

BRUNO LECLERC

Les hiérarchies de parties et leur demi-treillis

Mathématiques et sciences humaines, tome 89 (1985), p. 5-34

http://www.numdam.org/item?id=MSH_1985__89__5_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1985, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LES HIERARCHIES DE PARTIES ET LEUR DEMI-TREILLIS

Bruno LECLERC*

INTRODUCTION.

Les hiérarchies de parties sont l'un des objets fondamentaux de la classification mathématique depuis que Benzecri (1967) les a introduites de façon formalisée. Ces familles de parties, structurées en arborescences par l'inclusion, apparaissent dans toutes les classifications hiérarchiques ; sous diverses appellations, elles peuvent être considérées directement (arbres de classification, ou, en anglais, *dendrograms*, *n-trees*,...) ou par implication dans des structures plus fortes : hiérarchies stratifiées ou ultrapréordonnances (*ranked trees*, *ordered trees*...) d'une part, hiérarchies indicées ou ultramétriques, ou encore équivalence floues (*valued-trees*, *fuzzy equivalence relations*, ...) d'autre part [cf., par exemple, les ouvrages de Chandon et Pinson (1980), Hartigan (1975), Lerman (1970, 1981), Schader (1981), les articles de Boorman et Olivier (1973), Barthélemy, Leclerc et Monjardet (1984 b), et, plus particulièrement sur les ultrapréordonnances, de Schader (1980) et sur les ultramétriques, Leclerc (1981)].

D'autres types d'arbres s'apparentent sur certains points aux hiérarchies : arbres de recherche en informatique théorique, arbres intervenant comme modèles dans certaines grammaires formelles, en différenciation cellulaire, etc. D'un point de vue général, à tout arbre pointé ou arborescence correspond une hiérarchie sur ses feuilles, ou sommets pendants, qui en retient un aspect classificatoire.

* Centre d'Analyse et de Mathématiques Sociales - EHESS - Paris

On se propose ici d'étudier les hiérarchies en vue d'aborder, dans un second article, les problèmes de les comparer et de les évaluer. On considérera donc d'une part des propriétés et des paramètres des hiérarchies permettant la description d'une hiérarchie donnée, et d'autre part l'ensemble \mathcal{H}_X des hiérarchies sur un ensemble fini donné X , certaines relations de proximité ou de ressemblance entre hiérarchies, et la structure algébrique de demi-treillis de \mathcal{H}_X . L'ensemble des deux articles complète et met à jour un travail non publié de 1982.

Le fait de considérer d'abord les hiérarchies comme des familles de parties conduit à ordonner celles-ci par l'inclusion qui a un sens très concret, et à considérer d'emblée le demi-treillis des hiérarchies (par. 1.2.) et ses propriétés générales (par. 1.3.) ; on obtient notamment une première typologie, par niveaux, des hiérarchies, et un premier paramètre descripteur donné par la graduation.

Au paragraphe 2, nous passons à la description des hiérarchies, d'abord en les considérant en tant qu'arbres (par. 2.1., 2.2.), puis en introduisant divers paramètres (par. 2.3.) affectés à chacune de leurs classes, et en établissant quelques relations entre ces paramètres. On considère alors deux types de hiérarchies particulières : hiérarchies correspondant aux partitions de X d'une part, hiérarchies maximales (binaires) d'autre part (par. 2.4.). On définit une relation d'échange sur les hiérarchies binaires qui est en fait, sous un aspect nouveau, une relation de voisinage (*nearest neighbour interchange*) sur certains arbres dont l'étude a été récemment abordée par divers auteurs et s'est révélée difficile (par. 2.5.).

Aux paragraphes 2.6. et 2.7., on introduit deux paramètres plus élaborés décrivant une hiérarchie donnée. Dans le second, on fait d'abord une étude des préordres sur les paires d'éléments de X induits par les hiérarchies, que nous appelons préordonnances hiérarchiques.

Au paragraphe 3, on donne des codages des hiérarchies, le premier classique, par parenthésages, le second découlant de leur correspondance biunivoque avec certaines partitions [Comtet (1970 b)].

On considère au paragraphe 4 les hiérarchies non étiquetées (formes). Le demi-treillis \mathcal{H}_X , le filtrant semi-modulaires des formes à $|X|$ feuilles et le filtrant modulaire des partages de l'entier $|X|-1$ sont liés par des morphismes d'ordre injectifs préservant les niveaux.

Enfin, au paragraphe 5, on fait une revue, avec quelques compléments, des résultats connus sur le dénombrement des hiérarchies, notamment ceux de Comtet.

1. L'ENSEMBLE DES HIERARCHIES SUR X.

1.1. Définitions. On considère un ensemble X fini (avec $|X| = n$), et les ensembles $\mathcal{P}(X)$ des parties de X et $\mathcal{P}^2(X) = \mathcal{P}(\mathcal{P}(X))$. Une hiérarchie H sur X est un élément de $\mathcal{P}^2(X)$ vérifiant les axiomes (H1, H2, H3, H4) :

(H1) $X \in H$; (H2) $\emptyset \notin H$; (H3) $(\forall x \in X) \{x\} \in H$;
 (H4) $(\forall h, h' \in H) h \cap h' \in \{h, h', \emptyset\}$.

On pose $F = \mathcal{P}_1(X) = \{\{x\} / x \in X\}$. On écrira en général x pour $\{x\}$. F est l'ensemble des *feuilles* de H. $I = H - F$ est l'ensemble des éléments (ou classes) *intérieur(e)s* de H. Pour $h \in H$, on note $n_h = |h|$ et $H_h = \{h' \in H / h' \subseteq h\}$, la hiérarchie sur h incluse dans H et maximale avec cette propriété. Pour $i \in I$, on note S_i la partition unique de i en éléments de H strictement inclus dans i et maximaux avec cette propriété, et on pose $s_i = |S_i|$. Comme pour $H = H_X$ et $\underline{n} = \underline{n}_X$, on note $S = S_X$ et $\underline{s} = \underline{s}_X$. Il en sera de même pour les autres ensembles ou paramètres liés aux éléments de H.

Quand on aura à considérer simultanément plusieurs hiérarchies, on notera, par exemple, $I_H, s_{H,i}$ pour préciser I, s_i .

Pour $h \in H - \{X\}$, (H4) entraîne que l'ensemble $\{h' \in H / h \subset h'\}$ est totalement ordonné par l'inclusion. On note $p(h)$ l'élément minimum de cet ensemble, c'est-à-dire la classe de H contenant h et minimum avec cette propriété. $p(h)$ sera parfois appelé *père* de h.

1.2. Le demi-treillis des hiérarchies sur X. L'ensemble \mathcal{H}_X (noté parfois simplement \mathcal{H}) des hiérarchies sur X, ordonné par l'inclusion dans $\mathcal{P}^2(X)$, est stable pour l'intersection : $(\forall H_1, H_2 \in \mathcal{H}) H_1 \cap H_2 \in \mathcal{H}$.

Par contre, pour $\underline{n} \geq 3$, et pour $H_1, H_2 \in \mathcal{H}$, on a en général $H_1 \cup H_2 \notin \mathcal{H}$. Pour l'ordre d'inclusion, ou l'opération d'intersection, \mathcal{H} est donc un inf-demi-treillis, dont le minorant universel est la hiérarchie minimale $H_0 = \{X\} \cup F$.

La relation de *couverture* \prec dans (\mathcal{H}, \subseteq) , définie comme usuellement, est évidente. Pour $H, H', H'' \in \mathcal{H}$ on a :

$$[H' \prec H] \iff [H' \subseteq H'' \subseteq H \Rightarrow H'' \in \{H, H'\}] \iff [(\exists i \in I_H) H' = H - \{i\}].$$

Pour $i, j \in I_H - \{X\}$ (resp. $J \subseteq I_H - \{X\}$), on posera $H^i = H - \{i\}$ et $H^{ij} = H - \{i, j\}$, avec la convention $i \not\subseteq j$ (resp. $H^J = H - \{J\}$). On utilisera de même les notations S_j^i ou $S_{H,j}^i$, \underline{s}_j^i , etc., pour $S_{Hi,j}$, $\underline{s}_{Hi,j}$, etc.

L'ensemble $\min H = \{H' \in \mathcal{H} / I_{H'} \subseteq I_H\}$ des minorants de H dans \mathcal{H} est isomorphe pour l'ordre d'inclusion à l'ensemble $\mathcal{P}(I_H - \{X\})$. Il est stable pour l'union : $H_1, H_2 \in \min H \Rightarrow H_1 \cup H_2 \in \min H$.

L'intérêt qu'il y a à considérer la structure d'inclusion dans \mathcal{H}_X est qu'il est très fréquent qu'à une hiérarchie H , objet d'étude, soit substituée, volontairement ou non, une hiérarchie $H' \subseteq H$.

- Volontairement : par exemple, on a obtenu H par une méthode de classification hiérarchique. Le nombre de classes de H est trop grand pour qu'on puisse les prendre toutes en considération et H' est une simplification de H .

- Involontairement : H' est la partie connue d'une hiérarchie H hypothétique. H est par exemple un arbre phylogénétique et H' l'ensemble des classes suffisamment bien identifiées de H .

1.3. \mathcal{H}_X est un demi-treillis à médianes.

Un inf-demi-treillis T est un demi-treillis à médianes ssi il vérifie les deux propriétés suivantes : (M1) pour tout $t \in T$, $\min t$ est un treillis distributif ; (M2) pour tout $t_1, t_2, t_3 \in T$, si $t_1 \vee t_2, t_2 \vee t_3$ et $t_1 \vee t_3$ existent, alors $t_1 \vee t_2 \vee t_3$ existe aussi [Sholander (1954), Bandelt et Hedlikova (1983)]. Ces demi-treillis conservent un grand nombre des propriétés des treillis distributifs, notamment l'équivalence des définitions algébrique et métrique de la médiane [Bandelt et Barthélemy (1984) ; Barbut (1961) pour les treillis distributifs].

Une *charpente* d'un ensemble ordonné fini O est une partie O' de O en contenant tous les éléments minimaux et maximaux, et une chaîne au plus entre tout minimal et tout maximal. On montre facilement que l'ensemble des charpentes de O , ordonné par l'inclusion, constitue un demi-treillis à médianes [cf. Barthélemy, Leclerc et Monjardet (1984, b)].

Comme les hiérarchies sur X sont les charpentes de $\mathcal{P}(X) - \{\emptyset\}$, le demi-treillis \mathcal{H}_X est un demi-treillis à médianes. On en déduit des propriétés intéressantes de la médiane (c'est-à-dire de la règle majoritaire) comme solution au problème du consensus des hiérarchies [Margush et Mac Morris (1981), Barthélemy, Leclerc et Monjardet (1984, a)].

