RAIRO. INFORMATIQUE THÉORIQUE

LAURENT CHOTTIN
ROBERT CORI

Une preuve combinatoire de la rationalité d'une série génératrice associée aux arbres

RAIRO. Informatique théorique, tome 16, nº 2 (1982), p. 113-128
<http://www.numdam.org/item?id=ITA_1982__16_2_113_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 16, n° 2, 1982, p. 113 à 128)

UNE PREUVE COMBINATOIRE DE LA RATIONALITÉ D'UNE SÉRIE GÉNÉRATRICE ASSOCIÉE AUX ARBRES (*)

par Laurent CHOTTIN ([1]) et Robert CORI ([1])

Communiqué par la Rédaction

Résumé. — *Dans cet article, nous démontrons, de manière combinatoire qu'une famille de mots dérivant des systèmes de parenthèses est énumérée par* 4^n.

A cette occasion, nous présentons une factorisation particulière d'un mot écrit sur un alphabet à deux lettres que nous appelons factorisation de Catalan. Cette factorisation est un outil utile dans d'autres circonstances.

Abstract. — *We prove in a combinatorial way, using bijections, that a family of words related to parenthesis systems is enumerated by* 4^n.

For that aim, we introduce the concept of "Catalan factorisation" of words written on an alphabet of two letters; this factorisation is a general tool for other combinatorial problems.

L'énumération des arbres, des systèmes de parenthèses pointés ou de certaines familles de chemins du plan, figure en bonne place dans la plupart des traités de combinatoire [1], [9], [10]. Cette énumération fait intervenir les nombres de Catalan, et de nombreuses formules peuvent être obtenues à partir de ces nombres. Ceci est vraisemblablement du au fait que leur série génératrice vérifie une équation particulièrement simple ($xf^2-f+1=0$) et se trouve être, par conséquent, algébrique.

Dans cet article, nous nous intéressons à un comptage qui fait intervenir une série génératrice rationnelle, ainsi nous donnons une nouvelle preuve (par construction d'une bijection) du fait qu'une certaine famille généralisant les systèmes de parenthèses est comptée par 4^n. Cette situation est d'autant plus étonnante que les objets énumérés sont *a priori* bien plus complexes que les systèmes de parenthèses et que leur série n'a même aucune raison d'être algébrique. La première preuve de ce résultat a été obtenue par G. Kreweras [8], mais c'est semble-t-il, dans l'ouvrage de D. Knuth [6] qu'intervient pour la première fois une formule de ce type.

(*) Reçu juillet 1981.
([1]) Université de Bordeaux-I.

R.A.I.R.O. Informatique théorique/Theoretical Informatics, 0399-0540/1982/ 113/$ 5.00

Elle apparaît dans l'évaluation des performances d'algorithmes liés aux arbres binaires. Ceci est loin d'être fortuit, les systèmes de parenthèses, arbres et chemins constituent des structures de données fondamentales et ceux qui souhaitent établir des résultats sur les algorithmes qui les font intervenir sont naturellement amenés à considérer des formules où apparaissent les nombres de Catalan. L'ouvrage de D. Knuth et plus récemment, les travaux de P. Flajolet, J. Françon, G. Viennot et J. Vuillemin [3, 4 et 5] sur ces questions mettent en lumière le lien entre cette problématique de nature informatique et la combinatoire énumérative.

La preuve développée dans notre article repose sur la construction d'une bijection, elle utilise une propriété de factorisation des mots du monoïde libre $\{x, \bar{x}\}^*$ (ici appelée factorisation de Catalan). Cette propriété pourrait être déduite de la construction générale des bissections [10, 12]; nous avons préféré en donner une démonstration élémentaire. La même propriété est utilisée à d'autres fins dans [2] et [7].

L'article commence donc par l'étude en détail des factorisations de Catalan dont nous établissons l'existence et l'unicité. Une fois cet outil mis en place, nous nous intéressons dans la partie 2 à deux transformations G et D sur les mots de $\{x, \bar{x}\}^*$, transformations qui permettent la construction de la bijection B. La partie 3 présente une interprétation géométrique de la bijection B. Pour terminer, nous nous proposons de traduire ce résultat en termes d'arbres et d'arbres binaires, et retrouver ainsi le résultat original de Knuth [6].

FACTORISATIONS DE CATALAN

1.1. Systèmes de parenthèses

Soit X l'alphabet composé des deux lettres x et $\bar{x}$. Pour tout mot f du monoïde libre X^*, $|f|$ désigne la longueur du mot f, $|f|_x$ le nombre de lettres x qui figurent dans f et $|f|_{\bar{x}}$ le nombre de $\bar{x}$ figurant dans f. Ainsi : $|f|=|f|_x+|f|_{\bar{x}}$.

On considère l'application δ qui à un mot f fait correspondre l'entier $\delta(f)=|f|_x-|f|_{\bar{x}}$.

De cette définition, il découle que le mot vide $\mathbf{1}$ vérifie $\delta(\mathbf{1})=0$ et que $\delta(fg)=\delta(f)+\delta(g)$, ainsi δ est un morphisme de X^* dans $\mathbb{Z}$ considéré comme un monoïde pour l'addition.

DÉFINITION : *Un mot f est un système de parenthèses s'il vérifie* $\delta(f)=0$ *et si* $f=f'f''$ *implique* $\delta(f')\geq 0$.

L'ensemble de tous les systèmes de parenthèses est noté P dans la suite; les propriétés suivantes de P sont classiques et simples à démontrer. Nous les donnons sans preuve :

PROPRIÉTÉS 1.1 :

(i) Le mot vide et $x\bar{x}$ sont des éléments d P.

