RAIRO. INFORMATIQUE THÉORIQUE

P. MARCHAND

Grammaires parenthésées et bilangages réguliers

RAIRO. Informatique théorique, tome 14, nº 1 (1980), p. 3-38
<http://www.numdam.org/item?id=ITA_1980__14_1_3_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 14, n° 1, 1980, p. 3 à 38)

GRAMMAIRES PARENTHÉSÉES ET BILANGAGES RÉGULIERS (*)

par P. MARCHAND [1]

Communiqué par M. NIVAT

Résumé. — *Il est bien connu que les arbres peuvent s'écrire linéairement en utilisant des parenthèses. Dans cet article, on fait le lien entre les grammaires parenthésées et les grammaires d'arbres. La définition présentée ici des grammaires parenthésées est plus générale que la définition classique. Une caractérisation du cas particulier correspondant à la définition classique est donnée. Les résultats fondamentaux sont : on peut réduire ces grammaires à une forme très élémentaire et de là découvrir une grammaire minimale canonique pour un langage parenthésé donné. Toutes les constructions faites dans cet article sont effectives. Le dernier paragraphe est consacré à quelques applications à la théorie des langages.*

Abstract. — *It is well known that trees may be written linearly with the use of parentheses. In this paper a link is established between parenthesis grammars and tree grammars. The definition we give for parenthesis grammars is more general than the classical one. A characterization of the special case corresponding to the classical definition is given. The fondamental results are that these grammars can be reduced to a very elementary form from which a minimal canonical grammar for a given parenthesis language can be obtained. All constructions in the paper are effective. The last section is devoted to some applications to language theory.*

1. INTRODUCTION

Il existe une parenté étroite entre les arbres et les expressions parenthésées. Dans le paragraphe 1.1, on rappelle la formalisation de la notion d'arbre, dans 1.2 on redéfinit les expressions parenthésées et les langages de Dyck. Dans 1.3 on compare ces deux notions.

(*) Reçu en octobre 1978.

[1] Université de Nancy-I, Laboratoire d'Informatique, Nancy.

1.1. Ramification. Bilangages. Binoïdes

Nous donnerons ici des définitions peu formalisées. Pour une étude formelle des objets décrits ci-dessus le lecteur pourra se rapporter à [10 ou 7].

Une ramification sur un vocabulaire V est un arbre vide ou à une ou plusieurs racines dont les nœuds sont étiquetés par des éléments de V. On suppose de plus que, dans une ramification, les racines et les successeurs d'un point sont ordonnés. On note $\hat{V}$ l'ensemble des ramifications sur V. On appelle bilangage un sous-ensemble de $\hat{V}$.

On appelle mot des racines d'une ramification r le mot obtenu en prenant la suite des étiquettes des racines de r. On appelle feuille de r un nœud de r sans successeur et mot des feuilles de r le mot obtenu en prenant la suite des étiquettes des feuilles de r.

Exemple : le mot des racines de r représenté à la figure 1 est $\rho(r)=ab$, le mot des feuilles de r est $\varphi(r)=baca$:

Figure 1.

On peut munir $\hat{V}$ d'une structure algébrique en considérant les deux lois suivantes :

(i) une loi interne notée $+$ qui est la concaténation des ramifications. La figure 2 donne un exemple de calcul pour la loi $+$.

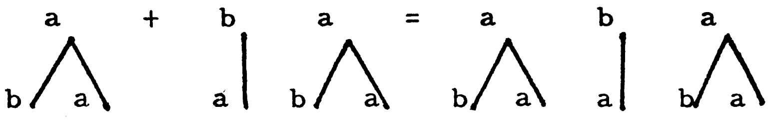

Figure 2.

la loi $+$ est associative et possède un élément neutre qui est la ramification vide notée Λ (arbre à zéro racine);

(ii) une loi externe notée $\times$ à opérateur dans V telle que $a \times r$ soit la ramification obtenue à partir de r en créant au-dessus de r une racine étiquetée par a. La figure 3 donne un exemple de tel calcul.

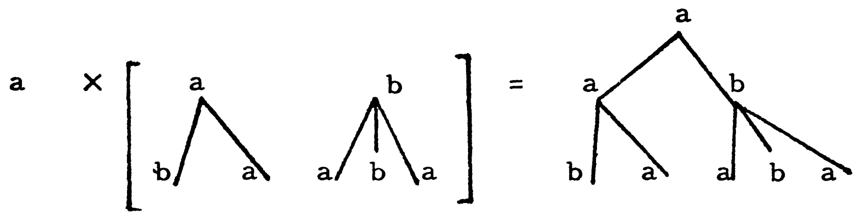

Figure 3.

Si r est différent de Λ il existe r', r'' uniques dans $\hat{V}$ et a unique dans V tels que $r = a \times r' + r''$.

Définition 1.1.1 [10] : On appelle V-binoïde un ensemble B muni de deux lois $+$ et $\times$ telles que la loi $+$ soit interne associative et possède un élément neutre et que la loi $\times$ soit externe à opérateur dans V.

Un homomorphisme entre deux V-binoïdes B et B' est une application $\psi : B \to B'$ telle que

$$\psi(x+y) = \psi(x) + \psi(y) \qquad \text{et} \qquad \psi(a \times x) = a \times \psi(x).$$

Proposition 1.1.2 [10] : *Le V-binoïde $\hat{V}$ est universel dans la catégorie des V-binoïdes. C'est-à-dire que pour tout V-binoïde B il existe un et un seul homomorphisme ψ de $\hat{V}$ dans B.*

Proposition 1.1.3 [10] : *Soit P une propriété, on a l'équivalence*

$$(\forall r \in \hat{V}),\ (P(r)) \iff P(\Lambda) \text{ et}$$

$$(\forall r'),\ (\forall r''),\ (\forall a \in V),\ P(r') \quad \text{et} \quad P(r'') \Rightarrow P(a \times r' + r'').$$

Soit E un ensemble, e_0 un élément de E et $g : \hat{V}^2 \times V \times E^2 \to E$, il existe une et une seule application $f : \hat{V} \to E$ telle que

$$f(\Lambda) = e_0,$$

$$(\forall a \in V),\ (\forall r' \in \hat{V}),\ (\forall r'' \in \hat{V}),$$

$$[f(a \times r' + r'') = g(r', r'', a, f(r'), f(r''))].$$

Cette proposition permet de faire des raisonnements pour récurrence sur $\hat{V}$ et de définir des fonctions pour des récurrences simples.

Par exemple $\rho : \hat{V} \to V^*$ est défini pour

$$\rho(\Lambda)=\Lambda,$$
$$\rho(a \times r' + r'')=a\,\rho(r'').$$

Dans la suite nous aurons souvent à utiliser un prédicat sur $V \times \hat{V}$ signifiant que $a \in V$ n'apparaît qu'aux feuilles de la ramification r. On notera Φ ce prédicat qui se définit pour la récurrence

$$\begin{aligned}&\Phi(a,\ \Lambda)=\text{vrai},\\ &\Phi(a,\ b \times r' + r'')=\mathit{si}\ b \neq a\ \mathit{alors}\ \Phi\ (a,\ r')\ \text{et}\ \Phi\ (a,\ r''),\\ &\qquad\qquad \mathit{sinon}\ \text{si}\ r' \neq \Lambda\ \mathit{alors}\ \text{faux},\\ &\qquad\qquad \mathit{sinon}\ \Phi\ (a,\ r'')\ \mathit{fin}\ \mathit{fin}.\end{aligned}$$

On utilisera aussi des fonctions $F_a : \hat{V} \to \mathcal{P}\ (\hat{V})$ donnant pour chaque a de V l'ensemble des mots juste en dessous de a dans une ramification r. F_a est définie pour la récurrence

$$\begin{aligned}&F_a(\Lambda)=\emptyset,\\ &F_a(b \times r' + r'')=\mathit{si}\ b \neq a\ \mathit{alors}\ F_a(r') \cup F_a(r''),\\ &\qquad\qquad \mathit{sinon}\ \{\rho(r')\} \cup F_a(r') \cup F_a(r'').\end{aligned}$$

On peut remarquer que $\Phi\ (a,\ r) \Leftrightarrow (F_a(r) \subseteq \{\Lambda\})$.

Si V et V' sont deux vocabulaires et si τ est une application de V dans V' on prolonge τ à $\hat{V}$ en posant

$$\bar{\tau}(\Lambda)=\Lambda \qquad \text{et} \qquad \bar{\tau}(a \times r' + r'')=\tau(a) \times \bar{\tau}(r') + \bar{\tau}(r').$$

Les applications obtenues par ce procédé s'appellent des transcriptions; leur rôle est de changer le nom des étiquettes des nœuds des ramifications.

1.2. Expressions parenthésées. Langage de Dyck. Grammaire parenthésée

Soit T, $\overline{T}$ et T' trois vocabulaires disjoints, T et $\overline{T}$ étant en bijection par l'application $a \mapsto \bar{a}$, on définit les expressions parenthésées sur ces vocabulaires comme étant les mots engendrés par la grammaire

$$G=(\{X\},\ T \cup \overline{T} \cup T',\ \to,\ X) \qquad \text{avec} \quad X \to a'X \mid aX\bar{a}X \mid \Lambda,$$

où a' est un élément quelconque de T' et a un élément quelconque de T.

Le langage de Dyck sur T noté $P(T)$ correspond au cas où T' est vide. Remarquons que l'ensemble des expressions parenthésées sur T, $\bar{T}$ et T' peut se plonger injectivement dans le langage de Dyck sur $T \cup T'$ en utilisant l'homomorphisme

$$h: \quad (T \cup \bar{T} \cup T')^* \to (T \cup \bar{T} \cup T' \cup \bar{T}')^*$$

défini par

$$a \in T \quad \Rightarrow \quad h(a) = a \qquad \text{et} \qquad h(\bar{a}) = \bar{a},$$
$$a' \in T \quad \Rightarrow \quad h(a') = a' \bar{a}'.$$

Nous utiliserons cette remarque pour travailler uniquement dans des langages de Dyck.

DÉFINITION 1.2.1 : Une grammaire parenthésée est un sextuplet $G = (N, T', T, \bar{T}, \to, X)$ tel que :

(i) $N, T', T, \bar{T}$ sont quatre vocabulaires deux à deux disjoints T et $\bar{T}$, étant une bijection pour $a \mapsto \bar{a}$;

(ii) $G' = (N, T' \cup T \cup \bar{T}, \to, X)$ est une grammaire algébrique dont les règles sont de la forme $A \to a \alpha \bar{a}$ avec $A \in N$, $a \in T$ et $\alpha \in (N \cup T')^*$.

Le langage $L(G)$ engendré par G est le langage engendré par G'. Il est évident que $L(G)$ est formé d'expressions bien parenthésées sur T, $\bar{T}$ et T'. Un langage parenthésé est un langage qui peut être engendré par une grammaire parenthésée.

Suivant les auteurs, T est réduit à un élément (*Parenthesis Grammar* [4, 11]) ou bien à chaque règle de G est associé un type de parenthèse (Bracheted context free langages [2]).

Au paragraphe 2.1 nous donnons une définition plus générale des grammaires parenthésées et au paragraphe 3.2 nous caractérisons le cas particulier défini en 1.2.1.

1.3. Représentation parenthésée des ramifications

Il est très facile de définir une bijection entre le V-binoïde universelle $\hat{V}$ et le langage de Dyck sur V. Il suffit de poser

$$\theta(\Lambda) = \Lambda \qquad \text{et} \qquad \theta(a \times r + s) = a\, \theta(r)\, \bar{a}\, \theta(s).$$

L'application θ est la représentation parenthésée infixée des ramifications. C'est la seule représentation parenthésée que nous utiliserons ici, mais il n'est pas difficile de montrer que les résultats de cet article subsistent quand on utilise les autres représentations parenthésées possibles (représentation préfixée, postfixée... que l'on trouvera par exemple dans [7]).

Cette correspondance entre les ramifications et les mots d'un langage de Dyck invite à comparer les sous-langages des langages de Dyck engendrés par les grammaires parenthésées par exemple et les bilangages.

En particulier, on introduit au paragraphe 1.4 la notion de bilangage régulier et on montre que les bilangages réguliers contiennent strictement les langages parenthésés définis en 1.2.1.

