JOURNAL OF COMPUTER AND SYSTEM SCIENCES 13, 223–244 (1976)

Un Théorème de Duplication pour les Forêts Algébriques

ANDRÉ ARNOLD ET MAX DAUCHET

Service Informatique, Université de Lille I, Villeneuve-D'Ascq, France

Received November 20, 1975; revised May 5, 1976

On caractérise les forêts algébriques dont tous les arbres sont de la forme $\alpha(t, t)$. On utilise cette caractérisation pour montrer que la classe des forêts algébriques n'est pas fermée par homomorphisme non linéaire, et pour montrer qu'il existe des forêts reconnaissables généralisées, au sens de Maibaum, dont le feuillage n'est pas une forêt algébrique.

INTRODUCTION

L'introduction, dans la théorie des langages, de systèmes grammaticaux où il est possible d'effectuer des dérivations en parallèle, tels que les systèmes ETOL de Rozenberg [13], a permis de démontrer des théorèmes de duplication ("copying theorems") de la forme: si $L' = \{w \# w \mid w \in L\}$ est un langage de la classe $\mathscr{F}_1$, alors L' et L sont des langages de la classe $\mathscr{F}_2$. Ce théorème a été montré par Skyum [14] (voir aussi Engelfriet et Skyum [8]) pour $\mathscr{F}_1 = \text{ETOL}$ et $\mathscr{F}_2 = \text{EDTOL}$. Fischer [9] a montré un théorème analogue: si $L' = \{w \# w \# w \mid w \in L\}$ est un langage indexé, alors L est un langage EDTOL. Engelfriet et Skyum [8] ont démontré d'autres théorèmes de duplication pour des classes de langages définis comme feuillages de certaines classes de forêts.

Dans cet article nous démontrons un théorème de duplication pour les forêts elles-mêmes. Les classes de forêts que nous considérons, outre la classe des forêts reconnaissables (Rounds [11]) sont la classe des forêts algébriques ("creative sets", Rounds [11], "context-free tree languages", Downey [4], Engelfriet et Schmidt [7], "context-free sets" Maibaum [10]) et celle des forêts que nous appelons corégulières définies dans [2].

Nous avons montré [2] que cette classe des forêts corégulières est identique à celle des transductés déterministes de forêts reconnaissables monadiques, c'est donc la même que les classes EDTOLT et ILBT de Downey [4, 5]. Notre théorème s'énonce: $F' = \{\alpha(t, t) \mid t \in F\}$ est une forêt algébrique si et seulement si F et F' sont des forêts corégulières.

Les feuillages des forêts algébriques étant les langages indexés [11, 12] et ceux des forêts corégulières étant les langages EDTOL [2, 5, 6], ce résultat est en quelque sorte très voisin de ceux énoncés ci-dessus concernant les langages.

Comme il existe une forêt F reconnaissable (et donc algébrique) qui n'est pas corégulière, la forêt $\{\alpha(t, t) \mid t \in F\}$ n'est pas algébrique. Nous en déduisons (i) que la classe des forêts algébriques n'est pas fermée par homomorphisme non linéaire, (ii) que les transductées déterministes de forêts reconnaissables ne sont pas toutes algébriques, ce qui répond à une question posée par Downey [5], (iii) qu'il existe des forêts reconnaissables généralisées au sens de Maibaum dont le feuillage n'est pas algébrique, ce qui contredit le "Yield theorem" de Maibaum [10]. Les propriétés (i) et (iii) ont été obtenues indépendamment et différemment par Engelfriet et Schmidt [7].

Notons que, comme dans le cas des théorèmes de duplication pour les langages, l'intérêt de notre théorème de duplication, ainsi que l'ont illustré les applications, est de permettre une démonstration facile de la non-algébricité de certaines forêts, puisqu'il la réduit à une démonstration de non-corégularité. Or celle-ci est plus souvent plus simple puisque les branches d'une forêt corégulière forment un langage rationnel.

Ce travail comprend cinq parties et une annexe. La première consiste en rappels sur les arbres, forêts, grammaires et forêts algébriques. La deuxième partie traite des forêts corégulières. Dans la troisième on définit les dérivations initiales dans une grammaire algébrique. Le théorème de duplication est démontré dans la quatrième partie et les applications sont faites dans la cinquième. L'annexe est consacrée à la démonstration d'une propriété technique des grammaires algébriques.

1. Préliminaires

Un *alphabet gradué* est un couple (Σ, d) où Σ est un ensemble fini et d une application de Σ dans N appelée le degré. Un alphabet gradué sera le plus souvent noté simplement Σ. Si Σ est un alphabet gradué, pour tout entier n, on note Σ_n l'ensemble $\{\sigma \in \Sigma \mid d(\sigma) = n\}$ des symboles de degré n.

L'ensemble des *arbres* sur un alphabet gradué Σ, noté T_Σ, est défini inductivement par:

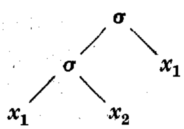
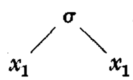
$$\Sigma_0 \subset T_\Sigma,$$
$$\text{si } \sigma \in \Sigma_n, \qquad t_1, \dots, t_n \in T_\Sigma \text{ alors } \sigma(t_1, \dots, t_n) \in T_\Sigma.$$

Si E est un ensemble quelconque, l'ensemble des *arbres indexés par* E, noté $T_\Sigma(E)$, est l'ensemble des arbres sur l'alphabet Σ' défini par $\Sigma_0' = \Sigma_0 \cup E$ et $\forall i > 0$, $\Sigma_i' = \Sigma_i$.

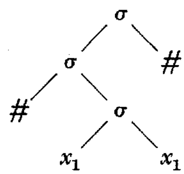
On dira qu'un arbre est *indexé* s'il est indexé par l'ensemble de *variables* $\mathscr{X} = \{x_1, \dots, x_n, \dots\}$. En particulier, tout arbre de T_Σ est un arbre indexé.

Soit t un arbre indexé de $T_\Sigma(\{x_1, \dots, x_n\})$, qu'on écrira parfois $t(x_{i_1}, \dots, x_{i_p})$ pour faire apparaître les variables qu'il contient, et soient $t_1, \ldots, t_n$ des arbres indexés; on note $t(t_{i_1}, \ldots, t_{i_p})$ l'arbre indexé obtenu en remplacant chaque occurence de la variable x_i par l'arbre t_i. Il faut remarquer que les variables $x_{i_1}, \ldots, x_{i_p}$ *ne sont pas* nécessairement distinctes.

EXEMPLE.

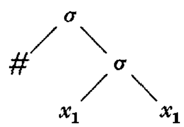
$t(x_1, x_2, x_1) =$  ; $t_1 = \#$; $t_2 =$  .

Alors

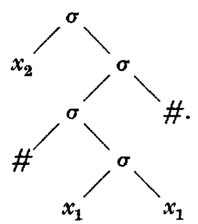
$t(t_1, t_2, t_1) =$ 

On dira qu'un arbre indexé t est un *sous-arbre terminal* de l'arbre indexé t', s'il existe un arbre indexé $t''(x_{i_1}, \ldots, x_{i_{j-1}}, x_{i_j}, x_{i_{j+1}}, \ldots, x_{i_q})$, où $k \neq j \Rightarrow i_k \neq i_j$, tel que $t' = t''(x_{i_1}, \ldots, x_{i_{j-1}}, t, x_{i_{j+1}}, \ldots, x_{i_q})$. Si t s'écrit $\sigma(t_1, \ldots, t_n)$ avec $\sigma \in \Sigma_n$, on dira que σ est *racine* de t.

EXEMPLE. L'arbre



est un sous-arbre terminal de



La relation "être un sous-arbre terminal de" est transitive. On appelle *forêt* sur Σ toute partie de T_Σ.

On appelle *grammaire algébrique* (ou "context-free") tout quadruplet $\langle V, \Sigma, X_0, \mathscr{R}\rangle$ où V (ensemble des symboles non terminaux) et Σ (ensemble des symboles terminaux) sont des alphabets gradués; X_0, l'axiome, est un symbole de V de degré 0 et $\mathscr{R}$ est un ensemble fini de *règles*.

Une règle est un couple $\langle X(x_1, ..., x_n), t(x_{i_1}, ..., x_{i_p})\rangle$ écrit $X(x_1, ..., x_n) \to t(x_{i_1}, ..., x_{i_p})$ où X est un symbole non-terminal de degré n et où t est un arbre de $T_{\Sigma \cup V}(\{x_1, ..., x_n\})$.

