RAIRO. INFORMATIQUE THÉORIQUE

CHRISTIANE FROUGNY

Grammaires algébriques et monoïdes simplifiables

RAIRO. Informatique théorique, tome 18, nº 3 (1984), p. 225-239

<http://www.numdam.org/item?id=ITA_1984__18_3_225_0>

R.A.I.R.O. Informatique Théorique/Theoretical Informatics
(vol. 18, n° 3, 1984, p. 225 à 239)

GRAMMAIRES ALGÉBRIQUES ET MONOÏDES SIMPLIFIABLES (*)

par Christiane FROUGNY ([1])

Communiqué par J.-E. PIN

Résumé. — *Étant donnée une grammaire algébrique G, le monoïde (resp. groupe) de Hotz de G est le quotient du monoïde (resp. groupe) libre engendré sur l'alphabet par les relations définies par les productions. Nous étudions le langage des mots ayant même image que l'axiome dans le monoïde de Hotz. Nous donnons une construction algébrique générale permettant de calculer le monoïde syntaxique de ce langage à partir du monoïde de Hotz. Si la grammaire est très simple à droite et à gauche, cette construction se simplifie : le monoïde de Hotz se plonge dans le groupe de Hotz, qui est libre, et le monoïde syntaxique est un quotient de Rees du monoïde de Hotz.*

Abstract. — *Given a context-free grammar G the Hotz monoid (group) of G is the quotient of the free monoid (group) on the alphabet by the relations defined by the productions. We study the langage of all words equal to the axiom in the Hotz monoid. We calculate the syntactical monoid of this language from the Hotz monoid by an algebraic and general construction. If the grammar is right and left very-simple, this construction is simpler: the Hotz monoid is embeddable in the Hotz group, which is free, and the syntactical monoid is a Rees quotient of the Hotz monoid.*

De nombreux travaux ont montré l'intérêt de l'étude des liens entre théorie des langages algébriques et théorie des groupes (citons en particulier [2, 16, 13]). Dans cette voie, Hotz [10] associe à une grammaire algébrique G un groupe, $H(G)$, qui est le quotient du groupe libre sur l'alphabet terminal et non terminal par la congruence engendrée par l'ensemble des productions de la grammaire. Il montre que $H(G)$ est un invariant pour le langage engendré par G.

Hotz a également défini le monoïde $M(G)$ d'une grammaire G comme le quotient du monoïde libre sur l'alphabet terminal et non terminal par la congruence engendrée par l'ensemble des productions.

En [7], nous avons mis en lumière le rôle clé joué dans ce domaine par les monoïdes simplifiables, et montré que le quotient simplifiable maximal du monoïde de Hotz, dénoté $C(MG)$), est un invariant pour le langage engendré par G. On retrouve ainsi le résultat de Hotz [11] : si $M(G)$ est simplifiable, alors c'est un invariant pour le langage engendré par G.

(*) Reçu en décembre 1982, révisé en février 1983.

([1]) L.I.T.P. et Université Paris-V, U.E.R. de Mathématiques, 12, rue Cujas, 75005 Paris.

Dans cet article, nous traitons deux questions. Tout d'abord, nous considérons le langage L_R, introduit dans [11], et qui est égal à l'ensemble des mots terminaux congrus à l'axiome de la grammaire G dans la congruence engendrée par l'ensemble des productions, en appliquant certaines méthodes de [7]. Nous caractérisons le monoïde syntaxique de L_R comme un quotient de Rees de $C(M(G))$. Ce résultat corrige le théorème 4 de [11].

Ensuite, nous proposons la définition de deux classes de grammaires dont le monoïde de Hotz est simplifiable. Grâce au critère de simplifiabilité d'Adjan [1] il est facile de voir que le monoïde de Hotz des grammaires très simples est simplifiable à gauche. D'où l'idée d'une extension bilatère de la définition des grammaires très simples.

Une première démarche possible consiste à symétriser la définition : une grammaire algébrique sera dite doublement très simple si elle est très simple à gauche (c'est-à-dire très simple au sens classique) et à droite. Le monoïde de Hotz de ces grammaires est alors non seulement simplifiable, mais encore plongeable dans le groupe de Hotz associé. Ce groupe n'est pas context-free en général. Les langages engendrés par les grammaires doublement très simples ne sont pas des langages très simples au sens habituel, mais sont des langages simples (donc déterministes).

Nous proposons ensuite une seconde définition, plus restrictive : une grammaire sera dite bi-très-simple si elle est très simple à gauche et à droite, et si, de plus, les lettres terminales figurant à gauche et à droite des membres droits de règles appartiennent à des ensembles disjoints. Un exemple en est la grammaire suivante du langage E :

$$G_E = (\{a, b, c, d\}, \{S, B\}, \{S \to aSBSc, S \to d, B \to b\}).$$

Nous montrons que ces grammaires sont *N.T.S.* [4] et engendrent une sous-classe stricte de la classe des langages très simples. Le monoïde syntaxique d'un langage engendré par une grammaire bi-très-simple G est un quotient de Rees de $M(G)$. Le monoïde de Hotz de ces grammaires est plongeable dans le groupe de Hotz associé, lequel groupe est libre.

1. NOTATIONS. DÉFINITIONS

Nous notons X^* le monoïde libre engendré par un ensemble X, $F(X)$ le groupe libre engendré par X, 1 l'élément neutre, $X^+ = X^* \setminus \{1\}$. Si Q est un sous-ensemble de $X^* \times X^*$, $[Q]$ dénote la congruence de X^* engendrée par les éléments de Q, $\langle Q \rangle$ la congruence de $F(X)$ engendrée par Q. La classe d'un mot w dans la congruence $[Q]$ est notée $[w]_Q$, dans la congruence $\langle Q \rangle$, $\langle w \rangle_Q$. Quand il n'y aura pas d'ambiguïté, l'indice Q sera omis.

