N° d'ordre : 618

UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE

THESE

présentée à

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE

pour obtenir

LE TITRE DE DOCTEUR DE 3ème CYCLE

(Informatique)

par

Georges GRIMONPREZ

APPLICATIONS DES THEORIES DE L'INFORMATION ET DES QUESTIONNAIRES AUX PROBLEMES DE CLASSIFICATION

Thèse soutenue le 14 décembre 1976, devant la Commission d'Examen

MEMBRES DU JURY
M. P. POUZET, Président
M. C.F. PICARD, Rapporteur
MM. G. JACOB
J. LOSFELD, Examinateurs

Je remercie Monsieur le Professeur POUZET pour le rôle qu'il a joué dans mon orientation actuelle en Informatique, pour l'intérêt qu'il a porté à mon travail et pour l' honneur qu'il me fait en acceptant la présidence de ce Jury.

Monsieur Claude PICARD, directeur de recherche au C.N.R.S., a bien voulu prendre la responsabilité de la direction de ce travail. Les membres de son équipe et lui-même m'ont apporté une contribution importante grâce à leurs remarques et à leurs critiques. Qu'ils trouvent ici l'expression de ma reconnaissance.

Je remercie également Monsieur le Professeur JACOB pour le caractère judicieux de ses critiques et pour avoir accepté de participer à ce Jury.

Monsieur Joseph LOSFELD m'a permis de débuter ces travaux, et ses conseils ont été, sur le plan amical et professionnel, déterminants dans les recherches que j'ai entreprises. Je tiens à rendre hommage à son obstination et à son dynamisme qui m'ont aidé à mener ce travail jusqu'à son terme.

J'ai toujours reçu de Monsieur le Professeur J. KAMPE DE FERIET des conseils précieux pour lesquels je lui dédie mes plus sincères remerciements. L'intérêt qu'il a porté à mon travail, en toutes circonstances, m'a fortement encouragé dans la direction choisie.

Ma gratitude ira également à mes amis G. COMYN et B. LEFEBVRE pour leur aide efficace : le premier dans la Théorie des Ensembles Ordonnés, le second sur l'Analyse des Données Statistiques.

Enfin, mes remerciements iront à Madame DESCARPENTRIES qui a tapé cette thèse avec compétence et à tous ceux qui ont permis la réalisation matérielle de cet ouvrage.

TABLE DES MATIERES

INTRODUCTION

I.- TREILLIS ET INFORMATION GENERALISEE

I.1.- RAPPELS.
I.2.- PERTE ET GAIN D'INFORMATION DANS UN TREILLIS.
I.3.- ÉTUDE DES PROPRIÉTÉS DE δ^+, δ^- ET δ.
I.4.- PROPRIÉTÉS DES APPLICATIONS J_R ET J_T
I.5.- EXPRESSION DES QUANTITÉS δ, J_R ET J_T.
I.6.- APPLICATION : RÉSUMÉS ET INFORMATIONS.

II.- INFORMATIONS ET QUESTIONNAIRES

II.1.- RAPPELS.
II.2.- INFORMATION GÉNÉRALISÉE SUR LE TREILLIS DES PARTITIONS D'UN ENSEMBLE Ω.
II.3.- ARBORESCENCE DE L'ORDRE DES QUESTIONS ASSOCIÉ À UN QUESTIONNAIRE.
II.4.- INFORMATIONS ASSOCIÉES À UN CHEMIN DE L'ARBORESCENCES DE L'ORDRE DES QUESTIONS.
II.5.- PROPRIÉTÉ DE LOCALISATION ET INFORMATIONS ASSOCIÉES À UN CHEMIN DE A_Q.
II.6.- LIEN ENTRE LES QUESTIONNAIRES ET NOS RÉSULTATS.
II.7.- APPLICATION AUX INFORMATIONS NE VÉRIFIANT PAS LA PROPRIÉTÉ DE LOCALISATION.

III.- APPLICATIONS AUX PROBLEMES DE CLASSIFICATION.

III.1.- PRINCIPES GÉNÉRAUX DES APPLICATIONS.
III.2.- ALGORITHME DE CONSTRUCTION DE CLASSES DE Ω.
III.3.- ALGORITHME DE CLASSIFICATION HIÉRARCHIQUE INDICÉE DE CLASSES DE Ω.
III.4.- QUESTIONNAIRE ASSOCIÉ À UNE HIÉRARCHIE INDICÉE.

INTRODUCTION

L'utilisation, dans la Théorie des Questionnaires, de mesures d'information nous a amené à étudier le gain d'information apporté par une question d'un questionnaire Q . Nous nous sommes intéressés plus particulièrement aux questionnaires posés sur un ensemble Ω et induisant sur celui-ci une structure de partitions : une question de Q est associée à des partitions de Ω et le gain d'information d'une question est relatif au passage d'une partition de Ω à une partition plus fine. Ceci nous a conduit à définir dans le treillis T des partitions de Ω la notion de séquence d'éléments comparables (s.e.c.) de T .

Nous plaçant alors dans le cadre général d'un treillis T quelconque muni d'une information généralisée dont nous utilisons uniquement les propriétés intrinsèques de l'axiome de monotonicité, nous évaluons les gains ou pertes d'information liés à une s.e.c. de T . De ces notions résultent la définition d'une distance sur T et d'une application J_R de $T \times T$ dans $\mathbb{R}^+$ mesurant le degré de ressemblance entre deux éléments quelconques du treillis.

Un questionnaire est associé naturellement à un ensemble de s.e.c. possibles. L'étude de l'importance prise par l'ordre dans lequel sont posées les questions nous a amené à étudier les conséquences de l'axiome de localisation énoncé par B. FORTE et N. PINTACUDA. Nous avons, à cette occasion, retrouvé, dans des définitions particulières d'information, les résultats classiques de la Théorie des Questionnaires.

Enfin, en utilisant les quantités ainsi définies, nous proposons un ensemble d'algorithmes adaptés à des problèmes de classification automatique relatifs à la détermination de classes homogènes de Ω .

CHAPITRE I

-=-=-=-=-=-=-=-=-=-

TREILLIS ET INFORMATION GENERALISEE

-=-=-=-=-=-=-=-=-=-

I.- TREILLIS ET INFORMATION GENERALISEE

Sur un treillis T muni d'une information généralisée J ; ne nous intéressant qu'aux propriétés intrinsèques de l'axiome de monotonicité, nous définissons les notions d'informations absorbée, transmise et apportée par une séquence d'éléments comparables du treillis. Des notions introduites résulte naturellement la définition d'un écart sur T dont nous déduisons une métrique lorsque J est strictement monotone.

Cette démarche nous conduit à définir une application J_R de $T \times T$ dans $\mathbb{R}$ possédant les propriétés essentielles de l'information de SHANNON définie sur le treillis des partitions d'un ensemble Ω .

L'introduction d'hypothèses particulières sur J nous permet d'obtenir des formules explicitant la métrique et l'application J_R.

Enfin, l'interprétation des quantités ainsi définies nous permet de mesurer le degré de ressemblance entre deux éléments quelconques d'un treillis.

Ce chapitre résulte d'un travail effectué en collaboration avec Jean-Claude VAN DORPE.

I.1.- RAPPELS

Nous rappelons, dans ce paragraphe, les définitions d'informations généralisées introduites par J. LOSFELD [11].

I.1.1.- INFORMATION SUR UN TREILLIS.

Soit T un treillis dont la relation d'ordre est notée $\leqslant$; une *mesure d'information généralisée* définie sur T est :

AXIOME 1.- une application J de T dans $\bar{\mathbb{R}}^+$

AXIOME 2.- monotone (nous la supposons décroissante)

$$\forall\ (x,y) \in T \times T \quad , \quad x \leqslant y \Longrightarrow J(x) \geq J(y)$$

I.1.2.- INFORMATIONS CONDITIONNELLE ET MUTUELLE.

. L'information *conditionnelle*, fournie par l'élément $y \in T$, connaissant l'information fournie par $x \in T$, est définie par :

$$J(x/y) = J(x \wedge y) - J(y) \quad \text{si} \quad J(y) < +\infty$$

$$J(x/y) = J(x) \qquad \text{si} \quad J(y) = +\infty$$

Notons que, pour tout $y \in T$ fixé, $J(./y)$ est une information généralisée, ce qui justifie la terminologie employée. Dans tout ce qui suit, cette définition servira uniquement de notation, commode pour les valeurs infinies.

. L'information *mutuelle* entre deux éléments x et y de T est définie par :

$$J(x;y) = J(x) + J(y) - J(x \wedge y) \quad \text{si} \quad J(x) < +\infty \quad \text{et} \quad J(y) < +\infty$$

$$J(x;y) = 0 \quad \text{sinon}$$

Notons qu'en général, l'information mutuelle peut prendre des valeurs négatives et qu'elle n'a aucune propriété de monotonicité. La terminologie employée, consacrée par l'usage, peut donc prêter à confusion.

I.2.- PERTE ET GAIN D'INFORMATION DANS UN TREILLIS.

Soit $(T,\leqslant,J)$ un *treillis* T muni d'une *information généralisée* J . Sur l'exemple de la figure 1, les lettres représentent les éléments de T, les nombres entre parenthèses représentent les valeurs de J associées aux éléments de T .

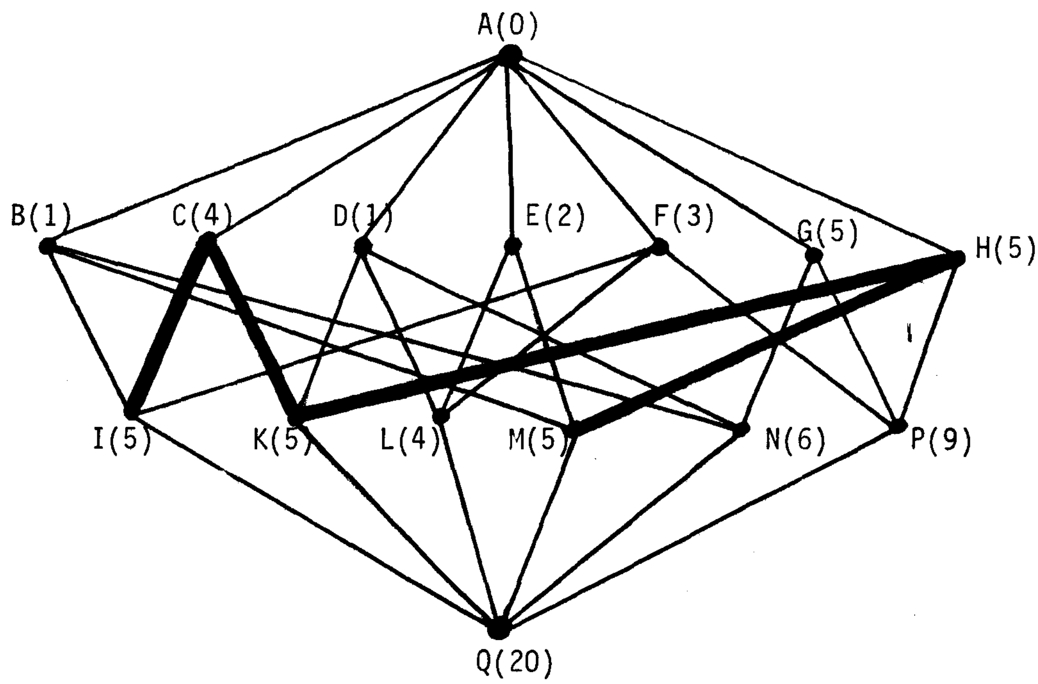


FIGURE 1

Considérons le treillis T comme représentant l'ensemble des états possibles d'un Système; l'information du système à l'état x est égale à la quantité J(x) . La suite (I, C, K, H, M) définit une séquence d'éléments comparables deux à deux de T ; cette suite formalise une évolution du système qui passe de l'état I à l'état M par une suite d'états intermédiaires.

I.2.1.- DEFINITION :

Une *séquence d'éléments comparables, de longueur* $n \in \mathbb{N}$, de T (en abrégé s.e.c.) est une séquence notée $[x_i]_0^n$ de n+1 éléments de T telle que, pour tout indice i = 0,1,...,n-1 , on ait :

$$x_i \leqslant x_{i+1} \quad \text{ou} \quad x_i \geqslant x_{i+1}$$

Si la propriété de transitivité est vérifiée pour tous les éléments d'une s.e.c., alors, conformément à la terminologie utilisée en Théorie des Treillis [3] , nous appelons *chaîne* une s.e.c. $[x_i]_0^n$ telle que :

$$\forall\ i = 0,\ldots,n-1 : x_i \leqslant x_{i+1}$$

ou

$$\forall\ i = 0,\ldots,n-1 : x_i \geqslant x_{i+1}$$

Lorsque, dans la s.e.c. $[x_i]_0^n$, nous avons $x_i \leqslant x_{i+1}$, l'information du système diminue. Nous considérons alors que le système perd, c'est-à-dire que la s.e.c. *absorbe* une certaine quantité d'information égale à $J(x_i) - J(x_{i+1}) = J(x_i/x_{i+1})$.

Inversement, si $x_i \geqslant x_{i+1}$, l'information du système augmente. Nous considérons alors que le système gagne une certaine quantité d'information égale à $J(x_{i+1}) - J(x_i) = J(x_{i+1}/x_i)$, c'est-à-dire que la s.e.c. *apporte* de l'information au système.

Le passage de l'état x_0 à l'état x_n s'accompagne d'une variation globale de l'information du système égale à $J(x_n) - J(x_0)$, variation ne dépendant que des extrêmités de la s.e.c.

Soit I l'ensemble des indices d'une s.e.c. $[x_i]_0^n$ donnée :

Notons $I^+ = \{i \in I | x_{i+1} \leqslant x_i\}$

$$I^- = \{i \in I | x_i \leqslant x_{i+1}\} .$$

Nous associons à une s.e.c. $[x_i]_0^n$ les quantités suivantes :

I.2.2.- DEFINITIONS.

. L'information *gagnée* par le système, c'est-à-dire *apportée* par la s.e.c. :

$$\Delta^+([x_i]_0^n) = \sum_{i \in I^+} J(x_{i+1}/x_i).$$

. L'information *perdue* par le système, c'est-à-dire *absorbée* par la s.e.c. :

$$\Delta^-([x_i]_0^n) = \sum_{i\in I^-} J(x_i/x_{i+1}) \ .$$

. La quantité

$$\Delta([x_i]_0^n) = \sum_{i=0}^{n-1} |J(x_{i+1}) - J(x_i)| = \Delta^+([x_i]_0^n) + \Delta^-([x_i]_0^n)$$

Δ^+ , Δ^- et Δ sont des quantités finies ou non, positives ou nulles. Si Δ^+ et Δ^- ne sont pas simultanément des valeurs infinies, nous définissons :

. La *variation d'information* du système, c'est-à-dire l'information *transmise* par la s.e.c., par :

$$V([x_i]_0^n) = J(x_n) - J(x_0) = \Delta^+([x_i]_0^n) - \Delta^-([x_i]_0^n) \ .$$

Cette quantité, finie ou non, peut être positive, négative ou nulle.

EXEMPLE : Les quantités d'information associées à la s.e.c. $C = (I, C, K, H, M)$ de la figure 1 , sont égales à :

$$V(C) = 0 \quad , \quad \Delta^+(C) = 1 \quad , \quad \Delta^-(C) = 1 \quad , \quad \Delta(C) = 2 \ .$$

Comme nous le montrons, au chapitre II, la terminologie ainsi introduite (informations absorbée, apportée, transmise), généralise celle utilisée en Théorie des Questionnaires [13] .

I.2.3.- ETUDE DES PROPRIETES DE Δ^+ , Δ^- ET Δ.

Pour tout couple fixé (x,y) de $T \times T$, notons $C(x,y)$ l'ensemble de toutes les s.e.c. reliant x à y .
Etant donnés deux éléments x et y de T , il existe une bijection entre $C(x,y)$ et $C(y,x)$:
A toute s.e.c. $C = (x_0 = x \ , x_1,\ldots,x_n = y)$ correspond une s.e.c. $\bar{C}$ appelée s.e.c. *inverse* de C reliant y à x :

$$\bar{C} = (y_0 = y \ , \ y_1 = x_{n-1},\ldots,y_n = x_0 = x),$$

il vient $\Delta^+(C) = \Delta^-(\bar{C})$, $\Delta^-(C) = \Delta^+(\bar{C})$, $\Delta(C) = \Delta(\bar{C})$

I.2.3.1.- PROPOSITION

Etant donnés deux éléments x *et* y *de* T *tels que* $J(x) \,.\, J(y) < +\infty$, C_1 *et* C_2 *deux s.e.c. reliant* x *à* y , *nous avons les équivalences :*

$$\Delta(C_1) \leq \Delta(C_2) \Longleftrightarrow \Delta^+(C_1) \leq \Delta^+(C_2) \Longleftrightarrow \Delta^-(C_1) \leq \Delta^-(C_2)$$

DEMONSTRATION.- Par hypothèse, $J(x)$ et $J(y)$ prennent des valeurs finies, Δ^+ , Δ^- et Δ prennent alors simultanément des valeurs finies ou infinies. Le cas infini étant trivial, considérons que Δ^+ , Δ^- et Δ prennent des valeurs finies :

Supposons que $\qquad \Delta(C_1) \leq \Delta(C_2)$

par définition $\qquad \Delta(C_1) = \Delta^+(C_1) + \Delta^-(C_1)$

$$\Delta(C_2) = \Delta^+(C_2) + \Delta^-(C_2)$$

On en déduit :

$$V(C_1) = \Delta^+(C_1) - \Delta^-(C_1) = J(y) - J(x) = \Delta^+(C_2) - \Delta^-(C_2) = V(C_2)$$

$$\left.\begin{array}{l} \Delta(C_1) \leq \Delta(C_2) \\ V(C_1) = V(C_2) \end{array}\right\} \Longrightarrow \begin{array}{l} \Delta^+(C_1) + \Delta^-(C_1) + \Delta^+(C_1) - \Delta^-(C_1) \leq \ldots \\ \ldots \leq \Delta^+(C_2) + \Delta^-(C_2) + \Delta^+(C_2) - \Delta^-(C_2) \end{array}$$

soit $\qquad \Delta^+(C_1) \leq \Delta^+(C_2)$.

On montre, de manière analogue :

$$\Delta^+(C_1) \leq \Delta^+(C_2) \Longrightarrow \Delta^-(C_1) \leq \Delta^-(C_2) \Longrightarrow \Delta(C_1) \leq \Delta(C_2)$$

I.2.3.2.- PROPOSITION.

Soient x *et* y *deux éléments comparables de* T, C *une chaîne quelconque reliant* x *à* y , *alors :*

(i) Si $x \lesssim y$: $\Delta^+(C) = 0$, $\Delta^-(C) = \Delta(C) = J(x/y)$

(ii) Si $x \gtrsim y$: $\Delta^-(C) = 0$, $\Delta^+(C) = \Delta(C) = J(y/x)$

La démonstration de cette proposition est immédiate.

Nous définissons :

. $\delta^+(x,y) = \inf\limits_{C \in \mathcal{C}(x,y)} \Delta^+(C)$

. $\delta^-(x,y) = \inf\limits_{C \in \mathcal{C}(x,y)} \Delta^-(C)$

. $\delta(x,y) = \inf\limits_{C \in \mathcal{C}(x,y)} \Delta(C)$

δ^+ , δ^- et δ sont des quantités finies ou non, positives ou nulles.

$\delta^+(x,y)$ représente la quantité minimale d'information que doit gagner le système pour passer de l'état x à l'état y .

$\delta^-(x,y)$ représente la quantité minimale d'information que doit perdre le système pour passer de l'état x à l'état y .

Des relations de définition de Δ , Δ^+ et Δ^- on tire :

$$\Delta([x_i]_0^n) = \Delta^+([x_i]_0^n) + \Delta^-([x_i]_0^n)$$

$$J(x_n) - J(x_0) = \Delta^+([x_i]_0^n) - \Delta^-([x_i]_0^n) \ ;$$

et de la proposition I.2.3.1., il vient immédiatement :

I.2.3.3.- THEOREME.

