MATHÉMATIQUES ET SCIENCES HUMAINES

B. LECLERC

La comparaison des hiérarchies : indices et métriques

Mathématiques et sciences humaines, tome 92 (1985), p. 5-40
<http://www.numdam.org/item?id=MSH_1985__92__5_0>

Math. Sci. hum. (*23*^{e} année, n°92, *1985*, p.5-40)

LA COMPARAISON DES HIERARCHIES : INDICES ET METRIQUES

B. LECLERC *

INTRODUCTION

Le problème considéré ici est celui de la comparaison des hiérarchies de parties définies sur un ensemble fini fixé X.

De nombreux travaux sur la comparaison et sur le consensus des classifications hiérarchiques ont été menés récemment. Ceci provient de la multiplicité des méthodes et des applications de la classification. Ainsi, l'utilisation successive, sur un même ensemble de données, de plusieurs des métriques et des stratégies d'agrégation proposées "à la carte" par les grands logiciels d'analyse de données donne des classifications hiérarchiques différentes ; on cherche ensuite à confronter celles-ci et, éventuellement, à les réunir en une seule classification "consensus". D'un autre côté, les données peuvent être différentes, comme c'est le cas pour des études décalées dans le temps des mêmes objets. En systématique, on trouve des études sur la reconstitution d'arbres phylogénétiques à partir de l'examen de chaînes protéiniques : chaque type de chaîne conduit à un arbre, d'où l'importance dans ce domaine du problème du consensus (voir les références du par.1.2.3. ci-dessous).

Pour la comparaison d'objets complexes, on peut distinguer deux approches principales : d'une part, la construction d'indices numériques, un indice servant à évaluer et à comparer les hiérarchies sous un aspect donné (comme la moyenne sert à comparer des distributions du point de vue de leur position centrale) ; d'autre part, la définition de similarités ou de dissimilarités sur les paires d'objets (métriques, corrélations) exprimant un degré de ressemblance ou de dissemblance entre les objets de la paire, également chacune sous un aspect particulier (comme la corrélation de deux vecteurs mesure leur distance angulaire).

* Centre d'Analyse et de Mathématique Sociales, E.H.E.S.S. Paris.

Dans l'importante littérature proposant des métriques (ou des corrélations) pour la comparaison des classifications hiérarchiques, de nombreux travaux se restreignent aux hiérarchies indicées (ultramétriques, équivalences floues, *valued trees*, *fuzzy partitions*) ou stratifiées (*ranked trees*) [Sokal et Rohlf(1962), Hartigan (1967), Jambu (1975), Hubert et Baker (1977), Leclerc (1981)] ou aux préordonnances ultramétriques [Schader (1980)]. Beaucoup d'autres cependant considèrent simultanément les divers types de classifications hiérarchiques [Barthélemy, Leclerc et Monjardet (1984a,b, 1986), Boorman et Olivier (1973)] ou étudient spécifiquement les hiérarchies [Farris (1969), Williams et Clifford (1971), Phipps (1971), Bobisud et Bobisud (1972), Waterman et Smith (1978), Margush (1981),...].

On trouve dans la littérature un certain nombre d'indices d'évaluation, signalés, sauf omissions, au paragraphe 2.1.. Ils sont en général définis à partir de considérations sur l'information, ou en vue d'exprimer la concordance de plusieurs hiérarchies par l'évaluation d'une hiérarchie consensus. Ce dernier usage est discuté par Rohlf (1982) et par Day et McMorris (1984).

Nous essayons ici de faire une approche un peu systématique, en définissant, à partir de propriétés simples, des familles d'indices. On peut ensuite s'adapter à des situations variées en construisant, dans la famille adéquate, le ou les indices correspondant le plus précisément à ce que l'on veut étudier. L'approche métrique n'est pas oubliée puisque deux des familles contiennent des valuations qui permettent de définir des distances.

Cet article utilise les résultats du précédent, *Les hiérarchies de parties et leur demi-treillis*, dont le propos était précisément de rassembler les éléments utiles pour la comparaison des hiérarchies (structures, paramètres, dénombrements, ...). L'ensemble complète et met à jour un travail non publié de 1982.

On définit au paragraphe 1 diverses propriétés d'indices intéressantes a priori (par.1.1.), et, en relation avec ces propriétés, diverses situations où des indices ont été, ou peuvent être utilisés (par.1.2.). On a ainsi une grille d'étude des douze indices, dont sept au moins attestés dans la littérature, présentés au par.2.1.. Parmi eux, sept sont des valuations et permettent de définir des métriques.

Les trois plus typiques de ces métriques sont étudiées en 2.2. : la distance du plus court chemin dans le graphe de couverture du demi-treillis des hiérarchies, en rappelant un résultat important pour le problème du consensus ; une distance assez naturelle, liée à l'indice le plus étudié jusqu'ici, et qui a une décomposition intéressante ; une métrique définie différemment par Margush (1982) et dont on complète l'étude.

On passe à la recherche de famille d'indices au paragraphe 3. Les hiérarchies sont des familles de parties (classes) ordonnées de façon arborescente par l'inclusion. Les "indices linéaires" affectent, à des évaluations de ces classes pouvant provenir de divers contextes (par.3.1.), des coefficients dépendant, eux, de la structure propres de la hiérarchie considérée (par.3.2.). On revoit au par.3.3., sur ces coefficients, certaines des propriétés envisagées en 1.1.. On peut ensuite définir des familles particulières d'indices linéaires dont les prototypes ont été vus en 2.1..

On étudie ainsi au paragraphe 4 deux familles de valuations, les mesures sur $\mathcal{P}^2(X)$ (par.4.1.) et les coûts (par.4.2.). On évalue ensuite au par.5 la qualité d'une classification en la reliant au fait que les classes adjacentes sont bien distinctes : les indices correspondant vérifient la "condition de la classe inutile" du par.5.1.. On établit en 5.2. quelques propriétés des coefficients correspondants (ils ne peuvent définir des valuations) et l'on se restreint à des familles particulières, "sommes" et "moyennes hiérarchiques", que l'on caractérise (par.5.3.). On termine en revenant sur l'étude de certains indices de ce type.

Pour la bibliographie générale, on renvoie aux livres de Birkhoff (1967) et de Barbut et Monjardet (1970) sur les ensembles ordonnés, de Harary (1969) et de Berge (1970) sur les graphes, et de Hartigan (1975) et de Chandon et Pinson (1980) sur la classification. On écrira, par exemple, HP 3.2. pour renvoyer au paragraphe 3.2. de l'article "Les hiérarchies de parties"... dont il a été question précédemment.

Certaines références données dans l'article se situent dans la controverse du cladisme [cf. Janvier et al (1980), et la présentation de la nouvelle revue scientifique *Cladistics*], propre au domaine de la systématique.
Fondées sur une structure intrinsèquement hiérarchique (liée à un processus historique de bifurcation) des caractères des sujets à classer, les méthodes du cladisme ne semblent pas immédiatement généralisables en analyse de données.

NOTATIONS ET DEFINITIONS

X étant un ensemble fini de cardinal $\underline{n}$, une famille H de parties de X est une *hiérarchie* sur X ssi elle vérifie :
(H1) $X \in H$; (H2) $\emptyset \notin H$; (H3) $(\forall x \in X)\ \{x\} \in H$; (H4) $(\forall h,h' \in H)$ $h \cap h' \in \{h,h',\emptyset\}$. Pour $h \in H$, $F = \{\{x\}/x \in X\}$ est l'ensemble des *feuilles*, ou singletons de H ; $I = H - F$ est l'ensemble des *classes intérieures* de H ; on note $\underline{n}_h = |h|$. Pour $i \in I$, on note $H_i = \{h \in H/h \subset i\}$ la hiérarchie sur i induite par H, S_i la partition de i en éléments de H inclus strictement dans i

et maximaux, $\underline{s}_i = |S_i|$ le nombre de ces éléments, et $\underline{t}_i = \underline{s}_i - 1$.

A tout $h \in H - X$, on associe le *père* $p(h)$ de h, unique, tel que $h \in S_{p(h)}$. On note B la partition de F en feuilles de même père. On associe à H le graphe $G_H = (H,U)$, où $U = \{(h,h') \in H \times H \ / \ h = p(h')\}$.
C'est une arborescence de racine X : pour tout $h \in H$, il y a dans U un chemin unique $L_h = \{(X,h_1), (h_1,h_2), \ldots, (h_{\ell-1},h_\ell)\}$ de X à h.
On pose $\underline{\ell}_h = |L_h|$; on écrit $\underline{n}, S, \underline{s}, L, \ldots$ pour $\underline{n}_h, S_h, \ldots$ lorsque $h = X$.
La figure 1 suivante illustre la plupart des notations précédentes.

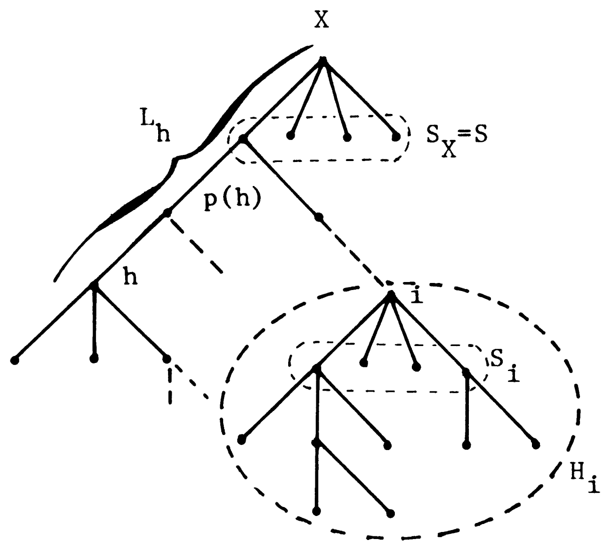

Fig.1

Ajoutons une $\underline{n}+1$-ième feuille r au graphe G_H, avec un arc (r,X). En oubliant l'orientation, on obtient un arbre réduit (sans sommet à deux voisins exactement) à n + 1 feuilles.
On définit ainsi une correspondance biunivoque entre ces arbres et les hiérarchies sur X (HP 2.2.).

L'ensemble $\mathcal{H}$ des hiérarchies sur X, stable pour l'intersection, est un demi-treillis, avec l'ordre d'inclusion (HP 1.3.). Pour $H \in \mathcal{H}$, on a $n+1 \leqslant |H| \leqslant 2\underline{n}-1$. Le nombre $\gamma(H) = |H| - (\underline{n}+1)$ des classes non triviales de H est la *graduation* de H. Pour $\underline{k} = 0,1,\ldots,\underline{n}-2$, on note $\mathcal{H}(\underline{k}) = \gamma^{-1}(\underline{k})$ le k-ième *niveau* de $\mathcal{H}$ (H.P 1.3). Le niveau $\mathcal{H}(0)$ ne contient que la hiérarchie minimale H_o (éventail, "bush",...) qui n'a que les classes triviales.
$\mathcal{H}(\underline{n}-2) = \mathcal{H}^m$ est l'ensemble des hiérarchies maximales, ou *binaires*, telles que, pour tout $i \in I$, on a $\underline{s}_i=2$ (HP 2.4.2.). A toute partition S de X, on associe la H-partition $H(S) \in \mathcal{H}$ en complétant S par X et par les singletons non présents dans (S) (H.P 2.4.1.)

Soient $H,H',H'' \in \mathcal{H}$. S'il existe $i \in I - X$ tel que $H' = H - \{i\}$, on écrit $H' = H^i$. H et H' sont alors liés par la relation de couverture $\prec$ définie ainsi sur $\mathcal{H}$. Le couple $(\mathcal{H},\prec)$ constitue le graphe de couverture du demi-treillis $(\mathcal{H}, \cap)$. Si de plus il existe $j \in H'$ tel que $H'' = H'^j$, on écrit $H'' = H^{ij}$; on adopte alors la convention selon laquelle $i \not< j$ (H.P 1.2.). L'ensemble (H,H^i,H^j,H^{ij}) constitue le *quadrilatère* $Q^{ij}(H)$.

1. INDICES SUR LES HIERARCHIES : PROPRIETES ET USAGES

1.1. Quelques propriétés

Un indice $\iota : \mathcal{H} \to R^+$ sur les hiérarchies est généralement défini pour un usage précis, et il importe, comme le soulignent Day et McMorris (1984) de savoir quelles propriétés de la hiérarchie H l'indice $\iota(H)$ "mesure" ; on peut aussi définir axiomatiquement ι à partir d'une liste de propriétés souhaitées.* Donnons d'abord quelques conditions envisageables pour ι, certaines classiques.

1.1.1. Monotonie

C'est le lien le plus élémentaire entre l'indice ι et la structure ordinale de $\mathcal{H}_X$ (HP 1.2.) ; il peut se présenter sous forme d'isotonie : croissance (condition C) ou non décroissance (C faible), ou d'antitonie : décroissance (D) ou non croissance (D faible).

(C) $(\forall\, H,H' \in \mathcal{H})$ $\quad H \subset H' \Rightarrow \iota(H) < \iota(H')$

(C faible) $\quad H \subseteq H' \Rightarrow \iota(H) \leqslant \iota(H')$

(D) et (D faible) sont définies dualement.

Pour étudier si un indice ι possède l'une de ces propriétés, il est souvent commode de déterminer l'expression générale de $c^i(\iota(H)) = \iota(H) - \iota(H^i)$, pour $H \in \mathcal{H}$ et $i \in I_H - \{X\}$, puis d'étudier le signe de $c^i(\iota(H))$. La valeur absolue de ce nombre est la valuation de l'arc (H,H^i) du graphe de couverture $(\mathcal{H}, \prec)$, valué selon ι.