Une façon de construire des sous-langages algébriques des langages de Dyck est d'en prendre l'intersection avec des langages réguliers. Dans le paragraphe 5 consacré aux applications on caractérise de manière effective les bilangages réguliers qui sont de ce type.

1.4. Polynôme. Bigrammaire régulière

Les définitions ci-dessus sont facilement généralisables. Une étude plus complète est faite dans [5].

DÉFINITION 1.4.1 : On introduit un marqueur $\mathbf{1}$ qui n'appartient sauf mention du contraire à aucun des vocabulaires manipulés dans la suite. Un polynôme sur V est une ramification r sur $V \cup \{\mathbf{1}\}$ telle que $\mathbf{1}$ ait une seule occurrence dans r que celle-ci soit une feuille de r. On note $\hat{V}[\mathbf{1}]$ l'ensemble des polynômes sur V. (La définition 1.1 correspond aux polynômes du 1[er] degré et à une variable dans [5].)

PROPOSITION 1.4.2 : *Soit B un W-binoïde et r un élément de $\hat{V}[\mathbf{1}]$ avec V contenu dans W. Soit $\psi : \hat{V} \to B$ l'unique homomorphisme de V-binoïde de $\hat{V}$ dans B. On peut associer à r une application notée r_B de B dans B par la récurrence suivante :*

$$(\mathbf{1})_B(b) = b,$$

$$r \in \hat{V}[\mathbf{1}] \quad et \quad a \in V \quad \Rightarrow \quad (a \times r)_B(b) = a \times r_B(b),$$

$$r \in \hat{V} \quad et \quad s \in \hat{V}[\mathbf{1}] \quad \Rightarrow \quad (r+s)_B(b) = \psi(r) + s_B(b) \ et$$

$$(s+r)_B(b) = s_B(b) + \psi(r).$$

Ce type d'application sera surtout utilisé dans les cas où $B = \hat{V}$, ou $B = \hat{V}[\mathbf{1}]$ dans ce cas, on notera encore r *l'application associée à* r. Dans ce cas, si $r \in \hat{V}[\mathbf{1}]$ et $s \in \hat{V}$, alors $r(s)$ est la ramification obtenue en « greffant » s à la place de $\mathbf{1}$ dans r. Remarquons aussi que

$$r \in \hat{V}[\mathbf{1}] \quad \text{et} \quad s \in \hat{V}[\mathbf{1}] \quad \Rightarrow \quad r(s) \in \hat{V}[\mathbf{1}].$$

DÉFINITION 1.4.3 [5] : Une C_0-bigrammaire est un quadruplet $G=(N, T, \twoheadrightarrow, X)$ tel que N et T sont deux vocabulaires disjoints appelés vocabulaire non terminal et vocabulaire terminal, X est un élément de N et $\twoheadrightarrow$ est une relation entre N et $\widehat{N \cup T}$ tel que :

- un nombre fini de couples est en relation pour $\twoheadrightarrow$;
- $A \twoheadrightarrow r \Rightarrow (\forall B \in N), (\Phi(B, r))$.

Intuitivement r peut être un second membre de règle si et seulement si les non terminaux de r n'apparaissent qu'aux feuilles.

On définit dans $(N \cup T)$ la relation « se réécrit dans G » notée $\underset{G}{\gg\!\!-}$ par

$$r \underset{G}{\gg\!\!-} s \;\Leftrightarrow\; (\exists u \in \widehat{N \cup T}[\mathbf{1}]), \quad (\exists A \in N), \quad (\exists t),$$
$$[A \underset{G}{\twoheadrightarrow} t \text{ et } r=u(A) \text{ et } s=u(t)].$$

On définit ensuite comme pour les grammaires algébriques la relation dérive notée $\overset{*}{\underset{G}{\gg\!\!-}}$ et le bilangage engendré $BL(G)=\{r;\, X \overset{*}{\underset{G}{\gg\!\!-}} r \text{ et } r \in T^*\}$.

Ce type de bigrammaires a été étudié en détail dans [5]. Si on traduit ces bigrammaires dans le langage de Dyck en utilisant la même technique qu'au paragraphe 2, on obtient les « ballanced grammars » introduites dans [3].

Dans cet article on utilise seulement un cas particulier de C_0-bigrammaire définie ci-dessous.

DÉFINITION 1.4.4 [5] : Une bigrammaire régulière est une C_0-bigrammaire $G=(N, T, \twoheadrightarrow, X)$ telle que

$$A \twoheadrightarrow r \;\Rightarrow\; \rho(r) \in T^*(N \cup \{\Lambda\}).$$

DÉFINITION 1.4.5 : Un bilangage régulier est un bilangage qui peut être engendré par une bigrammaire régulière.

REMARQUE : Cette définition des bilangages réguliers n'est pas celle donnée classiquement dans [10, 12] nous verrons au paragraphe 4.2 la coïncidence des deux définitions.

L'étude des bigrammaires régulières comportera en particulier un théorème de réduction très puissant (§ 3.1) qui conduit à construire une bigrammaire canonique associée à un bilangage régulier donné (§ 4.3). Cette bigrammaire canonique est utilisée au paragraphe 5.2 pour caractériser les intersections des langages de Dyck et des langages réguliers.

Des auteurs ont étudié des automates d'arbres et défini des familles d'arbres [13]. On obtient des notions qui ne sont pas très loin des bilangages réguliers, cependant ces auteurs travaillent en général non pas sur $\hat{V}$ mais sur un magma libre [13], c'est-à-dire que le nombre de fils d'un point d'un arbre est déterminé par ce point. Cela n'est pas le cas ici. On présente au paragraphe 4.2 une structure qui est voisine des automates d'arbre et qui est adaptée à $\hat{V}$.

2. BIGRAMMAIRES RÉGULIÈRES ET GRAMMAIRES PARENTHÉSÉES

2.1. Définition des grammaires parenthésées

DÉFINITION 2.1.1 : Soit T et $\overline{T}$ deux vocabulaires disjoints en bijection par l'application $a \mapsto \bar{a}$. Soit N un troisième vocabulaire. On note $R(N, T)$ l'ensemble des mots engendré par la grammaire algébrique

$$G_R(\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}, N \cup T \cup \overline{T}, \rightarrow, \mathbf{X})$$

telle que

$$\mathbf{X} \rightarrow \mathbf{Y} A \mid \mathbf{Y} \quad \text{pour tout } A \text{ de } N,$$

$$\mathbf{Y} \rightarrow a \mathbf{Z} \bar{a} \mathbf{Y} \mid \Lambda \quad \text{pour tout } a \text{ de } T,$$

$$\mathbf{Z} \rightarrow \mathbf{Y}\mathbf{Z} \mid A \mathbf{Z} \mid \Lambda \quad \text{pour tout } A \text{ de } N.$$

On montre facilement que cette grammaire G_R n'est pas ambiguë.

PROPOSITION 2.1.2 : *Soit ω un objet qui n'est pas un élément de $N \cup T \cup \overline{T}$ et*

$$\theta : \widehat{N \cup T} \rightarrow (N \cup T \cup \overline{T})^* \cup \{\omega\} \quad \textit{défini par}$$

$$\theta(\Lambda) = \Lambda,$$

$$\theta(a \times r + r') = \textit{si } \theta(r) \neq \omega \quad \textit{et} \quad \theta(r') \neq \omega$$
$$\textit{alors } a\,\theta(r)\,\bar{a}\,\theta(r') \textit{ sinon } \omega,$$

$$\theta(A \times r + r') = \textit{si } \theta(r') \neq \omega \textit{ et } r = \Lambda$$
$$\textit{alors } A\,\theta(r') \textit{ sinon } \omega.$$

On a alors l'équivalence suivante :

$$\alpha \in R(N, T) \quad \Leftrightarrow \quad (\exists r \in \widehat{N \cup T}),$$

$$[\theta(r) = \alpha \textit{ et } \rho(r) \in T^*(N \cup \{\Lambda\}) \text{ et } (\forall A \in N), (\Phi(A, r)].$$

De plus si $\alpha \in R(N, T)$ il existe une et une seule ramification r vérifiant $\theta(r) = \alpha$.

Démonstration : On commence par démontrer que

$$\theta(r) \neq \omega \quad \Leftrightarrow \quad (\forall A \in N), \quad (\Phi(A, r)),$$

ce qui se fait aisément par récurrence sur r.

L'équivalence cherchée se démontre alors facilement par récurrence.

L'application θ est en fait une généralisation de la représentation infixée des ramifications.

DÉFINITION 2.1.3 : On appelle grammaire parenthésée un sextuplet $G=(N, T', T, \overline{T}, \to, X)$ tel que :

– $N, T', T, \overline{T}$ sont des vocabulaires deux à deux disjoints;

– T et $\overline{T}$ sont en bijection par l'application $a \mapsto \bar{a}$;

– le quadruplet $G'=(N, T' \cup T \cup \overline{T}, \to, X)$ est une grammaire algébrique telle que si $A \to \alpha$ alors α appartient à $R(N \cup T', T)$. Le langage engendré par G est le langage engendré par la grammaire G'. Un langage est dit parenthésé s'il peut être engendré par une grammaire parenthésée.

REMARQUE 1 : La définition 2.1.3 est évidemment beaucoup plus générale que celle donnée en 1.2.1. Nous appellerons grammaires parenthésées classiques les grammaires définies en 1.2.1.

REMARQUE 2 : Soit $G=(N, T', T, \overline{T}, \to, X)$ une grammaire parenthésée, on peut construire une grammaire parenthésée $G''=(N, T'', T_1, \overline{T}_1, \leadsto, X)$ équivalente à G en un certain sens où T'' est vide, il suffit pour cela de choisir un vocabulaire $\overline{T}'$ en bijection avec T', de poser $T_1=T \cup T'$ et $\overline{T}_1=\overline{T} \cup \overline{T}'$ et de définir

$$A \leadsto \alpha \quad \Leftrightarrow \quad (\exists \alpha'),\ [A \to \alpha' \text{ et } \alpha=h(\alpha')],$$

où h est l'homomorphisme

$$h\colon (N \cup T' \cup T \cup \overline{T})^* \to (N \cup T' \cup \overline{T}' \cup T \cup \overline{T})^*$$

défini par

$$\begin{aligned} A \in N &\Rightarrow h(A)=A,\\ a \in T &\Rightarrow h(a)=a,\\ \bar{a} \in \overline{T} &\Rightarrow h(\bar{a})=\bar{a},\\ a' \in T' &\Rightarrow h(\bar{a}')=a'\bar{a}'. \end{aligned}$$

L'homomorphisme h remplace donc chaque élément de T' par un couple de parenthèses successives. Il est trivial de démontrer que

$$h(L(G))=L(G'')$$

et que de plus h est une bijection de $L(G)$ sur $L(G'')$.

Dans toute la suite nous supposerons donc $T' = \varnothing$ et une grammaire parenthésée, sera donc un quintuplet $G = (N, T, \overline{T}, \rightarrow, X)$. Remarquons que dans ce cas $L(G)$ est contenu dans le langage de Dyck construit sur T.

2.2. Correspondance entre bigrammaires régulières et grammaires parenthésées

THÉORÈME 2.2.1 : *Soit $N, T, \overline{T}$ trois vocabulaires, il existe une bijection entre les bigrammaires régulières ayant comme vocabulaires N et T et les grammaires parenthésées ayant comme vocabulaires N, T et $\overline{T}$ telle que si $G = (N, T, \twoheadrightarrow, X)$ et $G' = (N, T, \overline{T}, \rightarrow, X')$ sont associées alors G et G' ont le même axiome et*

$$A \underset{G}{\overset{*}{\succ\!\!\succ\!\!-}} r \Leftrightarrow A \underset{G'}{\overset{*}{\succ\!\!-}} \theta(r).$$

Démonstration : Soit $G = (N, T, \twoheadrightarrow, X)$ une bigrammaire régulière et $G' = (N, T, \overline{T}, \rightarrow, X')$ une grammaire parenthésée. On dira que G et G' sont associées si et seulement si elles ont le même axiome et

$$(\forall A \in N), \quad (\forall r \in \widehat{N \cup T}), \quad [A \underset{G}{\twoheadrightarrow} r \Leftrightarrow A \underset{G'}{\rightarrow} \theta(r)].$$

Cette association est évidemment bijective. Il est facile de démontrer par récurrence que si G et G' sont associés alors

$$(\forall A \in N), \quad (\forall r \in \widehat{N \cup T}), \quad [A \underset{G}{\overset{*}{\succ\!\!\succ\!\!-}} r \Leftrightarrow A \underset{G'}{\overset{*}{\succ\!\!-}} \theta(r)].$$

On en déduit que $L(G')$ et $BL(G)$ sont liés par la relation

$$\theta(BL(G)) = L(G').$$

[L'application θ sur $BL(G)$ est d'ailleurs dans ce cas la représentation infixée des ramifications de $BL(G)$.]

