il ne contient pas 1, le mot vide.

Nous allons étudier maintenant les propriétés du produit dans $\mathcal{M}(\Sigma)$.

Proposition 2 : *Le produit est une loi simplifiable à droite et à gauche, i.e. si $LL' = LL''$ ou $L'L = L''L$ alors $L' = L''$.*

Démonstration : Soient L, L', L'' des langages simples.

Si $LL' = LL''$ et si $u \in L$ alors $L' = (LL') \mid u = (LL'') \mid u = L''$.

Si $L'L = L''L$ et $u \in L'$ mais $u \notin L''$ (et donc $L' \neq L''$),

- ou bien il existe $w \neq 1$ tel que $uw \in L''$ mais alors

$$L = (L'L) \mid u = (L''L) \mid u = (L'' \mid u)L \qquad \text{et} \qquad \tau(L'' \mid u) \neq 0$$

car $w \neq 1$, on ne peut donc pas avoir $\tau(L) = \tau(L'' \mid u) + \tau(L)$ et ce cas est exclus ;

- ou bien $u = vw$, $v \in L''$ et est facteur gauche propre d'un mot de L'(i.e. $w \neq 1$). Alors $L = (L' \mid v)L$ et $\tau(L' \mid v) \neq 0$. On ne peut donc avoir $\tau(L) = \tau(L' \mid v) + \tau(L)$ et ce cas est exclus. Il en résulte que $L' = L''$. CQFD.

Un monoïde M est dit *équidivisible* s'il vérifie la condition suivante : pour tous $m_1, m_2, m_3, m_4 \in M$ tels que $m_1 m_2 = m_3 m_4$, il existe $m_5 \in M$ tel que : ou bien $m_1 = m_3 m_5$ et $m_5 m_2 = m_4$, ou bien $m_1 m_5 = m_3$ et $m_2 = m_5 m_4$.

Proposition 3 : *$\mathcal{M}(\Sigma)$ est équidivisible.*

Démonstration : Pour tout $L \subset \Sigma^*$ soit $\sigma(L)$ l'ensemble des mots de L de longueur minimale (i.e. égale à $\tau(L)$). Il est clair que $\sigma(LL') = \sigma(L)\sigma(L')$ Soient L_1, L_2, L_3, L_4 des langages simples tels que $L_1L_2 = L_3L_4$, tous différents de $\{1\}$ (sinon la démonstration est finie) ; supposons de plus que $\tau(L_1) \leqslant \tau(L_3)$. Alors $\sigma(L_1)\sigma(L_2) = \sigma(L_3)\sigma(L_4)$. Soit $u \in \sigma(L_1)$. C'est un facteur gauche d'un mot de L_3. Puisque $u \in L_1$ et grâce à la remarque $L_2 = (L_1L_2) \mid u = (L_3L_4) \mid u = (L_3 \mid u)L_4$.

Par le lemme 1, $L_3 \mid u \in M(\Sigma)$ et $L_1(L_3 \mid u)L_4 = L_3L_4$; la proposition 2 montre alors que $L_1(L_3 \mid u) = L_3$.

On appelle *atome* d'un monoïde M un élément qui n'est pas le produit de deux éléments de M différents de l'unité. Nous appellerons *langages atomiques* les atomes de $\mathcal{M}(\Sigma)$.

Le théorème suivant est dû à Levy.

Théorème : *Un monoïde équidivisible valué est librement engendré par l'ensemble de ses atomes.*

Démonstration : M est engendré par ses atomes : remarquons tout d'abord que si $\tau(m) = 1$ alors m est un atome. Cela résulte de ce que

$$\tau(m_1m_2) = \tau(m_1) + \tau(m_2) \geqslant 2$$

si m_1 et m_2 sont différents de 1. Soit $m \in M$ de valeur $\tau(m)$ minimale s'il en existe qui n'est pas un atome ni un produit d'atomes. Puisque m n'est pas un atome, $m = m_1m_2$, $\tau(m_1) < \tau(m)$ et $\tau(m_2) < \tau(m)$ donc m_1 et m_2 sont des produits d'atomes par hypothèse et il en est de même de m contrairement à l'hypothèse.

Il reste à montrer que si les a_i, et les b_i sont des atomes et si $a_1 \dots a_k = b_1 \dots b_l$ alors $k = l$ et $a_i = b_i$ pour $1 \leqslant i \leqslant k$. L'équidivisibilité de M implique qu'il existe c tel que $a_1 = b_1c$ ou $a_1c = b_1$. Puisque a_1 et b_1 sont des atomes, $c = 1$ et donc $a_1 = b_1$. Par suite $a_2 \dots a_k = b_2 \dots b_l$. Le même argument répété montre que $(a_i)_{1 \leqslant i \leqslant k} = (b_j)_{1 \leqslant j \leqslant l}$.

Nous pouvons reformuler ce théorème qui vaut également pour les langages préfixes :

Théorème 1 : *Tout langage simple est le produit d'une manière unique d'une suite finie de langages simples atomiques.*

2. EQUIVALENCE DES GRAMMAIRES SIMPLES

Soit $G(N, \Sigma)$ une grammaire simple.

Soit ψ l'homomorphisme $N^* \rightarrow \mathcal{M}(\Sigma)$ défini par $\psi(m) = L(G, m)$.

La communauté des images définit une *congruence* notée $\equiv$ (i.e. $m \equiv m'$ ssi $L(G, m) = L(G, m')$).