Le monoïde présenté par ensemble de générateurs X et ensemble de relations Q, noté classiquement $[X, Q]$, est égal au quotient $X^*/[Q]$; le groupe présenté par générateurs X et relations Q, noté $\langle X, Q \rangle$, est égal à $F(X)/\langle Q \rangle$.

Une *grammaire algébrique* [9] G est un quadruplet $G=(X, V, P, S)$, où X et V sont deux ensembles finis disjoints, où P est un sous-ensemble fini de $V \times (X \cup V)^*$ et S un élément de V. X est l'ensemble des lettres terminales, V l'ensemble des symboles non terminaux, P l'ensemble des productions (ou règles) et S l'axiome de la grammaire G. On écrit $u \underset{G}{\rightarrow} v$ (resp. $u \overset{*}{\underset{G}{\rightarrow}} v$) si v peut être dérivé de u par usage direct (resp. répété) des productions de P, et on note :

$$L(G)=\{ f \in X^* / S \overset{*}{\underset{G}{\rightarrow}} f \},$$

le langage engendré par G. Si $v \in V$, $L(G, v)=\{ f \in X^*/v \overset{*}{\underset{G}{\rightarrow}} f \}$ (quand il n'y aura pas d'ambiguïté, l'indice G sera omis).

Dans cet article ne seront considérées que des grammaires algébriques, appelées simplement grammaires.

Une grammaire $G=(X, V, P, S)$ est *réduite* si :

(i) $\forall A \in V, \exists f \in X^*, A \overset{*}{\rightarrow} f$;

(ii) $\forall A \in V, \exists f, g \in X^*, S \overset{*}{\rightarrow} f A g$.

2. MONOÏDE ET GROUPE DE HOTZ D'UNE GRAMMAIRE

Nous suivons pas à pas la construction présentée en [7] pour en tirer des résultats à propos du langage L_R défini dans l'introduction.

2.1. Rappels

Un monoïde M est dit *simplifiable* à *gauche* (resp. à *droite*) si, quels que soient x, y, z dans M, $xy=xz$ (resp. $yx=zx$) entraîne $y=z$. Un monoïde est *simplifiable* s'il est simplifiable à gauche et à droite.

Tout monoïde M possède un quotient simplifiable maximal, noté $C(M)$. De façon similaire, pour tout monoïde M, il existe un groupe $G(M)$ et un morphisme $\gamma: M \rightarrow G(M)$ tel que, pour tout morphisme $\alpha: M \rightarrow G$ où G est un groupe, il existe un morphisme $\beta: G(M) \rightarrow G$ tel que $\alpha=\beta \circ \gamma$.

Un *plongement* d'un monoïde M dans un groupe G est un morphisme injectif $\alpha : M \to G$. M est alors simplifiable, mais le contraire n'est pas vrai, sauf dans le cas d'un monoïde commutatif.

Soit L un langage de X^*. Le *monoïde réducteur* de L, noté $U(L)$ est par définition [7] le quotient de X^* par la congruence la plus fine telle que L soit contenu dans une seule classe, i. e. $U(L) = X^* / [L \times L]$. De façon analogue, on associe à L son *groupe réducteur* noté $I(L)$, et égal à $F(X)/\langle L \times L \rangle$.

Considérons maintenant une grammaire $G = (X, V, P, S)$. Le *monoïde* de Hotz de G est, par définition, $M(G) = (X \cup V)^* / [P]$, et le *groupe* de Hotz de G est $H(G) = (X \cup V)^* / \langle P \rangle$ (*cf.* [10]). Posons $L = L(G)$.

Les relations entre les éléments introduits plus haut sont décrites dans le diagramme commutatif et le théorème suivants :

$$
\begin{tikzcd}
F(X) \arrow[ddd, dash] & X^* \arrow[l] \arrow[r] \arrow[d, "\mu"'] & (X \cup V)^* \arrow[r] \arrow[d, "\psi"] & F(X \cup V) \arrow[ddd] \\
 & U(L) \arrow[r, "\beta"] \arrow[d, "\sigma"'] & (X \cup V)^*/[P] = M(G) \arrow[d, "\theta"] & \\
 & C(U(L)) \arrow[r, "\gamma"] \arrow[d] & C((X \cup V)^*/[P]) = C(M(G)) \arrow[d] & \\
I(L) & G(U(L)) \arrow[l] \arrow[r, "\delta"] & G((X \cup V)^*/[P]) \arrow[r] & F(X \cup V)/\langle P \rangle = H(G)
\end{tikzcd}
$$

Diagramme 1

où μ, σ, ψ et θ sont les morphismes canoniques, α l'injection canonique de X^* dans $(X \cup V)^*$, et où β existe par définition de $U(L)$ et γ parce que $C((X \cup V)^*/[P])$ est simplifiable.

THÉORÈME 1 [7] : *Si la grammaire G est réduite, alors, avec les notations du diagramme* 1 :

(i) β *est surjectif;*

(ii) $C(U(L))$ *et* $C(M(G))$ *sont isomorphes;*

(iii) $I(L)$ *et* $H(G)$ *sont isomorphes.*

Rappelons pour terminer une construction classique.