(ii) Si f et g sont deux mots de P, fg est un mot de P.

(iii) Sous les mêmes hypothèses que (ii) $xf\bar{x}g$ est un mot de P.

(iv) Si f et fg sont dans P, g est dans P.

(v) Tout mot h de P est soit vide soit se décompose sous la forme $h=xf\bar{x}g$ avec f et g dans P.

1.2. Définition et premières propriétés des factorisations de Catalan

Une factorisation (F) d'un mot f est une suite $f_1, f_2, \ldots, f_p$ de mots non vides dont le produit est égal à f. L'entier p est la longueur de la factorisation (F).

DÉFINITION : *Une factorisation (F) d'un mot f de X^* est une factorisation de Catalan si elle vérifie les deux conditions suivantes :*

1) *tout mot f_i de la factorisation est ou bien un élément de* X *ou bien un élément de P (différent de* **1**)*;*

2) *toute factorisation (G) du mot f vérifiant* (1) *est de longueur plus grande que (F).*

Exemple : Considérons le mot $f=\bar{x}x\bar{x}\bar{x}xx\bar{x}$; on vérifie facilement que $(\bar{x}, x\bar{x}, \bar{x}, x, x\bar{x})$ est une factorisation de Catalan de f.

Remarquons que tout mot admet au moins une factorisation de Catalan : en effet, si l'on choisit chaque f_i égal à une lettre de X, on obtient une factorisation vérifiant (1); la finitude du nombre de factorisations assure ensuite l'existence d'une factorisation de longueur minimale vérifiant (1).

Le but de la suite du paragraphe est de démontrer l'unicité de la factorisation de Catalan pour un mot f donné et de proposer une caractérisation de celle-ci.

Mais auparavant, nous établissons quelques propriétés simples qui seront utiles dans la suite.

LEMME 1.1 : *Si $F=(f_1, f_2, \ldots, f_p)$ est une factorisation de Catalan de f, alors f_i et f_{i+1} ne peuvent être à la fois dans P.*

En effet, $f_i f_{i+1}$ serait aussi dans P et :

$$G=(f_1, f_2, \ldots, f_{i-1}, f_i f_{i+1}, f_{i+2}, \ldots, f_p)$$

serait une factorisation vérifiant (1) et de longueur plus courte que (F).

LEMME 1.2 : *Si* $F=(f_1,f_2,\ldots,f_p)$ *est une factorisation de Catalan de* f, *alors* $f_i=x$ *implique que* $f_{i+1}\neq\bar{x}$.

En effet la factorisation (G) donnée plus haut de P vérifierait (1) aussi puisque $x\bar{x}$ est un élément de P.

LEMME 1.3 : *Si* $F=(f_1,f_2,\ldots,f_p)$ *est une factorisation de Catalan de* f, *alors* $f_i=x$ *et* $f_{i+1}\in P$ *implique* $f_{i+2}\neq\bar{x}$.

En effet, dans ce cas la factorisation de f :

$$H=(f_1,\ldots,f_{i-1},xf_{i+1}\bar{x},f_{i+3},\ldots,f_p),$$

vérifierait (1) car $f_{i+1}\in P$, $1\in P$ implique $xf_{i+1}\bar{x}\in P$ d'après la 3e partie de la propriété 1.1. Or, la longueur de (H) est plus petite que celle de (F) d'où la contradiction.

LEMME 1.4 : *Si* $F=(f_1,f_2,\ldots,f_p)$ *est une factorisation de Catalan de* f, *si* $f_i=x$ *et si* $f_j=\bar{x}$ *alors* i *est plus grand que* j.

Supposons qu'il n'en soit pas ainsi et que i soit plus petit que j. Soit alors i_0 le plus grand entier plus petit que j tel que $f_{i_0}=x$ (on a ainsi $i_0\geqq i$); et soit j_0 le plus petit entier plus grand que i_0 tel que $f_{j_0}=\bar{x}$ (on a ainsi $i\leqq i_0\leqq j_0\leqq j$). De par le choix de i_0 et j_0, f_k est nécessairement un élément de P pour tout k compris strictement entre i_0 et j_0; mais d'après le lemme 1.1, ceci implique que j_0 est égal à i_{0+1} ou à i_{0+2}. Le premier cas est exclus par le lemme 1.2 et le second par le lemme 1.3. On obtient ainsi la contradiction cherchée.

1.3. Unicité de la factorisation

Nous sommes maintenant en mesure de démontrer le théorème 1.1.

THÉORÈME 1.1 : *Tout mot* f *de* $\{x,\bar{x}\}^*$ *admet une factorisation de Catalan unique.*

Supposons qu'il existe un mot admettant deux factorisations de Catalan et soit f le mot le plus court qui soit deux fois décomposable. Ainsi, $f=f_1f_2\ldots f_p=g_1g_2\ldots g_p$, la minimalité de f assure alors $f_1\neq g_1$.

On a $g_1\neq\bar{x}$, car f_1 ne pourrait être alors ni un x ni élément de P (aucun ne commence par $\bar{x}$). Ainsi, g_1 est ou bien x ou bien un élément de P.

Si $g_1=x$, aucun g_i n'est un $\bar{x}$ d'après le lemme 1.4, ainsi $f=xg_2\ldots g_p$ implique que $\delta(g_2\ldots g_p)$ est positif ou nul et que $\delta(g')$ est aussi positif ou nul pour tout mot g' tel que $g_2\ldots g_p=g'g''$ dans ce cas, f_1 ne peut être un mot de P sinon $f_1=xf'_1$ et $g_2\ldots g_p=f'_1f_2\ldots f_p$ aurait comme facteur gauche f'_1 dont l'image par δ est $\delta(f_1)-\delta(x)=-1$. On obtient ainsi $f_1=x=g_1$, qui est la contradiction cherchée.