On peut donc passer de façon biunivoque des grammaires parenthésées aux bigrammaires régulières.

Dans la suite nous ferons les démonstrations en termes de bigrammaires régulières et nous nous contenterons de traduire les résultats en termes de grammaires parenthésées sous forme de corollaires.

3. RÉDUCTION DES BIGRAMMAIRES RÉGULIÈRES ET CARACTÉRISATION DES GRAMMAIRES PARENTHÉSÉES CLASSIQUES

3.1. Réduction des bigrammaires régulières

Deux bigrammaires sont équivalentes si et seulement si elles engendrent le même bilangage.

THÉORÈME 3.1.1 : *Soit* $G=(N, T, \rightarrow, X)$ *une bigrammaire régulière, il existe une bigrammaire régulière* $G'=(N', T, \mapsto, X)$ *équivalente à* G *et dont les règles sont de la forme*

$$A \mapsto \Lambda \quad ou \quad A \mapsto a \times B + C \qquad (A, B, C \in N', a \in T).$$

Démonstration : Nous donnerons seulement le schéma de la réduction, la démonstration complète se trouve dans [5]. Tout au cours de la réduction si on rencontre des règles de la forme $A \rightarrow B$ on les supprime en employant la méthode utilisée pour les grammaires algébriques.

1er *pas :* On peut trouver une bigrammaire équivalente à G où toutes les règles sont de la forme $A \rightarrow r+B$ ou $A \rightarrow \Lambda$ avec $\rho(r) \in T^*$.

Il suffit de remplacer les règles $A \rightarrow r$ par $A \rightarrow r+A'$ et $A' \rightarrow \Lambda$ où A' est un nouveau non terminal. Supposons cette condition réalisée.

2^{e} *pas :* On peut trouver une bigrammaire équivalente à G où toutes les règles sont de la forme $A \rightarrow a \times r+B$. Il suffit pour cela de remplacer chaque règle de la forme

$$A \rightarrow a_1 \times r_1 + a_2 \times r_2 + \ldots + a_k \times r_k + B$$

par

$$A \rightarrow a_1 \times r_1 + A_1, \qquad A_1 \rightarrow a_2 \times r_2 + A_2, \qquad \ldots, \qquad A_{k-1} \rightarrow a_k \times r_k + B,$$

où $A_1, A_2, \ldots, A_{k-1}$ sont de nouveaux non terminaux.

Supposons cette condition réalisée.

3^{e} *pas :* On peut trouver une bigrammaire équivalente à G où toutes les règles sont de la forme $A \rightarrow a \times (A_1 + A_2 + \ldots + A_k) + B$. Il suffit pour montrer cela de trouver un procédé permettant d'abaisser la hauteur d'une ramification qui est dans le second membre d'une règle. Or si $A \rightarrow a \times (r_1 + r_2 + \ldots + r_k) + B$ est une règle de G, on obtient une grammaire équivalente en remplaçant cette règle par $A \rightarrow a \times (A_1 + A_2 + \ldots + A_k) + B$, $A_1 \rightarrow r_1, \ldots, A_k \rightarrow r_k$ où $A_1, A_2, \ldots, A_k$ sont de nouveaux non terminaux. En réitérant ce procédé on obtient la condition annoncée que l'on suppose réalisée.

4^{e} *pas :* Les trois premiers pas étaient triviaux, celui-ci l'est moins. Pour arriver au résultat annoncé dans le théorème il suffit maintenant de trouver un procédé permettant de diminuer le nombre k_G défini par

$$k_G = \operatorname{Sup}(|\rho(r)|; (\exists A), (\exists B), (\exists a), (A \rightarrow a \times r + B)).$$

Si $k_G = 1$ la réduction est terminée.

Supposons $k_G \geqq 2$. Soit P_1 l'ensemble des règles de G de la forme $A \rightarrow a \times r + B$ avec $|\rho(r)| = k_G$ que l'on numérote de 1 à p, soit P_2 l'ensemble des autres règles de G, pour les i compris entre 1 et p on considère des vocabulaires N_i disjoints de N, disjoints entre eux et en bijection avec N par les applications $\tau_i : N_i \to N$. On définit alors une bigrammaire

$$G' = \left(N \cup \left(\bigcup_{i=1}^{p} N_i\right), T, \mapsto, X\right) \text{ équivalente à } G,$$

en posant

$$(A \to r) \in P_2 \quad \Rightarrow \quad A \mapsto r,$$

$$(A_i \to a_i \times (A_i^1 + A_i^2 + r_i) + B_i) \in P_1 \quad \Rightarrow \quad A_i \mapsto a_i \times (\tau_i^{-1}(A_i^1) + r_i) + B_i \text{ et}$$

$$\tau_i^{-1}(A_i) \mapsto a_i \times (\tau_i^{-1}(A_i^1) + r_i) + \tau_i^{-1}(B_i),$$

$$(A \to r + B) \in P_2 \quad \Rightarrow \quad \tau_i^{-1}(A) \mapsto r + \tau_i^{-1}(B),$$

$$A \mapsto \Lambda \quad \Rightarrow \quad \tau_i^{-1}(A) \to A_i^2.$$

On obtient alors une grammaire G' telle que $k_{G'} = k_G - 1$ et pour laquelle il n'est pas très difficile de montrer qu'elle est équivalente à G.

COROLLAIRE 3.1.2 : *Pour toute grammaire parenthésée il existe une grammaire équivalente dont les règles sont de la forme*

$$A \to a B \bar{a} C \qquad ou \qquad A \to \Lambda.$$

3.2. Lien entre les deux définitions des grammaires parenthésées

Dans ce paragraphe on caractérise de manière effective les grammaires parenthésées définies en 2.1.3 qui sont équivalentes aux grammaires parenthésées classiques définies en 1.2.1.

DÉFINITION 3.2.1 : Soit $G = (N, T, \to, X)$ une bigrammaire régulière. On associe à cette bigrammaire la grammaire régulière $G_\rho = (N, T, \mapsto, X)$ telle que

$$A \mapsto \alpha \quad \Leftrightarrow \quad (\exists r) \quad [A \to r \text{ et } \rho(r) = \alpha].$$

On dit que G est non récursive à l'horizontale (en abrégé NRH) si et seulement si tous les langages

$$L(G_\rho, A) = \left\{ \alpha ;\ A \overset{*}{\underset{G_\rho}{\succ\!\!-}} \alpha \ et\ \alpha \in T^* \right\}$$

sont finis.

REMARQUE : 1° On montre facilement que $L(G_\rho, A)=\rho(B(G, A))$.

2° Le fait que G soit NRH est décidable car cela signifie simplement que G_ρ n'est pas récursive à droite.

THÉORÈME 3.2.2 : *Les langages engendrés par les grammaires parenthésées au sens classique sont les représentations parenthétiques des bilangages engendrées par les bigrammaires régulières* NRH *et dont les règles ayant comme premier membre l'axiome sont de la forme* $X \to a \times r$ *avec* $a \in T$.

Démonstration : Si G est une grammaire parenthésée au sens classique, alors pour chaque A de N on a $L(G_\rho, A)$ réduit à des mots d'une lettre dans la bigrammaire régulière correspondante.

Réciproquement, soit G une bigrammaire régulière vérifiant les conditions du théorème, on peut remarquer que la construction de la réduite de G ne change pas son caractère NRH. Donc supposons G réduite (pour la suite de la démonstration il suffit d'ailleurs d'aller jusqu'au 3ᵉ pas de la réduction). Soit $A \to r$ une règle de G. Soit $r_1, r_2, \ldots, r_k$ les ramifications qui dérivent de r en appliquant des réécritures uniquement sur la racine et tels que $\rho(r_i)\in T^*$; l'hypothèse « G est NRH » est nécessaire et suffisante pour que ces ramifications soient en nombre fini. On obtient alors une bigrammaire équivalente en remplaçant chaque règle $A \to r$ par les règles $A \to r_1 \mid r_2 \mid \ldots \mid r_k$. Numérotons ces nouvelles règles sauf celles de premier membre X et celles de second membre Λ. Soit $A_i \to r_i$ la règle de numéro i avec $r_i = a_1^i \times r_1^i + \ldots + a_{k_i}^i r_{k_i}^i$. On introduit, pour chaque i, k_i nouveaux terminaux $A_1^i \ldots A_{k_i}^i$ et on remplace G par la bigrammaire

$$G'=(\{X\} \cup \bigcup_i \{A_1^i, \ldots, A_{k_i}^i\}, T, \rightsquigarrow, X)$$

telle que

$$X \rightsquigarrow r' \iff (\exists r), \ [X \to r \text{ et } r' \in h(r)],$$

$$A_j^i \rightsquigarrow r' \iff r' \in a_j^i \times h(r_j^i),$$

où h est la fonction de $\widehat{N \cup T}$ dans $\overline{\mathscr{P}(\{X\} \cup (\bigcup_i \{A_1^i \ldots A_{k_i}^i\} \cup T)}$ définie par

$$h(\Lambda)=\Lambda,$$

$$h(x \times r' + r'') = \text{si } x \in T \text{ alors } x \times h(r') + h(r'')$$

$$\text{sinon si } x \to \Lambda \text{ alors } h(r'') \cup \bigcup_{\substack{i \text{ tel que} \\ A_i = x}} A_1^i + \ldots + A_{k_i}^i + h(r'')$$

$$\text{sinon } \bigcup_{\substack{i \text{ tel que} \\ A_i = x}} A_1^i + \ldots + A_{k_i}^i + h(r'')$$

avec la convention

$$\bigcup_{\substack{i \text{ tel que} \\ A_i = X}} A_1^i \ldots A_{k_i}^i = X.$$

On obtient alors une bigrammaire régulière équivalente à G et qui correspond à une grammaire parenthésée au sens classique.

Exemple :

$$G=(\{X, A, B\}, \{a, b, c\}, \rightarrow, X),$$

avec

$$X \rightarrow a \times (A+B) \mid b \times B \mid a \times X,$$
$$A \rightarrow a \times A + B \mid \Lambda,$$
$$B \rightarrow c \times A \mid \Lambda.$$

On a

$$G_\rho=(\{X, A, B\}, \{a, b, c\}, \rightarrow, X),$$

avec

$$X \rightarrow a \mid b, \qquad A \mapsto a B \mid \Lambda, \qquad B \mapsto c \mid \Lambda.$$

Donc G_ρ n'est pas récursive à droite et donc G est NRH.

On obtient une bigrammaire équivalente à G en remplaçant les règles de G par

$$X \rightarrow a \times (A+B) \mid b \times B \mid a \times X,$$
$$A \rightarrow a \times A + c \times A \mid 1 \mid a \times A \mid 3 \mid \Lambda,$$
$$B \rightarrow c \times A \mid 3 \mid \Lambda,$$

la grammaire G' est alors

$$X \rightarrow a \mid a \times B_1^3 \mid a \times (A_1^1 + A_2^1) \mid a \times (A_1^1 + A_2^1 + B_1^3)$$
$$\mid a \times A_2^2 \mid a \times (A_2^2 + B_1^3) \mid b \mid b \times B_1^3 \mid a \times X,$$
$$A_1^1 \rightarrow a \mid a \times (A_1^1 + A_2^1) \mid a \times A_2^2,$$
$$A_2^1 \rightarrow c \mid c \times (A_1^1 + A_2^1) \mid c \times A_2^2,$$
$$A_2^2 \rightarrow a \mid a \times (A_1^1 + A_2^1) \mid a \times A_2^2,$$
$$B_1^3 \rightarrow c \mid c \times (A_1^1 + A_2^1) \mid c \times A_2^2.$$

En quelque sorte on peut dire que la généralisation que nous présentons ici des grammaires parenthésées et du même type que celle qui fait passer de l'étude des langages finis à l'étude des langages réguliers.