Soit $G = (X, V, P, S)$ une grammaire réduite. Pour tout non-terminal v de V, il existe un mot u de X^* tel que $u \equiv v [P]$. Posons $V_1 = V \setminus \{v\}$. Si l'on remplace toute occurrence de v par u dans les mots de P, on obtient un nouvel ensemble de relations P_1 sur $(X \cup V_1)^*$.

Par le théorème de Tietze [5], on a : $(X \cup V_1)^*/[P_1]$ est isomorphe à $(X \cup V)^*/[P]$ et $F(X \cup V_1)/\langle P_1 \rangle$ est isomorphe à $F(X \cup V)/\langle P \rangle$. En répétant ce procédé, on peut éliminer tous les éléments de V, et on obtient un sous-ensemble Q de $X^* \times X^*$ tel que $X^*/[Q]$ est isomorphe à $(X \cup V)^*/[P]$ et $F(X)/\langle Q \rangle$ est isomorphe à $F(X \cup V)/\langle P \rangle$.

2.2. Le langage L_R.

Nous allons étudier dans ce paragraphe le langage L_R des mots de X^* équivalents à l'axiome S. Plus formellement, on pose $L_R = [S]_P \cap X^*$ et on définit, comme ci-dessus, $U(L_R)$ et $I(L_R)$; on a alors le :

THÉORÈME 2 : *Si la grammaire G est réduite, alors :*

(i) *$M(G)$ est un quotient de $U(L_R)$;*

(ii) *$C(U(L_R))$ et $C(M(G))$ sont isomorphes;*

(iii) *$I(L_R)$ et $H(G)$ sont isomorphes.*

Preuve : la preuve suit celle du théorème 1. L_R étant égal à $[S]_P \cap X^*$, est contenu dans une seule classe modulo $[P]$; il s'ensuit que $M(G)$ est un quotient de $U(L_R)$.

Comme ci-dessus nous avons le diagramme commutatif :

$$\begin{tikzcd}
F(X) \arrow[dddd] & X^* \arrow[l] \arrow[r, "\alpha"] \arrow[d, "\mu'"'] & (X \cup V)^* \arrow[r] \arrow[d, "\psi"] & F(X \cup V) \arrow[dddd] \\
& U(L_R) \arrow[r, "\beta'"] \arrow[d, "\sigma'"'] & M(G) \arrow[d, "\theta"] & \\
& C(U(L_R)) \arrow[r, "\gamma'"] \arrow[d] & C(M(G)) \arrow[d] & \\
& G(U(L_R)) \arrow[r, "\delta'"] \arrow[dl, "\eta"'] & G(M(G)) \arrow[dr, "\zeta"] & \\
I(L_R) & & & H(G)
\end{tikzcd}$$

Diagramme 2

où μ', σ', ψ et θ sont les morphismes canoniques, β' existe par définition de $U(L_R)$, et γ' existe parce que $C(M(G))$ est simplifiable.

Si G est réduite, alors $M(G) = (X \cup V)^*/[P]$ est isomorphe à $X^*/[Q]$ (comme rappelé plus haut) et $\psi \circ \alpha$ est surjectif. β' et γ' le sont donc également.

Soit $(f, g) \in Q$. Par construction de Q, il existe $v \in V$ tel que $v \overset{*}{\to} f$ et $v \overset{*}{\to} g$. Comme G est réduite, il existe x et y dans X^* tels que $S \overset{*}{\to} xvy$. Ainsi, xfy et xgy sont tous deux dans L et donc dans L_R. Par suite, xfy et xgy ont même image dans $U(L_R)$ et f et g ont même image dans $C(U(L_R))$. $C(U(L_R))$ est donc un quotient de $X^*/[Q]$. Soit $\lambda : X^*/[Q] \rightarrow C(U(L_R))$ tel que $\sigma' = \lambda \circ \beta'$. Par définition, il existe $\xi : C(X^*/[Q]) \rightarrow C(U(L_R))$ tel que $\lambda = \xi \circ \theta$. On a alors $\sigma' = \xi \circ \theta \circ \beta' = \xi \circ \gamma' \circ \sigma'$. Puisque σ' est surjectif, ξ est l'inverse de γ' et γ' est injectif.

Rappelons le résultat suivant (*cf.* [3]) : Si R est inclus dans $X^* \times X^*$, alors $G(X^*/[R])$ est isomorphe à $F(X)/\langle R \rangle$. Il s'ensuit que η et ζ sont des isomorphismes. Par ailleurs, pour tout monoïde M, $G(C(M)) = G(M)$ (*cf.* [7]). On en déduit que, comme γ' est un isomorphisme, δ en est un également. ∎

2.3. Relations avec le monoïde syntaxique

Dans [7], on montrait que si le monoïde syntaxique $\mathrm{Synt}(L)$ d'un langage L est simplifiable, et si l'image de L dans $\mathrm{Synt}\, L$ est réduite à un seul élément, alors $\mathrm{Synt}(L)$ est isomorphe à $C(M(G))$. Par ailleurs, Hotz s'intéressait à des grammaires G dont le monoïde de Hotz $M(G)$ était simplifiable [11]. Nous donnons ici un résultat sans hypothèse sur $M(G)$ ou $\mathrm{Synt}(L)$.

Toutes les grammaires considérées seront désormais supposées réduites.

Rappelons une définition : le quotient de Rees d'un monoïde M par un idéal bilatère I, noté $M//I$, peut être décrit de la façon suivante : tous les éléments de I sont envoyés sur un unique élément (le zéro), tandis que les éléments de $M \setminus I$ restent inchangés.

