MATHÉMATIQUES ET SCIENCES HUMAINES

M. EYTAN

Une application élémentaire des grammaires génératives au dénombrement d'une certaine classe d'arbres

Mathématiques et sciences humaines, tome 21 (1968), p. 11-16

<http://www.numdam.org/item?id=MSH_1968__21__11_0>

UNE APPLICATION ÉLÉMENTAIRE DES GRAMMAIRES GÉNÉRATIVES AU DÉNOMBREMENT D'UNE CERTAINE CLASSE D'ARBRES

par

M. EYTAN [1]

1. — LE PROBLÈME.

Il s'agit de trouver le nombre A $(p; n)$ d'arbres biordonnés p — aires à $pn + 1$ nœuds, i.e. à n nœuds non-terminaux.

2. — DÉFINITIONS.

Par arbre biordonné [2] (appelé désormais arbre tout court) nous entendons le triplet (A, £, \$) d'un ensemble fini A, muni des deux ordres £ et \$ (dits resp. « hiérarchique » et « séquentiel ») vérifiant les axiomes suivants :

1) le £-prédécesseur immédiat de tout élément de A, s'il existe, est unique ;
2) \$ est un ordre total sur A ;
3) pour tout couple (x, y) d'éléments de A, x£y entraîne x\$y ;
4) quels que soient les éléments x, y, x′, y′ de A, x\$y et x£x′ et y£y′ entraînent x′\$y′ ;
5) il existe un élément unique de A sans £-prédécesseur.

L'élément déterminé par l'axiome 5) sera dit racine de l'arbre A (comme d'habitude, on fait l'abus de langage assimilant le triplet (A, £, \$) à l'ensemble sous-jacent A).

Un nœud terminal de A sera un élément sans £-successeur (i.e. un élément £-maximal). Un nœud qui n'est pas terminal sera dit non-terminal.

Un arbre sera dit p-aire si tout nœud non-terminal a p £-successeurs.

On remarquera qu'un arbre p-aire à n nœuds non-terminaux a en tout $pn + 1$ nœuds : car il y a autant de nœuds au niveau $q + 1$ (i.e. ensemble de nœuds qui sont reliés à la racine par une chaîne de $q + 1$ éléments dont chacun est £-successeur immédiat du précédent) que de branches partant du niveau q i.e. p fois le nombre de nœuds de ce niveau. En ajoutant la racine et remarquant que de tout nœud partent p branches, on obtient $pn + 1$.

Le lecteur connaît la représentation graphique d'un arbre. Nous utiliserons sans vergogne le schématisme et le langage correspondants. Toutefois, nous en introduisons une autre pour préciser complètement l'arbre par un mot d'un certain monoïde.

---

1. Centre de calcul de la Maison des sciences de l'homme. L'utilisation des grammaires m'a été suggéré par B. Jaulin, cf. aussi M. Gross, *Applications géométriques des langages formels* (C.N.R.S., Institut Blaise Pascal).
2. Cf. D. Guedj, Grammaires de constituants généraux (thèse de 3e cycle, ronéotypé).

3. — REPRÉSENTATION POLONAISE D'UN ARBRE A.

Soit $(s_1, s_2, \ldots s_k)$ la suite de tous les nœuds de A, rangés dans leur $-ordre (ceci est possible $ étant un ordre total).

Associons à chacun des s le symbole f si s est non-terminal, le symbole t s'il est terminal.

La suite qu'on en déduit, ou plutôt le résultat de la concaténation de ces éléments, est un mot du monoïde libre sur l'alphabet réduit aux deux seuls symboles f et t. C'est ce mot que nous appelons la représentation polonaise de A.

Exemple : Soit A l'arbre 4-aire dont la représentation graphique est :

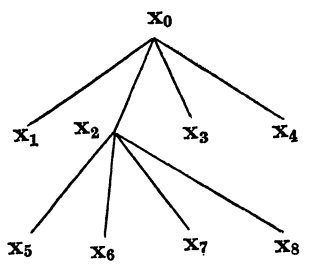

Fig. 1

La $-suite des nœuds de A est

$$(x_0, x_1, x_2, x_5, x_6, x_7, x_8, x_3, x_4)$$

La représentation polonaise est :

ftfttttt

On remarque que f est un symbole de poids 4 [1] (i.e. un symbole de fonction à 4 variables) et que la « représentation polonaise de A » est l'écriture de f (t, f (t,t,t,t,), t, t) en notation polonaise, d'où le nom.

4. — GRAMMAIRES GÉNÉRATIVES [2] (ou grammaires de constituants).