4. GRAMMAIRE PARENTHÉSÉE INVERSIBLE ET MINIMALE

4.1. Introduction, définitions et propriété des grammaires parenthésées inversibles

McNaughton [4] introduit la notion grammaire parenthésée inversible (backwards-deterministic parenthesis grammar). Une grammaire parenthésée est dite inversible si deux règles distinctes ont des seconds membres distincts. Dans la théorie classique, on déduit de cette définition qu'une grammaire inversible est non ambiguë. De plus, pour ce type de grammaire on peut commencer l'analyse syntaxique d'un mot par celle d'un facteur quelconque limité par deux parenthèses associées. Ces résultats ne se généralisent pas immédiatement à la forme générale des grammaires parenthésées. En effet, la non-ambïguité d'une grammaire inversible est due en particulier au fait qu'il est impossible, vu la forme particulière des règles dans le cas classique, de reconstruire un second membre de règle en utilisant d'autres règles. Ce n'est pas le cas avec les définitions données ici et même si on impose à une grammaire parenthésée générale de vérifier que des règles distinctes ont des seconds membres distincts, cette grammaire peut être ambiguë. Pour obtenir des résultats comparables à ceux de McNaughton il est donc nécessaire d'imposer une forme particulière au second membre des règles d'une grammaire parenthésée. Nous choisirons ici de prendre comme forme particulière $a\,A\,\bar{a}\,B$ ou Λ; cette forme semble particulièrement adaptée vu le théorème 3.1.1. On peut alors définir dans ce cadre une nouvelle notion de grammaire parenthésée inversible et obtenir des résultats comparables à ceux de McNaughton [4] mais plus puissants et plus généraux. En annexe on donne une définition plus générale de la notion de grammaire inversible qui recouvre ces deux cas.

DÉFINITION 4.1.1 : Soit $G=(N, T, \twoheadrightarrow, X)$ une bigrammaire régulière, on dit que G est inversible si elle vérifie les conditions suivantes :

(i) l'axiome X n'apparaît pas en partie droite des règles;

(ii) $$\begin{cases} (\forall A\in N),\ \ (\forall B\in N),\ \ (\forall \alpha\in \widehat{N\cup T}), \\ (A\twoheadrightarrow \alpha \text{ et } B\twoheadrightarrow \alpha \Rightarrow A=X \text{ ou } B=X \text{ ou } A=B); \end{cases}$$

(iii) G est réduite au sens du théorème 3.1.1.

Les conditions (i) et (ii) sont légèrement différentes de celles introduites par McNaughton [4] ceci est dû au fait que pour cet auteur une grammaire peut avoir plusieurs axiomes.

PROPOSITION 4.1.2 : *Soit* $G=(N, T, \twoheadrightarrow, X)$ *une bigrammaire régulière inversible alors* G *n'est pas ambiguë [c'est-à-dire que la grammaire parenthésée* $G'=(N, T, \overline{T}, \rightarrow, X)$ *n'est pas ambiguë].*

Démonstration : Soit P le langage de Dyck sur T et soit $g : P \rightarrow \mathscr{P}(N)$ défini par $g(\alpha)=\{A;\ A \overset{*}{\underset{G'}{\succ\!\!-}} \alpha\}$. Pour démontrer que G' est non ambiguë il suffit de montrer que

$$\left.\begin{array}{c}(\forall \alpha \in P),\quad (\exists A \in N),\\ {[g(\alpha)=\emptyset \text{ ou } g(\alpha)=\{A\} \text{ ou } g(\alpha)=\{A, X\}].}\end{array}\right\} \qquad \text{(I)}$$

Démontrons (I) par récurrence sur α.

Si $\alpha=\Lambda$; $g(\alpha)=\{A;\ A \in N \text{ et } A \rightarrow \Lambda\}$. D'après la condition (ii) de la définition 4.1.1, un seul non-terminal différent de X et éventuellement X vérifient $A \rightarrow \Lambda$. Donc (I) est vérifié pour $\alpha=\Lambda$. Supposons (I) vérifié pour $\alpha' \in P$ et $\alpha'' \in P$ et démontrons (I) pour $\alpha=a\,\alpha'\,\bar{a}\,\alpha''$. Si $g(\alpha)\neq\emptyset$, il existe une dérivation de la forme $A \underset{G}{\succ\!\!-} aB\bar{a}C \overset{*}{\underset{G}{\succ\!\!-}} \alpha$; donc $B \in g(\alpha')$ et $C \in g(\alpha'')$. D'après la condition (i) on a $B \neq X$ et $C \neq X$, donc B et C sont déterminés de manière unique grace à l'hypothèse de récurrence et donc on a A égal à X ou A déterminé de manière unique par a, B et C, d'où le résultat pour α.

REMARQUE : Soit $\alpha \in P$ de la forme $\alpha=\alpha'\,a\,\alpha''\,\bar{a}\,\alpha'''$ où a et $\bar{a}$ sont un couple de parenthèses associées, si α est engendré par G' alors il existe certainement une règle de G de la forme $A \rightarrow aB\bar{a}C$ telle que $B \overset{*}{\underset{G'}{\succ\!\!-}} \alpha''$. De plus on a $B \in g(\alpha'')$. Ceci signifie que pour faire l'analyse de α on peut commencer par celle de α'' qui nous donnera un non-terminal $B \neq X$ et faire ensuite l'analyse de $\alpha'\,a\,B\,\bar{a}\,\alpha''$. Dans le cadre des définitions de McNaughton [4], on peut aller un cran plus haut en trouvant B' tel que $B' \rightarrow aB\bar{a}$.

La suite du paragraphe va être consacrée à la construction de grammaires inversibles équivalentes à une grammaire donnée, les grammaires obtenues étant de plus en plus structurées jusqu'à obtenir une grammaire canoniquement associée au langage parenthésé considéré. Au passage nous démontrerons des résultats sur l'équivalence des bigrammaires régulières et sur le problème de l'ambiguïté des bigrammaires régulières.

4.2. Structure de T-semi-binoïde

DÉFINITION 4.2.1 : On appelle T-semi-binoïde un couple $(E, (f_a)_{a \in T})$ formé d'un ensemble E et de card (T) opérations binaires sur E en bijection avec T.

Exemple : Le T-binoïde universel $\hat{T}$ peut être muni d'une structure de T-semi-binoïde en posant

$$f_a(r, r') = a \times r + r',$$

quand on parlera du T-semi-binoïde $\hat{T}$ c'est à cette structure que l'on fera référence.

DÉFINITION 4.2.2 : Si E et E' sont deux T-semi-binoïdes, un homomorphisme de E dans E' est une application $h : E \to E'$ telle que

$$h(f_a(x, y)) = f'_a(h(x), h(y))$$

pour tout a de T et tout x et y de E.

PROPOSITION 4.2.3 : *Le T-semi-binoïde $\hat{T}$ est libre de base $\{\Lambda\}$ c'est-à-dire que si E est un T-semi-binoïde et si e est un élément quelconque de E alors il existe un et un seul homomorphisme h de $\hat{T}$ dans E tel que $h(\Lambda) = e$.*

Démonstration : Cela est une trivialité en effet h doit vérifier

$$h(\Lambda) = e \qquad \text{et} \qquad h(a \times r + r') = f_a(h(r), h(r'))$$

et il existe une et une seule application de $\hat{T}$ dans E vérifiant ces conditions.

DÉFINITION 4.2.4 : Soit $G = (N \cup \{X\}, T, \twoheadrightarrow, X)$ une bigrammaire régulière, on dit que G est associée à un T-semi-binoïde si et seulement si :

- $X \notin N$;
- X n'intervient pas en partie droite d'une règle;
- G est réduite;
- N a une structure de T-semi-binoïde pour la famille de la loi f_a;
- $$(\forall A \in N, \quad (\forall a \in T), \quad (\forall B \in N), \quad (\forall C \in N),$$

$$[A \twoheadrightarrow a \times B + C \Leftrightarrow A = f_a(B, C)];$$

- $(\exists!\ A \in N)(A \to \Lambda)$.

On peut remarquer qu'une telle bigrammaire est en particulier inversible.

PROPOSITION 4.2.5 : *Pour toute bigrammaire* $G = (N', T, \mapsto, X')$ *il existe une bigrammaire $G' = (N \cup \{X\}, T \twoheadrightarrow, X)$ équivalente à G et associée à un T-semi-binoïde.*

Démonstration : On peut supposer G réduite, prenons alors $N=\mathscr{P}(N')$ et posons

$$f_a(E, E')=\{A;\ A\in N' \text{ et } (\exists B\in E), (\exists C\in E'), (A\mapsto\!\!\!\rightarrow a\times B+C)\}.$$

On munit ainsi N d'une structure de T-semi-binoïde.

Soit

$$E_0=\{A;\ A\in N' \text{ et } A\mapsto\!\!\!\rightarrow\Lambda\}.$$

Soit

$$\mathscr{E}=\{E;\ E\in N \text{ et } X'\in E\}.$$

Prenons pour G' la grammaire définie pour

$$\begin{aligned} X\to a\times E+E' &\iff f_a(E, E')\in\mathscr{E},\\ E\to a\times E'+E'' &\iff f_a(E', E'')=E,\\ E\to\Lambda &\iff E=E_0. \end{aligned}$$

De façon évidente G' est associée à un T-semi-binoïde et il faut montrer maintenant que G' est équivalente à G. Pour cela il suffit de remarquer que

$$(\forall r\in\hat{T}),\quad [E \underset{G'}{\succ\!\!-} r \iff E=\{A;\ A\in N \text{ et } A \underset{G}{\succ\!\!-} r\}],$$

équivalence qui se démontre trivialement par récurrence sur r.

Commentaire : Cette démonstration est très proche de celle donnée par McNaughton [4] de l'existence d'une grammaire parenthésée inversible équivalente à une grammaire parenthésée donnée, l'idée nouvelle et d'introduire la notion de structure algèbrique qui va nous conduire à des recherches de structure algébrique minimale. On trouve dans [12] une démonstration tout à fait similaire mais avec des définitions légèrement différentes. Cette notion de T-semi-binoïde n'est pas très loin non plus de la notion d'automate d'arbre travaillant des feuilles vers la racine étudiée par Thatcher [13].

COROLLAIRE 4.2.6 : *Un bilangage sur T est régulier si et seulement s'il existe un T-semi-binoïde fini N, une partie N_1 de N et un homomorphisme de T-semi-binoïde $h : \hat{T}\to N$ tel que $h^{-1}(N_1)=L$.*

Démonstration : Si $L=h^{-1}(N_1)$ alors L est engendré par la bigrammaire régulière $G'=(N \cup \{X\}, T, \rightarrow, X)$ définie par

$$X \rightarrow a \times A+B \quad \Leftrightarrow \quad f_a(A, B) \in N_1,$$

$$A \rightarrow a \times B+C \quad \Leftrightarrow \quad f_a(B, C)=A,$$

$$A \rightarrow \Lambda \quad \Leftrightarrow \quad h(\Lambda)=A.$$

il est en effet facile de démontrer par récurrence que

$$(\forall r \in \hat{T}), \quad (A \rightarrow r) \quad \Leftrightarrow \quad (h(r)=A).$$

Réciproquement si L est engendré par une bigrammaire régulière alors il existe une bigrammaire régulière $G'=(N \cup \{X\}, T, \twoheadrightarrow, X)$ engendrant L et associé à un T-semi-binoïde. Soit $h : \hat{T} \rightarrow N$ l'homomorphisme de T-semi-binoïde tel que $h(\Lambda)=A$ où A est l'unique élément de N tel que $A \twoheadrightarrow \Lambda$. Il est alors facile de montrer que $L=h^{-1}(N_1)$ où

$$N_1=\{A; (\exists a, B, C), (X \rightarrow a \times B+C) \text{ et } A=f_a(B, C)\}.$$

COROLLAIRE 4.2.7 [10] : *La classe des bilangages réguliers est stable par les opérations booléennes et les holomorphismes inverses.*

Démonstration : Il suffit de calquer la démonstration faite pour les langages réguliers considérés comme image inverse de parties de monoïdes finis.