Soit G une grammaire algébrique $\langle V, \Sigma, X_0, \mathscr{R}\rangle$. On dira que *l'arbre indexé t' de $T_{\Sigma\cup V}(\mathscr{X})$ se dérive immédiatement en l'arbre indexé t'' par application de la règle $r = X(x_1, ..., x_n) \to t(x_{i_1}, ..., x_{i_p})$*, ce qui se notera $t' \Rightarrow_r t''$, si $t' = t_0(X(t_1, ..., t_n))$ et $t'' = t_0(t(t_{i_1}, ..., t_{i_p}))$. On dira que cette dérivation immédiate est *descendante* ("outermost", Downey [4], "*OI*", Engelfriet et Schmidt [7]) s'il n'existe aucun sous-arbre terminal t''' de t' de la forme $X'(t_1''', ..., t_p''')$ avec $X' \in V$, tel que $X(t_1, ..., t_n)$ est un sous-arbre terminal de t'''.

On appelle *dérivation de longueur n* toute séquence $t_0 \Rightarrow_{r_1} t_1 \Rightarrow_{r_2} t_2 \cdots \Rightarrow_{r_n} t_n$ telle que pour tout $i \in \{1, ..., n\}$ $t_{i-1} \Rightarrow_{r_i} t_i$ est une dérivation immédiate. On dira que cette dérivation est *descendante* si toutes les dérivations immédiates qui la composent sont descendantes. On dira que *t' se dérive en t''*, ce qu'on notera $t' \overset{*}{\Rightarrow}_G t''$, s'il existe une dérivation $t_0 \Rightarrow_{r_1} t_1 \cdots \Rightarrow_{r_n} t_n$ telle que $t_0 = t'$ et $t_n = t''$.

On sait [4, 7, 11] que si $t' \overset{*}{\Rightarrow}_G t''$ et si $t'' \in T_\Sigma(\mathscr{X})$, alors t' se dérive en t'' par une dérivation descendante. Aussi à partir de maintenant, on ne considérera plus que des dérivations descendantes.

On appelle *forêt engendrée par la grammaire G*, notée $F(G)$ l'ensemble $\{t \in T_\Sigma \mid X_0 \overset{*}{\Rightarrow}_G t\}$.

2. Forêts Corégulières

La classe des forêts corégulières introduite dans [2] peut aussi se définir comme celle des transductées déterministes de forêts reconnaissables monadiques. Elle est donc identique aux classes ILBT et EDTOLT de Downey [4, 5]. Intuitivement une forêt corégulière est une forêt algébrique engendrée par une grammaire où dans les dérivations on ne réécrit toujours que la racine des arbres.

Soit $G = \langle V, \Sigma, X_0, \mathscr{R}\rangle$ une grammaire algébrique. Cette grammaire sera dite *corégulière* si toutes ses règles sont de la forme $X(x_1, ..., x_n) \to t$ avec $t \in T_\Sigma(\mathscr{X})$, ou $X(x_1, ..., x_n) \to X'(t_1', ..., t_p')$ avec $X' \in V$ et $t_1', ..., t_p' \in T_\Sigma(\mathscr{X})$.

La dénomination "corégulière" vient de ce que ces grammaires sont en quelque sorte duales des grammaires régulières en ce sens que dans les unes on ne dérive que la racine d'un arbre, et dans les autres que les feuilles. Il faut noter que dans le cas monadique (i.e., celui des langages), ces grammaires correspondent aux grammaires linéaires à droite et à gauche qui engendrent la même classe de langages, les rationnels, ce qui n'est pas le cas dans le cas des arbres [2].

Une forêt est *corégulière* si elle est engendrée par une grammaire corégulière.

PROPOSITION 2.1. *La classe des forêts corégulières est fermée par homomorphisme.*

Cette proposition se démontre immédiatement en tenant compte du fait qu'une forêt est corégulière si elle est la transductée déterministe d'une forêt reconnaissable monadique, et que le composé d'un transducteur déterministe et d'un homomorphisme est encore un transducteur déterministe [11, 15] *cqfd.*

Nous allons montrer maintenant une condition nécessaire très simple pour qu'une forêt soit corègulière et qui porte sur l'ensemble de ses branches que nous définissons d'abord.

Soit Σ un alphabet gradué. On définit l'alphabet gradué Σ' par $\Sigma_0' = \Sigma_0$ et $\Sigma_1' = \{\sigma' \mid \sigma \in \Sigma - \Sigma_0\}$, et le transducteur linéaire non-déterministe à un seul état β de T_Σ dans $T_{\Sigma'}$ par

$$\text{si } \sigma \in \Sigma_0\,,\ q[\sigma] = \sigma,$$
$$\text{si } \sigma \in \Sigma - \Sigma_0\,,\ q[\sigma(x_1\,,\ldots,\,x_n)] = \sigma'(q[x_i]) \text{ pour tout } i \leqslant n.$$

Si F est une forêt de T_Σ *l'ensemble des branches de* F est $\beta(F)$ (Rounds [11]).

PROPOSITION 2.2. *Si F est une forêt corégulière, l'ensemble de ses branches est une forêt reconnaissable.*

Cette proposition est un corollaire immédiat du résultat de Rounds [11] qui démontre que l'ensemble des branches d'une forêt obtenue par transduction d'une forêt reconnaissable est reconnaissable.

PROPOSITION 2.3. *Il existe des forêts algébriques et même reconnaissables qui ne sont pas corégulières.*

Soit l'alphabet gradué $\Delta = \Delta_0 \cup \Delta_2$ avec $\Delta_2 = \{\delta\}$ et $\Delta_0 = \{\#\}$. Nous avons montré en [2] que la forêt reconnaissable T_Δ n'était pas corégulière. La démonstration que nous ne redonnons pas ici formalise les faits intuitifs suivants: si une forêt obtenue par transduction déterministe d'une forêt reconnaissable monadique contient des arbres arbitrairement "larges", ce qui est le cas de T_Δ, ces arbres contiennent nécessairement des sous-arbres terminaux suffisamment grands identiques et on peut montrer que ce n'est pas le cas de T_Δ. La proposition 2.2. fournit immédiatement des exemples de forêts algébriques non corégulières. Par exemple sur l'alphabet $A = A_1 \cup A_0$ avec $A_1 = \{a, b\}$ et $A_0 = \{\#\}$, la forêt $\{a^n b^n \# \mid n > 0\}$ est une forêt algébrique non corégulière.

3. Grammaires Quasi-Normales Complètes. Dérivations Initiales

Il est bien connu que tout langage algébrique sans mot vide peut être engendré par une grammaire ne contenant pas de règles $X \to \Lambda$. Par contre, dans le cas des forêts algébriques il n'est pas possible d'éliminer complètement les règles de la forme $X(x_1, ..., x_n) \to x_i$. Nous allons cependant montrer que toute forêt algébrique admet une grammaire d'une forme particulière telle que l'utilisation des règles $X(x_1, ..., x_n) \to x_i$ y est très limitée.

On dira qu'une grammaire algébrique $G = \langle V, \Sigma, X_0, \mathscr{R} \rangle$ est *sous forme quasi-normale* si toutes ses règles sont de la forme:

1. $X(x_1, ..., x_n) \to x_i$, où $X \in V_n$ (règles effacantes),
2. $X(x_1, ..., x_n) \to f(x_{i_1}, ..., x_{i_p})$ où $X \in V_n$ et $f \in \Sigma_p$ (règles finales),
3. $X(x_1, ..., x_n) \to X'(x_1, ..., x_n)$ où $X, X' \in V_n$ (règles monotones),
4. $X(x_1, ..., x_n) \to$

$$\begin{tikzcd}[column sep=small, row sep=small] & & & X \arrow[dll, dash] \arrow[drr, dash] & & & \\ & Y_1 \arrow[dl, dash] \arrow[dr, dash] & & & & Y_m \arrow[dl, dash] \arrow[dr, dash] & \\ x_1 & & x_n & & x_1 & & x_n \end{tikzcd}$$

où $X, Y_1, ..., Y_m \in V_n$ et $X' \in V_m$ (règles croissantes).

Cette forme quasi normale est une légère généralisation de la forme normale (Maibaum [10]).

On dira qu'une grammaire est *totale* si chaque symbole non terminal engendre une forêt non vide.