Rappelons qu'une grammaire générative est le quadruplet

$$G = (V, V_T, S, R) \text{ où}$$

1) V est un ensemble fini de symboles ;
2) V_T est un sous-ensemble de V ;
3) S est un élément distingué de $V - V_T$;
4) R est une partie finie du produit cartésien $(V - V_T)^* \times V^*$ (où X^* signifie monoïde libre engendré par X).

Le langage engendré par G, noté L(G), est l'ensemble des mots $x \in V_T^*$ obtenus par dérivation à partir du symbole S, où la dérivation est la fermeture transitive de la relation D définie par :

x D y ssi

il existe $r = (a, b) \in R$ et $g, d \in V^*$ tels que $x = gad$ et $y = gbd$.

5. — GRAMMAIRE GÉNÉRATIVE POUR LES ARBRES *p*-aires.

Voici une grammaire générative dont le langage engendré comprend les représentants polonais des arbres *p*-aires (et eux seuls comme le lecteur s'en convaincra) :

---

1. Cf. Bourbaki, livre I, *Théorie des ensembles*, chap. I, Appendice p. 51.
2. Cf. Chomsky, *Formal properties of grammars in Handbook of mathematical Psychology* dans la traduction française : N. Chomsky, *L'analyse formelle des langues naturelles*, Mouton, Paris, 1968 ; ou J. Friant « Les langages-Context-sensitive » in *Ann. Inst. H. Poincaré*, vol. III, n° 1, 1967 (p. 35-120).

$$G(p) = \left(V, V_T, S, R\,(p)\right) \text{ où}$$

- $V_T = \{f, t\}$
- $V - V_T = \{S\}$
- $R(p) = \{(S, fS^p), (S, t)\}$

où fS^p est fS ... S, p des symboles S suivant f.

La signification intuitive des éléments de $R(p)$ est la suivante : un arbre p-aire s'obtient en «suspendant» à un nœud quelconque un nœud (qui devient ainsi non terminal, donc étiqueté f) d'où «pendent» p branches qui peuvent devenir des nœuds à leur tour (élément (S, fS^p) de $R(p)$), ou bien en rendant ce nœud terminal (élément (S, t) de $R(p)$).

6. — SÉRIE FORMELLE ASSOCIÉE A LA GRAMMAIRE $G(p)$ [1].

Nous ne pouvons reprendre ici la théorie associant à toute grammaire une série formelle *non commutative* dont le support donne le langage L ($G(p)$). Contentons-nous de dire que la série formelle S′ associée à $G(p)$ doit vérifier

$$S' = fS'^p + t \tag{1}$$

et qu'elle s'obtient par approximations successives en écrivant :

$$\begin{cases} S'_0 = 0 \\ S'_{k+1} = fS'^p_k + t \qquad k \geqslant 0 \end{cases}$$
$$\text{avec } S' = \lim_k S'_k$$

En faisant $p = 2$ le lecteur retrouvera en calculant les trois-quatre premiers S_k la représentation polonaise de quelques $A(2 ; n)$. Nous l'engageons à le faire à titre d'exercice en comparant ce résultat avec la représentation graphique.

7. — FONCTION GÉNÉRATRICE POUR LES $A(p; n)$.

Dans la série formelle S′ (non-commutative), on retrouve, sous forme polonaise, tous les arbres p-aires. Parmi eux, en particulier, se trouvent ceux à n nœuds non-terminaux ; ils sont caractérisés par le fait qu'y figure n fois le symbole f. Etant donné que la représentation polonaise des arbres est fidèle (i.e. bijective) et que dans la série formelle S′ chaque arbre p-aire est représentée sous forme polonaise une fois et une seule (comme on s'en convraincra aisément), pour obtenir le nombre $A(p; n)$ d'arbres p-aires à n nœuds non-terminaux il suffirait de compter les termes de la série formelle S′ où figure n fois le symbole f.

Pour cela nous allons envoyer l'anneau des séries formelles non-commutatives en deux indéterminées f et t sur l'anneau des entiers rationnels Z [[f,t]], dans l'anneau des séries formelles commutatives en une indéterminée u sur l'anneau des entiers rationnels, Z [[u]]. Pour cela, il suffit on le sait, de se donner les images des indéterminées f et t. Posons donc F(f) = u, F(t) = u.

---

1. Cf. Chomsky-Schützenberger, *The algebraic theory of context-free languages in Computer Programming and Formal Systems* (North-Holland, 1963).