N^* est valué par $\tau \circ \psi$ (que nous noterons encore τ pour simplifier). La congruence $\equiv$ sur N^* respecte donc la valuation τ sur N^* (i.e. $m \equiv m'$ implique $\tau(m) = \tau(m')$). Nous appellerons $R(\tau)$ la congruence sur N^* définie par $\tau(m) = \tau(m')$.

Nous allons démontrer le résultat suivant :

Théorème 2 : *La congruence $\equiv$ sur N^* est décidable.*

Rappelons tout d'abord quelques propriétés des congruences engendrées sur un monoïde M.

Soit M un monoïde et A une partie de $M \times M$. $\bar{A}$ désignera sa fermeture réflexive et symétrique et $\tilde{A}$ sa fermeture transitive.

Si $B \subset M \times M$, $A \cdot B$ désignera $\{(m_1 m_2, m_3 m_4) \mid (m_1, m_3) \in A$ et $(m_2, m_4) \in B\}$. Soit A_i la suite croissante de relations sur M définie par :

$$\begin{aligned} A_1 &= \bar{A} \\ &\vdots \\ A_{n+1} &= \widetilde{A_n \cdot A_n} \\ &\vdots \\ \text{et soit } \hat{A} &= \bigcup_{n \geq 1} A_n \end{aligned}$$

Proposition 4 :

(i) *$\hat{A}$ est la congruence la plus grossière sur M contenant A.*

(ii) *Si $M = N^*$ est valué par τ et $A \subset R(\tau)$ alors $\hat{A} \subset R(\tau)$.*

(iii) *Si de plus A est fini, $\hat{A}$ est décidable* (i.e *pour chaque $(m, m') \in N^*$ on peut décider si $(m, m') \in \hat{A}$).*

Démonstration : (i) et (ii) sont laissés au lecteur. Prouvons (iii).

Montrons que pour tout $n \geq 0$ on peut calculer l'ensemble fini des $(m, m') \in \hat{A}$ tels que $\tau(m) = \tau(m') = n$.

Pour tout n, soit $B_n = \{ (m, m') \mid \tau(m) = \tau(m') = n$ et $(m, m') \in \bar{A}\}$. Définissons par induction la suite des $C_n \subset N^* \times N^*$:

$$C_0 = B_0 = \{ (1, 1) \}$$

$$\vdots$$

$$C_{n+1} = (B_{n+1} \cup \bigcup \{ C_m C_{m'} | m + m' = n + 1, m \text{ et } m' \geqslant 1 \})$$

$$C_\infty = \bigcup_{n \geqslant 0} C_n$$

Il résulte de cette construction que :

(i) $A \subset C_\infty$

(ii) pour tout $m \in N^*$, $(m, m) \in C_{\tau(m)}$

(iii) pour tout n, C_n est symétrique. (C'est clair pour $n = 0$; si c'est vrai jusqu'à n, $C_m C_{m'}$ est symétrique pour m et $m' \leqslant n$, ainsi que B_{n+1} ; il en est de même de $B_{n+1} \cup C_m C_{m'}$ et donc de sa clôture transitive, soit C_{n+1}),

(iv) pour tout n, C_n est transitive et donc C_∞ est une relation d'équivalence.

(v) pour tout n et n', $C_n C_{n'} \subset C_{n+n'}$ et donc C_∞ est une congruence, et donc avec (i) : $\hat{A} \subset C_\infty$,

(vi) pour tout n, $C_n \subset A_n$ et donc $\hat{A} = C_\infty$.

Chaque B_n et chaque C_n est fini et calculable et $(m, m') \in \hat{A}$ ssi $\tau(m) = \tau(m')$ et $(m, m') \in C_{\tau(m)}$. Donc $\hat{A}$ est décidable.

Revenons à la congruence $\equiv$ sur N^* définie par une grammaire simple $G(N, \Sigma)$; soit $\xi = \{ (S, m)/S \in N, m \in N^*$ et $S \equiv {}_G m \}$.

Proposition 5 : *La congruence $\equiv_G$ est engendrée par ξ.*

La démonstration résulte par une récurrence simple du corollaire suivant de la proposition 3 :

Corollaire 1 : *Le sous-monoïde de $\mathcal{M}(\Sigma)$, image par ψ de N^* est équidivisible*, que l'on peut formuler de la manière suivante :

Corollaire 1 *bis* : *Si $Sm \equiv S'm'$ pour $S, S' \in N$ et $m, m' \in N^*$*

alors (i) *si $\tau(S) = \tau(S')$*, alors $S \equiv S'$ *et* $m \equiv m'$,

(ii) *si $\tau(S) < \tau(S')$, il existe $m'' \in N^+$ tel que $S' \equiv Sm''$ et $m \equiv m''m'$.*

Démonstration : Il suffit de reprendre la démonstration de la proposition 3 avec $L_1 = L(G, S)$, $L_2 = L(G, m)$, $L_3 = L(G, S')$, $L_4 = L(G, m')$. Si $\tau(S) < \tau(S')$, $L_3 = L(G, m'')$, $m'' \in N^+$ (lemme 1) et si $\tau(S) = \tau(S')$ alors $L_3/u = \{ 1 \}$.

Il reste à montrer que l'on peut calculer ξ. Nous allons définir un ensemble fini et calculable de relations finies sur N^+, qui seront contenues dans ξ et dont ξ sera la plus grande : une partie A de $N^* \times N^*$ est *consistante* si :

(i) $A \subset \{ (S, m)/S \in N, m \in N^*, \tau(S) = \tau(m) \}$,

(ii) si $(S, S'm') \in A$ et $a \in \Sigma$ alors $m_{S,a}$ et $m_{S',a}$ sont simultanément définis et $(m_{S,a}, m_{S',a}m') \in \hat{A}$ }.