(i) $\delta(x,y) = \delta^+(x,y) + \delta^-(x,y)$

(ii) $J(y) - J(x) = \delta^+(x,y) - \delta^-(x,y)$.

La relation (ii) de ce théorème nous autorise à définir des applications J_R et J_T de $T \times T$ dans $\bar{\mathbb{R}}^+$ par :

. $J_T(x,y) = J(x) + \delta^+(x,y) = J(y) + \delta^-(x,y)$ pour tout couple (x,y) de $T \times T$.

. $J_R(x,y) = J(x) - \delta^-(x,y) = J(y) - \delta^+(x,y)$ si $J(x) \cdot J(y) < +\infty$

nous précisons, au paragraphe I-4, la définition pour les cas limites (valeur infinie pour $J(x)$ ou $J(y)$).

Nous donnons par la suite (paragraphe I-6), une interprétation de $J_T(x,y)$ et $J_R(x,y)$ ainsi que des quantités $\delta^+(x,y)$ et $\delta^-(x,y)$ introduites précédemment.

I.2.4.- REMARQUES.

Nous précisons, dans ce paragraphe, quelques résultats élémentaires concernant l'existence de s.e.c. particulières; enfin, nous étudions les cas limites liés aux valeurs infinies par les définitions des quantités de type δ et Δ .

. Etant donnée une s.e.c. $[x_i]_0^n \in C(x,y)$, il est toujours possible de construire une s.e.c. $[y_i]_0^m \in C(x,y)$ à partir de $[x_i]_0^n$ telle que :

- la s.e.c. $[y_i]_0^m$ soit *sans cycle,* c'est-à-dire :

$$\forall\ (i,j) \in \{0,1,\ldots,m\} \times \{0,1,\ldots,m\} \quad \text{l'on ait} \quad y_i \neq y_j$$

- la s.e.c. $[y_i]_0^m$ soit *alternée,* c'est-à-dire :

$$\forall\ i = 0,\ldots,m-2 : \quad (y_i \lneq y_{i+1} \Longrightarrow y_{i+1} \ngeq y_{i+2}) \quad \text{ou}$$

$$(y_i \gneq y_{i+1} \Longrightarrow y_{i+1} \nleq y_{i+2}) \ .$$

La construction de la s.e.c. alternée s'effectue par utilisation de la transitivité de la relation d'ordre.

- alors, nous obtenons :

$$\Delta([y_i]_0^m) \leq \Delta([x_i]_0^n) \ , \ \Delta^+([y_i]_0^m) \leq \Delta^+([x_i]_0^n) \ ,$$

$$\Delta^-([y_i]_0^m) \leq \Delta^-([x_i]_0^n) \ .$$

. Si le treillis T est *fini,* il existe donc au moins une s.e.c. C_0 de $C(x,y)$ telle que :

$$\delta(x,y) = \operatorname*{Inf}_{C \in C(x,y)} \Delta(C) = \Delta(C_0)$$

. Si $J(x) \, . \, J(y) < +\infty$ alors :

- Δ^+ , Δ^- et Δ prennent simultanément des valeurs finies ou infinies.
- δ^+ , δ^- et δ ont des valeurs finies.

. Si $J(x) = +\infty$ et $J(y) < +\infty$ alors :

- $\Delta^- = \Delta = +\infty$, $\Delta^+ \leq +\infty$
- $\delta^- = \delta = +\infty$, $\delta^+ < +\infty$

. Si $J(x) < +\infty$ et $J(y) < +\infty$ alors :

- $\Delta^+ = \Delta = +\infty$, $\Delta^- \leq +\infty$
- $\delta^+ = \delta = +\infty$, $\delta^- < +\infty$

. Si $J(x) = J(y) = +\infty$ alors :

- $\Delta^+ = \Delta^- = \Delta + \infty$
- $\delta^+ = \delta^- = \delta = +\infty$

I.3.- ÉTUDE DES PROPRIÉTÉS DE δ^+ , δ^- ET δ .

Les propriétés de δ étudiées dans ce paragraphe permettent d'étendre la notion de distance à un treillis quelconque.

I.3.1.- PROPOSITION.

L'application δ *de* $T \times T$ *dans* $\bar{\mathbb{R}}^+$ *définie par*
$\delta(x,y) = \operatorname*{Inf}_{C \in C(x,y)} \Delta(C)$ *est un écart sur* T , *c'est-à-dire une application de* $T \times T$ *dans* $\bar{\mathbb{R}}^+$ *vérifiant :*

(i) $\delta(x,x) = 0$ *pour tout* $(x,x) \in T \times T$

(ii) $\delta(x,y) = \delta(y,x)$ *pour tout* $(x,y) \in T \times T$

(iii) $\delta(x,y) \leq \delta(x,z) + \delta(z,y)$ *pour tout* $(x,y,z) \in T \times T \times T$
(inégalité triangulaire)

de plus

(iv) : s'il existe un minorant universel m *et un majorant*

universel M *dans* T , *alors :*

$$\forall\ (x,y) \in T \times T \ :\ \delta(x,y) \leq \delta(m,M)$$

DEMONSTRATION.

(i) $\delta(x,x) = 0$. Il suffit de considérer la s.e.c. de longueur 0 formée de l'élément x .

(ii) $\delta(x,y) = \delta(y,x)$. Cette égalité résulte immédiatement de la remarque suivant la définition de la s.e.c. inverse (paragraphe I-2-3)

(iii) $\delta(x,y) \leq \delta(x,z) + \delta(z,y)$.

Soient $[x_i]_0^n$ une s.e.c. quelconque reliant x à z ,
$[y_i]_0^m$ une s.e.c. quelconque reliant z à y .

La s.e.c. $C = (x_0,\dots,x_n = y_0,\dots,y_m)$ relie x à y et l'on a :

$$\delta(x,y) \leq \Delta(C)$$

Il est clair que $\Delta(C) = \Delta([x_i]_0^n) + \Delta([y_i]_0^m)$

Nous avons donc $\delta(x,y) \leq \Delta([x_i]_0^n) + \Delta([y_i]_0^m)$

Cette inégalité étant vérifiée par toute s.e.c. $[x_i]_0^n$ reliant x à z et toute s.e.c. $[y_i]_0^m$ reliant z à y , nous en déduisons :

$$\delta(x,y) \leq \delta(x,z) + \delta(z,y)$$

(iv) . Si $J(m) = +\infty$, alors $\delta(m,M) = J(m) - J(M) = +\infty$.

. Si $J(m) < +\infty$,

$$\delta(x,y) \leq \delta(x,m) + \delta(m,y) = 2\,J(m) - J(x) - J(y)$$

$$\delta(x,y) \leq \delta(x,M) + \delta(M,y) = J(x) + J(y) - 2\,J(M)$$

ce qui donne

$$2\,\delta(x,y) \leq 2\,J(m) - 2\,J(M) = 2\,\delta(m,M).$$

I.3.2.- PROPOSITION.

Les applications δ^+ *et* δ^- *de* $T \times T$ *dans* $\mathbb{R}^+$ *vérifient :*

(i) $\delta^+(x,x) = \delta^-(x,x) = 0$ *pour tout* $x \in T$

(ii) $\delta^+(x,y) = \delta^-(y,x)$ *pour tout* $(x,y) \in T \times T$

(iii) δ^+ *et* δ^- *vérifient l'inégalité triangulaire.*

La démonstration de cette proposition est analogue à celle de la proposition précédente.

I.3.3.- PROPOSITION.

Etant donné un treillis T *fini*, $T_0 = \{x \in T | J(x) < +\infty\}$, *une condition nécessaire et suffisante pour que la restriction de l'application* δ *à l'ensemble des couples* (x,y) *de* $T_0 \times T_0$ *définisse une distance est que l'application* J *soit strictement monotone.*

DEMONSTRATION.- En raison de la proposition I.3.1., il suffit de montrer :

$$J \text{ strictement monotone} \iff (\delta(x,y) = 0 \Longrightarrow x = y)$$

Démontrons la négation de cette équivalence :

. Hypothèse : $\delta(x,y) = 0 \Longrightarrow x = y$

Supposons J non strictement monotone :

$$\exists\ (x,y) \in T \times T \quad , \quad x \prec y \Longrightarrow J(x) = J(y)$$

soit C la s.e.c. de longueur 1 définie par (x,y) :

$$\Delta(C) = J(x/y) = J(x) - J(y) = 0$$

$$\Longrightarrow \delta(x,y) = 0 \Longrightarrow x = y$$

. Hypothèse : J strictement monotone.

Supposons qu'il existe un couple $(x,y) \in T \times T$ tel que :

$$\delta(x,y) = 0 \quad \text{et} \quad x \neq y$$

$$\left.\begin{array}{l}\delta(x,y) = 0 \\ \\ T \text{ fini (rem. I.2.4.)}\end{array}\right\} \Longrightarrow \exists\ [x_i]_0^n \in C(x,y) \text{ telle que}$$

$$\Delta([x_i]_0^n) = 0$$

d'oũ $J(x_0) = J(x_n) = J(x_i) \quad \forall\ i = 1,\ldots,n-1$

ce qui est impossible car J est strictement monotone et $x \neq y$ ·
On en déduit :

$$x \neq y : \quad \forall\ C \in C(x,y)\ , \quad \Delta(C) > 0$$

T étant supposé fini, $\delta(x,y)$ est strictement positif, c.q.f.d.

I.4.- PROPRIÉTÉS DES APPLICATIONS J_R ET J_T .

Rappelons les définitions des applications J_R et J_T de $T \times T$ dans $\bar{\mathbb{R}}^+$:

. $J_T(x,y) = J(x) + \delta^+(x,y) = J(y) + \delta^-(x,y)$ pour tout couple (x,y) de $T \times T$.

. $J_R(x,y) = J(x) - \delta^-(x,y) = J(y) - \delta^+(x,y)$ si $J(x) . J(y) < +\infty$

$= J(x) - \delta^-(x,y)$ si $J(y) = +\infty$ et $J(x) < +\infty$

$= J(y) - \delta^+(x,y)$ si $J(x) = +\infty$ et $J(y) < +\infty$

$J_R(x,y)$ n'est pas définie si $J(x) = +\infty$ et $J(y) = +\infty$

Remarquons que $J_R(x,y)$ prend toujours une valeur finie, alors que $J_T(x,y) = +\infty$ si et seulement si $J(x) . J(y) = +\infty$.

I.4.1.- PROPOSITION.

Les applications J_R *et* J_T *sont symétriques en* x *et* y *:*

$$J_R(x,y) = J_R(y,x) \quad , \quad J_T(x,y) = J_T(y,x)$$

et vérifient l'égalité :

$$J(x) + J(y) = J_R(x,y) + J_T(x,y) \; .$$

La symétrie résulte de l'égalité $\delta^+(x,y) = \delta^-(y,x)$ (prop. I.3.2, (ii)). L'égalité se déduit immédiatement du théorème I.2.3.3. (ii).

I.4.2.- THEOREME.

Les applications J_R *et* J_T *quand elles sont définies vérifient :*

(i) $$0 \leq J(x \vee y) \leq J_R(x,y) \leq \left\{ \begin{matrix} J(x) \\ \\ J(y) \end{matrix} \right\} \leq J_T(x,y) \leq J(x \wedge y)$$

(ii) pour tout y *fixé de* T *tel que* $J(y) < +\infty$, *les applications* $J_R(.,y)$ *et* $J_T(.,y)$ *sont décroissantes.*

(iii) Si l'application J *est strictement décroissante, alors*

- $J_R(x,y) = J(y) \iff x \preccurlyeq y$
- $J_T(x,y) = J(y) \iff x \succcurlyeq y$

DEMONSTRATION.-

(i) a) $J_T(x,y) \leq J(x \wedge y)$

Si $J(x) = +\infty$ ou $J(y) = +\infty$ alors $J(x \wedge y) = +\infty$

$J_T(x,y) = +\infty$, donc $J_T(x,y) = J(x \wedge y)$

Si $J(x) \,.\, J(y) < +\infty$, soit la s.e.c. $C = (x, x \wedge y, y)$

$\delta^+(x,y) \leq \Delta^+(C) = J(x \wedge y) - J(x)$

$\Longrightarrow \; J_T(x,y) = J(x) + \delta^+(x,y) \leq J(x \wedge y)$

b) $\left.\begin{matrix} J(x) \\ J(y) \end{matrix}\right\} \leq J_T(x,y)$

Si $J(x) = +\infty$ ou $J(y) = +\infty$ alors $J_T(x,y) = +\infty$
d'où l'inégalité (b) .

Si $J(x) \,.\, J(y) < +\infty$, l'inégalité (b) résulte du fait que $\delta^+(x,y)$ et $\delta^-(x,y)$ sont des quantités toujours ≥ 0 .

c) $J_R(x,y) \leq \left\{\begin{matrix} J(x) \\ J(y) \end{matrix}\right.$

Cette inégalité se déduit immédiatement de la définition de J_R

d) $J(x \vee y) \leq J_R(x,y)$

Si $J(x) < +\infty$ et $J(y) = +\infty$ alors $J(x \vee y) < +\infty$

$J_R(x,y) = J(x) - \delta^-(x,y)$.

Soit la s.e.c. $C = (x, x \vee y, y)$

$\Delta^-(C) = J(x) - J(x \vee y)$

$\delta^-(x,y) \leq \Delta^-(C) \Longrightarrow J_R(x,y) \geq J(x) - \Delta^-(C) = J(x \vee y)$.

La démonstration est analogue lorsque $J(x) = +\infty$ et $J(y) < +\infty$ et lorsque $J(x)\, J(y) < +\infty$.

(ii) Soient deux éléments x et x' de T tels que $x \lesssim x'$

Supposons que $J(x) = J(x') = +\infty$

alors $J_R(x,y) = J(y) - \delta^+(x,y)$

$J_R(x',y) = J(y) - \delta^+(x',y)$

$\delta^+(x,y) \leq \delta^+(x,x') + \delta^+(x',y)$ (proposition I.3.2. (iii))

$\delta^+(x,x) = 0$ (conséquence immédiate de la proposition I.2.3.2)

d'où

$\delta^+(x,y) \leq \delta^+(x',y) \Longrightarrow J_R(x,y) \geq J_R(x',y)$

On montre de manière identique que J_R est décroissante lorsque

$J(x) = +\infty$ et $J(x') < +\infty$, $J(x) < +\infty$ et $J(x') = +\infty$

et enfin lorsque $J(x)\,J(x') < +\infty$

La démonstration pour l'application J_T est analogue à celle de J_R .

(iii) Supposons $x \lessgtr y$.

. Si $J(x)\,J(y) < +\infty$ alors $\delta^+(x,y) = 0 \Longrightarrow J_R(x,y) = J(y)$

. Si $J(x) = +\infty$ et $J(y) < +\infty$ alors

$$J_R(x,y) = J(y) - \delta^+(x,y) = J(y)$$

. Si $J(x) < +\infty$ et $J(y) = +\infty$ alors

$$J_R(x,y) = J(x) - \delta^-(x,y)$$

$$x \lessgtr y \Longrightarrow \delta^-(x,y) = J(x) - J(y) \Longrightarrow J_R(x,y) = J(y)$$

Supposons maintenant que x et y soient tels que $J_R(x,y) = J(y)$.
$J_R(x,y)$ étant toujours fini, nous avons $J(y) < +\infty$. On en déduit :

$$J_R(x,y) = J(y) = J(x) - \delta^+(x,y) \Longrightarrow \delta^-(x,y) = 0$$

J étant strictement décroissante, on obtient $x \lessgtr y$.

La démonstration est analogue pour J_T .

<u>I.4.3.- PROPOSITION.</u>

Les applications J_R *et* J_T *vérifient :*

$J_R(x,y) \geq J(x) + J(y) - J(x \wedge y)$ *si* $J(x)\,J(y) < +\infty$

$J_T(x,y) \leq J(x) + J(y) - J(x \vee y)$ *pour tout couple* (x,y) *de* $T \times T$.

<u>DEMONSTRATION.</u>

. $J_R(x,y) = J(x) - \delta^-(x,y)$

soit la s.e.c. $C = (x, x \wedge y, y)$: $\Delta^-(C) = J(x \wedge y) - J(y)$

$$\delta^-(x,y) \leq \Delta^-(C) \Longrightarrow J_R(x,y) \geq J(x) - \Delta^-(C) \ , \quad \text{soit :}$$

$$J_R(x,y) \geq J(x) + J(y) - J(x \wedge y) \ .$$

. La démonstration est analogue pour J_T .

I.5.- EXPRESSION DES QUANTITÉS δ , J_R ET J_T .

I.5.1.- INTRODUCTION.

Nous proposons, dans ce paragraphe, des résultats permettant de fournir des formulations aisément calculables des quantités δ , J_R et J_T ...

Il résulte de la proposition I.2.3.1 qu'étudier les s.e.c. de $\mathcal{C}(x,y)$ minimisant les quantités Δ^+ , Δ^- ou Δ , revient à faire cette étude pour l'une d'entre elles, par exemple Δ^- . Nous disons alors qu'une *s.e.c. est J-minimale si et seulement si elle minimise l'une quelconque des quantités* Δ^+ , Δ^- *ou* Δ .

La première remarque du paragraphe I.2.4 nous permet, pour la suite, de ne plus considérer que des s.e.c. sans cycle et alternées.

Lorsque le treillis T *est fini,* il existe au moins une s.e.c. J-minimale entre deux éléments x et y donnés de T , ce qui n'est pas le cas si T est quelconque. D'autre part, l'ensemble des s.e.c. sans cycle et alternées reliant x et y est fini. Il est alors possible de mettre en oeuvre un algorithme du type Branch et Bound [5] permettant d'obtenir toutes les s.e.c. J-minimales reliant x à y , ce qui donne la possibilité de calculer les quantités $\delta(x,y)$, $J_R(x,y)$ et $J_T(x,y)$.

Cet algorithme n'offre cependant qu'un intérêt relatif, étant donné qu'il nécessite des temps d'exécution importants, aussi avons-nous été amenés [6] , *lorsque le treillis* T *est quelconque (fini ou infini),* à introduire des hypothèses particulières sur J permettant d'obtenir des expressions explicites des quantités δ , J_R et J_T .

I.5.2.- HYPOTHESES H

Il est clair que la valeur minimale de la quantité $\Delta^{-}(C)$, pour toutes les s.e.c. de longueur 2 reliant x à y , deux éléments donnés de T , est égale à :

(1) $\mathrm{Inf}(J(x/y), J(x/x \vee y))$, ce qui s'écrit, dans le cas fini :

(2) $\mathrm{Inf}(J(x \wedge y) - J(y) \; , \; J(x) - J(x \vee y))$

c'est-à-dire qu'il existe une s.e.c. J-minimale de longueur 2 de la forme $(x, x \vee y, y)$ notée $C_{\vee}(x,y)$, ou de la forme $(x, x \wedge y, y)$ notée $C_{\wedge}(x,y)$, ce qui nous amène à énoncer le théorème suivant :

I.5.2.1.- THEOREME.

Une condition nécessaire et suffisante pour que, pour tout couple (x,y) *de* $T \times T$ *, la s.e.c.* $C_{\wedge}(x,y) = (x, x \wedge y, y)$ *soit* J-*minimale est que* J *vérifie l'hypothèse* H1 *suivante :*

$$(H1) \quad \forall (x,y) \in T \times T \; , \; J(x) + J(y) \geq J(x \wedge y) + J(x \vee y)$$

DEMONSTRATION.

. Condition nécessaire

$\forall \, (x,y) \in T \times T \; , \; \forall \, C \in \mathcal{C}(x,y) : \Delta^{-}(C_{\wedge}(x,y)) \leq \Delta^{-}(C)$

soit $C = (x, x \vee y, y) : \Delta^{-}(C) = J(x) - J(x \vee y)$

$\Delta^{-}(C_{\wedge}(x,y)) = J(x \wedge y) - J(y)$

d'où $J(x \wedge y) - J(y) \leq J(x) - J(x \vee y)$ c.q.f.d.