L'application F s'étend de façon canonique en un homomorphisme d'anneaux F :

$$\mathbf{Z}\,[[f,t]] \to \mathbf{Z}\,[[u]].$$

Si l'on considère l'image F (S'), il est clair que chacun des termes de S' contenant n fois f (et par suite $pn + 1 - n$ fois t) a pour image le terme u^{pn+1} de F(S') ; et ce sont les seuls. Le nombre de termes de S' contenant n fois f, c.a.d.A $(p ; n)$, est donc égal au coefficient de u^{pn+1} dans la série formelle F(S'). Autrement dit la série formelle F(S') est la fonction génératrice des A$(p ; n)$.

Reste à déterminer effectivement F(S'). Or F est un homomorphisme d'anneaux. L'image par F de l'équation (1) donne donc l'équation :

$$F(S') = uF(S')^p + u \tag{2}$$

Posons $F(S') = \sum_{i \in \mathbf{N}} a_i\, u^i$. L'équation (2) s'écrit :

$$\sum_{i \in \mathbf{N}} a_i\, u^i = u \sum_{i \in \mathbf{N}} \left(\sum_{j_1 + j_2 \ldots + j_p = i} a_{j_1} a_{j_2} \ldots a_{j_p} \right) u^i + u$$

ce qui donne entre les a_i les relations

$$\begin{cases} a_0 = 0, \qquad a_1 = 1 \\ a_{i+1} = \sum_{j_1 + j_2 \ldots + j_p = i} a_{j_1} a_{j_2} \ldots a_{j_p} \qquad i > 1 \end{cases} \tag{3}$$

On montre facilement (par récurrence) que pour tout $i \neq pn$, $a_{i+1} = 0$.

Remarquant alors que $A\,(p ; n) = a_{pn+1}$

on obtient les relations :

$$A\,(p ; 0) = 1 \tag{4}$$

$$A\,(p ; n) = \sum_{j_1 + j_2 + \ldots + j_p = n-1} A\,(p ; j_1)\, A\,(p ; j_2) \ldots A\,(p ; j_p) \qquad n \geqslant 1$$

L'interprétation combinatoire de la relation (4) est facile : pour obtenir un arbre p-aire à n nœuds non-terminaux, il suffit de (et il faut) prendre $n - 1$ arbres p-aires à $j_1, j_2 \ldots j_p$ nœuds non-terminaux et de les réunir par leurs racines à un nœud qui est la racine du nouvel arbre. Pour que ce dernier ait n nœuds non-terminaux il faut (et il suffit) que $j_1 + j_2 + \ldots + j_p = n - 1$. En faisant décrire aux j_k tous les partages de l'entier $n - 1$ en p parts, on obtient une fois et une seule tous les arbres p-aires à n nœuds non-terminaux.

8. — CALCUL EXPLICITE DES A $(p ; n)$.

Les relations (4) permettent le calcul des A $(p ; n)$ de proche en proche. Nous allons cependant exhiber une formule close explicite qui donne les A $(p ; n)$.

1°) Une première façon de procéder et d'utiliser la formule de Lagrange [1]. Elle donne ici :

$$A\,(p ; n) = \left[\frac{d^{n-1}}{dy^{n-1}} (y^p + 1)^n \right]_{y = 0} \tag{5}$$

1. Goursat, *Cours d'analyse* (Gauthier-Villars, 1926) tome I, p. 471.

2°) Une autre procédure consiste à introduire une nouvelle représentation monoïdale de l'arbre en associant à chaque nœud sauf la racine, un des symboles $l_1, l_2, \ldots l_p$ par récurrence :

a) aux nœuds du niveau h, ($h > 0$) rangés dans leur \$-ordre, on associe les symboles $l_1, l_2, \ldots l_p$ dans l'ordre des indices.

b) les £-successeurs d'un nœud déjà « étiqueté» sont rangés dans leur \$-ordre et se voient attribuer les symboles $l_1, \ldots, l_p$ dans l'ordre des indices.

Puis on forme la suite $(s_1, \ldots, s_{pn})$ de tous les nœuds (sauf la racine) de l'arbre ; on les remplace par les l_i correspondants, et par concaténation on déduit un mot du monoïde libre construit sur l'alphabet $\{l_1, \ldots, l_p\}$.