1.1.2. La condition du quadrilatère

Nous résumons, dans le cas du demi-treillis $\mathcal{H}$, une méthode pour définir et calculer effectivement des métriques dans les ensembles ordonnés. Elle a été progressivement élaborée et étendue par divers auteurs [Haskins et Gudder (1972), Boorman et Olivier (1973), Grimonprez et van Dorpe (1976), Comyn et van Dorpe (1976), Bordes (1976), Duffus et Rival (1977), Barthélemy (1978, 1979a), Monjardet (1981)].

---

*Cette démarche sera illustrée aux paragraphes 4.1. et 5 suivants. Dans tous les cas, une étude fine de l'indice ι, quelle que soit la manière dont il a été obtenu, est nécessaire pour en atteindre une compréhension suffisante.

Tout indice $\iota : \mathcal{H} \to \mathbb{R}$ (ι désormais supposé - non strictement - positif) définit une métrique d_ι des plus courtes chaînes dans le graphe $(\mathcal{H}, \prec)$ dont les arêtes sont les $\{H,H'\} \subset \mathcal{H}$ tels que $H \prec H'$ ou $H' \prec H$, la longueur de l'arête $\{H,H'\}$ étant $|\iota(H) - \iota(H')|$. La longueur d'une chaîne (suite d'arêtes de la forme $(\{H,H_1\},\{H_1,H_2\},\ldots,\{H_{\underline{k}},H'\})$ entre H et H' est la somme des longueurs de ses arêtes et $d_\iota(H,H')$ est la longueur minimum d'une chaîne entre H et H'. d_ι n'est qu'un écart (une quasi-distance) s'il y a des arêtes de longueur nulle, i.e. s'il existe $H \in \mathcal{H}$, $i \in I_H$, avec $c^i(\iota(H)) = 0$. $d_\iota(H,H')$ est la longueur minimale (le coût minimal) de la transformation de H en H' par une succession d'opérations élémentaires d'adjonction ou de retrait d'une classe i exactement, le coût de chacune étant donné par c^i. Il reste à calculer effectivement d_ι dans un graphe $(\mathcal{H}, \prec)$ très vaste (HP 5.1.).

Si ι est monotone (non décroissant, par exemple), on a :

$$(\forall H, H' \in \mathcal{H})\ H \subseteq H' \Rightarrow d_\iota(H,H') = \iota(H) - \iota(H')$$

D'où pour H,H' quelconque, en considérant les chaînes entre H et H' passant par $H \cap H'$:

$$(\forall H,H' \in \mathcal{H})\ d_\iota(H,H') \leq \Big(\iota(H) - \iota(H \cap H')\Big) + \Big(\iota(H') - \iota(H \cap H')\Big) = \iota(H) + \iota(H') - 2\iota(H \cap H')$$

On trouve dans les références ci-dessus des conditions pour que l'inégalité précédente devienne une égalité, auquel cas le calcul de $d_\iota(H,H')$ devient facile. La condition la plus maniable dans le cas de $\mathcal{H}$ est la *condition du quadrilatère* [Barthélemy (1978, 1979a)].

Si dans une chaîne, on trouve successivement deux arêtes $\{H,H'\}$ et $\{H',H''\}$ avec $H \prec H'$ et $H'' \prec H'$ ("passage par le haut"), la semi-modularité (HP 1.3.) de $\mathcal{H}$ assure qu'il y a dans $(\mathcal{H}, \prec)$ deux arêtes $\{H, H \cap H''\}$ et $\{H \cap H'', H''\}$, qui peuvent être substituées aux deux précédentes ("passage par le bas"). La condition du quadrilatère assure que la nouvelle chaîne n'est pas plus longue que la première.

<u>Proposition 1.1.</u> [Cas particulier des résultats de Barthélemy]. Les deux conditions suivantes sont équivalentes si $\iota : \mathcal{H}_X \to \mathbb{R}$ vérifie (C faible) :

(Q) Pour tout quadrilatère $Q^{ij}(H)$ de $\mathcal{H}$ on a :

$$q^{ij}(\iota(H)) = \iota(H) + \iota(H^{ij}) - \iota(H^i) - \iota(H^j) \geq 0$$

(M) $(\forall H, H' \in \mathcal{H})\ d_\iota(H,H') = \iota(H) + \iota(H') - 2\iota(H \cap H')$.

On dira, si ι vérifie (Q) et (C faible), que ι est une *valuation* (inférieure) sur $(\mathcal{H}, \subseteq)$. Le lien de ι avec la structure semilatticielle de $\mathcal{H}$ est alors beaucoup plus fort que la simple isotonie.

Dualement, les conditions $(\overline{Q})$ et $(\overline{M})$ sont équivalentes pour ι non croissante :

$(\overline{Q})$ Pour tout quadrilatère $Q^{ij}(H)$, $q^{ij}(\iota(H)) \leqslant 0$

$(\overline{M})$ $(\forall\, H,H' \in \mathcal{H})$ $d_\iota(H,H') = 2\iota(H \cap H') - \iota(H) - \iota(H')$

Le fait pour l'indice ι de posséder simultanément les propriétés (C faible) et $(\overline{Q})$ (resp. (D faible) et (Q)) n'apporte rien en général au calcul de $d_\iota(H,H')$: ces propriétés correspondent à l'optimalité du "passage par le haut", mais l'existence de celui-ci n'est pas assurée. Cependant, si $H \cup H'$ existe, c'est-à-dire si H et H' sont deux sous-hiérarchies d'une même hiérarchie, on a dans ce cas :

$d_\iota(H,H') = 2\iota(H \cup H') - \iota(H) - \iota(H')$

(resp. $d_\iota(H,H') = \iota(H) - \iota(H') - 2\iota(H \cup H')$).

1.1.3. Symétrie

Un indice ι vérifie la condition (Sym) s'il est invariant par permutation des éléments de X. Pour H donnée, $\iota(H)$ ne dépend alors que de la *forme* (hiérarchie non étiquetée) T(H) (HP 4.1.). Il peut de plus ne dépendre que du partage de l'entier $\underline{n} - 1$ en nombres $\underline{t}_i$, $i \in I_H$ (H.P 4.2.).

1.1.4. Hiérarchies binaires extrémales

Il est souvent important de savoir pour quelles formes binaires un indice symétrique atteint ses valeurs extrémales. Intuitivement, les formes binaires les plus "typées" (H.P 2.4.2.) sont les peignes pour lesquels on a, pour tout $i \in I$, $\underline{n}_i = \underline{n}_{p(i)} - 1$, et les hiérarchies binaires équilibrées, pour lesquelles $\underline{n}_i = \underline{n}_{p(i)}/2$, définies pour $\underline{n}$ puissance de 2. On note, pour un indice symétrique ι, ι^p la valeur qu'il prend pour les peignes, ι^e celles pour les hiérarchies binaires équilibrées, et aussi $\iota_{min} = \min \{\iota(H)/H \in \mathcal{H}^m\}$, et ι_{max} définie dualement. On a quatre conditions d'extrémalité possibles pour le peigne :

(P) $(\forall\, H \in \mathcal{H})$ $\quad \iota(H) = \iota_{min} \Rightarrow$ H est un peigne

(P faible) $\quad \iota_{min} = \iota^p$

$(\overline{P})$ $\quad \iota(H) = \iota_{max} \Rightarrow$ H est un peigne

$(\overline{P}$ faible) $\quad \iota_{max} = \iota^p$

Pour les hiérarchies binaires équilibrées (HBE), on définit de même les propriétés (E), (E faible), ($\overline{E}$), ($\overline{E}$ faible), en se restreignant à n puissance de 2.

1.2. Quelques usages

1.2.1. Descriptions de la forme des hiérarchies

Un indice ι mesurant le degré de possession $\iota(H)$ par une hiérarchie H (ou par l'arborescence G_H associée) d'une propriété de forme (symétrie entre les classes, équilibre, centralité, ...) doit être symétrique au sens du par.1.1.3 ci-dessus. Le lien entre indices et propriétés n'est pas toujours évident, comme le savent bien les statisticiens ; plus proche de cette étude, la recherche d'une bonne mesure de la centralité d'un réseau (cf. Parlebas (1972)) illustre bien cette difficulté. Il est important de repérer les hiérarchies binaires extrémales : un indice d'équilibre, par exemple, doit posséder les propriétés (P) et ($\overline{E}$), ou leurs duales.

1.2.2. Mesures d'information

De telles mesures peuvent se faire dans des optiques bien différentes. Ainsi, la quantification de l'information fournie par la restriction des configurations d'un certain type à celles compatibles avec une hiérarchie H donnée (e.g. la restriction des parties de X aux classes de H, celle des partitions de X à celles incluses dans H, celle des hiérarchies binaires sur X à celles incluant H) repose sur des dénombrements.

D'un autre côté, on peut considérer une fonction (e.g. une moyenne) des informations correspondant à chacune des classes de H. Enfin, les exemples de la littérature les plus connus sont ceux d'indices traduisant la structuration de l'information liée à la forme arborescente de H.

On peut rapprocher de ces considérations la recherche d'une mesure de la qualité classificatoire d'une hiérarchie. En classification, on a généralement tendance à préférer des hiérarchies équilibrées à des peignes, cette préférence étant exprimée plus ou moins explicitement. De plus une hiérarchie non binaire peut être perçue comme incomplète.

Un indice symétrique de qualité classificatoire doit donc avoir simultanément les propriétés (C) et ($\overline{E}$), ou leurs duales (D) et (E). La construction d'un tel indice peut s'appuyer sur des considérations sur l'information du type de celles évoquées ci-dessus. Au paragraphe 5, on trouve une autre approche, fondée sur la recherche de classes adjacentes bien différenciées. De ce point de vue, les hiérarchies binaires équilibrées seront évidemment préférées aux peignes. Avec cette approche, on est à même de définir des indices non symétriques en relation avec un contexte particulier.

1.2.3. Mesures d'accord prises sur une hiérarchie consensus

Pour évaluer l'accord entre les éléments d'un $\underline{v}$-uple de hiérarchies $(H_1,\ldots,H_{\underline{v}}) \in \mathcal{H}^{\underline{v}}$, on peut déterminer d'abord une hiérarchie consensus H. La littérature sur l'obtention d'un tel consensus s'est récemment développée (Adams (1972), Barthélémy, Leclerc et Monjardet (1984a,b,1986), Finden et Gordon(1984), Margush et McMorris (1981), McMorris et Neumann (1983), Nelson (1979), Neumann et Norton (1985a,b), Sokal et Rohlf (1981), Stinebrickner (1984a,b). L'accord est ensuite mesuré par la valeur $\iota(H)$ prises sur H par un indice bien choisi ι, évidemment croissant. Mickevich(1978), Colless (1980), Nelson et Platnick (1981), Schuh et Farris (1981), Rohlf (1982) ont proposé des indices pour cet usage.

Selon Rohlf, un tel indice doit être constant (égal à 1 par normalisation) sur $\mathcal{H}^m$, pour ne pas dépendre en fait d'aspects particuliers des éléments du $\underline{v}$-uple considéré. Pour Day et McMorris (1984), un indice d'accord normalisé doit prendre la valeur 1 dès qu'il y a unanimité, c'est-à-dire identité des $H_1,\ldots,H_{\underline{v}}$, ce qui conduit à récuser les indices d'accord du type précédent. On peut cependant attribuer une plus grande importance à une unanimité sur une hiérarchie binaire qu'à une unanimité sur une hiérarchie ayant peu de classes non triviales, voire pas du tout si les éléments du v-uple sont tous égaux à H_o.

1.2.4. Valuations et métriques associées

La proposition 1.1. ci-dessus établit comme cas particulier de résultats généraux le fait que, pour un indice ι vérifiant les conditions (C) et (Q), la distance d_ι du plus court chemin dans le graphe de couverture $(\mathcal{H}, \prec)$ valué selon ι est donnée, pour tous $H,H' \in \mathcal{H}$, par :

$$d_\iota(H,H') = \iota(H) + \iota(H') - 2\iota(H \cap H')$$

Si ι ne satisfait que (C), l'expression ci-dessus définit seulement une dissimilarité sur $\mathcal{H}$, qui ne satisfait pas en général l'inégalité triangulaire des distances.

Inversement, toute métrique d sur $\mathcal{H}$ fait correspondre, à toute hiérarchie particulière H_1 un indice ι^{d,H_1} défini, pour tout $H \in \mathcal{H}$, par $\iota^{d,H_1}(H) = d(H,H_1)$. Posons $\iota^d = \iota^{d,H_o}$; une condition minimale de compatibilité de d avec le graphe $(\mathcal{H},\prec)$ est que ι^d soit un indice croissant. Monjardet (1981) donne des conditions nécessaires et suffisantes pour que ι^d soit une valuation sur le demi-treillis $(\mathcal{H},\cap)$ et que l'on aie $d_\iota d = d$.

On peut utiliser une métrique d (ou une dissimilarité) pour la construction d'indices d'accord sur les v-uples de hiérarchies, satisfaisant à la condition de Day et McMorris présentée au paragraphe précédent : moyenne normalisée des $d(H_k, H_{k'})$, $k, k' = 1, \ldots, v$, par exemple, ou éloignement normalisé d'une hiérarchie consensus H (cf. Barthélémy et Monjardet (1981)), l'éloignement étant la quantité $\Sigma\, d(H, H_k)$.