Soit $J(L_R)$ l'idéal maximal de X^* qui a une intersection vide avec L_R. $\mathrm{Synt}\,(L_R)$ a un zéro si et seulement si $J(L_R) \neq \emptyset$. Soit K l'image de $J(L_R)$ dans $C(U(L_R))$.

THÉORÈME 3 : *Synt* (L_R) *est isomorphe à* $C(U(L_R))//K$ *(et donc à* $C(M(G))//K$*).*

Preuve : Soient f et g deux mots de L_R. Alors $f \equiv S \equiv g\,[P]$, et il s'ensuit que f et g sont égaux dans $\mathrm{Synt}\,(L_R)$. Ainsi, l'image de L_R dans son monoïde syntaxique est réduite à un seul élément. Par définition de $U(L_R)$, $\mathrm{Synt}\,(L_R)$ est alors un quotient de $U(L_R)$ et donc de $C(U(L_R))$.

Prenons maintenant deux mots f et g de X^* qui ont la même image dans $\mathrm{Synt}\,(L_R)$. Si f et g sont tous deux dans $J(L_R)$, leur image dans $\mathrm{Synt}\,(L_R)$ est zéro, et dans $C(U(L_R))//K$ également. Si f et g n'appartiennent pas à $J(L_R)$, il existe x et y de X^* tels que xfy et xgy appartiennent à L_R. xfy et xgy ont donc même image dans $U(L_R)$ et $C(U(L_R))$. Comme $C(U(L_R))$ est simplifiable, f et g sont égaux dans $C(U(L_R))$. ∎

Ce résultat élargit la proposition 5 de [7] et rectifie le théorème 4 de [11].

Rappelons que nous connaissons plusieurs classes de grammaires qui engendrent des langages L égaux à L_R.

1. Les grammaires *N.T.S.* : une grammaire $G=(X, V, P, S)$ est dite *N.T.S.* (à non terminaux séparés) [4] si, pour tout v de V, le langage élargi engendré par v, $\hat{L}(G, v)=\{ f \in (X \cup V)^* / v \overset{*}{\rightarrow} f \}$ est égal à $[v]_P$. En particulier, on a $L=L(G, S)=[S]_P \cap X^* = L_R$.

2. Les langages et les grammaires à groupes. On appelle *noyau* d'un groupe H un langage L de X^* tel qu'il existe un morphisme surjectif φ de X^* sur H tel que $L=\varphi^{-1}(1_H)$. Un groupe dont le noyau est un langage algébrique est appelé *groupe context-free* [2]. On a le :

THÉORÈME 4 [7] : *Soit* $G=(X, V, P, S)$ *une grammaire réduite qui engendre un langage* L *noyau de groupe. Alors* $L=[S]_P \cap X^* = L_R$.

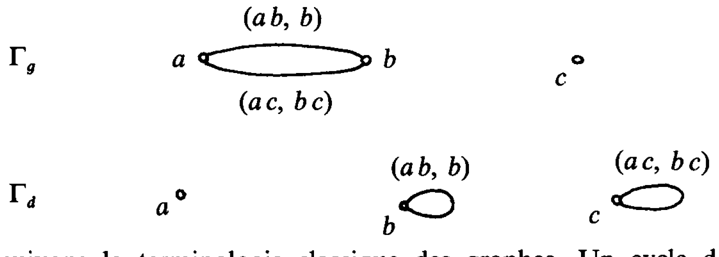
3. SIMPLIFIABILITÉ DU MONOÏDE DE HOTZ D'UNE GRAMMAIRE

Dans son article déjà cité, Hotz faisait l'hypothèse que le monoïde $M(G)$ d'une grammaire G était simplifiable. Ceci nous a donné l'idée de rechercher des grammaires dont le monoïde de Hotz soit simplifiable. Dans cette optique, le critère d'Adjan de simplifiabilité pour les monoïdes présentées par générateurs et relations s'applique aisément aux monoïdes de Hotz des grammaires très simples.

3.1. Critère de simplifiabilité d'Adjan

Soit un semi-groupe $T=[Y, R]$, où $R \subseteq Y^+ \times Y^+$. On associe à la présentation (Y, R) deux graphes non dirigés : le graphe gauche $\Gamma_g(Y, R)$ et le graphe droit $\Gamma_d(Y, R)$. Ces deux graphes ont Y comme ensemble de sommets et R comme ensemble d'arêtes. Dans $\Gamma_g(Y, R)$, une relation (r, s) de R relie les lettres initiales de r et de s. Dans $\Gamma_d(Y, R)$, une relation (r, s) de R relie leurs lettres terminales.

Exemple : $T=[a, b, c; (ab, b), (ac, bc)]$



Nous suivons la terminologie classique des graphes. Un cycle dans un graphe est un chemin fermé. Si tous les sommets d'un cycle sont différents, un tel cycle est dit élémentaire.

On dit que la présentation (Y, R) est sans cycle gauche (resp. droit) si $\Gamma_g(Y, R)$ [resp. $\Gamma_d(Y, R)$] n'a pas de cycle élémentaire.

PROPOSITION 1 (Adjan [1]) : *Si* (Y, R) *est sans cycle gauche* (*resp. droit*) *alors le semi-groupe* $[Y, R]$ *est simplifiable à gauche* (*resp. droite*). *Si* (Y, R) *n'a ni cycle gauche ni cycle droit, alors* $[Y, R]$ *est plongeable dans le groupe* $\langle Y, R \rangle$.

3.2. Monoïde de Hotz d'une grammaire très simple

Dans toute la suite, nous nous limiterons à des grammaires sans règles de la forme $v \to 1$, $v \in V$.

