ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/238846354

Un theoreme de Chomsky-Schutzenberger pour les forets algebriques

Article *in* Calcolo · June 1977

DOI: 10.1007/BF02575971

CITATIONS
8

READS
116

2 authors:



André Arnold
Université Bordeaux 1
104 PUBLICATIONS **3,082** CITATIONS
SEE PROFILE



Max Dauchet
Université de Lille
93 PUBLICATIONS **2,394** CITATIONS
SEE PROFILE

All content following this page was uploaded by André Arnold on 15 November 2016.

The user has requested enhancement of the downloaded file.

Estratto da *Calcolo* Vol. XIV, fasc. II (Aprile - Giugno) 1977

UN THEOREME DE CHOMSKY - SCHÜTZENBERGER POUR LES FORETS ALGEBRIQUES

A. ARNOLD - M. DAUCHET (1)

RÉSUMÉ - On démontre un théorème de Chomsky - Schützenberger dans le cas des forêts algébriques. On définit les forêts de semi-Dyck et les forêts de semi-Dyck simples et on montre qu'une forêt est algébrique si et seulement si elle est l'image par un homomorphisme linéaire de l'intersection d'une forêt de semi-Dyck [ou de semi-Dyck simple] et d'une forêt locale [ou reconnaissable] ou équivalemment si elle est la transductée linéaire complète ascendante ou descendante d'une forêt de semi-Dyck ou de semi-Dyck simple.

Introduction.

Le théorème de Chomsky-Schützenberger [5], [17] (1963) qui caractérise les forêts algébriques comme image homomorphe de l'intersection d'un langage de Dyck ou de semi-Dyck et d'un langage local a été l'un des premiers théorèmes importants de la théorie des langages. Il a fait appaître le rôle important que jouent les langages de Dyck ou de semi-Dyck dans la théorie des langages algébriques et a permis de montrer que la classe des langages algébriques était une AFL principale, définie par Ginsburg et Greibach [9].

Ce résultat a été le point de départ d'un grand nombre de travaux portant sur les relations entre la classe des langages de Dyck et celle des langages algébriques comme ceux de Nivat [11], Boasson [4], Autebert et Beauquier [3], et qui ont amené une meilleure connaissance de la classe des langages algébriques.

Reçu le 2 Mars 1976.

(1) Laboratoire de Calcul Université de Lille I Boîte Postale 36 59650 Villeneuve d'Ascq.

En comparaison, on connaît relativement peu de choses sur la classe des forêts algébriques définie par Rounds [15] en 1970, en dépit de leur intérêt dans le cas des forêts algébriques. Et comme les langages locaux et les homo- Nivat [12], [13] et Courcelle [6]. Rounds [15], [16], Maibaum [10] et Engelfriet et Schmidt [8].

Il est intuitivement évident, quand on considère une construction qui permet de démontrer le théorème de Chomsky-Schützenberger, — par exemple celle de Rhodes et Shamir [14] — qu'on peut faire une construction analogue dans le cas des forêts algébriques. Et comme les languages locaux et les homomorphismes ont leur équivalent dans la théorie des forêts, il suffit de définir l'équivalent des langages de Dyck ou de semi-Dyck pour pouvoir énoncer un théorème de Chomsky-Schützenberger dans le cas des forêts algébriques. Sa démonstration sera analogue à la démonstration du théorème classique.

Précisons tout d'abord que nous nous sommes limités à définir, dans le cas des forêts, l'équivalent des langages de semi-Dyck, pour des raisons de simplicité, et pour la raison que le rôle symétrique que jouent les parenthèses ouvrantes et fermantes dans le langage de Dyck n'a aucune raison d'être dans le cas des forêts car la notion d'arbre est une notion « orientée » contrairement à celle de mot.

Un alphabet gradué de semi-Dyck est un alphabet formé de parenthèses ouvrantes et fermantes tel qu'il existe une bijection entre l'ensemble des parenthèses ouvrantes et une *partition* de l'ensemble des parenthèses fermantes. Un arbre de semi-Dyck est un arbre dont toutes les branches portent un mot de semi-Dyck. Toute forêt algébrique est alors l'image par un homomorphisme linéaire de l'intersection d'une forêt de semi-Dyck et d'une forêt *locale.*

C'est uniquement pour faciliter les démonstrations qu'on considère un alphabet où à une même parenthèse ouvrante sont associées plusieurs parenthèses fermantes. Si on se restreint au cas où à chaque parenthèse ouvrante est associée une seule parenthèse fermante, on obtient les alphabets et les forêts de semi-Dyck simples. Un théorème s'énonce alors: une forêt est algébrique si et seulement si elle est l'image par un homomorphisme linéaire de l'intersection d'une forêt de semi-Dyck *simple* et d'une forêt *reconnaissable.*

Nous démontrons aussi une autre formulation de ce théorème: toute forêt algébrique est l'image par une transduction linéaire complète d'une forêt de semi-Dyck ou d'une forêt de semi-Dyck simple. Or la classe des transductions linéaires complètes est une des très rares classes de transductions à posséder les deux propriétés suivantes:

(i) il n'y a pas de différence entre le cas ascendant et le cas descendant;

(ii) la classe est fermée par composition (Engelfriet [7]); ce qui rend donc cette caractérisation très intéressante.

Enfin nous démontrons que Alg_n, la classe des forêts algébriques sur alphabets gradués dont tous les symboles sont de arité inférieure ou égale à n peut être engendrée par transduction linéaire complète à partir d'une seule forêt de semi-Dyck simple D_n. Ce résultat-qui est le meilleur qu'on peut obtenir car il est bien clair qu'une forêt contenant des symboles de arité strictement supérieure à n ne pourra jamais être obtenue par transduction *linéaire* à partir de forêts ne contenant que des symboles de arité inférieure ou égale à n-ressemble au résultat bien connu que la classe des langages algébriques est un « cône rationnel principal » [4] et semble pouvoir être le point de départ d'une étude approfondie de la classe des forêts algébriques analogue à celle faite en [3] pour les langages algébriques et à celle faite en [1] pour les forêts reconnaissables.

La première partie de ce travail rappelle très brièvement quelques notions utilisées. La deuxième partie est consacrée à la définition des forêts de semi-Dyck et à quelques unes de leurs propriétés. Le théorème de Chomsky-Schützenberger est démontré dans la troisième partie et ses autres formulations dans la quatrième.

1. Préliminaires.

1.1. *Notions de base.*

Nous supposons déjà connus du lecteur, les définitions, notations, et résultats élémentaires de la théorie des langages d'arbres, qu'il pourra retrouver dans l'article de Rounds [15], et dans l'excellente introduction d'Engelfriet [7]. En particulier nous ne définissons pas les notions et notations suivantes:

— alphabet gradué, arbre, forêt, arbre indexé.

— $\Sigma^*, \Sigma^*(X)$, où Σ est un alphabet gradué et où $X=\{x_i \mid i\in N\}$ est l'ensemble des variables.

— sous-arbre initial, sous-arbre final, profondeur.

— noeud, racine, feuille, étiquette.

— homomorphisme linéaire, homomorphisme alphabétique.

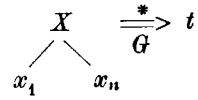
— automate, forêt reconnaissable.

— transducteur linéaire complet ascendant ou descendant.

PROPOSITION 1.1. (*Engelfriet* [7]). *La classe des transductions linéaires ascendantes est identique à celle des transductions linéaires descendantes. Elle est*

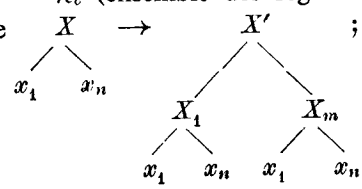
fermée par composition et contient l'intersection avec une forêt reconnaissable.

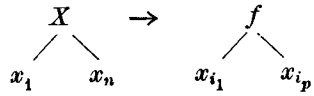
On notera T_{LC} cette classe de transductions.

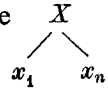
Nous supposons également connues les notions de forêt algébrique, grammaire algébrique, dérivation dans une grammaire et la notation  où G est une grammaire algébrique et X un symbole non-terminal de G. (voir [8], [10], [15], [16]).

PROPOSITION 1.2. (*Rounds* [16]). *La classe des forêts algébriques est fermée par transduction linéaire complète, par intersection avec une forêt reconnaissable et par homomorphisme linéaire.*

Une grammaire algébrique $G=\langle \Sigma_N, \Sigma, X_0, \mathcal{R}\rangle$ est dite sous forme normale [10], si $\mathcal{R} = \mathcal{R}_c \cup \mathcal{R}_f \cup \mathcal{R}_e$ où

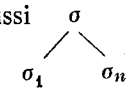
— $\mathcal{R}_c$ (ensemble des règles croissantes) ne contient que des règles de la forme  ;

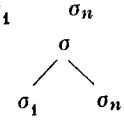
— $\mathcal{R}_f$ (ensemble des règles finales) ne contient que des règles de la forme  où $f\in\Sigma_p$, p pouvant être nul;

— $\mathcal{R}_e$ (ensemble des règles effaçantes) ne contient que des règles de la forme .