Pour la comparaison des hiérarchies, on utilisera plutôt le fait que \mathcal{H} est *semi-modulaire* (inférieurement), c'est-à-dire qu'il vérifie :

(SMI) $(\forall H, H_1, H_2 \in \mathcal{H}) [H > H_1] \text{ et } [H > H_2] \Rightarrow (\exists H' \in \mathcal{H}) [H_1 > H'] \text{ et } [H_2 > H']$

Avec $H_1 = H^i$ et $H_2 = H^j$, on a $H' = H^{ij}$. Posant $Q^{ij}(H) = \{H, H^i, H^j, H^{ij}\}$, $Q^{ij}(H)$ est un *quadrilatère* de \mathcal{H} .

Depuis leur introduction par Ore (1943), les ensembles ordonnés semi-modulaires ont été l'objet de nombreux travaux, portant notamment sur leurs propriétés métriques [Haskins et Gudder (1972), Comyn et van Dorpe (1976), Monjardet (1976), Barthélemy (1978), entre autres ; cf. la revue récente de Monjardet (1981)]. Ils sont gradués, ce qui signifie, pour (\mathcal{H}, \subseteq) qu'il existe une fonction (*graduation*) $\gamma = \mathcal{H} \rightarrow \{0, 1, 2, \dots\}$ vérifiant $\gamma(H_0) = 0$ et $[H' \prec H] \Rightarrow [\gamma(H) = \gamma(H') + 1]$. En fait, on a $\gamma(H) = |I_H - \{X\}| = |I_H| - 1$. Pour $0 \leq k \leq \max \{\gamma(H) / H \in \mathcal{H}\}$, le k -ième niveau $\mathcal{H}(k)$ de \mathcal{H} est l'ensemble $\gamma^{-1}(k)$, avec k entier.

2. DESCRIPTION D'UNE HIERARCHIE.

Avant de pousser plus loin l'étude de l'ensemble \mathcal{H} , il est nécessaire de définir des paramètres permettant la description de toute hiérarchie particulière.

2.1. Arbre G_H associé à une hiérarchie H .

C'est le graphe $G_H = (H, U_H)$, où $U_H = \{(h, h') \in H \times H / h' \in S_h\}$. La suite $p(h), p(p(h)), \dots$, aboutit à X . Il existe donc dans G_H , un chemin unique de la forme $\{(X, h_1), (h_1, h_2), \dots, (h_{\ell-1}, h)\}$ entre X et h . On note L_h ce chemin, et on pose $\ell_h = |L_h|$. ℓ_h est le nombre d'éléments de H contenant strictement h .

Par l'existence et l'unicité de L_h , on voit que le graphe G_H est une *arborescence*, c'est-à-dire un arbre orienté possédant une racine (qui est X). Pour les définitions et propriétés des arbres et arborescences, nous renvoyons aux livres de Berge (1970) et de Harary (1969).

Les feuilles, éléments de F , sont les sommets pendants (i.e. de degré 1, c'est-à-dire ayant exactement un voisin) de G_H . Tout autre sommet $i \in I_H$ a un demi-degré extérieur $s_i \geq 2$ et, sauf X , un degré $s_i + 1$ au moins égal à 3.

Réciproquement, considérons une arborescence $G = (Y, V)$ telle que :

- (A1) l'ensemble des sommets pendants de G est F .
- (A2) la racine de G est un élément de $Y - F$.
- (A3) aucun demi-degré extérieur n'est égal à 1.

Alors tout sommet intérieur $i \in Y - F$ peut être identifié à l'ensemble des éléments $y \in F$ tels qu'il y a un chemin de G de i à Y , et à G correspond une hiérarchie unique sur X . Seul l'étiquetage des sommets pendants importe. Deux arborescences isomorphes dont l'étiquetage ne diffère que sur des sommets intérieurs définissent la même hiérarchie.

La figure 1 suivante donne l'arbre de la hiérarchie H sur $X = \{a, b, c, d, e, f, g, h\}$, telle que $H = \{X\} \cup F \cup \{\{a, b, d, f, g\}, \{a, f\}, \{b, g\}, \{c, e, h\}, \{e, h\}\}$.

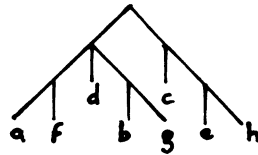


Figure 1

La figure 2 illustre un certain nombre des notations précédentes sur les hiérarchies.

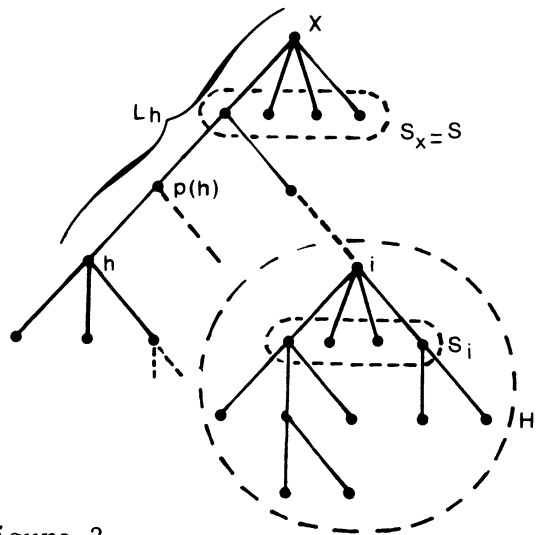


Figure 2

2.2. Position des hiérarchies dans la typologie des arbres.

Un arbre est *pointé* lorsqu'un de ses sommets est distingué, et *planté* si ce sommet est pendent. Il y a une correspondance bi-univoque entre arbres pointés et arborescences (rooted trees).

Un arbre est *ordonné* (ordered tree) lorsqu'on s'est donné un ordre circulaire sur les voisins de tout sommet intérieur i . S'il est pointé en r , le voisin unique de i situé sur la chaîne entre r et i est distingué et il revient au même de se donner un ordre total sur les autres voisins de i . Les arbres considérés en informatique théorique sont en général de ce dernier type [cf. Knuth (1969 , 1973)].

Un arbre est (*homéomorphiquement*) *réduit* (homeomorphically reduced, series-reduced) quand il n'a pas de sommets de degré 2. On réduit un graphe quelconque par la répétition de l'opération : remplacer un sommet de degré 2 et ses deux arêtes incidentes par une arête unique. Un arbre réduit est *ternaire* quand ses sommets intérieurs sont tous de degré 3. Les arbres réduits ont parfois été appelés *graphes de Halin*.

Un arbre est *étiqueté* (labelled), ou *marqué*, quand ses sommets sont les éléments d'un ensemble donné. L'utilité de cette notion est de définir les arbres non (ou partiellement) étiquetés, qui sont des classes d'isomorphisme d'arbres étiquetés.

Le graphe G_H associé à une arborescence H est donc *pointé* en X , *non planté*, *non ordonné*, *presque réduit*, en ce sens que seul X peut être de degré 2, *partiellement étiqueté* quant aux feuilles ; les sommets intérieurs, on l'a vu, sont a posteriori assimilés à des ensembles d'au moins deux feuilles, ce qui leur attribue au plus $\underline{n} - 1$ étiquettes parmi $2^{\underline{n}} - \underline{n} - 1$ possibles. A noter que Harary, Mowshovitz et Riordan (1969) dénombrent des arbres dont, à l'opposé, seuls les sommets intérieurs sont étiquetés, par nombres de sommets pendants et intérieurs (sans autre restriction). Ajoutons à G_H un sommet pendent supplémentaire r , avec l'arête (r, X) et plantons en r le graphe G_H' obtenu. On obtient une correspondance bi-univoque entre arbres associés aux hiérarchies et *arbres réduits plantés*, partiellement étiquetés en leurs $\underline{n} + 1$ sommets pendants. L'arbre de la figure 3 correspond à la hiérarchie de la figure 1.

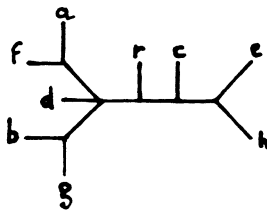


Figure 3

2.3. Couverture et contraction.

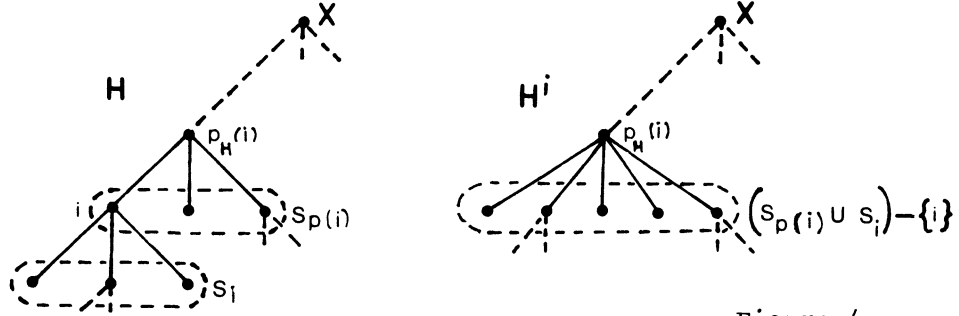


Figure 4

La figure 4 illustre le fait que, pour $i \in I - \{X\}$, au passage de H à H^i correspond, pour le passage de G_H à G_H^i , une opération classique sur les graphes : la contraction de l'arc $(p(i), i)$. On a $U_H^i = U_H - \{(p(i), i)\}$ et $S_{p(i)}^i = S_{p(i)} \cup S_i - \{i\}$. Pour $h \in H - \{i\}$, $\ell_h^i = \ell_h - 1$ si $h \in H_i$ et $\ell_h^i = \ell_h$ sinon. Posons, pour $i \in I$, $\underline{t}_i = \underline{s}_i - 1$. Les nombres \underline{t}_i joueront un rôle important.

PROPOSITION 2.1. Les nombres \underline{t}_i , $i \in I$, vérifient :

- (1) $(\forall i \in I) \quad \underline{t}_{p(i)}^i = \underline{t}_i + \underline{t}_{p(i)} \quad (\text{et } (\forall j \in I - \{i, p(i)\}) \quad \underline{t}_j^i = \underline{t}_j)$
- (2) $\sum_{i \in I} \underline{t}_i = \underline{n} - 1$

(1) provient de $\underline{s}_{p(i)}^i = \underline{s}_{p(i)} + \underline{s}_i - 1$. Par suite la somme des \underline{t}_i est constante dans \mathcal{H} , et égale à $\underline{n} - 1$, comme on le voit en considérant H_0 . \square

La propriété (2) se montre aussi en effectuant deux dénombrements de U_H : $|U_H| = \sum \underline{s}_i = \underline{n} + |I| - 1$ (le nombre des arcs de tout arbre est inférieur d'une unité à celui des sommets).