COROLLAIRE 4.2.8 [9] : *L'égalité, l'inclusion sont des problèmes décidables pour les bilangages réguliers.*

Démonstration : Si L et L' sont des bilangages réguliers, on sait construire de manière effective une bigrammaire engendrant la différence symétrique $L \Delta L'$ et une autre engendrant $L' \cap \complement_{\hat{T}} L$.

Or

$$L=L' \quad \Leftrightarrow \quad L \Delta L'=\emptyset,$$

$$L' \subset L \quad \Leftrightarrow \quad L' \cap \complement_{\hat{T}} L=\emptyset.$$

Donc ces deux problèmes sont décidables.

Ce corollaire a été démontré dans un cas particulier par McNaughton [4] et dans un cas plus général par Paull et Unger [11], ce problème a été repris dans toute sa généralité et en terme de bilangage par Pair [9, 8]. Dans [8] on trouve une autre application astucieuse des bilangages aux morphismes de grammaire. Remarquons que cette façon d'aborder le problème de l'égalité de deux bilangages réguliers est beaucoup plus efficace que celle rencontrée dans la littérature. En effet, en général on traite ce problème en construisant une grammaire canonique associée à un langage parenthésé donné [4] qui n'est pas nécessaire pour ce problème. Par exemple, pour comparer deux langages réguliers il n'est pas nécessaire de calculer les automates minimaux de ces deux langages.

COROLLAIRE 4.2.9 : *Le problème de l'ambiguïté est décidable pour les bigrammaires régulières.*

Démonstration : Les réductions conduisant à la forme réduite énoncée au théorème 3.1.1 ne changent pas le caractère ambigu ou non d'une bigrammaire sauf peut être la toute première réduction qui consiste à supprimer les règles de la forme $A \;\rightarrowtail\; B$ et qui peut supprimer l'ambiguïté d'une grammaire. Mais si l'ambiguïté provient de cette réduction elle est facile à tester directement. Donc on peut se contenter de traiter le cas où $G=(N,\ T,\ \twoheadrightarrow,\ X)$ est sous forme réduite.

Pour chaque $A \in N$ notons $\mathscr{B}(A)=\{r;\ A \overset{*}{\underset{G}{\twoheadrightarrow}} r$ et $r \in \hat{T}\}$. Pour que G soit ambigü il faut et il suffit que G vérifie

$$(\exists a \in T),\quad (\exists A),\quad (\exists B),\quad (\exists B'),\quad (\exists C),\quad (\exists C'),$$

$$[A \twoheadrightarrow a \times B + C \text{ et } A \twoheadrightarrow a \times B' + C' \text{ et } a \times B + C \neq a \times B' + C'$$
$$\text{et } (a \times \mathscr{B}(B) + \mathscr{B}(C)) \cap (a \times \mathscr{B}(B') + \mathscr{B}(C')) \neq \emptyset].$$

Or la vérification de cette condition peut se faire de manière effective car elle revient à tester si des bilangages réguliers sont vides ou non.

Ce corollaire a été énoncé et démontré dans [11] à propos des grammaires parenthésées classiques à un seul type de parenthèse. La technique de démonstration proposée dans [11] est beaucoup plus lourde quelle celle mise en évidence ici.

4.3. Structure de T-binoïde, bigrammaire minimale associée

Nous allons déjà voir dans ce paragraphe que l'on peut améliorer le résultat de la proposition 4.2.5 en introduisant des bigrammaires associées à des T-binoïdes.

DÉFINITION 4.3.1 : Soit $G=(N \cup \{X\},\ T,\ \twoheadrightarrow X)$ une bigrammaire régulière, on dit que G est associé à un T-binoïde si et seulement si G vérifie les conditions suivantes :

- $X \notin N$ et X n'intervient pas en partie droite d'une règle;
- G est réduite;

— N a une structure de T-binoïde, l'élément neutre de $+$ étant E;

— $(\forall A \in N)$, $(\forall a \in T)$, $(\forall B \in N)$, $(\forall C \in N)$,

$$(A \twoheadrightarrow a \times B + C \Leftrightarrow A = a \times B + C);$$

— $A \to \Lambda \Leftrightarrow A = E$.

Proposition 4.3.2 [10] : *Un bilangage L sur T est régulier si et seulement si il existe un T-binoïde fini B et une partie B' de B telle que pour l'homomorphisme ψ : $\hat{T} \to B$ on ait*

$$L = \psi^{-1}(B').$$

Démonstration : Si $L = \psi^{-1}(B')$, on peut définir sur B une structure de T-semi-binoïde en posant $f_a(b, b') = a \times b + b'$. Soit $h : \hat{T} \to B$ l'homomorphisme de T-semi-binoïde défini par $h(\Lambda) = e$ (e est l'élément neutre de B pour $+$), alors il est facile de montrer par récurrence que $h = \psi$. Donc

$$L = h^{-1}(B').$$

Si L est régulier, d'après le corollaire 4.2.6 il existe un T-semi-binoïde B, h : $\hat{T} \to B$ et $B' \subset B$ tel que $L = h^{-1}(B')$. Soit $B_1 = B^B$ l'ensemble des applications de B dans B, soit $e = h(\Lambda)$, on définit dans B_1 la structure de T-binoïde suivante :

$$F + G = F \circ G \quad \text{(composition des applications)},$$

$$(a \times F)(b) = f_a(F(e), b).$$

Il existe alors un unique homomorphisme de T-binoïde $\psi : \hat{T} \to B_1$. Montrons que $\psi(r)(e) = h(r)$ en faisant une récurrence sur r :

$$\psi(\Lambda)(e) = e = h(\Lambda),$$

$$\begin{aligned}\psi(a \times r + r')(e) &= (a \times \psi(r) + \psi(r'))(e) = (a \times \psi(r)) \circ \psi(r')(e)\\ &= (a \times \psi(r))(h(r')) = f_a(\psi(r)(e), h(r')) = f_a(h(r), h(r')) = h(a \times r + r').\end{aligned}$$

Soit $B'_1 = \{F, F(e) \in B'\}$, on obtient alors $L = \psi^{-1}(B'_1)$.

Cette proposition montre en particulier que la définition des bilangages réguliers donnée ici est compatible avec celle donnée dans [9, 10, 12].

Corollaire 4.3.3 : *Pour toute bigrammaire $G = (N, T, \twoheadrightarrow, X)$ il existe une bigrammaire $G = (N \cup \{X\}, T, \twoheadrightarrow, X)$ équivalente à G' et associée à un T-binoïde.*

Démonstration : Soit L le bilangage engendré par G'. Soit B un T-binoïde tel que $L=\psi^{-1}(B')$ avec $B'\subset B$ et $\psi : \hat{T}\to B$, alors L est engendré par la bigrammaire $G=(B\cup\{X\}, T, \twoheadrightarrow, X)$ avec

$$X\to a\times s+s' \quad\Leftrightarrow\quad s\in B \text{ et } s'\in B \text{ et } a\times s+s'\in B',$$

$$s\to a\times s'+s'' \quad\Leftrightarrow\quad s\in B \text{ et } s'\in B \text{ et } s''\in B \text{ et } a\times s'+s''=s,$$

$$s\to\Lambda \quad\Leftrightarrow\quad s=e=\text{l'élément neutre de } B \text{ pour } +.$$

Nous allons montrer que parmi les T-binoïdes associés à un bilangage régulier il y en a un de minimal qui est entièrement déterminé par le bilangage régulier.

DÉFINITION 4.3.4 : Soit L un bilangage sur T. On appelle contexte pour L l'application $C_L : \hat{T}\to\mathscr{P}(T[\mathbf{1}])$ définie par

$$C_L(r)=\{t;\ t\in\hat{T}[\mathbf{1}] \text{ et } t(r)\in L\}.$$

Cette notion a été introduite dans [4] sous un aspect légèrement différent et a été réintroduite indépendamment sous cette forme dans [5].

PROPOSITION 4.3.5 : *L'ensemble $\mathscr{C}_L=\{C_L(r);\ r\in\hat{T}\}$ peut être muni d'une structure de T-binoïde telle que C_L soit l'homomorphisme de $\hat{T}$ dans $\mathscr{C}_L$. De plus on a*

$$L=D_L^{-1}(\mathscr{C}'_L) \qquad \textit{avec} \quad \mathscr{C}'_L=\{C_L(r);\ \mathbf{1}\in C_L(r)\}.$$

Démonstration : Pour la première partie de la démonstration, il suffit de montrer que :

$$\text{(i)} \qquad \left\{\begin{array}{l} C_L(r)=C_L(r') \\ \text{et} \\ C_L(s)=C_L(s') \end{array}\right. \quad\Rightarrow\quad C_L(r+s)=C_L(r'+s');$$

$$\text{(ii)} \qquad C_L(r)=C_L(r') \quad\Rightarrow\quad C_L(a\times r)=C_L(a\times r').$$

Montrons (i) :

$$\begin{aligned} t\in C_L(r+s) &\Leftrightarrow t(r+s)\in L \Leftrightarrow t(\mathbf{1}+s)(r)\in L \\ &\Leftrightarrow t(\mathbf{1}+s)\in C_L(r) \Leftrightarrow t(\mathbf{1}+s)\in C_L(r') \\ &\Leftrightarrow t(r'+s)\in L \Leftrightarrow t(r'+\mathbf{1})\in C_L(s) \\ &\Leftrightarrow t(r'+\mathbf{1})\in C_L(s') \Leftrightarrow t\in C_L(r'+s'). \end{aligned}$$

Même type de démonstration pour (ii).

La deuxième partie est triviale.

PROPOSITION 4.3.6 : *Soit B un T-binoïde et* $L=\psi^{-1}(B')$ *avec B' contenu dans B et* $\psi : \hat{T} \to B$. *Alors il existe un homomorphisme surjectif* $\chi : \psi(\hat{T}) \to \mathscr{C}_L$ *tel que*

$$(\forall r \in \hat{T}), \quad (C_L(r)=\chi(\psi(r)) \qquad et \qquad \mathscr{C}'_L=\chi(B' \cap \psi(\hat{T})).$$

Démonstration : Pour trouver χ il suffit de démontrer que

$$\psi(r)=\psi(r') \quad \Rightarrow \quad C_L(r)=C_L(r').$$

Pour cela il suffit de démontrer que

$$\psi(r)=\psi(r') \quad \Rightarrow \quad (\forall h \in \hat{T}[\mathbf{1}]) \quad , [\psi(h(r))=\psi(h(r'))],$$

ce qui est trivial pour récurrence sur h. Le reste des vérifications est trivial.

THÉORÈME 4.3.7 : *Un bilangage L est régulier si et seulement si l'ensemble* $\mathscr{C}_L$ *est fini.*

Démonstration : Cela résulte immédiatement des propositions 4.3.5 et 4.3.6.

DÉFINITION 4.3.8 : Si L est régulier, $\mathscr{C}_L$ est le T-binoïde minimal associé à L; de plus la grammaire $G=(\mathscr{C}_L \cup \{X\}, T, \twoheadrightarrow, X)$ avec

$$X \to a \times C + C' \quad \Leftrightarrow \quad a \times C + C' \in \mathscr{C}'_L,$$

$$C \to a \times C' + C'' \quad \Leftrightarrow \quad a \times C' + C'' = C,$$

$$C \to \Lambda \quad \Leftrightarrow \quad C = C_L(\Lambda).$$

est appelée grammaire minimale de L.

PROPOSITION 4.3.9 : *Le T-binoïde minimal associé à L peut être construit de manière effective à partir d'une bigrammaire régulière engendrant L.*

Démonstration : Tout au long des démonstrations les constructions ont été faites de manière effective sauf la construction de $\mathscr{C}_L$. Donc à partir d'une bigrammaire régulière engendrant L, il est possible de construire de manière effective un T-binoïde B et une partie B' de B tel que $L=\psi^{-1}(B')$ avec $\psi : \hat{T} \to B$. Commençons par construire $\psi(\hat{T})$. Or $\psi(\hat{T})$ est le sous-ensemble de B formé des éléments que l'on peut obtenir à partir de e en appliquant la loi externe.