On dira qu'une grammaire quasi-normale G est *complète* si elle vérifie

(i) si $X(x_1, ..., x_n) \to$

$$\begin{tikzcd}[column sep=small, row sep=small] & & & X' \arrow[dll, dash] \arrow[drr, dash] & & & \\ & Y_1 \arrow[dl, dash] \arrow[dr, dash] & & & & Y_m \arrow[dl, dash] \arrow[dr, dash] & \\ x_1 & & x_n & & x_1 & & x_n \end{tikzcd}$$

est une règle croissante de G et $X'(x_1, ..., x_n) \to x_i$ une règle effaçante de G, alors la règle monotone $X(x_1, ..., x_n) \to Y_i(x_1, ..., x_n)$ est une règle de G.

(ii) si $X(x_1, ..., x_n) \to X'(x_1, ..., x_n)$ est une règle monotone de G et $X'(x_1, ..., x_n) \to x_i$ est une règle effaçante de G, la règle $X(x_1, ..., x_n) \to x_i$ est une règle de G.

Proposition 3.1. *Quelle que soit la forêt algébrique F, il existe une grammaire quasi-normale totale complète G telle que $F = F(G)$.*

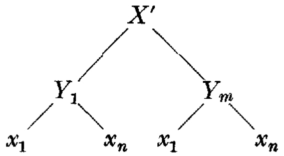
Toute forêt algébrique admet une grammaire normale [10], nous montrons en Annexe qu'on peut la supposer totale; elle admet donc à fortiori une grammaire quasi-normale totale. Soit donc $G = \langle V, \Sigma, X_0, \mathscr{R}\rangle$ cette grammaire telle que $F = F(G)$. Nous allons montrer qu'il existe une grammaire quasi-normale totale complète G^* telle que $F(G) = F(G^*)$.

On pose $\mathscr{R}_0 = \mathscr{R}$ et on considère la suite de grammaires $G_i = \langle V, \Sigma, X_0, \mathscr{R}_i\rangle$ où G_{i+1} est définie à partir de G_i de la façon suivante:

on définit $\mathscr{R}_i'$ par

$X(x_1, ..., x_n) \to Y_j(x_1, ..., x_n) \in \mathscr{R}_i'$ ssi il existe

$X(x_1, ..., x_n) \to$



et $X'(x_1, ..., x_m) \to x_j$ dans $\mathscr{R}_i$;

et $\mathscr{R}_i''$ par

$$X(x_1, ..., x_n) \to x_j \in \mathscr{R}_i'' \quad \text{ssi il existe} \quad X(x_1, ..., x_n) \to X'(x_1, ..., x_n)$$

et $X'(x_1, ..., x_n) \to x_j$ dans $\mathscr{R}_i$.

On pose alors $\mathscr{R}_{i+1} = \mathscr{R}_i \cup \mathscr{R}_i' \cup \mathscr{R}_i''$. Comme on forme $\mathscr{R}_{i+1}$ en ajoutant à $\mathscr{R}_i$ des règles monotones ou effaçantes, G_{i+1} est encore une grammaire quasi-normale. De plus, il est facile de voir que le nombre de règles effaçantes sur l'alphabet V est au plus $\sum_{X \in V} d(X)$ qui est un nombre fini, et celui des règles monotones est $\sum_0^\infty \text{Card}\,(V_n)^2 < (\text{Card}\, V)^2$ qui est aussi un nombre fini. La suite $\mathscr{R}_i$ qui est croissante et majorée devenir stationnaire à partir d'un certain rang n_0. On pose alors $G^* = G_{n_0}$. Il reste à montrer que pour tout $i, F(G_i) = F(G_{i+1})$. Comme $\mathscr{R}_i \subset \mathscr{R}_{i+1}$, $F(G_i) \subset F(G_{i+1})$. D'autre par, d'après la construction de $\mathscr{R}_{i+1}$, si $X(x_1, ..., x_n) \to t \in \mathscr{R}_{i+1} - \mathscr{R}_i$, alors $X(x_1, ..., x_n) \overset{*}{\Rightarrow}_{G_i} t$ d'où $F(G_{i+1}) \subset F(G_i)$. Comme on n'a pas introduit de nouveaux symboles non terminaux, si G est totale, G^* est encore totale. *cqfd.*

Soit $G = \langle V, \Sigma, X_0, \mathscr{R}\rangle$ une grammaire quasi-normale et soit une dérivation $X_0 \Rightarrow_{r_1} t_1 \Rightarrow_{r_2} \cdots \Rightarrow_{r_n} t_n$ dans G. Comme cette dérivation est descendante, si la racine de l'arbre t_i est un symbole non terminal, c'est ce symbole qui sera réécrit par la règle r_{i+1}. Cette constatation nous permet de définir les sous-dérivations initiales.

Soit $\delta = X_0 \Rightarrow_{r_1} t_1 \Rightarrow_{r_2} \cdots \Rightarrow_{r_n} t_n$ une dérivation dans G. On dira que la sous-dérivation $\delta' = X_0 \Rightarrow_{r_1} t_1 \Rightarrow_{r_2} \cdots \Rightarrow_{r_i} t_i$ est *initiale* si quelque soit $j \in \{1, ..., i-1\}$, t_j est de la forme $X(t_1', ..., t_p')$ avec $X \in V_p$. Elle sera dite initiale *maximale* si elle est initiale et si, de plus, $t_n = f(t_1', ..., t_p')$ avec $f \in \Sigma$.

PROPOSITION 3.2. (i) *Toute dérivation* $X_0 \Rightarrow_{r_1} t_1 \cdots \Rightarrow_{r_n} t_n$, où $t_n \in T_\Sigma$ *admet une seule sous-dérivation initiale maximale* $X_0 \Rightarrow_{r_1} t_1 \cdots \Rightarrow_{r_{i_0}} t_{i_0}$.

(ii) *Une sous-dérivation* $X_0 \Rightarrow_{r_1} \cdots \Rightarrow_{r_j} t_j$ *est alors initiale* ssi $j \leqslant i_0$.

(iii) *Pour tout* $i < i_0$, $t_i \in T_V$ et $t_{i_0} = f(t_1', \ldots, t_p')$ *avec* $t_1', \ldots, t_p' \in T_V$.

Comme $X_0 \in T_V$ et qu'on ne considère que des dérivations descendantes, la proposition découle immédiatement du fait que si $t_i \in T_V$, alors $t_{i+1} \in T_V$ si r_{i+1} est une règle croissante, monotone ou effaçante et $t_{i+1} = f(t_1', \ldots, t_p')$ avec $t_1', \ldots, t_p' \in T_V$ si r_{i+1} est une règle finale. Comme $t_n \in T_\Sigma$, la dérivation comporte au moins une règle finale et il suffit de prendre i_0 le plus petit entier i tel que r_i est une règle finale. *cqfd.*

Nous allons montrer maintenant qu'on peut supposer qu'une sous-dérivation initiale dans une grammaire quasi-normale complète ne contient pas de règles effaçantes, ce qui permettra par la suite de faire un rapprochement avec les forêts corégulières.

PROPOSITION 3.3. *Soit* $G = \langle V, \Sigma, X_0, \mathscr{R} \rangle$ *une grammaire quasi-normale complète. Soit* $X_0 \overset{*}{\Rightarrow}_G t$, *ou* $t \in T_\Sigma$ *une dérivation dans* G *et soit* $X_0 \Rightarrow_{r_1} t_1 \cdots \Rightarrow_{r_n} t_n$ *sa sous-dérivation initiale maximale. Alors il existe une sous-dérivation initiale maximale* $X_0 \Rightarrow_{r_1'} t_1' \cdots \Rightarrow_{r_p'} t_p' = t_n$ *qui ne contient aucune règle effaçante.*

Soit $\delta = X_0 \Rightarrow_{r_1} t_1 \cdots \Rightarrow_{r_n} t_n$ une dérivation initiale maximale. On va montrer que si le nombre de règles effaçantes utilisées dans δ est $p \geqslant 1$, il existe une dérivation initiale maximale $\delta' = X_0 \Rightarrow_{r_1'} t_1' \cdots \Rightarrow_{r_i'} t_i' = t_n$ qui utilise $p - 1$ règles effaçantes. La proposition se déduit alors immédiatement de ce résultat.