PROPOSITION 1.3. (*Maibaum* [10]). *Toute forêt algébrique peut être engendrée par une grammaire sous forme normale.*

1.2. FORETS LOCALES.

Soit Σ un alphabet gradué. On appelle *transition* tout uple $(\sigma, \sigma_1, \ldots, \sigma_n)$ où $n=d(\sigma)$ qu'on écrira aussi .

On dira que la transition  figure dans l'arbre indexé t de $\Sigma^*(X)$ si t s'écrit $t'(\sigma(t_1, \ldots, t_n))$ où la racine de t_i est étiquetée par σ_i.

Soit $\Sigma' \subset \Sigma$, $\Sigma_0' \subset \Sigma_0$ et T un ensemble de transitions. La forêt *locale* définie par $\langle \Sigma', \Sigma_0', T \rangle$ est l'ensemble des arbres de Σ^* dont la racine est étiquetée par un symbole de Σ', dont les feuilles sont étiquetées par des symboles de Σ_0', et dont toutes les transitions figurent dans T.

Il est clair que toute forêt locale est reconnaissable.

1.3. Branches d'arbres.

On note $d(\Sigma)$, appelé degré maximum de l'alphabet gradué Σ, le plus grand entier n tel que $\Sigma_n \neq \varnothing$.

Soit $t \in \Sigma^*(X)$. On appelle *branche* de t tout élément de l'ensemble $B_0(t) \subset (\{0, \ldots, d(\Sigma)\} \times (\Sigma \cup X))^*$ défini inductivement par:

$$B_i(x_j) = \langle i; x_j \rangle$$

$$B_i(\sigma(t_1, \ldots, t_n)) = \langle i; \sigma \rangle (B_1(t_1) \cup B_2(t_2) \ldots \cup B_n(t_n)).$$

Lemme 1.1. *Soit t un arbre de $\Sigma^*(X)$ dont une seule feuille est étiquetée par x. Alors $\forall i\, B_i(t)$ n'a qu'un seul mot se terminant par $\langle j; x \rangle$.*

Ce lemme se démontre très facilement par induction sur la profondeur de t.

Lemme 1.2. *Soit $t(x) \in \Sigma^*(X)$ dont une seule feuille est étiquetée par la variable x et soit $t' \in \Sigma^*(X)$. Alors, $\forall j$, $B_j(t(t'))$ est obtenu en substituant $B_i(t')$ à $\langle i; x \rangle$ dans $B_j(t)$.*

Cette propriété se démontre aisément par induction sur la longueur de l'unique branche de $t(x)$ se terminant par $\langle i; x \rangle$ et en utilisant l'associativité de la substitution dans les langages.

Lemme 1.3. *Soit $t \in \Sigma^*(X)$ et soit $b_1 \langle i; \sigma \rangle b_2 \in B_j(t)$ avec $b_1 \neq \Lambda$. Alors il existe une variable x et un arbre indexé $t_1(x)$ dont une seule feuille est étiquetée par x, un arbre indexé t_2 tels que:*

$$t = t_1(t_2);$$

$b_1 \langle i, x \rangle$ est l'unique élément de $B_j(t_1(x))$ se terminant par $\langle i; x \rangle$;

$$\langle i; \sigma \rangle b_2 \in B_i(t_2).$$

Cette propriété se démontre par induction sur la longueur de b_1.

- Si la longueur de $b_1 = 1$, alors $b_1 = \langle j; \gamma \rangle$ et $t = \gamma(t_1', \ldots, t_n')$. On pose $t_1(x)$, où x est une variable ne figurant pas dans t, égal à $\gamma(t_1', \ldots, t'_{i-1}, x, t'_{i+1}, \ldots, t_n')$ et $t_2 = t_i'$. On a bien $\langle j; \gamma \rangle \langle i; x \rangle \in B_j(t_1(x))$. De plus $B_j(t) = \langle j; \gamma \rangle \bigcup_k B_k(t_k') = \langle j; \gamma \rangle \bigcup_{k \neq i} B_k(t_k') \cup \langle j; \gamma \rangle B_i(t_i')$ qui est bien obtenu

en substituant $B_i(t_i')$ à $\langle j;\gamma\rangle \bigcup_{k\neq i} B_k(t_k') \cup \langle j;\gamma\rangle\langle i;x\rangle = B_j(t_1(x))$.

· Si la longuer de b_1 est supérieure à 1, $b_1 = b_1'\langle k;\gamma\rangle$. t s'écrit donc $t_1(t_2)$ avec $b_1'\langle k;x\rangle \in B_j(t_1(x))$ et $\langle k;\gamma\rangle\langle i;\sigma\rangle\, b_2 \in B_k(t_2)$. Mais t_2 s'écrit $t_3(t_4)$ avec $\langle k;\gamma\rangle\langle i;x\rangle \in B_k(t_3(x))$ et $\langle i;\sigma\rangle\ b_2 \in B_i(t_4)$ d'où $b_1'\langle k;\gamma\rangle\langle i;x\rangle \in B_j(t_1(t_3(x)))$ et $\langle i;\sigma\rangle\, b_2 \in B_i(t_4)$ et $t = t_1(t_3(t_4))$

cqfd.

L'introduction de cette notion de branche va permettre de définir dans la partie suivante les mots portés par une branche et de définir les arbres d'une forêt de semi-Dyck comme ceux dont les branches portent des mots d'un langage de semi-Dyck. Les lemmes 1.2 et 1.3 permettront la décomposition d'un arbre de semi-Dyck en arbres Dyck-premiers en utilisant la décomposition en mots Dyck-premiers des mots de semi-Dyck portés par les branches.

2. Forêts de semi-Dyck.

2.1. Alphabet gradue de semi-Dyck.

Soit I un ensemble fini, soit n une application de I dans $\mathbb{N}$ et d une application de I dans $\mathbb{N}-\{o\}$.

On appelle *alphabet gradué de semi-Dyck* associé à I, n et d, noté $\Delta(I,n,d)$, l'alphabet contenant les symboles suivants:

· $\widehat{o}$ et $\breve{o}$ où le degré de $\widehat{o}$ est 1 et celui de $\breve{o}$ est 0.

· pour chaque $i\in I$, $\widehat{i}$, $\breve{i}^0, \ldots, \breve{i}^{n(i)}$ où le degré de $\widehat{i}$ est $d(i)$ et pour chaque $j\in\{0,\ldots,n(i)\}$ celui de $\breve{i}^j$ est 1.

On appellera parenthèses fermantes les symboles de la forme $\breve{i}^j$ ou $\breve{o}$ et parenthèses ouvrantes ceux de la forme $\widehat{i}$ ou $\widehat{o}$. Contrairement au cas des langages de semi-Dyck, ici à une même parenthèse ouvrante peuvent être associées plusieurs parenthèses fermantes ceci afin de faciliter les démonstrations. Cependant on considère plus loin le cas où $n(i)=1$, c'est-à-dire où à chaque parenthèse ouvrante est associée une seule parenthèse fermante. On verra que $\widehat{o}$ et $\breve{o}$ introduits aussi pour des raisons de commodité seront utilisés comme seules étiquettes possibles pour la racine et les feuilles des arbres d'une forêt de semi-Dyck.

A tout alphabet gradué de semi-Dyck $\Delta(I,n,d)$ on associe l'alphabet non gradué $\Delta'(I,n,d)$ qui contient les symboles

— $\widehat{o}$, $\breve{o}$

— pour chaque $i\in I$, $\widehat{i}$ et $\breve{i}$.

On note $D'_{\Delta'(I,n,d)}$ le langage de semi-Dyck sur $\Delta'(I,n,d)$ engendré par les congruences $\widehat{o}\cdot\widecheck{o}\equiv\Lambda$ et $\widehat{i}\cdot\widecheck{i}\equiv\Lambda$ pour tout $i\in I$.

2.2. Arbres de Dyck.

Soit $\Delta=\Delta(I,n,d)$ un alphabet gradué de semi-Dyck, et $\Delta'=\Delta'(I,n,d)$ l'alphabet non gradué qui lui est associé. On considère l'homomorphisme de monoïde μ de $(\{o,\dots,d(\Delta)\}\times(\Delta\cup X))^*$ dans Δ'^* défini par

$$\mu(\langle j;\ \widehat{i}\rangle)=\widehat{i} \text{ pour tout } i\in I\cup\{o\}$$

$$\mu(\langle j;\ \widecheck{i}^{k}\rangle)=\widecheck{i} \text{ pour } i\in I \text{ et } k\in\{o,\dots,n(i)\}$$

$$\mu(\langle j;\ \widecheck{o}\rangle)=\widecheck{o}$$

$$\mu(\langle j;x\rangle)=\Lambda \text{ pour } x\in X.$$

On dira qu'un arbre indexé $t\in\Delta^*(X)$ est un arbre indexé de semi-Dyck si $\mu(B_0(t))\subset D'_{\Delta'}$. En d'autres termes, un arbre indexé est de semi-Dyck, si pour chaque branche de t, le mot obtenu en prenant les étiquettes des noeuds de cette branche de la racine à la feuille est un mot de semi-Dyck si on ne tient pas compte des indices des parenthèses fermantes. En particulier l'arbre indexé réduit à une variable est un arbre de semi-Dyck puisque $\mu(B_0(x))=\mu(\langle o;x\rangle)=\Lambda$.