2. QUELQUES INDICES SYMETRIQUES

2.1. Indices

2.1.1. La graduation γ

$\gamma(H) = |I_H| - 1$ est le nombre de classes non triviales de H et est compris entre 0 (pour $H = H_o$) et $n - 2$ (obtenu si et seulement si H est une hiérarchie binaire).

Avec $c^i(\gamma(H)) = 1$ et $q^{ij}(\gamma(H)) = 0$, γ est croissant et est une valuation. Constant sur les hiérarchies de même niveau, ce n'est pas un indice de forme. Proposé comme indice d'accord par Nelson (1979) et par Schuh et Polhemus (1980) et, sous forme normalisée $\gamma/n - 2$ par Colless (1980) et Sokal et Rohlf (1981), il satisfait la condition de Rohlf, qui met l'accent sur sa simplicité et sa facilité d'interprétation.

2.1.2. Trois éloignements de la racine X dans l'arborescence G_H

Comme indices de forme, ils ont une signification précise. Deux d'entre eux sont, de plus, des valuations.

2.1.2.1. $\rho(H) = \max_{x \in X} \ell_x$, *écartement* de X.

L'indice ρ vérifie (C faible) avec $c^i(\rho(H)) \in \{0,1\}$, mais n'est pas une valuation, car $q^{ij}(\rho(H))$ prend les valeurs -1,0 et 1. Il vérifie $(\overline{P})$ avec $\rho^p = n - 1$ et (E) avec $\rho^e = \log_2 n$. C'est l'un des indices d'équilibre les plus simples.

2.1.2.2. $\sigma(H) = \sum_{x \in X} \ell_x = \sum_{i \in I} n_i$, *éloignement partiel* de X.

Avec $c^i(\sigma(H)) = n_i$ et $q^{ij}(\sigma(H)) = 0$, c'est une valuation. Son double calcul (H.P 2.3.)le rattache aux deux familles de valuations signalées ci-dessous : les mesures sur $\mathcal{P}^2(X)$ (par. 4.1.) et les éloignements (par. 4.2.). Comme indice de forme, il vérifie $(\overline{P})$ avec $\sigma^p = (n+2)(n-1)/2$ et (E) avec $\sigma^e = n \log_2 n$ (cf. Knuth (1973)).

La dernière propriété constitue un résultat célèbre sur l'organisation de l'information : le nombre moyen des choix dichotomiques pour sélectionner un élément parmi n est minimal quand ces choix sont organisés suivant une hiérarchie binaire équilibrée (le nombre de choix pour sélectionner un élément x particulier étant ℓ_x). Sous cet aspect, il a été d'abord étudié par Burge (1958). Sous le nom de "distropy", il est aussi étudié par Moriconi et Rizzi (1981), toujours dans le cas binaire. Nelson et Platnick (1981) l'appellent "total information" et l'utilisent comme indice de concordance entre hiérarchies.

2.1.2.3. $\mu(H) = \sum_{h \in H} \underline{\ell}_h$, *éloignement* de X.

Dans le cas binaire, les indices μ et σ sont liés par l'égalité $\mu(H) = 2\sigma(H) - 2\underline{n} - 2$ (HP 2.3.), d'où les propriétés $(\overline{P})$ et (E) pour μ. On montre plus loin (par. 4.2.) que $c^i(\mu(H)) = \underline{\ell}_i + \underline{n}_i + \gamma(H_i)$ et que $q^{ij}(\mu(H))$ est égal à 0 ou à 1. L'indice μ est donc une valuation.

2.1.3. Quelques indices liés à σ.

2.1.3.1. L' "entropie des chemins" (path entropy) de Green (1973).

$$\eta(H) = \sum_{i \in I} \underline{n}_i \log_2 \underline{s}_i .$$

L'utilisation de σ pour évaluer l'organisation de l'information dans une arborescence binaire peut difficilement être généralisée à H quelconque. Green propose de valuer chaque arc ((p(h),h) du chemin L_x joignant la racine X à la feuille x par la quantité $\log_2 \underline{s}_{p(h)}$ mesurant l'information correspondant au choix d'un successeur de p(h) parmi $\underline{s}_{p(h)}$. Par sommation sur les arcs de L_x, puis sur les $x \in X$, on obtient l'indice ci-dessus. Dans le cas binaire, on retrouve $\eta = \sigma$.

Green montre que les bornes de η dans $\mathcal{H}^m$ sont aussi celles dans $\mathcal{H}$. Pour $\underline{n}$ puissance de 2, on a $\eta(H_o) = \eta^e = \underline{n} \log_2 \underline{n}$, ce qui peut être décevant: l'absence d'organisation est optimale ! Pour $\underline{n} \geqslant 5$, η n'est pas monotone.

2.1.3.2. Deux mesures de liaison globale des éléments de X dans H.

Soient $x,y \in X$, $H \in \mathcal{H}$. On note xHy l'élément unique de H contenant x et y et minimal pour l'inclusion avec cette propriété. $\underline{r}_H(x,y)$ est le nombre minimal d'arcs d'un chemin de G_H entre x et y. On a :

$\underline{r}_H(x,y) = \underline{\ell}_x + \underline{\ell}_y - 2\underline{\ell}_{xHy}$. On pose : $\beta(H) = \sum_{x,y \in X,\, x \neq y} \underline{r}_H(x,y)$, et on trouve :

$\beta(H_o) = \underline{n}(\underline{n}-1)$ et $c^i(\beta(H)) = \underline{n}_i(\underline{n}-\underline{n}_i)$, d'où :

$$(\forall H \in \mathcal{H}) \quad \beta(H) = \underline{n}(\underline{n}-1) + \sum_{i \in I} \underline{n}_i(\underline{n}-\underline{n}_i) .$$

On voit sur des exemples que ni les peignes ni les HBE ne sont des hiérarchies extrémales pour β (qui ne peut donc être utilisée comme indice d'équilibre). Pour aller plus loin, on remarque que $\beta(H) + \sigma(H) + \underline{n}$ est la somme des distances entre les n+1 sommets pendants de l'arbre réduit à $\underline{n}+1$ feuilles associé à G_H (HP 2.2. et introduction) et ne dépend que de l'arbre réduit non étiqueté associé à T(H) (HP 4.1.). En particulier, toutes les hiérarchies binaires H telles que, si S = (h,h'), H_h et $H_{h'}$ sont des peignes, ont le même arbre réduit planté non étiqueté, donc la même valeur de $\beta + \sigma$. On minimise $\sigma(H)$ parmi ces hiérarchies, donc on maximise $\beta(H)$ en posant $\underline{n}_h = \left\lfloor \frac{\underline{n}}{2} \right\rfloor$.

L'idée de comparer deux hiérarchies H et H' en comparant les distances $\underline{r}_H$ et $\underline{r}_{H'}$ est ancienne et diverses méthodes, autres que celles présentées ici, ont été proposées [Farris (1969), Williams et Clifford (1971), Phipps (1971), Bobisud et Bobisud (1972)] .

Un indice apparenté est la "levels sum" de Schuh and Farris (1981), égale, à une constante près, à $\beta'(H) = \sum_{x,y \in H, x \neq y} \underline{\ell}_{xHy}$. On compte pour chaque paire x,y d'éléments distincts de X le nombre de classes de H contenant simultanément x et y, et on somme sur toutes les paires.

De $\beta'(H_o) = \underline{n}(\underline{n} - 1) / 2$ et $c^i(\beta'(H)) = \underline{n}_i(\underline{n}_i - 1) / 2$, on tire

$$2\,\beta'(H) = \sum_{i \in I} \underline{n}_i(\underline{n}_i - 1) ,$$

Comme γ et σ, β et β' sont obtenus en donnant des poids aux classes de H et en sommant ces poids. On étudie au par. 4.1. ci-dessous la famille des valuations constituée des indices de ce type.

2.1.3.3. Cette famille contient aussi deux des indices proposés comme indices d'accord dans la littérature :

$\sigma' = \sum_{i \in I} \underline{n}_i - 1 = \sigma - \gamma + 1$, "term information" de Nelson et Platnick (1981).

Pour $\underline{n}$ suffisamment grand, σ' diffère peu de σ.

$\sigma'' = \sum_{i \in I} \min(\underline{n}_i - 1, \underline{n} - \underline{n}_i)$, coefficient de similarité de Mickevich (1978), dont l'étude est disséminée dans plusieurs papiers et thèse, où il est notamment établi que $\sigma''_{max} = \sigma''^{p} = \lfloor n-1/2 \rfloor \lfloor \underline{n}-2/2 \rfloor$. Comme indice d'accord, il ne satisfait pas la condition de Rohlf, qui critique aussi le fait qu'il majore l'importance des classes d'effectif moyen.

2.1.4. Indices basés sur des dénombrements.

2.1.4.1. On a abordé (HP 2.6.) l'étude du nombre $\varepsilon(H)$ des partitions de X incluses dans H. L'indice ε est croissant et vérifie $\varepsilon_{min} = \varepsilon^{p} = \underline{n}$. On a conjecturé aussi que $\varepsilon_{max} = \varepsilon^{e}$. Cet indice aurait alors les propriétés d'un indice de qualité classificatoire, difficile à interpréter cependant.

La façon la plus simple de le calculer semble être l'emploi de la récurrence $\varepsilon(H) = 1 + \prod_{h \in S} \varepsilon(H_h)$.

2.1.4.2. Le nombre $\alpha(H)$ des arbres minimaux de la préordonnance hiérarchique P_H induite par H sur les paires d'éléments de X est

$$\alpha(H) = \frac{1}{\underline{n}} \prod_{i \in I} \underline{n}_i^{\underline{t}_i} \qquad \text{(HP 2.7.2.).}$$

On a établi que α est un indice décroissant et que $\alpha_{max} = \alpha^{p} = (\underline{n} - 1)!$, tandis que $\alpha_{min} = \alpha^{e} = 4^{\underline{n}-1} / \underline{n}^{2}$.
Nous laissons au lecteur le soin d'établir que $q^{ij}(\alpha(H)) \geqslant 0$, et donc que α n'est pas une valuation. C'est par contre un indice de qualité classificatoire, liant celle-ci à la restriction du nombre des arbres minimums de la classification. Un arbre minimum est une relation binaire sur X résumant certains aspects de la hiérarchie F, ou des ultramétriques liées à H ; cf. Leclerc (1981).

La proportion des arbres minimums pour P_H parmi tous les arbres sur X est $\pi(H) = \alpha(H) / \underline{n}^{\underline{n}-2}$, à partir de laquelle on définit :

$$\delta(H) = -\frac{1}{\underline{n}-1} \log_2 \pi(H) = \sum_{i \in I} \frac{\underline{t}_i}{\underline{n}-1} \log_2 \frac{\underline{n}}{\underline{n}_i}$$

L'indice δ est fonctionnellement lié à α, mais il a une autre interprétation, en terme d'information. Puisque $\Sigma\, \underline{t}_i = \underline{n} - 1$, c'est une moyenne pondérée des gains d'information de Wiener-Shannon correspondant aux classes $i \in I$.

Des bornes $(\underline{n} - 1)! \geqslant \alpha(H) \geq \frac{4^{\underline{n}-1}}{\underline{n}^2}$ établies pour H binaire, on déduit :

$$\log_2 \underline{n} - \frac{1}{\underline{n}-1} \log_2(\underline{n}!) \leq \delta(H) \leq \frac{\underline{n}}{\underline{n}-1} \log_2 \underline{n} - 2$$

Le terme de gauche correspond aux peignes et tend vers $\log_2 e$ quand $\underline{n}$ croît (e base des logarithmes néperiens). L'indice δ est aussi croissant, avec $c^{i}(\delta(H)) = (\underline{t}_i/\underline{n} - 1) \log_2 (\underline{n}_{p(i)}/\underline{n}_i)$, mais n'est pas une valuation, car $q^{ij}(\delta(H))$ est égal à 0 si $i \neq p(j)$ et à $(\underline{t}_j/\underline{n} - 1) \log_2 (\underline{n}_i/\underline{n}_{p(i)})$, nombre négatif, si $i = p(j)$ (cf. par. 5.1. ci-dessous).

C'est par contre un indice de qualité classificatoire, interprétable en termes d'information. Il appartient à la famille d'indices étudiée au paragraphe 5, où son étude est complétée.

2.1.4.3. Rohlf (1982) propose, comme indice d'accord, de prendre la proportion $\nu'(H)$ des hiérarchies binaires contenant H parmi toutes les hiérarchies binaires sur X. Il donne l'expression de $\nu'(H)$:

$$\nu'(H) = \frac{1}{\underline{Hb}(\underline{n})} \prod_{i \in I} \underline{Hb}\ (\underline{s}_i),$$

Où $\underline{Hb}(\underline{k})$ (noté $\underline{H}(\underline{k},\underline{k} - 2)$ dans HP 5.1.) est le nombre de hiérarchies binaires à $\underline{k}$ feuilles étiquetées. C'est le produit des $\underline{k}$ - 1 premiers entiers impairs (Harding (1971)) :

$$\underline{Hb}(\underline{k}) = 1.3.5 \ldots (2\underline{k} - 3) = \frac{(k - 2)!}{2^{\underline{k}-1}(\underline{k}-1)!},$$

d'où l'on tire une forme de ν', un peu plus condensée que celle donnée par Rohlf :

$$\nu\ H) = \frac{(n - 1)!}{2(\underline{n} - 1)!} \prod_{i \in I} \frac{(2\underline{t}_i)!}{\underline{t}_i!}$$

L'indice ν' ne dépend donc que du partage de l'entier $\underline{n}$ - 1 en parts égales aux $\underline{t}_i$, elles-mêmes liées aux extensions binaires de H (HP 4.2.). Il est décroissant et constant sur $\mathcal{H}_m$.