Proposition 6 : *Si A est consistante et $(m, m') \in \hat{A}$, alors $m \equiv m'$.*

Démonstration : En vue de cette preuve, introduisons les relations d'équivalence suivantes sur N : pour s entier positif, m et $m' \in N^*$, $m \equiv {}_s m'$ ssi les mots de longueur inférieure ou égale à s de $L(G, m)$ et $L(G, m')$ sont les mêmes.

Soit s_0 le plus petit entier s'il en existe, tel qu'il existe n_0, m et m' vérifiant $(m, m') \in C_{n_0}$ (notations de la proposition 4) et $m \not\equiv {}_{s_0} m'$.

Il existe dans $D = B_{n_0} \cup \bigcup_{n+n'=n_0} C_n C_{n'}$ un tel couple (m, m'). Sinon, puisque C_{n0} est la clôture transitive de D et que $\equiv {}_{s_0}$ est transitive, tous les éléments de C_{n0} seraient équivalents modulo $\equiv {}_{s_0}$. Alors :

- ou bien $(m, m') \in C_n C_{n'}$, et $n + n' = n_0$, et donc

$$m = m_1 m_2, m' = m'_1 m'_2, (m_1, m'_1) \in C_n \text{ et } (m_2, m'_2) \in C_{n'}.$$

A cause du choix de s_0 que nous avons fait, $m_1 \equiv {}_{s_0-1} m'_1$ et $m_2 \equiv {}_{s_0-1} m'_2$. Il est clair alors que $m_1 m_2 \equiv {}_{s_0} m'_1 m'_2$ contrairement à l'hypothèse;

- ou bien $(m, m') \in B_{n_0}$. On ne peut pas avoir $m' = m$. Donc $(m, m') = (S, S'm_1)$, soit $\alpha \in \Sigma^*$ de longueur s_0 tel que

$$\alpha \in L(G, S) \text{ ssi } \alpha \notin L(G, S'm_1).$$

Distinguons encore deux cas :

- ou bien $|\alpha| = s_0 = 1, \alpha = a \in \Sigma$. Alors $a \in L(G, S)$ et $a \notin L(G, S'm_1)$ Puisque $\hat{A}$ est consistante, $(m_{S,a}, m_{S',a}m_1) \in \hat{A}$. Mais par ailleurs $m_{s,a} = 1$. et $m_{S,a}m_1 \neq 1$. Donc $\hat{A} \not\subset R(\tau)$ contrairement à la proposition 4 (ii) ;
- ou bien $\alpha = a\beta$, $(m_{s,a}, m_{s',a}m_1) \in \hat{A}$ et

$$\beta \in L(G, m_{s,a}) \text{ ssi } \beta \notin L(G, m_{s',a}m_1).$$

Puisque $|\beta| = s_0 - 1$, ceci contredit la définition de s_0. CQFD.

Théorème 2 : *$m \equiv m'$ ssi il existe une partie consistante de $N^* \times N^*$ telle que $(m, m') \in \hat{A}$. La congruence $\equiv$ est décidable.*

Démonstration : La condition est suffisante (proposition 6). Il est clair que ξ est consistante et la condition est nécessaire (proposition 5). La condition (i) de la définition des parties consistantes entraîne qu'elles sont en nombre fini et la condition (ii) est décidable (proposition 4 (iii)). On peut donc énumérer les parties consistantes A de $N^* \times N^*$ et décider pour chacune d'elles si $(m, m') \in \hat{A}$. La congruence $\equiv$ est donc décidable.

Cet algorithme grossier permet de décider l'équivalence des grammaires simples : soient $G_1(N_1, \Sigma, \sigma_1)$ et $G_2(N_2, \Sigma, \sigma_2)$. Quitte à remplacer N_2 par un ensemble équipotent, supposons que $N_1 \cap N_2 = \emptyset$. Soit alors

$$G_3(N_1 \cup N_2, \Sigma) = G_1(N_1, \Sigma) \cup G_2(N_2, \Sigma).$$

On peut décider si $\sigma_1 \equiv_{G_3} \sigma_2$ (i.e si $L(G_1, \sigma_1) = L(G_2, \sigma_2)$), grâce au théorème 2.

Nous allons décrire un algorithme efficace pour décider si $m \equiv m'$ pour tout couple $(m, m') \in N^* \times N^*$. Il est dû à Hopcroft et Korenjak [2].

Lemme : *Soit $G(N, \Sigma)$ une grammaire simple. Il existe un algorithme se terminant toujours qui, avec $(m, m') \in N^* \times N^*$ pour donnée, construit $F(m, m') \subset N \times N^*$ ou échoue à construire cet ensemble et tel que :*

$$P \begin{cases} \text{(i) } \textit{si la construction échoue, } m \not\equiv m', \\ \text{(ii) } \textit{sinon, } (m, m') \in \hat{F}(m, m') \text{ et } (m \equiv m' \text{ ssi } F(m, m') \subset \equiv). \end{cases}$$

Ce lemme est une version « effective » de la proposition 5.

Démonstration : Définissons récursivement cet algorithme par les règles suivantes :

1) si $\tau(m) \neq \tau(m')$ alors la construction échoue,

2) $F(1, 1) = \emptyset$,

3) $F(Sm_1, Sm'_1) = F(m_1, m'_1)$,

4) $F(Sm_1, S'm'_1)$ avec $\tau(S) \leqslant (S')$. Soit $u \in L(G, S)$ de longueur $\tau(S)$. Calculons $m_2 \in N^*$ tel que $S' \xrightarrow[G]{*} um_2$. S'il n'en existe pas, la construction échoue.