Remarque. D'après la définition de B_i et μ, il est clair que

$$\forall i,j,\mu(B_i(t))=\mu(B_j(t)).$$

Proposition 2.1. *Si $t(x)$ et t' sont des arbres indexés de semi-Dyck, $t(t')$ est un arbre indexé de semi-Dyck.*

D'après le lemme 1.1, $B_0(t(t'))=(B_0(t(x))-b\langle j;x\rangle)\cup b\,B_j(t')$ où $b\langle j;x\rangle$ est l'unique branche de t se terminant par x. On en déduit que $\mu(B_0(t(t'))\subset\mu(B_0(t(x))\cup\mu(b)\cdot\mu(B_j(t'))$.

Or

$$\mu(B_0(t(x))\subset D'_{\Delta'}\ ;\ \mu(B_j(t'))\subset D'_{\Delta'}$$

et

$$\mu(b)=\mu(b\langle j;x\rangle)\subset\mu(B_0(t(x)))\subset D'_{\Delta'}. \qquad \textit{cqfd.}$$

Un mot de $D'_{\Delta'}$ est Dyck-premier si c'est un mot non vide de $D'_{\Delta'}$ qui n'est pas le produit de plusieurs mots non vides de $D'_{\Delta'}$. L'ensemble des mots Dyck-premiers de $D'_{\Delta'}$ est donc $D'^{p}_{\Delta'}=(D'_{\Delta'}-\{\Lambda\})-(D'_{\Delta'}-\{\Lambda\})^+$. Par analogie, on dira qu'un arbre indexé de semi-Dyck est Dyck-premier si $\mu(B_0(t))\subset D'^{p}_{\Delta'}$. Un arbre réduit à une seule variable n'est donc pas Dyck-premier.

Comme dans le cas des langages on peut décomposer un arbre indexé de semi-Dyck en arbres Dyck-premiers.

LEMME 2.1. *Si t est un arbre indexé de semi-Dyck non Dyck-premier, il existe un arbre indexé de semi-Dyck, Dyck-premier, $t_0(x_1, \dots, x_n)$ et n arbres indexés de Dyck non réduits à des variables $t_1, \dots, t_n$ tels que $t_0(t_1, \dots, t_n)=t$. Cette décomposition est unique.*

Pour démontrer ce lemme, il suffit de montrer que si t n'est pas Dyck-premier, il existe deux arbres indexés de semi-Dyck $t_1(x)$ et t_2, où t_2 n'est pas réduit à une variable, tels que $t=t_1(t_2)$.

Si t n'est pas Dyck-premier, il existe une branche b de t telle que $\mu(b)$ n'est pas un mot Dyck-premier, ce qui entraine que $b=b_1 b_2$ avec $\mu(b_1)$ et $\mu(b_2)$ appartenant à $D'_\Delta-\{\Lambda\}$. b_2 est donc égal à $\langle i;\sigma\rangle b_2'$. D'après le lemme 1.3 il existe $t_1(x)$ et t_2 tels que $t=t_1(t_2)$; $b_1\langle i;x\rangle$ est l'unique élément de $B_0(t_1(x))$ se terminant par $\langle i;x\rangle$ et $\langle i;\sigma\rangle b_2 \in B_i(t_2)$.

D'après le lemme 1.2 $B_0(t_1(t_2))=(B_0(t_1(x))-b_1\langle i;x\rangle)\cup b_1 B_i(t_2)$.

Comme $\mu(B_0(t_1(t_2)))\subset D'_\Delta$, $\mu(B_0(t_1(x))-b_1\langle i;x\rangle)\subset D'_\Delta$ et $\mu(b_1(B_i(t_2))=$ $=\mu(b_1)\mu(B_i(t_2))\subset D'_\Delta$.

Mais $\mu(b_1)\subset D'_\Delta$ ce qui entraine que $\mu(B_i(t_2))=\mu(B_0(t_2))\subset D'_\Delta$ et donc t_2 est un arbre indexé de semi-Dyck. Par ailleurs $\mu(b_1)=\mu(b_1\langle i;x\rangle)$ d'où $\mu(B_0(t_1(x))=\mu(B_0(t_1(x))-b_1\langle i;x\rangle)\cup\mu(b_1\langle i;x\rangle)\subset D'_\Delta$ est donc $t_1(x)$ est aussi un arbre indexé de semi-Dyck.

En réitérant cette décomposition, on obtient bien un sous-arbre initial Dyck-premier de l'arbre t. Ce sous-arbre initial est unique car sinon il existerait une branche b de t telle que $\mu(b)$ aurait plusieurs décompositions en mots Dyck-premiers. *cqfd.*

PROPOSITION 2.2. *Si t est un arbre indexé Dyck-premier, toutes ses feuilles sont étiquetées par o ou toutes étiquetées par des variables, selon que la racine est étiquetée par $\widehat{o}$ ou non.*

Si t est de la forme $\widehat{o}(t')$, tous les mots de $\mu(B_0(t))$ sont de la forme $\widehat{o}\,W$. Comme ils sont tous Dyck-premiers, ils sont tous de la forme $\widehat{o}\,W'\,\breve{o}$. Toutes les feuilles de t sont donc étiquetées par le symbole o-aire $\breve{o}$. Si t n'est pas de la forme $\widehat{o}(t')$ toutes ses feuilles sont étiquetées par des variables, car sinon il aurait des feuilles étiquetées par l'unique symbole o-aire $\breve{o}$; $\mu(B_0(t))$ contiendrait alors au moins un mot $w\,\breve{o}$ et donc un mot $\widehat{o}\,w\,\breve{o}$ ce qui est impossible si la racine de t n'est pas étiquetée par $\widehat{o}$. *cqfd.*

LEMME 2.2. *Si t est un arbre indexé Dyck-premier de la forme $\sigma(t_1, \dots, t_n)$ dont toutes les feuilles sont étiquetées par des variables, alors*

- *il existe* $i \in I$ *tel que* $\sigma = \widehat{i}$
- *il existe* n *arbres indexés de semi-Dyck* $t_1', \ldots, t_n'$ *dont les feuilles sont toutes étiquetées par des variables tels que* t_j *est obtenu en remplaçant les feuilles de* t_j' *par des arbres de la forme* $\widetilde{i}^k(x_p)$.

Tous les mots de $\mu(B_0(t))$ commencent par $\sigma' = \mu(\langle o; \sigma \rangle)$ et sont dans $D'_{\Delta'}$. σ' est donc égal à $\widehat{o}$ ou $\widehat{i}$, et donc $\sigma = \widehat{o}$ ou $\widehat{i}$. Le cas $\sigma = \widehat{o}$ est exclu d'après la proposition 2.2. Comme les mots de $\mu(B_0(t))$ sont Dyck-premiers ils se terminent tous par $\widetilde{i}$; les branches de t se terminent donc toutes par $\langle m; \widetilde{i}^k \rangle \cdot \langle 1, x_p \rangle = B_m(\widetilde{i}^k(x_p))$-puisque $\widetilde{i}^k$ est un symbole de degré 1 - les arbres t_j' sont obtenus par application du lemme 1.3. De plus il est immédiat que $\mu(B_0(t)) = \widehat{i} \cdot \mu(B_1(t_1') \cup \ldots \cup B_n(t_n')) \cdot i$. Comme $\mu(B_0(t)) \subset D'_{\Delta'}$, il s'ensuit immédiatement que, quel que soit j, $\mu(B_j(t_j')) \subset D'_{\Delta'}$ et donc que les t_j' sont des arbres indexés de semi-Dyck. *cqfd.*

LEMME 2.3. *Si* t *est un arbre indexé Dyck-premier dont toutes les feuilles sont étiquetées par* $\widetilde{o}$, *il existe un arbre indexé de semi-Dyck* t' *dont les feuilles sont étiquetées par des variables tel que* $t = \widehat{o}(t'')$ *ou* t'' *est obtenu en remplaçant les variables par* $\widetilde{o}$ *dans* t'.

Ce lemme se démontre de la même façon que le précédent.

3. FORETS DE SEMI-DYCK.

La *forêt de semi-Dyck* D_Δ sur l'alphabet gradué de semi-Dyck $\Delta = \Delta(I, n, d)$ est l'ensemble des arbres de Δ^* qui sont des arbres de semi-Dyck et dont la racine est étiquetée par $\widehat{o}$.

PROPOSITION 2.3. *Tout arbre de la forêt de semi-Dyck* D_Δ *est Dyck-premier.*

Soit t un arbre de D_Δ. Si t n'est pas Dyck-premier, alors, d'après le lemme 2.1, t s'écrit $t_0(t_1, \ldots, t_n)$ où t_0 est Dyck premier. L'arbre t_0 a des feuilles étiquetées par des variables et la racine étiquetée par $\widehat{o}$ ce qui contredit la proposition 2.2. *cqfd.*

PROPOSITION 2.4. *L'arbre* t *appartient à la forêt de semi-Dyck* D_Δ *ssi il existe un arbre indexé de semi-Dyck* $t'(x_1, \ldots, x_n)$ *tel que* $t = \widehat{o}(t'(\widetilde{o}, \ldots, \widetilde{o}))$.

