MATHÉMATIQUES ET SCIENCES HUMAINES

I. C. LERMAN
R. GRAS
H. ROSTAM

Élaboration et évaluation d'un indice d'implication pour des données binaires. 2

Mathématiques et sciences humaines, tome 75 (1981), p. 5-47
<http://www.numdam.org/item?id=MSH_1981__75__5_0>

Math. Sci. hum. (19^{e}année, n°75, 1981, p.5-47)

ELABORATION ET EVALUATION D'UN INDICE D'IMPLICATION POUR DES DONNEES BINAIRES. 2 *

I.C. LERMAN, R. GRAS et H. ROSTAM **

Préambule.

Nous avions, dans le précédent article, abouti à l'expression de deux indices de "mesure" de l'implication a => b où a et b sont deux comportements donnés. Ψ étant la fonction de répartition de la loi normale centrée réduite, ces indices se mettent respectivement sous la forme $\Psi[q_p(a,\bar{b})]$ et $\Psi[q_p(\bar{a},b)]$, avec

$$q_p(a,\bar{b}) = \sqrt{n}\ \frac{[p(a\wedge\bar{b}) - p(a)p(\bar{b})]}{\sqrt{p(a)p(\bar{b})}}$$

$$q_p(\bar{a},b) = \sqrt{n}\ \frac{[p(\bar{a}\wedge b) - p(\bar{a})p(b)]}{\sqrt{p(\bar{a})p(b)}}$$

(cf. formules (11) et (12) § III.2.) où p(x) désigne la proportion de sujets qui ont le comportement x.

Avec l'adoption d'un seuil Ψ_0, on associe à un même indice un graphe discret d'implication (A,U) sur l'ensemble A des comportements. (a,b) est un arc de ce graphe ($(a,b) \in U$) si et seulement si p(a)< p(b) et la valeur de l'indice d'implication supérieure à Ψ_0 (cf. formules (28) et (29) § III.3.).

---

* La partie 1 a paru dans Math. Sci. hum. n°74 .

** Laboratoire de Statistique, IRISA, Université de Rennes-I et IREM de Rennes, "Campus Beaulieu", 35042 RENNES Cedex.

Pour arriver à une même base de comparaison entre les deux indices, on peut être conduit à substituer à $q(\bar{a},b)$ ou bien à la fois à $q_p(a,\bar{b})$ et à $q_p(\bar{a},b)$, les indices $q'_p(a,\bar{b})$ et $q'_p(\bar{a},b)$, centrés et réduits "globalement" (cf. formules (30) et (31) § III.3.).

Conformément à ce que nous avons annoncé dans l'introduction générale et repris à la fin du premier article, nous allons représenter ici le dessin des deux graphes discrets d'implication respectivement associés à chacun des deux indices $q_p(a,\bar{b})$ et $q_p(\bar{a},b)$, dans le contexte d'un cas réel fourni par un test mathématique élaboré autour du concept de la symétrie centrale. L'interprétation comparée des résultats permettra de mettre en évidence l'intérêt relatif de chacun des deux indices. Cette interprétation aura comme point de référence une taxinomie d'objectifs cognitifs.

Dans un deuxième temps, à partir d'un indice très général de comparaison entre deux relations pondérées sur un même ensemble, nous mettrons en relation chacun des deux graphes (pondéré ou discret, associé à l'un ou à l'autre des deux indices) avec un préordre total sur l'ensemble des comportements qui est construit par le chercheur comme image de la taxinomie. Ce dernier indice jouera enfin le rôle d'un critère d'adéquation.

IV - COMPORTEMENT DES DEUX INDICES SUR UN EXEMPLE REEL

IV.1 - Problème didactique

Le problème de la complexité se pose de façon cruciale dans l'enseignement des Mathématiques ; en effet, différents composants y interviennent et interfèrent. On peut par exemple citer

. des éléments de complexité intrinsèque : épistémologie du ou des concepts en question, topologie des algorithmes, nature des représentations, nature des tâches de résolution, etc...

. des éléments de complexité extrinsèque : formulation, didactique du ou des concepts, intervalle de temps séparant de l'apprentissage, etc...

L'expérience pédagogique permet peu ou prou la connaissance et donc la maîtrise de la deuxième catégorie des éléments de complexité. Par contre, l'ambition du didacticien est de poursuivre sans relâche la connaissance des

éléments de la première catégorie.