. Condition suffisante

Si x est égal à y le problème est résolu.
Supposons x différent de y :
La démonstration s'effectue par récurrence sur la longueur de la s.e.c. reliant x à y .
Il est facile de montrer que $C_{\wedge}(x,y)$ est J-minimale pour l'ensemble des s.e.c. de longueur 1 et 2 reliant x à y .

Soit C une s.e.c. de longueur 3 reliant x à y :

$$C = (x,z,t,y) \quad \text{avec} \quad (x \lessgtr z \ , \ z \gtrless t \ , \ t \lessgtr y)$$

$$\text{ou} \quad (x \gtrless z \ , \ z \lessgtr t \ , \ t \gtrless y)$$

Supposons que $(x \lessgtr z \ , \ z \gtrless t \ , \ t \lessgtr y)$

$$\left.\begin{array}{l} t \lessgtr y \Longrightarrow x \wedge t \lessgtr x \wedge y \\ J \text{ décroissante} \end{array}\right\} \Longrightarrow J(x \wedge t) \geq J(x \wedge y)$$

$$\Delta^-(C_\wedge(x,y)) = J(x \wedge y) - J(y) \leq J(x \wedge t) - J(y)$$

$$\Delta^-(C_\wedge(x,y)) \leq J(x \wedge t) - J(t) + J(t) - J(y)$$

HYPOTHESE H1 $\Longrightarrow J(x \wedge t) - J(t) \leq J(x) - J(x \vee t)$

$$\Delta^-(C_\wedge(x,y)) \leq J(x) - J(x \vee t) + J(t) - J(y)$$

$$z \gtrless x \vee t \Longrightarrow J(z) \leq J(x \vee t)$$

ce qui donne :

$$\Delta^-(C_\wedge(x,y)) \leq J(x) - J(z) + J(t) - J(y) = \Delta^-(C)$$

On montre, de la même manière, que si C est de la forme :

$$(x \gtrless z \ , \ z \lessgtr t \ , \ t \gtrless y) \quad \text{alors} \quad \Delta^-(C_\wedge(x,y) \leq \Delta^-(C)$$

Ainsi, la s.e.c. $C_\wedge(x,y)$ est J-minimale pour toute s.e.c. de longueur 3. Supposons, maintenant, que, pour tout couple (x,y) de $T \times T$, la s.e.c. $C_\wedge(x,y)$ soit J-minimale pour toute s.e.c. de longueur n-1 reliant x à y. Soit $[x_i]_0^n$ une s.e.c. de longueur n reliant x à y.

L'hypothèse de récurrence entraîne :

$$\Delta^-(C_\wedge(x,x_{n-1})) \leq \Delta^-([x_i]_0^{n-1})$$

$$\Delta^-([x_i]_0^n) = \Delta^-([x_i]_0^{n-1}) + \Delta^-(C_1)$$

où C_1 est définie par la séquence (x_{n-1},x_n) , d'où :

$$\Delta^-([x_i]_0^n) \geq \Delta^-(C_\wedge(x,x_{n-1})) + \Delta^-(C_1) = \Delta^-(C_2)$$

où C_2 est définie par la séquence $(x,x \wedge x_{n-1},x_{n-1},y)$.

Le résultat sur les s.e.c. de longueur 3 entraîne :

$$\Delta^-(C_2) \geq \Delta^-(C_\wedge(x,y))$$

ce qui donne :

$$\Delta^-(C_\wedge(x,y)) \leq \Delta^-([x_i]_0^n) \qquad \text{c.q.f.d.}$$

On démontre de la même manière les résultats suivants :

I.5.2.2.- THEOREME.

Une condition nécessaire et suffisante pour que, pour tout couple, (x,y) *de* $T \times T$, *la s.e.c.* $C_\vee(x,y) = (x,x \vee y,y)$ *soit* J-*minimale est que* J *vérifie l'hypothèse H2 :*

(H2) $\quad \forall\ (x,y) \in T \times T \ , \ J(x) + J(y) \leq J(x \wedge y) + J(x \vee y)$

I.5.2.3.- COROLLAIRE

Une condition nécessaire et suffisante pour que, pour tout couple (x,y) *de* $T \times T$, *les s.e.c.* $C_\wedge(x,y)$ *et* $C_\vee(x,y)$ *soient* J-*minimales est que* J *vérifie l'hypothèse HO :*

(HO) $\quad \forall\ (x,y) \in T \times T \ , \ J(x) + J(y) = J(x \wedge y) + J(x \vee y)$.

Ces résultats permettent de généraliser immédiatement la proposition I.3.3.

I.5.2.4.- PROPOSITION.

Etant donné un treillis T *quelconque, si* J *vérifie l'une quelconque des hypothèses H0 , H1, ou H2, une condition nécessaire et suffisante pour que la restriction de* δ *à l'ensemble des couples* (x,y) *de* $T_0 \times T_0$ *définisse une distance est que l'application* J *soit strictement monotone.*

Ces résultats nous permettent également de donner une expression explicite des quantités δ , J_R et J_T . Nous donnons ces expressions dans le tableau I.5.2.5. Le tableau I.5.2.6. donne les formes particulières obtenues lorsque les éléments x et y de T sont tels que $J(x) < +\infty$ et $J(y) < +\infty$.

	H1	H2
$\delta^+(x,y)$	$J(y/x)$	$J(y/x \vee y)$
$\delta^-(x,y)$	$J(x/y)$	$J(x/x \vee y)$
$\delta(x,y)$	$J(y/x) + J(x/y)$	$J(y/x \vee y) + J(x/x \vee y)$
$J_R(x,y)$	$J(x) + J(y) - J(x \wedge y)$	$J(x \vee y)$
$J_T(x,y)$	$J(y) + J(y/x)$ $=$ $J(x) + J(x/y)$	$J(x) + J(y/x \vee y)$ $=$ $J(y) + J(x/x \vee y)$

TABLEAU I.5.2.5. (cas général)

.../...

TABLEAU I.5.2.6. (Lorsque $J(x)\ J(y) < +\infty$)

	H1	H2	H0
$\delta^+(x,y)$	$J(x \wedge y) - J(x)$	$J(y) - J(x \vee y)$	$J(x \wedge y) - J(x)$ = $J(y) - J(x \vee y)$
$\delta^-(x,y)$	$J(x \wedge y) - J(y)$	$J(x) - J(x \vee y)$	$J(x \wedge y) - J(y)$ = $J(x) - J(x \vee y)$
$\delta(x,y)$	$2J(x \wedge y) - J(x) - J(y)$	$J(x) + J(y) - 2J(x \vee y)$	$J(x \wedge y) - J(x \vee y)$
$J_R(x,y)$	$J(x) + J(y) - J(x \wedge y)$	$J(x \vee y)$	$J(x) + J(y) - J(x \wedge y)$ = $J(x \vee y)$
$J_T(x,y)$	$J(x \wedge y)$	$J(x) + J(y) - J(x \vee y)$	$J(x) + J(y) - J(x \vee y)$ = $J(x \wedge y)$

Pour l'hypothèse H0, nous pouvons prendre indifféremment les expressions de ces quantités relatives à la condition H1 ou à la condition H2.

REMARQUES : L'information J définie sur la figure 1 (page I-4), ne vérifie ni H1, ni H2.

En effet :

$$J(I) + J(K) < J(I \vee K) + J(I \wedge K)$$

$$J(C) + J(H) > J(C \vee H) + J(C \wedge H)$$

Dans cet exemple, la s.e.c. J-minimale reliant I à M est la s.e.c. I C K H M de longueur 4 soulignée en gras sur la figure.

I.6.- APPLICATION : RÉSUMÉS ET INFORMATIONS [8].

I.6.1.- Notons H l'information de SHANNON sur les partitions finies Π d'un espace probabilisé (Ω,S,P) définie par :

$$H(\Pi) = \sum_{A\in\Gamma} P(A) \text{ Log } 1/P(A)$$

Nous savons [12] que l'information mutuelle $H(\Pi;\Pi')$ apportée par les partitions Π et Π' est définie par :

$$H(\Pi;\Pi') = \sum_{(A,B)\in\Pi\times\Pi'} P(A\cap B) \text{ Log } \frac{P(A\cap B)}{P(A)P(B)},$$

et possède les propriétés :

$\forall\ \Pi$, Π' et Π'' partitions de Ω :

. $H(\Pi;\Pi') \geq 0$

. $\Pi' \leq \Pi'' \Longrightarrow H(\Pi;\Pi') \geq H(\Pi;\Pi'')$

. $H(\Pi;\Pi') \leq H(\Pi)$ avec égalité si et seulement si $\Pi' \leq \Pi$.

La partition Π étant supposée fixée, l'interprétation de l'application $H(.;\Pi)$ est la suivante : $H(.;\Pi)$ est une information généralisée mesurant ce que la connaissance d'une partition Π' apporte sur Π . En effet : $H(\Pi;\Pi')$ est une quantité positive, inférieure à $H(\Pi)$ mais d'autant plus grande que la partition Π' se "rapproche" de la partition Π.

On montre [7] que l'information de SHANNON sur les partitions vérifie l'hypothèse H1 et donc $H_R(.,\Pi)$ est identique à $H(.;\Pi)$ pour cet exemple.

I.6.2.- De manière générale, soit H une mesure d'information généralisée quelconque définie sur le treillis des partitions d'un ensemble fini, non vide. Le Théorème I.4.2. montre que l'application $H_R(.,\Pi)$ possède les propriétés intéressantes de l'information mutuelle. Cette application, égale à $H(.;\Pi)$ lorsque H vérifie l'hypothèse H1 se présente donc bien comme une extension de l'information mutuelle.

Considérons une partition Π comme ensemble de référence; lorsque l'on cherche à expliquer Π par Π' ; c'est-à-dire à considérer Π' comme un résumé (modèle) de Π alors :

. Nous interprétons $H_R(\Pi,\Pi')$ comme la quantité d'information en commun entre Π et Π' et $H_T(\Pi,\Pi')$ comme la quantité totale d'information apportée par Π et Π' .

. $\delta^-(\Pi,\Pi') = H(\Pi) - H_R(\Pi,\Pi') = H_T(\Pi,\Pi') - H(\Pi')$

définit la quantité d'information apportée par Π et non expliquée par Π' .

. $\delta^+(\Pi,\Pi') = H(\Pi') - H_R(\Pi,\Pi') = H_T(\Pi,\Pi') - H(\Pi)$

définit la quantité d'information "parasite" apportée par Π' et ne servant pas à la connaissance de Π.

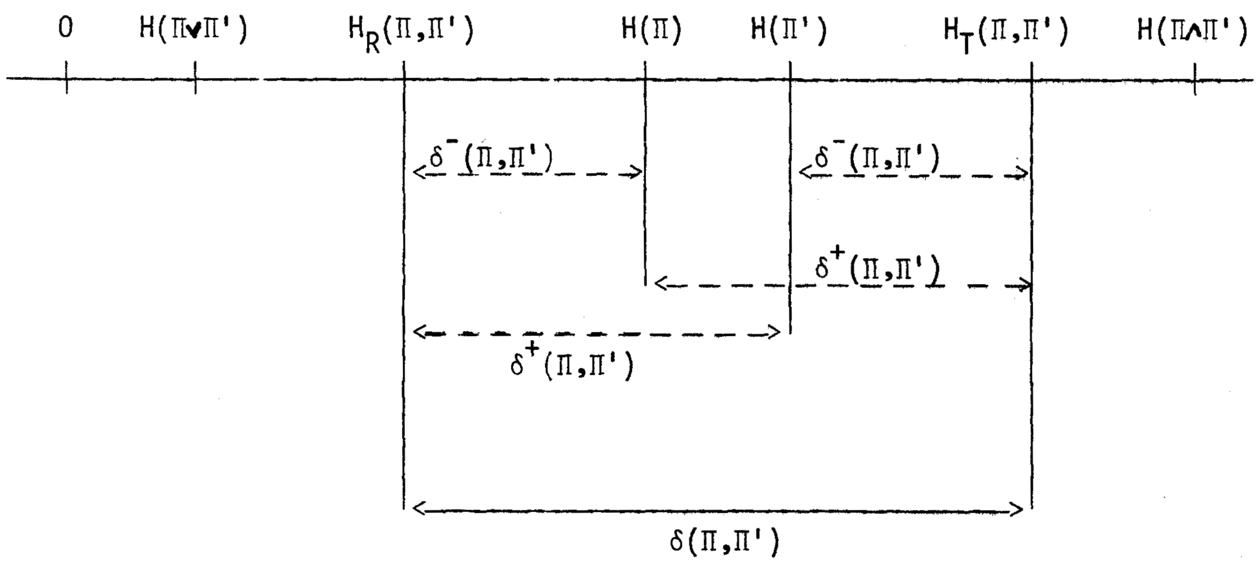


FIGURE I.6.1.

Certains algorithmes utilisés en analyse des données [2] (par exemple les méthodes de classification hiérarchique), se ramènent à la donnée d'une partition Π et à la construction d'une séquence $(\Pi_i \mid i = 1,\ldots,)$ de partitions de plus en plus fines, cherchant à approcher Π de mieux en mieux.

Les inégalités $\Pi_1 \gtrless \Pi_2 \gtrless \dots \gtrless \Pi_i \gtrless \Pi_{i+1} \gtrless \dots$ impliquent :

$$H_R(\Pi,\Pi_1) \le \dots \le H_R(\Pi,\Pi_i) \le H_R(\Pi,\Pi_{i+1}) \le \dots$$

$$H_T(\Pi,\Pi_1) \le \dots \le H_T(\Pi,\Pi_i) \le H_T(\Pi,\Pi_{i+1}) \le \dots$$

$$\delta^-(\Pi,\Pi_1) \ge \dots \ge \delta^-(\Pi,\Pi_i) \ge \delta^-(\Pi,\Pi_{i+1}) \ge \dots$$

$$\delta^+(\Pi,\Pi_2) \le \dots \le \delta^+(\Pi,\Pi_i) \le \delta^+(\Pi,\Pi_{i+1}) \le \dots$$

ce qui formalise le fait que la suite de partitions $\{\Pi_i\}$ approche de mieux en mieux l'information de Π , mais elle introduit aussi de plus en plus d'information parasite (inutile dans la connaissance de Π).

Nous remarquons que $\delta(\Pi,\Pi') = H_T(\Pi,\Pi') - H_R(\Pi,\Pi')$ *; la fonction* $\delta(.,\Pi)$ *n'est pas monotone en général. Les minima locaux de* δ *représentent alors de "bons" résumés de* Π *dans le sens où ils correspondent à des positions d'équilibre concernant les variations respectives de* H_R *et* H_T .

Les remarques et interprétations précédentes ont été présentées dans le cadre du treillis des partitions d'un ensemble Ω et sont utilisées dans le chapitre III pour les applications (algorithme ARBRE). Rappelons que les résultats et propriétés fondamentaux permettant ces interprétations ont été établis ici dans le cadre général d'un treillis T quelconque muni d'une information généralisée J :

Soient $(x,y) \in T \times T$.
$J(x)$ *(respectivement* $J(y)$ *) représente la quantité d'information apportée par la connaissance de la réalisation de l'état* x *(resp.* y*) du système :*

- $J_R(x,y)$ *représente la quantité d'information commune à* x *et* y *;*

- $\delta^-(x,y) = J(x) - J_R(x,y)$ *la quantité d'information spécifique à* x *(dans le contexte* y*)*

. $\delta^+(x,y) = J(y) - J_R(x,y)$ *la quantité d'information spécifique à* y *(dans le contexte* x*).*

Ou encore, si x *est l'état du système, si* y *est l'explication qu'on en donne (résumé ou modèle).*

. $J_T(x,y)$ *est la quantité totale d'information apportée par le résumé* y, *comprenant en particulier une certaine quantité inutile, dans la connaissance de* x .

. $J_R(x,y)$ *est la quantité d'information qu'apporte le résumé* y *sur la la connaissance de* x .

. $\delta^-(x,y)$ *la quantité d'information de* x *non "expliquée" par le résumé* y ;

. $\delta^+(x,y)$ *une quantité d'information "parasite" apportée par le résumé* y *sans rapport avec l'information de* x .

C H A P I T R E II

-=-=-=-=-=-=-=-=-=-=-=

INFORMATIONS ET QUESTIONNAIRES

-=-=-=-=-=-=-=-=-=-=-=

II.- INFORMATIONS ET QUESTIONNAIRES.

L'utilisation de la mesure d'information de SHANNON en Théorie des Questionnaires a permis d'introduire les notions d'information *apportée :* $J(Q)$, *absorbée :* $e(Q)$ et *transmise :* $I(Q)$ en obtenant le théorème fondamental [13] :

$$J(Q) = I(Q) + e(Q)$$

$$J(Q) \geq I(Q)$$

Par la suite, d'autres mesures d'information ont été utilisées telles que la mesure d'AGGARWAL-CESARI-PICARD, ou encore celle de BELIS et GUIASCU. Les définitions particulières de $J(Q)$, $I(Q)$ et $e(Q)$, adaptées à chacune des mesures ont permis de retrouver le théorème précédent. Les résultats du premier chapitre nous ont amenés à définir les notions d'information *apportée* $\Delta^+(C)$, *absorbée* $\Delta^-(C)$, et *transmise* $V(C)$ d'une s.e.c. C d'éléments d'un treillis muni d'une information généralisée.

Nous proposons une définition de la "localisation" englobant la propriété de branchement, et équivalente à l'axiome A3 de B. FORTE et N. PINTACUDA. Les résultats nous permettent, quand on les applique au treillis des partitions d'un ensemble fini non vide, muni d'une information généralisée vérifiant cette propriété de localisation d'unifier les notions d'information dans les questionnaires en retrouvant les résultats fondamentaux cités ci-dessus. Nous montrons que la terminologie utilisée pour $\Delta^+(C)$, $\Delta^-(C)$ et $V(C)$ est cohérente avec celle utilisée en Théorie des Questionnaires. Enfin, nous étudions brièvement l'information de RENYI qui ne vérifie pas la propriété de localisation. Il est néanmoins possible pour cette mesure d'information, de définir les notions d'information dans un questionnaire.

II.1.- RAPPELS.

II.1.1.- QUESTIONNAIRES [13]

Rappelons quelques définitions :

. Un graphe $G = (X,\Gamma)$ orienté est *quasi fortement connexe inférieurement* (q.f.c.i.) s'il existe un sommet ascendant de tous autres; explicitons :

$\exists\ x_0 \in X$ tel que :

$\forall\ x \in X$, $x \neq x_0$, il existe un chemin reliant x_0 à x .

. Un *latticiel* est un graphe fini q.f.c.i. et sans circuit.

. Une *valuation* d'un graphe (X,Γ) est une application de l'ensemble des arcs ou des sommets dans un ensemble Y .

. Un *graphe valué* est un triplet (X,Γ,v) où (X,Γ) est un graphe et v une valuation.

. Un *questionnaire* est un graphe valué q.f.c.i. , $Q = (X,\Gamma,\mu)$ tel que l'ensemble des sommets admet la partition $X = E \cup F$ où E est formé des sommets terminaux appelés *réponses,* F est formé des sommets non terminaux appelés *questions,* et vérifiant les axiomes :

(q_1) : (X,Γ) est un graphe fini sans circuit.

(q_2) : Aucun sommet de X n'est origine d'un seul arc.

(q_3) : Il existe une application μ de l'ensemble des arcs dans $\bar{\mathbb{R}}^+$ telle que, pour tout i appartenant à F tel que Γ_i^{-1} soit différent de l'ensemble vide, l'on ait :

$$\sum_{j\in\Gamma_i} \mu(i,j) = \sum_{h\in\Gamma_i^{-1}} \mu(h,i) = \mu(i)$$

(q_4) : Cette application μ est telle que :

$$\mu(\alpha) = \sum_{i\in E} \sum_{h\in\Gamma_i^{-1}} \mu(h,i) \quad \text{où } \alpha \text{ est la}$$

racine du questionnaire.

. Pour tout $i \in X$, notons $E(i) = \hat{\Gamma}_i \cap E$ l'ensemble des descendants de i qui sont terminaux.