La condition nécessaire pour que t appartienne à L_Δ est une conséquence de la proposition 2.3 et du lemme 2.3. La condition suffisante découle du fait que $\mu(B_0(t)) = \widehat{o}\, \mu(B_0(t'))\, \widetilde{o}$. *cqfd.*

THEOREME 2.1. *La forêt de semi-Dyck* D_Δ *est une forêt algébrique.*

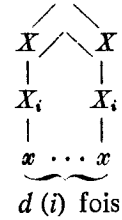
Soit $\Delta=\Delta(I, n, d)$ un alphabet gradué de semi-Dyck.

Soit la grammaire algébrique G dont l'ensemble des symboles non-terminaux est $\{X_0, X\}\cup\{X_i \mid i\in I\}$ avec $d(X_0)=0$, $d(X)=1$, $\forall i\in I\ d(X_i)=1$; dont l'axiome est X_0 et dont les règles sont

r_0: $X_0 \to \widehat{o}\,(X\,(\breve{o}\,))$

r_1: $X(x) \to X(X(x))$

r_2: $X(x) \to x$

$r_{3,i}$: $X(x) \to$  pour tout $i\in I$

$d(i)$ fois

$r_{4,i,j}$: $X_i(x) \to \breve{i}^{j}(x)$ pour tout $i\in I$ et $j\in\{o, \ldots, n(i)\}$.

Soit F la forêt algébrique engendrée par cette grammaire. Il est immédiat d'après la construction de G, que $F\subset D_\Delta$. Pour montrer l'inclusion $D_\Delta\subset F$, on montre d'abord que si t est un arbre indexé de semi-Dyck dont toutes les feuilles sont étiquetées par x, alors $X(x) \underset{G}{\Longrightarrow} t$, en raisonnant par induction sur le nombre $p(t)$ de parenthèses ouvrantes qui figurent dans t.

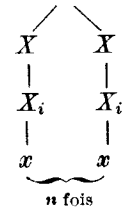
Si $p=o$ alors $t=X$ et $X(x) \overset{*}{\underset{G}{\Longrightarrow}} x$.

Si $p\neq o$

— si t n'est pas Dyck-premier, d'après le lemme 2.1, $t=t_0(t_1, \ldots, t_n)$ avec $p(t_i)<p(t)$ $\forall i\in\{o, \ldots, n\}$ où $t_0(x_1, \ldots, x_n)$, $t_1, \ldots, t_n$ sont des arbres indexés de semi-Dyck. On a donc $X(x) \overset{*}{\underset{G}{\Longrightarrow}} t_i$ pour $1\leq i\leq n$ et $t_0(x, \ldots, x)$ étant aussi un arbre de semi-Dyck $X(x) \overset{*}{\Longrightarrow} t_0(\overline{x, \ldots, x})$ d'où $X(\overline{x}) \overset{*}{\Longrightarrow} X(X(x)) \overset{*}{\Longrightarrow} \overset{*}{\Longrightarrow} t_0(X(x), \ldots, X(x)) \overset{*}{\Longrightarrow} t_0(t_1, \ldots, t_n)=t$

— si t est Dyck-premier, alors, d'après le lemme 2.2, il existe des arbres indexés de semi-Dyck $t_1', \ldots, t_n'$ dont toutes les feuilles sont étiquetées par x, tels que t est obtenu en remplaçant x par $\breve{i}^{j}(x)$ dans $\widehat{i}\,(t_1', \ldots, t_n')$. On a donc

$X(x) \overset{*}{\Longrightarrow} t_k'(x)$ et $X(x) \Longrightarrow \widehat{i} \overset{*}{\Longrightarrow} \widehat{i}(t_1'', \ldots, t_n'')$



où t_k'' est obtenu en remplaçant x par $X_i(x)$ dans t_k' d'où $X(x) \overset{*}{\Longrightarrow} t$.

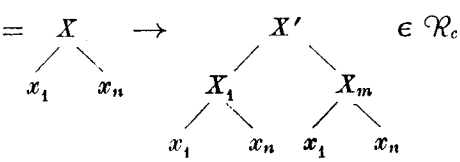
Si maintenant $t \in D_\Delta$, on applique la proposition 2.4. Il existe donc un arbre indexé de semi-Dyck t' dont toutes les feuilles sont étiquetées par x tel que $t = \widehat{o}(t'(\breve{o}, \ldots, \breve{o}))$. Comme $X(x) \overset{*}{\Longrightarrow} t'$, $X_0 \Longrightarrow \widehat{o}(X(\breve{o})) \overset{*}{\Longrightarrow} t$. *cqfd.*

3. Caractérisation des forêts algébriques.

Soit F une forêt algébrique engendrée par une grammaire $G = \langle \Sigma_N, \Sigma, X_0 \mathcal{R} \rangle$ sous forme normale-donc $\mathcal{R} = \mathcal{R}_c \cup \mathcal{R}_f \cup \mathcal{R}_e$.

Soit pg l'application de $\mathcal{R}$ dans Σ_N qui à toute règle r associe le symbole non terminal qui figure en partie gauche de cette règle.

On considère l'ensemble E et l'application val de E dans Σ_N définis comme suit:

- $Z_0 \in E$ et $\text{val}(Z_0) = X_0$
- pour toute règle $r =$  $\in \mathcal{R}_c$

$Z_r, Z_{\bar{r}_1}, \ldots, Z_{\bar{r}_m} \in E$ et $\text{val}(Z_r) = X'$; $\text{val}(Z_{\bar{r}_1}) = X_1, \ldots, \text{val}(Z_{\bar{r}_m}) = X_m$.

3.1. Definition de l'alphabet gradue de semi-Dyck.

On forme l'alphabet gradué de semi-Dyck $\Delta = \Delta(I, n, d)$ défini par

$$I = \{\langle r, Z \rangle \in \mathcal{R} \times E \mid \text{val}(Z) = pg(r)\}$$

$$\forall \langle r, Z \rangle \in I, \; n(\langle r, Z \rangle) = d(\text{val}(Z))$$

$$\forall\, \langle r, Z\rangle \in I,\ d(\langle r, Z\rangle) = \begin{cases} 1 \text{ si } r \in \mathcal{R}_e \cup \mathcal{R}_c \\ 1 \text{ si } r = \overset{X}{\underset{x_1}{\diagup}\ \underset{x_n}{\diagdown}} \rightarrow a \in \mathcal{R}_f \\ p \text{ si } r = \overset{X}{\underset{x_1}{\diagup}\ \underset{x_n}{\diagdown}} \rightarrow \overset{f}{\underset{x_{i_1}}{\diagup}\ \underset{x_{i_p}}{\diagdown}} \in \mathcal{R}_f. \end{cases}$$

3.2. DEFINITION DE L'HOMOMORPHISME.

Soit # un symbole de degré 0 n'appartenant pas à Σ. On définit l'homomorphisme Φ de Δ^* dans $(\Sigma \cup \{\#\})^*$ par

. $\Phi(\widehat{o}(x)) = x$.

. $\Phi(\breve{o}) = \#$.

. $\Phi(\widehat{i^j}(x)) = x$ pour tout $i \in I$ et $j \in \{o, \ldots, n(i)\}$.

— si $r \in \mathcal{R}_e \cup \mathcal{R}_c$:

. $\Phi(\widehat{\langle r, Z\rangle}(x)) = x$ pour tout Z tel que $\mathrm{val}(Z) = pg(r)$.

— si $r = \overset{X}{\underset{x_1}{\diagup}\ \underset{x_m}{\diagdown}} \rightarrow a \in \mathcal{R}_f$:

. $\Phi(\widehat{\langle r, Z\rangle}(x)) = a$ pour tout Z tel que $\mathrm{val}(Z) = pg(r)$.

— si $r = \overset{X}{\underset{x_1}{\diagup}\ \underset{x_n}{\diagdown}} \rightarrow \overset{f}{\underset{x_{i_1}}{\diagup}\ \underset{x_{i_p}}{\diagdown}} \in \mathcal{R}_f$:

. $\Phi(\widehat{\langle r, Z\rangle}(x_1, \ldots, x_p)) = f(x_1, \ldots, x_p)$.

Par définition cet homomorphisme est linéaire.

3.3. DEFINITION DE LA FORET LOCALE K.

Soit T l'ensemble de transitions défini par

1. pour tout $\langle r, Z\rangle \in I$ tel que $r \in \mathcal{R}_c$, $r', r'' \in \mathcal{R}$

1.1. $\widehat{\langle r, Z\rangle}$ appartient à T.

$$\Big|$$

$$\widehat{\langle r', Z_r\rangle}$$

1.2. $\widetilde{\langle r', Z_{\bar{r}_i}\rangle}^{j}$ appartient à T $\forall i \in \{1, \ldots, d(\mathrm{val}(Z_r))\}$ et $j \in \{0, \ldots, \inf(n(\langle r', Z_{\bar{r}_i}\rangle), n(\langle r, Z\rangle))\}$.

$$\Big|$$

$$\widetilde{\langle r, Z\rangle}^{j}$$

1.3. $\widetilde{\langle r', Z_r\rangle}^{i}$ appartient à T $\forall i \in \{1, \ldots, n(\langle r', Z_n\rangle)\}$.

$$\Big|$$

$$\widehat{\langle r'', Z_{\bar{r}_i}\rangle}$$

1.4. $\widetilde{\langle r', Z_r\rangle}^{0}$ appartient à T.

$$\Big|$$

$$\widetilde{\langle r, Z\rangle}^{0}$$

2. pour tout $\langle r, Z\rangle \in I$ tel que $r = X \to x_i$, $\widehat{\langle r, Z\rangle}$ appartient à T.

$$\Big|$$

$$\widetilde{\langle r, Z\rangle}^{i}$$

3. pour tout $\langle r, Z\rangle \in I$ tel que $r = X \to a$, $\widehat{\langle r, Z\rangle}$ appartient à T.

$$\Big|$$

$$\widetilde{\langle r, Z\rangle}^{0}$$

4. pour tout $\langle r, Z\rangle \in I$ tel que $r = X \to f$, $\langle r, Z\rangle$

$$\widetilde{\langle r, Z\rangle}^{i_1} \quad \widetilde{\langle r, Z\rangle}^{i_p}$$

appartient à T.