Rappelons la définition : une grammaire $G=(X, V, P, S)$ est *simple* si :

(i) G est sous forme normale de Greibach, i. e. $P \subseteq V \times XV^*$;

(ii) si $v \to xm$ et $v \to xm'$ sont deux productions de P, avec v dans V, x dans X, m et m' dans V^*, alors $m=m'$.

Un langage est *simple* s'il existe une grammaire simple qui l'engendre.

Une grammaire $G=(X, V, P, S)$ est *très simple* si elle est simple et si, de plus, pour tout x de X, il existe une unique règle de la forme $v \to xm$. C'est-à-dire que si $v \to xm$, $v' \to xm'$, avec v et v' dans V, m et m' dans V^*, alors $m=m'$ et $v=v'$.

Un langage est *très simple* s'il existe une grammaire très simple qui l'engendre.

PROPOSITION 2: *Le monoïde de Hotz d'une grammaire très simple est simplifiable à gauche.*

Preuve : Soit $G=(X, V, P, S)$ une grammaire très simple. Montrons que $\Gamma_g(X \cup V, P)$ est sans cycle élémentaire.

Supposons que $\Gamma_g(X \cup V, P)$ contienne une boucle, c'est-à-dire un cycle de longueur 1, de la forme a. Il est clair que le fait que G soit sous forme normale de Greibach interdit cette situation.

Nous montrons maintenant qu'il n'existe pas de chemin de longueur 2 dans $\Gamma_g(X \cup V, P)$. Soit un chemin (de longueur 2) qui relie a_0, a_1 et a_2. Si a_0 et a_1 sont liés, c'est qu'il existe une règle de P de la forme $a_0 \to a_1 m$ (ou bien $a_1 \to a_0 m$), où $a_0 \in V$ et $a_1 \in X$. Si a_1 et a_2 sont liés, alors il existe une règle $a_2 \to a_1 m'$. La grammaire étant très simple, on a : $a_0=a_2$ et $m=m'$. Ainsi, il n'y a pas de chemin de longueur 2 dans $\Gamma_g(X \cup V, P)$, et *a fortiori* pas de cycle. Ainsi, $M(G)=[X \cup V, P]$ est simplifiable à gauche. ■

Rappelons que les grammaires très simples sont $N.T.S.$ [4, 6], et donc vérifient $L=L_R$.

La proposition 2 conduit naturellement à l'idée d'une bilatéralisation des grammaires très simples. Nous proposons ici deux extensions possibles.

3.3. Grammaires doublement très simples

On sait [15] que pour toute grammaire algébrique il existe une grammaire équivalente sous forme normale de Greibach double, c'est-à-dire telle que l'ensemble P des productions de la grammaire est un sous-ensemble de $V \times (XV^* X \cup X)$.

Nous proposons alors la définition suivante : une grammaire $G=(X, V, P, S)$ est *doublement très simple* si :

(i) G est sous forme normale de Greibach double;

(ii) G est « très simple à gauche », i. e. :

$$\left.\begin{array}{llll} v \rightarrow xmy, & v \in V, & m \in V^*, & x, y \in X \\ v' \rightarrow xm'y', & v' \in V, & m' \in V^*, & y' \in X \end{array}\right\} \Rightarrow \left\{\begin{array}{l} v=v' \\ m=m', \\ y=y' \end{array}\right.$$

$$\left.\begin{array}{l} v \rightarrow xmy \\ v' \rightarrow x \end{array}\right\} \Rightarrow v=v', m=1, y=1,$$

$$\left.\begin{array}{l} v \rightarrow x \\ v' \rightarrow x \end{array}\right\} \Rightarrow v=v';$$

(iii) de manière symétrique, G est « très simple à droite ».

Un langage est *doublement très simple* s'il existe une grammaire de ce type qui l'engendre.

Dans de telles grammaires, la connaissance d'une lettre terminale et de sa position (gauche ou droite) suffit à caractériser la règle (unique) qui permet de l'engendrer.

PROPOSITION 3 : *Le monoïde de Hotz $M(G)$ d'une grammaire G doublement très simple est plongeable dans le groupe de Hotz $H(G)$.*

Preuve : De façon analogue à la démonstration de la proposition 2, on montre que la présentation $(X \cup V, P)$ n'a ni cycle droit ni cycle gauche, et le théorème d'Adjan permet de conclure. ■

Dire que $[X, Q]$ est plongeable dans $\langle X, Q \rangle$ peut s'exprimer ainsi : $\forall u \in X^*$, $[u]_Q = \langle u \rangle_Q \cap X^*$.

Nous donnons maintenant quelques propriétés vérifiées par les grammaires doublement très simples.

REMARQUE 1 : Les grammaires doublement très simples engendrent des langages simples, mais généralement pas très simples.

Preuve : 1. Soit G une telle grammaire. On construit une grammaire G' équivalente à G ainsi : si $v \to xmy$ est une règle de G, on aura dans G' deux règles : $v \to xmy$ et $Y \to y$, Y étant un nouveau non terminal. Il est clair que G' est simple.

2. Considérons la grammaire G doublement très simple de règles $S \to a\,S\,a$, $S \to b$. Le langage $\{\, a^n ba^n / n \geqq 1 \,\}$ engendré par G n'est pas très simple à cause du lemme suivant :

LEMME 1 : *Un mot d'un langage très simple ne peut commencer et finir par la même lettre.*

Preuve : Soit $f = xgx \in X^*$ où $x \in X$ un mot d'un langage engendré par une grammaire très simple $G = (X, V, P, S)$.