Reste à étudier le cas où g_1 est un élément de P;f_1 est aussi dans P (sinon on se retrouverait dans une situation étudiée plus haut).

Supposons g_1 de longueur plus petite que f_1, on écrit alors $f_1=g_1f'_1$. Par le lemme 1.1, g_2 est nécessairement soit un $\bar{x}$ soit un x; ce ne peut être un $\bar{x}$ car alors $g_1\bar{x}$ serait un facteur gauche de f_1 d'image par δ négative. Ce ne peut être non plus un x, car alors tout facteur gauche de f aurait une image par δ positive et il en serait de même pour $g_3\ldots g_p$, ceci est incompatible avec $f_1=g_1xf''_1$;f_1, $g_1\in$ P et f''_1 facteur gauche de $g_3\ldots g_p$. Cette dernière contradiction termine la preuve du théorème.

PROPRIÉTÉ 1.2 : *Caractérisation des factorisations de Catalan. Soit* $(F)=(f_1,f_2,\ldots,f_p)$ *une factorisation du mot f qui vérifie la condition* 1 *et les conditions* (3) *et* (4) *suivantes :*

3) f_i *et* f_{i+1} *ne sont jamais tous les deux dans P;*

4) $f_i=x$ *et* $f_j=\bar{x}$ *implique* $i>j$;

(F) *est alors la factorisation de Catalan de f.*

La preuve de cette propriété très semblable à celle du théorème précédent est laissée au lecteur.

1.4. Factorisation de Catalan des facteurs gauches et droits des mots de P

A tout mot f de X^* on associe le couple d'entiers $\chi(f)=(a_f, b_f)$ tels que si $F=(f_1,f_2,\ldots,f_p)$ est la factorisation de Catalan de F :

$$a_f=\text{Card}\{i\,|\,f_i=\bar{x}\}\qquad \text{et}\qquad b_f=\text{Card}\{j\,|\,f_j=x\}.$$

Ainsi $\delta(f)=\sum_{i=1}^{p}\delta(f_i)=b_f-a_f$ [car $\delta(f_i)=0$ si $f_i\in$ P].

Avec cette notation, on a la propriété suivante :

PROPRIÉTÉ 1.3 : *Étant donné un mot f, les trois conditions suivantes sont équivalentes :*

(i) $a_f=0$.

(ii) *Pour tout* f' *tel que* $f=f'f''$ *on a* $\delta(f')\geqq 0$.

(iii) *Il existe g tel que fg appartienne à P.*

Preuve :

(i)$\Rightarrow$(ii) En effet, soit $f=f_1f_2\ldots f_p$ la factorisation de Catalan de f qui ne comporte alors aucun f_i égal à $\bar{x}$; si f est alors égal à $f'f''$ on a $f'=f_1\ldots f_{k-1}f'_k$; le fait que f_i soit ou bien un x ou bien un élément de P implique $\delta(f_i)\geqq 0$ et $\delta(f'_k)\geqq 0$ et (ii) découle de $\delta(f')=\sum_{i=1}^{k-1}\delta(f_i)+\delta(f'_k)$.

(ii) $\Rightarrow$ (iii) Considérons le mot $g=\bar{x}^{\delta(f)}$, il est aisé de montrer que fg est bien un système de parenthèses.

(iii) $\Rightarrow$ (i) Si f contient un f_i égal à $\bar{x}$, d'après le lemme 1.4 on a ou bien $f_1=\bar{x}$, ou bien $f_1 \in P$ et $f_2=\bar{x}$, dans les 2 cas, f admet un facteur gauche (f_1 ou $f_1 f_2$) ayant une image par δ négative.

On démontrerait de la même façon la propriété 1.4 suivante :

PROPRIÉTÉ 1.4 : *Les trois conditions suivantes sont équivalentes pour un mot f :*

(i) $b_f=0$.

(ii) *Pour tout f'' tel que $f=f'f''$ on a* $\delta(f'')\leqq 0$.

(iii) *Il existe un mot h tel que hf appartienne à P.*

2. LA BIJECTION B ENTRE SYSTÈMES DE PARENTHÈSES POINTES ET MOTS DE $\{x, \bar{x}\}^*$ DE LONGUEUR PAIRE

2.1. Les transformations D et G

Soit f un mot de $\{x, \bar{x}\}^*$ tel que a_f soit différent de zéro, il contient ainsi au moins un $\bar{x}$ dans sa factorisation $F=(f_1, f_2, \ldots, f_p)$ de Catalan; notons alors f_α le dernier mot égal à $\bar{x}$ dans cette factorisation ($f_\alpha=\bar{x}$ et $f_i\neq\bar{x}$ pour $i>\alpha$), on définit alors le mot $D(f)$ par :

$$D(f)=f_1 f_2 \ldots f_{\alpha-1}\, x f_{\alpha+1} \ldots f_p.$$

Ainsi $D(f)$ est obtenu à partir de f en changeant le dernier $\bar{x}$ de la factorisation en un x.