. Un questionnaire *arborescent* est un questionnaire dont le support est arborescent, sinon le questionnaire est dit latticiel.

II.1.2.- INFORMATIONS AU SENS DE SHANNON DANS LES QUESTIONNAIRES.

Soit Q un questionnaire de type probabiliste [13] : la valuation μ définie sur Γ est une application notée p vérifiant :

$$\sum_{e \in E} p(e) = 1 \quad \text{où} \quad \forall\, e \in E : p(e) = \sum_{h \in \Gamma_e^{-1}} p(h,e)$$

$$\forall\, e \in E \;:\; 0 < p(e) \leq 1$$

. L'information *transmise par un questionnaire* Q est définie par :

$$I(Q) = \sum_{e \in E} p(e) \operatorname{Log} \frac{1}{p(e)}$$

. L'information *traitée par une question* i définie par :

$$J(i) = \sum_{j \in \Gamma_i} \frac{p(i,j)}{p(i)} \operatorname{Log} \frac{p(i)}{p(i,j)} \quad = \sum_{i \in \Gamma_i} p(j/i) \log \frac{1}{p(j/i)}$$

$I(i) = p(i)\, J(i)$ représente l'information *apportée par la question* i .

. L'information *traitée par le questionnaire* Q est la somme :

$$J(Q) = \sum_{i \in F} p(i)\, J(i) \, .$$

. L'information *absorbée par une question* différente de la racine α , ou une réponse, est définie par :

$$e(i) = \sum_{h \in \Gamma_i^{-1}} p(h,i) \operatorname{Log} \frac{p(i)}{p(h,i)}$$

L'information *absorbée par le questionnaire* Q est égale à :

$$e(Q) = \sum_{i \in E \cup \bar{F}} e(i) \quad \text{où} \quad \bar{F} = F - \{\alpha\}$$

II.1.2.1.- THEOREME [13] .

Soit Q *un questionnaire, avec les notations précédentes, on a :*

(i) $J(Q) \geq I(Q)$

(ii) $J(Q) = I(Q) + e(Q)$

(iii) $e(Q) = 0 \Longleftrightarrow Q$ *est arborescent.*

II.2.- INFORMATION GÉNÉRALISÉE SUR LE TREILLIS DES PARTITIONS D'UN ENSEMBLE Ω .

II.2.1.- Soient Ω un ensemble quelconque, non vide, d'événements élémentaires ω , S une algèbre de Boole de parties de Ω . Une *mesure d'information généralisée sur les événements* [9] de (Ω,S), est une fonction d'ensemble I définie sur S et satisfaisant aux axiomes :

Axiome 1 : $I : S \to \mathbb{R}^+$

Axiome 2 : I est monotone pour l'inclusion, c'est-à-dire :

$$(A,B) \in S \times S \quad \text{et} \quad A \subseteq B \Longrightarrow I(A) \geq I(B)$$

Nous prendrons comme valeurs universelles $I(\emptyset) = +\infty$ et $I(\Omega) = 0$.

II.2.2.- Soient Ω un ensemble fini non vide, notons T l'ensemble des partitions de Ω ordonné par la relation de "finesse" :

$$(\Pi_1,\Pi_2) \in T \times T \quad , \quad \Pi_1 \preccurlyeq \Pi_2 \Longleftrightarrow \forall A \in \Pi_1 \quad , \exists B \in \Pi_2$$

$$\text{tel que} \quad A \subseteq B$$

pour cette relation d'ordre, T a une structure de treillis [1] et possède un majorant universel $\Pi_M = \{\Omega\}$ et un minorant universel $\Pi_m = \{(\omega) | \omega \in \Omega\}$.

DEFINITION.- Une *information généralisée* H [12] *définie sur* T est une application vérifiant les axiomes :

A1 : $H : T \to \bar{\mathbb{R}}^+$

A2 : $\Pi_1 \preccurlyeq \Pi_2 \Longrightarrow H(\Pi_1) \geq H(\Pi_2)$

nous supposerons de plus que $H(\Pi_m) < +\infty$.

Supposons définies sur une algèbre de Boole S de parties de Ω , une mesure de P de probabilités et une information I [9].

. L'*information moyenne* d'une partition Π de Ω est définie [12] par :

$$H(\Pi) = \sum_{A \in \Pi} p(A)\ I(A)$$

Cette application H est une information généralisée au sens où nous venons de la définir.

II.2.3.- PRINCIPE DE LOCALISATION DE FORTE ET PINTACUDA [4] .

FORTE et PINTACUDA ont défini [4] l'axiome A3 de principe de localisation pour une mesure d'information apportée par une expérience. Reprenons les notations de FORTE et PINTACUDA dans le cas d' "expériences complètes", c'est-à-dire des partitions de Ω :

Soient (A,B) une partition de Ω en deux éléments,

Π_A et Π'_A deux partitions de A telles que $\Pi_A \preccurlyeq \Pi'_A$,

Π_B et Π'_B deux partitions de B telles que $\Pi_B \preccurlyeq \Pi'_B$,

adoptons les notations :

$\Pi_1 = (\Pi_A, B)$

$\Pi_2 = (A, \Pi_B)$

$\Pi'_1 = (\Pi'_A, B)$

$\Pi'_2 = (A, \Pi'_B)$.

Quand il s'agit de partitions de Ω , l'axiome A3 de FORTE ET PINTACUDA se traduit comme suit :

(A3)' Pour tout quadruplet $(\Pi_1,\Pi'_1,\Pi_2,\Pi'_2)$ défini par les notations précédentes, on a l'égalité :

$$H(\Pi_1 \wedge \Pi_2) + H(\Pi'_1 \wedge \Pi'_2) = H(\Pi'_1 \wedge \Pi_2) + H(\Pi_1 \wedge \Pi'_2)$$

où l'on a posé, pour deux partitions Π et Π' de Ω :

$$\Pi \wedge \Pi' = \{C \cap D | C \in \Pi \text{ , } D \in \Pi'\}$$

L'application (paragraphe II-5) du principe de localisation nous conduit à proposer une définition équivalente de la propriété de localisation; cette définition facilitant l'utilisation de cette propriété pour les mesures d'information utilisées dans les questionnaires.

II.2.3.1.- DEFINITION.

Soit (A,B) une partition quelconque de Ω en deux éléments;
Soient $(B_1,B_2,\ldots,B_m)$ une partition quelconque de B ,
(A_1,A_2) une partition quelconque de A en deux éléments.
Posons :

$$\Pi_1 = (A,B_1,\ldots,B_m)$$

$$\Pi_2 = (A_1,A_2,B_1,\ldots,B_m)$$

$$\Pi'_1 = (A,B)$$

$$\Pi'_2 = (A_1,A_2,B)$$

DEFINITION.- Nous dirons qu'une information généralisée H vérifie la *propriété de localisation* si et seulement si : quels que soient Π_1 , Π_2 , Π'_1 , Π'_2 définies par les notations précédentes, on a :

$$H(\Pi_2) - H(\Pi_1) = H(\Pi'_2) - H(\Pi'_1)$$

II.2.3.2.- PROPOSITION.

Toute information moyenne vérifie la propriété de localisation.

II.2.3.3.- PROPOSITION.

La propriété de localisation est équivalente à l'axiome (A3)' de B. FORTE et N. PINTACUDA.

DEMONSTRATION.-

1) (A3)' $\Longrightarrow$ P.L. (Propriété de localisation)

Soient $\Pi_A = (A_1,A_2)$

$\Pi'_A = A$

$\Pi_B = (B_1,B_2,\dots,B_n)$

$\Pi'_B = B$

Posons $\Pi_1 = (A_1,A_2,B)$, $\Pi_2 = (A,B_1,\dots,B_n)$

$\Pi'_1 = (A,B)$, $\Pi'_2 = (A,B)$

$\Longrightarrow$ $\Pi_1 \wedge \Pi_2 = A_1,A_2,B_1,\dots,B_n)$, $\Pi_1 \wedge \Pi'_2 = (A_1,A_2,B)$,

$\Pi'_1 \wedge \Pi_2 = (A,B_1,\dots,B_n)$, $\Pi'_1 \wedge \Pi'_2 = (A,B)$

d'où :

$$H(A_1,A_2,B_1,\dots,B_n) - H(A,B_1,\dots,B_n) = H(A_1,A_2,B) - H(A,B)$$

2) P.L. $\Longrightarrow$ (A3)'

Soient Π_1,Π_2,Π_3,Π_4 des partitions de Ω définies par :

.../...

$$\Pi_1 = (A_1,\dots,A_n \, , \, B_1,\dots,B_m)$$

$$\Pi_2 = (A,B_1,\dots,B_m)$$

$$\Pi_3 = (A_1,\dots,A_n,B)$$

$$\Pi_4 = (A,B)$$

avec $\quad A = \bigcup_{i=1}^{n} A_i \; , \; B = \bigcup_{j=1}^{m} B_j$

Montrons que $\;$ P.L. $\Longrightarrow H(\Pi_1) - H(\Pi_2) = H(\Pi_3) - H(\Pi_4)$

Montrons d'abord que $\forall \; k = 1,\dots,n$

$$H(A_1,A_2,\dots,A_n,B_1,\dots,B_m) - H(\bigcup_{i=1}^{k} A_i,A_{k+1},\dots,A_n,B_1,\dots,B_m) =$$

$$= H(A_1,\dots,A_n,B) - H(\bigcup_{i=1}^{k} A_i,A_{k+1},\dots,A_n,B)$$

L'égalité est vraie pour $k = 1$. Supposons-la vraie à l'ordre $k-1$

$$H(A_1,\dots,A_n,B) - H(\bigcup_{1}^{k=1} A_i,A_k,\dots,A_n,B) =$$

$$= H(A_1,\dots,A_n,B_1,\dots,B_m) - H(\bigcup_{1}^{k-1} A_i,A_k,\dots,A_n,B_1,\dots,B_m)$$

Posons $\quad \alpha = H(A_1,\dots,A_n,B_1,\dots,B_m) - H(\bigcup_{1}^{k} A_i,A_{k+1},\dots,A_n,B_1,\dots,B_m)$

$$\text{P.L.} \Longrightarrow H(\bigcup_{1}^{k-1} A_i,A_k,\dots,A_n,B_1,\dots,B_m) - H(\bigcup_{1}^{k} A_i,A_{k+1},\dots,A_n,B_1,\dots,B_m)$$

$$= H(\bigcup_{1}^{k-1} A_i,A_k,\bigcup_{k+1}^{n} A_i \cup B) - H(\bigcup_{1}^{k} A_i,\bigcup_{k+1}^{n} A_i \cup B)$$

$$\Longrightarrow \quad \alpha = H(A_1,\dots,A_n,B_1,\dots,B_m) - H(\bigcup_{1}^{k-1} A_i,A_k,\dots,A_n,B_1,\dots,B_m)$$

$$+ H(\bigcup_{1}^{k-1} A_i,A_k,\bigcup_{k+1}^{n} A_i \cup B) - H(\bigcup_{1}^{k} A_i,\bigcup_{k+1}^{n} A_i \cup B)$$

L'hypothèse de récurrence entraîne :

$$\alpha = H(A_1,\dots,A_n,B) - H(\bigcup_{1}^{k-1} A_i,A_k,\dots,A_n,B) + H(\bigcup_{1}^{k-1} A_i,A_k,\bigcup_{k+1}^{n} A_i \cup B)$$

$$- H(\bigcup_{1}^{k} A_i,\bigcup_{k+1}^{n} A_i \cup B)$$

soit :

$$H(\bigcup_{1}^{k-1} A_i, A_k, \ldots, A_n, B) - H(\bigcup_{1}^{k} A_i, A_{k+1}, \ldots, A_n, B) = \gamma$$

$$= H(\bigcup_{1}^{k-1} A_i, A_k, \bigcup_{k+1}^{n} A_i \cup B) - H(\bigcup_{1}^{k} A_i, \bigcup_{k+1}^{n} A_i \cup B)$$

$$\text{P.L.} \Longrightarrow \gamma = H(\bigcup_{1}^{k-1} A_i, A_k, \bigcup_{k+1}^{n} A_i \cup B) - H(\bigcup_{1}^{k} A_i, \bigcup_{k+1}^{n} A_i \cup B)$$

d'oū

$$\alpha = H(A_1, \ldots, A_n, B) - H(\bigcup_{1}^{k} A_i, A_{k+1}, \ldots, A_n, B) \qquad \text{c.q.f.d.}$$

On en déduit, pour $k = n$

$$H(A_1, A_2, \ldots, A_n, B_1, \ldots, B_m) - H(A, B_1, \ldots, B_m)$$

$$= H(A_1, \ldots, A_n, B) - H(A, B).$$

II.2.4.- PROPRIETES DE LOCALISATION ET DE BRANCHEMENT.

Soit Φ_N une fonction d'information définie sur des questionnaires [13] :

$$\Phi_N(Q) = \Phi_N(\mu_1, \mu_2, \ldots, \mu_N)$$

où μ_i est la mesure à la réponse e_i.

Soient Q_0 et Q_1 deux questionnaires admettant respectivement $N-\beta$ et $\beta+1$ réponses, on dit que Φ_N possède la propriété de branchement si :

$$\Phi_N(Q_0 \lozenge Q_1) = \Phi_{N-\beta}(Q_0) + \Psi_{\beta+1}(Q_1)$$

où $\Psi_{\beta+1}$ est une fonction des $\beta+1$ mesures définies aux réponses de Q_1.

Soit Π_0 la partition de Ω associée aux réponses de Q_0, $\text{card}(\Pi_0) = N-\beta$. Prolongeons Q_0 par Q_1 à partir d'une réponse de Q_0, nous obtenons les partitions suivantes :

$\Pi_0 \quad |\overset{A_1}{\rule{2cm}{0.4pt}}|\overset{A_2}{\rule{2cm}{0.4pt}}| \;\cdots\cdots\; |\overset{A_k}{\rule{4cm}{0.4pt}}| \;\cdots\; |\overset{A_{N-\beta-1}}{\rule{2cm}{0.4pt}}|\overset{A_{N-\beta}}{\rule{2cm}{0.4pt}}|$

$\Pi_1 \quad |\overset{A_1}{\rule{2cm}{0.4pt}}|\overset{A_2}{\rule{2cm}{0.4pt}}| \;\cdots\cdots\; |\overset{B_1}{\rule{0.8cm}{0.4pt}}|\cdots|\overset{B_k}{\rule{0.8cm}{0.4pt}}|\cdots|\overset{B_{\beta+1}}{\rule{0.8cm}{0.4pt}}| \;\cdots\; |\overset{A_{N-\beta-1}}{\rule{2cm}{0.4pt}}|\overset{A_{N-\beta}}{\rule{2cm}{0.4pt}}|$

$\mathrm{card}(\Pi_1) = N$

Nous avons :

$$\Phi_N(Q_0 \Diamond Q_1) = \Phi_N(\Pi_1)$$

$$\Phi_{N-\beta}(Q_0) = \Phi_{N-\beta}(\Pi_0)$$

Supposons que Φ_N vérifie la propriété de localisation, nous obtenons :

$$\Phi_N(\Pi_1) - \Phi_{N-\beta}(\Pi_0) = \Phi_{\beta+2}(B_1,\ldots,B_{\beta+1},\bar{B}) - \Phi_2(\bigcup_{1}^{B+1} \beta_i,\bar{B})$$

où $\bar{B}$ est le complémentaire relativement à Ω de la réunion des B_i , ce qui nous donne :

$$\Phi_N(\Pi_1) - \Phi_{N-\beta}(\Pi_0) = \Psi_{\beta+1}(B_1,\ldots,B_{\beta+1})$$

où $\Psi_{\beta+1}$ est une fonction des $\beta+1$ mesures définies aux réponses de Q_1 .

Ainsi, la propriété de branchement est un cas particulier de la propriété de localisation, plus précisément :

Pour une fonction d'information de la forme $\Phi_N(Q) = \Phi_N(\mu_1,\ldots,\mu_N)$, la propriété de localisation entraîne la propriété de branchement.

II.3.- ARBORESCENCE DE L'ORDRE DES QUESTIONS ASSOCIÉ À UN QUESTIONNAIRE.

Etant donné un questionnaire Q quelconque, soit Ω un ensemble sur lequel on applique le questionnaire ; nous appelons Ω : "Domaine du Questionnaire".

Le fait de poser le questionnaire sur un ensemble Ω nous apporte une certaine information sur cet ensemble; cette information est liée à la structure introduite par le questionnaire sur Ω , structure définie en fait par l'ensemble des réponses du questionnaire : *nous nous intéressons dans ce qui suit, aux questionnaires induisant une structure de partition sur* Ω .

Nous interprétons alors une question comme un opérateur qui effectue une partition d'un sous-ensemble de Ω : tout arc $(i,j) \in \Gamma$ étant associé à une issue de la question i , notons Ω_i^j le sous-ensemble de Ω correspondant à cet arc.

Nous appelons *domaine* d'une question i , noté $\mathcal{D}(i)$, le sous-ensemble de Ω sur lequel est posé la question i :

$$\mathcal{D}(i) = \bigcup_{j \in \Gamma_i} \Omega_i^j = \bigcup_{h \in \Gamma_i^{-1}} \Omega_h^i$$

Si le questionnaire Q possède n questions, le graphe G , support du questionnaire, nous donne un certain nombre de contraintes sur l'ordre dans lequel sont posées les questions.

Soit une permutation $C = (i_1,\dots,i_n)$ de questions de Q , C est dite *compatible* avec Q si et seulement si :

$$\forall\ j = 1,\dots,n : \quad (\forall\ i \in \hat{\Gamma}_{i_j}^{-1} \ , \ \exists\ i_k \in C \ , \ k \leq j \ \text{ tel que } \ i_k = i)$$

cette définition exprime le fait que, pour poser une question i_j (j fixé $\in [1,n]$) , il faut avoir posé obligatoirement toutes les questions ascendantes de i_j .

<u>EXEMPLE</u> : Soit le graphe G suivant, support d'un questionnaire Q :

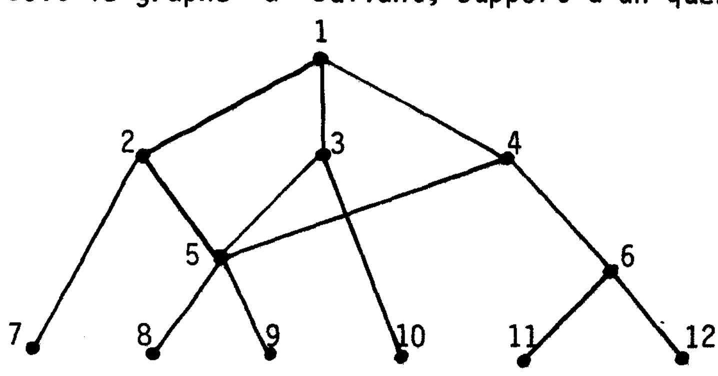


FIGURE II-1

ainsi, l'on ne peut poser, par exemple, la question 5 qu'après avoir posé les questions 1, 2, 3 et 4. La question 6 ne peut être posée qu'après les questions 1 et 4 .

Considérons alors l'ensemble de toutes les permutations de questions compatibles avec le questionnaire Q . Cet ensemble se représente en construisant ce que nous appelons *l'arborescence* A_Q *de l'ordre des questions associée à un questionnaire* Q *:* un chemin de l'arborescence joignant la racine à un sommet pendant définit une permutation de questions compatibles avec le questionnaire.

<u>EXEMPLE</u> : Arborescence de l'ordre des questions associée au questionnaire de la figure II-1.

Voir FIGURE II-2 page suivante.