PROPOSITION 2.2. Les nombres $\underline{\ell}_x$, $x \in X$; $\underline{\ell}_i$, \underline{n}_i , \underline{t}_i , $i \in I$, vérifient :

$$\sum_{x \in X} \underline{\ell}_x = \sum_{i \in I} \underline{n}_i = \gamma(H) + 1 + \sum_{i \in I} \underline{t}_i (\underline{\ell}_i + 1).$$

DEMONSTRATION. La première égalité provient de ce que, pour $i \in I - \{X\}$, le nombre des $x \in X$ tels que $(p(i), i) \in L_x$ est \underline{n}_i , et que le nombre d'arcs $(p(x), x)$ est \underline{n} . Pour la seconde, on vérifie que les deux quantités sont égales à \underline{n} pour $H = H_0$, et diminuent de \underline{n}_i quand on passe de H à H^i . \square

On peut aussi mettre la seconde égalité sous les formes :

$$\sum_{i \in I} (\underline{n}_i - 1 - \underline{t}_i (\underline{\ell}_i + 1)) = \sum_{i \in I} (\underline{s}_i \underline{\ell}_i + \underline{s}_i - \underline{n}_i - \underline{\ell}_i) = 0$$

Grandes hiérarchies. Dans certaines applications (classifications animales ou végétales, par exemple), l'individualisation des éléments de X , de cardinal très grand, a peu d'intérêt. Pour avoir une définition des hiérarchies n'utilisant pas les éléments de F , on peut considérer la partition B induite par p sur X . On suppose que toute classe $b \in B$ est de cardinal grand, en tout cas supérieur à 1. Le sous-graphe $G_I = (I, U_I)$, restriction de G_H à I est une arborescence, non forcément réduite, et à chaque classe $b \in B$ correspond un de ses sommets $i = p(x)$, pour tout $x \in b$.

Notons \underline{s}_i le demi-degré extérieur de i dans G_I , et soit $J \subseteq I$, avec $i \in J$ ssi $\underline{s}_i = 0$. Posant, pour tout $i \in I$, $\underline{t}_i = \underline{s}_i - 1$ on a $\sum_{i \in I} \underline{t}_i = -1$. La condition $\underline{s}_i \geq 2$ pour les hiérarchies est remplacée par :

$$(\forall i \in I) [\underline{t}_i \leq 0 \Rightarrow (\exists b \in B) i = p'(b)].$$

Cette présentation des hiérarchies en permet certaines généralisations. Prenons par exemple $I = \mathbb{Z}$, $U_I = \{(\underline{z}, \underline{z} - 1) / \underline{z} \in \mathbb{Z}\}$. Le graphe infini G_I satisfait certaines définitions des arborescences. Avec $X = \mathbb{R}$, $B = \{] \underline{z} - 1, \underline{z}] / \underline{z} \in \mathbb{Z}\}$, $p'(] \underline{z} - 1, \underline{z}]) = \underline{z}$, le quadruplet (I, U_I, B, p') peut être considéré comme une hiérarchie infinie sur \mathbb{R} .

2.4. Hiérarchies particulières.

2.4.1. H-partitions. Soit S une partition de X ; on lui associe la hiérarchie $H(S) = S \cup \{X\} \cup F$. Pour les partitions $S_\wedge = \{\{x\} / x \in X\}$, la plus fine, et $S_\vee = \{\{X\}\}$, la moins fine, on a $H(S_\wedge) = H(S_\vee) = H_0$. A ceci près, la correspondance $S \leftrightarrow H(S)$ est bi-univoque, et on dira que $H(S)$ est une *H-partition*. Les H-partitions sont les hiérarchies telles que : $\max \{\underline{l}_x / x \in X\} \leq 2$.

Si $H \in \mathcal{H}$ est une H-partition, il en est de même de tout $H' \in \min H$. La graduation $\gamma(H(S))$ est le nombre de classes de S ayant plus d'un élément, et est comprise entre 0 et $\lfloor \frac{n}{2} \rfloor$ (le plus grand entier au plus égal à $\frac{n}{2}$). Les niveaux $\mathcal{H}(0)$ et $\mathcal{H}(1)$ de \mathcal{H} sont les seuls à ne contenir que des H-partitions.

L'ordre de finesse des partitions [cf. par exemple, Barbut et Monjardet (1970)] et l'ordre d'inclusion des hiérarchies ne sont que peu liés.

2.4.2. Hiérarchies binaires. Le fait bien connu que $|I_H|$ est maximum et égal à $\underline{n} - 1$ (et $\gamma(H)$ à $\underline{n} - 2$) quand on a, pour tout $i \in I$, $s_i = 2$, donc $t_i = 1$, est une conséquence immédiate de l'égalité $\sum t_i = \underline{n} - 1$. La hiérarchie H est alors dite *binnaire*. L'arbre réduit planté associé est ternaire. Les hiérarchies binaires sont les éléments maximaux du demi-treillis \mathcal{H} , et on note $\mathcal{H}^m = \mathcal{H}(\underline{n} - 2)$ leur ensemble.

La proposition 2.2. généralise une égalité donnée par Knuth (1969) pour les hiérarchies binaires : en prenant les t_i tous égaux à 1, on trouve pour celles-ci :

$$\sum_{x \in X} \ell_x = 2\underline{n} - 2 + \sum_{i \in I} \ell_i.$$

Parmi les hiérarchies binaires, deux configurations apparaissent comme extrêmes dans de nombreux problèmes.

Définitions. Une hiérarchie binaire est un *peigne* (resp. une *hiérarchie binnaire équilibrée* (HBE)), ssi elle vérifie, pour tout $i \in I - \{X\}$,

$$\underline{n}_i = \underline{n}_{p(i)} - 1 \quad (\text{resp. } \underline{n}_i = \frac{1}{2} \underline{n}_{p(i)}).$$

Il existe des peignes pour tout $\underline{n} \geq 2$, et des HBE seulement si \underline{n} est une puissance de 2. La notion d'équilibre d'un arbre donnée par Knuth (1973) est moins restrictive.

2.5. Le graphe d'échanges des hiérarchies binaires.

Sur l'ensemble \mathcal{H}_X^m des hiérarchies binaires sur X , on définit la relation de voisinage $\mathcal{V}_X \subseteq (\mathcal{H}_X^m)^2$ suivante :

$$\mathcal{V} = \left\{ (H, H') \in \mathcal{H}^m / |H \Delta H'| = 2 \right\},$$

où Δ est la différence symétrique dans $\mathcal{P}^2(X)$. H et H' sont donc voisines si elles ne diffèrent que par l'échange d'une classe : il existe $i \in H$,

$j \in H' : H' = (H - \{i\}) \cup \{j\}$. H et H' sont voisines ssi $H, H' > H \cap H'$.

Posons $H \cap H' = H^i$, $S_{H,p(i)} = \{i, h_1\}$, $S_{H,i} = \{h_2, h_3\}$. Alors $H \cap H' = H'^j$,

avec $j = h_1 \cup h_2$ ou $j = h_1 \cup h_3$. Avec les $\underline{n} - 2$ classes intérieures

$i \in I - \{X\}$ possibles, H a $2\underline{n} - 4$ voisines dans \mathcal{H}^m . Nous dirons que le graphe $\mathcal{G}_X = (\mathcal{H}_X^m, \mathcal{V}_X)$ est le *graphe d'échanges* du demi-treillis $(\mathcal{H}_X, \subseteq)$.

La figure 5 montre ce graphe pour $n = 4$. Notons que c'est le graphe aux arêtes du célèbre graphe de Petersen.

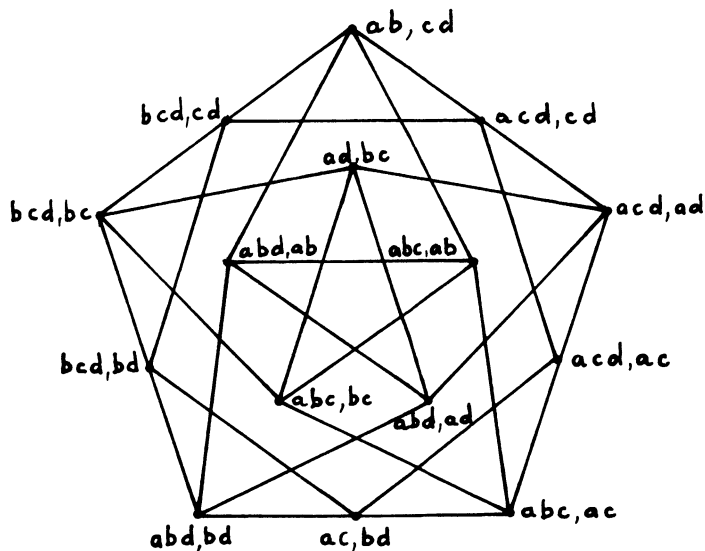


Figure 5. Les hiérarchies binaires sont données par leurs classes non triviales.

Un graphe défini de manière analogue dans un autre demi-treillis, est le permutoèdre des ordres totaux [Guilbaud et Rosenstiehl 1960] dans l'inf-demi-treillis des ordres [cf. Barthélemy (1979)].

Sous une forme assez différente, la relation \mathcal{U} a été introduite indépendamment par Robinson (1971) et par Moore, Goodman et Barnabas (1973). Considérons l'arbre ternaire réduit à $n + 1$ sommets pendants associé à G_H (par. 2.2.) et, dans cet arbre, l'arête v non pendante correspondant à $(p(i), i)$. Soient B_1 et B'_1 (resp. B_2 et B'_2) les deux branches de l'arbre ne contenant pas v incidentes à $p(i)$ (resp. à i). L'arbre ternaire réduit correspondant à $G_{H'}$, s'obtient par échange de B'_1 avec B_2 ou B'_2 . La figure 6 ci-dessous illustre la relation de voisinage pour deux hiérarchies (a) et pour deux arbres ternaires réduits (b).

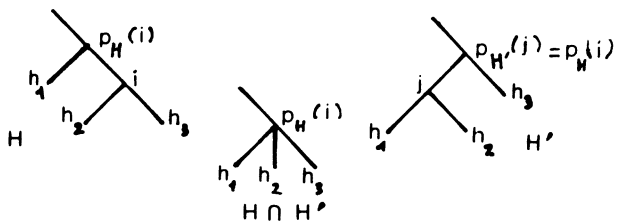


Figure 6 a

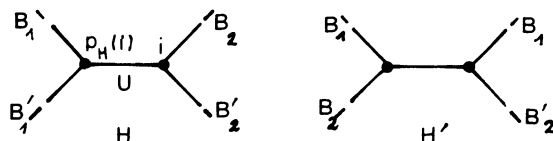


Figure 6 b

Le passage d'un arbre ternaire à un arbre voisin est le plus souvent appelé "Nearest neighbour interchange (NNI)", depuis l'article de Waterman et Smith (1978). Le graphe d'échanges des hiérarchies binaires sur X est isomorphe au graphe des NNI des arbres ternaires réduits à $\underline{n} - 1$ sommets pendants. De nombreux auteurs se sont récemment intéressés au calcul de la distance du plus court chemin dans ce graphe, mais ce problème s'est révélé très difficile [cf. par exemple Jarvis, Luedeman et Shier (1983), Brown et Day (1984)].