Soit donc

$$r_i = \begin{array}{c} X \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \to x_i$$

la première règle effaçante utilisée dans δ. Comme X_0 ne peut être la partie gauche d'une règle effaçante, i est strictement supérieur à 1; la règle r_{i-1} n'étant ni effaçante, ni finale, elle est de la forme (1)

$$\begin{array}{c} X' \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \to \begin{array}{c} X \\ \diagup \qquad \diagdown \\ \begin{array}{c} Y_1 \\ \diagup \ \diagdown \\ x_1 \quad x_n \end{array} \quad \begin{array}{c} Y_m \\ \diagup \ \diagdown \\ x_1 \quad x_n \end{array} \end{array}$$

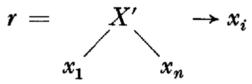
ou (2)

$$\begin{array}{c} X' \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \to \begin{array}{c} X \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array}.$$

Dans le cas 1, la règle

$$r = \begin{array}{c} X' \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \to \begin{array}{c} Y_i \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array}$$

appartient aussi à G puisque G est complète et la dérivation initiale maximale $X_0 \Rightarrow_{r_1} t_1 \cdots \Rightarrow_{r_{i-2}} t_{i-2} \Rightarrow_r t_i \Rightarrow_{r_{i+1}} t_{i+1} \cdots \Rightarrow_{r_n} t_n$ contient bien $p - 1$ règles effaçantes. Dans le cas 2, la règle



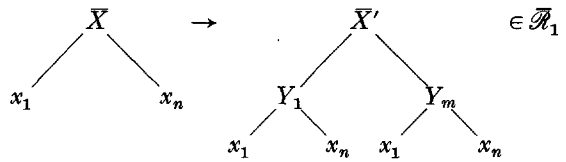
appartient à G et la dérivation initiale maximale $X_0 \Rightarrow_{r_1} \cdots \Rightarrow_{r_{i-2}} t_{i-2} \Rightarrow_r t_i \cdots \Rightarrow_{r_n} t_n$ contient p règles effaçantes, mais cette fois-ci, la première règle effaçante utilisée est la $(i - 1)$ème. On recommence alors le procédé de "réduction" utilisé ci-dessus. Au bout d'un certain nombre d'étapes (inférieur à i), on va certainement se trouver dans le cas 1, car sinon on obtiendrait une dérivation initiale maximale dont la première règle effaçante serait la première ce qui, nous l'avons vu ci-dessus, est impossible. *cqfd*

4. Le Théorème de Duplication

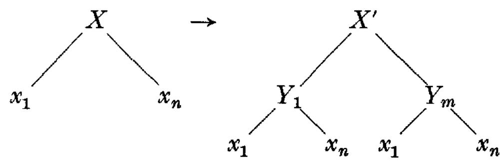
Avant de démontrer le théorème de duplication, nous allons utiliser les résultats de la partie précédente pour obtenir une relation entre forêts algébriques et forêts corégulières.

Soit $G = \langle V, \Sigma, X_0, \mathscr{R} \rangle$ une grammaire quasi-normale complète. On forme la grammaire corégulière $\bar{G}$ de la façon suivante:

1. L'alphabet non-terminal de $\bar{G}$ est $\bar{V} = \{\bar{X} \mid X \in V\}$ avec $\forall \bar{X} \in \bar{V}, d(\bar{X}) = d(X)$;
2. L'axiome de $\bar{G}$ est $\bar{X}_0$;
3. L'alphabet terminal de $\bar{G}$ est $V \cup \Sigma$;
4. L'ensemble $\bar{\mathscr{R}}$ des règles de $\bar{G}$ est $\bar{\mathscr{R}}_1 \cup \bar{\mathscr{R}}_2 \cup \bar{\mathscr{R}}_3$ où



ssi



est une règle croissante de $\mathscr{R}$.

$$\begin{array}{ccc} & \bar{X} & \\ / & & \backslash \\ x_1 & & x_n \end{array} \rightarrow \begin{array}{ccc} & \bar{X}' & \\ / & & \backslash \\ x_1 & & x_n \end{array} \in \bar{\mathscr{R}}_2 \quad \text{ssi} \quad \begin{array}{ccc} & X & \\ / & & \backslash \\ x_1 & & x_n \end{array} \rightarrow \begin{array}{ccc} & X' & \\ / & & \backslash \\ x_1 & & x_n \end{array}$$

est une règle monotone de $\mathscr{R}$.

$$\begin{array}{ccc} & \bar{X} & \\ / & & \backslash \\ x_1 & & x_n \end{array} \rightarrow \begin{array}{ccc} & f & \\ / & & \backslash \\ x_{i_1} & & x_{i_p} \end{array} \in \bar{\mathscr{R}}_3 \quad \text{ssi} \quad \begin{array}{ccc} & X & \\ / & & \backslash \\ x_1 & & x_n \end{array} \rightarrow \begin{array}{ccc} & f & \\ / & & \backslash \\ x_{i_1} & & x_{i_p} \end{array}$$

est une règle finale de $\mathscr{R}$.

Il est immédiat par construction que la grammaire $\bar{G}$ est corégulière.

PROPOSITION 4.1. *Avec les notations ci-dessus,* $\forall t \in T_\Sigma$, $X_0 \overset{*}{\Rightarrow}_G t$ ssi *il existe* $t' \in T_{\bar{V} \cup \Sigma}$ *tel que* $\bar{X}_0 \overset{*}{\Rightarrow}_{\bar{G}} t'$ *et* $t' \overset{*}{\Rightarrow}_G t$. *De plus* t' *est de la forme* $f(t_1, \ldots, t_p)$ avec $f \in \Sigma_p$ *et* $t_1, \ldots, t_p \in T_V$.

A. *Condition Suffisante*

Considérons la projection Π de $\bar{V} \cup V \cup \Sigma$ dans $V \cup \Sigma$ définie par

$$\forall \bar{X} \in \bar{V}, \Pi(\bar{X}) = X,$$
$$\forall Z \in V \cup \Sigma, \qquad \Pi(Z) = Z.$$

D'après la construction de $\bar{G}$, il est clair que si $t \overset{*}{\Rightarrow}_{\bar{G}} t'$ alors $\Pi(t) \overset{*}{\Rightarrow}_G \Pi(t')$. D'où, si il existe $t' \in T_{\bar{V} \cup \Sigma}$ tel que $\bar{X}_0 \overset{*}{\Rightarrow}_{\bar{G}} t'$ alors $X_0 \overset{*}{\Rightarrow}_G \Pi(t')$, et comme $t' \in T_{V \cup \Sigma}$, $\Pi(t') = t'$. On en déduit donc que $X_0 \overset{*}{\Rightarrow}_G t'$ et $t' \overset{*}{\Rightarrow}_G t$ et donc que $X_0 \overset{*}{\Rightarrow}_G t$.

B. *Condition Nécessaire*

Soit $X_0 \Rightarrow_{r_1} t_1 \cdots \Rightarrow_{r_n} t_n$ la sous-dérivation initiale maximale de la dérivation $X_0 \overset{*}{\Rightarrow}_G t$. D'après la proposition 3.3, on peut supposer qu'elle ne contient aucune règle effaçante. Pour $i < n$, d'après la proposition 3.2, la racine de t_i est un symbole non terminal X; on forme l'arbre $\bar{t}_i$ en remplaçant cette racine par $\bar{X}$. Si $t_i \Rightarrow_{r_{i+1}} t_{i+1}$, il est immédiat d'après la construction de $\bar{G}$ et en tenant compte du fait que r_{i+1} n'est pas une règle effaçante, qu'il existe une règle r_{i+1} de $\bar{G}$ telle que $\bar{t}_i \Rightarrow_{\bar{r}_{i+1}} \bar{t}_{i+1}$, d'où $\bar{X}_0 \overset{*}{\Rightarrow}_{\bar{G}} t_n$ et comme $t_n \overset{*}{\Rightarrow}_G t$ on a bien le résultat proposé. D'après la proposition 3.2, t_n a bien la forme indiquée. *cqfd.*

Nous sommes maintenant en mesure de démontrer le théorème de duplication.

THÉORÈME 4.1. (*Théorème de duplication*). *Soit* Σ *un alphabet gradué,* F *une forêt de* T_Σ, α *un symbole de degré 2 n'appartenant pas à* Σ. *La forêt* $F' = \{\alpha(t, t) \mid t \in F\}$ *est algébrique si et seulement si* F *et* F' *sont des forêts corégulières.*

L'idée qui est à la base de la démonstration est la suivante: si une grammaire algébrique engendre $\alpha(t, t)$ on doit pouvoir engendrer t avant de le dupliquer; comme la duplication se fait nécessairement par dérivation de la racine, il faut donc que t puisse aussi se dériver par la racine, et donc que t appartienne à une forêt corégulière.