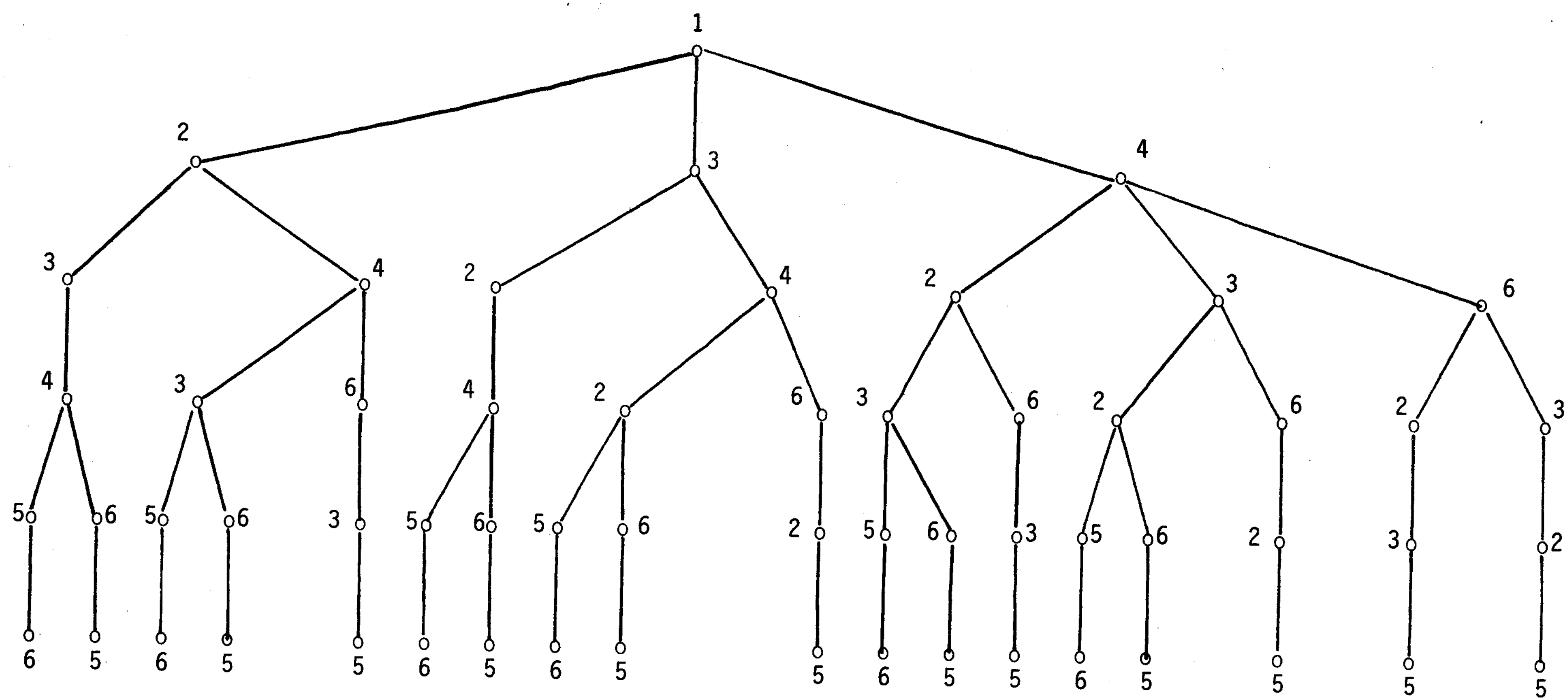


FIGURE II - 2

II.3.1.- ALGORITHME DE CONSTRUCTION DE A_Q .

Soit $G = (X,U)$ le graphe support du questionnaire Q . $X = E \cup F$ où F représente l'ensemble des questions de Q. Notons A_Q^1 le graphe défini par :

$$A_Q^1 = (X^1,U^1) \quad \text{où} \quad X^1 = \{\alpha\} \quad \text{racine du questionnaire } Q$$

$$U^1 = \{\emptyset\}$$

Supposons construit

$$A_Q^{k-1} = (X^{k-1}, U^{k-1})$$

Notons E_{k-1} l'ensemble des sommets pendants de A_Q^{k-1}

C_{k-1} l'ensemble des chemins reliant α à un élément de E_{k-1}

$\forall\, C \in C_{k-1}$, soit X_C l'ensemble des sommets de C , notons ξ_C l'élément de X_C appartenant à E_{k-1} , effectuons la construction suivante :

Pour tout $x \in F - X_C$, si $\hat{\Gamma}^{-1}(x) \subset X_C$, alors prolonger A_Q^{k-1} de la manière suivante :

$$X^k = X^{k-1} \cup x$$

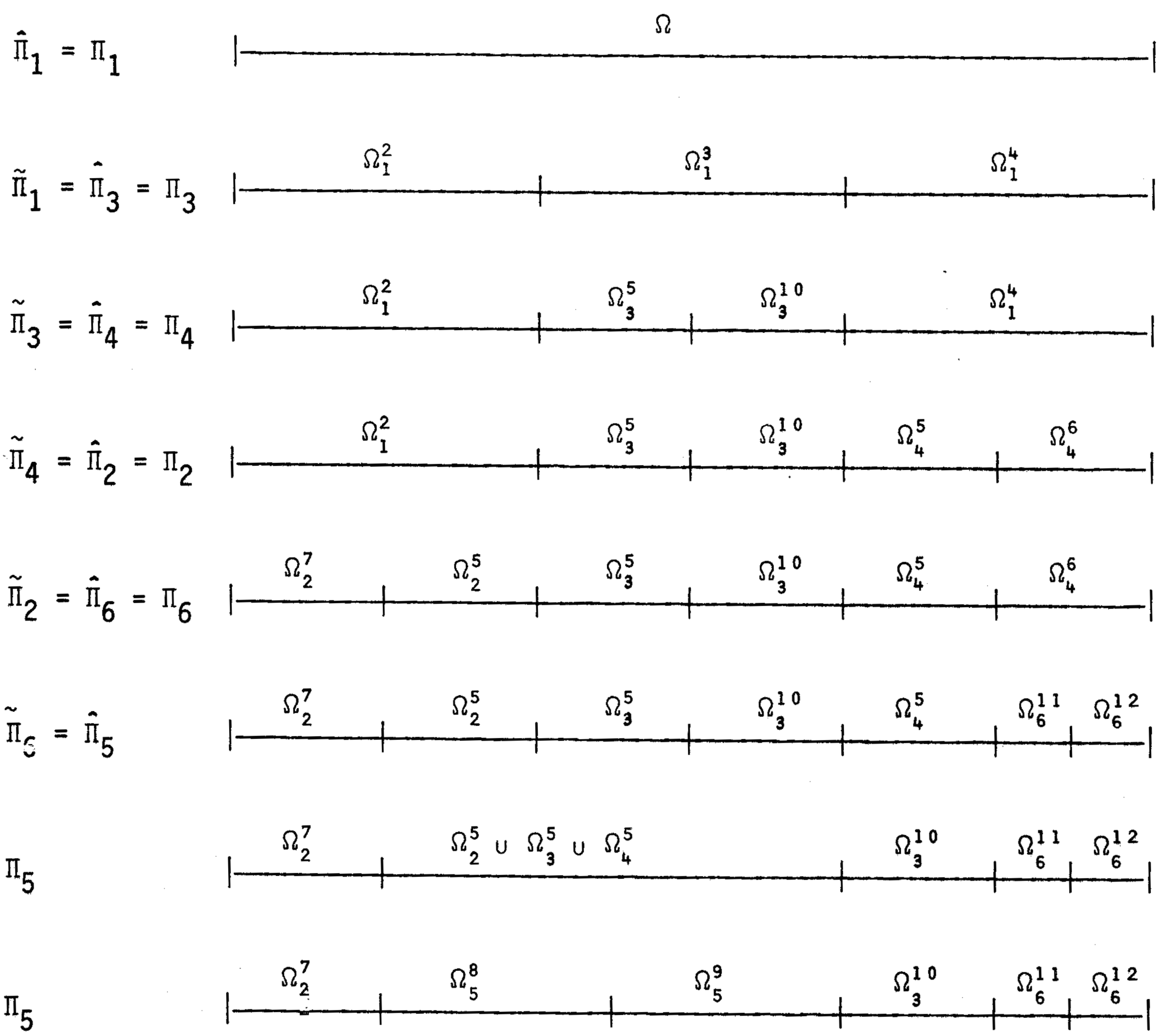
$$U^k = U^{k-1} \cup (\xi_C,x)$$

Ceci nous permet de construire A_Q^k en fonction de A_Q^{k-1} , le processus est itératif et se termine lorsque k est égal au nombre des sommets non terminaux (questions) de Q .

II.4.- INFORMATIONS ASSOCIÉES À UN CHEMIN DE L'ARBORESCENCE DE L'ORDRE DES QUESTIONS.

Soit, par exemple, le chemin (1,3,4,2,6,5) de la figure II-2.

Représentons, sur la figure suivante, toutes les partitions associées au chemin (1,3,4,2,6,5) :



Ainsi, à tout sommet de ce chemin, nous associons trois partitions de Ω : considérons le sommet 5 :

$$\hat{\Pi}_5 = (\Omega_2^7 \, , \, \Omega_2^5 \, , \, \Omega_3^5 \, , \, \Omega_3^{10} \, , \, \Omega_4^5 \, , \, \Omega_6^{11} \, , \, \Omega_6^{12})$$

$$\Pi_5 = (\Omega_2^7 \, , \, \Omega_2^5 \cup \Omega_3^5 \cup \Omega_4^5 \, , \, \Omega_3^{10} \, , \, \Omega_6^{11} \, , \, \Omega_6^{12})$$

$$\tilde{\Pi}_5 = (\Omega_2^7 \, , \, \Omega_5^8 \, , \, \Omega_5^9 \, , \, \Omega_3^{10} \, , \, \Omega_6^{11} \, , \, \Omega_6^{12})$$

$\hat{\Pi}_5$ est la partition de Ω obtenue après avoir posé les questions 1, 3, 4, 2 et 6 .

Π_5 est la partition de Ω obtenue après le regroupement, lié à la question 5 , de certains éléments de $\hat{\Pi}_5$.

$\tilde{\Pi}_5$ est la partition de Ω obtenue après avoir posé la question 5 .

La suite $(\hat{\Pi}_1 \, , \, \Pi_1 \, , \, \tilde{\Pi}_1 \, , \, \hat{\Pi}_3 \, , \, \Pi_3 \, , \, \tilde{\Pi}_3, \ldots, \hat{\Pi}_5 \, , \, \Pi_5 \, , \, \tilde{\Pi}_5)$ définit alors une s.e.c. de partitions de Ω associée au chemin (1,3,4,2,6,5).

De manière générale, nous associons à tout chemin C de l'arborescence A_Q , une s.e.c. [C] définie par la suite :

$$(\hat{\Pi}_{i_1} \, , \, \Pi_{i_1} \, , \, \tilde{\Pi}_{i_1} \, , \, \ldots \, , \, \hat{\Pi}_{i_n} \, , \, \Pi_{i_n} \, , \, \tilde{\Pi}_{i_n})$$

où n est le nombre de questions de Q :

Notons $a(i)$ le nombre de réponses d'une question i :

$$\hat{\Pi}_{i_1} = \Pi_1 = (\{\Omega\}) \quad , \quad \tilde{\Pi}_{i_1} = (\Omega_{i_1}^{j_1} \, , \, \ldots \, , \, \Omega_{i_1}^{j_{a(i_1)}})$$

où $\forall \, p \leq j_1 \, , \, \ldots \, , \, j_{a(i_1)} \quad , \quad j_p \in \Gamma_{i_1}$

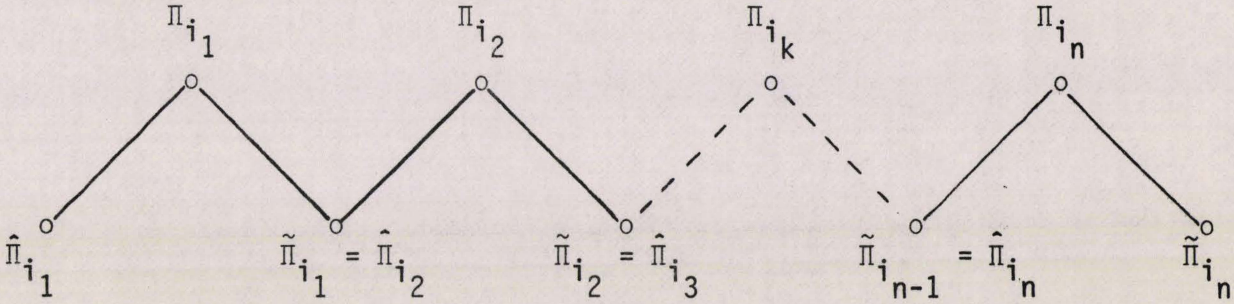
$\forall \, k = 2,\ldots,n$; posons $\hat{\Pi}_{i_k} = \tilde{\Pi}_{i_{k-1}}$

$\Pi_{i_k} = (\mathcal{D}(i_k) \, , \, \hat{\Pi}_{i_k} - \mathcal{D}(i_k))$ où $\hat{\Pi}_{i_k} - \mathcal{D}(i_k)$ représente l'ensemble des éléments de $\hat{\Pi}_{i_k}$ non inclus dans $\mathcal{D}(i_k)$.

$$\tilde{\Pi}_{i_k} = (\Omega_{i_k}^{j_1}, \ldots, \Omega_{i_k}^{j_{a(i_k)}}, \hat{\Pi}_{i_k} - \mathcal{D}(i_k))$$

où $\forall\ p = 1,\ldots,j_{a(i_k)} : j_p \in \Gamma_{i_k}$

Remarquons que toute s.e.c. [C] de partitions de Ω associée à un chemin C de A_Q, où Q est un questionnaire ayant n questions, est de la forme :



plus précisèment, $\forall$ k = 1,...,n nous avons les relations :

$$\hat{\Pi}_{i_k} \preccurlyeq \Pi_{i_k} \quad , \quad \Pi_{i_k} \succcurlyeq \tilde{\Pi}_{i_k}$$

ainsi, la s.e.c. [C] est alternée.

Si nous supposons maintenant que l'ensemble des partitions de Ω *est muni d'une information généralisée* H *,* nous avons les quantités d'information suivantes :

$$\Delta^-([C]) = \sum_{k=1}^{n} [H(\hat{\Pi}_{i_k}) - H(\Pi_{i_k})]$$: information absorbée par la s.e.c. [C]

$$\Delta^+([C]) = \sum_{k=1}^{n} [H(\tilde{\Pi}_{i_k}) - H(\Pi_{i_k})]$$: information apportée par la s.e.c. [C]

$$V([C]) = H(\tilde{\Pi}_{i_n}) - H(\{\Omega\})$$
$$= H(\tilde{\Pi}_{i_n}) \quad \text{si} \quad H(\{\Omega\}) = 0$$
: information transmise par la s.e.c. [C].

$$\Delta([C]) = \Delta^+([C]) + \Delta^-([C]) .$$

REMARQUES.-

- La valeur de $V([C])$ est indépendante de la s.e.c. $[C]$, donc du chemin C de A_Q ; ainsi l'*information transmise par le questionnaire est indépendante de l'ordre dans lequel l'on pose les questions*. Par contre, l'exemple du paragraphe II-7 montre qu'il n'en est généralement pas de même pour les informations absorbée et apportée.

- Localement les quantités suivantes sont définies :

$\Delta^-(i_k) = H(\hat{\Pi}_{i_k}) - H(\Pi_{i_k})$ information absorbée par la question i_k

$\Delta^+(i_k) = H(\tilde{\Pi}_{i_k}) - H(\Pi_{i_k})$ information apportée par la question i_k

$\Delta(i_k) = \Delta^-(i_k) + \Delta^+(i_k)$

$V(i_k) = H(\tilde{\Pi}_{i_k}) - H(\hat{\Pi}_{i_k})$ information transmise par la question i_k

- Nous avons clairement les propriétés :

$$\Delta^+(i_k) = V(i_k) + \Delta^-(i_k)$$

information apportée par la question = information transmise par la question
+ information absorbée par la question.

$$\Delta^+(i_k) \geq V(i_k)$$

information apportée par la question $\geq$ information transmise par la question.

- Globalement, nous avons les relations suivantes :

$$\Delta^-([C]) = \sum_{k=1}^{n} \Delta^-(i_k)$$

$$\Delta^+([C]) = \sum_{k=1}^{n} \Delta^+(i_k)$$

$$\Delta([C]) = \sum_{k=1}^{n} \Delta(i_k)$$

$$V([C]) = \sum_{k=1}^{n} V(i_k)$$

- Propriétés :

P1 : $\Delta^+([C]) = V([C]) + \Delta^-([C])$

Information apportée = information transmise + information absorbée

P2 : $\Delta^+([C]) \geq V([C])$

Information apportée $\geq$ information transmise

P3 : $\Delta^-([C]) = 0$ si et seulement si le support de Q est arborescent.

II.5.- PROPRIÉTÉ DE LOCALISATION ET INFORMATIONS ASSOCIÉES À UN CHEMIN DE A_Q.

Etant donnés un questionnaire Q de racine α , A_Q l'arborescence de l'ordre des questions associé à Q , et i une question de Q , soient C_1 et C_2 deux chemins de A_Q reliant α à un sommet pendant :

Notons i_1 (resp. i_2) le sommet de C_1 (resp. C_2) correspondant à la question i .

II.5.1.- THEOREME.

Si H vérifie la propriété de localisation, alors noue avons les égalités :

$\Delta^+(i_1) = \Delta^+(i_2)$

$\Delta^-(i_1) = \Delta^-(i_2)$

$\Delta(i_1) = \Delta(i_2)$

$V(i_1) = V(i_2)$

DEMONSTRATION.-

Montrons que $\Delta^+(i_1) = \Delta^+(i_2)$

$$\Delta^+(i_1) = H(\tilde{\Pi}_{i_1}) - H(\Pi_{i_1})$$

$$\Delta^+(i_1) = H(\Omega_{i_1}^{j_1},\dots,\Omega_{i_1}^{j_{a(i_1)}} , \hat{\Pi}_{i_1} - \mathcal{D}(i_1)) - H(\mathcal{D}(i_1) , \hat{\Pi}_{i_1} - \mathcal{D}(i_1))$$

Notons $$B = \bigcup_{A \in \hat{\Pi}_{i_1} - \mathcal{D}(i_1)} A$$

H vérifie la propriété de localisation entraîne que :

$$\Delta^+(i_1) = H(\Omega_{i_1}^{j_1},\ldots,\Omega_{i_1}^{j_{a(i_1)}},B) - H(\mathcal{D}(i_1),B)$$

or
$$\mathcal{D}(i_1) = \mathcal{D}(i_2) .$$

ce qui donne :

$$\Delta^+(i_1) = H(\Omega_{i_2}^{j_1},\ldots,\Omega_{i_2}^{j_{a(i_2)}},B) - H(\mathcal{D}(i_2),B)$$

H vérifiant la propriété de localisation entraîne que :

$$\Delta^+(i_1) = H(\Omega_{i_2}^{j_2},\ldots,\Omega_{i_2}^{j_{a(i_2)}}, \hat{\Pi}_{i_2} - \mathcal{D}(i_2)) - H(\mathcal{D}(i_2), \hat{\Pi}_{i_2} - \mathcal{D}(i_2))$$

d'où :

$$\Delta^+(i_2) = \Delta^+(i_2)$$

La démonstration est analogue pour Δ^- , Δ et V .

<u>REMARQUE</u>.- Si $\Delta^+(i_1) = \Delta^+(i_2)$, nous en déduisons que H vérifie la propriété de localisation pour toutes les partitions "réalisées" par Q , c'est-à-dire pour l'ensemble des partitions de Ω pouvant être obtenues par les s.e.c. de partitions associées aux chemins de A_Q .

<u>II.5.2.- COROLLAIRE</u>.

Si H *vérifie la propriété de localisation, alors, quels que soient* C_1 , C_2 , *chemins de* A_Q *reliant* α *à un sommet pendant, l'on a :*

$$\Delta^+([C_1]) = \Delta^+([C_2]) \ , \ \Delta^-([C_1]) = \Delta^-([C_2]) \ , \ \Delta([C_1]) = \Delta([C_2]) \ .$$

Ainsi, H *vérifie la propriété de localisation entraîne que les notions d'information absorbée et apportée relatives à une question* i *ou à un questionnaire* Q *sont indépendantes de l'ordre dans lequel sont posées les questions.*

Ce résultat fondamental nous permet alors de définir dans un cadre général les notions d'Informations d'un questionnaire :

II.5.3.- INFORMATIONS D'UN QUESTIONNAIRE.