2.6. Le nombre de partitions incluses dans une hiérarchie.

L'étude de ce nombre va contribuer à situer les H-partitions par rapport aux autres hiérarchies, notamment aux hiérarchies binaires. Il a un intérêt évident en taxonomie mathématique.

Notons $\varepsilon(H)$ le nombre de partitions distinctes incluses dans H . Alors $\varepsilon(H) - 1$ est le nombre de H-partitions incluses dans $\min H$. Par définition, la fonction $\varepsilon : \mathcal{H} \rightarrow \mathbb{N}$ est strictement croissante : $H \subset H' \Rightarrow \varepsilon(H) < \varepsilon(H')$.

Pour construire une partition S' incluse dans H , on peut soit prendre $S' = \{X\}$, soit prendre une partition $S'(h)$ dans chaque sous-hiérarchie H_h , $h \in S$ et poser $S' = \bigcup_{h \in S} S'(h)$. D'où la récurrence :

$$\varepsilon(H) = 1 + \prod_{h \in S} \varepsilon(H_h)$$

Elle permet le calcul de $\varepsilon(H)$, pour tout $H \in \mathcal{H}$, à partir du fait que, pour $\underline{n} = 1, 2, 3$, on a $\varepsilon(H) = \underline{n}$.

Pour $\underline{n} = 4$, on a $\varepsilon(H) \geq 4$ et $\varepsilon(H) = 4$ ssi H est un peigne. On en déduit que, pour \underline{n} quelconque, le minimum de $\varepsilon(H)$ pour $H \in \mathcal{H}^m$ est \underline{n} , atteint ssi H est un peigne.

Notons $\hat{\varepsilon}(\underline{n})$ le maximum de $\varepsilon(H)$ dans \mathcal{H}^m , donc dans \mathcal{H} . On a :

$$\hat{\varepsilon}(\underline{n}) = 1 + \max_{1 \leq \underline{k} \leq \underline{n}-1} \hat{\varepsilon}(\underline{k}) \hat{\varepsilon}(\underline{n}-\underline{k}). \text{ Donnons les premières valeurs de } \hat{\varepsilon}(\underline{n}) :$$

\underline{n}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\hat{\varepsilon}(\underline{n})$	1	2	3	5	7	11	16	26	36	56	81	131	183	287	417	677

Cette suite ne figure pas dans le *Handbook of Integer Sequences* de Sloane (1973). On constate de plus que pour les premières puissances de 2 ($\underline{n} = 1, 2, 4, 8, 16$), la borne supérieure est atteinte par les HBE, et elles seules. On peut conjecturer qu'il en est de même pour toutes les puissances de 2 .

Du point de vue combinatoire, $\varepsilon(H)$ est le nombre des parties libres maximales du graphe (H, \hat{U}_H) , où \hat{U}_H est la fermeture transitive de U_H [cf. Berge (1970), ch.13].

2.7. Préordonnances hiérarchiques.

2.7.1. A une hiérarchie H sur X on associe une préordonnance P_H sur X , c'est-à-dire un préordre sur l'ensemble $\mathcal{P}_2(X)$ des parties de X à deux éléments. Notons, pour $x, y \in X$ (distincts), xHy la classe unique de H contenant x et y et minimale avec cette propriété. Alors, nous posons

$(\{x, y\}, \{x', y'\}) \in P_H \iff xHy \subseteq x'Hy'$. Nous appelons *préordonnance hiérarchique* toute préordonnance pouvant être obtenue ainsi à partir d'une hiérarchie. Une telle préordonnance P n'est pas totale en général : il existe $u, v \in \mathcal{P}_2(X)$ tels que $(u, v) \notin P$ et $(v, u) \notin P$ (nous dirons alors que u et v sont incomparables pour P_H). Elle a cependant toujours une classe maximum : l'ensemble des paires $\{x, y\}$ telles que $xHy = X$, c'est-à-dire que $x \in h$ $y \in h'$, avec $h, h' \in S$ et $h \neq h'$. La préordonnance $(\mathcal{P}_2(X))^2$ est hiérarchique, associée à la hiérarchie minimale H_0 .

Dans la suite, pour $u, v \in \mathcal{P}_2(X)$, nous écrirons $u \leq v$ pour $(u, v) \in P$ et $u \simeq v$ pour $u \leq v$ et $v \leq u$.

Soit $K_X = (X, \mathcal{P}_2(X))$ le graphe simple complet sur X . Un arbre A de K_X est une partie de $\mathcal{P}_2(X)$ telle que le graphe partiel (X, A) est un arbre. Pour $x, y \in X$ (distincts) on note $A(x, y)$ la chaîne unique de A joignant x à y . Un arbre A de K_X est *minimum* pour une préordonnance P sur X ssi, pour tous $x, y \in X$ (distincts) et pour tout $a \in A(x, y)$, on a $a \leq \{x, y\}$. Si P est totale, un tel arbre existe toujours et a de fortes propriétés d'extrémalité [Rosenstiehl (1969) ; cf. Berge (1970)]. Ces propriétés sont conservées par un arbre minimum pour une préordonnance partielle, mais l'existence n'en est plus assurée [Flament et Leclerc (1983)].

Nous donnons plusieurs caractérisations des préordonnances hiérarchiques :

THEOREME 2.3. Pour une préordonnance P sur X, les quatre conditions suivantes sont équivalentes :

- (1) P est une préordonnance hiérarchique.
- (2) $\forall x, y, z \in X$ (distincts), $\{x, y\} \leq \{x, z\}$ ou $\{x, y\} \leq \{y, z\}$.
- (3) $\forall x_1, x_2, \dots, x_k \in X$ (tous distincts), il existe $\underline{j} \in \{1, \dots, k-1\}$ tel que $\{x_1, x_k\} \leq \{x_{\underline{j}}, x_{\underline{j}+1}\}$.
- (4) $\forall u \in \mathcal{P}_2(X)$, il existe un arbre A de K_X minimum pour P tel que $u \in A$.

DEMONSTRATION. (1) \Rightarrow (2). L'ensemble des $h \in H$ tels que $x \in h$ étant totalement ordonné par inclusion, on a $xHy \subseteq xHz$, donc $\{x, y\} \leq \{x, z\}$, ou $xHz \subseteq xHy$ d'où $yHz = xHz$ et $\{y, z\} \simeq \{x, z\}$.

(2) \Rightarrow (3). D'après (2), (3) est vraie pour $\underline{k} = 3$. Supposons (3) vraie jusqu'à $\underline{k} - 1$ et montrons la pour \underline{k} . On a, d'après (2), $\{x_1, x_{\underline{k}}\} \leq \{x_1, x_{\underline{k}-1}\}$ ou $\{x_1, x_{\underline{k}}\} \leq \{x_{\underline{k}-1}, x_{\underline{k}}\}$. Dans le second cas (3) est vérifiée. Dans le premier, il existe par hypothèse $\underline{j} \in \{1, \dots, \underline{k}-2\}$ tel que $\{x_1, x_{\underline{k}-1}\} \leq \{x_{\underline{j}}, x_{\underline{j}+1}\}$ et (3) est encore vraie.

(3) \Rightarrow (4). Soit un arbre A_0 de K_X et $\{x, y\} \notin A_0$. Si (3) est vraie, il existe $a \in A_0(x, y)$ tel que $\{x, y\} \leq a$. Dans le cas où $a \not\leq \{x, y\}$, on remplace A_0 par l'arbre $A_1 = (A_0 - \{a\}) \cup \{x, y\}$. Par itération, on arrive à un arbre A tel que que cet échange n'est possible pour aucun $\{x, y\} \notin A$. Montrons que les chaînes de A ont des éléments maximums pour P. Sinon il existe une chaîne $A(x, y)$ ayant exactement deux éléments maximaux a, a' incomparables. Mais on doit avoir, par exemple, $(x, y) \simeq a$. Alors la chaîne $(A - a') \cup \{x, y\}$ ne vérifie pas (3). Donc toute chaîne $A(x, y)$ a au moins un élément a maximum pour P et l'on a $a \simeq \{x, y\}$. Alors A est bien un arbre minimum et il en est de même de l'arbre $(A - \{a\}) \cup \{x, y\}$.

(4) \Rightarrow (1). Si (4) est vrai et si a est maximal pour P dans un arbre minimum A, on a $a \in A(x, y) \Rightarrow a \simeq \{x, y\}$, car sinon $\{x, y\}$ serait maximum unique dans le cycle $\{x, y\} \cup A\{x, y\}$, et ne pourrait être élément d'aucun arbre minimum [cf. Flament et Leclerc]. Alors, il ne peut y avoir deux $a, a' \in A$ maximums pour P et incomparables, et A contient un ensemble A^M non vide d'éléments maximums pour P. Soit $a \in A^M$. Construisons comme suit la partition S de X : S est la partition en classes connexes du graphe $(X, A - A^M)$. Pour $x, y \in X$ (distincts), on a $\{x, y\} \simeq a$ si il existe $a' \in A(x, y) \cap A^M$, c'est-à-dire si x et y sont dans des classes distinctes de S et $\{x, y\} \leq a$ sinon. Toute hiérarchie H telle que $S_{H, X} = S$ a donc bien une préordonnance P_H de même classe maximum que P.

On considère ensuite les restrictions de P et de A à chaque classe $h \in S$ telle que $\underline{n}_h \geq 2$ et l'on construit de même S_h et ainsi de suite jusqu'à l'obtention de toute la hiérarchie H telle que $P_H = P$. \square

La condition (2) est analogue à l'inégalité triangulaire des ultramétriques [cf. Barthélemy, Leclerc et Monjardet (1984,b)]. Celles-ci se caractérisent aussi, parmi les dissimilarités, par la condition (4) [Leclerc (1981)]. Ceci est dû à l'équivalence bien connue des ultramétriques et des hiérarchies "indicées", c'est-à-dire munies d'une valuation dans \mathbb{R}^+ préservant l'ordre.

2.7.2. Le nombre $\alpha(H)$ d'arbres minimums pour P_H possède des propriétés intéressantes. Pour calculer ce nombre, nous utilisons un résultat de Moon (1967) :

Soit S une partition de X et, pour chaque $h \in S$, un arbre A_h de K_h . Alors le nombre d'arbres distincts de K_X contenant tous les A_h , $h \in S$ est égal à $\underline{n}^{\underline{s}-2} \prod_{h \in S} \underline{n}_h$.

THEOREME 2.4. Le nombre des arbres de K_X minimums pour la préordonnance P_H est $\alpha(H) = \frac{1}{\underline{n}} \prod_{i \in I} \frac{t_i}{\underline{n}_i}$.

DEMONSTRATION. Montrons le par récurrence sur \underline{n} . Le résultat est vrai, avec $\alpha(H) = 1$, pour $\underline{n} = 1, 2$.