Rappelons le résultat suivant [6] : si, dans une grammaire très simple on a la dérivation $v \xrightarrow{*} gh$, où $v \in V$, g et $h \in X^*$, alors il existe un unique $m \in V^*$ tel que $v \xrightarrow{*} gm$, $m \xrightarrow{*} h$.

Comme $S \xrightarrow{*} f = xgx$, on sait alors qu'il existe $m \in V^*$ tel que $S \xrightarrow{*} xgm$, $m \xrightarrow{*} x$. Mais puisque la grammaire est très simple, il existe une unique règle faisant apparaître x, à savoir $S \to xp$, $p \in V^*$. Ceci est contradictoire avec $m \xrightarrow{*} x$. Ainsi f ne peut commencer et finir par la même lettre. ■

REMARQUE 2 : Une grammaire doublement très simple n'engendre pas nécessairement un langage L égal à L_R, ensemble des mots terminaux congrus à l'axiome modulo P.

Preuve : Il suffit de considérer la grammaire G suivante, de règles :

$$\begin{cases} S \to a\,S\,U\,T\,W\,b + c, \\ U \to \bar{x}\,x, \\ W \to y\,\bar{y}, \\ T \to x\,T\,y + d. \end{cases}$$

La grammaire G est doublement très simple. Le mot $a\,c\,\bar{x}\,d\,\bar{y}\,b$ est congru à S modulo P, et n'appartient pas à L. ■

REMARQUE 3 : Il existe des grammaires doublement très simples, vérifiant $L = L_R$, et telles que le groupe de Hotz associé ne soit pas un groupe context-free.

Preuve : Reprenons la grammaire G doublement très simple, de règles $\{S \to aSa,\ S \to b\}$. Il est clair que $L = L_R$.

Le groupe de Hotz est $H(G) = \langle a, b;\ abab^{-1} = 1 \rangle$. De $abab^{-1} = 1$, on tire la relation $abb = bba$. Soit C le sous-groupe de $H(G)$ engendré par les éléments a et bb. C est commutatif, et est donc isomorphe à un quotient de $\mathbb{Z}^2$. Rappelons le théorème suivant :

FREIHEITSSATZ [12] : *Soit $H = \langle X, r \rangle$ un groupe à un relateur r cycliquement réduit. Si M est un sous-ensemble de X qui omet un générateur occurant dans r, alors le sous-groupe de H engendré par M est librement engendré par M.*

Une conséquence de l'application de ce résultat à $H(G)$ est que a et bb sont deux éléments d'ordre infini. C n'est donc pas fini.

Si C était un quotient propre de $\mathbb{Z}^2$, a et bb serait liés. On aurait donc $a^p = b^{2q}$, pour p et q entiers, et $b^{2q}a^{-p} = 1$. Mais toute relation de la forme $w = 1$ dans $H(G)$ a la propriété que la somme des exposants des b est nulle. On aurait alors $q = 0 = p$. Il est donc impossible que a et bb soient liés.

Ainsi, C est isomorphe à $\mathbb{Z}^2$. Si $H(G)$ est un groupe context-free, alors C étant un sous-groupe finiment engendré, l'est également [16]. Or il est connu (*cf.* [14]) que $\mathbb{Z}$ n'est pas un groupe context-free. $H(G)$ n'est donc pas context-free. ■

3.4. Grammaires bi-très-simples

Les résultats, plutôt négatifs, qui terminent le paragraphe précédent, nous ont amenée à considérer des grammaires plus contraintes que les grammaires doublement très simples.

Ce sont des grammaires qui sont encore très simples des deux côtés mais où l'on impose, de plus, que les lettres apparaissant à droite et à gauche du membre droit des règles soient toujours distinctes. On pose alors la définition suivante : une grammaire $G = (X, V, P, S)$ est dite *bi-très-simple* si :

(i) G est sous forme normale de Greibach double :

(ii) l'ensemble des terminaux est partitionné en $X = X_g \cup X_d \cup X_c$;

(iii) les règles de P sont de la forme :

• $v \to xmy,\ v \in V,\ m \in V^*,\ x \in X_g,\ y \in X_d,$

ou :

• $v \to x,\ x \in X_c;$

(iv) G est « très simple à gauche » (comme dans la définition des grammaires doublement très simples) ;

(v) G est « très simple à droite ».

Un langage est dit *bi-très-simple* s'il peut être engendré par une grammaire de ce type.

Ces contraintes sont très fortes. Cependant, certaines grammaires bien connues sont de ce type. Ainsi la grammaire G_E de règles $\{S \to aSBSc,\ S \to d,\ B \to b\}$, qui engendre le langage E, amplement étudié (*voir* [3]).

Luc Boasson nous a signalé l'existence de grammaires très voisines des nôtres : les « bracketed context-free grammars » de [8] (en français, grammaires complètement parenthésées). Rappelons-en la définition : une grammaire algébrique $G=(X, V, P, S)$ est *complètement parentésée* si :

(i) X est partitionné en $X_g \cup X_d \cup X_c$;

(ii) toute règle de P est de la forme :

$$v \to xmy, \qquad \text{où } v \in V, \qquad x \in X_g, \qquad y \in X_d, \qquad m \in (V \cup X_c)^*;$$

(iii)
$$\left.\begin{array}{r} v \to xmy \\ v' \to xm'y' \end{array}\right\} \Rightarrow \left\{\begin{array}{l} v=v' \\ m=m'. \\ y=y' \end{array}\right.$$

Notons une différence essentielle avec les grammaires bi-très-simples : la condition (iii) signifie que dans deux règles distinctes, les parenthèses gauches sont différentes, mais les parenthèses droites peuvent être les mêmes.