Soient Q un questionnaire posé sur un ensemble Ω , H une mesure d'information généralisée définie sur le treillis T des partitions de Ω et vérifiant la propriété de localisation , $[C]$ une s.e.c. quelconque de partitions associée au questionnaire Q , alors nous avons les définitions :

$I(Q) = V([C])$: *information transmise par* Q

$J(Q) = \Delta^+([C])$: *information apportée (traitée) par* Q

$e(Q) = \Delta^-([C])$; *information absorbée par* Q

nous avons, de plus, les relations :

- $J(Q) = I(Q) + e(Q)$
- $J(Q) \geq I(Q)$
- $e(Q) = 0$ *si et seulement si le questionnaire est arborescent.*

II.6.- LIEN ENTRE LES QUESTIONNAIRES ET NOS RÉSULTATS.

A.- *Soit H une information moyenne* définie sur l'ensemble des partitions de Ω , par $H(\Pi) = \sum_{A\in\Pi} p(A)\, I(A)$

Notons

$$D(i) = \bigcup_{j\in\Gamma_i} \Omega_i^j = \bigcup_{h\in\Gamma_i^{-1}} \Omega_h^i$$

en raison des propriétés de localisation de l'information moyenne, avons les relations suivantes :

$\Delta^-(i) = \sum_{h\in\Gamma_i^{-1}} p(\Omega_h^i)\, I(\Omega_h^i) - p(D(i))\, I(D(i))$: information absorbée par le question i .

$\Delta^+(i) = \sum_{j\in\Gamma_i} p(\Omega_i^j)\, I(\Omega_i^j) - p(D(i))\, I(D(i))$: information apportée par le question i .

$V(i) = \Delta^+(i) - \Delta^-(i)$: information transmise par la question i .

Ces relations s'écrivent avec les notations utilisées dans les questionnaires :

$$e(i) = \Delta^{-}(i) = \sum_{h \in \Gamma_i^{-1}} p(h,i)\ I(\Omega_h^i) - p(i)\ I(D(i))$$

$$J(i) = \frac{1}{p(i)}\ \Delta^{+}(i) = \frac{1}{p(i)} \sum_{j \in \Gamma_i} p(i,j)\ I(\Omega_i^j) - I(D(i))$$

$$I(i) = p(i)\ \ J(i) - e(i) = V(i)$$

B.- *<u>Si H n'est pas une information moyenne</u>* mais vérifiant la propriété de localisation, nous avons

$$e(i) = \Delta^{-}(i) = H(\hat{\Pi}_i) - H(\Pi_i)$$

$$J(i) = \frac{1}{p(i)}\ \Pi^{+}(i) = \frac{1}{p(i)} [H(\tilde{\Pi}_i) - H(\Pi_i)]$$

$$I(i) = p(i)\ J(i) - e(i) = V(i) = H(\tilde{\Pi}_i) - H(\hat{\Pi}_i)$$

Le vocabulaire introduit dans les s.e.c. est celui utilisé en Théorie des Questionnaires, à une différence près :

$\Delta^{+}([C])$, information *apportée* par la s.e.c. [C] est associé à J(Q) information *traitée* par le questionnaire Q au sens de C.F. PICARD.

Appliquons les résultats à qu elques mesures d'information utilisées en Théorie des Questionnaires.

voir tableau page suivante

.../...

MESURE	I(A)	INFORMATION APPORTÉE PAR UNE QUESTION	INFORMATION ABSORBÉE PAR UNE QUESTION
SHANNON	$\text{Log}\,\frac{1}{P(A)}$	$\sum_{j\in\Gamma_i} p(i,j)\,\text{Log}\,\frac{p(i)}{p(i,j)}$	$\sum_{h\in\Gamma_i^{-1}} p(h,i)\,\text{Log}\,\frac{p(i)}{p(h,i)}$
HYPERBOLIQUE	$\frac{1}{p^2(A)} - 1$	$\sum_{j\in\Gamma_i} p(i,j)\left[\frac{1}{p^2(i,j)} - 1\right]$ $- p(i)\left[\frac{1}{p^2(i)} - 1\right]$	$\sum_{h\in\Gamma_i^{-1}} p(h,i)\left[\frac{1}{p^2(h,i)} - 1\right]$ $- p(i)\left[\frac{1}{p^2(i)} - 1\right]$
AGGARWAL CESARI PICARD	$\alpha - 1$ $P(A)$ $0<\alpha<1$	$\sum_{j\in\Gamma_i} p^{\alpha}(i,j) - p^{\alpha}(i)$	$\sum_{h\in\Gamma_i^{-1}} p^{\alpha}(h,i) - p^{\alpha}(i)$
HAVRDA CHARVAT	$\alpha - 1$ $\frac{P(A) - 1}{a^{1-\alpha} - 1}$ $a \neq 1$ $0<\alpha$	$\frac{\sum_{j\in\Gamma_i} p^{\alpha}(i,j) - p^{\alpha}(i)}{a^{1-\alpha} - 1}$	$\frac{\sum_{h\in\Gamma_i^{-1}} p^{\alpha}(h,i) - p^{\alpha}(i)}{a^{1-\alpha} - 1}$
BELIS	$\text{Log}\,\frac{1}{P(A)}$	$\sum_{j\in\Gamma_i} w(i,j)\,\text{Log}\,\frac{p(i)}{p(i,j)}$	$\sum_{h\in\Gamma_i^{-1}} w(h,i)\,\text{Log}\,\frac{p(i)}{p(h,i)}$

II.7.- APPLICATION AUX INFORMATIONS NE VÉRIFIANT PAS LA PROPRIÉTÉ DE LOCALISATION.

L'information de RENYI ne vérifie pas la propriété de localisation; soit le questionnaire de la figure II-1, considérons les chemins de A_Q (figure II-2) suivants :

$$C_1 = (1,3,4,2,6,5)$$

$$C_2 = (1,2,3,4,5,6)$$

sur C_1 :

$$\hat{\Pi}_5 = \left|\ \Omega_2^7\ \middle|\ \Omega_2^5\ \middle|\ \Omega_3^5\ \middle|\ \Omega_4^5\ \middle|\ \Omega_3^{10}\ \middle|\ \Omega_6^{11}\ \middle|\ \Omega_6^{12}\ \right|$$

$$\Pi_5 = \left|\ \Omega_2^7\ \middle|\ \Omega_2^5 \cup \Omega_3^5 \cup \Omega_4^5\ \middle|\ \Omega_3^{10}\ \middle|\ \Omega_6^{11}\ \middle|\ \Omega_6^{12}\ \right|$$

Prenons $p(2,7) = p(2,5) = p(3,5) = 0.2$

$p(4,5) = p(3,10) = p(6,11) = p(6,12) = 0.1$

ce qui donne $\Delta^-(5) = 1$ avec $\alpha = 2$ pour l'information de RENYI

sur C_2 :

$$\hat{\Pi}_5 = \left|\ \Omega_2^7\ \middle|\ \Omega_2^5\ \middle|\ \Omega_3^5\ \middle|\ \Omega_4^5\ \middle|\ \Omega_3^{10}\ \middle|\ \Omega_4^6\ \right|$$

$$\Pi_5 = \left|\ \Omega_2^7\ \middle|\ \Omega_2^5 \cup \Omega_3^5 \cup \Omega_4^5\ \middle|\ \Omega_3^{10}\ \middle|\ \Omega_4^6\ \right|$$

Les probabilités prises pour C_1 entraînent que :

$p(2,7) = p(2,5) = p(3,5) = P(4,6) = 0.2$

$p(4,5) = p(3,10) = 0.1$

ce qui donne $\Delta^-(5) = 0.91754$ avec $\alpha = 2$ pour l'information de RENYI. Ainsi, sur cet exemple, l'information absorbée par la question 5 dépend de toutes les questions posées antérieurement. L'arborescence de l'ordre où sont posées les questions nous fournit l'ensemble de toutes les s.e.c. de partitions associées au questionnaire. Les s.e.c. de partitions nous donnant un moyen de mesurer toutes les informations du questionnaire, nous avons donc, si la mesure de l'information ne vérifie pas la propriété de localisation, la possibilité de déterminer le questionnaire optimal dans le sens où l'information absorbée sera minimale.

C H A P I T R E   III

-=-=-=-=-=-=-=-=-=-=-=

APPLICATIONS AUX PROBLEMES DE CLASSIFICATION

-=-=-=-=-=-=-=-=-=-=-=

III.- APPLICATIONS AUX PROBLEMES DE CLASSIFICATION.

Etant donné un tableau à double entrée $(\mathcal{O},\Omega)$ fournissant les réponses d'un ensemble Ω d'individus à un ensemble $\mathcal{O}$ d'observateurs (caractères), le but de la classification automatique est de rechercher des sous-ensembles (classes) d'individus "homogènes". L'utilisation de mesures d'information adaptées à ces classes nous permet d'élaborer des moyens originaux de traitement.

Nous considérons le tableau $(\mathcal{O},\Omega)$ comme un produit de partitions de Ω ou, dans d'autres cas, comme un produit de sous-ensembles (événements) de Ω . Cette dernière éventualité implique que la matrice de donnée soit codée à l'aide de deux symboles, en général 0,1 . Nous avons donc, suivant des techniques bien connues [2] , démultiplié chaque caractère en caractères élémentaires (2 modalités de réponses par caractère).

III.1.- PRINCIPES GÉNÉRAUX DES APPLICATIONS.

L'ensemble des applications réalisées se décompose en 3 étapes :

- CONSTRUCTION DE CLASSES DE Ω .

  Nous proposons un algorithme (programme GENER) permettant, à partir du tableau de données $(\mathcal{O},\Omega)$, de construire une partition de Ω pour laquelle :

  . le nombre de classes de cette partition n'est pas imposé.

  . chaque classe est "homogène" au sens intuitif du terme, précisé lors de la description du programme correspondant, où à chaque classe correspond un sous-ensemble d'observateurs pour lesquels l'ensemble des individus de la classe a le même comportement, c'est-à-dire "pratiquement" les mêmes réponses.

  L'algorithme nous permet alors de mettre en évidence, pour chaque classe de la partition de Ω , l'ensemble des observateurs caractérisant cette classe.

- CLASSIFICATION HIERARCHIQUE D'UN ENSEMBLE DE CLASSES DE Ω .

Ayant obtenu, par une méthode quelconque, un ensemble de classes homogènes de Ω , nous construisons par la méthode classique du "plus proche voisin" [10] , une *classification hiérarchique indicée* [10] de ces classes en formant une suite de partitions suggérée au chapitre I . (programme ARBRE). La distance servant à indicer la classification permet de déterminer les niveaux "intéressants" de l'arbre associé à cette hiérarchie.

- CONSTRUCTION D'UN QUESTIONNAIRE ASSOCIE A UNE CLASSIFICATION HIERARCHIQUE.

Le programme (QUEST) construit alors un questionnaire (hiérarchie d'observateurs) permettant, en partant de Ω , de retrouver les partitions de Ω intéressantes mises en relief par l'algorithme précédent.
De plus, à chaque question (observateur sélectionné) sont fournis l'ensemble des observateurs strictement équivalents ou équivalents à un pourcentage fixé.

III.2.- ALGORITHME DE CONSTRUCTION DE CLASSES DE Ω . (PROGRAMME GENER).

III.2.1.- DESCRIPTION.

Le principe de la méthode est le suivant : à partir du tableau de données initial $(\mathcal{O},\Omega)$, choisissons arbitrairement un élément ω_0 de Ω .

L'ensemble des réponses de ω_0 à $\mathcal{O}$ définit deux sous-ensembles de $\mathcal{O}$:

- $\mathcal{O}_1^{\omega_0}$: l'ensemble des tests auxquels ω_0 répond 1 .
- $\mathcal{O}_0^{\omega_0}$: l'ensemble des tests auxquels ω_0 répond 0 .

Un individu ω de Ω sera "voisin" de ω_0 si et seulement si $\mathcal{O}_1^{\omega}$ et $\mathcal{O}_0^{\omega}$ sont "presque identiques" à $\mathcal{O}_1^{\omega_0}$ et $\mathcal{O}_0^{\omega_0}$.

Plus précisèment, soit $F = P(\mathcal{O})$, alors [7] la distance $\delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega})$ construite à partir de la mesure d'information de SHANNON sur les événements nous permet de mesurer le degré de ressemblance entre $\mathcal{O}_1^{\omega_0}$ et $\mathcal{O}_1^{\omega}$.

$$\delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega}) = \mathrm{Log}\ (\frac{P^2(\mathcal{O}_1^{\omega_0} \cup \mathcal{O}_1^{\omega})}{P(\mathcal{O}_1^{\omega_0}) \ .\ p(\mathcal{O}_1^{\omega})})$$

Cette quantité s'écrit :

$$\delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega}) = \mathrm{Log}\ (\frac{[P(\mathcal{O}_1^{\omega_0}) + P(\mathcal{O}_1^{\omega}) - P(\mathcal{O}_1^{\omega_0} \cap \mathcal{O}_1^{\omega})]^2}{P(\mathcal{O}_1^{\omega_0}) \ .\ P(\mathcal{O}_1^{\omega})})$$

ω_0 étant l'individu de référence, nous pouvons considérer que $P(\mathcal{O}_1^{\omega_0})$ est égal à une constante α .

Posons

$$y = P(\mathcal{O}_1^{\omega}) \quad , \quad z = P(\mathcal{O}_1^{\omega} \cap \mathcal{O}_1^{\omega_0})$$

z *mesure le nombre de réponses 1 communes à* ω *et* ω_0 .

$$\delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega}) = \mathrm{Log}\ (\frac{(\alpha + y - z)^2}{\alpha y}) \quad ,$$

On montre facilement que ,

. pour y fixé, δ décroit lorsque z croit.

Ceci s'interprète (figure III-2-1-1-(A)) : les deux individus ω et ω' vérifient $y = y'$ et $z_{\omega} > z_{\omega'}$

$$\delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega}) < \delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega'})$$

ce qui exprime bien le fait qu'intuitivement ω est de plus proche de ω_0

. pour z fixé.

- Si $z \geq \frac{\alpha}{2}$, δ croit lorsque y croit.

Les deux individus ω et ω' de la figure III-2-1-1 (B) vérifient

$$z_\omega = z_{\omega'} \quad , \quad z_\omega \geq \frac{\alpha}{2} \text{ et } y_\omega < y_{\omega'} \quad : \quad \delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega}) < \delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega'})$$

d'où ω est meilleur voisin de ω_0 que ω' .

- Si $z \leq \frac{\alpha}{2}$, δ est minimum lorsque $y = \alpha - z$

Dans ce cas δ prend la valeur $\mathrm{Log}\,\frac{4(\alpha-z)}{\alpha}$ (figure III-2-1-1 (C))

$\delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega}) \leq \delta(\mathcal{O}_1^{\omega_0} , \mathcal{O}_1^{\omega'})$ pour tout ω' vérifiant :

$$z_\omega = z_{\omega'} \quad , \quad z_\omega \leq \frac{\alpha}{2} .$$

De plus, $\delta = \mathrm{Log}\,\frac{4(\alpha-z)}{\alpha}$ est minimale quand $z = \frac{\alpha}{2}$ et est alors égale à 1 .

L'individu ω_m représenté sur la figure III-2-1-1 (C) représente donc le "meilleur" individu pouvant exister. Nous remarquons que $\mathcal{O}_1^{\omega}$ et $\mathcal{O}_1^{\omega_m}$ sont relativement éloignés, par conséquent cette valeur 1 pour δ nous donne un seuil minimum de voisinage entre ω_0 et un individu ω quelconque, au delà duquel il n'est pas réaliste de considérer ω comme voisin de ω_0 .

L'ensemble de ces propriétés assure donc que $\delta(\mathcal{O}_1^{\omega} , \mathcal{O}_1^{\omega_0})$ mesure bien le degré de ressemblance entre $\mathcal{O}_1^{\omega}$ et $\mathcal{O}_1^{\omega_0}$.

Nous décidons que $\mathcal{O}_1^{\omega_0}$ est voisin à k près de $\mathcal{O}_1^{\omega_0}$ si :

$$\delta(\mathcal{O}_1^{\omega} , \mathcal{O}_1^{\omega_0}) \leq k < 1 .$$

L'algorithme consiste à agglomérer tous les individus ω tels que $\delta(\mathcal{O}_1^{\omega} , \mathcal{O}_1^{\omega_0}) \leq k$, ce qui nous donne un noyau $N \in \mathcal{P}(\Omega)$; l'étape suivante détermine, à partir de ce noyau, l'ensemble des observateurs caractérisant celui-ci :

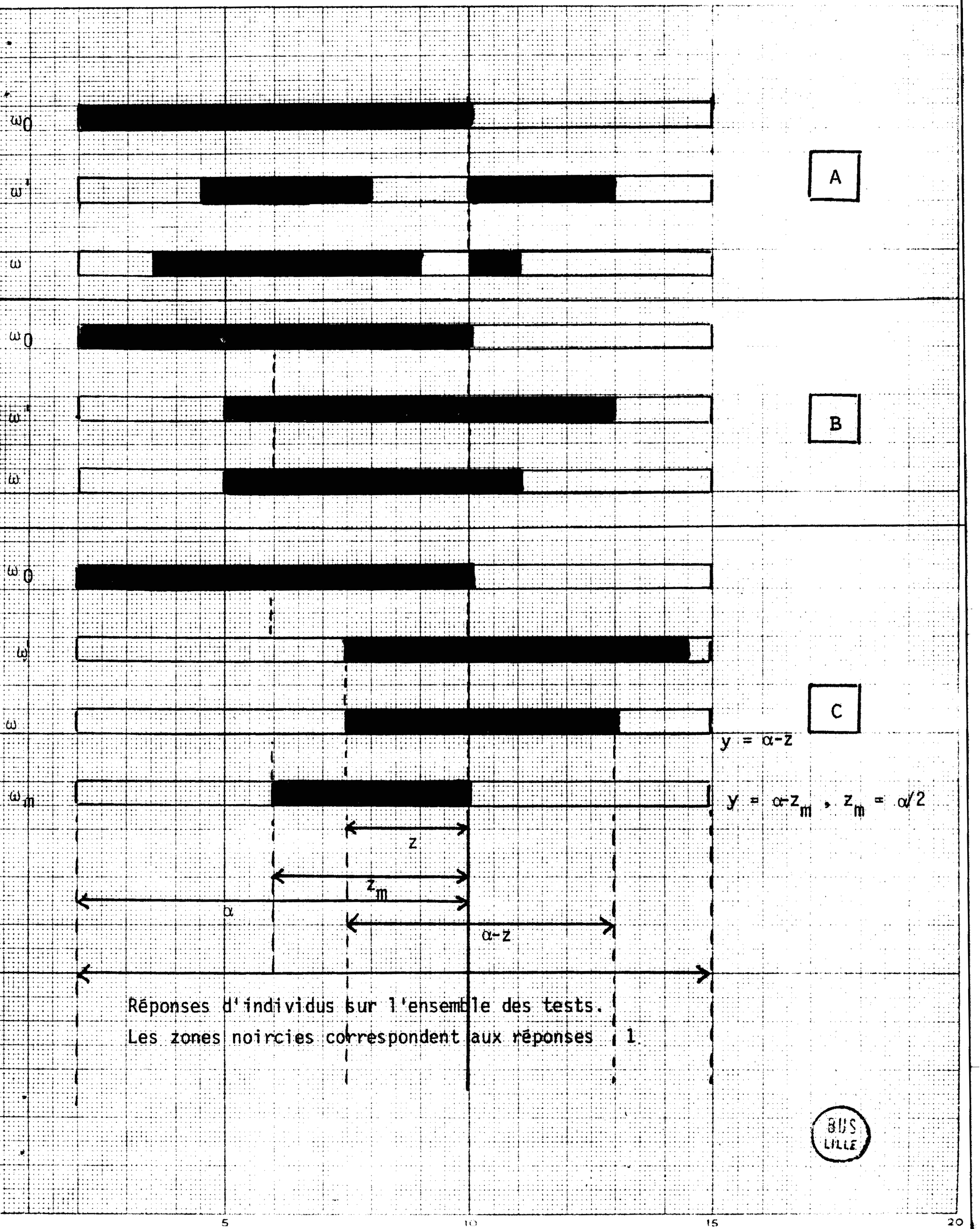


FIGURE III.2.1.1.