Un arbre A de K_X est minimum ssi il contient un arbre minimum de chacune des restrictions de P_H aux $h \in S$ (nous admettons ici ce fait qui repose sur les propriétés générales des arbres minimums). Pour $h \in S$, posons $\alpha_h = \alpha(H_h)$ et $I_h = I_{H_h}$. Avec le résultat de Moon et l'hypothèse de récurrence, on a :

$$\begin{aligned} \alpha(H) &= \underline{n}^{\underline{s}-2} \prod_{h \in S} \underline{n}_h \prod_{h \in S} \alpha_h = \underline{n}^{\underline{s}-2} \prod_{h \in S} \underline{n}_h \frac{1}{\underline{n}_h} \prod_{i \in I} \frac{t_i}{\underline{n}_i} \\ &= \frac{1}{\underline{n}} \underline{n}^{\underline{t}} \prod_{i \in I - \{X\}} \frac{t_i}{\underline{n}_i} = \frac{1}{\underline{n}} \prod_{i \in I} \frac{t_i}{\underline{n}_i} . \square \end{aligned}$$

On en déduit que l'application $\alpha : \mathcal{H} \rightarrow \mathbb{N}$ est strictement décroissante. On a en effet, pour tous $H \in \mathcal{H}$ et $i \in I_H$, $\frac{\alpha(H^i)}{\alpha(H)} = \left(\frac{\underline{n}_i}{\underline{n}_{p(i)}} \right)^{t_i} < 1$.

Si H est une H -partition, $\alpha(H) = \frac{1}{\underline{n}^{s-2}} \prod_{h \in S} \frac{\underline{n}_h^{-1}}{\underline{n}_h}$. En particulier $\alpha(H_0) = \frac{1}{\underline{n}^{n-2}}$: c'est le nombre des arbres de K_X , car tous sont minimum pour la préordonnance $(\mathcal{P}_2(X))^2$.

Si H est binaire, $\alpha(H) = \frac{1}{\underline{n}} \prod_{i \in I} \underline{n}_i$. Dans ce cas, on peut préciser les valeurs extrêmes de $\alpha(H)$.

PROPOSITION 2.5. Si H est une hiérarchie binaire, on a $4^{\frac{\underline{n}-1}{2}} / \underline{n}^2 \leq \alpha(H) \leq (\underline{n}-1)!$ La borne inférieure est atteinte ssi H est équilibrée et la borne supérieure ssi H est un peigne.

DEMONSTRATION. Nous effectuons encore une récurrence sur \underline{n} . On vérifie le résultat facilement pour $\underline{n} = 1, 2, 3, 4$. Posons $S = \{h_1, h_2\}$, $\underline{n}_1 = \frac{\underline{n}}{h_1}$, $\underline{n}_2 = \frac{\underline{n}}{h_2}$, H_1, H_2 , α_1, α_2 définis de même.

On a $\alpha(H) = \frac{1}{\underline{n}_1 \underline{n}_2} \alpha_1 \alpha_2$ et, avec l'hypothèse de récurrence, $4^{\frac{\underline{n}-2}{2}} / \underline{n}_1 \underline{n}_2 \leq \alpha(H) \leq (\underline{n}_1)! (\underline{n}_2)!$. Le terme de gauche est minimum et égal à $4^{\frac{\underline{n}-1}{2}} / \underline{n}^2$ pour $\underline{n}_1 = \underline{n}_2 = \frac{\underline{n}}{2}$. Les hiérarchies H_1 et H_2 sont alors équilibrées et H aussi ; le terme de droite est maximum (en posant, sans perte de généralité, $\underline{n}_1 \geq \underline{n}_2$) pour $\underline{n}_1 = \underline{n} - 1$ et $\underline{n}_2 = 1$, et on a alors $\alpha(H) = (\underline{n} - 1)!$, H étant un peigne. \square

Puisque α est décroissant la borne inférieure sur \mathcal{H} est la même que celle, obtenue ci-dessus, sur \mathcal{H}^m .

3. DEUX CODAGES DES HIERARCHIES.

3.1. Parenthésages commutatifs.

On fait correspondre à tout $h \in H$ un mot M_h dont les lettres sont prises dans l'alphabet $X \cup \{(\,,)\}$, avec les règles :

- (1) $h = \{x\} \in F \Rightarrow M_h = x$
- (2) $i \in I, S_i = \{h_1, \dots, h_k\}$, avec $\underline{k} = \underline{s}_i \Rightarrow M_i = (M_{h_1} M_{h_2} \dots M_{h_k})$

Alors le mot $M = M_X$ représente la hiérarchie H . En fait, prenant le mot M' tel que $M = (M')$, il est connu que la représentation de H par M' est bi-univoque quand on ordonne l'arborescence G_H . Pour la représentation des hiérarchies, les mots M_{h_1}, \dots, M_{h_k} peuvent commuter à l'intérieur des parenthèses correspondant à i , et de nombreux mots représentent la même hiérarchie.

Ils sont tous de longueur $\underline{n} + 2 \gamma(H)$. De telles représentations de la hiérarchie de la figure 1 sont, par exemple, ((af)d(bg))(c(eh)) ou ((he)c)((af)(gb)d).

3.2. Représentation des hiérarchies par des partitions.

Comtet (1970,b) a établi l'existence d'un codage (i.e. une correspondance bi-univoque n'impliquant apparemment pas de lien structurel fort) entre les hiérarchies et un ensemble de partitions défini classiquement.

Une *partition associée* d'un ensemble est une partition sans classe de cardinal 1 (singleton, hapax). Les ensembles que nous allons considérer sont l'ensemble $\mathcal{H}(\underline{k})$ des hiérarchies de niveau \underline{k} ($0 \leq \underline{k} \leq \underline{n}-2$) d'une part, et l'ensemble $\mathcal{C}_{\underline{n}+\underline{k}, \underline{k}+1}$ des partitions associées de l'intervalle entier $[1, \underline{n}+\underline{k}] = \{1, 2, \dots, \underline{n}+\underline{k}\}$ en $\underline{k}+1$ classes, d'autre part.

Tout d'abord, on construira un numérotage bijectif $\nu : H - \{X\} \rightarrow [1, \underline{n}+\underline{k}]$. On posera alors, pour $i \in I$, $\nu(S_i) = \{\nu(h) / h \in S_i\}$, puis $D(H) = \{\nu(S_i) / i \in I\}$. La partition cherchée sera $D(H)$.

Le numérotage ν se fait comme suit : on se donne d'abord un ordre d'introduction $x_1, x_2, \dots, x_{\underline{n}}$ des éléments de X ; on aura $\underline{p} < \underline{q} \Rightarrow \nu(x_{\underline{p}}) < \nu(x_{\underline{q}})$. Pour $\underline{n}=2$, il n'existe qu'une hiérarchie H sur X , et qu'une partition associée $D(H) = \{1, 2\}$. Le numérotage ν est unique : $\nu(x_1) = 1$, $\nu(x_2) = 2$.

On construit ν par récurrence sur \underline{n} : on pose $x = x_{\underline{n}}$ et on suppose qu'un bon numérotage ν_1 a été construit sur la hiérarchie $H_1 = \{h - x / h \in H\} \in \mathcal{H}_{X-x}$. Deux cas sont à considérer, selon la valeur de $\underline{s}_{H, \underline{p}}(x)$:

Cas 1. $\underline{s}_{H, \underline{p}}(x) \geq 3$, donc $\gamma(H_1) = \gamma(H) = \underline{k}$. Seul le numéro $\underline{n} + \underline{k}$ est à attribuer.

On pose $\nu(x) = \underline{n} + \underline{k}$, et, pour $h \in H - \{x\}$:

- si $x \notin h$, $\nu(h) = \nu_1(h)$.
- si $x \in h$, $\nu(h) = \nu_1(h-x)$.

Cas 2. $\underline{s}_{H, \underline{p}}(x) = 2$, donc $\gamma(H_1) = \gamma(H) - 1 = \underline{k} - 1$, et les numéros $\underline{n} + \underline{k} - 1$ et $\underline{n} + \underline{k}$ sont à attribuer. Il y a deux sous-cas :

(a) $\underline{p}_H(x) \neq X$. On pose $\nu(x) = \underline{n} + \underline{k}$, $\nu(\underline{p}_H(x)) = \underline{n} + \underline{k} - 1$, et, pour $h \in H - \{x, \underline{p}_H(x)\}$, on fait comme dans le cas 1.

(b) $p_H(x) = X$, c'est-à-dire $H_1 = H - \{X, \{x\}\}$. On pose $v(X - \{x\}) = \underline{n} + \underline{k} - 1$, $v(x) = \underline{n} + \underline{k}$, et, pour tout $h \in H_1$, $v(h) = v_1(h)$.

Le numérotage v , donc la partition $D(H)$, étant bien défini et unique pour X ordonné, avec $\underline{n} = 2$, l'est encore pour X ordonné, \underline{n} quelconque, $H \in \mathcal{H}_X$.

Pour la reconstruction de H à partir de $D(H)$, on remarque qu'à chacun des cas précédents correspond un type différent de classe $D(H, x)$ de $D(H)$ contenant $x_{\underline{n}} = x$.

Si $|D(H, x)| \geq 3$, le passage de v_1 à v s'est effectué comme dans le cas 1.

Si $D(H, x) = \{\underline{n} + \underline{k}, n'\}$, avec $\underline{n}' \in [1, \underline{n} + \underline{k} - 2]$, il s'est effectué comme dans le cas 2a, et comme dans le cas 2b, si $D(H, x) = \{\underline{n} + \underline{k} - 1, \underline{n} + \underline{k}\}$.

Pour $D(H)$ connu, $D(H_1)$ est unique, et, en itérant on construit une suite de partitions $D(\{x_1, \dots, x_{\underline{n}'}\})$, pour $\underline{n}' = \underline{n}, \underline{n}-1, \dots, 2$. H se reconstruit en reprenant cette suite dans le sens de \underline{n}' croissant.

La figure 7 montre la hiérarchie de la figure 1 réétiquetée, la partition $v(H)$ étant alors 1 9/2 11/3 6/4 12/5 8 10/7 13.

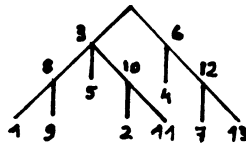


Figure 7

La hiérarchie H est représentée par un mot de même longueur $\underline{n} + 2\gamma(H)$ que le parenthésage du paragraphe 3.1., mais à lettres prises dans un alphabet à $\underline{n} + \gamma(H) + 1$ éléments. Avec les conventions supplémentaires : écriture des classes de D dans l'ordre de leurs premiers éléments, omission de ceux-ci, on réduit le mot à la longueur $\underline{n} + \gamma(H)$. Pour l'exemple ci-dessus, ce mot est 9/11/6/12/8 10/13.

Un problème intéressant est de trouver d'autres codages de $\mathcal{H}(k)$ dans $\mathcal{C}_{\underline{n}+\underline{k}, \underline{k}-1}$, possédant de bonnes propriétés.

4. FORMES ET PARTAGES.

4.1. Le filtrant des formes sur \underline{n} éléments. Soient X et X' deux ensembles de même cardinal \underline{n} , et une hiérarchie $H \in \mathcal{H}_X$. Si φ est une application de X dans X' , on pose $\varphi(H) = \{\varphi(h)/h \in H\}$. Si φ est bijective, $\varphi(H)$ est une hiérarchie sur X' .