5. pour tout $r \in \mathcal{R}$ tel que $pg(r) = X_0$, $\widehat{o}$ et $\widetilde{\langle r, Z_0\rangle}^{0}$ appartiennent à T.

$$\Big|$$

$$\widehat{\langle r, Z_0\rangle} \qquad \widetilde{o}$$

La forêt locale K est la forêt locale définie par $<\{\{\widehat{o}\}, \{\widetilde{o}\}, T\}$. On note $K_1(X)$ l'ensemble des arbres indexés dont toutes les feuilles sont étiquetées par des variables et dont toutes les transitions sont dans $T \cup \{\widetilde{\langle r, Z\rangle}^{i} \mid \langle r, Z\rangle \in I$ et

$$\Big|$$

$$x_i$$

$I \in \{o, \ldots, n(\langle r, Z\rangle)\}\}$ et on note $K_D(X)$ l'ensemble des arbres de $K_1(X)$ qui sont des arbres de semi-Dyck-premiers.

PROPOSITION 3.2. *Soit t un arbre de Δ^*. L'arbre t appartient à $D_\Delta \cap K$ ssi il existe $r \in \mathcal{R}$ et un arbre t' de $K_D(X)$ dont la racine est étiquetée par $\langle \widehat{r, Z_0}\rangle$ et toutes les feuilles par x_0 tels que $t = \widehat{o}(t'')$ où t'' est obtenu en remplaçant x_0 par $\breve{o}$ dans t'.*

Remarquons d'abord que $\langle \widehat{r, Z_0}\rangle$ est un symbole du degré 1. En effet si la règle r a val $(Z_0) = X_0$ en partie gauche, $r \in \mathcal{R}_c$ et $d(\langle \widehat{r, Z_0}\rangle) = 1$, ou r est de la forme $X_0 \to a$ et le degré de $\langle \widehat{r, Z_0}\rangle$ est encore 1. D'autre part $n(\langle r, Z_0\rangle) = d(X_0) = 1$.

Démonstration de la condition suffisante.

Soit un arbre t' appartenant à $K_D(X)$ dont la racine est étiquetée par $\langle \widehat{r, Z_0}\rangle$, et toutes les feuilles par x_0. D'après le lemme 2.2 il existe un arbre indexé de semi-Dyck t''' tel que $t' = \langle r, Z_0\rangle (t''' (\langle \widetilde{r, Z_0}\rangle^0 (x_0), \ldots, \langle \widetilde{r, Z_0}\rangle^0 (x_0)))$. Les transitions $\widehat{o}$ et $\langle \widetilde{r, Z_0}\rangle^0$ appartenant à T l'arbre $t = \widehat{o}(\langle \widehat{r, Z_0}\rangle (t'''$ $\cdot (\langle \widetilde{r, Z_0}\rangle^0 (\breve{o}), \ldots, \langle \widetilde{r, Z_0}\rangle^0 (\breve{o})))) = \widehat{o}(t'')$ appartient bien à $D_\Delta \cap K$ et t'' est bien obtenu en remplaçant les variables par $\breve{o}$ dans t'.

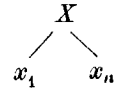
Démonstration de la condition nécessaire.

Si t appartient à $D_\Delta \cap K$ alors d'après la proposition 2.4 $t = \widehat{o}(t'(\breve{o}, \ldots, \breve{o}))$. Comme t appartient aussi à K, $t = \widehat{o}(\langle \widehat{r, Z_0}\rangle (t'' (\langle \widetilde{r_1, Z_0}\rangle^0 (\breve{o}), \ldots, \langle \widetilde{r_n, Z_0}\rangle^0 (\breve{o}))))$ et $t' = \langle \widehat{r, Z_0}\rangle (t'' (\langle \widetilde{r_1, Z_0}\rangle^0 (x_0), \ldots, \langle \widetilde{r_n, Z_0}\rangle^0 (x_0))$ appartient à $K_1(X)$. Il reste à montrer que t' est un arbre Dyck-premier (et donc $r = r_1 \ldots = r_n$).

Si t' n'est pas Dyck-premier il existe une branche b de $B_0(t')$ telle que $\mu(b)$ n'est pas un mot Dyck-premier. Or $\mu(b) = \langle \widehat{r, Z_0}\rangle \omega \langle \widetilde{r_i, Z_0}\rangle^0$ donc comme $\mu(b)$ n'est pas Dyck-premier $\mu(b) = \langle \widehat{r, Z_0}\rangle \omega_1 \langle \widetilde{r, Z_0}\rangle^0 \omega_2 \langle \widehat{r_i, Z_0}\rangle \omega_3 \langle \widetilde{r_i, Z_0}\rangle^0$.

D'après la définition de T, $\langle \widehat{r_i, Z_0}\rangle$ ne peut qu'être précédé de $\widehat{o}$ et $\langle \widetilde{r, Z_0}\rangle^0$ suivi de $\breve{o}$ d'où $\omega_2 = \breve{o}\, \omega_2' \breve{o}$, ce qui est impossible car $\breve{o}$ ne peut pas avoir de successeur ni $\widehat{o}$ de prédécéceur. *cqfd.*

LEMME 3.1. *Si* $X \overset{*}{\Longrightarrow} t(x_{i_1}, \ldots, x_{i_p}) \in \Sigma^*(X)$ *alors il existe* $r \in \mathcal{R}$

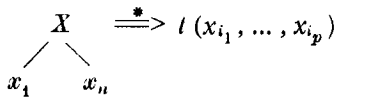


tel que $pg(r)=X$ *et* $\forall Z$ *tel que* $val(Z)=X$ *il existe un arbre* $t'(x_0, \ldots, x_0, x_{i_1}, x_0, \ldots, x_0, x_{i_2}, x_0, \ldots, x_0, x_{i_p}, x_0, \ldots, x_0)=t' \in K_D(X)$ *tel que*

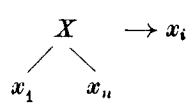
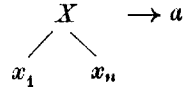
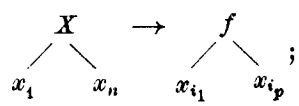
i) — *la racine de* t' *est étiquetée par* $\langle \widehat{r, Z} \rangle$

ii) — $t(x_{i_1}, \ldots, x_{i_p})=\Phi(t')$

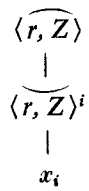
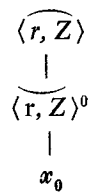
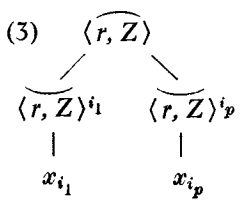
On va démontrer le résultat par induction sur la longueur de la dérivation



— si $l=1$, comme $t(x_{i_1}, \ldots, x_{i_p}) \in \Sigma^*(X)$ la dérivation se réduit à une règle de $\mathcal{R}_e \cup \mathcal{R}_f$. Soit r cette règle, elle a la forme

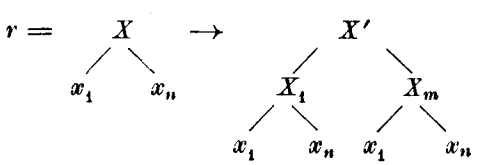
(1) $r=$  ou (2) $r=$  ou (3) $r=$ ;

alors $\forall Z$ tel que $val(Z)=X$ on pose $t_Z=$

(1)  ou (2)  ou (3) 

où t_Z appartient à $K_D(X)$ et vérifie (i) et (ii).

— si $l>1$ la dérivation $X(x_1, .., x_n) \overset{*}{\Longrightarrow} t(x_{i_1}, \ldots, x_{i_p})$ commence nécessairement par la règle



avec

- $X' \overset{*}{\Longrightarrow} t_0(x_{i_1}, \ldots, x_{i_k}) = t_0$ (arbre X' : racine, feuilles x_1 et x_m)
- pour $j \in \{1, \ldots, k\}$ $X_{i_j} \overset{*}{\Longrightarrow} t_j(x_{p_j(1)}, \ldots, x_{p_j(q_j)}) = t_j$ (arbre X_{i_j} : feuilles x_1 et x_n)
- $t(x_{i_1}, \ldots, x_{i_p}) = t_0(t_1(x_{p_1(1)}, \ldots, x_{p_1(q_1)}), \ldots, t_k(x_{p_k(1)}, \ldots, x_{p_k(q_k)}))$.