Si $(S', Sm_2) \in \hat{F}(m_1, m_2m'_1)$ alors $F(Sm_1, S'm'_1) = F(m_1, m_2m'_1)$. Sinon $F(Sm_1, S'm'_1) = \{ (S', Sm_2) \} \cup F(m_1, m_2m'_1)$,

5) Si $\tau(S) > \tau(S')$, $F(Sm_1, S'm'_1) = F(S'm'_1, Sm_1)$.

Cet algorithme se termine toujours : en effet, ou bien la construction de $F(m, m')$ échoue, ou bien elle utilise $F(m_1, m'')$ avec $\tau(m_1) < \tau(m)$. La suite des appels de la procédure F est donc finie.

Montrons que pour tout $(m, m') \in N^* \times N^*$, la condition P est remplie : si $\tau(m) \neq \tau(m')$ alors $m \not\equiv m'$ et la construction échoue donc P est vraie, sinon, montrons-le par récurrence sur $n = \tau(m) = \tau(m')$.

Si $n = 0$, $(m, m') = (1, 1)$ et P est vraie.

Si $(m, m') = (Sm_1, Sm_1')$ et $P(m_1, m_1')$ est vraie alors si la construction de $F(m, m')$ échoue, c'est parce que celle de (m_1, m_1') échoue, donc

$$m_1 \not\equiv m_1' \text{ et } Sm_1 \not\equiv Sm_1';$$

si elle réussit, $(Sm_1, Sm_1') \in \hat{F}(m_1, m_1') = \hat{F}(m, m')$ car $(m_1, m_1) \in \hat{F}(m_1, m_1)$; d'autre part, $m \equiv m'$ ssi $m_1 \equiv m_1{}'$ ssi $F(m_1, m_1{}') \subset \equiv$ ssi $F(m, m') \subset \equiv$ puisque $F(m, m') = F(m_1, m_1{}')$.

Si $(m, m') = (Sm_1, S'm_1')$ et $\tau(S) \leqslant \tau(S')$, $m \equiv m'$ ssi il existe $u \in \sigma(L(G, S))$ et $m_2 \in N^*$ tels que $S' \xrightarrow{*} um_2$, $S' \equiv Sm_2$ et $m_1 \equiv m_2 m_1'$. P en résulte.

Si $(m, m') = (Sm_1, S'm_1')$ et $\tau(S) > \tau(S')$, P est vraie pour $(Sm_1, S'm_1')$ ssi elle est vraie pour $(S'm_1', Sm_1)$ et on est ramené au cas précédent. CQFD.

Nous pouvons maintenant écrire l'algorithme :

Étant donnée une grammaire simple $G(N, \Sigma)$, $(m_0, m_0') \in N^* \times N^*$, construisons la suite croissante de parties de $N \times N^*$ définie par :

(i) $\qquad E_0 = F(m_0, m_0')$

(ii) $\qquad E_{i+1} = E_i \cup \bigcup \{ F(m_{S,a}, m_{S',a}m)/a \in \Sigma \text{ et } (S, S'm) \in E_i \}$

L'algorithme s'arrête dès que, pour un $i \geqslant 0$:
ou bien il existe (S, m') dans E_i et $\tau(S) \neq \tau(m')$,
ou bien la construction de E_i échoue et *il répond* $m_0 \not\equiv m_0'$ dans ces deux cas,
ou bien $E_{i+1} = E_i$ et il *répond* $m_0 \equiv m_0'$.

Proposition : *L'algorithme ci-dessus s'arrête toujours. Il est correct ; si $m_0 \equiv m_0'$ et $E_{i+1} = E_i$ alors E_i est consistante et $(m_0, m_0') \in \hat{E}_i$.*

Démonstration : L'algorithme termine toujours car l'ensemble des parties E de $N \times N^*$ contenues dans $R(\tau)$ est fini. Si l'algorithme s'arrête et dit que $m_0 \equiv m_0'$ alors il existe i tel que $E_{i+1} = E_i$ et, puisque la construction n'a pas échoué, $E_i \subset R(\tau)$ et pour tous a et $(S, S'm') \in E_i$, $m_{S,a}$ et $m_{S',a}$ sont simultanément définis ; $F(m_{S,a}, m_{S',a}m') \subset E_{i+1} = E_i$, donc $(m_{S,a}, m_{S',a}m') \in \hat{E}_i$.

Nous venons de vérifier que E_i est consistante. $(m_0, m_0') \in \hat{E}_0 \subset \hat{E}_i$ et donc $m_0 \equiv m_0'$ (proposition 6). Si $m_0 \equiv m_0'$ montrons par récurrence que la construction de E_i ne peut pas échouer et que $E_i \subset \equiv$. Cela résulte du lemme pour E_0. Si $S \equiv S'm$ et $a \in \Sigma$, $m_{S,a}$ et $m_{S',a}$ sont simultanément définis, et alors $m_{S,a} \equiv m_{S',a}m$. La construction de E_{i+1} ne peut donc pas échouer et $E_{i+1} \subset \equiv$ (lemme). L'algorithme ne peut donc pas répondre $m_0 \not\equiv m_0'$.

CQFD.