Définition : deux hiérarchies H sur X , H' sur X' sont dites *isomorphes* si il existe une bijection $\varphi : X \rightarrow X'$ telle que $H' = \varphi(H)$. On dira comme Harding (1971) que H_1 et H_2 ont la même *forme* (shape) $T(H) = T(H')$. On écrira aussi $H' \in T(H)$.

Une forme T est donc une classe d'équivalence de hiérarchies. On note $\mathcal{T}_{\underline{n}}$ l'ensemble des formes à \underline{n} feuilles. Le graphe G_T associé à une forme $T \in \mathcal{T}_{\underline{n}}$ est une arborescence *non étiquetée*, à \underline{n} sommets pendants ne comprenant pas la racine, sans sommets de demi-degré extérieur égal à 1. On lui associe, comme à la fin du paragraphe 2.2., un arbre planté réduit non étiqueté.

Ordre sur les formes. On note encore \subseteq la relation sur les formes induite par l'ordre d'inclusion des hiérarchies :

$$(\forall T, T' \in \mathcal{T}_{\underline{n}}) [T \subseteq T' \Leftrightarrow (\exists H, H' \in \mathcal{H}_X) H \in T, H' \in T', H \subseteq H']$$

PROPOSITION 4.1. La relation \subseteq sur $\mathcal{T}_{\underline{n}}$ et l'application $H \rightarrow T(H)$ vérifient :

- (1) \subseteq est un ordre filtrant inférieurement sur $\mathcal{T}_{\underline{n}}$.
- (2) $H \prec H' \Rightarrow T(H) \prec T(H')$.
- (3) $(\mathcal{T}_{\underline{n}}, \subseteq)$ est un ensemble ordonné semi-modulaire inférieurement.
- (4) La graduation $\gamma(T)$ dans $\mathcal{T}_{\underline{n}}$ est donnée par : $H \in T \Rightarrow \gamma(H) = \gamma(T)$.

DEMONSTRATION. La relation \subseteq sur $\mathcal{T}_{\underline{n}}$ est transitive : soient $T, T', T'' \in \mathcal{T}_{\underline{n}}$ tels que $T \subseteq T' \subseteq T''$. Il existe $H, H'_1, H'_2, H'' \in \mathcal{H}_X$ et φ permutation de X avec $H \in T, H'_1 \in T', H'_2 \in T', H'' \in T'', h'_1 = \varphi(H'_2), H \subseteq H'_1, H'_2 \subseteq H''$. D'où $H \subseteq H'_1 = \varphi(H'_2) \subseteq \varphi(H'') \in T''$ et $T \subseteq T''$.

Toutes les hiérarchies H de même forme ont le même nombre de classes intérieures $\gamma(H) + 1$, et si T est la forme de H , on peut poser $\gamma(T) = \gamma(H)$. Par suite, avec $H \prec H'$ et $H' \in T' \in \mathcal{T}$, on a $\gamma(T') = \gamma(T) + 1$. On en déduit (2), (4) et l'antisymétrie de \subseteq sur \mathcal{T} , d'où (1).

Pour la semi-modularité, soient $T, T'_1, T'_2 \in \mathcal{T}$ avec $T'_1 \prec T$ et $T'_2 \prec T$. Il existe alors $H_1, H_2, H'_1, H'_2 \in \mathcal{H}$ tels que $H_1, H_2 \in T, H'_1, H'_2 \in T', H'_1 \prec H_1,$

$H'_2 \prec H_2$. Soit π la permutation de X telle que $H_1 = \pi(H_2)$. On a $\pi(H'_2) \prec \pi(H_2) = H_1$, et il existe $H'' \in \mathcal{H}$ tel que $H'_1 \succ H''$ et $\pi(H'_2) \succ H''$. D'où $T'_1 \succ T(H'')$ et $T'_2 \succ T(H'')$. \square

La figure 8 a,b donne les diagrammes de Hasse des ensembles ordonnés $(\mathcal{T}_{\underline{n}}, \subseteq)$, pour $\underline{n} = 4$ et $\underline{n} = 5$. Les traits renforcés mettent en évidence le fait que pour $\underline{n} = 5$ (et $\underline{n} > 5$), $\mathcal{T}_{\underline{n}}$ n'est pas un treillis.

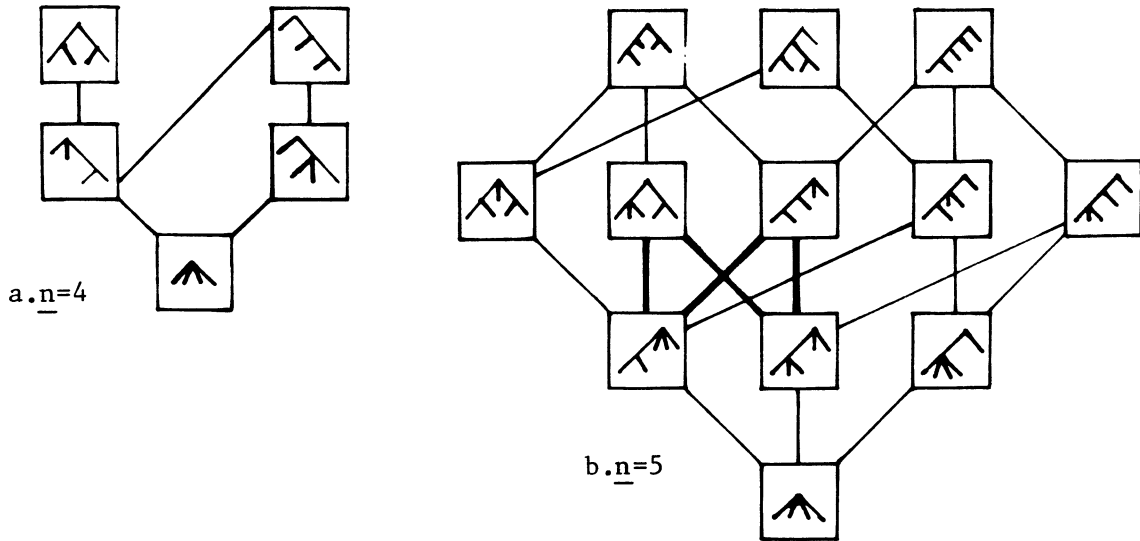


Figure 8

4.2. Relation avec l'ensemble ordonné des partages de l'entier $\underline{n} - 1$. Un *partage* d'un entier $\underline{n} \geq 1$ est une écriture de \underline{n} comme somme non croissante d'entiers positifs (les *parts*). Nous préférons, suivant Guilbaud (1966), dire *partage* plutôt que "partition d'entier" d'autant plus que nous avons ici à considérer des partitions (d'ensembles).

Ordre sur les partages de \underline{n} . Soient $\underline{n} = \underline{n}_1 + \underline{n}_2 + \dots + \underline{n}_k = (\underline{n})$ et

$\underline{n} = \underline{n}'_1 + \underline{n}'_2 + \dots + \underline{n}'_k = (\underline{n})'$ deux partages de \underline{n} . On définit un ordre \leq sur l'ensemble $\mathcal{P}_{\underline{n}}$ des partages de \underline{n} comme suit : $(\underline{n}) \leq (\underline{n})'$ ssi il existe une

partition $\{D_1, D_2, \dots, D_k\}$ de $\{1, \dots, k\}$ telle que $\underline{n}_p = \sum_{q \in D_p} \underline{n}'_q$, pour

$p = 1, \dots, k$. Cet ordre, qui est modulaire [Barthélemy (1978)], a pour relation de couverture \prec : $(\underline{n}) \prec (\underline{n})'$ ssi $(\underline{n}) \leq (\underline{n})'$ et (\underline{n}) a un élément exactement de moins que $(\underline{n})'$.

Les nombres \underline{t}_i , $i \in I_H$, rangés par ordre non croissant, associent à toute hiérarchie $H \in \mathcal{H}$ un partage de l'entier $\underline{n} - 1$. A deux hiérarchies de même forme T correspondent le même partage, noté $(\underline{n} - 1)_T$.

PROPOSITION 4.2. L'application $T \mapsto (\underline{n} - 1)_T$ est un épimorphisme d'ordre de $\mathcal{C}_{\underline{n}, \subseteq}$ sur $(\mathcal{P}_{\underline{n}-1}, \leq)$, préservant la couverture.

DEMONSTRATION. Epimorphisme d'ordre signifie que l'application est surjective et que $T \subseteq T' \Rightarrow (\underline{n} - 1)_T \leq (\underline{n} - 1)_{T'}$. Seule la surjectivité n'est pas une conséquence immédiate de la proposition 1.1. Nous allons montrer que si un partage $(\underline{n} - 1)$ est image d'une forme $T \in \mathcal{C}$, soit $(\underline{n} - 1) = (\underline{n} - 1)_T$, tout partage $(\underline{n} - 1)'$ tel que $(\underline{n} - 1)' \succ (\underline{n} - 1)$ est image d'une forme T' . Comme le minorant universel à une seule part de l'ordre des partages est image de $T(H_0)$, ceci établira la surjectivité.

Soit donc $(\underline{n} - 1) : \underline{n} - 1 = \underline{t}_1 + \underline{t}_2 + \dots + \underline{t}_k$, et $(\underline{n} - 1)' : \underline{n} - 1 = \underline{t}'_1 + \underline{t}'_2 + \dots + \underline{t}'_{k+1}$, les \underline{t}_p , $p = 1, \dots, k$, étant égaux aux \underline{t}'_q , $q = 1, \dots, k + 1$, sauf, par exemple, $\underline{t}_1 = \underline{t}'_{q_1} + \underline{t}'_{q_2}$. Il existe une hiérarchie $h \in T$ telle que les \underline{t}_i , $i \in I_H$ réordonnés constituent le partage $(\underline{n} - 1)$ et $h \in I_H$ tel que $\underline{t}_h = \underline{t}_1$, donc $\underline{s}_h = \underline{t}_1 + 1$. Posons $S_h = \{h_1, \dots, h_{\underline{t}_1+1}\}$, avec une indiciation arbitraire des éléments de S_h et $h' = h_1 \cup h_2 \cup \dots \cup h_{\underline{t}'_{q_1}+1}$. Alors $H \cup \{h'\}$ est une hiérarchie dont le partage de $\underline{n} - 1$ associé est $(\underline{n} - 1)'$ \square

En fait, les \underline{t}_i représentent le nombre des classes intérieures de toute hiérarchie binaire contenant H qui sont incluses dans la classe $i \in I_H$:

PROPOSITION 4.3. Soient $H \in \mathcal{H}$ et $H' \in \mathcal{H}^m$, avec $H \subseteq H'$. Posons :

$$(\forall i \in I_H) \quad I'(i) = \{i' \in I_{H'} / [i' \subseteq i] \text{ et } [i' \subseteq j, j \in I_H \Rightarrow i \subseteq j]\}.$$

Alors, les $I'(i)$ forment une partition de $I_{H'}$, et, pour tout $i \in I_H$, on a $|I'(i)| = \underline{t}_i$.