D'après l'hypothèse de récurrence, il existe r_0 *et*

$$t_0'(x_0, \ldots, x_0, x_{i_1}, x_0, \ldots, x_0, x_{i_2}, x_0, \ldots, x_0, x_{i_k}, x_0, \ldots, x_0) = t_0'$$

$\in K_D(X)$ dont la racine est étiquetée par $\langle \widehat{r_0, Z_r} \rangle$ et tels que $\Phi(t_0') = t_0$ et pour chaque $j \in \{1, \ldots, k\}$ il existe r_j et

$$t_j' = t_j'(x_0, \ldots, x_0, x_{p_j(1)}, x_0, \ldots, x_0, x_{p_j(2)}, x_0, \ldots, x_{p_j(q_j)}, x_0, x_0 \ldots) \in K_D(X)$$

dont la racine est étiquetée par $\langle \widehat{r_j, Z_{\bar{r}_{i_j}}} \rangle$ et tel que $t_j = \Phi(t_j')$.

Comme t_0' appartient à $K_D(X)$, la variable x_{i_j} est précédée de $\langle \widetilde{r_0, Z_r} \rangle^{i_j}$ La transition $\langle \widetilde{r_0, Z_n} \rangle^{i_j}$ (avec $\langle \widehat{r_j, Z_{\bar{r}_{i_j}}} \rangle$ au-dessous) appartenant à T l'arbre $u = t_0'(x_0, \ldots, x_0, t_1', x_0, \ldots, x_0, t_2', \ldots, t_k', x_0, \ldots, x_0) \in K_1(X)$ et c'est un arbre de semi-Dyck.

Les transitions $\langle \widehat{r, Z} \rangle$ ($\langle \widehat{r_0, Z_r} \rangle$), $\langle \widetilde{r_j, Z_{\bar{r}_{i_j}}} \rangle^k$ ($\langle \widetilde{r, Z} \rangle^k$), $\langle \widetilde{r_0, Z_r} \rangle^0$ ($\langle \widetilde{r, Z} \rangle^0$), $\langle \widetilde{r_j, Z_{\bar{r}_{i_j}}} \rangle^0$ ($\langle \widetilde{r, Z} \rangle^0$) appartenant à

pour tout Z tel que $\mathrm{val}(Z) = pg(r) = X$ alors l'arbre $\langle \widehat{r, Z} \rangle(u')$ où u' e obtenu en remplaçant x_j par $\langle \widehat{r, Z} \rangle^j(x_j)$ dans u appartient à $K_D(X)$. De pl on a bien $\Phi(\langle \widehat{r, Z} \rangle(u')) = \Phi(u')$; et comme

$$\Phi(\langle \widetilde{r, Z} \rangle^j(x_j)) = x_j, \Phi(u') = \Phi(u) =$$

$$= t_0(\Phi(t_1'), \ldots, \Phi(t_n')) = t_0(t_1, \ldots, t_n) = t.$$

De plus il est immédiat par construction que les variables de $\widehat{\langle r, Z\rangle}(u')$ différentes de x_0 sont, dans l'ordre

$$x_{p_1(1)}, \ldots, x_{p_1(q_1)}, x_{p_2(1)}, \ldots, x_{p_2(q_2)}, \ldots, x_{p_k(1)}, x_{p_k(q_k)}.$$

cqfd.

LEMME 3.2. *Si t est un arbre de $K_D(X)$ dont la racine est étiquetée par $\langle r, Z\rangle$, alors* $\begin{array}{c}\mathrm{val}\,(Z)\\ \diagup\quad\diagdown\\ x_1\qquad x_n\end{array} \overset{*}{\Longrightarrow} \Phi(t)$. *De plus les variables de $\Phi(t)$ sont, dans l'ordre, celles de t qui sont différentes de x_0.*

On démontre le résultat par induction sur le nombre p de parenthèses ouvrantes figurant dans t.

· Si $p=1$ t' est nécessairement de la forme

(1) $\begin{array}{c}\widehat{\langle r, Z\rangle}\\ \diagup\quad\diagdown\\ \widetilde{\langle r, Z\rangle}^{i_1}\qquad \widetilde{\langle r, Z\rangle}^{i_p}\\ |\qquad\qquad |\\ x_{i_1}\qquad\quad x_{i_p}\end{array}$ ou (2) $\begin{array}{c}\widehat{\langle r, Z\rangle}\\ |\\ \widetilde{\langle r, Z\rangle}^{i}\\ |\\ x_i\end{array}$, avec $i\neq 0$, ou (3) $\begin{array}{c}\widehat{\langle r, Z\rangle}\\ |\\ \widetilde{\langle r, Z\rangle}^{0}\\ |\\ x_0\end{array}$

Dans le cas 1) $r=\begin{array}{c}\mathrm{Val}\,(Z)\\ \diagup\quad\diagdown\\ x_1\qquad x_n\end{array} \Rightarrow \Phi(t)=f(x_{i_1}, \ldots, x_{i_n})$.

Dans le cas 2) $r=\begin{array}{c}\mathrm{val}\,(Z)\\ \diagup\quad\diagdown\\ x_1\qquad x_n\end{array} \Rightarrow \Phi(t)=\begin{cases} f(x_i) \text{ ou}\\ x_i.\end{cases}$

Dans le cas 3) $r=\begin{array}{c}\mathrm{val}\,(Z)\\ \diagup\quad\diagdown\\ x_1\qquad x_n\end{array} \Rightarrow \Phi(t)=a.$

Dans ces trois cas les conclusions du lemme sont vérifiées.

$p>1$.

Remarquons d'abord que dans ce cas $r\in\mathcal{R}_c$ et donc $\widehat{\langle r, Z\rangle}$ est un symbole de degré 1 car si $r\notin\mathcal{R}_c$ les transitions permises à partir de $\widehat{\langle r, Z\rangle}$ sont $\begin{array}{c}\widehat{\langle r, Z\rangle}\\ |\\ \widetilde{\langle r, Z\rangle}^{i}\end{array}$ ou $\begin{array}{c}\widehat{\langle r, Z\rangle}\\ \diagup\quad\diagdown\\ \widetilde{\langle r, Z\rangle}^{i_1}\qquad \widetilde{\langle r, Z\rangle}^{i_p}\end{array}$ et si $t\in K_D(X)$, t n'aurait qu'une seule parenthèse ouvrante, ce qui est exclu.

On applique à t le lemme 2.2 et en tenant compte du fait que $t \in K_1(X)$ on en déduit qu'il existe un arbre indexé de semi-Dyck t' appartenant à $K_1(X$ dont la racine est étiquetée par $\langle \widehat{r_0, Z_r} \rangle$ tel que t est obtenu en remplaçan x_i par $\langle \widetilde{r, Z} \rangle^i (x_i)$ dans $\langle \widehat{r, Z} \rangle (t')$. De plus en tenant compte du lemme 2.1, t peut s'écrire $t_0(t_1, \ldots, t_p)$ (p pouvant être nul si t' est Dyck-premier) où $t_1, \ldots, t_p$ sont des arbres indexés de semi-Dyck (éventuellement réduits à une variable). Comme t_0 est un arbre Dyck-premier dont la racine est étiquetée par $\langle \widehat{r_0, Z_r} \rangle$, la variable x_j est précédée de $\langle \widetilde{r_0, Z_r} \rangle^{k_j}$ et $t_0(x_{k_1}, \ldots, x_{k_p}) \in K_D(X)$

1) Supposons que l'arbre t_i soit réduit à la variable x_{q_j}. Comme en substituant t_j à x_j dans $t_0(x_1, \ldots, x_p)$ on obtient un arbre de $K_1(X)$ et si $t_i = x_{q_i}$ la variable x_i est précedée de $\langle \widetilde{r', Z_r} \rangle^{q_i}$. Mais en substituant $\langle \widetilde{r, Z} \rangle^i (x_i)$ à x dans $\langle r, Z \rangle (t')$ on obtient encore un arbre de $K_1(X)$; la transition $\langle \widetilde{r_0, Z_r} \rangle^{q}$ / $\langle \widetilde{r, Z} \rangle^{q_i}$ appartient à T ce qui n'est possible que si $q_i = 0$.

2) Si t_i n'est pas réduit à une variable, sa racine est étiquetée par $\langle \widehat{r_i, Z_i} \rangle$ et la transition $\langle \widetilde{r_0, Z_r} \rangle^{k_i} \in T$ ce qui nécessite que $Z_i = Z_{\bar{r}_{k_j}}$. Comme $t \in K_1(X)$ / $\langle \widehat{r_i, Z_i} \rangle$ $t_i \in K_1(X)$. Si t_i n'est pas Dyck-premier t_i pourra se décomposer ou t_i ($\langle \widetilde{r_i', Z_{\bar{r}_{k_i}}} \rangle^j (t'')$ où la racine de t_i'' est étiquetée par $\langle \widehat{r_i'', Z_i''} \rangle$ ce qui est impossible car T ne contient aucune transition de la forme $\langle \widetilde{r', Z_{\bar{r}_i}} \rangle^j$.