Un indice plus maniable est $\nu = -\log_2 \nu'$, qui s'interprète comme le gain d'information correspondant à la restriction de $\mathcal{H}^m$ aux hiérarchies binaires contenant H. Pour $H \in \mathcal{H}$, on a :

$$\nu(H) = \log_2 \frac{(\underline{n} - 1)!}{(2\underline{n} - 1)!} + \sum_{i \in I} \log_2 \frac{(2\underline{t}_i)!}{\underline{t}_i!}$$

2.2. Trois métriques sur l'ensemble des hiérarchies.

2.2.1. $\underline{d}_\gamma$: la distance de la différence symétrique.

d_γ est la distance de la différence symétrique Δ dans $\mathcal{P}(X)$, avec $H \Delta H' = (H \cup H') - (H \cap H')$, pour $H,H' \in \mathcal{H}$.

Proposition 2.1. Pour tout $H,H' \in \mathcal{H}$, on a :

$$d_\gamma(H,H') = \gamma(H) + \gamma(H') - 2\gamma(H \cap H') = |H \Delta H'|.$$

En effet, $d_\gamma(H,H') = |I_H| - 1 + |I_H| - 1 + 2(|I_{H\cap H'}| - 1) = |I_H \Delta I_{H'}| = |H \Delta H'|$. □

$d_\gamma(H,H')$ est le nombre minimum d'adjonctions ou de suppressions de classes nécessaires pour passer de H à H', c'est-à-dire la longueur du plus court chemin entre H et H' dans le graphe de couverture $(\mathcal{H}, \prec)$. Cette distance peut être vue comme une restriction de la distance naturelle sur les "arbres phylogénétiques" étudiée par Robinson et Foulds (1981) ; cf. Barthélémy et Leclerc (1986). Du fait que $\mathcal{H}$ est un demi-treillis à médianes (HP 1.3.), en connaît, pour tout $\underline{v}$-uple de hiérarchies $(H_1,\ldots,H_{\underline{v}})$, l'expression algébrique des hiérarchies consensus H qui, minimisant $\sum_{\underline{k}=1}^{v} d_\gamma(H,H_{\underline{k}})$, sont *médianes* pour d_γ (Margush et McMorris (1981) ; cf. Bandelt et Barthélemy (1984), Barthélemy, Leclerc et Monjardet (1984a,b, 1986)).

On a $d_\gamma(H,H') \leq \gamma(H)+\gamma(H') \leq 2\underline{n}-2$, avec la première égalité ssi $H\cap H' = H_o$, et la seconde si, de plus, $H,H' \in \mathcal{H}^m$.

Il y a dans $\mathcal{H}^m$ beaucoup de paires de hiérarchies H,H' à distance $2\underline{n}-2$, car (HP 2.2.) H et H' sont de "petits" ensembles dans $\mathcal{P}(X)$. La distance d_γ est peu discriminante entre les couples (H,H') de hiérarchies. Elle se caractérise bien, car le demi-treillis $(\mathcal{H}, \subseteq)$ est un sous-inf-demi-treillis semi-modulaire de $(\mathcal{P}^2(X), \subseteq)$, avec la même graduation. Cette situation est analogue à celle du demi-treillis des ordres dans $\mathcal{P}(X^2)$ [cf. Barthélemy (1979b)], et on a la même caractérisation axiomatique de la distance de la différence symétrique sur les ordres [Bogart (1973, Barthélemy (1979b) ; cf. aussi Barthélemy, Flament et Monjardet (1982)] :

<u>Proposition 2.2.</u> d_γ est la seule application d : $\mathcal{H}^2 \to \mathbb{R}$ vérifiant :

(Δ1) $H \subseteq H' \Rightarrow d(H,H') = d(H',H)$

(Δ2) $H \subseteq H' \subseteq H'' \Rightarrow d(H,H'') = d(H,H') + d(H',H'')$

(Δ3) $d(H,H') = d(H,H\cap H') + d(H\cap H',H')$

(Δ4) $H \prec H' \Rightarrow d(H,H') = 1$

Le fait que d_γ est une distance est une conséquence des axiomes précédents.

2.2.2. <u>La distance d_σ</u>

Elle est plus discriminante que d_γ, et le fait même qu'elle soit liée à σ fait son intérêt. En particulier, elle appartient simultanément aux deux familles de distances étudiées au paragraphe 4. On a, pour $H,H' \in \mathcal{H}$, en prenant successivement la forme générale, le cas où $H\cap H' = H_o$, et celui où, de plus, H et H' sont des peignes :

$$d_\sigma(H,H') = \sigma(H) + \sigma(H') - 2\sigma(H\cap H') \leq \sigma(H) + \sigma(H') - 2\underline{n} \leq (\underline{n}-2)(\underline{n}+1).$$

$d_\sigma(H,H')$ peut être décomposée comme suit selon les éléments $x \in X$:

$\underline{\ell}_{H,x}$ étant le nombre d'éléments de I_H contenant x, on peut écrire

$d_\sigma(H,H') = \sum_{x \in X} \left(\underline{\ell}_{H,x} + \underline{\ell}_{H',x} - 2\underline{\ell}_{H \cap H',x}\right) = \sum_{x \in X} d_\sigma^x(H,H')$, où $d_\sigma^x(H,H')$ est le nombre d'éléments de $H \Delta H'$ contenant x.

On a aussi $d_\sigma(H,H') = \sum_{h \in H \Delta H'} \underline{n}_h$.

2.2.3. $\underline{d}_\mu$: la distance de Margush.

Considérons la relation $U_H \subset (\mathcal{P}(X))^2$ associée à la hiérarchie H (HP 2.1.) et sa fermeture transitive $\hat{U}_H$. Margush (1982) étudie la distance d_υ sur $\mathcal{H}$, dont il donne une caractérisation axiomatique, définie par :

$$(\forall H, H' \in \mathcal{H}) \quad d_\upsilon(H,H') = |\hat{U}_H \,\Delta\, \hat{U}_{H'}| \ .$$

Nous allons montrer que $d_\mu = d_\upsilon$. Tout d'abord :

Proposition 2.3. En posant $\mu'(H) = |\hat{U}_H| = d_\upsilon(H,H_o) + \underline{n}$ on a :

$$(\forall H \in \mathcal{H}) \quad \mu'(H) = \mu(H).$$

Démonstration. On a d'abord $\mu'(H_o) = \mu(H_o) = \underline{n}$. Explicitons $c^i\left(\mu'(H)\right)$. En passant de H à H^i, on supprime les $\underline{\ell}_i$ arcs (h,i), $h \in L_i$, et les $(|H_i| - 1)$ arcs (i,h), $h \in H_i - \{i\}$, d'où $c^i(\mu'(H)) = c^i(\mu(H)) = \underline{\ell}_i + |H^i| - 1$. Si l'égalité est vraie pour H, elle l'est pour tout $H' \succ H$, et donc pour tout $H \in \mathcal{H}$. □

Proposition 2.4. $d_\mu = d_{\mu'}$.

Démonstration. Il suffit de montrer [Monjardet (1981), proposition 5 (par dualité)] que d_υ possède les propriétés d'intermédiarité (1) et (2) :

(1) $H \subseteq H' \subseteq H'' \Rightarrow d_\upsilon(H,H'') = d_\upsilon(H,H') + d_\upsilon(H',H'')$

(2) $d_\upsilon(H,H') = d_\upsilon(H,H \cap H') + d_\upsilon(H \cap H',H)$

(1) L'application $H \to \hat{U}_H$ est monotone : $H \subseteq H' \Rightarrow U_H \subseteq U_{H'} \Rightarrow \hat{U}_H \subseteq \hat{U}_{H'}$. (1) est alors ramenée à une propriété d'intermédiarité classique de la distance de la différence symétrique. Il en est de même pour (2), après avoir vérifié que, de plus, $\hat{U}_H \cap \hat{U}_{H'} = \hat{U}_{H \cap H'}$. □

Le calcul de la borne de d_μ découle de celui fait pour σ :

$d_\mu(H,H') \le \mu(H) + \mu(H') - 2\underline{n} \le 2\underline{n}(\underline{n}-2)$, la dernière valeur étant atteinte si H et H' sont des peignes et $H \cap H' = H_o$.

3. LES INDICES LINEAIRES

3.1. La prise en compte d'une valeur des classes.

Dans de nombreux cas, il est intéressant pour l'évaluation des hiérarchies de prendre en considération une fonction donnant une valeur aux classes, soit $f : \mathcal{P}(X) \to \mathbb{R}$.

Exemples.

1) $f = 1_{\mathcal{P}(X)}$, i.e. pour tout $h \subseteq X$, $f_h = 1$.

2) $f_h = 1$ si $h = \{x\} \in F$, $f_h = 0$, sinon. Nous notons alors $f = 1_F$.

3) $f_h = \underline{n}_h$ (notation $f = n$). Plus généralement, $f_h = m_h$, où m est une *mesure* (par exemple une probabilité) sur $\mathcal{P}(X)$ définie par des m_x non nuls, et, pour $h \subseteq X$, $m_h = \sum_{x \in h} m_x$ (et $m_\emptyset = 0$).

4) $f_h = \log_2 \underline{n}_h$, information de Wiener-Shannon (on peut considérer une autre mesure d'information).

5) f_h fonction de *rareté* des éléments de h dans X. Par exemple
$f_h = \frac{\underline{n}}{\underline{n}_h} - 1$, $f_h = \log_2 \frac{\underline{n}}{\underline{n}_h}$, $f_h = 1 - \frac{\underline{n}_h}{\underline{n}}$ [cf. Patil et Taillie (1979)].

6) f_h est l'un des indices de cohésion de la classe h utilisés en classification mathématique. Par exemple, on s'est donné un indice de dissemblance $R : X^2 \to \mathbb{R}^+$ et on prend
$f_h = \max \{R(x,x') \,/\, x,x' \in H\}$, $f_h = \min \{R(x,x') \,/\, x \in h, x' \in X - h\}$,
$f_h = \sum_{x,x' \in h} R(x,x')$, etc.

3.2 . Définition des indices linéaires.

Nous allons donc considérer des indices $\iota : \mathcal{H} \times \mathbb{R}^{\mathcal{P}(X)} \to \mathbb{R}$.
Pour f fixé, on a un indice $\iota(f)$, qui donne une évaluation $\iota(H,f)$ de la hiérarchie H. Comme il est nécessaire de se limiter, nous allons borner ici notre étude aux indices linéaires en f avec de plus, deux propriétés naturelles. L'indice $\iota(H,f)$ sera dit *linéaire* s'il vérifie les propriétés suivantes :

(L1) $\left(\forall f, f' \in \mathbb{R}^{\mathcal{P}(X)}, \lambda, \lambda' \in \mathbb{R}\right) \iota(H, \lambda f + \lambda' f') = \lambda\,\iota(H,f) + \lambda'\iota(H,f')$

(L2) Si les restrictions de f et f' à H sont égales, $\iota(H,f) = \iota(H,f')$

(L3) Si f est non négative, l'indice $\iota(f)$ est non négatif.

Alors d'après (L1), ι est de la forme $\iota(H,f) = \sum_{h \in \mathcal{P}(X)} a_{H,h} f_h$, avec $h \notin H \Rightarrow a_{H,h} = 0$ (par L2), et, pour tout $h \in H \in \mathcal{H}$, $a_{H,h} \geq 0$ (par L3).

Finalement, nous considérons des indices $\iota(H,f) = \sum_{h \in H} a_{H,f}\, f_h$, entièrement définis par la donnée d'une fonction $a : \underset{H \in \mathcal{H}}{\times} H \to \mathbb{R}^+$ qui associe à tout $H \in \mathcal{H}$ une fonction $a_H : H \to \mathbb{R}^+$. On notera si possible a_h pour $a_{H,h}$, a_h^i pour a_{H^i}, etc. Pour $H \in \mathcal{H}$, $i,j \in I_H$ (avec la convention $i \not\subseteq j$), on a :

$$c^i\Big(\iota(H,f)\Big) = a_i f_i + \sum_{h \in H^i} \left(a_h - a_h^i\right) f_h \qquad (3.1.1.)$$

$$q^{ij}\Big(\iota(H,f)\Big) = (a_i - a_i^j) f_i + (a_j - a_j^i) f_j + \sum_{h \in H^{ij}} (a_h + a_h^{ij} - a_h^i - a_h^j)\, f_h \qquad (3.1.2.)$$

Nous supposerons désormais f non négative, ce qui est le cas dans les exemples 1) à 6) précédents et simplifiera l'étude des signes de c^i et q^{ij}

3.3. Bonnes propriétés pour a.

L'intérêt de tels indices est qu'ils permettent l'évaluation de toute hiérarchie H en tenant compte simultanément de la structure propre de H et d'un contexte extérieur à H. La fonction f, qui correspond à un angle particulier d'étude de H (mesure, information, rareté, cohésion), est définie sur $\mathcal{P}(X)$ indépendamment de la hiérarchie H particulière considérée, tandis que la prise en compte de la structure propre de celle-ci se fait par les coefficients $a_{H,h}$. Dans cette perspective, on reprend ici pour celle-ci certaines des propriétés envisagées paragraphe 1.1.

Symétrie (LSym)

Permutabilité relative (LP). La fonction a_H est invariante par permutation des éléments de X.