Posons

$$B_0=\{e\},$$

$$B_{n+1}=(a \times B_n + B_n) \cup B_n.$$

On obtient immédiatement

$$\psi(\hat{T})=\bigcup_{n \geqq 0} B_n.$$

Mais comme B est fini, on a, pour le premier n_0 tel que $B_{n_0}=B_{n_0+1}$, $\psi(\hat{T})=B_{n_0}$. On ne change rien au problème en supposant que B est égal à B_{n_0} et que B' est égal à $B'\cap B_{n_0}$. Il faut maintenant construire dans B la relation d'équivalence

$$b\sim b' \quad\Leftrightarrow\quad \chi(b)=\chi(b'),$$

où $\chi : B\to\mathscr{C}_L$ est l'homomorphisme tel que $\chi\circ\psi=C_L$. L'ensemble quotient $B/\sim$ est alors isomorphe au binoïde minimal de L. On obtient :

$$\begin{aligned}
b\sim b' \quad&\Leftrightarrow\quad (\exists\, r\in\hat{T}),\quad (\exists\, s\in\hat{T}),\quad [\psi(r)=b \text{ et } \psi(s)=b' \text{ et } \chi\circ\psi(r)=\chi\circ\psi(s)]\\
&\Leftrightarrow\quad (\exists\, r\in\hat{T}),\quad (\exists\, s\in\hat{T}),\quad [\psi(r)=b \text{ et } \psi(s)=b' \text{ et } C_L(r)=C_L(s)]\\
&\Leftrightarrow\quad (\exists\, r\in\hat{T}),\quad (\exists\, s\in\hat{T}),\\
&\qquad [\psi(r)=b \text{ et } \psi(s)=b' \text{ et } (\forall\, t\in\hat{T}[\mathbf{1}]\ (t(r)\in L\Leftrightarrow t(s)\in L)]\\
&\Leftrightarrow\quad (\exists\, r\in\hat{T}),\quad (\exists\, s\in\hat{T}),\\
&\qquad [\psi(r)=b \text{ et } \psi(s)=b' \text{ et } (\forall\, t\in\hat{T}[\mathbf{1}])(\psi(t(r)\in B'\Leftrightarrow\psi(t(s))\in B')].
\end{aligned}$$

Or si on appelle t_B l'application de B dans B associée au polynôme t on a

$$(\forall\, r\in\hat{T}),\quad (\forall\, t\in\hat{T}[\mathbf{1}]),\quad (\psi(t(r))=t_B(\psi(r))).$$

(Démonstration facile par récurrence sur t et en utilisant le fait que ψ est un homomorphisme.)

On en déduit donc que

$$b\sim b' \quad\Leftrightarrow\quad (\forall\, h\in\hat{T}[\mathbf{1}]),\quad [h_B(b)\in B'\Leftrightarrow h_B(b')\in B'].$$

Notons L_b le bilangage sur $T\cup\{\mathbf{1}\}$ défini par

$$L_b=\{h;\ h\in\hat{T}[\mathbf{1}] \text{ et } h_B(b)\in B'\}.$$

On a donc $b\sim b'\Leftrightarrow L_b=L_{b'}$. Pour montrer que cette relation d'équivalence peut se calculer de manière effective, il suffit donc maintenant de montrer que L_b est un bilangage régulier.

Considérons les trois $(T\cup\{1\})$ binoïdes B_1, B_2, B_3 définis par :

1° $B_1=B\cup\{\omega\}$ où ω est un symbole qui n'est pas dans B et les lois de B_1 sont définies par les tables de la figure 4. Dans ces tables b' et b'' désignent des éléments quelconques de B, e est l'élément neutre de B, a est quelconque dans T, b est l'élément de B pour lequel on calcule L_b, les lois $+$ et $\times$ utilisées dans les tables sont celles de B.

$+$	b'	ω
b''	$b'+b''$	ω
ω	ω	ω

$\times$	a	1
e	$a\times e$	b
$b'\neq e$	$a\times b'$	ω
ω	ω	ω

Figure 4.

2° $B_2=\{\perp, \top, \omega\}$ les lois de B_2 sont définies par les tables de la figure 5.

$+$	$\perp$	$\top$	ω
$\perp$	$\perp$	$\top$	ω
$\top$	$\top$	$\top$	ω
ω	ω	ω	ω

$\times$	a	1
$\perp$	$\top$	$\top$
$\top$	$\top$	ω
ω	ω	ω

Figure 5.

3° $B_3=\{\perp, \top, \omega\}$ les lois de B_3 sont définies par les tables de la figure 6.

$+$	$\perp$	$\top$	ω
$\perp$	$\perp$	$\top$	ω
$\top$	$\top$	ω	ω
ω	ω	ω	ω

$\times$	a	1
$\perp$	$\perp$	$\top$
$\top$	$\top$	ω
ω	ω	ω

Figure 6.

Soit $\psi_1 : \widehat{T\cup\{1\}} \to B_1, \psi_2 : \widehat{T\cup\{1\}} \to B_2, \psi_3 : \widehat{T\cup\{1\}} \to B_3$. Il est facile de démontrer par récurrence que

$$t\in \hat{T}[1] \quad\Rightarrow\quad \psi_1(t)=t(b),$$
$$t\in \hat{T} \quad\Rightarrow\quad \psi_1(t)=\psi(t),$$

$$\psi_2^{-1}(\{\top\})=\{t;\ t\in \widehat{T\cup\{1\}} \text{ et 1 n'apparaît qu'aux feuilles de } t\};$$

$$\psi_3^{-1}(\{\top\})=\{t;\ t\in \widehat{T\cup\{1\}} \text{ et 1 a une seule occurrence dans } t\}.$$

On en déduit donc que

$$L_b=\psi_1^{-1}(B')\cap\psi_2^{-1}(\{\top\})\cap\psi_3^{-1}(\{\top\}),$$

et donc que L_b est un bilangage régulier comme intersection de bilangage régulier (*cf.* prop. 4.1.7).

THÉORÈME 4.3.10 : *La bigrammaire minimale de L qui est canonique parmi les bigrammaires associées à des T-binoïde et engendrant L se construit de manière effective à partir d'une bigrammaire régulière quelconque engendrant L.*

Démonstration : Résulte immédiatement de la proposition 4.3.9.

5. QUELQUES APPLICATIONS A LA THÉORIE DES LANGAGES ALGÉBRIQUES

Nous allons voir trois applications à la théorie des langages, la première n'est pas nouvelle et n'est traitée que pour montrer la puissance du théorème 3.1.1. La seconde est originale à notre connaissance et caractérise de manière effective parmi les langages parenthésés ceux qui sont des intersections d'un langage de Dyck et d'un langage régulier. La dernière application éclaire quelque peu les problèmes d'ambiguïtés inhérentes, le résultat démontré ici peut se déduire des résultats de McNaughton [4] en généralisant quelque peu les résultats de cet auteur, nous donnerons ici une démonstration fondée sur la théorie des bilangages réguliers. J. Thatcher [13] aborde aussi des problèmes similaires.

5.1. Application aux réduites de Greibach

On démontre que toute grammaire algébrique est équivalente à une grammaire $G=(N, T, \rightarrow, X)$ où les règles sont de la forme $A \rightarrow a\alpha \mid \Lambda$ avec $a \in T$ et $\alpha \in N^*$. Cette forme réduite est appelée réduite de Greibach. Il est bien connu que l'on peut améliorer cette réduite en imposant en plus que $A \rightarrow a\alpha \Rightarrow |\alpha|=2$. On va donner une démonstration utilisant le théorème 3.1.1 du passage de la forme générale à la forme améliorée des réduites de Greibach.

Démonstration : Soit $G=(N, T, \rightarrow, X)$ une grammaire sous forme réduite de Greibach. Associons à G la grammaire parenthésée $G'=(N, T, \overline{T}, \mapsto, X)$ avec $\mapsto$ défini par

$$A \mapsto a\alpha\bar{a} \quad \Leftrightarrow \quad A \rightarrow a\alpha,$$

$$A \mapsto \Lambda \quad \Leftrightarrow \quad A \rightarrow \Lambda.$$

Il est évident que $L(G)=h(L(G'))$ avec $h:(T \cup \overline{T})^* \rightarrow T^*$ défini par $h(a)=a$ et $h(\bar{a})=\Lambda$. D'après le théorème 3.1.1, G' est équivalente à une grammaire $G''=(N', T, \overline{T}, \rightsquigarrow, X)$ dont les règles sont de la forme $A \rightsquigarrow a\, B\, \bar{a}\, C \mid \Lambda$ on en déduit donc que $L(G)=h(L(G''))$ est engendré par la grammaire $G'''=(N', T, \overline{T}, \twoheadrightarrow, X)$ où $\twoheadrightarrow$ est défini par

$$A \twoheadrightarrow aBC \quad \Leftrightarrow \quad A \rightsquigarrow a\, B\, \bar{a}\, C,$$

$$A \twoheadrightarrow \Lambda \quad \Leftrightarrow \quad A \rightsquigarrow \Lambda.$$

D'où l'existence de la réduite de Greibach améliorée.

5.2. Caractérisation des langages parenthésés qui sont des intersections de langage de Dyck avec des langages réguliers.

On rappelle déjà les liens existants entre les arbres syntaxiques engendrés par une grammaire algébrique et les bilangages réguliers. Ces rappels serviront aussi dans le paragraphe 5.3.

DÉFINITION 5.2.1 [10] : On appelle grammaire algébrique généralisée un quadruplet $G=(N, T, \rightarrow, X)$ tel que X est un langage régulier sur $N \cup T$ et $\rightarrow$ est une relation entre N et $(N \cup T)^*$ telle que $K_A=\{\alpha;\ A \rightarrow \alpha\}$ est un langage régulier pour chaque A de N.

On démontre sans difficulté que ces grammaires algébriques généralisées engendrent les langages algébriques et eux seuls.

DÉFINITION 5.2.2 [10] : On appelle bilangage engendrée par une grammaire algébrique généralisée $G=(N, T, \rightarrow, X)$ le bilangage

$$B(G)=\{r;\ r \in \widehat{N \cup T} \text{ et } \rho(r) \in X \text{ et } (\forall A \in N), (\forall \alpha \in F_A(r))(A \rightarrow \alpha)\}.$$

Il est facile de voir que le langage engendré par G est l'ensemble des mots des feuilles des ramifications de $B(G)$.

On démontre alors facilement les deux résultats suivants.

PROPOSITION 5.2.3 [10] : *Soit L un bilangage régulier sur T, il existe une grammaire algébrique généralisée $G=(N', T', \rightarrow, X)$* et *une transcription $\tau : \widehat{N' \cup T'} \rightarrow \hat{T}$ telles que $\tau(B'(G))=L$.*

PROPOSITION 5.2.4 [7] : *Soit L un bilangage régulier sur T, parmi les grammaires algébriques $G=(N, T, \rightarrow, X)$ telles que $\tau(B(G))=L$, il existe des grammaires telles que si $P(G)$ est la représentation parenthétique de $B(G)$ alors $P(G)$ est l'intersection du langage de Dyck sur $N \cup T$ et d'un langage régulier sur $N \cup T \cup \overline{N \cup T}$.*

On peut se demander si en fait les bilangages réguliers sont aussi de la forme $P \cap K$. Cette opinion semble confirmée par la forme réduite des bigrammaires régulières énoncées au théorème 3.1.1.

Malgré cela, cette première impression est fausse comme le montre l'exemple suivant :

Exemple : Le langage $L=\{a^n\, a^n\, \bar{a}^p\, \bar{a}^p;\ n \geqq 0,\ p \geqq 0\}$ est un langage parenthésé qui n'est pas de la forme $P \cap K$. La démonstration de ce fait sera donnée après le théorème 5.2.5 à titre d'exemple d'utilisation de ce théorème. Le lecteur pourra facilement faire une démonstration directe de ce fait.

THÉORÈME 5.2.5 : *Soit L un bilangage régulier. Une représentation parenthésée de L est de la forme $P \cap K$ si et seulement si le T-binoïde minimal $\mathscr{C}_L$ de L vérifie*

$$(\exists \theta : T \to \mathscr{C}_L), \quad (\exists \bar{\theta} : T \to \mathscr{C}_L), \quad (\forall a \in T), \quad (\forall C \in \mathscr{C}_L),$$
$$[a \times C = \theta(a) + C + \bar{\theta}(a)].$$

En particulier il est décidable de savoir si un langage parenthésé est de la forme $P \cap K$.

Démonstration : La deuxième affirmation est évidemment une conséquence de la première.