Exemple (tiré de [2]) :

$\Sigma = \{ a, b \}, N = \{ A, B, C, D, E, F, G \}.$

$$G \begin{cases} A = bAB + aB \\ B = b \\ C = a \\ D = bDF + a \\ E = bG \\ F = b \\ G = a \end{cases}$$

Soit à décider si $AC \equiv DE$.

1) $\tau(AC) = \tau(DE)$

2) $E_0 = F(AC, DE) = F(DE, AC) = \{ (A, DB) \} \cup F(E, BC)$
$= \{ (A, DB), (E, BC) \}$ (car $D \rightarrow a$ et $A \rightarrow aB$).

3) $E_0 \subset R(\tau)$; il faut continuer et construire E_1.

4) $E_1 = E_0 \cup F(AB, DFB) \cup F(B, B) \cup F(G, C) = E_0 \cup \{ (A, DF), (G, C) \}$

5) $E_1 \subset R(\tau)$.

6) $E_2 = E_1 \cup F(B, F) \cup F(AB, DFF) \cup F(1, 1) = E_1 \cup \{ (B, F) \}$.

7) $E_3 = E_2 \cup F(B, F) = E_2$.

Nous venons de prouver que $AC \equiv DE$.

3. GRAMMAIRES SIMPLES ENGENDRANT UN LANGAGE DONNE

Une *grammaire algébrique* (resp. *simple*) *avec axiome* est la donnée $G(N, \Sigma, \sigma)$ d'une grammaire algébrique (resp. simple) $G(N, \Sigma)$ et d'un *axiome* $\sigma \in N^*$. Le *langage engendré* est $L(G) = L(G, \sigma)$. Deux grammaires sont *équivalentes* si elles engendrent le même langage. On dit que G est *réduite par rapport à* σ, si pour tout $S \in N$, $L(G, S) \neq \varnothing$ et il existe u et $v \in \Sigma^*$ tels que $\sigma \xrightarrow[G]{*} uSv$.

Nous allons définir une classe de transformations des grammaires algébriques conservant le langage engendré. Les grammaires simples engendrant un langage donné sont les images par ces transformations d'une unique grammaire simple appelée *grammaire canonique* du langage. On peut construire effectivement cette grammaire et décider si une grammaire donnée dérive de la grammaire canonique.

Définition : Soient $G(N, \Sigma)$ et $G(N', \Sigma)$ deux grammaires algébriques en forme normale de Greibach et l un homomorphisme de $N'^* \rightarrow N^*$. Nous dirons que *G' dérive de G par l,* ce que nous noterons $G \Rightarrow {}_l G'$, si

1. l est continu (i.e $l(S') \in N^+$ pour tout $S' \in N'$),

2. pour tous $a \in \Sigma$, $S \in N$, $S' \in N'$ et $m_1 \in N^*$ tels que $l(S') = Sm_1$, $l(M_{S'a}) = M_{S,a}m_1$ [1]. Pour deux grammaires avec axiome, $G(N, \Sigma, \sigma)$ et $G(N', \Sigma, \sigma')$ on ajoutera la condition :

3. $l(\sigma') = \sigma$.

L'exemple suivant éclaire la signification de la relation $G \Rightarrow {}_l G'$: soit G_1 la grammaire (simple) :

$G_1 : S = aSSS + b$. Soit $N_2 = \{ T, U, V \}$ un nouvel ensemble de non-terminaux.

Nous allons chercher une grammaire $G_2(N_2, \Sigma)$ telle que

$$L(G_2, T) = L(G_1, SS), L(G_2, U) = L(G_1,S), L(G_2, V) = L(G_1, S),$$

les langages engendrés par S, T, U, V doivent vérifier :

$$G_1' \left\{ \begin{array}{l} S = a\,SSS + b \\ T = a\,SSSS + bS \\ U = a\,SSS + b \\ V = a\,SSS + b \end{array} \right.$$

Appliquons le corollaire du théorème de Schützenberger : T, U, V peuvent être définis par :

$$\begin{array}{l} S = a\,SSS + b \\ G_2 \left\{ \begin{array}{l} T = a\,TT + bU \\ U = a\,VT + b \\ V = a\,VUV + b \end{array} \right. \end{array}$$

Cette nouvelle grammaire G_2 dérive de G_1 par $l : N_2^* \rightarrow N_1^*$ défini par $l(T) = SS$, $l(U) = l(V) = S$. Plus généralement, si $G \Rightarrow {}_l G'$, on peut construire G' à partir de G de manière semblable.

Proposition 7 : *Si $G'(N', \Sigma, \sigma')$ dérive de $G(N, \Sigma, \sigma)$ par l alors pour tout $S' \in N'$, $L(G', S') = L(G, l(S'))$. G et G' sont équivalentes.*

---

(1) Rappelons que $M_{S,a} = \{ m \in N^*/S \rightarrow {}_{am} \}$.

Démonstration : Puisque G' est en forme normale de Greibach, elle est propre. D'après le théorème de Schützenberger, il suffit de vérifier que $\langle L(G, l(S')) \rangle_{S' \in N'}$ est une solution du système G'.

Soit $S' \in N'$ et $S' = \sum_{i=1}^{l} a_i m'_i$ l'équation associée et $l(S') = Sm''$. L'équation de G associée à S est $S = \sum_{i=1}^{l} a_i m_i$ et pour tout i, $l(m'_i) = m_i m''$. Or $L(G, S) = \bigcup_{i=1}^{l} a_i L(G, m_i)$. Donc

$$L(G, l(S')) = L(G, S)L(G, m'') = \bigcup_{i=1}^{l} a_i L(G, m_i m'') = \bigcup_{i=1}^{l} a_i L(G, l(m'))$$
$$= \bigcup_{i=1}^{l} a_i L(G, l(S'_{i,1})) \dots L(G, l(S'_{i,r_i}))$$

si pour tout $i, m'_i = S'_{i,1} \dots S'_{i,r_i}$. Ceci n'est autre que l'équation à vérifier.