Ainsi, le monoïde de Hotz de ces grammaires n'est pas simplifiable à droite en général.

Il est clair que toute grammaire bi-très-simple est doublement très simple, mais l'inverse n'est pas vrai (*cf.* les grammaires données en exemple dans les remarques 2 et 3).

Corollaire 1 : *Le monoïde de Hotz d'une grammaire bi-très-simple est plongeable dans le groupe de Hotz de la grammaire.*

En particulier, reprenons la grammaire G_E de E.

$$M(G_E)=[a, b, c, d;\ (adbdc, d)].$$

Notons $Q=(adbdc, d)$. On sait que E est égal à $[d]_Q$. Comme $M(G_E)$ est plongeable dans le groupe $H(G_E)$, on a alors : $[d]_Q=\langle d\rangle_Q \cap X^*$, ce qui signifie que le langage E n'est autre que la classe des mots terminaux congrus à d modulo Q dans le *groupe* $H(G_E)$.

Nous donnons maintenant quelques propriétés vérifiées par ces grammaires.

Proposition 4 : *Si G est bi-très-simple, alors $L(G)$ est un langage très simple.*

Preuve : On construit une grammaire très simple $G'=(X, V', P', S)$ de la façon suivante : si $v \to xmy$ est dans P, avec $x \in X_g$, $y \in X_d$, $m \in V^*$, on met dans P' les règles $v \to xm\,Y$, $Y \to y$ où $Y \in V'$ est un nouveau non terminal. Les règles terminales de la forme $v \to x$ restent inchangées. Il est clair que $L(G')=L(G)$. G' est une grammaire très simple : si $v \to xm\,Y$ et $v' \to xm'\,Y'$ dans G', c'est que $v \to xmy$ et $v' \to xm'y'$ dans G, où $Y \underset{G'}{\to} y$, $Y' \underset{G'}{\to} y'$. Comme G est bi-très-simple, $v=v'$, $m=m'$ et $y=y'$. Par définition de G', $Y=Y'$ et donc G' est très simple. ■

REMARQUE 4 : Il existe des langages très simples qui ne sont pas bi-très-simples.

Preuve : Soit L le langage engendré par la grammaire :

$$G=(\{a, b, c, t\}, \{S, T\}, P, S),$$

très simple, où :

$$P=\begin{cases} S \to a\,S\,T+b\,S\,T+c \\ T \to t \end{cases},$$

$$L=\{a^n ct^n / n \geqq 0\} \cup \{b^n ct^n / n \geqq 0\}.$$

L n'est pas un langage bi-très-simple, à cause du lemme suivant :

LEMME 2 : *Soit G une grammaire bi-très-simple, f et g deux mots de $L(G)$ de longueur $\geqq 2$. Si f et g finissent par la même lettre, alors leurs lettres initiales sont identiques (et symétriquement).*

Preuve : Écrivons $f=xuy$, avec $x, y \in X$, $u \in X^*$ et $g=zty$, avec $z \in X$, $t \in X^*$, et $z \neq x$.

Comme G est bi-très-simple, il existe une unique règle qui fasse apparaître la lettre y, à savoir $S \to amy$, où $a \in X$, $m \in V^*$. Si $S \overset{*}{\to} xuy$ et $S \overset{*}{\to} zty$, alors nécessairement, $x=z=a$. ■

PROPOSITION 5 : *Une grammaire bi-très-simple est une grammaire N.T.S.*

Preuve : Il suffit [4, 6] de montrer que, pour tout v, w de V, si $v \overset{*}{\to} \alpha u \beta$, $w \overset{*}{\to} u$ avec α, u et β dans $(X \cup V)^*$, alors $v \overset{*}{\to} \alpha w \beta$. Procédons par récurrence sur la longueur de la dérivation de w en u :

- $w \to u$: comme la grammaire est bi-très-simple, la règle $w \to u$ est l'unique façon d'engendrer u. Alors, dans une dérivation $v \overset{*}{\to} u$, on a nécessairement $v \overset{*}{\to} \alpha w \beta$.

• $w \overset{n+1}{\rightarrow} u$: alors $w \overset{n}{\rightarrow} tw'z \rightarrow tu'z = u$, où t, $z \in (X \cup V)^*$, $w' \in V$. On a $v \overset{*}{\rightarrow} \alpha u \beta = \alpha tu'z \beta$, et $w' \rightarrow u'$, d'où $v \overset{*}{\rightarrow} \alpha tw'z \beta$, et par hypothèse de récurrence, $v \overset{*}{\rightarrow} \alpha w \beta$. ∎

Revenons aux notions introduites dans le chapitre 2. Soit L un langage engendré par une grammaire bi-très-simple, $J(L)$ l'idéal maximal de X^* qui a une intersection vide avec L, K l'image de $J(L)$ dans $C(U(L))$.

PROPOSITION 6 : *Soit L un langage engendré par une grammaire bi-très-simple G. Synt (L) est isomorphe au quotient de Rees de M (G) par K.*

Preuve : Comme G est *N.T.S.*, $L = L_R$. De plus $M(G)$ est simplifiable, et donc égal à $C(U(L))$. La proposition est alors une simple conséquence du théorème 3. ∎

En particulier, $\text{Synt}\, E = [a, b, c, d; (adbdc, d)]//K$.

Terminons par le groupe de Hotz d'une grammaire bi-très-simple.