Soit L un bilangage régulier et L' le langage parenthésé associé, supposons L' de la forme $P \cap K$. Soit M un monoïde fini, $\psi : (T \cup \bar{T})^* \to M$ et $M' \subset M$ tel que $K = \psi^{-1}(M')$.

L' est engendré par la grammaire parenthésée $G = (X \cup M, T, \bar{T}, \to, X)$ telle que

$$X \to as'\bar{a}s'' \quad \Leftrightarrow \quad \psi(a) + s' + \psi(\bar{a}) + s'' \in M',$$
$$s \to as'\bar{a}s'' \quad \Leftrightarrow \quad \psi(a) + s' + \psi(\bar{a}) + s'' = s,$$
$$s \to \Lambda \quad \Leftrightarrow \quad s = e \quad (e \text{ est l'élément neutre de } M).$$

Donc si on munit M de la structure de T-binoïde définie par $a \times s = \psi(a) + s + \psi(\bar{a})$ la loi $+$ étant la même que dans M, on démontre facilement que $L = \psi_1^{-1}(M)$ où $\psi_1 : \hat{T} \to M$ et l'homomorphisme de T-binoïde de $\hat{T}$ dans B. Comme le T-binoïde minimal de L est homomorphe à tout T-binoïde permettant de reconnaître L, on en déduit le théorème dans ce sens.

Réciproquement si le T-binoïde minimal $\mathscr{C}_L$ de L vérifie la condition du théorème, soit $\mathscr{C}' \subseteq \mathscr{C}_L$ et $\psi_0 : \hat{T} \to \mathscr{C}_L$ tel que $L = \psi^{-1}(\mathscr{C}')$, considérons la structure de monoïde de $\mathscr{C}_L$, soit $\psi : (T \cup \bar{T})^* \to \mathscr{C}_L$ l'homomorphisme de monoïde tel que $\psi(a) = \theta(a)$ et $\psi(\bar{a}) = \bar{\theta}(a)$. Il est facile de montrer que $L' = P \cap \psi^{-1}(\mathscr{C}')$.

Exemple : Reprenons le bilangage L associé au langage parenthésé $L' = \{ a^n \bar{a}^n a^p \bar{a}^p \}$. Dans $\hat{T}$, on note a^n la ramification $r = \underbrace{a \times a \times \ldots, \times a \times}_{n \text{ fois}} \Lambda$. On a donc

$$L = \{ a^n + a^p ;\ n \geqq 0,\ p \geqq 0 \}.$$

On en déduit facilement que

$$(\exists n), \quad (r = a^n) \quad \Rightarrow \quad C_L(r) = \{ a^n \times \mathbf{1} + a^p ;\ n \geqq 0,\ p \geqq 0 \}$$
$$\cup \{ a^n + a^p \times \mathbf{1},\ n \geqq 0,\ p \geqq 0 \} = C_0,$$
$$(\exists n > 0), \quad (\exists p > 0), \quad (r = a^n + a^p) \quad \Rightarrow \quad C_L(r) = \{ \mathbf{1} \} = C_1,$$
$$r \text{ d'une autre forme} \quad \Rightarrow \quad C_L(r) = \emptyset = C_2.$$

Donc $\mathscr{C}_L=(C_0, C_1, C_2)$ admet comme structure de binoïde celle décrit par les deux tableaux de la figure 7.

$+$	C_0	C_1	C_2
C_0	C_0	C_1	C_2
C_1	C_1	C_2	C_2
C_2	C_2	C_2	C_2

$\times$	a
C_0	C_0
C_1	C_2
C_2	C_2

Figure 7

Si L était de la forme $P \cap K$ on devrait avoir :

$$(\exists\, i \in [0, 2]),\quad (\exists\, j \in [0, 2]),\quad (\forall\, k \in [0, 2]),\quad (C_i + C_k + C_j = a \times C_k)$$

en prenant $k=0$ on en déduit $i=j=0$, et en prenant $k=1$ on en déduit une contradiction. Donc L n'est pas de la forme $P \cap K$.

5.3. Application aux langages algébriques ayant une ambiguïté inhérente

Classiquement on dit qu'un langage algébrique présente une ambiguïté inhérente si et seulement si toute grammaire algébrique qui l'engendre est ambigüe. Nous allons montrer en particulier qu'une condition nécessaire et suffisante pour qu'un langage algébrique L présente une ambiguïté inhérente est que pour tout bilangage régulier L' tel que $\varphi(L')=L$ alors la fonction $\varphi : L' \to L$ n'est pas injective. Cette propriété montre que le phénomène de l'ambiguïté inhérente est lié aux structures des marqueurs de phrase et non à leur étiquetage (*cf.* en particulier le corollaire 5.3.6). En effet, la classe des bilangages réguliers est stable par transcription. Ce résultat n'est pas surprenant, car le principal outil dont on dispose pour démontrer sur des exemples l'ambiguïté inhérente de langage algébrique est le lemme d'Ogden [6] qui précisément permet de construire pour un mot du langage plusieurs arbres syntaxiques associés à ce mot et de structures différentes.

Nous montrerons d'autre part que les degrés d'ambiguïté des langages algébriques sont conservés en prenant comme ensembles de marqueur de phrases les bilangages réguliers.

Ces résultats peuvent se déduire des résultats de [4] en généralisant les théorèmes de cet article à des grammaires algébriques généralisées.

Le lemme qui suit servira dans la démonstration du théorème 4.3.3.

LEMME 5.3.1 : *Soit L un bilangage régulier sur T. On peut imposer à un T-binoïde B tel que $L=\psi^{-1}(B')$ avec $\psi : \hat{V} \to B$ et $B' \subset B$ de vérifier la condition suivante :*

$$(\forall a \in V),\quad (\forall s \in B),\quad (\forall s' \in B),\quad (e \neq a \times s + s'),$$

où e désigne l'élément neutre de B.

Démonstration : Soit B un T-binoïde et soit $\bar{B} = B \cup \{\bar{e}\}$ où $\bar{e}$ est symbole qui n'est pas dans B. Donnons à $\bar{B}$ une structure de T-binoïde en prolongeant les lois de B par

$$(\forall s \in \bar{B}),\quad (\bar{e}+s=s+\bar{e}=s),$$
$$(\forall a \in T),\quad (a \times \bar{e} = a \times e),$$

où e est l'élément neutre de B.

Il est évident que $\bar{B}$ vérifie la condition du lemme 5.3.1.

On montre facilement par récurrence sur r que si $\bar{\psi}$ est l'unique homomorphisme de $\hat{T}$ dans $\bar{B}$ on a

$$\bar{\psi}(\Lambda)=\bar{e},$$
$$r \neq \Lambda \quad \Rightarrow \quad \bar{\psi}(r)=\psi(r) \quad (\psi : \hat{T} \to B).$$

Donc si B' est contenu dans B on a

$$e \notin B' \quad \Rightarrow \quad \bar{\psi}^{-1}(B')=\psi^{-1}(B'),$$
$$e \in B' \quad \Rightarrow \quad \bar{\psi}^{-1}(B' \cup \{\bar{e}\})=\psi^{-1}(B').$$

On en déduit alors de façon immédiate le lemme annoncé.

DÉFINITION 5.3.2 : Soit L un bilangage régulier et $L'=\varphi(L)$ l'ensemble des mots des feuilles des ramifications de L. Le degré d'ambiguïté $A_L(\alpha)$ d'un mot de L' par rapport à L est le nombre de ramification de L ayant α comme mot des feuilles. Si G est une grammaire algébrique engendrant L' on définit $A_G(\alpha)$ par

$$A_G(\alpha)=A_{B(G)}(\alpha)$$

[$B(G)$ est le bilangage régulier engendré par G].

REMARQUE : Si L est un bilangage régulier sur T, si $G=(N, T', \to, X)$ est une grammaire algébrique et si $\tau : N \cup T' \to T$ est une application conservant les éléments de T' et telle que $L=\bar{\tau}(B(G))$ alors $\varphi(L)$ et $L(G)$ sont égaux et pour tout α de $L(G)$ on a $A_G(\alpha) \geqq A_L(\alpha)$.

THÉORÈME 5.3.3 : *Soit L un bilangage régulier de $\hat{T}$ et $L'=\varphi(L)$. Il existe une grammaire algébrique G engendrant L' telle que*

$$(\forall \alpha \in L'), \quad [A_G(\alpha) \leqq A_L(\alpha)].$$

Ce théorème signifie que pour tout bilangage régulier L il existe une grammaire engendrant le langage $\varphi(L)$ et diminuant ou conservant le degré d'ambiguïté de chaque mot de $\varphi(L)$.

Démonstration : Comme L est un bilangage régulier il existe un T-binoïde fini B et une partie B' de B tel que pour l'unique homomorphisme $\psi : \hat{V} \to B$ on ait $L=\psi^{-1}(B')$. On peut de plus imposer à B de vérifier la condition du lemme 5.3.1. Considérons la grammaire

$$G=(B, T, \to, B'),$$

où la relation $\to$ est définie par

$$s \to s's'' \quad \Leftrightarrow \quad (\exists a \in T), \quad (s=a \times s'+s'' \text{ et } s' \neq e)$$

(e est l'élément neutre de B et le calcul $a \times s'+s''$ est effectué dans B)

$$s \to as'' \quad \Leftrightarrow \quad s=a \times e+s'',$$
$$e \to \Lambda.$$

Il est facile de montrer par récurrence que

$$s \underset{G}{\overset{*}{\rightarrowtail}} \alpha \quad \text{et} \quad \alpha \in T^* \quad \Leftrightarrow \quad (\exists r \in \hat{V}), \quad [\varphi(r)=\alpha \text{ et } \psi(r)=s].$$

(on fait une récurrence sur $|\alpha|$ pour démontrer la proposition dans le sens direct et une récurrence sur r pour démontrer la réciproque).

D'où en déduit que G engendre L'. Montrons que G vérifie les conditions du théorème 5.3.3. Considérons l'application $g : \hat{T} \to \widehat{B \cup T}$ définie par

$$g(\Lambda)=e,$$
$$g(a \times r'+r'')=\text{si } r' \neq \Lambda \text{ alors } \psi(a \times r'+r'') \times (g(r')+g(r''))$$
$$\text{sinon } \psi(a+r'') \times (a+g(r'')).$$

On remarque que

$$g(r)=r' \quad \text{et} \quad r \neq \Lambda \quad \Rightarrow \quad \varphi(r)=\varphi(r') \quad \text{et} \quad \rho(r')=\psi(r).$$

Soit G' la grammaire $(B, T, \rightarrow, B)$ avec $\rightarrow$ défini comme dans G on montre facilement par récurrence sur r dans le premier cas et sur r' dans le deuxième que

$$g(r)=r' \quad \Rightarrow \quad r' \in B(G'),$$
$$r' \in B(G') \quad \Rightarrow \quad (\exists\, r \in L), \quad [g(r)=r'].$$

D'où on en déduit que $B(G)=g(L)$.

Comme g conserve le mot des feuilles il est alors immédiat que le nombre de ramification de $B(G)$ ayant α comme mot des feuilles est inférieur ou égal au nombre de ramification de L ayant α comme mot des feuilles.

COROLLAIRE 5.3.4 : *Une condition nécessaire et suffisante pour qu'un langage algébrique L' présente une ambiguïté inhérente est que pour tout bilangage régulier L tel que $L'=\varphi(L)$ l'application, $\varphi : L \rightarrow L'$ ne soit pas injective.*

Démonstration : Ceci est une conséquence immédiate du théorème 5.3.3.

DÉFINITION 5.3.5 : Soit $G=(N, T, ::=, X)$ une grammaire algébrique généralisée, on appelle ensemble des structures de G le bilangage régulier sur $T \cup \{ \# \}$, où $\#$ est un symbole qui n'est pas dans T, qui est obtenu par transcription à partir de $B(G)$ par la transcription $\tau : \widehat{N \cup T} \rightarrow T \cup \{ \# \}$ définie par $\tau \mid T=\mathrm{Id}_T$ et $(\forall\, A \in N)$, $[\tau(A)= \#]$.

Le terme de structure est justifié par le fait que cette transcription élimine complètement l'étiquetage des arbres syntaxiques engendrés par G en ne gardant que leurs structures.