De même, si f est un mot possédant au moins un x dans sa factorisation de Catalan (soit si $b_f\neq 0$), on note f_β le premier mot égal à x ($f_\beta=x$ et $f_i\neq x$ pour $i<\beta$). $G(f)$ est alors obtenu en remplaçant ce premier x par un $\bar{x}$; soit :

$$G(f)=f_1 f_2 \ldots f_{\beta-1}\, \bar{x} f_{\beta+1} \ldots f_p$$

Par exemple, si $f=x\bar{x}\bar{x}x\bar{x}\bar{x}x\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}x\bar{x}xx\bar{x}xx\bar{x}$ qui se factorise en ($x\bar{x}$, $\bar{x}$, $x\bar{x}$, $\bar{x}$, $x\bar{x}$, $\bar{x}$, $\bar{x}$, $\bar{x}$, $\bar{x}$, $x\bar{x}$, x, $x\bar{x}$, x, $x\bar{x}$) on obtient $a_f=6$, $b_f=2$, $\alpha=9$, $\beta=11$ ce qui donne :

$$D(f)=x\bar{x}\bar{x}x\bar{x}\bar{x}x\bar{x}\bar{x}\bar{x}\bar{x}xx\bar{x}xx\bar{x}xx\bar{x},$$

$$G(f)=x\bar{x}\bar{x}x\bar{x}\bar{x}x\bar{x}\bar{x}\bar{x}\bar{x}\bar{x}x\bar{x}\bar{x}x\bar{x}xx\bar{x}.$$

En gardant les notations précédentes, on obtient :

PROPRIÉTÉ 2.1 : *Soit f un mot tel que $a_f\neq 0$; si $(f_1, f_2, \ldots, f_p)$ est la factorisation de Catalan de f alors :*

1) *La factorisation de Catalan de* $D(f)$ *est :*

$$(F')=(f_1, f_2, \ldots, f_{\alpha-1}, x, f_{\alpha+1}, \ldots, f_p);$$

2) $\chi(D(f))=(a_f-1, b_f+1)$.
3) $G(D(f))=f$.

Preuve : 1) est une conséquence immédiate de la propriété caractéristique 1.2; en effet (F') satisfait les conditions nécessaires car elle s'obtient à partir de (F) par un changement du dernier $\bar{x}$ en x;

2) est immédiate car il y a changement d'un $\bar{x}$ en x;

3) est une conséquence de la définition de G et de ce que le premier x dans (F) occupe la α-ième position.

On obtient de même la propriété 2.2.

PROPRIÉTÉ 2.2 : *Soit* f *un mot vérifiant* $b_f \neq 0$ *et admettant* $(F)=(f_1 \ldots f_p)$ *comme factorisation de Catalan, alors :*

1) $(F'')=(f_1, \ldots, f_{\beta-1}, \bar{x}, f_{\beta+1}, \ldots, f_p)$ *est la factorisation de Catalan de* $G(f)$;
2) $\chi(G(f))=(a_f+1, b_f-1)$;
3) $D(G(f))=f$.

NOTATION : Pour tout mot f et tout entier k inférieur ou égal à a_f on peut définir $D^k(f)$ comme le k-ième itéré par D de f; il s'obtient par échange des k derniers $\bar{x}$ de la factorisation de f en des x. De même on peut définir $G^h(f)$ pour tout h plus petit que b_f.

2.2. Système de parenthèses pointé

On considère l'alphabet Y contenant x, $\bar{x}$, et tous les entiers $\underline{0}, \underline{1}, \ldots, \underline{n}, \ldots$ que nous soulignerons quand ils seront considérés comme une lettre de Y.

DÉFINITION : Un mot w écrit sur l'alphabet Y est un système de parenthèses pointé s'il vérifie :

1) $w=f\underline{k}g, fg \in \mathrm{P}$;
2) $\delta(f) \geqq k$.

L'ensemble des systèmes de parenthèses pointés est noté $\underline{Q}$. Ainsi il y a 4 systèmes de parenthèses pointés de longueur 3 qui sont $\underline{0}x\bar{x}$, $x\underline{0}\bar{x}$, $x\underline{1}\bar{x}$ et $x\bar{x}\underline{0}$ et 16 systèmes de parenthèses pointés de longueur 5 : $\underline{0}xxx\bar{x}$, $x\underline{1}x\bar{x}\bar{x}$, $x\underline{0}x\bar{x}\bar{x}$, $xx\underline{2}\bar{x}\bar{x}$, $xx\underline{1}\bar{x}\bar{x}$, $xx\underline{0}\bar{x}\bar{x}$, $xx\bar{x}\underline{1}\bar{x}$, $xx\bar{x}\underline{0}\bar{x}$, $xx\bar{x}\bar{x}\underline{0}$, $\underline{0}x\bar{x}x\bar{x}$, $x\underline{1}\bar{x}x\bar{x}$, $x0\bar{x}x\bar{x}$, $x\bar{x}0x\bar{x}$, $x\bar{x}x\underline{1}\bar{x}$, $x\bar{x}x\underline{0}\bar{x}$, $x\bar{x}x\bar{x}\underline{0}$.

Pour un système de parenthèses pointé $w=f\underline{k}g$, fg est le *mot sous-jacent* et w est un *pointage* de fg. Le nombre de pointages du mot h est alors donné par la fonction $\sigma(h)=\sum_{h=fg}(\delta(f)+1)$. Ainsi le nombre de mots de Q de longueur $2n+1$ est $\tau(n)=\sum_{\substack{h\in P\\ |h|=2n}}\sigma(h)$.

Nous nous proposons d'établir que $\tau(n)$ est égal à 4^n en construisant une bijection entre les mots de Q de longueur $2n+1$ et les mots de $\{x, \bar{x}\}^*$ de longueur $2n$.

2.3. La bijection B

Soit $w=f\underline{k}g$ un système de parenthèses pointé; d'après les propriétés 1.3 et 1.4 on a $a_f=0$ et $b_g=0$; d'autre part $\delta(f)=-\delta(g)=b_f=a_g$, notons m cet entier qui est plus grand ou égal à k.