A. *Condition Suffisante*

Si F est une forêt corégulière engendrée par la grammaire corégulière G, F' sera engendrée par la grammaire G' obtenue en remplaçant les règles de G de la forme $X(x_1, ..., x_n) \to t'$ avec $t' \in T_\Sigma(\mathscr{X})$ par les règles $X(x_1, ..., x_n) \to \alpha(t', t')$. Comme dans une dérivation dans une grammaire corégulière seule la dernière règle appliquée peut être de ce type on aura bien

$$X_0 \overset{*}{\underset{G}{\Rightarrow}} \begin{array}{c} X \\ / \quad \backslash \\ t_1 \qquad t_n \end{array} \underset{G}{\Rightarrow} t'(t_{i_1}, ..., t_{i_p}) = t$$

$$\text{ssi} \quad X_0 \overset{*}{\underset{G'}{\Rightarrow}} \begin{array}{c} X \\ / \quad \backslash \\ t_1 \qquad t_n \end{array} \overset{*}{\underset{G'}{\Rightarrow}} \alpha(t'(t_{i_1}, ..., t_{i_p}), \qquad t'(t_{i_1}, ..., t_{i_p})) = \alpha(t, t).$$

De plus toute forêt corégulière est bien algébrique.

B. *Condition Nécessaire*

Soit F une forêt de T_Σ et soit $G = \langle V, \Sigma \cup \{\alpha\}, X_0, \mathscr{R}\rangle$ une grammaire quasi-normale totale complète qui engendre $F' = \{\alpha(t, t) \mid t \in F\}$. D'après la proposition 4.1, il existe une forêt corégulière F'' de $T_{V \cup \Sigma \cup \{\alpha\}}$ telle que $\alpha(t, t) \in F'$ ssi il existe $t' \in F''$ tel que $t' \overset{*}{\Rightarrow}_G \alpha(t, t)$. De plus t' est de la forme $\alpha(t_1, t_2)$ avec $t_1, t_2 \in T_V$, toujours d'après la proposition 4.1.

Pour que $\alpha(t_1, t_2) \overset{*}{\Rightarrow}_G \alpha(t, t)$, il faut que $t_1 \overset{*}{\Rightarrow}_G t$ et $t_2 \overset{*}{\Rightarrow}_G t$. De plus il faut aussi que $\{t' \in T_\Sigma \mid t_1 \overset{*}{\Rightarrow}_G t'\} = \{t\}$. Car si il existait $t' \neq t$ tel que $t_1 \overset{*}{\Rightarrow}_G t'$ alors $\alpha(t', t)$ serait dans F ce qui est exclu. Pour les mêmes raisons, $\{t' \in T_\Sigma \mid t_2 \overset{*}{\Rightarrow}_G t'\} = \{t\}$.

Associons à chaque symbole non-terminal X de V un arbre t_X de $T_\Sigma(X)$ tel que $X(x_1, ..., x_n) \overset{*}{\Rightarrow}_G t_X$, ce qui est possible puisque la grammaire G est totale. On définit l'homomorphisme ϕ de $T_{V \cup \{\alpha\}}$ dans $T_{\Sigma \cup \{\alpha\}}$

$$\phi(\alpha(x, y)) = \alpha(x, y),$$

$$\phi(X(x_1, ..., x_n)) = t_X.$$

On a donc bien, pour tout arbre t de T_V, $t \overset{*}{\Rightarrow}_G \phi(t)$, d'où si $\alpha(t_1, t_2) \in F''$ alors $\alpha(\phi(t_1), \phi(t_2)) \in F'$. D'autre part si $\alpha(t_1, t_2) \overset{*}{\Rightarrow}_G \alpha(t, t)$, alors $t = \phi(t_1) = \phi(t_2)$ puisque $\{t' \mid t_1 \overset{*}{\Rightarrow} t'\} = \{t\} = \{t' \mid t_2 \overset{*}{\Rightarrow} t'\} = \{t\}$ On déduit que $\alpha(t, t) \in F'$ ssi il existe $\alpha(t_1, t_2) \in F''$ tel que $\alpha(t, t) = \alpha(\phi(t_1), \phi(t_2)) = \phi(\alpha(t_1, t_2))$.

Autrement dit F' est l'image par ϕ de la forêt corégulière F''; c'est donc une forêt corégulière d'après la proposition 2.1.

Pour terminer on considère l'homomorphisme ψ de $T_{\Sigma\cup\{\alpha\}}$ dans T_Σ qui vaut l'identité sur T_Σ et tel que $\psi(\alpha(x, y)) = x$. L'image par ψ de $F' = \{\alpha(t, t) \mid t \in F\}$ est la forêt F qui est donc une forêt corégulière. *cqfd.*

5. Applications

PROPOSITION 5.1. *La classe des forêts algébriques n'est pas fermée par homomorphisme non-linéaire.*

Soit Σ un alphabet gradué, F une forêt algébrique de T_Σ qui n'est pas corégulière (voir la proposition 2.3), β un symbole de degré 1 et α un symbole de degré 2 qui ne sont pas dans Σ. La forêt $\{\beta(t) \mid t \in F\} = F'$ est encore algébrique. On considère l'homomorphisme ϕ de $T_{\Sigma\cup\{\beta\}}$ dans $T_{\Sigma\cup\{\alpha\}}$ vaut l'identité sur T_Σ et tel que $\phi(\beta(x)) = \alpha(x, x)$. L'image par ϕ de F' est $\{\alpha(t, t) \mid t \in F\}$. Si cette forêt était algébrique, F serait corégulière ce qui est exclu. *cqfd.*

Ce résultat a été obtenu aussi par Engelfriet et Schmidt [7].

PROPOSITION 5.2. *La classe des transductées déterministes de forêts reconnaissables n'est pas incluse dans celle des algébriques.*

Tout homomorphisme étant un cas particulier de transducteur déterministe, cette proposition se démontre exactement comme la précédente en prenant pour F la forêt reconnaissable non corégulière T_Δ (voir proposition 2.3).

Ce résultat répond à une question soulevé par Downey [5] et montre donc l'incomparabilité vis à vis de l'inclusion de la classe des forêts algébriques et de celle des transductées de forêts reconnaissables.

Maibaum [10] a défini des algébres dérivées qui permettent de considérer tout arbre sur un alphabet gradué comme le feuillage d'un arbre dans cette algébre dérivée. En utilisant les notations de Maibaum nous obtenons:

PROPOSITION 5.3. *Il existe des forêts reconnaissables dont le feuillage n'est pas une forêt algébrique.*

Soit F_0 la forêt reconnaissable engendrée par la grammaire régulière

$$\mathscr{R} = \begin{cases} X \to C_{\langle 2,0\rangle}(\delta, X, X), \\ X \to \#. \end{cases}$$

Le feuillage de F_0 est bien T_Δ. Si à l'ensemble $\mathscr{R}$ on rajoute la règle:

$$X_0 \to C_{\langle 1,0\rangle}(C_{\langle 2,1\rangle}(\alpha, \delta_1{}^1, \delta_1{}^1), X),$$

la forêt reconnaissable F_1 engendrée par ces règles en prenant X_0 comme axiome a pour feuillage $\{\alpha(t, t) \mid t \in T_\Delta\}$ qui n'est pas algébrique. *cqfd.*

Engelfriet et Schmidt [7] ont également exhibé un contre-exemple au "Yield theorem" de Maibaum.

ANNEXE

Soit $G = \langle V, \Sigma, X_0, \mathscr{R}\rangle$ une grammaire sous forme normale. On montre qu'il existe une grammaire $G' = \langle V', \Sigma, X_0, \mathscr{R}'\rangle$ sous forme normale, qui engendre la même forêt, et qui vérifie: $\forall n$, $\forall X$ symbole non terminal de degré n, il existe $t \in T_\Sigma(x_1, ..., x_n)$ tel que

$$\begin{array}{ccc} & X & \\ \swarrow & & \searrow \\ x_1 & & x_n \end{array} \overset{*}{\underset{G}{\Rightarrow}} t.$$

A chaque symbole non-terminal de V on associe l'ensemble

$$F(G, X) = \left\{ t \in T_\Sigma(x_1, ..., x_n) \;\middle|\; \begin{array}{ccc} & X & \\ \swarrow & & \searrow \\ x_1 & & x_n \end{array} \overset{*}{\underset{G}{\Rightarrow}} t \right\}$$

et on pose $I(G) = \{X \in V \mid F(G, X) = \varnothing\}$.