Monotonie relative (LC). $H \subseteq H' \Rightarrow (\forall h \in H)\; a_{H,h} \leq a_{H',h}$.

Condition du quadrilatère relative (LQ).

$(\forall H \in \mathcal{H})(\forall i, j \in H, h \in H^{ij})\; a_{H,h} + a_{H,h}^{ij} - a_{H,h}^{i} - a_{H,h}^{j} \geq 0$

Proposition 3.1. On considère les propriétés suivantes pour un indice linéaire $\iota(H,f)$:

(1) $(\forall f : \mathcal{P}(X) \to \mathbb{R}^+)$ $\iota(f)$ vérifie (C faible)

(2) $(\forall f : \mathcal{P}(X) \to \mathbb{R}^+)$ $\iota(f)$ vérifie (Q)

Alors, on a les implications (LC) $\Longleftrightarrow$ (1) et [(LC) $\wedge$ (LQ)] $\Longleftrightarrow$ (2) .

La démonstration est sans difficulté, à partir des égalités (3.1.1.) et (3.1.2.). □

Les propriétés (LC) et (LQ) assurent donc que, pour toute fonction f non négative, l'indice $\iota(f)$ est une valuation.

Exemples de fonctions a.

1) $(\forall h \in H \in \mathcal{H})\ a_{H,h} = 1$ (vérifie (LP), (LC), (LQ)).

pour $f = 1_{\mathcal{P}(X)}$, $\iota(H,f) = \gamma(H) + \underline{n} + 1$; pour $f = n$, $\iota(H,f) = \sigma(H) + \underline{n}$.

2) $a_{H,h} = \underline{\ell}_{H,h}$ (LP) (LC) (LQ) [cf. 3.4.]

pour $f = 1_F$, $\iota(H,f) = \sigma(H)$; pour $f = 1_{\mathcal{P}(X)}$, $\iota(H,f) = \mu(H)$.

3) $a_{H,i} = \underline{s}_i$ et $a_{H,x} = 0$, $x \in X$ [(LP) (LD)].

avec $f = n - 1_F$, on trouve $\delta + \sigma$, et pour $a_{H,i} = \underline{t}_i$, δ (η pour $a_{H,i} = \log_2 \underline{s}_i$) .

Les indices β, β', σ', σ'' et ν sont aussi, de façon immédiate et éventuellement à une constante près, des indices linéaires.

4. DEUX FAMILLES DE VALUATIONS SUR $\mathcal{H}$.

4.1. Mesures sur $\mathcal{P}^2(X)$

A une constante près, γ et σ sont des cas particuliers de l'exemple 1) ci-dessus. En même temps, ils vérifient tous deux $q^{ij}(\iota(H)) = 0$. Cette propriété est caractéristique des indices de la forme $\iota(H,f) = \sum_{h \in H} f_h$, sommation de f sur H, c'est-à-dire, pour lesquels $\iota(H,f)$ est la *mesure* de H dans $\mathcal{P}(X)$ induite par les pondérations f_h. Comme il a été noté en 2.1.3.2., β, β', σ' et σ'' sont de ce type.

Proposition 4.1. Les trois conditions suivantes sont équivalentes pour un indice $\iota : \mathcal{H} \to \mathbb{R}^+$ vérifiant (C).

(Q forte) $(\forall H \in \mathcal{H})(\forall i,j \in I_H)\quad q^{ij}(\iota(H)) = 0$

(Σ1) $(\exists f : \mathcal{P}(X) \to \mathbb{R}^+ - \{0\})(\forall H \in \mathcal{H})\ \iota(H) = \sum_{h \in H} f_h$

(Σ2) $(\exists f : \mathcal{P}(X) \to \mathbb{R}^+ - \{0\})(\forall H, H' \in \mathcal{H})\ d_\iota(H,H') = \sum_{h \in H \Delta H'} f_h$.

Démonstration. L'équivalence des conditions (Σ1) (Σ2), et de la condition suivante : $\left(\forall H, H' \in \mathcal{P}^2(X)\right)$ $\iota(H \cup H') + \iota(H \cap H') - \iota(H) - \iota(H') = 0$, dont (Q forte) est une restriction, est classique dans les treillis booléens [cf. Barbut et Monjardet (1970), ch.IV ; Birkhoff (1967), ch.X]. D'où, par restriction à $(\mathcal{H},\subseteq)$, $(\Sigma 1) \Longleftrightarrow (\Sigma 2) \Rightarrow$ (Q forte).

(Q forte) $\Rightarrow (\Sigma 1)$. Posons, pour $i \in \mathcal{P}(X)$, et $\underline{n} > \underline{n}_i \geq 2$, $f_i = \iota(H_0 \cup \{i\}) - \iota(H_o)$, $f_X = \iota(H_o)$ et, pour $x \in X$, $f_x = 0$. Soit $H \in \mathcal{H}$, tel que $I_H = \{X, i, i_1, i_2, \ldots, i_{\gamma(H)-1}\}$, et soient les hiérarchies $H^{(o)} = H$, $H^{(1)} = H^{i_1}, \ldots, H^{(\underline{k})} = H - \{i_1, \ldots, i_{\underline{k}}\}$, pour $1 \leq \underline{k} \leq \gamma(H) - 1$. On a $H^{(\gamma(H)-1} = H_o \cup \{i\}$. Appliquons (Q forte) à la suite des quadrilatères $\left\{H^{(\underline{k})}, H^{(\underline{k})} - \{i\}, H^{(\underline{k}+1)}, H^{(\underline{k}+1)} - \{i\}\right\}$. On trouve :

$\iota(H) - \iota(H^i) = \iota(H^{(1)}) - \iota(H^{(1)i}) = \ldots = \iota(H_o \cup \{i\}) - \iota(H_o) = f_i$.

On remplace ensuite H et i par H^i et i_1, puis H^{ii_1} et i_2, etc., pour obtenir finalement $\iota(H) = \sum_{h \in H} f_h$. □

On notera $\sigma(H,f) = \sum_{h \in H} f_H$ la mesure de H dans $\mathcal{P}(X)$ définie par la mesure $\sigma(f)$ induite par f sur $\mathcal{P}^2(X)$. La condition $(\Sigma 1)$ équivaut à dire que $c^i(\iota(H))$ dépend de i mais non de H. On sera amené à considérer une mesure dans $\mathcal{P}(X)$ chaque fois que l'on voudra cette propriété.

On a ainsi $\beta(H) = \sigma(H, n(\underline{n} - n) + \underline{n}(\underline{n} - 1)$, $\beta'(H) = \sigma(H, n(n-1)/2)$, $\sigma''(H) = \sigma(H, \min(n - 1, \underline{n} - n))$, etc., et, pour les métriques associées, par exemple, $d_{\sigma''}(H,H') = \sum_{i \in H \Delta H'} , \min(\underline{n}_i - 1, \underline{n} - n_i)$.

Notons une propriété particulière de d_β. Reprenant les notations du par.2.2.2., et posant $d_\sigma^{xy}(H,H') = \underline{\ell}_{H,xHy} + \underline{\ell}_{H',xH'y} - 2\underline{\ell}_{H\cap H', xH\cap H'y}$ nombre d'éléments de $H \Delta H'$ contenant simultanément x et y, on décompose d_β :

$$d_\beta(H,H') = \sum_{i \in H\Delta H'} \underline{n}_i(\underline{n} - \underline{n}_i) = \sum_{x,y \in X,\ x \neq y} \left(d_\sigma^x(H,H') + d_\sigma^y(H,H') - 2d_\sigma^{xy}(H,H')\right).$$

Il serait aussi naturel de considérer, après la sommation, la moyenne arithmétique de f sur H, soit $\varphi(H,f) = \frac{1}{|H|} \sum_{h \in H} f_h$, indice linéaire correspondant à $a_{H,h} = \frac{1}{|H|}$, pour tous $h \in H \in \mathcal{H}$. On vérifie que cette fonction a satisfait (LD) et (LQ), mais le calcul de c^i, soit $c^i\left(\varphi(H,f)\right) = \frac{1}{|H|}\left(f_i - \varphi(H^i,f)\right)$, montre que $\varphi(f)$ n'est monotone que si f est constante sur $\mathcal{P}(X) - \{\emptyset, X\} - F$.

4.2. <u>Généralisation des éloignements de X : les coûts des hiérarchies</u>.

On suppose maintenant que $a_{H,h}$ est la longueur du chemin $L_{H,h}$ calculée à partir des longueurs d'arcs données par une fonction v :

$$\left\{(h_1, h_2) \in (\mathcal{P}(X))^2 \;/\; \emptyset \subset h_1 \subset h_2 \subseteq X\right\} \to \mathbb{R}^+, \text{ soit : } a_{H,h} = \sum_{(h_1,h_2)\in L_h} v(h_1, h_2)$$

On pose alors $\mu(H,v,f) = \sum_{h\in H} a_{H,h} f_h$. C'est le coût de la hiérarchie H pour la longueur v et la valeur des classes f. De tels coûts généralisent μ (avec $f = 1_{\mathcal{P}(X)}$ et v constante égale à 1) et σ (avec $f = 1_F$ et le même v).

Un problème classique en informatique théorique est la recherche de $H \in \mathcal{H}$ minimisant $\mu(H,v,f)$ sous diverses contraintes [cf. par exemple, Knuth (1973), Hu (1973)].

Notons, pour $h \subset i \in I_H$, h(i) l'élément unique de S_i contenant h.

On a, pour $h \in H$, dans le cas d'un coût :

$$\begin{aligned}(a_{H,h} - a^i_{H,h}) &= 0 && \text{si } h \not\subset i\\ &= v(p(i),i) + v(i,h(i)) - v(p(i),h(i)) && \text{si } h \subset i\end{aligned}$$

$$\begin{aligned}(a_{Hh} + a^{ij}_{H,h} - a^i_{H,h} - a^j_{H,h}) &= v(p(i),h(j)) + v(p(i),j) - v(i,h(j)) && \text{si } h \subset j \subset i = p(j)\\ &= 0 && \text{sinon}\end{aligned}$$

Les égalités (3.1.1.) et (3.1.2.) deviennent :

$$c^i(\mu(H,v,f)) = a_{H,i} f_i + \sum_{h\in S_i} \left(a_{H,h} - a^i_{H,h}\right)\left(\sigma(H_{h,f}) + \sum_{x\in h} f_x\right)$$

$$\begin{aligned}q^{ij}(\mu(H,v,f)) &= 0 && \text{si } j \not\subset i\\ &= (a_{H,j} - a^i_{H,j}) f_j && \text{si } j \subset i \text{ et } i \neq p(j)\\ &= (a_{H,j} - a^i_{H,j}) f_j\\ &+ \sum_{h\in S_j}\Big(v(p(i),h) + v(i,j) - v(p(i),j) - v(i,h)\Big)\left(\sigma(H_{h,f}) + \sum_{x\in h} f_x\right) && \text{si } i = p(j)\end{aligned}$$

D'où, en particulier, les valeurs de $c^i(\mu(H))$ et $q^{ij}(\mu(H))$ données en 2.1.2.3..

Les signes de c^i et q^{ij} sont fixés par les propriétés (EC) et EQ) :

(EC) $(\forall\, i,j,k \subseteq X)\ [\emptyset \subset k \subset j \subset i \Rightarrow v(i,j) + v(j,k) \geq v(i,k)]$

(EQ) $(\forall\, k,i,j,h \subseteq X)\ [\emptyset \subset k \subset j \subset i \subset h \Rightarrow v(h,k) + v(i,j) \geq v(h,j) + v(i,k)]$

(EC) est l'inégalité triangulaire la plus classique, assurant que la suppression de la classe $j \in I_H - \{X\}$ n'allonge aucun chemin. (EQ) est une condition quaternaire du type de (Q), mais sur les chaînes de $(\mathcal{P}(X), \subseteq)$. Les équivalences (EC) $\Longleftrightarrow$ (LC) et (EQ) $\Longleftrightarrow$ (LQ) sont immédiates pour un coût et, d'après la proposition 3.1., les conditions (EC) et (EQ) sur v garantissent que, pour toute f non négative, $\mu(H,v,f)$ est une valuation.

Exemples.

1) On se donne trois constantes $\lambda,\lambda',\lambda'' \in \mathbb{R}$ et deux fonctions $w,z : \mathcal{P}(X) \to \mathbb{R}^+$, respectivement fonction de sortie et d'entrée, et on pose pour

$\emptyset \subset j \subset i \subseteq X$, $v(i,j) = \lambda w_i + \lambda' z_j + \lambda''$. La condition (EQ) est alors vérifiée, avec de plus, l'égalité ainsi que la condition (EC) si l'on a pris λ'' suffisamment grand. Comme cas particuliers, on a $a_{H,h} = \ell_{H,h}$ pour $\lambda = \lambda' = 0$, $\lambda'' = 1$; pour $\lambda = 1$, $\lambda' = -1$, $\lambda'' = 0$, $a_{H,h} = w_X - z_h$ n'est fonction que de h.

2) Soit λ une constante suffisamment grande et $w : \mathcal{P}(X) \to \mathbb{R}^+$, monotone. Avec $v(i,j) = \lambda - w_i w_j$, on trouve, pour $k \subset j \subset i \subset h \subseteq X$, $v(i,j) + v(j,k) - v(i,k)$

$= \lambda - w_i w_j + w_j(w_i - w_j) = r - w_j w_k + w_i(w_k - w_j) > 0$ et $v(h,k) + v(i,j) - v(h,j) - v(i,k)$

$= (w_h - w_i)(w_j - w_k) \geq 0$

Cas des H-partitions : La condition (EQ) est toujours vérifiée par la restriction aux H-partitions (qui forment un sous-inf demi-treillis de $\mathcal{H}$). Pour celles-ci, on ne peut avoir pour $i,j \in I_H$, $j \subset i \subset X$ et q^{ij} est toujours nul.