REMARQUES : 1) Si $G \Rightarrow {}_l G'$ et l est un *isomorphisme* de N'^* sur N^*, les grammaires ne diffèrent que par le nom des non-terminaux. On les considérera comme identiques.

2) Si $G \Rightarrow {}_l G'$ et $G' \Rightarrow {}_{l'} G''$ alors $G \Rightarrow {}_{l''} G''$ avec $l'' = lol'$.

Une grammaire simple $G(N, \Sigma)$ (resp. avec axiome) est dite *canonique* si

(i) tout non-terminal S de N engendre un langage atomique,

(ii) pour tous S et S' dans N, $S \equiv S'$ implique $S = S'$.

Cette définition a la conséquence immédiate suivante : $m \equiv {}_G m' \Rightarrow m = m'$.

Théorème 3 : *Soit $G'(N', \Sigma, \sigma')$ une grammaire simple équivalente à une grammaire canonique $G(N, \Sigma, \sigma)$ et l la relation sur $N'^* \times N^*$ définie par $m'lm \Leftrightarrow L(G', m') = L(G, m)$. l est un homomorphisme $N'^* \to N^*$ et G' dérive de G par l. Si G' est canonique, l est un isomorphisme.*

Démonstration : l est une relation fonctionnelle car $m'lm$ et $m'lm''$ impliquent $m \equiv {}_G m''$ et, puisque G est canonique $m = m''$. C'est un morphisme pour le produit sur N'^* et N^*. Montrons que tout $S' \in N'$ a une image dans N^* : puisque G' est réduite par rapport à σ et d'après le lemme 1, il existe $u \in \Sigma^*$ et $m' \in N'^*$ tels que $\sigma' \xrightarrow[G']{*} uS'm'$. Puisque $L(G') = L(G)$, u est facteur gauche propre d'un mot de $L(G)$. Il existe donc $S_1 \dots S_k \in N^+$ tel que $\sigma \xrightarrow[G]{*} uS_1 \dots S_k$. Il en résulte que

$$L(G) \mid u = L(G, S_1) \dots L(G, S_k) = L(G', S')L(G', m').$$

Puisque les $L(G, S_i)$ sont atomiques, $L(G', S') = L(G, S_1) \dots L(G, S_r)$ avec $1 \leqslant r \leqslant k$. Donc $l(S') = S_1 \dots S_r$. Il est clair que $l(S') \neq 1$ et que $l(\sigma') = \sigma$ d'après les définitions. Il reste à montrer que si $l(S') = Sm_1$ et $a \in \Sigma$ alors

$$l(m_{S',a}) = m_{S,a} m_1.$$

Or

$$L(G', S') \mid a = L(G', m_{S',a})\ (^1)$$
$$= L(G, Sm_1) \mid a = (L(G, S) \mid a)L(G, m_1) = L(G, m_{S,a})L(G, m_1).$$

Autrement dit $l(m_{S',a}) = m_{S,a}m_1$. Si de plus G' est canonique, l est bijective et c'est un isomorphisme. CQFD.

Nous appellerons *taille* d'une grammaire simple $G(N, \Sigma)$ l'entier

$$\tau(G) = \sum_{S \in N} \tau(S) = \sum_{S \in N} \tau(L(G, S)).$$

Théorème 4 : *Il existe une grammaire canonique engendrant un langage simple donné : C'est la grammaire de taille minimum engendrant le langage.*

Soit G une grammaire, S et T deux non-terminaux. Nous dirons que *S est divisible par T* s'il existe un langage simple L_0 tel que $L(G, S) = L_0L(G, T)$. Il résulte de la proposition 2 que L_0 est unique; nous l'appellerons le quotient de S par T.

Lemme 2 : *Tout non-terminal non atomique est divisible par un non-terminal atomique de la même grammaire.*

Démonstration : Soit S un non-terminal non atomique, non divisible par un non-terminal atomique de G et de taille $\tau(S)$ minimale, s'il en existe. $L(G, S) = L_0L_1$; soit $u \in L_0$ et $S_1 \dots S_k \in N^*$, tels que $S \overset{*}{\underset{G}{\rightarrow}} uS_1 \dots S_k$. Alors $L_1 = (L_0L_1) \mid u = L(G, S) \mid u = L(G, S_1 \dots S_k)$. Il en résulte que

$$\tau(S_1) + \dots + \tau(S_k) = \tau(L_1) < \tau(S) \text{ et } \tau(S_k) < \tau(S).$$

Donc S_k est soit atomique, soit divisible par un atomique de G. Il en est donc de même de $L_1 = L(G, S_1 \dots S_k)$ et par suite de $L(G, S)$ contrairement à l'hypothèse initiale. CQFD.

Lemme 3 : *Si S est divisible par T, atomique et si $S \underset{G}{\rightarrow} aS_1 \dots S_k$ alors, ou bien $S_k \equiv T$ ou bien S_k est divisible par T.*

Démonstration : Soit $L(G, S) = L_0L(G, T)$ et $S \underset{G}{\rightarrow} aS_1 \dots S_k$.

$$L(G, S) \mid a = (L_0 \mid a)L(G, T) = L(G, S_1 \dots S_{k-1})L(G, S_k).$$

Puisque $L(G, T)$ est un langage atomique, il existe L_1 dans $\mathcal{M}(\Sigma)$ tel que $L(G, S_k) = L_1L(G, T)$. CQFD.