. Soit t un observateur quelconque de O ; notons t_1^N l'ensemble des éléments de N répondant 1 à t , t_0^N l'ensemble des éléments de N répondant 0 à t .

Comme ci-dessus, nous définissons :

$$\delta(t_1^N , \hat{t}_1^N) \quad \text{et} \quad \delta(t_0^N , \hat{t}_0^N)$$

où $\hat{t}_1^N$ et $\hat{t}_0^N$ sont deux tests fictifs auxquels tous les individus de N répondraient respectivement I et 0 .

Si $\delta(t_1^N , \hat{t}_1^N) = 0$ (resp. $\delta(t_0^N , \hat{t}_0^N)$) , il est clair que l'observateur t caractérise le groupe t_1^N (resp. t_0^N) . Si aucune de ces quantités n'est nulle, on peut choisir la quantité minimale; mais dans ce cas on risque de prendre un observateur "caractéristique" de manière arbitraire ($\delta(t_1^N , \hat{t}_1^N)$ voisin de $\delta(t_0^N , \hat{t}_0^N)$) . Aussi, à un observateur t et un noyau N , nous associons un observateur non caractéristique $\tilde{t}$ construit de telle sorte qu'il vérifie les propriétés suivantes :

- le nombre de réponses 1 de $\tilde{t}$ est égale à $\mathrm{card}(N)/2$.
- $\delta(t_1^N , \tilde{t}_1^N)$ est minimum.

Dans ces conditions, la quantité β égale à :

$$\min(\delta(t_1^N , \hat{t}_1^N) , \delta(t_0^N , \hat{t}_0^N) , \delta(t_1^N , \tilde{t}_1^N))$$

nous permet de décider si l'observateur est caractéristique ou non du noyau N , de la manière suivante :

- Si $\beta = \delta(t_1^N , \hat{t}_1^N)$ alors l'observateur t est caractéristique (réponse 1) du noyau N .

- Si $\beta = \delta(t_0^N , \hat{t}_0^N)$ alors l'observateur t est caractéristique (réponse 0) du noyau N

- Si $\beta = \delta(t_1^N , \tilde{t}_1^N)$ alors l'observateur t n'est pas caractéristique du noyau N .

Ayant construit l'ensemble F^N des observateurs caractéristiques du noyau N , nous distinguons les sous-ensembles F_1^N et F_0^N des observateurs caractéristiques pour la réponse 1 ou pour la réponse 0 .
F_1^N et F_0^N définissent un individu type ω_1 (élément de Ω) (cet individu étant éventuellement fictif) répondant 1 sur F_1^N et 0 sur F_0^N .

Le processus devient maintenant itératif en opérant avec ω_1 comme on a fait précédemment avec ω_0 .

L'algorithme se termine lorsque le noyau obtenu à l'itération n est identique à celui obtenu à l'itération n-1 , nous avons alors obtenu une classe homogène de Ω .

On reprend alors au hasard un individu ne se trouvant pas dans l'ensemble des classes formées et l'on réitère le processus pour former une nouvelle classe en ne considérant que le sous-ensemble des individus non classés.

Le programme s'arrête lorsque tous les individus de Ω se trouvent dans les classes.

STRUCTURE DU PROGRAMME GENER.

1.- Lecture des paramètres
2.- RESIDU = Ω
3.- Tirage de ω_0 au hasard dans RESIDU
4.- Constitution du noyau N
5.- Recherche des observateurs caractéristiques de N
6.- Constitution du nouveau noyau M
7.- Si M n'est pas identique à N , faire : N devient M et aller en 5
8.- Oter du RESIDU la classe M formée
9.- Si le RESIDU n'est pas vide, aller en 3
10.- FIN.

III.2.2.- RESULTATS.

Le programme GENER a été appliqué sur des données fournies par des biologistes : la population étudiée (ENTEROBACTER) comporte 175 souches de bactéries dont on connaît les réponses à 199 tests en 0,1. Le programme a été exécuté pour différentes valeurs de k (seuil d'agrégation) :

k variant de 0.05 à 0.50 avec un pas de 0.05.

Les résultats bruts sont représentés, sur la figure III.2.2.1 (page III.11), dans la partie droite où figurent les regroupements d'individus effectués pour les différentes valeurs de k . L'analyse de cette représentation met en évidence la structure suivante :

- existence de classes élémentaires : A_1 , A_2 , C_1 , C_2 , B , D, E, ELAT , EAAT , EV , EAG .
- emboîtement des classes élémentaires : figure III.2.2.2. et figure III.2.2.1. (partie gauche).

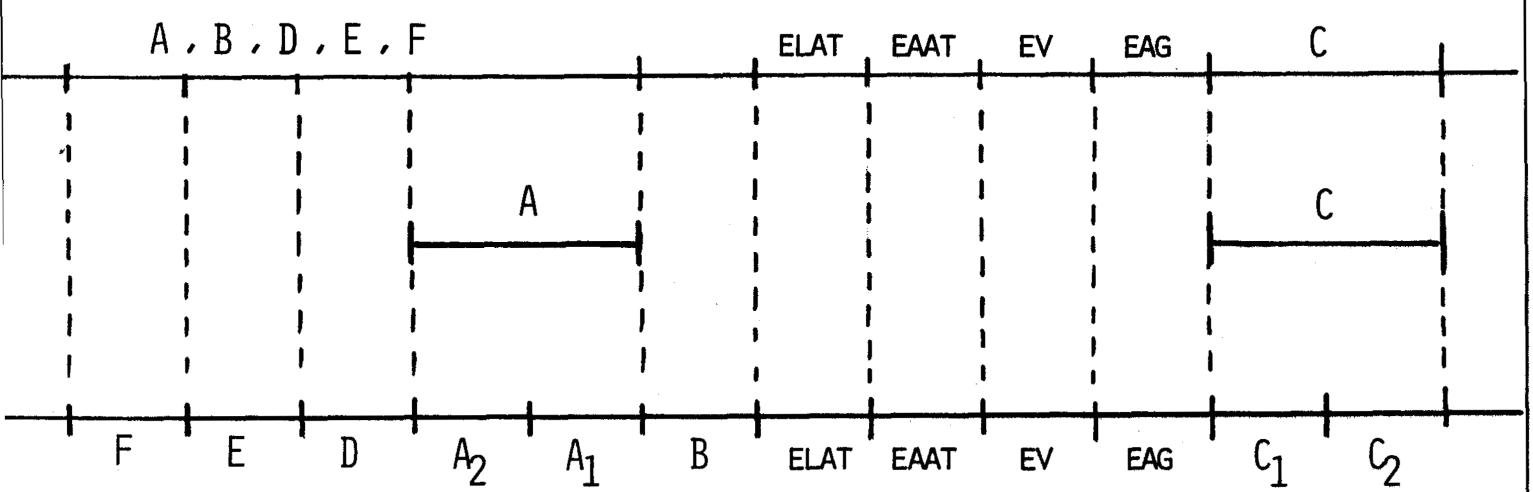


FIGURE III.2.2.2.

Afin de s'assurer de la cohérence de ces résultats, nous avons soumis ces mêmes données à des méthodes "classiques" : classification hiérarchique et Analyse factorielle. La synthèse des résultats obtenus par celles-ci est représentée sur la figure III.2.2.3 (page 1[illegible]), nous remarquons que le programme GENER met en évidence des classes homogènes que retrouvent les méthodes classiques, ainsi que les regroupements de classes importants.

De plus, pous chaque classe ou regroupement de classes, nous avons la liste de tous les tests caractérisant ceux-ci.

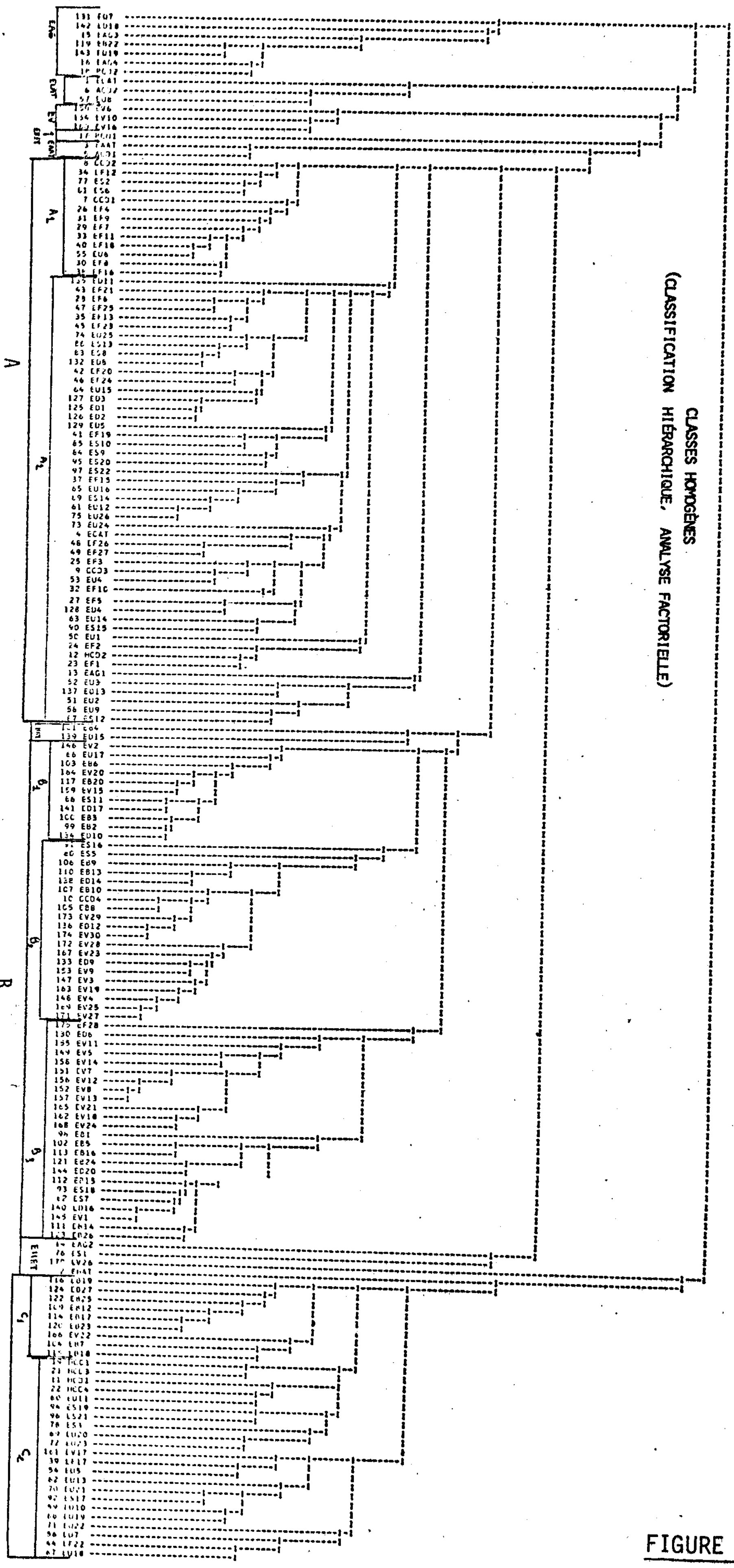
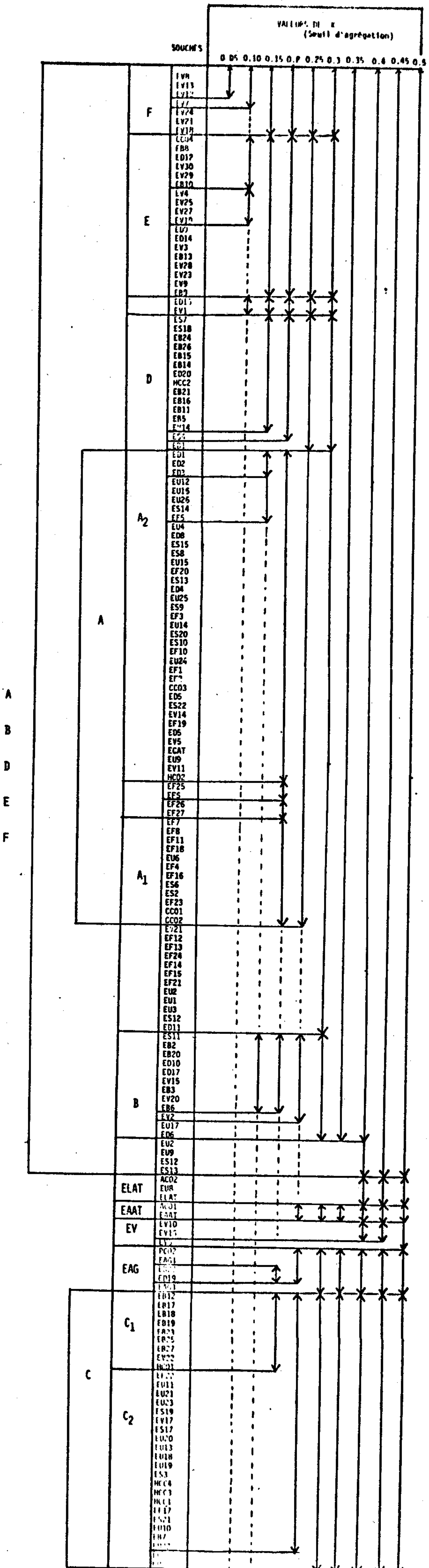


FIGURE III.2.2.3.



ENTEROBACTER

FIGURE III.2.2.4.

III.3.- ALGORITHME DE CLASSIFICATION HIÉRARCHIQUE INDICÉE DE CLASSE DE Ω .

III.3.1.- RAPPELS [10] .

Une hiérarchie sur un ensemble fini C de classes de Ω est un sous-ensemble H de $P(C)$ tel que :

1 - $x \in C : \{x\} \in H$

2 - $C \in H$

3 - $\forall\ (A,B) \in H \times H$

$A \cap B \neq \emptyset \Longrightarrow A \subset B$ ou $B \subset A$

La hiérarchie est dite indicée s'il existe de plus une application m de H dans $\bar{\mathbb{R}}^+$ telle que :

4 - $\forall\ x \in C\ ,\ m(\{x\}) = 0$

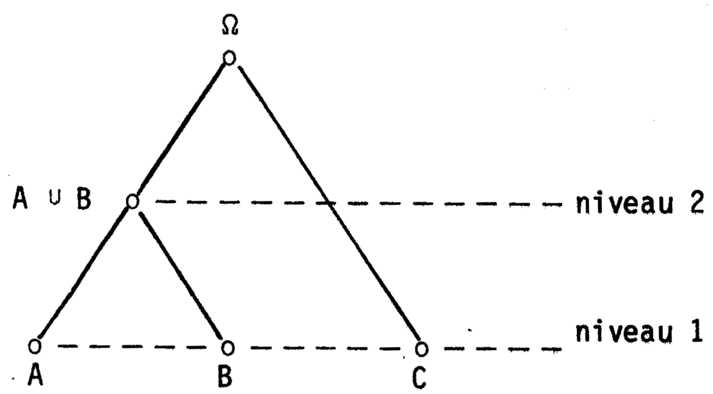
5 - $\forall\ (A,B) \in H \times H$

$A \subset B$ et $A \neq B \Longrightarrow m(A) < m(B)$.

III.3.2.- PROGRAMME ARBRE

Etant donné une partition Π (ensemble de classes homogènes) de Ω, nous désirons construire une hiérarchie de ces classes en regroupant des sous-ensembles restant homogènes par rapport aux classes non modifiées de Ω .

EXEMPLE :



Cet arbre est considéré sémantique dans le sens où le passage du niveau 1 au niveau 2 correspond à la construction d'une nouvelle classe $A \cup B$ restant homogène par rapport à C : nous utiliserons pour cette raison le terme de "hiérarchie sémantique" pour signifier que les regroupements se font en fonction du caractère d'homogénéïté des classes à regrouper.

L'algorithme classique de classification hiérarchique (méthode ascendante du plus proche voisin) consiste à construire une suite de partitions de moins en moins fines :

$$\Pi \leqslant \Pi_1 \leqslant \dots \leqslant \Pi_k \leqslant \dots \leqslant \Pi_n = \{\Omega\}$$

Toute classe de Π_k est réunion de classes de Π_{k-1} ; la cohésion des classes diminuant lorsque k augmente.

Le programme ARBRE effectue une telle construction en utilisant un tableau de distances δ défini sur $\Omega \times \Omega$ avec $\delta(i,j)$ = distance entre l'individu ω_i et l'individu ω_j, où la distance utilisée est celle du paragraphe III.2.1. (δ définie à partir de la mesure d'information de SHANNON sur les évênements).

Soient A et B deux éléments de Π_k (c'est-à-dire deux classes homogènes), nous mesurons la distance $d(A,B)$ entre A et B en utilisant une expression classique de la distance de HAUSDORF entre ensembles :

$$d(A,B) = \mathrm{Max}\{\mathop{\mathrm{Max}}_{a\in A}\mathop{\mathrm{Min}}_{b\in B} \delta(a,b) \; , \; \mathop{\mathrm{Max}}_{b\in B}\mathop{\mathrm{Min}}_{a\in A} \delta(a,b)\}$$

avec :

$$\delta(a,b) = \delta(\mathcal{O}_1^a \; , \; \mathcal{O}_1^b) = \mathrm{Log}\,\frac{P^2(\mathcal{O}_1^a \cup \mathcal{O}_1^b)}{P(\mathcal{O}_1^a)\; P(\mathcal{O}_1^b)}$$

L'emploi de cette distance d se justifie par les résultats expérimentaux intéressants qu'elle apporte : nous pensons que ces résultats sont également dûs aux propriétés mises en relief (paragraphe III.2.1.) de la distance δ utilisée pour définir d .

Soit H la hiérarchie des classes de Π (partitions de Ω), notons $\delta^*(\Pi,\Pi')$ la distance obtenue à partir de la mesure d'information H^* de SHANNON sur l'ensemble des partitions de Ω [7] ,

$$\delta^*(\Pi,\Pi') = \sum_{(A,B)\in\Pi\times\Pi'} P(A \cap B) \operatorname{Log}\left(\frac{P(A) \cdot P(B)}{P^2(A \cap B)}\right)$$

H^* vérifie l'hypothèse H_1 du chapitre I d'où :

$$\delta^*(\Pi,\Pi') = H^*(\Pi) + H^*(\Pi') - H^*(\Pi \wedge \Pi')$$

on en déduit que $\delta^*(\Pi,.)$ est une application strictement croissante pour la suite de partitions Π_i $(i = 1,\ldots,n)$ construites par le programme ARBRE, car :

$$\Pi \prec \Pi_1 \prec \cdots \prec \Pi_k \prec \cdots \prec \Pi_n = \{\Omega\}$$

Les conclusions du paragraphe I-6 nous permettent d'en déduire que la suite de partitions $\{\Pi_i\}$ résume de moins en moins l'information de Π , δ étant monotone, une variation importante de la quantité

$$\delta^*(\Pi,\Pi_{i+1}) - \delta^*(\Pi,\Pi_i)$$

signifie que le niveau Π_i est stable (homogène) : le passage de Π_i à Π_{i+1} provoque un regroupement de classes relativement éloignées les unes des autres.

Enfin, δ^* permet d'indicer la classification hiérarchique H : en effet, soit $A \in H$, notons A_i , $i \in I = \{1,\ldots,p\}$, les éléments (classes) de Π .