Les "taxinomies d'objectifs cognitifs" se donnent pour but de hiérarchiser, pour un concept donné, les capacités en jeu face à des problèmes. On prétend de la sorte rendre compte, le long d'une même hiérarchie, d'une croissance de la complexité d'exécution des tâches demandées.

Certaines taxinomies comme celle de Guilford (1967) (cf. dans [22]) sont à caractère multidimensionnel et intègrent ainsi des facteurs non ordonnables les uns par rapport aux autres, tels que les contenus et les opérations mentales. Cependant, le praticien souffre de la lourdeur de leur emploi. D'autres taxinomies comme celle de Bloom (1956) ou de la N.L.S.M.A. (National Longitudinal Study of Mathematical Abilities) (1969) présentent une seule dimension, compromis entre la finalité et la pratique. R. Gras (cf. [7] (1979)) élabore une nouvelle taxinomie, inspirée des précédentes, mais s'appuyant plus explicitement sur, d'une part, les acquis de la psychologie de l'apprentissage (Piaget, Brüner, Brousseau, Vergnaud,...) et sur, d'autre part, la spécificité des concepts mathématiques.

Cette taxinomie se matérialise par un préordre total sur l'ensemble des objectifs et on admet implicitement dans sa construction que les capacités de la classe j dominent celle de la classe i ($i < j$) au sens suivant : les capacités développées au niveau i ne sont pas suffisantes au niveau j, alors que la réciproque est vraie. Plus précisément, pour résoudre une question illustrant un objectif de la classe j, il est nécessaire d'arriver à le faire pour toute question relevant de la classe i. La complexité va donc croissant de la première à la dernière classe de la taxinomie. On y distingue cinq classes notées A,B,C,D et E dont nous reprenons dans le tableau ci-dessous l'expression schématique. Pour une analyse plus détaillée nous renvoyons à la thèse de R. Gras (cf. [7] (1979)).

TABLEAU D'OBJECTIFS COGNITIFS

CLASSES	RUBRIQUES	OBJECTIFS	ACTIVITES ATTENDUES
A Connaissance des outils de préhension de l'objet et du fait mathématiques	A_1	Connaissance de la terminologie et du fait spécifique	associer assembler
	A_2	Capacité d'action intériorisée sur l'évocation d'une forme physique du concept	simuler observer
	A_3	Capacité de lire des cartes, des tableaux, des graphiques, des notices	déchiffrer décrire
	A_4	Effectuation d'algorithmes simples	organiser, calculer
B Analyse de faits et transposition	B_1	Substitution d'une démarche représentative à une manipulation Anticipation graphique	abstraire prolonger induire
	B_2	Reconnaissance et usage d'une relation implicite simple où intervient l'objet mathématique connu	analyser comparer
	B_3	Traduction d'un problème d'un mode dans un autre avec interprétation	schématiser,traduire transposer
C Compréhension des relations et des structures	C_1	Compréhension du concept, de ses relations avec les autres objets mathématiques	reconnaître construire
	C_2	Compréhension d'un raisonnement mathématique : justification d'un argument	justifier
	C_3	Choix et ordonnancement d'arguments	déduire
	C_4	Application dans des situations familières	analyser,abstraire appliquer,interpoler
D Synthèse et Créativité	D_1	Effectuation et découverte d'algorithmes composites et de nouvelles relations	organiser,calculer optimiser
	D_2	Construction de démonstrations et d'exemples personnels	illustrer,démontrer valider,créer, inventer
	D_3	Découverte de généralisations	généraliser,induire prévoir,extrapoler
	D_4	Reconnaissance du modèle et applications dans des situations non routinières	modéliser,identifier,différencier, classifier,résumer
E Critique et évaluation	E_1	Distinction du nécessaire et du suffisant	formuler des hypothèses, déduire
	E_2	Critique de données et de méthodes ou de modèles résolvants	contrôler,optimiser prévoir, critiquer questionner,vérifier
	E_3	Critique d'argumentation et construction de contre-exemple	critiquer,tolérer contredire