Remarquons que comme on suppose que la forêt engendrée par G n'est pas vide (sinon le résultat n'aurait pas d'intérêt), $X_0 \in I(G)$. D'autre part la propriété $F(G, X) = \varnothing$ étant décidable [12], la construction qui va suivre est effective.

On remarque également que si $Y \in I(G)$ toute règle de la forme

$$\begin{array}{ccc} & X & \\ \swarrow & & \searrow \\ x_1 & & x_n \end{array} \quad \to \quad \begin{array}{ccccccc} & & & Y & & & \\ & & \swarrow & & \searrow & & \\ & Y_1 & & & & Y_m & \\ \swarrow & & \searrow & & \swarrow & & \searrow \\ x_1 & & x_n & & x_1 & & x_n \end{array}$$

ne sera jamais utilisée dans une dérivation d'un arbre de la forêt $F(G, X_0)$ engendrée par G. On peut donc retirer de telles règles et admettre que G vérifie la propriété

$$\text{si} \quad \begin{array}{ccc} & X & \\ \swarrow & & \searrow \\ x_1 & & x_n \end{array} \quad \to \quad \begin{array}{ccccccc} & & & Y & & & \\ & & \swarrow & & \searrow & & \\ & Y_1 & & & & Y_m & \\ \swarrow & & \searrow & & \swarrow & & \searrow \\ x_1 & & x_n & & x_1 & & x_n \end{array} \qquad (*)$$

est une règle de $\mathscr{R}$, et si $X \notin I(G)$, alors $Y \notin I(G)$.

On forme l'alphabet $V' = \{X_\alpha \mid X \in V \text{ et } \alpha \subset \{1,\dots, d(X)\}\}$. Intuitivement α sera l'ensemble des indices des variables "interdites" dans la dérivation de X, c'est-à-dire des variables auxquelles il sera possible de substituer des symboles non terminaux de $I(G)$.

On construit l'ensemble de règles $\mathscr{R}'$ de la façon suivante

1. $X_\alpha(x_1, \dots, x_n) \to Y_{\alpha'}(X_{1\alpha_1}(x_1, \dots, x_n), \dots, X_{m\alpha_m}(x_1, \dots, x_n))$ est dans $\mathscr{R}'$ ssi

   (a) $X(x_1, \dots, x_n) \to Y(X_1(x_1, \dots, x_n), \dots, X_m(x_1, \dots, x_n))$ est dans $\mathscr{R}$,

   (b) si $X_i \in I(G)$ alors $i \in \alpha'$,

   (c) si $i \notin \alpha'$ alors $\alpha_i = \alpha$.

2. $X_\alpha(x_1, \dots, x_n) \to x_i$ est dans $\mathscr{R}'$ ssi,

   $X(x_1, \dots, x_n) \to x_i$ est dans $\mathscr{R}$ et $i \notin \alpha$.

3. $X_\alpha(x_1, \dots, x_n) \to f(x_{i_1}, \dots, x_{i_p})$ est dans $\mathscr{R}'$ ssi,

   $X(x_1, \dots, x_n) \to f(x_{i_1}, \dots, x_{i_p})$ et $\forall j \in \{1,\dots, p\}, \quad i_j \notin \alpha$.

4. $Y_\alpha(x_1, \dots, x_n) \to \#$ (où $\#$ est un nouveau symbole terminal) si $Y \in I(G$

On obtient ainsi une nouvelle grammaire $G' = \langle V', \Sigma \cup \{\bar{X}\}, X_0\Phi, \mathscr{R}' \rangle$ dont l'axiome est $X_{0\varnothing} = X_0$ (puisque la partie vide est la seule partie de l'ensemble $\{1,\dots, d(X_0)\} = \varnothing$).

On forme pour cette grammaire l'ensemble $I(G')$ et on définit $\mathscr{R}_I'$ comme l'ensemble des règles de $\mathscr{R}'$ contenant des symboles de $I(G')$.

Alors nous affirmons que la grammaire $G'' = \langle V' - I(G'), \Sigma \cup \{\bar{X}\}, X_0, \mathscr{R}' - \mathscr{R}_I' \rangle$ est la grammaire cherchée, c'est-à-dire

1. elle est sous forme normale,
2. $I(G'') = \varnothing$,
3. $F(G'', X_0) = F(G, X_0)$.

Le point 1 est vérifié immédiatement par construction de G' et G''. Pour montrer que les deux autres le sont aussi nous avons besoin de deux lemmes.

Lemme A1. Pour tout $X \notin I(G)$, si

$$\begin{array}{ccc} & X & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \overset{*}{\underset{G}{\Rightarrow}} t(x_{i_1}, \dots, x_{i_p}),$$

alors pour tout α inclus dans $\{1, \dots, n\} - \{i_1, \dots, i_p\}$,

$$\begin{array}{ccc} & X_\alpha & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t(x_{i_1}, \dots, x_{i_p}).$$

On raisonne par induction sur la longueur de la dérivation.

1. Dans le cas où la dérivation se réduit à une règle de la forme

$$\begin{array}{ccc} & X & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \rightarrow \begin{array}{ccc} & f & \\ \diagup & & \diagdown \\ x_{i_1} & & x_{i_p} \end{array} \quad \text{ou} \quad \begin{array}{ccc} & X & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \rightarrow x_i ,$$

le résultat est évident par construction de $\mathscr{R}'$.

2. Si

$$\begin{array}{ccc} & X & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \Rightarrow \begin{array}{ccc} & Y & \\ \diagup & & \diagdown \\ \begin{array}{ccc} & X_1 & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} & & \begin{array}{ccc} & X_m & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \end{array} \overset{*}{\underset{G}{\Rightarrow}} t(x_{i_1}, \dots, x_{i_p})$$

alors il existe $t_0(x_{j_1}, \dots, x_{j_q})$, $t_1, \dots, t_q$ tels que

$$t(x_{i_1}, \dots, x_{i_p}) = t_0(t_1, \dots, t_q);$$

$$\begin{array}{ccc} & Y & \\ \diagup & & \diagdown \\ x_1 & & x_m \end{array} \overset{*}{\underset{G}{\Rightarrow}} t_0(x_{j_1}, \dots, x_{j_q});$$

$$\forall_i \in \{1, \dots, q\}, \qquad \begin{array}{ccc} & X_{j_i} & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \overset{*}{\underset{G}{\Rightarrow}} t_i .$$

D'après l'hypothèse de récurrence, $\forall \alpha \subset \{1,\ldots, n\} - \{i_1, \ldots, i_p\}$, $\forall_i \in \{1,\ldots, q\}$,

$$\begin{array}{c} X_{j_i\alpha} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t_i \, .$$

Toujours d'après l'hypothèse de récurrence,

$$\begin{array}{c} Y_{\alpha'} \\ \diagup \quad \diagdown \\ x_1 \qquad x_m \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t_0(x_{j_1}, \ldots, x_{j_q}) \qquad \text{où} \quad \alpha' = \{1,\ldots, m\} - \{j_1, \ldots, j_q\}.$$

De plus, quel que soit $i \in \{1,\ldots, q\}$,

$$\begin{array}{c} X_{j_i} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \overset{*}{\underset{G}{\Rightarrow}} t_i$$

et donc $X_{j_i} \notin i(G)$. D'après la construction de $\mathscr{R}'$ la règle

$$\begin{array}{c} X_{\alpha} \\ \diagup \qquad \diagdown \\ x_1 \qquad\qquad x_n \end{array} \quad \rightarrow \quad \begin{array}{c} Y_{\alpha'} \\ \diagup \qquad\qquad \diagdown \\ \begin{array}{c} X_{1\alpha_1} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \quad \begin{array}{c} X_{m\alpha_m} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \end{array}$$

est dans $\mathscr{R}'$, où $\alpha_i = \alpha$ si $i \notin \alpha'$, d'où

$$\begin{array}{c} X_{\alpha} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t_0(t_1, \ldots, t_q) = t(x_{i_1}, \ldots, x_{i_p}).$$

cqfd.