/ $\langle \widehat{r', Z} \rangle$

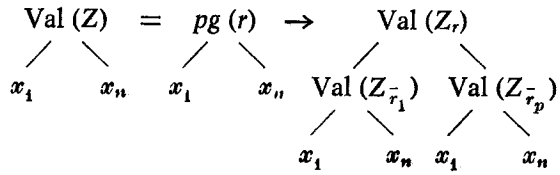
En conséquence, il existe donc $t_0 = t_0(x_0, \ldots, x_0, x_{k_1}, x_0, \ldots, x_0, x_{k_2}, x_0, \ldots, x_{k_m}$ $x_0, \ldots, x_0) \in K_D(X)$ dont la racine est étiquetée par $\langle \widehat{r_0, Z_r} \rangle$ et m arbres de $K_D(X)$, $t_1, \ldots, t_m$ dont les racines sont étiquetées respectivement par $\langle \widehat{r_i, Z_{\bar{r}_{k_i}}} \rangle$. D'apres l'hypothèse de récurrence $\mathrm{Val}(Z_r) \overset{*}{\Longrightarrow} \Phi(t_0) = \bar{t}_0(x_{k_1}, \ldots, x_{k_m})$.

$x_1 \quad x_n$

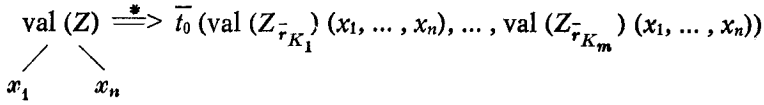
$\mathrm{Val}(Z_{\bar{r}_{k_i}}) \overset{*}{\Longrightarrow} \Phi(t_i)$.

$x_1 \quad x_n$

Par ailleurs, par définition de l'ensemble E,



d'où



$$\overset{*}{\Longrightarrow} \bar{t}_0 (\Phi (t_1), \dots, \Phi (t_m)) = \Phi (t_0 (x_0, \dots, x_0, t_1, x_0, \dots, x_0, t_2, \dots, t_m, x_0, \dots, x_0)) =$$

$$= \Phi (t) \text{ puisque } \Phi (\widehat{\langle r, Z\rangle} (x)) = x \text{ et } \Phi (\widetilde{\langle r, Z\rangle}^i (x)) = x.$$

De plus Φ ne fait bien disparaître que les variables égales à x_0.

cqfd.

PROPOSITION 3.3. *La forêt algebrique F est égale à l'ensemble* $\Phi (K \cap D_A)$.

Si $t \in F$, alors $X_0 \overset{*}{\Longrightarrow} t$. D'après le lemme 3.1, il existe r et un arbre $t' (x_0, \dots, x_0)$ de $K_D (X)$ dont la racine est étiquetée par $\widehat{\langle r, Z_0\rangle}$ et les feuilles par x_0, tel que $t = \Phi (t')$. D'après la proposition 3.2 $\widehat{o} (t' (\breve{o}, \dots, \breve{o})) \in K \cap D_A$ et $\Phi (\widehat{o} (t' (\breve{o}, \dots, \breve{o})) = \Phi (t' (\breve{o}, \dots, \breve{o})) = t$.

Réciproquement, si $t \in K \cap D_A$, il existe un arbre $t' (x_0, \dots, x_0)$ de $K_D (X)$ dont la racine est étiquetée par $\widehat{\langle r, Z_0\rangle}$ tel que $t = \widehat{o} (t' (\breve{o}, \dots, \breve{o}))$. D'après le lemme 3.2 $X_0 = \text{val} (Z_0) \overset{*}{\Longrightarrow} \Phi (t' (x_0, \dots, x_0)) \in \Sigma^*$ et $\Phi (t' (x_0, \dots, x_0)) = \Phi (\widehat{o} (t' (\breve{o}, \dots, \breve{o}))) = \Phi (t)$ d'où $\Phi (t) \in F$.

cqfd.

THEOREME 3.1. *Une forêt F est algébrique ssi il existe une forêt de semi-Dyck* D_A, *une forêt locale K et un homomorphisme linéaire* Φ *tel que* $F = \Phi (K \cap D_A)$.

On vient de démontrer la condition nécessaire. La condition suffisante est une conséquence du fait que D_A est une forêt algébrique (th 2.1) et des propriétés de fermeture de la classe des forêts algébriques (Proposition 1.3) *cqfd.*

4. Autres caractérisations équivalentes.

4.1. CARACTERISATION PAR TRANSDUCTION LINEAIRE COMPLETE.

THEOREME 4.1. *Une forêt est algébrique ssi elle est la transductée par une transduction linéaire complète d'une forêt de semi-Dyck.*

La condition suffisante est une conséquence immédiate du théorème 2.1 et de la proposition 1.3.

D'après le théorème 3.1 toute forêt algébrique est égale à $\Phi(K \cap D_\Delta)$. Comme l'intersection avec une forêt reconnaissable est une transduction linéaire complète et que ces transductions linéaires complètes se composent, il suffit de démontrer que l'homomorphisme Φ construit en 3.2 est égal à une transduction linéaire complète sur $K \cap D_\Delta$, ce que nous démontrons dans le lemme suivant, et l'on obtient la condition nécessaire. *cqfd.*

LEMME 4.1. *Soient D_Δ la forêt de semi-Dyck, F la forêt locale et Φ l'homomorphisme définis dans la partie 3. Alors il existe un transducteur linéaire complet τ_Φ tel que $\Phi(D_\Delta \cap K) = \tau_\Phi(D_\Delta \cap K)$.*

Par définition, le seul cas où l'homomorphisme Φ n'est pas complet est le cas

$$\Phi(\widehat{\langle r, Z\rangle}(x)) = a \text{ où } r = \begin{array}{c} \text{val}(Z) \to a. \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array}$$

Mais dans ce cas, si $t(\widehat{\langle r, Z\rangle}(t')) \in K$, alors t' est un arbre monadique de la forme $\widetilde{\langle r, Z\rangle}^0 \widetilde{\langle r_1, Z_1\rangle}^0 \ldots \widetilde{\langle r_n, Z_n\rangle}^0 \widetilde{\langle r_{n+1}, Z_0\rangle}^0 \tilde{o}$.

On va construire un transducteur linéaire complet qui mémorisera $a = \Phi(\widehat{\langle r, Z\rangle})$, lira t' et réécrira a en lisant $\tilde{o}$. De façon plus précise, on considère le transducteur τ_Φ dont l'ensemble des états est $\{q_0\} \cup \{q_a \mid a \in \Sigma_0\}$, dont l'état initial est q_0 et dont les règles sont pour tout $\sigma \in \Delta$.

$q_0[\sigma(x_1, \ldots, x_n)] = t(q_0[x_1], \ldots, q_0[x_n])$ si $\Phi(\sigma(x_1, \ldots, x_n)) = t(x_1, \ldots, x_n) \notin \Sigma_0$

$q_0[\widehat{\langle r, Z\rangle}(x)] = q_a[x]$ si $\Phi(\widehat{\langle r, Z\rangle}(x)) = a$

$q_a[\widetilde{\langle r, Z\rangle}^0(x)] = q_a[x] \quad \forall a \in \Sigma_0$

$q_a[\tilde{o}] = a.$

Ce transducteur est bien linéaire complet et vérifie $\tau_\Phi(D_\Delta \cap K) = \Phi(D_\Delta \cap K)$. *cqfd.*

4.2. ALPHABET GRADUE DE SEMI-DYCK SIMPLE.

Dans la définition des forêts de semi-Dyck, on utilise un alphabet gradué de semi-Dyck où à une parenthèse ouvrante peuvent être associées plusieurs

parenthèses fermantes, ces différentes parenthèses fermantes permettant de « marquer » différemment les divers sous-arbres descendants d'un même noeud. On va démontrer que cette pluralité de parenthèses n'intervient que pour des raisons techniques dans les démonstrations de la partie 3-en particulier elle ne joue aucun rôle dans la définition des forêts de semi-Dyck-et qu'on peut se limiter à considérer des alphabets gradués de semi-Dyck où à toute parenthèse ouvrante est associée une seule parenthèse fermante, comme dans le cas des langages de Dyck.

On dira qu'un alphabet gradué de semi-Dyck $\Delta=\Delta(I, n, d)$ est *simple* si $n(i)=o$ pour tout $i\in I$. Une forêt de semi-Dyck est *simple* si c'est une forêt de semi-Dyck sur un alphabet simple.

Si $\Delta=\Delta(I, n, d)$ est un alphabet gradué de semi-Dyck, le sous-alphabet $\Delta^s\subset\Delta$ défini par $\Delta^s=\{\widehat{o}, \breve{o}\}\cup\{\widehat{i}, \breve{i}^0 \mid i\in I\}$ est simple.

On définit l'homomorphisme linéaire complet Π de Δ dans Δ^s par

$$\Pi(\widehat{o})=\breve{o}$$

$$\Pi(\widehat{i})=\breve{i}$$

$$\Pi(\breve{i}^j)=\breve{i}^0$$

$$\Pi(\breve{o})=\breve{o}.$$

Il est immédiat par définition des forêts de semi-Dyck que $\Pi(D_\Delta)=D_{\Delta^s}$ et que $D_\Delta=\Pi^{-1}(D_{\Delta^s})$.