L'expression de la distance $d_\mu(H(S),H(S'))$ est, pour v,f données :

$$d_{\mu(v,f)}(H(S),H(S')) = \sum_{i \in S \Delta S' - F} [v(X,i)f_i + \sum_{x \in i} (v(X,i) + v(i,x) - v(X,x)\, f_x]$$

L'expression entre crochets peut être vue comme le coût de la création ou de la suppression d'une classe (non singleton) de S, et la distance représente le coût minimal de la transformation de S en S' par suppressions ou adjonctions de classes.

5. LA CONDITION DE LA CLASSE INUTILE

5.1. La condition forte et le cas des indices linéaires.

On cherche un indice $\iota(H,f)$ évaluant l'intérêt classificatoire d'une hiérarchie H en fonction d'une valeur des classes $f : \mathcal{P}(X) \to \mathbb{R}^+$. Si $i,p(i) \in I_H$ sont telles que f_i et $f_{p(i)}$ sont proches, alors la présence de i dans H ajoute peu à l'intérêt de la présence de p(i). On va donc demander que $\iota(H)$ et $\iota(H^i)$ soient proches aussi, d'où la condition (CI forte) :

(CI forte) $\left(\forall H \in \mathcal{H},\ i \in I_H - \{X\}\right) \left(\forall f \in (\mathbb{R}^+)^{\mathcal{P}(X) - \{i\}}\right) \lim_{f_i \to f_{p(i)}} \iota(H,f) = \iota(H^i,f)$

Un affaiblissement de (CI forte) est la *condition de la classe inutile* (CI) :

$$\text{(CI)}\quad \left(\forall H \in \mathcal{H},\ i \in I_H - \{X\}\right) \left(\forall f \in (\mathbb{R}^+)^{\mathcal{P}(X)}\right)\quad f_i = f_{p(i)} \Rightarrow \iota(H^i,f) = \iota(H,f)\ .$$

(CI) exprime que pour $f_i = f_{p(i)}$ la classe i "inutile" peut être supprimée sans changement de la valeur de l'indice ι .

Proposition 5.1. Soit $\iota(H,f) = \sum_{h \in H} a_{H,h} f_h$ un indice linéaire sur $\mathcal{H}$. Les quatre conditions (CI forte), (CI) et (1), (2) suivantes sont équivalentes pour ι .

$$(1)\quad \left(\forall H \in \mathcal{H},\ i \in I_H - \{X\}\right) \left[a^i_{H,p(i)} = a_{H,p(i)} + a_{H,i}\right] \text{et} \left[\left(\forall h \in H - \{i,p(i)\}\right) a^i_{H,h} = a_{H,h}\right]$$

$$(2)\quad \left(\forall H \in \mathcal{H},\ i \in I_H - \{X\}\right) c^i\left(\iota(H,f)\right) = a_{H,i}\left(f_i - f_{p(i)}\right)\ ;$$

Démonstration : (1) $\Rightarrow$ (2) découle de l'expression (3.1.1.) de c^i ;
(2) $\Rightarrow$ (CI forte) $\Rightarrow$ (CI) sont évidents.

(CI) $\Rightarrow$ (1). $c^i(H,f)$ est une fonction affine de f_i et de $f_{p(i)}$ qui, pour que (CI) soit vraie, doit s'annuler pour $f_i = f_{p(i)}$, donc être de la forme $\lambda(a)(f_{p(i)} - f_i)$. Ceci entraîne, pour $h \in H - \{i,p(i)\}$, $a^i_{H,h} - a_{H,h} = 0$ et :

$\lambda(a) = -a_{H,i} = a_{H,p(i)} - a^i_{H,p(i)}$. □

Le calcul de q^{ij} pour ι linéaire vérifiant (CI) donne, à partir de (3.1.2.) :

$$q^{ij}\left(\iota(H,f)\right) = a_{H,j}\left(f_{p(i)} - f_j\right) \qquad \text{si } i = p(j)$$

$$= 0 \qquad \text{sinon.}$$

Si f est non croissante sur $(\mathcal{P}(X), \subseteq)$, $\iota(f)$ est non décroissant et vérifie $(\overline{Q})$ et dualement, non croissant et vérifiant (Q) si f est non décroissante. Les indices linéaires vérifiant (CI) ne sont pas des valuations sur $(\mathcal{H}_X, \subseteq)$.

5.2. Les fonctions a possibles.

Soit A l'ensemble des fonctions a telles que $\iota(H,f) = \sum_{h\in H} a_{H,h} f_h$ vérifie (CI).

Pour a donné, les coefficients $a_{H,x}$ sont invariants en H et on les supposera tous nuls. Pour $a, a' \in A, \lambda, \lambda' \in \mathbb{R}^+$, on a : $\lambda a + \lambda' a' \in A$. Les résultats suivants établissent que la dimension du vectoriel dont A est la partie non négative est au plus $2^{\underline{n}} - (\underline{n}+1)$.

Proposition 5.2. Soient $a \in A$ et $\underline{k}$ entier avec $2 \leq \underline{k} \leq \underline{n}-2$. Les $a_{H,h}$, pour H de niveau $\underline{k}$, sont déterminés par les $a_{H',h}$, pour H' de niveau $\underline{k}-1$ et $\underline{k}-2$.

Pour H de niveau $\underline{k}$, on peut trouver $h, i, j \in I_H$ tels que l'on a, soit $h = p(i)$ et $i = p(j)$, soit $h = p(i) = p(j)$.

Premier cas : des trois égalités $a_{H,i} + a_{H,j} = a^j_{H,i}$; $a_{H,h} + a_{H,i} = a^i_{H,h}$; $a_{H,h} + a_{H,i} + a_{H,j} = a^{ij}_{H,h}$ on tire $a_{H,h} = a^{ij}_{H,h} - a^j_{H,i}$; $a_{H,j} = a^{ij}_{H,h} - a^i_{H,h}$; $a_{H,i} = a^i_{H,h} + a^j_{H,i} - a^{ij}_{H,h}$.

Deuxième cas : on a : $a_{H,h} + a_{H,i} = a^i_{H,h}$; $a_{H,h} + a_{H,j} = a^j_{H,H}$; $a_{H,h} + a_{H,i} + a_{H,j} = a^{ij}_{H,h}$. D'où : $a_{H,i} = a^{ij}_{H,h} - a^j_{H,h}$; $a_{H,j} = a^{ij}_{H,j} - a^i_{H,h}$ et $a_{H,h} = a^i_{H,h} + a^j_{H,h} - a^{ij}_{H,h}$. □

Corollaire 5.3. Si $a \in A$, a est entièrement déterminée par les $a_{H,X}$, pour H de niveau 0 ou 1.

Pour $\alpha(H) = 0$, $H = H_o$, $I_H = \{X\}$ et la donnée de $a_{H,X}$ équivaut à celle de a_H. Pour $\alpha(H) = 1$, $H = H_o \cup \{i\}$, avec $\underline{n} > \underline{n}_i \geq 2$. On a $a_{H_o,X} = a^i_{H,X} = a_{H,i} + a_{H,X}$, d'où $a_{H,i} = a_{H_o,X} - a_{H,X}$. Ensuite, d'après la proposition précédente, on peut déterminer H pour $\alpha(H) = 2,3,\ldots,\underline{n}-2$. □

5.3. Sommes et moyennes hiérarchiques.

5.3.1. Définitions. Une *somme hiérarchique* de f sur H ∈ est un indice linéaire $\tau(H,f)$ vérifiant (CI) et : $(\forall H \in \mathcal{H}^m)$ $\tau(H,f) = \sigma(H,f) = \sum_{i\, I_H} f_i$. L'indice $\tau(H,f)$

coïncide donc avec la mesure $\sigma(H,f)$ quand H est binaire; les algorithmes de classification hiérarchique, ou les modèles de bifurcation donnent le plus souvent des hiérarchies binaires. Les conditions ci-dessus déterminent $\tau(f)$.

Proposition 5.4. La somme hiérarchique $\tau(H,f)$ a pour expression

$$\tau(H,f) = \sum_{i\in I} \underline{t}_i f_i$$

Démonstration. D'après les propriétés des t_i, $i\in I$, établies en HP 2.3.(prop.2.1.) $\Sigma \underline{t}_i f_i$ est bien une somme hiérarchique. L'équivalence (CI) $\Leftrightarrow$ (1) de la prop. 4.1. montre que a est entièrement déterminé par les $a_{H,h}$, $h \in H \in \mathcal{H}^m$, d'où l'unicité. □

Evaluer une hiérarchie H quelconque par $\tau(H,f)$ revient à pondérer chaque classe intérieure i par $\underline{t}_i = \underline{s}_i - 1$, donc à considérer que i représente t_i classes. Précisément, la prop. 4.3. dans HP 4.2. établit que $\underline{t}_i = |I'(i)|$ est le nombre de classes distinctes remplaçant i dans toute hiérarchie binaire H' telle que $H \subseteq H'$.

La *moyenne hiérarchique* $\Theta(H,f)$ de f sur H est donnée par $\Theta(H,f) = \frac{1}{\underline{n}-1}\tau(H,f) = \sum_{i\ I} \frac{\underline{t}_i}{\underline{n}-1} f_i$. C'est une moyenne pondérée des f_i, puisque $\sum_{i\in I} \underline{t}_i = \underline{n} - 1$.

On est amené à considérer, plutôt que f_i, le gain $f_X - f_i$ obtenu pour la classe i par rapport à l'ensemble non classifié X, d'où le gain hiérarchique $\sum_{i\in T} t_i (f_X - f_i) = \tau(H,f_X - f)$ et le gain hiérarchique moyen $\Theta(H,f_X - f)$. Si f est croissante, $\tau(H,f)$ et $\Theta(H,f)$ sont décroissants, et $\tau(H,f_X - f)$ et $\Theta(H,f_X - f)$ sont croissants, avec $\tau(H_o,f_X - f) = \Theta(H_o,f_X - f) = 0$. On a aussi $\Theta(H,f_X - f) = f_X - \Theta(H,f)$.

Une autre forme pour τ, se prêtant à des généralisations à des graphes quelconques (dont les graphes de couverture pour l'inclusion d'hypergraphes) est $\tau(H,f) = - f_X + \sum_{x\in X} f_x + \sum_{(h,h')\in U_H} (f_h - f_{h'})$.

5.3.2. Autres caractérisations axiomatiques de τ et de θ. τ et Θ peuvent être définis au moyen d'un nombre réduit d'axiomes. En particulier, une seule propriété fixe presque τ :

Proposition 5.5. Soit un indice $\tau'(H,f)$ vérifiant la condition

(Δ) $(\forall H \in \mathcal{H},\ i\in I_H - \{X\})\ t_i = 1 \Rightarrow c^i(\tau'(H,f)) = f_i - f_{p(i)}$

Alors, on a $\tau'(H,f) = \sum_{i\in I} \underline{t}_i f_i - (n-1)f_X + \tau'(H_o,f)$

Démonstration. Si H est binaire, on a $\tau'(H,f) - \tau'(H_o,f) = \sigma(H,f) - (\underline{n} - 1)f_X$. On le voit en enlevant successivement de H tous les éléments de I - {X}, en choisissant toujours un élément maximal.

Pour H non binaire, on considère H'binaire contenant H et la partition $\{I'(i)/i\in I_H\}$ de $I_{H'}$ (cf. HP 4.2.). On a $\tau'(H',f) - \tau'(H,f) = \sum_{i\in I_H} \sum_{j\in I'(i)} (f_j - f_i)$

$= \sigma(H',f) - \sum_{i\in I_H} \underline{t}_i f_i$, d'où $\tau'(H,f) - \tau'(H_o,f)$

$= (\tau'(H,f) - \tau'(H',f)) - (\tau(H_o,f) - \tau'(H',f))$

$= \sum_{i\in I_H} \underline{t}_i f_i - (\underline{n} - 1)f_X$. □

La condition (Δ) part du souhait que, pour $H,H' \in \mathcal{H}$ et $i \in I_H \cap I_{H'}$ tels que $p_H(i) = p_{H'}(i)$, on ait $c^i(\tau'(H,f)) = c^i(\tau'(H',f))$, indépendamment des autres propriétés de H et H'. Mais cette condition ne peut, sans contradiction, être vérifiée pour n'importe quels H,H',i vérifiant les conditions ci-dessus.

La proposition précédente montre qu'on obtient un bon résultat en se restreignant aux classes i qui ont deux successeurs dans H. Il reste à fixer les quantités $\tau'(H_o,f)$.

On peut par exemple le faire des deux façons suivantes :

Corollaire 5.6. $\tau(H,f)$ est le seul indice $\tau'(H,f)$ vérifiant (Δ) et

1) $\tau'(H_o,f)$ ne dépend que de f_X;

2) pour f constante égale à λ, H binaire, on a $\tau'(H,\lambda) = |I_H|\lambda = (\underline{n} - 1)\lambda$.