LEMME A2. *Pour tout $X \notin I(G)$, et tout α tels que*

$$\begin{array}{c} X_{\alpha} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t(x_{i_1}, \ldots, x_{i_p})$$

alors $\alpha \subset \{1,\ldots, n\} - \{i_1, \ldots, i_p\}$ *et*

$$\begin{array}{c} X \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \overset{*}{\underset{G}{\Rightarrow}} t(x_{i_1}, \ldots, x_{i_p}).$$

Comme précédemment on raisonne par induction sur la longueur des dérivations.

1. Si la dérivation se réduit à une règle

$$\begin{array}{ccc} & X_\alpha & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \rightarrow \begin{array}{ccc} & f & \\ \diagup & & \diagdown \\ x_{i_1} & & x_{i_p} \end{array} \quad \text{ou} \quad \begin{array}{ccc} & X_\alpha & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \rightarrow x_i ,$$

le résultat est immédiat par construction.

2. Si

$$\begin{array}{ccc} & X_\alpha & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \underset{G'}{\Rightarrow} \begin{array}{ccc} & Y_{\alpha'} & \\ \diagup & & \diagdown \\ \begin{array}{ccc} & X_{1\alpha_1} & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} & & \begin{array}{ccc} & X_{m\alpha_m} & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t(x_{i_1}, \ldots, x_{i_p})$$

alors il existe $t_0(x_{j_1}, \ldots, x_{j_q})$ et pour chaque $i \in \{1, \ldots, q\}$ un arbre $t_i(x_{k_1}(i), \ldots, x_{k_i}(i))$ tels que

$$\begin{array}{ccc} & Y_{\alpha'} & \\ \diagup & & \diagdown \\ x_1 & & x_m \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t_0(x_{j_1}, \ldots, x_{j_q}),$$

pour chaque $i \in \{1, \ldots, q\}$,

$$\begin{array}{ccc} & X_{j_i\alpha_{j_i}} & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \overset{*}{\underset{G'}{\Rightarrow}} t_i ,$$

$t(x_{i_1}, \ldots, x_{i_p}) = t_0(t_1, \ldots, t_q)$.

De plus la règle

$$\begin{array}{ccc} & X & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \rightarrow \begin{array}{ccc} & Y & \\ \diagup & & \diagdown \\ \begin{array}{ccc} & X_1 & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} & & \begin{array}{ccc} & Y_m & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \end{array}$$

est une règle de G.

D'après la propriété (*), puisque $X \notin I(G)$, $Y \notin I(G)$, et en appliquant l'hypothèse de récurrence, on obtient que

$$\begin{array}{ccc} & Y & \\ \diagup & & \diagdown \\ x_1 & & x_m \end{array} \overset{*}{\underset{G}{\Rightarrow}} t_0(x_{j_1}, \ldots, x_{j_q}) \text{ et } \alpha' \subset \{1, \ldots, m\} - \{j_1, \ldots, j_q\}.$$

D'après la construction de $\mathscr{R}'$, on en déduit que $\forall i \in \{1, \ldots, q\}$, $x_{j_i} \notin I(G)$. On peut donc appliquer à ces X_{j_i} l'hypothèse de récurrence, ce qui donne $\forall i \in \{1, \ldots, q\}$

$$\begin{array}{ccc} & X_{j_i} & \\ \diagup & & \diagdown \\ x_1 & & x_n \end{array} \overset{*}{\underset{G}{\Rightarrow}} t_i \text{ et } \alpha_{j_i} \subset \{1, \ldots, n\} - \{k_1^{(i)}, \ldots, k_i^{(i)}\}.$$

Mais comme $j_i \notin \alpha'$, $\alpha_{j_i} = \alpha$. On a bien

$$\begin{array}{ccc} & X & \\ / & & \backslash \\ x_1 & & x_n \end{array} \quad \overset{*}{\underset{G}{\Rightarrow}} t_0(t_1, \dots, t_q) = t(x_{i_1}, \dots, x_{i_p})$$

et

$$\begin{aligned} \alpha \subset \bigcap_{i=1}^{q} (\{1, \dots, n\} - \{k_1^{(i)}, \dots, k_i^{(i)}\}) &= \{1, \dots, n\} - \bigcup_{i=1}^{q} \{k_1^{(i)}, \dots, k_i^{(i)}\} \\ &= \{1, \dots, n\} - \{i_1, \dots, i_p\}. \end{aligned} \qquad \text{cqfd.}$$

LEMME A3. Pour tout $X_\alpha \in V'$, si

$$\begin{array}{ccc} & X_\alpha & \\ / & & \backslash \\ x_1 & & x_n \end{array} \quad \overset{*}{\underset{G'}{\Rightarrow}} t(x_{i_1}, \dots, x_{i_p})$$

alors

$$\begin{array}{ccc} & X_\alpha & \\ / & & \backslash \\ x_1 & & x_n \end{array} \quad \overset{*}{\underset{G''}{\Rightarrow}} t(x_{i_1}, \dots, x_{i_p}).$$

Si $X \in I(G)$ alors, pour tout α,

$$\begin{array}{ccc} & X_\alpha & \\ / & & \backslash \\ x_1 & & x_n \end{array} \quad \to \#$$

est une règle de G'; donc $X_\alpha \notin I(G')$ et cette règle est encore dans G''.

Si $X \notin I(G)$ alors d'après le lemme A2,

$$\begin{array}{ccc} & X & \\ / & & \backslash \\ x_1 & & x_n \end{array} \quad \overset{*}{\underset{G}{\Rightarrow}} t(x_{i_1}, \dots, x_{i_p}) \qquad \text{et} \qquad \alpha \subset \{1, \dots, n\} - \{i_1, \dots, i_p\}.$$

Par un raisonnement analogue à celui fait dans le lemme A1 on démontre que

$$\begin{array}{ccc} & X & \\ / & & \backslash \\ x_1 & & x_n \end{array} \quad \overset{*}{\underset{G''}{\Rightarrow}} t(x_{i_1}, \dots, x_{i_p})$$

il suffit de faire la remarque suivante: si

$$\begin{array}{ccc} & X & \\ / & & \backslash \\ x_1 & & x_n \end{array} \quad \underset{G}{\Rightarrow} \quad \begin{array}{ccccccc} & & & Y & & & \\ & & / & & \backslash & & \\ & X_1 & & & & X_m & \\ / & & \backslash & & / & & \backslash \\ x_1 & & x_n & & x_1 & & x_n \end{array} \quad \overset{*}{\underset{G}{\Rightarrow}} t_0(X_{i_1}(x_1, \dots, x_n), \dots, X_{i_p}(x_1, \dots, x_n)),$$

alors en posant $\alpha' = \{1,\ldots, m\} - \{i_1,\ldots, i_p\}$, $Y\alpha' \notin I(G')$; de même $X_{i_j}\alpha$ n'appartient pas à $I(G')$ ainsi que X_α. Pour un symbole non terminal X_j autre que les X_{i_j}, soit il est dans $I(G)$ et alors on vient de voir que $X_{j\alpha''}$ n'est pas dans $I(G')$, soit il existe $t_j(x_{j_1},\ldots, x_{j_q})$ tel que

$$\begin{array}{c} X_j \\ \diagup \quad \diagdown \\ x_1 \qquad\quad x_n \end{array} \ \overset{*}{\underset{G}{\Rightarrow}}\ t_j(x_{j_1},\ldots, x_{j_q}).$$

En posant $\alpha_j' = \{1,\ldots, n\} - \{j_1,\ldots, j_q\}$, on en déduit que $X_{j\alpha_j'} \notin I(G')$. D'où tous les symboles de

$$\begin{array}{c} X_\alpha \\ \diagup \qquad \diagdown \\ x_1 \qquad\qquad x_n \end{array} \ \longrightarrow\ \begin{array}{c} Y_{\alpha'} \\ \diagup \qquad\qquad \diagdown \\ \begin{array}{c} X_{1\alpha_1} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \quad \begin{array}{c} X_{m\alpha_m} \\ \diagup \quad \diagdown \\ x_1 \qquad x_n \end{array} \end{array}$$

où

$$\begin{aligned} \alpha_i &= \alpha && \text{si } i \notin \alpha', \\ &= \alpha_i' && \text{sinon et si } X_i \notin I(G), \\ &= \alpha && \text{dans les autres cas,} \end{aligned}$$

sont dans $I(G')$; cette règle est donc dans $\mathscr{R}''$. cqfd.