Enfin si Φ est l'homomorphisme défini en 3.2, on a $\Phi(\breve{i}^j(x))=\Phi(\breve{i}^0(x))=x$ d'où $\Phi=\Phi\circ\Pi$. De même si τ_Φ est le transducteur défini dans le lemme 4.1, $\tau_\Phi=\tau_\Phi\circ\Pi$ sur $K\cap D_\Delta$.

THEOREME 4.2. *Soit F une forêt. Les trois propriétés suivantes sont équivalentes:*

i) *F est une forêt algébrique;*

ii) *il existe une forêt de semi-Dyck* simple D_{Δ^s}, *une forêt reconnaissable F_R et un homomorphisme linéaire Φ tels que $F=\Phi(D_{\Delta^s}\cap F_R)$;*

iii) *il existe une forêt de semi-Dyck* simple D_{Δ^s} *et un transducteur linéaire complet τ tels que $F=\tau(D_{\Delta^s})$.*

Les implications (ii) $\Rightarrow$ (i) et (iii) $\Rightarrow$ (i) sont immédiates.

D'après le théorème 3.1 et le lemme 4.1

$$F = \Phi(D_\Delta \cap K) = \tau_\Phi(D_\Delta \cap K)$$

et donc

$$F = \Phi(\Pi(D_\Delta \cap K)) = \tau_\Phi(\Pi(D_\Delta \cap K)).$$

Or $\Pi(D_\Delta \cap K) = D_{\Delta^g} \cap \Pi(K)$ car $\Pi(D_\Delta \cap K) \subset \Pi(D_\Delta) \cap \Pi(K) =$ $= D_{\Delta^g} \cap \Pi(K) \subset \Pi(\Pi^{-1}(D_{\Delta^g}) \cap K) = \Pi(D_\Delta \cap K)$.

D'où $F = \Phi(D_{\Delta^g} \cap \Pi(K)) = \tau_\Phi(D_{\Delta^g} \cap \Pi(K))$ et comme K est une forêt locale, $\Pi(K)$ est une forêt reconnaissable. *cqfd.*

4.3. GENERATEURS.

Dans le cas des langages, on sait que la classe des algébriques peut être engendrée par transduction rationnelle à partir d'un seul langage de Dyck. Dans le cas des forêts un tel résultat est impossible. En effet supposons qu'il existe une forêt de semi-Dyck D telle que quelle que soit la forêt algébrique F il existe un transducteur linéaire complet τ_F tel que $\tau_F(D) = F$; alors si un arbre t de F contient un symbole de degré n, il existe un arbre de D contenant un symbole de degré au moins égal à n puisque τ_F est linéaire. Donc si D engendre la classes des forêts algébriques par transduction linéaire complète, l'alphabet de D contient des symboles de degré arbitrairement grand et est donc infini.

Par contre si on ne considère que les forêts algébriques formées sur des alphabets dont le degré des symboles est borné, on pourra obtenir un théorème de génération.

Pour tout entier $n > 1$, on note donc Alg_n la classe des forêts algébriques formées sur les alphabets gradués Σ vérifiant $d(\Sigma) \leq n$ et on définit la forêt de semi-Dyck D_n sur l'alphabet gradué de semi-Dyck simple $\Delta(I_n, d)$ où $I_n = \{a, \omega_1, \ldots, \omega_n\}$ et où $d(a) = 1$, $d(\omega_i) = i$ pour $i \in \{1, \ldots, n\}$.

THÉORÈME 4.3. *La classe* Alg_n *est engendrée par transduction linéaire à partir de la forêt de semi-Dyck simple* D_n.

Si τ est une transduction linéaire complète, $\tau(D_n)$ est clairement une forêt de Alg_n.

Réciproquement soit F une forêt algébrique. D'après le théorème 4.2, il existe une forêt de semi-Dyck simple D_Δ sur l'alphabet $\Delta = \Delta(I, d)$ et un transducteur linéaire complet τ tel que $F = \tau(D_\Delta)$. D'après la construction de D_Δ, $d(\Delta) = d(F) \leq n$.

On définit l'homomorphisme ρ de Δ dans D_n de la façon suivante: à tout élément i de I on associe injectivement un entier c_i (i. e. $i \neq i' \Rightarrow c_i \neq c_i'$) et on pose

$$\rho(\widehat{o})=\widehat{o}$$

$$\rho(\breve{o})=\breve{o}$$

pour $i\in I$, $\rho(\widehat{i})=\underbrace{\widehat{a}\ldots\widehat{a}}_{c_i \text{ fois}}\widehat{\omega}_{d(i)}$

$$\rho(\breve{i})=\breve{\omega}_{d(i)}\underbrace{\widehat{a}\ldots\widehat{a}}_{c_i \text{ fois}}$$

D'après la construction de ρ, il est clair que $\rho(D_\Delta)=D_n\cap\rho(\Delta^*)$.

Considerons le transducteur linéaire complet σ dont l'ensemble des états $\{q_0\}\cup\{q_j \mid j\leq \max\{c_i \mid i\in I\}\}\cup\{\bar{q}_j \mid 1\leq k\leq n,\ j\leq \max\{c_i \mid i\in I\}\}$, dont l'état initial est q_0 et dont les règles sont:

$$q_0(\widehat{o}(x))=\widehat{o}(q_0[x])$$

$$q_0(\breve{o})=\breve{o}$$

$$q_i[\widehat{a}(x)]=q_{i+1}[x] \text{ si } i\neq o$$

$$q_{c_i}[\widehat{\omega}_{d(i)}(x_1,\ldots,x_{d(i)})]=\widehat{i}(q_0[x_1],\ldots,q_0[x_{d(i)}])$$

$$q_0[\breve{\omega}_{d(i)}(x)]=\bar{q}_1[x]$$

$$\bar{q}_i[\breve{a}(x)]=\bar{q}_{i+1}[x]$$

$$\bar{q}_i[\breve{a}(x)]=\breve{j}(q_0[x]) \text{ si } c_j=i.$$

Il est immédiat par construction que $\sigma(\rho(t))=t$ d'où

$$D_\Delta=\sigma(\rho(D_\Delta))=\sigma(D_n\cap\rho(\Delta^*)) \text{ et } \tau(D_\Delta)=\tau(\sigma(D_n\cap\rho(\Delta^*))).$$

Comme τ, σ sont des transducteurs linéaires complets, il existe un transducteur linéaire τ' tel que $\tau(D_\Delta)=\tau(\sigma(D_n\cap\rho(\Delta^*)))=\tau'(D_n)$. *cqfd.*

On remarque que dans le cas particulier ou $n=1$ c'est à dire le cas des langages on retrouve le théorème bien connu de génération de la classe des langages algébriques.

REFERENCES

[1] ARNOLD A., DAUCHET M., *Une relation d'équivalence décidable sur la classe des forêts régulières.* Publications du Labo de Calcul de Lille n° 52 (1975).

[2] ARNOLD A., DAUCHET M., *Un théorème de duplication pour les forêts algébriques.* J. Comput. System Sci. **13** (1976), 223-244.

[3] AUTEBERT J. M., BEAUQUIER J., *Une caractérisation des générateurs standard.* RAIRO R 1 (1974), 63-83.

[4] BOASSON L., *Two iteration theorems for some families of languages.* J. Comput. System Sci. **7** (1973), 583-596.

[5] CHOMSKY N., SCHUTZENBERGER M. P., *The algebraic theory of context-free languages.* In « Computer programming and Formal systems » (Braffort et Hirschberg Eds) North-Holland, Amsterdam (1963), 118-161.

[6] COURCELLE B., *Recursive schemas, algebraic trees and deterministic languages.* Rapport IRIA n° 78 (1974).

[7] ENGELFRIET J., *Tree automata and tree grammars.* Daimi-Report FN-10. University of Aarhus, Danemark (1975).

[8] ENGELFRIET J., SCHMIDT E. M., *IO and OI.* Daimi-Report PB 47. University of Aarhus, Danemark (1975).

[9] GINSBURG S. GREIBACH S., *Principal AFL.* J. Comput. System Sci. **4** (1970), 308-338.

[10] MAIBAUM T. S. E., *A generalized approach to formal languages.* J. Comput. System Sci. **8** (1974), 409-439.

[11] NIVAT M., *Transductions des langages de Chomsky.* Ann. Inst. Fourier (Grenoble) **18** (1968), 339-456.

[12] NIVAT M., *Langages algébriques sur le magma libre et sémantique des schémas de programme.* in « Automata, Languages and Programming » (Nivat Ed) North-Holland, Amsterdam (1973), 293-307.

[13] NIVAT M., *On the interpretation of recursive program schemes.* Rapport IRIA n° 84 (1974).

[14] RHODES J., SHAMIR E., *Complexity of grammars by group-theoretic method.* Combinatorial theory **4** (1968), 222-239.

[15] ROUNDS W. C., *Mappings and grammars on trees.* Math. Systems Theory **4** (1970), 257-287.

[16] ROUNDS W. C., *Tree oriented proofs of some theorems on context-free and indexed languages.* 2nd ACM Symp. on theory of computing (1970), 109-116.

[17] SCHUTZENBERGER M. P., *On context-free languages and push-down automata.* Information and Control **6** (1963), 246-264.

Gubbio - Tipografia Oderisi Editrice - 1977

View publication stats