DEMONSTRATION. Que les $I'(i)$ forment une partition de $I_{H'}$, découle de leur définition et de la propriété (H4) des hiérarchies. Montrons que si $|I'(i)| = \underline{t}_i$ est vrai pour $\gamma(H) = \underline{k}$, c'est encore vrai pour $\gamma(H) = \underline{k} - 1$. Comme c'est vrai pour H binaire, i.e. $\gamma(H) = \underline{n} - 2$, puisqu'alors $H' = H$ et, pour $i \in I_H$, $\underline{t}_i = 1$, la démonstration sera complète.

Il existe une hiérarchie $H_1 \in \mathcal{H}$, telle que $H \prec H_1 \subseteq H'$; soit $i_1 \in H_1$ avec $H = H_1^{i_1}$. Notant $I'_1(i)$, \underline{t}_{1i} les équivalents de $I'(i)$, \underline{t}_i obtenus en substituant H_1 à H , on a par hypothèse, pour $i \in I_{H_1}$, $\underline{t}_{1i} = |I'_1(i)|$. Soit

$p(i_1)$ le père de i_1 dans H_1 . Si $i \in I_H - \{p(i_1)\}$ on a $I'(i) = I'_1(i)$ et $t_i = t_{1i}$, donc $t_i = |I'(i)|$. Comme $t_{p(i_1)} = t_{1i} + t_{1p(i_1)}$ et $I'(p(i_1)) + I'_1(i_1)$, on a aussi $t_{p(i_1)} = |I'(p(i_1))|$ \square

5. DENOMBREMENTS.

Nous nous appuyerons dans ce paragraphe sur deux ouvrages de base :

A handbook of integer sequences de Sloane (1973). [HIS 1217] par exemple sera mis pour "séquence 1217 du manuel de Sloane".

Analyse combinatoire de Comtet (1970,a). [AC 2-57] renverra au tome 2, p.57, de cet ouvrage fondamental.

La littérature sur les dénombrements d'arbres de types divers est importante. Nous nous limiterons aux résultats concernant les hiérarchies. Certains d'entre eux figurent dans Dobson (1974), Murtagh (1984). Des arbres vérifiant des conditions d'étiquetage moins strictes que pour les hiérarchies ont été dénombrés, sous le nom d'"arbres phylogénétiques" par Foulds et Robinson (1980,1982).

5.1. Dénombrement des hiérarchies.

Le nombre de *Stirling associé* (ou 2-associé) de seconde espèce $\underline{S}_2(\underline{n}, \underline{k})$ est le nombre de partitions d'un ensemble à \underline{n} éléments en \underline{k} classes, toutes de cardinal au moins égal à 2 [AC 2-57].

PROPOSITION 5.1. [Comtet (1970,b)]. Pour \underline{k} entier compris entre 0 et $\underline{n} - 2$, le nombre $\underline{H}(\underline{n}, \underline{k})$ des hiérarchies de niveau \underline{k} sur un ensemble à \underline{n} éléments est $\underline{S}_2(\underline{n} + \underline{k}, \underline{k} + 1)$.

Cette égalité découle de la correspondance bijective du paragraphe 3.2. qui est liée à une récurrence triangulaire sur les $\underline{H}(\underline{n}, \underline{k})$.

Pour construire un élément de $\mathcal{H}_X(\underline{k})$, on peut choisir un $x \in X$, et :

- soit prendre un élément $H \in \mathcal{H}_{X-\{x\}}(\underline{k})$ et "accrocher" la feuille $\{x\}$

à l'une quelconque des $\underline{k} + 1$ classes intérieures de H .

- soit prendre $H \in \mathcal{H}_{X-\{x\}}(\underline{k} - 1)$ et accrocher la feuille $\{x\}$ au-dessus de

l'une des $\underline{n} + \underline{k} - 1$ classes $h \in H$, créant la nouvelle classe intérieure $h \cup \{x\}$.

La récurrence triangulaire cherchée est donc :

$$\underline{H}(\underline{n}, \underline{k}) = (\underline{k} + 1) \underline{H}(\underline{n} - 1, \underline{k}) + (\underline{n} + \underline{k} - 1) \underline{H}(\underline{n} - 1, \underline{k} - 1).$$

Avec $\underline{p} = \underline{n} + \underline{k}$, $\underline{q} = \underline{k} + 1$, c'est bien celle des nombres de Stirling associés :

$$\underline{S}_2(\underline{p}, \underline{q}) = \underline{q} \underline{S}_2(\underline{p} - 1, \underline{q}) + (\underline{p} - 1) \underline{S}_2(\underline{p} - 1, \underline{q} - 1).$$

On constate dans la table suivante que la proposition est vraie pour les petites valeurs de \underline{n} et \underline{k} .

\underline{n}	\underline{k}						$\underline{H}(\underline{n})$
	0	1	2	3	4	5	
2	1						1
3	1	3					4
4	1	10	15				26
5	1	25	105	105			236
6	1	56	490	1260	945		2752
7	1	119	1918	9450	17325	10395	39208

La somme en lignes $\underline{H}(\underline{n}) = \sum_{\underline{k}=0}^{\underline{n}-2} \underline{H}(\underline{n}, \underline{k}) = |\mathcal{K}_X|$ donne bien les *nombre*s du quatrième problème de Schröder [HIS 1465, AC 2-60, Schröder (1870)], dont Comtet (1970b) donne quelques propriétés.

La diagonale $\underline{H}(\underline{n}, \underline{n}-2)$ donne les nombres de hiérarchies binaires. Ce sont, comme le montre Harding (1971), les produits des $\underline{n} - 1$ premiers entiers naturels impairs.

$$\underline{H}(\underline{n}, \underline{n}-2) = 1.3.5.7\dots(2\underline{n} - 3) = \frac{(2\underline{n}-3)!}{2^{\underline{n}-2}(\underline{n}-2)!} = \frac{(2\underline{n}-2)!}{2^{\underline{n}-1}(\underline{n}-1)!}$$

Le graphe d'échanges des hiérarchies binaires, défini au paragraphe 2.5., a $\underline{H}(\underline{n}, \underline{n}-2)$ sommets et est régulier de degré $2\underline{n}-4$. Il a donc

$$|\mathcal{V}| = \frac{(2\underline{n}-3)!}{2^{\underline{n}-2}(\underline{n}-3)!} = 1.3.5\dots(2\underline{n}-3)(\underline{n}-2) \text{ arêtes.}$$

5.2. Le problème du dénombrement des formes.

Par des méthodes seulement partiellement généralisables, on trouve, pour les petites valeurs de \underline{n} , les nombres $\underline{T}(\underline{n}, \underline{k})$ de formes de niveau \underline{k} dans $\mathcal{C}_{\underline{n}}$ suivants :

\underline{n}	\underline{k}	0	1	2	3	4	5	6	7	8	$\underline{T}(\underline{n})$
2		1									1
3		1	1								2
4		1	2	2							5
5		1	3	5	3						12
6		1	4	10	12	6					33
7		1	5	16	29	28	11				90
8		1	6	24	57	84	66	23			261
9		1	7	33	99		...		46		...
10		1	8	44	157			...		98	...

On découvre dans HIS que les suites $\underline{T}(\underline{n}, \underline{k})$, \underline{k} fixé ont été, au moins pour $\underline{k} = 2$ et $\underline{k} = 3$, étudiées par Riordan ["Communication personnelle" ; HIS 531, 1171] ; la suite $\underline{T}(\underline{n}) = \sum_{\underline{k}=0}^{\underline{n}-2} \underline{T}(\underline{n}, \underline{k}) = |\mathcal{C}_{\underline{n}}|$, a été signalée par Cayley (1857) [HIS 558].

La diagonale $\underline{w}_{\underline{n}} = \underline{T}(\underline{n}, \underline{n}-2)$ des nombres de formes binaires vérifie la récurrence :

$$\begin{aligned} \underline{w}_{\underline{n}} &= \underline{w}_1 \underline{w}_{\underline{n}-1} + \underline{w}_2 \underline{w}_{\underline{n}-2} + \dots + \underline{w}_{\underline{m}} \underline{w}_{\underline{m}+1} & \text{si } \underline{n} = 2\underline{m} + 1 \\ &= \underline{w}_1 \underline{w}_{\underline{n}-1} + \underline{w}_2 \underline{w}_{\underline{n}-2} + \dots + \frac{1}{2} \underline{w}_{\underline{m}} \left(\underline{w}_{\underline{m}} + 1 \right) & \text{si } \underline{n} = 2\underline{m} \end{aligned}$$

Ce sont les nombres, bien connus, de Wedderburn-Etherington [Wedderburn (1922), Etherington (1939), HIS 298, AC 1-67]. Leur récurrence rappelle celle des nombres de Catalan $\underline{u}_{\underline{n}}$ célèbres par leurs applications variées [HIS 557, AC 1-64, Gardner (1976) et Gould (1971) pour une revue de leurs domaines d'application], qui dénombrent (entre autres) les hiérarchies binaires ordonnées (au sens du paragraphe 1.22), et vérifient :

$$\underline{u}_{\underline{n}} = \sum_{\underline{m}=1}^{\underline{n}-1} \underline{u}_{\underline{m}} \underline{u}_{\underline{n}-\underline{m}} .$$

Améliorant un travail de Otter (1948), Harding (1971) donne une formule asymptotique pour les $w_{\underline{n}}$. Il détermine ensuite les probabilités des formes, puis des hiérarchies, à partir d'un modèle simple de bifurcations aléatoires. Cet important travail ne concerne que les formes et les hiérarchies binaires.

5.3. Hiérarchies de forme ou de partage donnés.

Pour obtenir le nombre $\underline{H}(T)$ des éléments de \mathcal{H}_X dont la forme est $T \in \mathcal{T}_{\underline{n}}$, cherchons quelles permutations des étiquettes des feuilles de $H \in T$ laissent H invariante. Soit $I(\underline{k})$ la partie de I_H définie, pour $\underline{k} = 2, \dots, \underline{n}$, par :

$$I(\underline{k}) = \{i \in I_H / \underline{k} \text{ exactement des } H_h, h \in S_i, \text{ ont la même forme}\}$$

Les $I(\underline{k})$, $\underline{k} = 2, \dots, \underline{n}$, ne sont pas nécessairement disjoints. Si $i \in I(\underline{k})$, on peut permuer entre elles les \underline{k} sous-arborescences de même forme sans changer la forme $T(H)$. Posant $\text{sym}(T, \underline{k}) = |I(\underline{k})|$ (ce nombre ne dépend que de T), on a finalement :

$$\underline{H}(T) = \frac{\underline{n}!}{\prod_{\underline{k}=2}^{\underline{n}} (\underline{k}!)^{\text{sym}(T, \underline{k})}}$$

Exemples :

1) Pour $T = T(H_0)$, on a $\text{sym}(T, \underline{n}) = 1$, et $\text{sym}(T, \underline{k}) = 0$, pour $\underline{k} \neq \underline{n}$. On trouve bien $\underline{H}(T(H_0)) = \underline{n}! / \underline{n}! = 1$.