Des lemmes A1 et A3 on déduit que

$$\begin{array}{c} X \\ \diagup \quad \diagdown \\ x_1 \qquad\quad x_n \end{array} \ \overset{*}{\underset{G}{\Rightarrow}}\ t$$

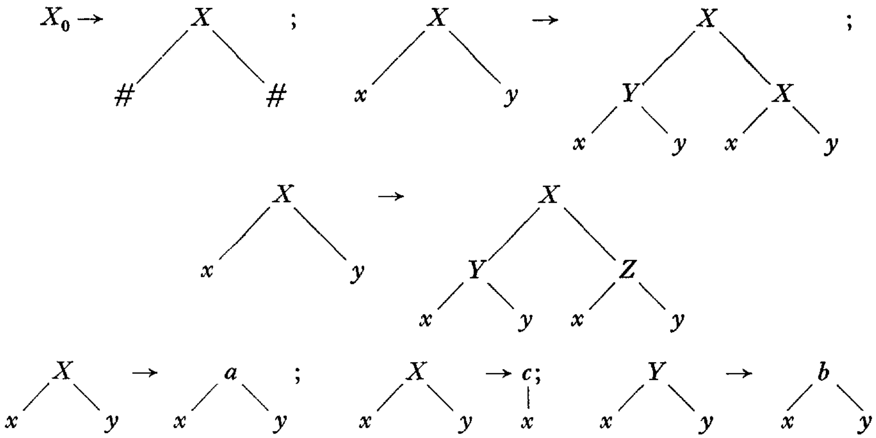
entraîne que

$$\begin{array}{c} X_\alpha \\ \diagup \quad \diagdown \\ x_1 \qquad\quad x_n \end{array} \ \overset{*}{\underset{G''}{\Rightarrow}}\ t.$$

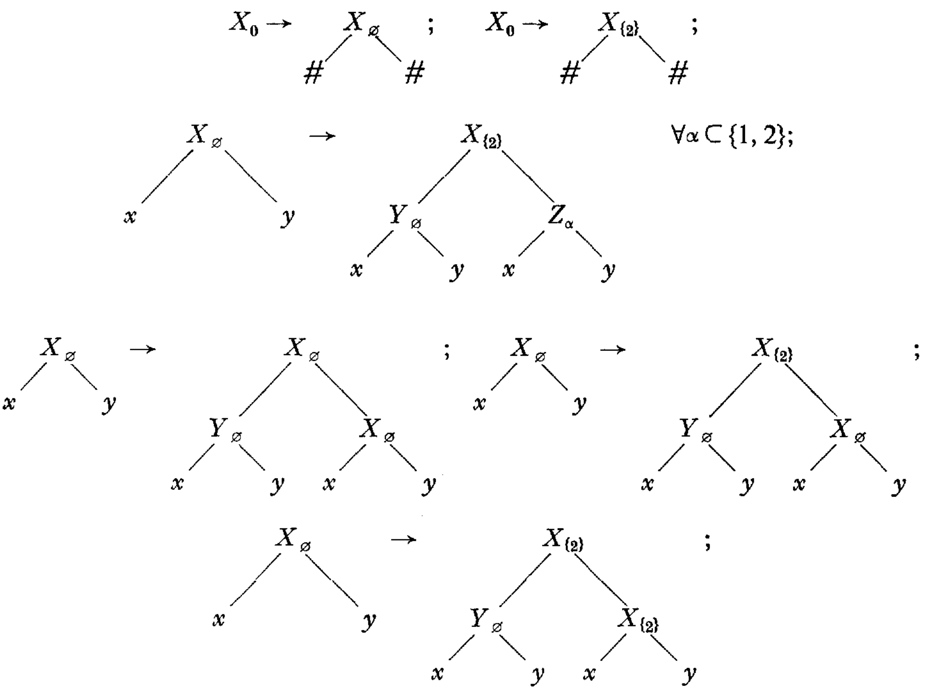
Du lemme A2 et du fait que G'' est une sous-grammaire de G', l'implication inverse est vraie; il s'ensuit immédiatement que $F(G, X_0) = F(G'', X_0)$. Le point 3 est donc vérifié.

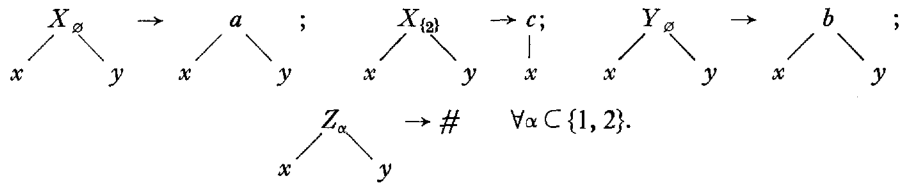
Du lemme A3 on déduit que si $X_\alpha \notin I(G'')$, autrement dit $V' - I(G') \subset V' - I(G'')$. Or $I(G'') \subset V' - I(G') \subset V - I(G'')$ d'où $I(G'') = \varnothing$. Ce qui achève cette démonstration.

EXEMPLE. Soit la grammaire G:

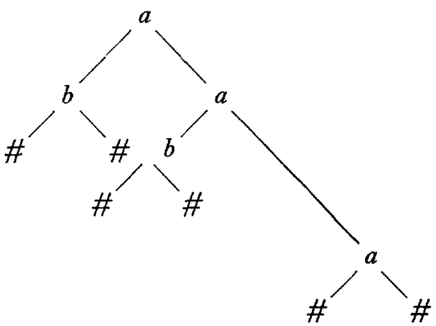


Il est immédiat que $I(G) = \{Z\}$.
La grammaire équivalente G'' fournie par notre construction est:

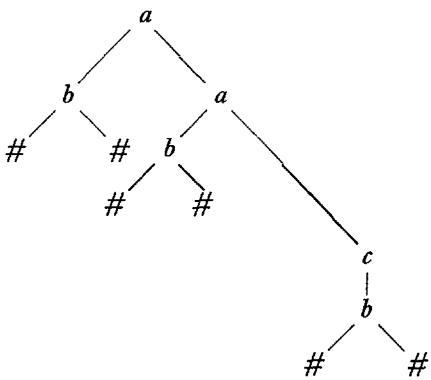




Cette deuxième grammaire ne contient aucun symbole non terminal engendrant une forêt vide et elle engendre bien la même forêt que la première, à savoir l'ensemble des arbres de la forme:



ou



REFERENCES

1. A. V. AHO, Indexed grammars, an extension of context-free grammars, *J. Assoc. Comput. Mach.* **15** (1968), 647–671.
2. A. ARNOLD ET M. DAUCHET, Transductions de forêts reconnaissalbes monadiques. Forêts corégulières. À paraître dans *Rev. Française Automat. Informat. Rech. Opérat*, série rouge.
3. B. S. BAKER, Tree transductions and families of tree langages, *in* 5th ACM Symp. on Theory of Computing, 1973, pp. 200–206.
4. P. DOWNEY, Formal languages and recursion schems, Ph. D. Dissertation, Harvard University, 1974.
5. P. DOWNEY, Tree transducers and ETOL tree systems, Conference on formal languages, automata and development. Noordwykerhout, The Netherlands, 1975.
6. J. ENGELFRIET, Surface tree languages and parallel derivation trees. Daimi-Report PB. 44, University of Aarhus, Danemark, 1975.
7. J. ENGELFRIET ET E. M. SCHMIDT, IO and OI Daimi-Report PB. 47, University of Aarhus, Danemark, 1975.
8. J. ENGELFRIET ET S. SKYUM, Copying theorems. Daimi-Report PB. 48, University of Aarhus, Danemark, 1975.
9. M. J. FISCHER, Grammars with macro-like productions. Doctoral Dissertation, Harvard University, 1968.

10. T. S. E. MAIBAUM, A generalized approach to formal languages, *J. Comput. System Sci.* **8** (1974), 409–439.
11. W. C. ROUNDS, Mapping and grammars on trees. *Math. Systems Theory* **4** (1970), 257–287.
12. W. C. ROUNDS, Tree oriented proofs of some theorems on context-free and indexed languages, *in* "2nd ACM Symp. on Theory of Computing (1970)," 109–116.
13. G. ROZENBERG, Extensions of tabled OL-systems and languages, *Internat. J. Comput. Information Sci.* **2** (1973), 311–336.
14. S. SKYUM, Decomposition theorems for various kinds of languages parallel in nature, à paraître dans *SIAM J. Computing.*
15. J. W. THATCHER, Generalized² sequential machines, *J. Comput. System Sci.* **4** (1970), 339–367.

Printed in Belgium