(2) implique avec (Δ) $\tau'(H_o,\lambda) = (\underline{n} - 1)\lambda$, d'où, avec (1) quelle que soit f, $\tau'(H_o,f) = (\underline{n} - 1) f_X$ et $\tau'(H,f) = \tau(H,f)$ pour $H \in \mathcal{H}$. □

Corollaire 5.7. $\theta(H,f)$ est le seul indice $\theta'(H,f)$ vérifiant simultanément :

1) $(\exists K \in \mathbb{R}^+)(\forall H \in \mathcal{H}, i \in I_H - \{X\})\ \underline{t}_i = 1 \Rightarrow c^i(\theta'(H,f)) = \kappa(f_i - f_{p(i)})$

2) $\theta'(H_o,f) = f_X$

3) pour f constante égale à λ, $H \in \mathcal{H}^m$, $\theta'(H,f) = \lambda$

D'après la proposition 4.5., (1) implique $\theta'(H,f) = \sum_{i\in I} \underline{t}_i f_i - \kappa(\underline{n} - 1)f_X + \theta'(H_o,f)$

(2) fixe $\theta'(H_o,f)$ et (3) donne $\kappa = \dfrac{1}{\underline{n} - 1}$. □

5.4. Sous hiérarchies et partitions associées à une hiérarchie.

5.4.1. Sous hiérarchies. Les fonctions τ et θ ont été élaborées en vue de la comparaison des hiérarchies. $\tau(f_X - f)$ et $\theta(f_X - f)$ en particulier ont un sens précis et, s'annulant pour $H = H_o$, donnent de bonnes évaluations de l'apport classificatoire, relativement à f, de H par rapport à la donnée de X. Elles ont de plus de bonnes propriétés pour la simplification des hiérarchies. Soit $H' \in \mathcal{H}$ et $J \subseteq I_{H'} - \{X\}$ tel que le graphe partiel engendré par J dans $G_{H'}$ est connexe (fig.2). J a un élément maximum i et en posant $H = H' - J \cup \{i\}$, on a en reprenant la démonstration de la proposition 5.5. :

$$\tau(H',f) = \tau(H,f) - \sum_{j\in J} t_j(f_i - f_j) \qquad (5.4.1.)$$

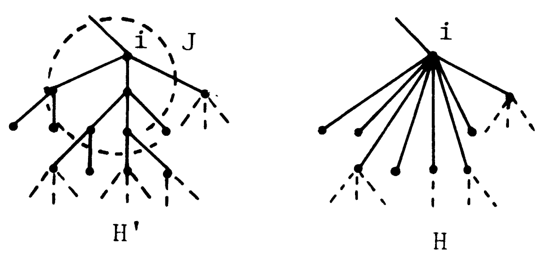

fig.2

Pour H binaire, $H' \subseteq H$, on retrouve :

$$\tau(H',f) = \tau(H,f) - \sum_{i\in I_H} \sum_{j\in I'(i)} t_j(f_i - f_j) \qquad (5.4.2.)$$

Cette égalité peut servir à borner $|\tau(H',f) - \tau(H,f)|$ quand H est une simplification de H', par exemple dans le cas où H' est inconnu. Reprenons (5.4.1.) dans le cas où $J = H_i - F$. On obtient une *additivité des gains hiérarchiques* quand on substitue à une classe intérieureminimale i une hiérarchie sur i :

$$\tau(H',f_X - f) = \tau(H,f_X - f) + \tau_i(H_i',f_i - f) \qquad (5.4.3.)$$

τ_i est l'indice somme hiérarchique pour les hiérarchies éléments de $\mathcal{H}_i$.

Pour le gain moyen hiérarchique θ, H_i représente $\underline{n}_i - 1$ classes parmi $\underline{n} - 1$. On a bien, avec $H = (H' - H_i') \cup \{i\} \cup F$, $i \in I_{H'}$, l'additivité pondérée :

$$\theta(H',f_X - f) = \theta(H,f_X - f) + \frac{\underline{n}_i - 1}{\underline{n} - 1}\,\theta(H_i',f_i - f) \qquad (5.4.4.)$$

Exemple : On considère une hiérarchie binaire équilibrée H', avec $\underline{n} = 2^{10} = 1024$, et la hiérarchie $H_1 - (H' - H_i') \cup \{i\} \cup F$, avec $i \in S$, donc $\underline{n}_i = 512$. H_1 est

donc obtenu par la suppression de 510 (sur 1023) classes intérieures de H'. On a, avec $f = n$, $\tau(H',\underline{n} - n) = 1\ 037\ 312$, et :
$\tau(H',\underline{n} - n) - \tau(H_1,\underline{n} - n) = \tau(H_i,512 - n) = 257\ 024$.

Si on prend plutôt H_2 obtenue en enlevant toutes les classes intérieures de cardinal au plus égal à 32 (soit 992 classes), on trouve, si H_2' est une HBE correspondant à $\underline{n} = 2^6 = 64$:

$$\tau(H',\underline{n} - n) - \tau(H_2,\underline{n} - n) = 16\,\tau\,(H_2',64 - n) = 58\ 368.$$

Pour $f = n$, la seconde simplification est de loin préférable à la première.

Cet exemple introduit le problème de la *simplification optimale d'une hiérarchie* relativement à une valeur de classes f. La recherche du minimum de $|\tau(H',f) - \tau(H,f)|$ (sous des contraintes convenables, par exemple pour $|H' - H|$ fixé) est une façon d'aborder ce problème, mais sa solution générale ne semble pas immédiate. L'égalité (5.2.1.) peut servir de point de départ au moins dans la pratique. L'heuristique "gloutonne" consistant à itérer la suppression de la classe $i \in I - \{X\}$ maximisant $\underline{t}_i\left(f_{p(i)} - f_i\right)$ ne donne pas en général un optimum global.

5.4.2. Expressions de τ et θ à partir de la partition B de X en feuilles de même père. On considère la partition B de F (ou de X) en feuilles de même père (HP 2.3.) et pour $i \in I$, les nombres $\underline{n}_i' = |F \cap S_i|$, $\underline{t}_i' = \underline{t}_i - \underline{n}_i'$.

$\tau(H,f)$ se met sous la forme :

$$\tau(H,f) = \sum_{i\in I} t_i' f_i + \sum_{b\in B} \underline{n}'_{p'(b)} f_{p'(b)} = \sum_{i\in I} \underline{t}_i' f_i + \sum_{b\in B} \underline{n}_b f_{p'(k)}$$

Si $p'(b) = b$, pour tout $b \in B$ (ou de manière approchée si $\underline{n}_b$ est grand devant $\underline{n}_{p'(b)} - \underline{n}_b$), $\tau(H,f)$ s'écrit comme somme d'un terme hiérarchique et d'un terme ne dépendant que de la partition B :

$$\tau(H,f) = \sum_{i\in I} \underline{t}_i' f_i + \sum_{b\in B} \underline{n}_b f_b \qquad (5.4.5.)$$

Pour la moyenne hiérarchique θ, en posant $m_i = \underline{n}_i / \underline{n}$, la forme générale est :

$$\theta(H,f) = \sum_{i\in I} \frac{\underline{t}_i'}{\underline{n} - 1} f_i + \frac{\underline{n}}{\underline{n} - 1} \sum_{b\in B} m_b f_{p'(b)}$$

Et, dans les mêmes conditions que (5.4.5.), on a, avec $\overline{f}_B = \Sigma\, m_b f_b$:

$$\theta(H,f) = \frac{\underline{n}}{\underline{n} - 1} \overline{f}_B + \sum_{i\in I} \frac{\underline{t}_i'}{\underline{n} - 1} f_i \qquad (5.4.6.)$$

Dans le cas d'une H-partition H(S), on a d'abord :

$$\theta\Big(H(S),f\Big) = \sum_{i\in I} \underline{t}_i\, f_i = \frac{\underline{s}-1}{\underline{n}-1} f_X + \sum_{h\in X} \frac{\underline{n}_h - 1}{\underline{n}-1} f_h \qquad (5.4.7.)$$

$$\theta\Big(H(S),f_X - f\Big) = \sum_{h\in S} \frac{\underline{n}_h - 1}{\underline{n}-1} (f_X - f) \qquad (5.4.8.)$$

Cette dernière expression s'accorde bien avec l'idée que seules les classes autres que X ou {x}, $x\in X$, ont un intérêt classificatoire.

Pour S sans singletons, on obtient à partir de (5.4.6.) ou de (5.4.7.) :

$$\theta\Big(H(S),f\Big) = \frac{\underline{s}-1}{\underline{n}-1} f_X - \frac{1}{\underline{n}-1} \sum_{h\in S} f_h + \frac{1}{\underline{n}-1} \overline{f}_S \qquad (5.4.9.)$$

5.4.3. La "dendropie" $\delta(H)$. L'indice $\delta(H)$ présenté en 2.1.4.2. est aussi égal à $\theta(H, f_X - f)$ pour $f_i = \log_2 \underline{n}_i$, quantité d'information de Wiener-Shannon nécessaire pour identifier un élément de i. On définit axiomatiquement $\delta(H)$ en considérant $\log_2 \underline{n}_{p(i)} - \log_2 \underline{n}_i = \log_2 \frac{\underline{n}_{p(i)}}{\underline{n}_i}$, quantité d'information apportée par la restriction de p(i) à i.

Proposition 5.8. $\delta(H)$ est le seul indice $\delta : \mathcal{H} \to \mathbb{R}$ vérifiant simultanément :

(1) $\Big(\forall H \in \mathcal{H},\ i \in I_H - \{X\}\Big)\ \underline{t}_i = 1 \Rightarrow c^i\Big(\delta(H)\Big) = \frac{1}{\underline{n}-1} \log_2 \frac{\underline{n}_{p(i)}}{\underline{n}_i}$

(2) $\delta(H_o) = 0$

Démonstration. De (1) on déduit comme dans la proposition 5.5. (en prenant $f_i = \log_2 \frac{\underline{n}}{\underline{n}_i}$) $\delta(H) = \delta(H_o) + \sum \frac{\underline{t}_i}{\underline{n}-1} \log_2 \frac{\underline{n}}{\underline{n}_i}$, d'où le résultat. □

La pondération $\frac{1}{\underline{n}-1}$ se justifie par $|I_H| = \underline{n}-1$, pour $H \in \mathcal{H}^m$.

Comme l'entropie de Shannon pour les partitions, δ est un gain d'information moyen. A ce titre, c'est un analogue pour les hiérarchies de l'entropie des partitions. Notamment, l'égalité (5.3.4.) donne une propriété d'additivité pondérée proche de celle de l'entropie [cf. par exemple Picard (1972), Guiasu (1977)]. Pour $H' \in \mathcal{H}$, $i \in I_{H'}$, $I_H = I_{H'} - H_i \cup \{i\}$:

$$\delta(H') = \delta(H) + \frac{\underline{n}_i - 1}{\underline{n}-1} \delta_i(H'_i)$$

Les formes (5.4.6.) et (5.4.7.) montrent de plus que si $\underline{n}$ croît en laissant invariant $|I|$ et les $\frac{\underline{n}_i}{\underline{n}}$, $\delta(H)$ tend vers l'entropie $\chi(B)$ de la partition B. Pour une H-partition H(S), on a :

$$\delta\Big(H(S)\Big) = \frac{\underline{n}}{\underline{n}-1} \quad \chi(S) - \frac{1}{\underline{n}-1} \sum_{h\in S} \log_2 \frac{\underline{n}}{\underline{n}_h} \quad .$$

Alors si H' est une hiérarchie binaire équilibrée que l'on simplifie par la suppression des $2^{\underline{k}} - 2^{\underline{k}-\underline{k}'+1}$ classes intérieures de cardinal inférieur à $2^{\underline{k}'}$ (avec $2 \le \underline{k}' \le \underline{k}-1$) pour obtenir une hiérarchie simplifiée $H_{(\underline{k}')}$ on a, pour $\underline{k}$ assez grand :

$$\delta(H') - \delta(H_{(\underline{k}')}) = 2^{\underline{k}-\underline{k}'} \frac{2^{\underline{k}'}-1}{2^{\underline{k}}-1} \left(\frac{2^{\underline{k}'}}{2^{\underline{k}'}-1}\ \underline{k}'-2\right) \simeq \underline{k}'-2$$

Si par contre, on prend $\underline{n}_i = 2^{\underline{k}'}$ et $H = (H' - H_i) \cup \{i\} \cup F$, on trouve :

$$\delta(H') - \delta(H) = \frac{2^{\underline{k}'}-1}{2^{\underline{k}-1}} \delta_i(H_i) \simeq \frac{\underline{k}'-2}{2^{\underline{k}-\underline{k}'}} \quad .$$

Dans l'exemple du par.5.4.1.,on trouve : $\delta(H') \simeq 8{,}01$; $\delta(H') - \delta(H_1) \simeq 3{,}505$; $\delta(H') - \delta(H_2) \simeq 4{,}035$.

La simplification H_1 fait perdre moins d'information que H_2 (mais en supprimant moins de classes). Si on considère H_3 obtenu par suppression seulement des 960 classes de cardinal au plus égal à 16, H_3 est meilleure que H_1 :

$$\delta(H') - \delta(H_3) \simeq 3{,}0655.$$

5.4.4. La moyenne hiérarchique des effectifs.

C'est l'indice ξ qui est défini par $\xi(H) = \xi(H) = \theta\left(H, \frac{\underline{n}}{\underline{n}}\right) = \sum_{i\in I} \frac{\underline{t}_i\, \underline{n}_i}{\underline{n}(\underline{n}-1)}$.

Pour H binaire, $\xi = \sigma / \underline{n}(\underline{n}-1)$, d'où les bornes $(\underline{n}+2)/2\underline{n} \le \xi(H) \le \log_2 \underline{n} / \underline{n}-1$.