(1) Ceci contient aussi le cas où $L(G', S') \mid a = \emptyset$ i.e. $m_{S',a}$ n'est pas défini.

Proposition 8 : *Soit $G(N, \Sigma, \sigma)$ une grammaire simple non canonique. Il existe une grammaire simple $G_1(N_1, \Sigma, \sigma_1)$ équivalente et de taille strictement inférieure à $\tau(G)$.*

Démonstration : Pour tout non-terminal T soit $E(T) \subset N$ l'ensemble des non-terminaux équivalents à T et différents de T et $D(T)$ l'ensemble de ceux que T divise.

Si G n'est pas canonique, il existe au moins un non-terminal atomique T_0 tel que $E(T_0) \cup D(T_0) \neq \emptyset$ (lemme 2).

Nous appellerons N' l'ensemble des nouveaux symboles non-terminaux $\{ [S \mid T_0]/S \in D(T_0) \}$. Il résulte du lemme 3 que chaque équation de G associée à $S \in D(T_0)$ est de la forme :

$S = \sum_{i=1}^{l} a_i m_i S_i + \sum_{i=l+1}^{k} a_i m_i R_i$ où $S_i \in D(T_0)$, $R_i \in E(T_0) \cup \{ T_0 \}$, $m_i \in N^*$ et $a_i \in \Sigma$. Associons alors à $[S \mid T_0]$ l'équation

$$[S|T_0] = \sum_{i=1}^{l} a_i m_i [S_i|T_0] + \sum_{i=l+1}^{k} a_i m_i.$$

Ajoutées aux équations de $G(N, \Sigma)$ ces équations constituent une grammaire simple $G'(N \cup N', \Sigma, \sigma)$.

Il est clair que pour tout $S \in N$, $L(G', S) = L(G, S)$. Montrons que pour $S \in D(T_0)$, $L(G', S) = L(G', [S \mid T_0])L(G', T_0)$ en appliquant la forme forte du théorème de Schützenberger (voir § 1) : il suffit de vérifier que

$$\langle L(G', [S \mid T_0]).L(G', T_0) \rangle_{S \in D(T_0)}$$

satisfait les équations :

$S = \sum_{i=1}^{l} a_i \bar{m}_i S_i + \sum_{i=l+1}^{k} a_i \bar{m}_i \bar{R}_i$ (où $\bar{m}_i$ est le langage $L(G', m_i)$) pour chaque $S \in D(T_0)$.

Or c'est vrai d'après la forme des équations correspondant aux $[S \mid T_0]$ et le fait que $\bar{R}_i = L(G', T_0)$.

Les non-terminaux de G' vérifient donc les équations suivantes :

(i) $S \equiv [S \mid T_0]T_0$ pour tout $S \in D(T_0)$,

(ii) $S \equiv T_0$ pour tout $S \in E(T_0)$.

Remplaçons dans l'axiome σ et dans les membres de droite des équations de G', chaque $S \in D(T_0)$ par $[S \mid T_0]T_0$ et chaque $S \in E(T_0)$ par T_0. D'après le corollaire du théorème de Schützenberger, la grammaire G'' ainsi obtenue est équivalente à G' non-terminal à non-terminal i.e, pour tout $S \in N \cup N'$,

$L(G', S) = L(G'', S))$ et $L(G', \sigma) = L(G'', \sigma'')$. Puisque les non-terminaux $S \in D(T_0) \cup E(T_0)$ ne figurent pas dans les membres de droite des équations de G'' ni dans l'axiome σ'', la grammaire $G_1(N_1, \Sigma, \sigma'')$ où

$$N_1 = N \cup N' - (D(T_0) \cup E(T_0)),$$

obtenue en supprimant les équations correspondant aux variables de $D(T_0) \cup E(T_0)$ est une grammaire simple équivalente à G''. Calculons sa taille :

$$\tau(G_1) = \sum_{S \in N} \tau(S) + \sum_{S \in D(T_0)} \tau([S | T_0]) - \sum_{S \in D(T_0)} \tau(S) - \sum_{S \in E(T_0)} \tau(S)$$

(Dans cette équation, la taille d'un non-terminal est calculée par rapport à l'une quelconque des grammaires G, G', G'', G_1 dont il dépend. Ceci est correct car les langages engendrés sont les mêmes lorsqu'un non-terminal dépend de deux grammaires.) D'autre part, $\tau([S \mid T_0]) = \tau(S) - \tau(T_0)$ pour tout $S \in D(T_0)$. Donc $\tau(G_1) = \tau(G) - \tau(T_0)(|E(T_0)| + |D(T_0)|)$. Nous avions supposé au départ que $D(T_0) \cup E(T_0) \neq \varnothing$, donc $\tau(G_1) < \tau(G)$. CQFD.

Démonstration du théorème 4 : Soit $G(N, \Sigma, \sigma)$ une grammaire simple non canonique.

Il existe alors (proposition 8) une suite de grammaires G_1, G_2 ... équivalentes à G et dont les tailles décroissent strictement. Puisque la taille est supérieure ou égale à 1, la suite G_i est finie et la dernière grammaire est canonique. La grammaire canonique d'un langage simple est donc la grammaire de taille minimale engendrant ce langage.