Le mot $B(w)$ se construit de deux manières différentes suivant que g commence par un x ou par un $\bar{x}$. Plus précisément :

DÉFINITION : *Le mot $B(w)$ est égal à $G^k(f)\,D^m(g)$ si g commence par un $\bar{x}$ ou si g est le mot vide; et à $G^m(f)\,\bar{x}\,D^k(g')$ si g commence par x, (où g' est défini par $g=xg'$).*

Il est simple de construire la factorisation de Catalan de $h=B(w)$ à partir de celles de f et g, on peut ainsi démontrer la propriété suivante :

PROPRIÉTÉ 2.3 : *Soit $w=f\underline{k}g$ un système de parenthèses pointé; son image h par B vérifie : [où l'on a posé $\delta(f)=m$] :*
$\chi(h)=(k, 2m-k)$ si g commence par $\bar{x}$ ou s'il est vide, et
$\chi(h)=(2m+2-k, k)$ si g commence par x.

Preuve : Soient $(f_1, f_2, \ldots, f_p)$ et $(g_1, g_2, \ldots, g_q)$ les factorisations de Catalan de f et g respectivement. Par les propriétés 1.3 et 1.4, f_i est ou bien un x ou bien un élément de P, de même g_j est ou bien un $\bar{x}$ ou bien dans P. Soient $f_{i_1}, f_{i_2}, \ldots, f_{i_m}$ les mots égaux à x dans la factorisation de f, et $g_{j_1}, g_{j_2}, \ldots, g_{j_m}$ ceux égaux à $\bar{x}$ dans la factorisation de g. Par construction, si g commence par $\bar{x}$, h est égal à $f'_1 f'_2 \ldots f'_p g'_1 g'_2 \ldots g'_q$ où f'_i est égal à f_i si f_i appartient à P ou si i est plus grand que i_k sinon $f_i=\bar{x}$; de même g'_i est égal à g_i si g_i appartient à P et $g'_{j_1}=g'_{j_2}=\ldots=g'_{j_k}=x$. La factorisation $(f'_1, f'_2, \ldots, f'_p, g'_1, g'_2, \ldots, g'_q)$ vérifie alors les conditions de la propriété 1.2; ainsi, c'est la factorisation de Catalan de h et la première relation en découle. Pour démontrer la seconde, examinons le cas où g commence par x, g_1 (premier élément de la factorisation de Catalan de g)

est alors dans P qui par la propriété 1.1 s'écrit $g_1 = xg'_1 \bar{x} g''_1$ où g'_1 et g''_1 sont des éléments de P. On peut alors écrire $h = B(w)$ sous la forme :

$$h = f'_1 f'_2 \dots f'_p \bar{x} g'_1 \bar{x} g''_1 g'_2 \dots g'_q$$

où :

$-f'_i = f_i$ si f_i appartient à P et $f'_{i_1} = f'_{i_2} = \dots = f'_{i_m} = \bar{x}$;

$-g'_j = g_j$ si g_j appartient à P ou si j est inférieur ou égal à j_{m-k} et $g'_{j_m} = g'_{j_{m-1}} = \dots = g'_{j_{m-k+1}} = x$.

Encore une fois la condition suffisante I.2 assure que $(f'_1, f'_2, \dots, f'_p, \bar{x}, g'_1, \bar{x}, g''_1, g'_2, g'_3, \dots, g'_q)$ est la factorisation de Catalan de h, et la propriété en découle.

Nous sommes maintenant en mesure de démontrer le résultat central de cet article.

THÉORÈME 2 : *L'application B est une bijection de l'ensemble des systèmes de parenthèses pointés sur l'ensemble des mots de $\{x, \bar{x}\}^*$ de longueur paire. De plus si w est de longueur $2n+1$, $B(w)$ est de longueur $2n$, ainsi le nombre de mots de Q de longueur $2n+1$ est 4^n.*

Il nous faut démontrer que la donnée d'un mot h de $\{x, \bar{x}\}^*$ de longueur paire détermine de manière unique un système de parenthèse pointé w tel que $B(w) = h$. Cette démonstration utilise de manière cruciale la factorisation de Catalan de h.

On peut tout d'abord remarquer que l'entier k (et donc la lettre $\underline{k}$ intervenant dans w) est d'après la propriété 2.3 déterminé de manière unique par :

$$k = \min(a_h, b_h).$$

Reste donc à construire les mots f et g; pour cela il faut déterminer si h vérifie :
(i) $h = G^k(f) D^m(G)$
ou bien (ii) $h = G^m(f) \bar{x} D^k(g')$.

Une fois encore la propriété 2.3 répond à cette question :

(i) est satisfaite si $\delta(h)$ est positif ou nul;

(ii) l'est par contre si $\delta(h)$ est négatif.

Examinons séparément ces deux cas :

(i) $\delta(h) \geqq 0$ la factorisation de Catalan (F) de h s'écrit alors :

$$(F) = (h_1, h_2, \dots, h_r)$$

et la propriété 2.3 permet de déterminer l'entier $\delta(f) = m$ par $m = (b_h + a_h)/2$. Soit alors p l'entier tel que $f' = h_1 h_2 \dots h_p$ vérifie :

$$\chi(f') = \left(a_h, \frac{b_h - a_h}{2}\right) \qquad \text{et} \qquad h_{p+1} = x,$$

f'contient donc tous les $\bar{x}$ de la factorisation de Catalan de h et $(b_h-a_h)/2$ de ces x.

On obtient alors :

$$h_1 h_2 \ldots h_p = G^k(f) \qquad \text{et} \qquad h_{p+1} \ldots h_r = D^m(g),$$

f et g sont finalement donnés par :

$$f = D^k(h_1 h_2 \ldots h_p), \qquad g = G^m(h_{p+1} \ldots h_r)$$

du fait des propriétés 2.1 et 2.2.