L'exemple précédent illustre le fait que ξ donne plus d'importance relative que δ aux classes de cardinal élevé. Pour une H-partition H(S), on a :

$$\xi\Big(H(S)\Big) = \frac{\underline{s}-1}{\underline{n}-1} + \sum_{h\in S} \frac{\underline{n}_h\,(\underline{n}_h-1)}{\underline{n}\,(\underline{n}-1)} = \frac{\underline{s}-1}{\underline{n}-1} + \frac{\beta'(H(S))}{2\underline{n}(\underline{n}-1)}$$

β' est l'indice de liaison des paires d'éléments de X du par.2.1.3.2. et ξ apparaît comme une sorte d'intermédiaire entre σ et β'. Pour $\underline{s}$ fixé, ξ ne dépend que du nombre de paires liées dans la partition S.

BIBLIOGRAPHIE

ADAMS E.N. III, "Consensus techniques and the comparison of taxonomic trees", *Syst. Zool.*, 21 (1972), 390-397.

BANDELT H.J., BARTHELEMY J.P., "Medians in median graphs", *Discrete Applied Math.*, 8 (1984), 131-142.

BARBUT M., MONJARDET B., *Ordre et classification. Algèbre et Combinatoire*,, tome 2, Hachette, Paris, 1970.

BARTHELEMY J.P., "Remarques sur les propriétés métriques des ensembles ordonnés", *Math. Sci. hum.*, 61 (1978), 39-60.

BARTHELEMY J.P., *Propriétés métriques des ensembles ordonnés. Comparaison et agrégation des relations binaires*, thèse, Université de Franche-Comté, Besançon, 1979a.

BARTHELEMY J.P., "Caractérisations axiomatiques de la distance de la différence symétrique entre les relations binaires", *Math. Sci. hum.*, 67 (1979b), 85-113.

BARTHELEMY J.P., FLAMENT C., MONJARDET B., "Ordered sets and social sciences", *in Proc. Symposium on ordered sets* (Banff 1981), Reidel, Dordrecht, 1982.

BARTHELEMY J.P., LECLERC B., "Median consensus for tree representations of dichotomous variables", en préparation, (1986).

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.

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.

BARTHELEMY J.P., LECLERC B., MONJARDET B., *On the use of ordered set in problems of comparison and consensus of classifications*, rapport CAMS, P.018, 1986.

BARTHELEMY J.P., MONJARDET B., "The median procedure in cluster analysis and social choice theory", *Mathematical Social Sciences*, 1 (1981), 235-267.

BIRKHOFF G., *Lattice theory*, 3rd edition, American Mathematical Society, Providence, N.J., 1967.

BERGE C., *Graphes et hypergraphes*, Dunod, Paris, 1970. *Graphs and hypergraphs*, North-Holland, Amsterdam, 1973.

BOBISUD H.M., BOBISUD L.E., "A metric for classifications", *Taxon*, 21 (1972), 607-613.

BOGART K.P., "Preference structure I : distance between transitive preference relations", *J. Math. Sociol.*, 3 (1973), 49-67.

BOORMAN S.A., OLIVIER D.C., "Metrics on spaces of finite trees", *J. Math. Psychol.*, 10 (1973), 26-59.

BORDES G., "Métriques bornées définies par des valuations sur un demi-treillis", *Math. Sci. hum.*, 56 (1976), 89-95.

BURGE W.H., "Sorting, trees and measures of order", *Information and control*, 1 (1958), 181-197.

CHANDON J.L., PINSON S., *Analyse typologique*, Masson, Paris, 1980.

COLLESS, "Congruence between morphometric and allozyme data for Menidia species : a reappraisal", *Syst. Zool.*, 29 (1980), 288-299.

COMYN G., Van DORPE J.C., "Valuation et semi-modularité dans les demi-treillis", *Math. Sci. hum.*, 56 (1976), 63-75.

DAY W.H.E., MCMORRIS F.R., *A formalization of consensus index methods*, Technical Report 8405, Memorial University of Newfoundland, St John's, Nfld, 1984.

DUFFUS O., RIVAL I., "Path length in the covering graph of a lattice", *Discrete Math.*, 19 (1977), 139-158.

FARRIS J.S., "A successive approximations approach to character weighting", *Syst. Zool.*, 18 (1969), 374-385.

FINDEN C.R., GORDON A.D., "Obtaining common pruned trees", *J. of Classification*, 2 (1985), 255-276.

GREEN C.D., "A path entropy function for rooted trees", *J.A.C.M.*, 20 (1973), 378-384.

GRIMONPREZ G., Van DORPE J.C., "Distance définie par une application monotone sur un treillis", *Math. Sci. hum.*, 56 (1976), 47-62.

GUIASU S., *Information theory with applications*, Mc Graw-Hill, New York, 1977.

HARARY F., *Graph theory*, Addison-Wesley, Reading, Mass., 1969.

HARDING E.S., "The probabilities of rooted tree-shapes generated by random bifurcations", *Adv. Appl. Prob.*, 3 (1971), 44-77.

HARTIGAN J.A., "Representation of similarity matrices by trees", *J. Amer. Statist. Ass.*, 62 (1967), 1140-1158.

HARTIGAN J.A., *Clustering algorithms*, Wiley, New York, 1975.

HASKINS L., GUDDER S., "Height on posets and graphs", *Discrete Math.*, 2 (1972), 357-382.

HU T.C., "Some results and problems in binary trees", *in* R. RUSTIN, ed., *Combinatorial algorithms*, Algorithmic Press, New York, 1973, 11-15.

HUBERT L.J., BAKER F.B., "The comparison and fitting of given classification schemes", *J. Math. Psychol.*, 16 (1977), 233-253.

JAMBU M., "Quelques critères de comparaison des hiérarchies indicées produites en classification automatique", *Consommation*, (1975) n°1, 55-85.

JANVIER Ph., TASSY P., THOMAS H., "Le cladisme", *La Recherche*, 117 (1980), 1396-1406.

KNUTH D.E., *The art of computer programming*. Vol.3 : *Sorting and searching*, Addison-Wesley, Reading, Mass., 1973.

LECLERC B., "Description combinatoire des ultramétriques", *Math. Sci. hum.*, 73 (1981), 5-37.

LECLERC B. , *Description, évaluation et comparaison des hiérarchies de parties*, rapport CAMS-EHESS, 1982.

MCMORRIS F.R., NEUMANN D.A., "Consensus functions defined on trees", *Math. Social Sciences*,4 (1983), 131-136.

MARGUSH T., "Distances between trees", *Discrete Applied Math.*, 4 (1982), 281-290.

MARGUSH T., MCMORRIS F.R., "Consensus n-trees", *Bull. of Math. Biology*, 43 (1981), 239-244.

MICKEVICH M.F., "Taxonomic congruence", *Syst. Zool.*, 27 (1978), 143-158.

MONJARDET B., "Metrics on a partially ordered set. A survey", *Discrete Math.*, 35 (1981), 173-181.

MORICONI F. RIZZI B., "Distropy in the dichotomic trees", *Boll. Un. Mat. Ital.*, Ser. V, XIV-A n°1 (1978), 180-186.

NELSON G., "Cladistic analysis and synthesis : principles and definitions, with a historical note on Adanson's Familles des plantes (1763-1764)", *Syst. Zool.*, 28 (1979), 1-21.

NELSON G., PLATNICK N., *Systematics and biogeography. Cladistics and vicariance*, Columbia University Press, New York, 1981.

NEUMANN D.A., "Faithful consensus methods for n-trees", *Math. Biosci.*, 63 (1983), 271-287.

NEUMANN D.A., NORTON V., *On lattice consensus methods*, Bowling Green State University, Bowling Green, 1985a.

NEUMANN D.A., NORTON V., *Consensus on partitions with applications to other structures*, Bowling Green State University, Bowling Green, 1985b.

PARLEBAS P., "Centralité et compacité d'un graphe", *Math. Sci. hum.*, 39 (1972), 5-26.

PATIL G.P., TAILLIE C., "An overview of diversity", *in* : GRASSLE J.F. et al, eds, *Ecological diversity in theory and practice*, (International Cooperative Publ. House, Fairland, MD), 1979, 3-27.

PHIPPS J.B., "Dendrogram topology", *Syst. Zool.*, 20 (1971), 306-308.

PICARD C.F., *Graphes et questionnaires*. t.2 : *Questionnaires*, Gauthier-Villars, Paris, 1972. *Graphs and questionnaires*, North-Holland, Amsterdam, 1980.

ROBINSON D.F., FOULDS L.R., "Comparison of phylogenetic trees", *Math. Biosci.*, 53 (1981), 131-147.

ROHLF F.J., "Consensus indices for comparing classifications", *Math. Biosci*, 59 (1982), 131-144.

SCHADER M., "Hierarchical analysis : classification with ordinal object dissimilarities", *Metrika*, 27 (1980), 127-132.

SCHUH R.T., FARRIS J.S., "Methods for investigating taxonomic congruence and their application to the Leptopodomorpha", *Syst. Zool.*, 30 (1981), 331-351.

SCHUH R.T., POLHEMUS J.T., "Analysis of taxonomic congruence among morphologigal, ecological and biogeographic data sets for the Leptopodomorpha (Hemiptera)", *Syst. Zool.*, 29 (1980), 1-26.

SOKAL R.R., ROHLF F.J., "The comparison of dendrograms by objective methods", *Taxon*, 11 (1962), 33-39.

SOKAL R.R., ROHLF F.J., "Taxonomic congruence in the Leptopodomorpha Reexamined", *Syst. Zool.*, 30 (1981), 309-325.

STINEBRICKNER R., "s-consensus trees and indices", *Bull. of math. Biology*, 46 (1984a), 923-935.

STINEBRICKNER R., "An extension of intersection methods from trees to dendrograms", *Syst. Zool.*, 33 (1984b), 381-386.

WATERMAN M.S., SMITH T.F., "On the similarity of dendrograms", *J. Theor. Biology*, 73 (1978), 789-800.

WILLIAMS W.T., CLIFFORD H.T., "On the comparison of two classifications on the same set of elements", *Taxon*, 20 (1971), 519-522.

GLOSSAIRE

On rappelle que HP 2.1. renvoie au paragraphe 2.1. de l'article précédent, signalé dans l'introduction. La lettre r est ajoutée lorsque le terme est rappelé à la fin de l'introduction du présent article.

arbres associés à une hiérarchie	HP 2.1.r, HP 2.2.r, HP 4.1.
arbre minimum	HP 2.7.
classe intérieure	HP 1.1.r
condition du quadrilatère	1.1.2, 3.3.
- de la classe inutile	5.
couverture	HP 1.2.r
demi-treillis	HP 1.2.
- à médianes, semi-modulaire	HP 1.3.
éloignement	1.2.4.
feuille	HP 1.1.r
forme	HP 4.1., 1.1.3.
graduation	HP 1.3.r
graphe d'échanges des hiérarchies binaires	HP 2.5.
hiérarchie	HP 1.1.r
- binaire	HP 2.4.2.r
- binaire équilibrée	HP 2.4.2., 1.1.4.
H-partition	HP 2.4.1.r
indice	1.1.
- linéaire	3.2.
niveau	HP 1.3.r
partage associé à une hiérarchie	HP 4.2.
peigne	HP 2.4.2., 1.1.4.
père	HP 1.1.r
préordonnance hiérarchique	HP 2.7.1.
quadrilatère	HP 1.3.r
valuation	1.1.2.

NOTATIONS

$a, a_{H,h}$	3.2.	α	HP 2.7.2., 2.1.4.2.
B	HP 2.3.r	β, β'	2.1.3.2.
c^i	1.1.	γ	HP 1.3.r, 2.1.1.
d_ι	1.1.2.	δ	2.1.4.2.
F	HP 1.1.	ε	HP 2.6., 2.1.4.1.
f	3.1.	η	2.1.3.1.
G_H	HP 2.1.	$\theta(H,f)$	5.3.
$\mathcal{H}, \mathcal{H}_X$	HP 1.2.r	ι	1.1.
$\mathcal{H}(k)$	HP 1.3.r	$\iota^p, \iota^e, \iota_{min}, \iota_{max}$	1.1.4.
$\mathcal{H}^m$	HP 2.4.2.r	ι^d	1.2.4.
H, h, H_h	HP 1.1.r	μ	2.1.2.3.
H^i, H^{ij}, H_o	HP 1.2.r	$\mu(H,v,f)$	4.2.
$H(S)$	HP 2.4.1.	ν, ν'	2.1.4.3.
xHy	HP 2.7.1., 2.1.3.2.	ξ	5.4.4.
I, I_H, i	HP 1.1.r	π	2.1.4.2.
$L, \underline{\ell}$	HP 2.1.r	ρ	2.1.2.1.
$\underline{n}, \underline{n}_h$	HP 1.1.r	σ	2.1.2.2.
n	3.1.	σ', σ''	2.1.3.3.
$p(h)$	HP 1.1.r	$\sigma(H,f)$	4.1.
$Q^{ij}(H)$	HP 1.3.r	$\tau(H,f)$	5.3.
q^{ij}	1.1.2.	χ	5.4.3.
$\underline{r}_H$	2.1.3.2.	$\prec$	HP 1.2.r
$S, S_i, \underline{s}, \underline{s}_i$	HP 1.1.r		
$\underline{t}_i$	HP 2.3.r		
U_H	HP 2.1.r		
$\hat{U}_H$	HP 2.6., 2.2.3.		