2) Pour la hiérarchie H_1 de la figure 5, on a $\text{sym}(T, 3) = 1$, et $\text{sym}(T, 2) = 3$. On trouve $\underline{H}(F) = 720 / 48 = 15$, aisément vérifiable.

Pour les formes binaires, on a $1 \leq \text{sym}(T, 2) \leq \underline{n} - 1$, la borne inférieure n'étant atteinte que pour la forme du peigne, et la borne supérieure que pour la forme binaire équilibrée. On peut dire que, *grosso modo*, $\underline{H}(T)$ est d'autant plus élevé que la forme T est déséquilibrée. Ainsi le souhait fréquent en taxinomie d'obtenir des hiérarchies assez équilibrées correspond à une demande de hiérarchies "rares".

Pour les formes quelconques, on a : $1 \leq \sum_{\underline{k}=2}^{\underline{n}} \text{sym}(T, \underline{k}) \leq \underline{n}$, et

$$\sum_{\underline{k}=2}^{\underline{n}} \underline{k} \cdot \text{sym}(T, \underline{k}) \leq \sum_{i \in I} \underline{s}_i = \underline{n} + \gamma(H).$$

Considérons aussi un partage $(\underline{n}-1) = \underline{t}_1 + \underline{t}_2 + \dots + \underline{t}_{\underline{k}} = \underline{n}-1$ de $\underline{n}-1$ en \underline{k} parts, $1 \leq \underline{k} \leq \underline{n}-1$, et cherchons le nombre des hiérarchies H dont les \underline{t}_i , $i \in I$, sont les éléments de $(\underline{n}-1)$.

La donnée des \underline{t}_i entraîne celle des $\underline{s}_i = \underline{t}_i + 1$, donc du type de la partition construite en 3.2., et le calcul du nombre de partitions de type donné, est classique. Soit, pour $q = 1, \dots, \underline{n} - 1$, p_q le nombre de parts de $(\underline{n} - 1)$ égales à q . Alors :

$$\underline{H}((\underline{n} - 1)) = \frac{(\underline{n} + \underline{k})!}{(2!)^{p_1} (3!)^{p_2} \dots (n!)^{p_{\underline{n} - 1}} p_1! p_2! \dots p_{\underline{n} - 1}!}$$

Pour $(\underline{n} - 1) = 1 + 1 + \dots + 1$, c'est-à-dire $p_1 = \underline{n} - 1$ et $p_2 = p_3 = \dots = p_{\underline{n} - 1} = 0$; $\underline{k} = \underline{n} - 1$, on retrouve les nombres doubles factoriels $\underline{H}((\underline{t})) = (2\underline{n} - 1)! / 2^{\underline{n} - 1} (\underline{n} - 1)!$

Comme celui du dénombrement des formes, le problème du dénombrement des formes de partage donné, paraît requérir des techniques de dénombrement plus élaborées.

BIBLIOGRAPHIE

- BANDELT H.J., BARTHELEMY J.P., "Medians in median graphs", Discrete Applied Math., 8 (1984), 131-142.
- BANDELT H.J., HEDLIKOVA J., "Median Algebras", Discrete Math., 45 (1983), 1-30.
- BARBUT M., "Mediane, distributivité, éloignements" (1961), Math. Sci. hum., 70 (1980), 5-31.
- BARBUT M., MONJARDET B., Ordre et classification. Algèbre et Combinatoire, Paris, Hachette, 1970.
- BARTHELEMY J.P., "Remarques sur les propriétés métriques des ensembles ordonnés", Math. Sci. hum., 61 (1978), 39-60.
- BARTHELEMY J.P., "Caractérisations axiomatiques de la distance de la différence symétrique entre les relations binaires", Math. Sci. hum., 67 (1979), 85-113.
- BARTHELEMY J.P., LECLERC B., MONJARDET B., "Quelques aspects du consensus en classification", in : DIDAY E., et al., eds, Data Analysis and Informatics 3, Amsterdam, North-Holland, 1984a, 307-316.
- BARTHELEMY J.P., LECLERC B., MONJARDET B., "Ensembles ordonnés et taxonomie mathématique", in : POUZET M., RICHARD D., eds, Orders : description and roles, Amsterdam, North-Holland, 1984b, 523-548.
- BENZECRI J.P., "Description mathématique des classifications" (1967), in : L'analyse des données I, La Taxinomie, Paris, Dunod, 1973.
- BERGE C., Graphes et hypergraphes, Paris, Dunod, 1970. Graphs and hypergraphs, Amsterdam, North-Holland, 1973.
- BOORMAN S.A., OLIVIER D.C., "Metrics on spaces of finite trees", J. Math. Psychol., 10 (1973), 26-59.
- BROWN E.K., DAY W.H.E., "A computationally Efficient Approximation to the Nearest Neighbour Interchange Metric", J. of Classification 1 (1984), 93-124.
- CAYLEY A., "On the theory of the analytic forms called trees", Phil. Magazine, XIII, (1857), 172-176. Collected mathematical papers, vol.3, p.242, Cambridge University Press.
- CHANDON J.L., PINSON S., Analyse typologique, Paris, Masson, 1980.

- COMTET L., Analyse combinatoire, Paris, Presses Universitaires de France, 1970a. Advanced Combinatorics, Dordrecht, Reidel, 1974.
- COMTET L., "Sur le quatrième problème et les nombres de Schröder", C.R. Acad. Sci. Paris A-271 (1970 b) 913-916.
- COMYN G., VAN DORPE J.C., "Valuation et semi-modularité dans les demi-treillis", Math. Sci. hum., 56 (1976), 63-75.
- DOBSON A.J., "Unrooted trees for numerical taxonomy", J. Appl. Prob., 11 (1974), 32-42.
- ETHERINGTON I.M.H., "On non-associative combinations", Proc. Roy. Soc. Edinburgh, 59 (1939), 153-162.
- FLAMENT C., LECLERC B., "Arbres minimaux d'un graphe préordonné", Discrete Math., 46 (1983), 159-171.
- FOULDS L.R., ROBINSON R.W., "Determining the asymptotic number of phylogenetic trees", in : Combinatorial Mathematics VII, Lecture Notes in Mathematics 829, Berlin, Springer-Verlag, 1980.
- FOULDS L.R., ROBINSON R.W., "Enumeration of binary phylogenetic trees", in : Combinatorial Mathematics VIII, Lecture Notes in Mathematics 884, Berlin, Springer-Verlag, 1982.
- GARDNER M., "Mathematical games : Catalan numbers", Scientific American 235, (June 1976), 120-125.
- GOULD H., "Research bibliography of two special numbers sequences", Mathematica Monongaliae, 12 (1971).
- GUILBAUD G.Th., "Un problème leibnitzien : les partages en nombres entiers", Math. Sci. hum., 17 (1968), 13-36.
- GUILBAUD G.Th., ROSENSTIEHL P., "Analyse algébrique d'un scrutin", Math. Sci. hum., 4 (1960), 9-33.
- HARARY F., Graph Theory, Reading, Mass., Addison-Wesley, 1969.
- HARARY F., MOWSHOWITZ A., RIORDAN J., "Labelled trees with unlabelled endpoints", J. Comb. Theory, 6 (1969), 60-64.
- HARDING E.S., "The probabilities of rooted tree-shapes generated by random bifurcations", Adv. Appl. Prob., 3 (1971), 44-77.
- HARTIGAN J.A., Clustering algorithms, New-York, Wiley, 1975.
- HASKINS L., GUDDER S., "Height on posets and graphs", Discrete Math., 2, (1972), 357-382.

- JARVIS J.P., LUEDEMAN J.K., SHIER D.R., "Counterexamples in Measuring the distance between binary trees", Mathematical Social Sciences, 4 (1983), 271-274.
- KNUTH D.E., The art of computer programming, vol.1 : Fundamental algorithms, vol.3 : Sorting and searching, Reading, Mass., Addison-Wesley, 1969- 1973.
- LECLERC B., "Description combinatoire des ultramétries", Math. Sci. hum., 73 (1981), 5-37.
- LECLERC B., Description, évaluation et comparaison des hiérarchies de parties, rapport CAMS-EHESS, 1982.
- LERMAN I.C., Les bases de la classification mathématique, Paris, Gauthier-Villars, 1970.
- LERMAN I.C., Classification et analyse ordinale des données, Paris, Dunod, 1981.
- MARGUSH T., Mc MORRIS F.R., "Consensus n-trees", Bull. Math. Biology, 43 (1981), 239-244.
- MONJARDET B., "Caractérisations métriques des ensembles ordonnés semi-modulaires", Math. Sci. hum., 56 (1976), 77-87.
- MONJARDET B., "Metrics on a partially ordered set. A survey", Discrete Math., 35 (1981), 173-181.
- MOON J.W., "Enumerating labelled trees", in : HARARY F., ed., Graph Theory and Theoretical Physics, London, Academic Press, 1967.
- MOORE G.W., GOODMAN M., BARNABAS J., "An iterative approach from the standpoint of the additive hypothesis to the dendrogram problem posed by molecular data sets", J. Theor. Biol., 38 (1973), 423-457.
- MURTAGH F., "Counting dendrograms : a survey", Discrete Applied Math., 7 (1984), 191-199.
- ORE O., "Chains in partially ordered sets", Bull. Amer. Math. Soc., 49 (1943), 558-566.
- OTTER R., "The number of trees", Ann. Math., 49, n°3 (1948), 583-599.
- ROBINSON D.F., "Comparison of Labeled Trees with valency three", J. of Comb. Theory 11 (1971), 105-119.
- ROSENSTIEHL P., "L'arbre minimum d'un graphe", in : ROSENSTIEHL P., ed., Théorie des graphes (Rome, 1966), Paris, Dunod, 1967.

- SCHADER M., "Hierarchical analysis : classification with ordinal object dissimilarities", Metrika, 27 (1980), 127-132.
- SCHADER M., Scharfe und unscharfe Klassifikation Qualitativer Daten, Mathematic Systems in Economics 65, Verlagsgruppe Athenäum, 1981.
- SHOLANDER M., "Medians, Lattices and Trees", Proc. Amer. Math. Soc., 5 (1954), 808-812.
- SCHRÖDER E., "Vier combinatorische probleme", S. für Mat. Phys., 15 (1870), 361-376.
- SLOANE N.J.A., A handbook of integer sequences, New-York, Academic Press, 1973.
- WATERMAN M.S., SMITH T.F., "On the similarity of dendrograms", J. Theor. Biol. 73 (1978), 789-800.
- WEDDERBURN J.H., "The functional equation $g(x^2) = 2x + [g(x)]^2$ ", Ann. Math., 24 (1922), 121-140.