PROPOSITION 7 : *Le groupe de Hotz d'une grammaire bi-très-simple est libre.*

Preuve : Soit $G = (X, V, P, S)$ une grammaire bi-très-simple, $H(G) = \langle X \cup V, P \rangle$ son groupe de Hotz. Notons $p_1, p_2, \ldots, p_n$ les éléments de P. Alors :

$$H(G) \simeq ((F(X \cup V)/\langle p_1 \rangle)/\langle p_2 \rangle \ldots)/\langle p_n \rangle.$$

Nous allons montrer que $F(X \cup V)/\langle p_1 \rangle$ est libre. Remarquons tout d'abord que si $p_1 = (v, m)$, alors mv^{-1} est cycliquement réduit, parce que G est sous-forme normale de Greibach double.

Nous examinons maintenant cas par cas la règle p_1.

1. p_1 est de la forme $v \rightarrow x_c$, où $x_c \in X_c$. D'après le Freiheitssatz, le sous-groupe $\langle (X \cup V) - \{x_c\} \rangle$ est libre, et il est clairement isomorphe à $\langle X \cup V, p_1 \rangle$.

2. p_1 est de la forme $v \rightarrow x_g m x_d$, où $x_g \in X_g$, $x_d \in X_d$, $m \in V^*$. Dans le groupe, on a $x_g = v x_d^{-1} m^{-1}$. D'après le Freiheitssatz, $\langle (X \cup V) - \{x_g\} \rangle$ est libre. Or il est isomorphe à $\langle X \cup V, p_1 \rangle$. Ainsi, $\langle X \cup V, p_1 \rangle$ est toujours libre.

Remarquons maintenant que dans le premier cas, c'est-à-dire $p_1 = v \rightarrow x_c$, x_c n'apparaît jamais comme lettre d'une des règles $p_2, \ldots, p_n$, car la grammaire étant bi-très-simple, seule p_1 fait apparaître x_c. De même dans le second cas, $p_1 = v \rightarrow x_g m x_d$, x_g n'est jamais une lettre d'une règle $p_2, \ldots, p_n$. On voit alors que, en recommençant pour p_2, puis p_3, *etc.* p_n, ce que l'on a fait pour p_1, on obtient, à chaque étape, un groupe libre.

Ainsi $H(G)$ est libre, et isomorphe au groupe libre engendré par $(X \cup V) \setminus \{n \text{ lettres terminales}\}$.

On pourrait également montrer que $H(G)$ est isomorphe au groupe libre engendré par un sous-ensemble de X.

REMERCIEMENTS

Je tiens à remercier Jacques Sakarovitch pour ses nombreuses remarques et suggestions.

BIBLIOGRAPHIE

1. S. I. ADJAN, *Defining Relations and Algorithmic Problems for Groups and Semigroups*, Proc. Steklov Inst., vol. 85, 1966, Amer. Math. Soc. Transl., vol. 152, 1967.
2. A. V. ANISIMOV et F. D. SEIFERT, *Zur algebraischen Charakteristik der durch kontextfreie Sprachen definierten Gruppen*, Elektronische Informationsverarbeitung und Kybernetik, vol. 11, 1975, p. 695-702.
3. J. BEAUQUIER, *Contribution à l'étude de la complexité structurelle des langages algébriques*, Th. Sc. Math., Univ. Paris-VII, 1979.
4. L. BOASSON, *Dérivations et réductions dans les grammaires algébriques*, Proc. of the 7th I.C.A.L.P., Lecture Notes in Computer Science, vol. 85, 1980, p. 109-118.
5. A. CLIFFORD et G. PRESTON, *The Algebraic Theory of Semigroups*, Amer. Math. Soc., vol. 1, 1961; vol. 2, 1967.
6. Ch. FROUGNY, *Une famille de langages algébriques congruentiels : les langages à non-terminaux séparés*, Thèse 3e cycle, Univ. Paris-VII, 1980.
7. Ch. FROUGNY, J. SAKAROVITCH et E. VALKEMA, *On the Hotz Group of a Context-Free Grammar*, Acta Informatica, vol. 18, 1982, p. 109-115.
8. S. GINSBURG et M. HARRISON, *Bracketed Context-Free Languages*, Journal of Computer and System Sciences, vol. 1, 1967, p. 1-23.
9. M. HARRISON, *Introduction to Formal Language Theory*, Addison Wesley, 1978.
10. G. HOTZ, *Eine neue Invariante für kontextfreie Sprachen*, Theoret. Computer Sc., vol. 11, 1980, p. 107-116.
11. G. HOTZ, *Über die Darstellbarkeit des syntaktischen Monoides Kontextfreier Sprachen*, R.A.I.R.O. Informatique Théorique, vol. 13, 1979, p. 337-345.
12. R. C. LYNDON et P. E. SCHUPP, *Combinatorial Group Theory*, Springer, 1977.
13. D. E. MULLER et P. E. SCHUPP, *Pushdown Automata, Graphs, Ends, Second-Order logic, and reachability Problems*, Proc. of the 13th Symposium on Theory of Computing, 1981, p. 46-54.
14. J. F. PERROT, *Monoïdes syntactiques des langages algébriques*, Acta Informatica, vol. 7, 1977, p. 399-413.
15. D. J. ROSENKRANTZ, *Matrix Equations and Normal Forms for Context-Free Grammars*, Journal of the Association for Computing Machinery, vol. 14, 1967, p. 501-507.
16. J. SAKAROVITCH, *Sur les groupes infinis, considérés comme monoïdes syntaxiques de langages formels*, Séminaire Dubreil 1975-1976, Lecture Notes 586, 1977, p. 168-179.