Théorème 5 :

(i) *Étant données deux grammaires en F.N.G., G et G', on peut calculer l'ensemble fini des $l : N'^* \to N^*$ tels que $G \Rightarrow {}_l G'$.*

(ii) *Étant donnée une grammaire simple $G(N, \Sigma, \sigma)$, on peut construire la grammaire canonique équivalente à G.*

Démonstration :

(i) Si l est tel que $G \Rightarrow {}_l G'$, alors (proposition 7) $L(G', S') = L(G, l(S'))$, pour tout $S' \in N'$ et donc $\tau(S') = \tau(l(S'))$. Puisque pour tout $S \in N$, $\tau(S) \geqslant 1$, il existe un nombre fini d'homomorphismes continus respectant cette condition. On peut les déterminer et dire ensuite pour chacun d'eux si $G \Rightarrow {}_l G'$.

(ii) Il suffit de rendre effective la construction de la proposition 8 et c'est l'objet de lemme suivant :

Lemme 4 : *Il existe un algorithme se terminant toujours, qui pour toute grammaire simple $G(N, \Sigma)$ dit si elle est canonique, et si elle ne l'est pas, déter-*

mine un non-terminal atomique T_0 tel que $D(T_0) \cup E(T_0) \neq \emptyset$ ainsi que les ensembles $D(T_0)$ et $E(T_0)$.

Démonstration : Soit $G(N, \Sigma)$ une grammaire simple et T_0 un non-terminal atomique. Soit $R_i \subset N$ la suite croissante définie par :

(i) $R_0 = \{ S \in N / S \not\equiv T_0 \text{ et } \tau(S) \leqslant \tau(T_0) \}$

(ii) $R_{i+1} = R_i \cup \{ S \in N \mid S \not\equiv T_0$ et il existe $S' \in R_i, m \in N^*$ et $a \in \Sigma$ tels que $m_{S,a} = mS' \}$.

Puisque N est fini, il existe l tel que $R_{l+1} = R_l$. D'autre part puisque $\equiv$ est décidable, on peut calculer cette suite R_i et donc R_l. Montrons que pour tout $i, R_i \subset N - (D(T_0) \cup \{ T_0 \} \cup E(T_0))$. C'est clair pour $i = 0$. Grâce au lemme 3, si c'est vrai pour i, c'est vrai pour $i + 1$. Donc

$$R_l \subset N - (D(T_0) \cup E(T_0) \cup \{ T_0 \})$$

et

$$D(T_0) \subset D'(T_0) = N - (R_l \cup E(T_0) \cup \{ T_0 \}).$$

Par construction de R_l, $D'(T_0)$ vérifie le lemme 3, i.e.: si $S \in D'(T_0)$, $m_{S,a}$ se termine par un non-terminal de $D'(T_0)$ ou de $E(T_0) \cup \{ T_0 \}$. Cette condition suffit pour appliquer la construction de la proposition 8, qui montre que $D(T_0) = D'(T_0)$.

Revenons au problème initial : pour tout entier positif k, soit N_k (resp. A_k) l'ensemble des non-terminaux $S \in N$ (resp. non-terminaux atomiques) tels $\tau(S) \leqslant k$.

Pour tout k, $A_k \subset N_k$ et si $A_k = N_k$, alors $A_{k+1} = N_{k+1}$ ssi pour tout $T \in A_k$, $D'(T) = \emptyset$ (lemme 2) et ceci est décidable d'après ce qui précède. D'autre part $A_1 = N_1$. On peut donc déterminer le premier k tel que $A_k \neq N_k$ ou il existe S et T dans N_k tels que $S \equiv T$ et $S \neq T$. Si on n'en a pas trouvé après avoir épuisé N, $G(N, \Sigma)$ est canonique. Sinon, l'algorithme aura permis de calculer T atomique tel que $D(T) \cup E(T) \neq \emptyset$ ainsi que $D'(T)$. On peut calculer $E(T)$ puisque M est décidable. CQFD.

Exemple : Reprenons la grammaire G_2 de l'exemple de ce § en prenant comme axiome $\sigma_2 = TV$.

$$G_2 \begin{cases} T = aTT + bU \\ U = aVT + b \\ V = aVUV + b \quad \text{donc} \quad \tau(G_2) = 4. \end{cases}$$

Il est clair que U est atomique et que T est divisible par U (algorithme du lemme 4). $[T \mid U] = aT[T \mid U] + b$. D'autre part $V \equiv U$.

La grammaire G_3 suivante est équivalente à G_2.

$$G_3 \begin{cases} \sigma_3 = [T \mid U]UU \\ [T \mid U] = a[T \mid U]U[T \mid U] + b \\ U = aU[T \mid U]U + b \end{cases}$$

On s'aperçoit alors que $U \equiv [T \mid U]$ et on obtient la grammaire G_4 canonique,

$$G_4 \begin{cases} \sigma_4 = UUU \\ U = aUUU + b \end{cases}$$

Je remercie M. Nivat qui a suivi l'élaboration de ce travail et guidé sa rédaction, ainsi que P. Butzbach pour ses remarques sur une version préliminaire.

REFERENCES

[1] Butzbach P., *Sur l'équivalence des grammaires simples.* A paraître dans les actes de l'école de printemps sur les langages algébriques, Bonascre, 1973.

[2] Hopcroft J. E. et Korenjak A. J., *Simple Deterministic Languages,* SWAT, 1966, pp. 36-46.

[3] Nivat M., *Transduction des langages de Chomsky,* Annales de l'Institut Fourier, vol. 18 1968, p. 339-456.