(ii) $\delta(h)<0$ Dans ce cas aussi la propriété 3 permet de déterminer l'entier $\delta(f)=m$, cette fois-ci par $m=(a_h-b_h)/2-1$.

A partir de la factorisation de Catalan $(h_1, h_2, \ldots, h_r)$ de h on peut alors construire le mot $h_1 h_2 \ldots h_p$ qui contient exactement m parmi les $a_h \bar{x}$ intervenant dans cette factorisation et qui vérifie de plus $h_{p+1}=\bar{x}$.

On obtient alors $G^m(f)=h_1 \ldots h_p D^k(g')=h_{p+2} \ldots h_r$; les propriétés 2.1 et 2.2 permettent d'écrire : $f=D^m(h_1 \ldots h_p)$ et $g'=G^k(h_{p+2}, \ldots, h_r)$; ainsi, f et $g(=xg')$ sont parfaitement déterminés.

Ce qui termine la preuve de la bijectivité de B. Le fait que $|B(w)|=|w|-1$ est contenu dans la définition de B, enfin le nombre de mots de $\{x, \bar{x}\}^*$ de longueur $2n$ est $2^{2n}=4^n$, l'existence de la bijection B assure que c'est aussi le nombre de mots de Q de longueur $2n+1$.

3. LIEN ENTRE LES SYSTÈMES DE PARENTHÈSES ET LES CHEMINS DU PLAN : UNE INTERPRÉTATION GÉOMÉTRIQUE DE LA BIJECTION B

3.1. Systèmes de parenthèses pointés et aires de chemins

A tout mot de longueur n de $\{x, \bar{x}\}^*$ on peut associer un chemin du plan joignant les points $M_0, M_1, \ldots, M_n$, dont les coordonnées (x_i, y_i) vérifient :

$$(x_0, y_0)=(0, 0), \qquad x_{i+1}=x_i+1$$

et

$$y_i=y_{i-1}+1 \qquad \text{ou} \qquad y_i=y_{i-1}-1$$

suivant que la i-ième lettre de f est un x ou un $\bar{x}$.

Le mot f est un système de parenthèses si et seulement si le chemin est au-dessus de l'axe des x et si $y_n=0$.

A un système de parenthèses pointé $f\underline{k}g$ est associé un chemin et un point de coordonnées $(|f|, k)$. Ce point se trouve entre le chemin et l'axe des x; ainsi, pour un système de parenthèses h, le nombre $\tau(h)$ de pointages de h est aussi le nombre de points à coordonnées entières situées à l'intérieur ou sur la figure formée par l'axe des x et le chemin associé à h. Ce nombre peut être interprété comme « l'aire » du chemin (voir *fig.* 1).

Ainsi, la propriété prouvée peut être reformulée ainsi : la somme des aires des chemins de longueur $2n$ situés au-dessus de l'axe des x est 4^n.

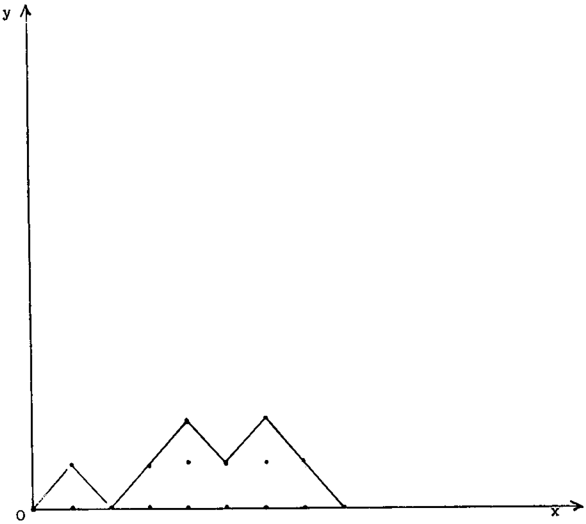

Fig. 1. — Le chemin associé à $x\bar{x}xx\bar{x}x\bar{x}\bar{x}$ **dont l'aire vaut 17.**

3.2. Interprétation géométrique de la factorisation de Catalan et de la bijection B

La factorisation de Catalan d'un mot f de $\{x, \bar{x}\}^*$ peut être considérée comme la sélection de certaines lettres du mot f (les $\bar{x}$ sélectionnés dans le mot doivent apparaître avant les x) de telle façon que les mots situés entre ces lettres constituent des systèmes de parenthèses. Si on représente le mot f par un chemin, la sélection de ces lettres consiste à distinguer d'une certaine manière certains pas du chemin; le choix de ces pas peut être présenté de la façon imagée suivante :

Imaginons deux sources de lumière, l'une située sur la portion de l'axe des y en des points de coordonnées négatives, l'autre sur une parallèle à l'axe des y menée du point qui termine le chemin et située en dessous de ce point; les pas sélectionnés sont ceux éclairés par ces sources (voir *fig.* 2).

Étant donné un chemin, et un point situé entre ce chemin et l'axe des x, chemin représentant le système de parenthèses pointé $f\underline{k}g$ on peut trouver le chemin associé à $B(w)$ en procédant en deux étapes : on sélectionne tout d'abord un certain nombre de pas de ce chemin, on inverse ensuite le sens (montée-descente des pas sélectionnés dans ce chemin).