COROLLAIRE 5.3.6 : *Soit L' un langage algébrique présentant une ambiguïté inhérente et G une grammaire engendrant L' alors pour au moins un mot α de L' deux éléments au moins de l'ensemble des structures de G ont α comme mot des feuilles.*

Démonstration : Comme l'ensemble des structures d'une grammaire est un bilangage régulier le corollaire 5.3.6 est un cas particulier du corollaire 5.3.4. A propos de ce corollaire montrons comment on peut le déduire des théorèmes de McNaughton [4]. (Je dois cette démonstration à Perrot.)

Soit G une grammaire engendrant L', soit L l'ensemble des structures de G, L est un langage parenthésé et donc il existe une grammaire inversible,

$$G'=(N, (\#, \overline{\#}\} \cup T \cup \overline{T}, \rightarrow, X)$$

au sens de [4] engendrant L. La fonction $\varphi : L \rightarrow L'$ est l'homomorphisme défini par $\varphi(\#)=\varphi(\overline{\#})=\varphi(\overline{a})=\Lambda$ et $\varphi(a)=a$. Donc si φ est injective alors la grammaire $G''=(N, T, \mapsto, X)$ définie par

$$A \mapsto \alpha \quad \Leftrightarrow \quad (\exists\, \beta), \quad [A \rightarrow \beta \text{ et } \varphi(\beta)=\alpha]$$

engendre L' et est non ambiguë. Donc L' n'a pas d'ambiguïté inhérente. Remarquons que cette démonstration n'est valable que si G n'est pas une grammaire généralisée, et que pour obtenir le résultat général du corollaire 5.3.6 il est nécessaire de démontrer indépendamment que si L' est engendré par une grammaire généralisée non ambiguë alors L' l'est aussi par une grammaire non généralisée non ambiguë.

Le théorème 5.3.3 peut aussi se déduire des résultats de McNaughton en employant un raisonnement similaire à celui qui précède, mais là pour obtenir le résultat dans toute sa généralité il est nécessaire de généraliser les résultats de [4] aux grammaires généralisées. Cela ne pose pas de gros problèmes, mais il faut remarquer que si on conserve la forme des règles des grammaires parenthésées de McNaughton, on introduira des grammaires parenthésées avec une infinité de règles pour lesquelles il sera impossible de trouver une grammaire parenthésée équivalente de la même forme avec un nombre fini de règles (*cf.* th. 3.2.2).

DÉFINITION 5.3.7 : On appelle degré d'ambiguïté d'un langage algébrique L' l'entier $A(L')$ éventuellement infini défini par

$$A(L')=\operatorname*{Inf}_{G}(\sup_{\alpha\in L'}(A_G(\alpha))).$$

où G parcourt la classe des grammaires algébriques engendrant L' et $A_G(\alpha)$ désigne le degré d'ambiguïté de α par rapport à G.

On peut définir de même un entier $A'(L')$ en posant

$$A'(L')=\operatorname*{Inf}_{L}(\sup_{\alpha\in L'}(A_L(\alpha))).$$

où L parcourt la classe des bilangages réguliers tels que $\varphi(L')=L$. On obtient alors la conséquence immédiate suivante du théorème 5.3.3.

COROLLAIRE 5.3.8 : *Pour tout langage algébrique L' les deux entiers $A(L')$ et $A'(L')$ sont égaux.*

ANNEXE

GÉNÉRALISATION DE LA NOTION DE GRAMMAIRE PARENTHÉSÉE INVERSIBLE

Dans cette annexe on pose quelques définitions qui permettent de définir des grammaires parenthésées inversibles plus générales que celles étudiées au paragraphe 4. On montre en particulier que ces grammaires inversibles sont non-ambiguës.

DÉFINITION 1 : Soit r une ramification sur $\widehat{N\cup T}$ où les éléments de N n'apparaissent qu'aux feuilles. Soit n le nombre d'occurrence d'élément de N dans r. On numérote entre 1 et n ces occurrences. On note f_r la fonction de

$(\hat{T})^n \to \hat{T}$ telle que $f_r(r_1, r_2, \ldots, r_n)$ est la ramification de $\hat{T}$ obtenu en remplaçant pour chaque i la i-ième occurrence d'un élément de N par r_i. On dit que r est injectif sur X si la restriction de f_r sur X^n est une fonction injective.

Exemple :

– si $r = a \times A + B$, r est injectif sur $\hat{T}$;

– si $r = a \times (A_1 + A_2 + \ldots + A_n)$, r est injectif sur $X = \{ r; \mid \rho(r) = 1 \}$.

PROPOSITION 1 : *Une condition nécessaire et suffisante pour que r soit injectif sur $\hat{T}$ est que $\rho(r)$ contient au plus un élément de N et que deux éléments de N n'ait pas le même prédecesseur. Formellement on obtient :*

$$r \textit{ injectif } \Leftrightarrow (\forall a \in T), \ [\forall \alpha \in F_a(r)], \ [\alpha \in T^* (N \cup \{ \Lambda \}) T^*] \textit{ et }$$
$$\rho(r) \in T^* (N \cup \{ \Lambda \}) T^*.$$

Démonstration : Supposons que r ne vérifie pas cette condition. Soit i et j le numéro de deux occurrences d'élément de N dans r apparaissant toutes les deux dans la racine ou toutes les deux dans une même famille de prédécesseur a. Soit r' la ramification qui est entre ces deux occurrences dans r et r'' la ramification obtenue en remplaçant dans r' tout élément de N par Λ.

Si r'' est non vide on obtient :

$$r(\Lambda, \Lambda, \ldots, \underset{\substack{\uparrow \\ i\text{-ième} \\ \text{place}}}{r''}, \Lambda, \ldots, \underset{\substack{\uparrow \\ j\text{-ième} \\ \text{place}}}{\Lambda}, \ldots, \Lambda) = r(\Lambda, \ldots, \Lambda, \underset{\substack{\uparrow \\ i\text{-ième} \\ \text{place}}}{\Lambda}, \ldots, \underset{\substack{\uparrow \\ j\text{-ième} \\ \text{place}}}{r''}, \ldots, \Lambda),$$

donc r est non injectif.

Si r'' est vide alors on obtiendra le même résultat que ci-dessus mais en utilisant une ramification s quelconque à la place de r''.

Réciproquement si la condition est vérifiée par r on montre que r est injectif en raisonnant par récurrence sur r.

DÉFINITION 2 : Soit $\mathscr{R}$ un ensemble de ramifications sur $N \cup T$ où les éléments de N n'apparaissent qu'aux feuilles, on dit que $\mathscr{R}$ est indécomposable sur X si et seulement si

$$(\forall r \in \mathscr{R}), \ (\forall r' \in \mathscr{R}), \ [f_r \neq f'_r \Rightarrow \forall r_1, \ldots, r_n \in X,$$
$$\forall r'_1, \ldots, r'_p \in X, f_r(r_1, \ldots, r_n) \neq f_{r'}(r'_1, \ldots, r'_p)].$$

Exemple :

– si $r \in \mathscr{R} \Rightarrow r = \Lambda$ ou $r = a \times A + B$, alors $\mathscr{R}$ est indécomposable sur $\hat{T}$;

– si $r \in \mathscr{R} \Rightarrow r = a \times (A_1 + A_2 + \ldots + A_n)$ alors $\mathscr{R}$ est indécomposable sur $X = \{ r; \mid \rho(r) = 1 \}$.

DÉFINITION 3 : Soit $G=(N, T, \rightarrow, X)$ une bigrammaire régulière (ou une C_0-bigrammaire) G est dite inversible si et seulement si :

(i) X n'apparaît pas en partie droite des règles;

(ii) $A \rightarrow r$ et $B \rightarrow r \Rightarrow A=X$ ou $B=X$ ou $A=B$;

(iii) pour chaque A de N l'ensemble $\mathscr{R}_A=\{r; A \rightarrow r\}$ est indécomposable sur X_G et chaque second membre de règle est injectif sur X_G avec X_G défini par

$$X_G=\bigcup_{A \in N}\{r; \mathrm{A} \overset{*}{\succ\!\!-} r \text{ et } r \in \hat{T}\};$$

(iv) il existe au plus un non-terminal différent de X dont dérive Λ.

THÉORÈME : *Une bigrammaire inversible est non ambiguë.*

Démonstration : Soit $g : \hat{T} \rightarrow \mathscr{P}(N)$ défini par

$$g(r)=\{A; A \overset{*}{\succ\!\!-} r \text{ et } r \in \hat{T}\}.$$

Montrons que

$$(\forall r \in \hat{T}), \quad (\exists A \in N), \quad [g(r) \subseteqq \{A, X\}]. \tag{I}$$

Pour cela faisons une récurrence sur le nombre de point $n(r)$ de r. Si $n(r)=0$ alors $r=\Lambda$ et l'hypothèse (iv) donne le résultat. Supposons (I) démontrer pour $n(r) \leqq k$ et supposons que $n(r)=k+1$.

Soit $A \in g(r)$. On a $A \succ\!\!- s \overset{*}{\succ\!\!-} r$.

Donc $r=f_s(r_1, r_2, \ldots, r_k)$ et comme $\mathscr{R}_A$ est indécomposable et f_s injectif f_s, $r_1, r_2, \ldots, r_k$ sont déterminés de manière unique donc en appliquant l'hypothèse de récurrence à $r_1, r_2, \ldots, r_k$ et en utilisant l'hypothèse (i). On obtient que s est déterminé de manière unique et donc grâce à (i) et (ii) on en déduit (I) pour r.

Pour démontrer que G est non ambiguë, il suffit de montrer que :

$$A \overset{*}{\succ\!\!-} r \quad \text{et} \quad r \in \hat{T} \quad \Rightarrow \quad (\exists !\, r_1, \ldots, r_k),$$

$$[A \rightarrow s \text{ et } f_s(r_1, \ldots, r_k)=r].$$

Cela se fait aisément par récurrence sur la longueur de la dérivation en employant le même type de raisonnement que pour démontrer (I).

Exemple : Les grammaires inversibles de McNaughton [4] et celles définies au paragraphe 4 sont des cas particuliers de la définition 3.

Je remercie M. Pair qui a suivi l'élaboration de ce travail et guidé sa rédaction.

Je remercie aussi MM. Nivat et Perrot qui par leurs conseils m'ont permis de développer largement l'article initialement prévu.

BIBLIOGRAPHIE

1. B. Courcelle, *Une forme canonique pour les grammaires simples déterministes*, R.A.I.R.O., vol. R 1, 1974, p. 19-36.
2. S. Ginsburg et M. Harrison, *Bracketed Context Free Languages*, J. Comp. Syst. Sc., vol. 1, 1967, p. 1-23.
3. D. E. Knuth, *A Characterization of Parenthesis Languages*, Information and Control, vol. 11, 1967, p. 269-289.
4. R. McNaughton, *Parenthesis Grammar*, J. Assoc. Comp. Mach, vol. 14, 1967, p. 490-500.
5. P. Marchand, *Étude et classification des bigrammaires. Application à l'étude des systèmes transformationnels*, Thèse de spécialité, Nancy, 1974.
6. W. F. Ogden, *A Hepful Result for Proving Inherent Ambiguity*, Math. Systems Theory, vol. 2, 1968, p. 191-194.
7. C. Pair, *Sur les notions algébriques liées à l'analyse syntaxique*, Rev. Fr. Inf. R.O. 4, vol. R. 13, 1970, p. 3-29.
8. C. Pair, *Application des bilangages*, Journée sur les arbres, Lille, 1976.
9. C. Pair, *Application de la théorie des ramifications au problème de l'équivalence structurale de deux C-grammaires*, Rev. Fr. Inf. R.O. 5, vol. R. 12, 1971, p. 130-136.
10. C. Pair et A. Quere, *Définition et Étude des Bilangages réguliers*, Information and Control, vol. 13, 1968, p. 565-593.
11. M. Paull et S. Unger, *Structural Equivalence of Context-Free Grammars*, J. Comp. Syst. Sc., vol. 2, 1968, p. 427-463.
12. A. Quere, *Étude des Ramifications et des Bilangages*, Thèse de Spécialité, Nancy, 1969.
13. J. Thatcher, *Tree Automata: an Informal Survey Currents in the Theory of Computing*, A. V. Aho, éd., Prentice-Hall, 1973, p. 143-172.