Le regroupement A de classes de Π s'écrit :

$$A = \bigcup_{\substack{i_j \in J \\ J \subset I}} A_{i_j} \ ,$$

Notons :

$$\Pi_A = \{A, \{A_{i_j} \mid i_j \in C_I\ J\}\}$$

nous définissons une application m de H dans $\bar{\mathbb{R}}^+$ par :

$$A \in H \quad , m(A) = \delta^*(\Pi_A, \Pi)$$

. $\forall\ A_i \in \Pi\ ,\quad m(\{A_i\}) = \delta^*(\Pi_{A_i}, \Pi) = \delta^*(\Pi, \Pi) = 0$

. $\forall\ (A,B) \in H \times H\ ,$

$$A \subset B \text{ et } A \neq B \Longrightarrow \Pi_A < \Pi_B \Longrightarrow \delta^*(\Pi_A, \Pi) < \delta^*(\Pi_B, \Pi) \Longrightarrow m(A) < m(B)$$

ainsi, l'application m que nous venons de définir permet bien d'indicer la hiérarchie.

STRUCTURE DU PROGRAMME ARBRE

1 - Lecture de la partition Π de Ω

2 - Calcul du tableau δ défini sur $\Omega \times \Omega$

3 - Calcul du tableau des distances de HAUSDORF entre deux éléments quelconques de Π .

4 - Construction d'une nouvelle partition Π' en regroupant les deux classes situées à distance de HAUSDORF minimale.

5 - Si card(Π') $\neq 1$, alors faire : Π devient Π' ; aller en 3

6 - FIN.

III.3.3.- RESULTATS

Nous avons exécuté le programme ARBRE sur l'ensemble précédent de bactéries ENTEROBACTER. La partition Π (classes homogènes de Ω) nous a été fournie par les biologistes. La figure III.3.3.1 représente l'arbre obtenu par le programme ARBRE :

En ordonnée figure la valeur de $\delta^*(\Pi_k, \Pi)$; le calcul des différences Δ successives de δ^* nous donne le tableau suivant :

$$\Delta(i) = \delta^*(i) - \delta^*(i-1)$$

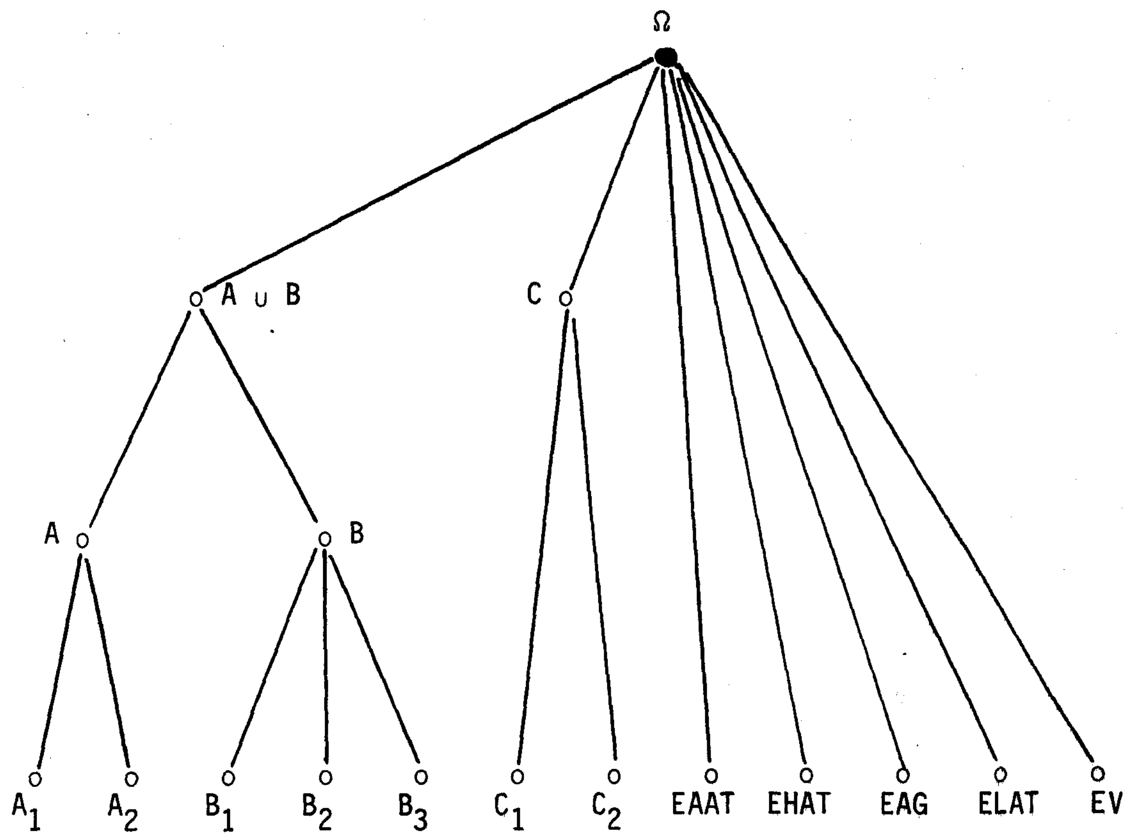
ETAPE	δ^*	Δ
0	0	
1	0,17	0,17
2	0,34	0,17
3	0,65	0,31
4	1,29	0,64
5	1,61	0,32
6	1,76	0,15
7	1,87	0,11
8	2	0,13
9	2,13	0,13
10	2,24	0,11
11	3,01	0,77

Ce tableau met en relief deux niveaux intéressants dans l'arbre : ceux correspondant aux plus grands sauts pour δ^* , c'est-à-dire les niveaux des étapes 3 et 10 .

L'arbre suivant, non indicé, fourni par les biologistes, a été établi en faisant une synthèse des résultats fournis par les méthodes classiques (classification hiérarchique et analyse factorielle).
Cet arbre est compatible avec celui obtenu par le programme ARBRE : nous avons bien la même structure fondamentale.

voir page suivante

.../...



Les résultats du programme ARBRE permettent de plus de constater que la classe $A = A_1 \cup A_2$ ne paraît pas homogène par rapport à B . En effet $B = B_1 \cup B_2 \cup B_3$ ayant été généré à l'itération 3 , l'itération 4 provoque le regroupement de A_2 à B et non, ce que l'on aurait pu penser, la formation de $A = A_1 \cup A_2$ (réalisée à l'itération 6) : en fait, ces résultats ne sont pas surprenants dans la mesure où :

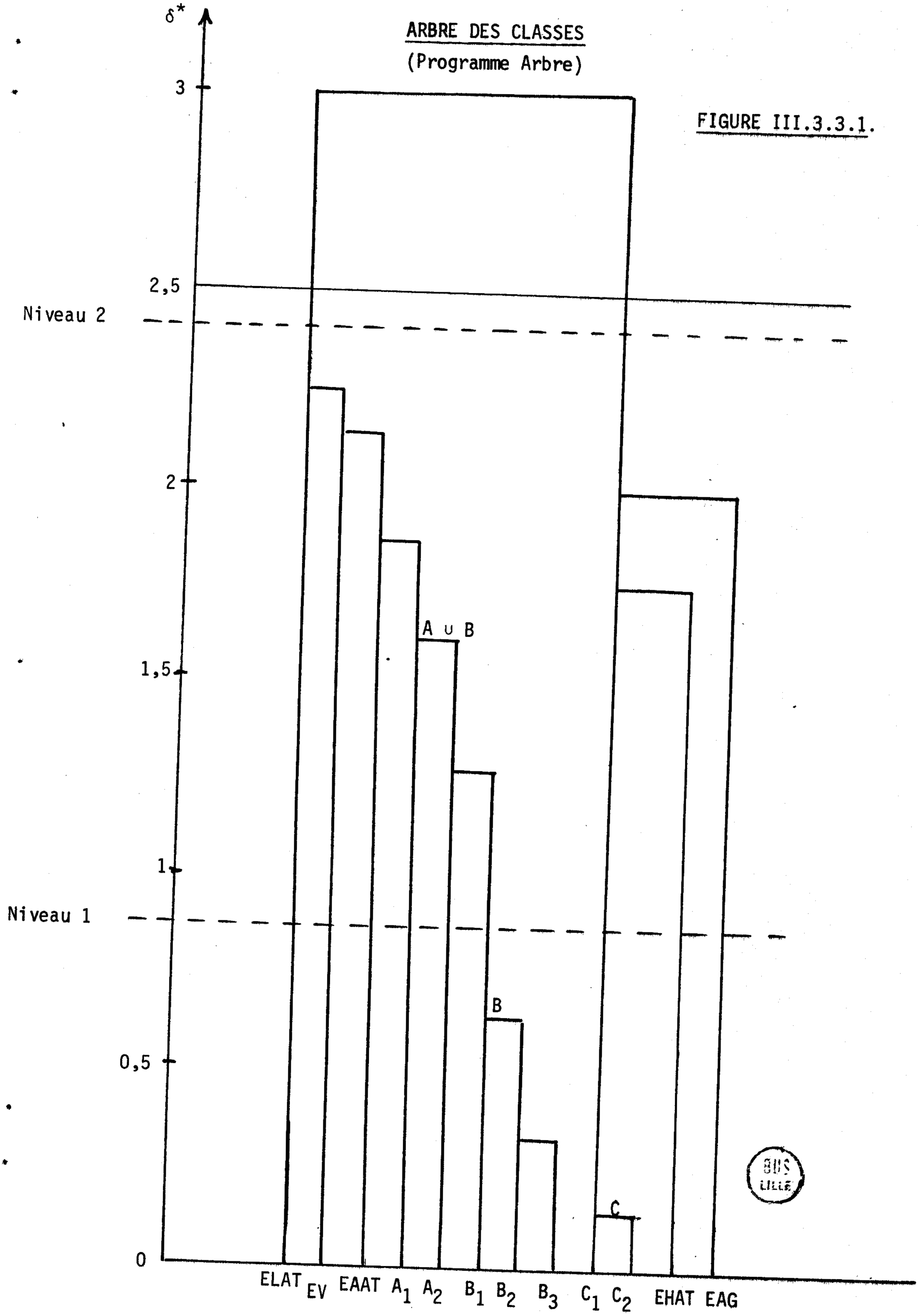
- l'analyse factorielle ne montre pas une séparation nette entre A_2 et B_3
- la classification hiérarchique faite par agrégation à distance maximum provoque le regroupement de A_2 à B_3 .

Enfin, la hiérarchie des classes obtenue peut bien être considérée comme sémantique, car tous les regroupements effectués ont respecté l'homogénéïté des classes formées.

ARBRE DES CLASSES

(Programme Arbre)

FIGURE III.3.3.1.



III-4.- QUESTIONNAIRE ASSOCIÉ À UNE HIÉRARCHIE INDICÉE. (PROGRAMME QUEST).

III.4.1.- ALGORITHME.

La hiérarchie sémantique fournir une certaine structure sur l'ensemble Ω .

Si nous ajoutons un élément ω' à Ω , nous désirons déterminer la classe de rangement pour que la structure ne soit pas modifiée.
Il s'agit donc, si nous ne voulons pas faire subir à ω' l'ensemble des observateurs de $\mathcal{O}$, de sélectionner un nombre restreint d'observateurs qui, appliqués à ω' , permettent de ranger ω' dans l'une des classes de Ω .

La hiérarchie sémantique nous est fournie sous la forme d'une suite de partitions de plus en plus fines :

$$\Pi_0 \lneq \Pi_1 \lneq \cdots \lneq \Pi_n = \{\Omega\}$$

Le principe du programme QUEST est le suivant :

a) partant de la partition courante $\Pi_C = \{\Omega\}$, nous cherchons à atteindre la partition objectif $\Pi_{OB} = \Pi_{n-1}$ en posant le moins de questions (observateurs) possibles : nous recherchons l'observateur t et l'élément A de Π_C tels que la nouvelle partition Π'_C que l'on obtiendrait par la division de A à l'aide des réponses de A à l'observateur t , soit à distance (δ^*) minimum de Π_{OB} . Si un tel observateur n'existe pas, alors l'itération est terminée, sinon avec la nouvelle partition Π_C obtenue, nous recommençons en a).
De plus, pour chaque observateur sélectionné, nous construisons l'ensemble des observateurs équivalents, strictement ou à un pourcentage d'erreurs fixé.

Il est à noter que le questionnaire obtenu n'est pas unique puisque nous associons à chaque observateur un ensemble d'observateurs équivalents.
Enfin, la minimisation effectuée est locale (établie lors du choix d'un observateur) et n'assure donc pas l'obtention du "meilleur" questionnaire dans le sens où le nombre d'observateurs sélectionnés serait minimum.

Structure du programme QUEST.

1 - Lecture des données

2 - $\Pi_C = \{\Omega\}$

3 - Lecture de Π_{OB}

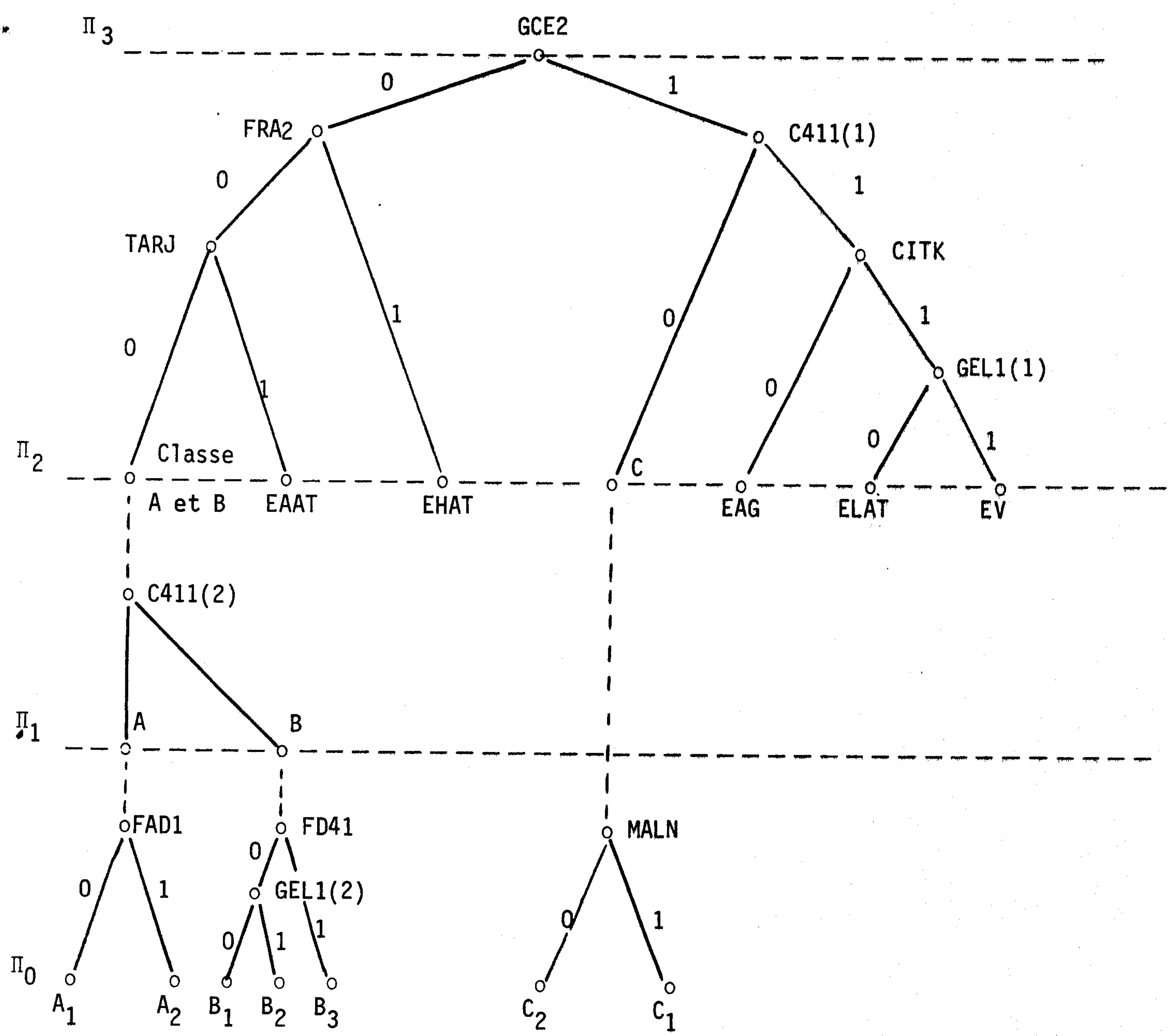
4 - Sélection des tests permettant de construire Π_C à distance minimum de Π_{OB}

5 - S'il reste une partition objectif à lire, alors aller en 3

6 - FIN.

III.4.2.- RESULTATS.

Le programme QUEST appliqué aux ENTEROBACTERS en se donnant trois partitions objectifs (Π_0 , Π_1 , Π_2) construit un questionnaire (figure III.4.2.1.) ayant 11 questions ; ce nombre de questions (tests) est donc minimum puisque nous avons 12 classes à différencier. De plus, pour chaque test, sont fournis au biologiste les tests équivalents, strictement ou à un pourcentage fixé.



QUESTIONNAIRE LIE A L'ARBRE DES CLASSES

FIGURE III.4.2.1.

B I B L I O G R A P H I E

[1] BARBUT et MONJARDET — *ORDRE ET CLASSIFICATION, ALGEBRE ET COMBINATOIRE.*
HACHETTE UNIVERSITE, PARIS, (1970)

[2] J.B. BENZECRI et collaborateurs — *L'ANALYSE DES DONNEES TOME I : LA TAXINOMIE*
DUNOD, PARIS, (1973)

[3] G. BIRKHOF — *LATTICE THEORY*
AMERICAN MATH. SOCIETY Colloquium publications Vol. XXV, (1967)

[4] B. FORTE — *MEASURES OF INFORMATION : THE GENERAL AXIOMATIC THEORY*
R.I.R.O. (3ème Année, n° R-2, (1969) pp. 63-90)

[5] H. GREENBERG — *INTEGER PROGRAMMING*
p. 177, Vol. 76, MONTEREY, Academic Press, 1971

[6] G. GRIMONPREZ, J. LOSFELD J. Cl. VAN DORPE — *PERTE ET GAIN D'INFORMATION SUR UNE SEQUENCE D'ELEMENTS COMPARABLES D'UN TREILLIS ET METRIQUE ASSOCIEE.*
C.R. Acad. Sc. PARIS, t. 280 (14-4-1975), Série A , p. 957.

[7] G. GRIMONPREZ J. Cl. VAN DORPE — *PERTE ET GAIN D'INFORMATION SUR UNE SEQUENCE D'ELEMENTS COMPARABLES D'UN TREILLIS (APPLICATIONS).*
C.R. Acad. Sc. PARIS, t. 280 (16-5-1075), Série A, p. 1381

[8] G. GRIMONPREZ J. Cl. VAN DORPE — *CONSEQUENCES DE L'AXIOME DE MONOTONICITE. NOTION DE RESUME.*
Journées Lyonnaises des Questionnaires (12 - 13 Juin 1975) Publications du Groupe de Recherche 22 : Structure de l'information, C.N.R.S.

[9] J. KAMPE DE FERIET B. FORTE — *INFORMATION ET PROBABILITE*
C.R. Acad. Sc. PARIS, t. 265 (1967) Série A - pp. 110-114

[10] J. LEMAIRE, J. POUGET — *INTRODUCTION A L'ANALYSE DES DONNEES : EXPOSE DE QUELQUES METHODES*

ECOLE D'ETE D'INFORMATIQUE, A.F.C.E.T., LANNION (Juillet 1976)

[11] J. LOSFELD — *INFORMATION GENERALISEE ET RELATION D'ORDRE.*

RENCONTRES EN THEORIE DE L'INFORMATION, MARSEILLE-LUMINY, Juin 1973

[12] J. LOSFELD — *INFORMATION FOURNIE PAR UN ENSEMBLE D'OBSERVATEURS ET APPLICATIONS AUX QUESTIONNAIRES ET A L'ANALYSE DES DONNEES.*

THESE DE DOCTORAT D'ETAT Juin 1974 - LILLE

[13] C.F. PICARD — *GRAPHES ET QUESTIONNAIRES*

GAUTHIER VILLARS, PARIS, 1972