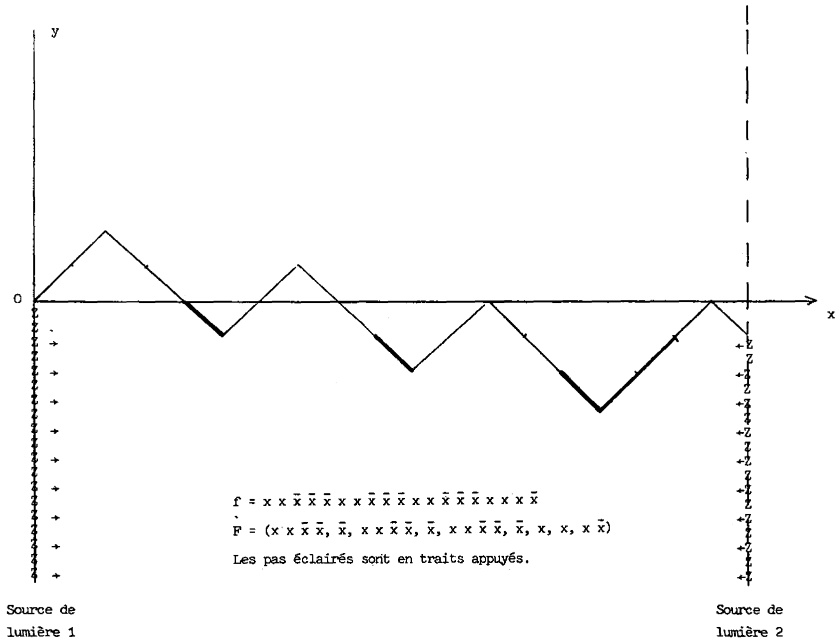

Fig. 2

Pour sélectionner les pas, on procède comme suit : on trace la droite Δ d'équation $x=|f|$ ensuite :

— si le premier pas du chemin à droite de Δ est descendant, ou si tout le chemin est à gauche de Δ, on considère une source lumineuse qui va de l'axe des x jusqu'au point M, à gauche de Δ; et de l'axe des x jusqu'au chemin, à droite de Δ. Les pas sélectionnés sont ceux éclairés par les sources;

— si le premier pas à droite de Δ est montant, on sélectionne ce pas ainsi que ceux éclairés par les sources allant de l'axe des x jusqu'à M, à droite de Δ; et de l'axe des x jusqu'au chemin à gauche de Δ (voir *fig.* 3).

4. APPLICATION AUX ARBRES

Nous nous proposons d'examiner ici comment se traduit le résultat obtenu, en utilisant le codage bien connu des arbres dessinés dans le plan et des arbres binaires par des systèmes de parenthèses.

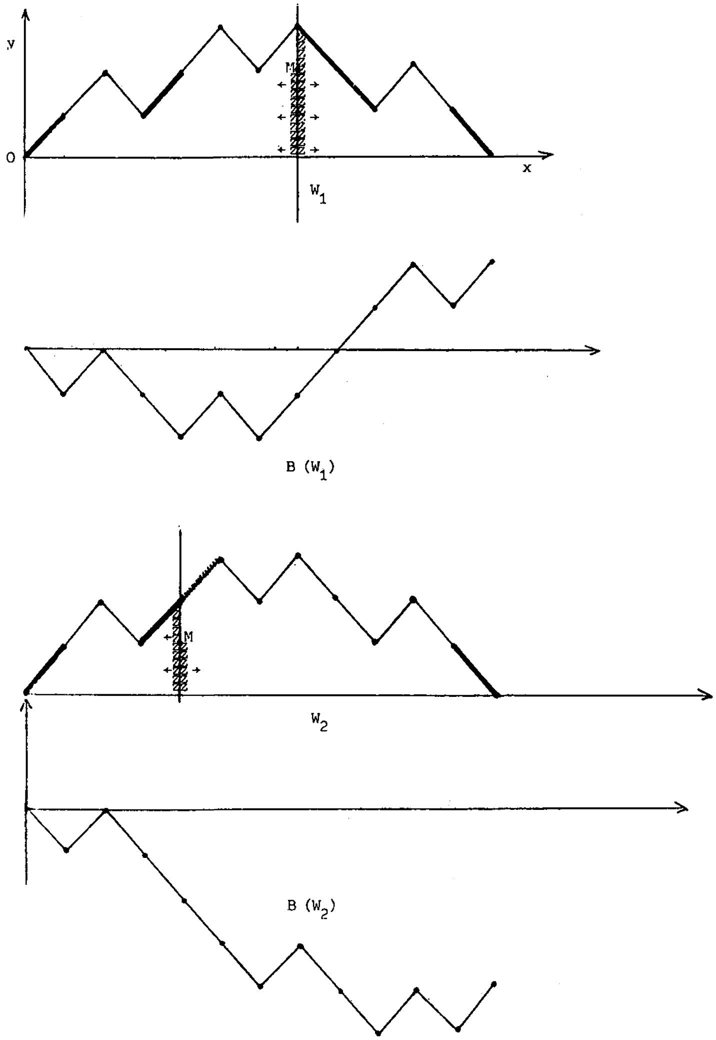

Fig. 3. — La bijection B sur deux cas génériques.

4.1. Arbres dessinés dans le plan

A un arbre dessiné dans le plan ayant une arête distinguée est classiquement associé un système de parenthèses par le procédé suivant : on parcourt la face infinie à partir de l'arête distinguée et on inscrit un x lorsqu'on parcourt une nouvelle arête, un $\bar{x}$ si on emprunte une arête déjà parcourue.

On peut alors associer à tout facteur gauche g du système de parenthèses, un chemin de l'arbre code. Ce chemin joint la racine à un sommet s_g de l'arbre, il peut passer plusieurs fois par une arête donnée, la distance du sommet s_g à la racine (ou la hauteur de s_g) est donnée par $\delta(g)$.