Les mots ainsi construits (i.e. les représentations correspondantes d'arbres p-aires) ont une propriété remarquable (qui se démontre aisément par récurrence sur la construction) :

(LP) tout segment initial d'un tel mot a une image commutative

$$l_1^{i_1}\, l_2^{i_2} \ldots l_p^{i_p} \text{ vérifiant } i_1 \geqslant i_2 \geqslant \ldots \geqslant i_p.$$

On retrouve les « lattice permutations» de McMahon [1], ou plutôt le cas particulier des « lattice permutations» dont tous les indices sont identiques et égaux à n. La formule donnée par McMahon permet d'écrire immédiatement :

$$A(p; n) = \frac{(pn)!}{(n+p-1)!\,(n+p-2)!\ldots n!} \prod_{1 \leqslant j < k \leqslant p} (k-j) \qquad (6)$$

$$= \frac{(pn)!\,(p-1)!!}{(n+p-1)!\,(n+p-2)!\ldots n!} \quad (\text{où } p!! = p!\,(p-1)!\ldots 1!)$$

Remarquons à ce propos que les résultats précédents i.e. les relations (4) permettent d'écrire une formule de récurrence pour les « lattice permutations» connue de McMahon seulement dans le cas $p = 2$ et qui dans sa notation s'écrit :

$$\underbrace{(nn \ldots n;)}_{p \text{ fois}} = \sum_{j_1 + j_2 \ldots + j_p = n-1} \underbrace{(j_1 \ldots j_1;)}_{p \text{ fois}} \ldots \underbrace{(j_p \ldots j_p;)}_{p \text{ fois}}$$

D'autre part si pour le cas $p = 2$ McMahon obtient facilement la relation que vérifie la fonction génératrice associée à $(nn;)$, pour $p = 3$ déjà il n'obtient qu'une fonction génératrice qu'il appelle redondante (tous les $(nnn;)$ se trouvent parmi ses coefficients, mais certains des coefficients ne sont pas des $(nnn;)$). Nous connaissons cette relation, elle est donnée par la formule (2).

9. — UNE GÉNÉRALISATION.

Dans un texte intitulé [2] « Unicité du foncteur degré », J. P. Benzecri pose un problème très général qui dans la catégorie des ensembles et applications revient au suivant :

Etant donné une règle r $(n_1, \ldots, n_p ; m)$ qui fabrique une application f à m variables à partir de p applications, la première f_1 de $n_1, \ldots$, la p-ème f_p de n_p variables chacune utilisée un nombre quelconque de fois, calculer le nombre N $(n_1, \ldots, n_p ; m ; q_1, \ldots, q_p)$ d'applications fabriquées selon la règle r $(n_1, \ldots, n_p ; m)$, la j-ème fonction f_j étant utilisée q_j fois.

---

1. Mc Mahon, *Combinatory analysis* (Chelsea, New York, 1960) chap. V. Elles interviennent à propos du problème de vote cohérent (désigné par G. Kreweras dans sa thèse sous le nom de « problème de Bertrand») d'où l'ordre final des candidats a été respecté tout au long du scrutin.

2. Ronéoté, non publié.

On remarque que se donner la règle r devient, en supprimant parenthèses et virgules (i.e. en utilisant la notation polonaise) la représentation polonaise d'un arbre. Le nombre à calculer n'est donc rien d'autre que le nombre d'arbres correspondant.

Tout d'abord par un raisonnement qui est l'analogue de celui de la fin du § 1, on voit que le nombre total de nœuds est $\sum_j q_j n_j + 1$ (le nœud étiqueté f_j est d'ordre q_j). Le nombre total de nœuds terminaux est $\sum_j q_j$. Donc le nombre de nœuds non-terminaux, i.e. m est $\sum_j q_j n_j + 1 - \sum_j q_j$ soit

$$m = \Sigma (q_j - 1) n_j + 1$$

La donnée de m est donc superflue.

La grammaire générative engendrant les arbres r donnés est :

$$G = (V, V_T, S, R) \text{ où}$$

- $V_T = \{f_1, \ldots, f_p, t\}$
- $V - V_T = \{S\}$
- $R = \{(S, f_1 S^{n_1}), \ldots, (S, f_p S^{n_p}), (S,t)\}$

La série formelle non-commutative associée à G est solution de l'équation

$$S' = f_1 S'^{n_1} + \ldots + f_p S'^{n_p} + t$$

et $N(n_1, \ldots, n_p ; m ; q_1, \ldots, q_p)$ est le coefficient du monôme $f_1^{q_1} \ldots f_p^{q_p}$ dans la série formelle commutative F(S') déduite de S' par l'homomorphisme engendré par F qui ne change pas $f_1, \ldots, f_p$, t.

Nous ne connaissons pas de formule permettant d'expliciter ce coefficient.