Dans la liste des sommets s_g on s'aperçoit qu'un même sommet intervient un nombre de fois égal à son degré; par contre si l'on convient d'associer s_g à g uniquement dans le cas où g se termine par x et $s_{g'}$ à $g=g'\bar{x}$ on vérifie qu'un sommet s est alors rencontré exactement deux fois (à part la racine qui n'est jamais rencontrée). On obtient ainsi en notant $H(\mathscr{A})$ la somme des hauteurs de tous les sommets de l'arbre $\mathscr{A}$ codé par f :

$$2H(\mathscr{A})=\sum_{f=gh}\delta(g)+|f|_{\bar{x}}.$$

En remarquant que :

$$\sigma(f)=\sum_{f=gh}\delta(g)+|f|+1,$$

on a :

$$2H(\mathscr{A})=\sigma(f)-\frac{1}{2}|f|-1.$$

En appliquant le résultat précédent et en utilisant le fait que le nombre de systèmes de parenthèses de longueur $2n$ est le nombre de Catalan $(2n)!/n!(n+1)!$ on obtient que la somme des $H(\mathscr{A})$ étendue à tous les arbres ayant n arêtes est donnée par :

$$\frac{1}{2}\left(4^n-(n+1)\frac{(2n)!}{n!\,n+1!}\right).$$

Soit :

$$\frac{1}{2}\left(4^n-\binom{2n}{n}\right).$$

4.2. Arbres binaires [6, p. 315]

Il s'agit d'arbres dans lesquels tout sommet a 0 ou 2 successeurs (un fils à gauche et un fils à droite). L'algorithme qui permet de les coder par un système de parenthèses consiste tout en parcourant l'arbre dans l'ordre préfixé (passer par la racine puis par le sous-arbre gauche, enfin par le sous-arbre à droite), à inscrire dans l'ordre où sont rencontrés les sommets un $\bar{x}$ chaque fois que l'on passe par une feuille, un x lorsqu'on rencontre un sommet ayant deux fils. Pour obtenir effectivement un système de parenthèses, il faut supprimer l'$\bar{x}$ final.

On se convainc facilement sur un exemple, que pour tout système de parenthèse f codant un arbre binaire $\delta(g)$ pour un facteur gauche g de f dénote alors le nombre d'arêtes « gauches » qu'emprunte le chemin joignant la racine au sommet correspondant à la dernière lettre de g. On désigne cette quantité par $h_g(s)$; symétriquement $h_d(s)$ désigne le nombre d'arêtes droites qu'emprunte le chemin défini plus haut. La somme de ces quantités pour tous les sommets de l'arbre $\mathscr{A}$ est désignée par $H_g(\mathscr{A})$ et $H_d(\mathscr{A})$ respectivement et $H(\mathscr{A})=H_g(\mathscr{A})+H_d(\mathscr{A})$ est la valeur cumulée des hauteurs des sommets de l'arbre binaire $\mathscr{A}$.

Le même type de remarque que le paragraphe précédent donne pour le mot f est l'arbre binaire $\mathscr{A}$ associé :

$$\sigma(f)=H_g(\mathscr{A})+n+1 \qquad (\text{ou } |f|=2n).$$

L'opération qui consiste à retourner un arbre binaire est une involution qui échange H_g et H_d. Ainsi :

$$\sum H_g(\mathscr{A})=\sum H_d(\mathscr{A})=\frac{1}{2}\sum H(\mathscr{A})$$

où les sommes sont étendues à tous les arbres binaires ayant $2n+1$ sommets. On obtient alors grâce au résultat du paragraphe 2 :

$$\sum H(\mathscr{A})=2\left(4^n-(2n+1)\frac{2n!}{n!(n+1)!}\right).$$

Dans [6] Knuth remarque que si $H'(\mathscr{A})$ dénote la somme des hauteurs des sommets qui ne sont pas des feuilles on a $2H'(\mathscr{A})=H(\mathscr{A})-2n$; on retrouve ainsi la valeur de la somme de $H'(\mathscr{A})$ étendue à tous les arbres binaires ayant n feuilles :

$$4^n-(3n+1)\frac{2n!}{n!(n+1)!}.$$

BIBLIOGRAPHIE

1. L. Comtet, *Analyse combinatoire*, vol. 1 et 2, Presses Universitaires de France, Paris, 1970.
2. R. Cori, *Fichiers inverses et propriétés combinatoires des systèmes de parenthèses*, Séminaire d'Informatique théorique, Université Paris-VI, 1979.
3. P. Flajolet, *Combinatorial Aspects of Continued Fractions*, à paraître dans Discrete Mathematics.
4. P. Flajolet, J. Françon et J. Vuillemin, *Analysis of Data Structures Under Sequence of Operations* (à paraître).
5. J. Françon, *Histoires de fichiers*, R.A.I.R.O., Informatique théorique, vol. 12, 1978, p. 49-62.

6. D. KNUTH, *The Art of Computer programming*, Addison-Werley, Reading, vol. 13, 1973, p. 404 et 590.
7. I. GESSEL, *A Non Commutative Generalization and q Analog of Lagrange Inversion Formula*, Trans. Amer. Soc., vol. 257, 1980, p. 455-482.
8. G. KREWERAS, *Aires de chemins surdiagonaux et application à un problème économique*, Cahiers du B.U.R.O., vol. 24, 1976, p. 1-8.
9. J. RIORDAN, *An Introduction to Combinatorial Analysis*, John Wiley and Sons, New York, 1958.
10. J. RIORDAN, *Combinatorial Identities*, John Wiley and Sons, New York, 1968.
11. M. P. SCHÜTZENBERGER, *On a Factorisation of Free Monoïds*, Proc. Amer. Math. Soc., vol. 16, 1965, p. 21-24.
12. G. VIENNOT, *Springer Lectures Notes in Mathematics*, n° 691.