La taxinomie va constituer notre point d'ancrage pour l'interprétation comparée des deux graphes discrets d'implication (cf. formules (28) et (29), § III.3. partie 1) respectivement associés aux indices $q_p(a,\bar{b})$ et $q_p(\bar{a},b)$ (cf. (31), § III.3). La base expérimentale de cette analyse est le comportement réel d'un échantillon d'élèves à un test défini par un questionnaire établi autour d'un même concept. Ce type d'interprétation suppose par conséquent l'affectation d'un même item du questionnaire à une classe donnée de la taxinomie ; mais une telle affectation peut être souvent difficile et ne manque pas de comporter des éléments subjectifs d'appréciation. C'est au paragraphe V suivant que nous aborderons en un certain sens, le problème de la validation de la taxinomie. Notre intention ici est, sur la base d'un questionnaire construit autour du concept de la symétrie centrale , d'une part, de décrire une démarche heuristique de représentation du graphe discret d'implication ; d'autre part, d'examiner si des différences sensibles de structure apparaissent entre les deux graphes respectivement associés aux deux indices, en cherchant à dégager des remarques didactiques sur l'appropriation des élèves et sur la construction du questionnaire.

IV.2. Les données

A partir de la taxinomie et sur le concept de symétrie centrale enseigné en classe de 4ème des Collèges d'Enseignement Secondaire, on construit un questionnaire à 40 items , chacun de ceux-ci étant censé illustrer une capacité exprimée dans une classe de la taxinomie. Mais, comme nous venons de le signaler, cette opérationalisation demeure subjective. Ce n'est donc qu'en toute hypothèse que l'on dira que le questionnaire est une image de la taxinomie.

Il serait trop long dans le cadre de cet article de fournir le détail des questions et nous renvoyons à cet égard à la thèse précitée (cf.[7]). Toutefois, pour chercher à se faire tant soit peu comprendre, nous donnons ci-dessous le sens général de chacun des items et l'affectation choisie à la classe de la taxinomie. Cette répartition taxinomique est la suivante :

9 items correspondant à la classe d'objectifs A :

A_{10}, A_{11}, A_{12}, A_{13}, A_{14}, A_{20}, A_{30}, A_{40}, A_{41}.

7 items correspondant à la classe B :

B_{10}, B_{11}, B_{20}, B_{26}, B_{27}, B_{30} et B_{31}.

9 items ramenés à 7 correspondant à la classe C :

C_{10}, C_{11} et C_{12}(ramenés à C_{10}), C_{21}, C_{22}, C_{30}, C_{31}, C_{40} et C_{41}.

12 items correspondant à la classe D :

D_{10}, D_{15}, D_{16}, D_{20}, D_{30}, D_{31}, D_{32}, D_{40}, D_{41}, D_{42}, D_{43} et D_{44}.

enfin, 5 items correspondant à la classe E :

E_{10}, E_{20}, E_{21}, E_{30} et E_{31}.

Signalons que dans cette indexation, le premier indice signale la sous-classe taxinomique ; ainsi par exemple, A_{40} et A_{41} illustrent la sous-classe taxinomique A_4.

Le sens général des différents items est le suivant :

A_{10}, A_{11}, A_{12}, A_{13}, A_{14} : "reconnaissance d'un centre de symétrie d'une figure formée d'un couple de points rentrant dans la composition d'une configuration géométrique".

A_{20} : "reconnaissance, parmi trois montages, de celui qui transforme une figure F en une figure F' translatée, homothétique ou symétrique de F , les figures F et F' étant données".

A_{30}: " reconnaissance de la composition de symétries centrales en évoquant un montage réalisant ce type de transformation géométrique"

A_{40},A_{41}: "application d'une expression algorithmique de la moyenne"

B_{10}: "construction du symétrique d'un triangle par rapport à un point"

B_{11}: "reconnaissance de la parité de la composition de symétries centrales mais sans évocation explicite d'un montage de réalisation".

B_{20}: "parmi un ensemble de figures données, reconnaissance de celles qui admettent un centre de symétrie".

B_{26},B_{27}: "reconnaissance d'une opération de moyenne arithmétique sur des nombres donnés à partir de deux réalisations de cette dernière".

B_{30}: "reconnaissance de la symétrie centrale de centre le milieu d'un segment à partir d'une suite de réalisations de cette transformation".

B_{31}: "par rapport à un même sommet d'une configuration donnée, reconnaître quels sont les points de cette configuration qui admettent leurs symétriques respectifs dans la configuration".

C_{10}: "par rapport à différents couples de figures donnés de la forme (F,F') reconnaître l'existence ou non d'une symétrie centrale transformant F en F' et la caractériser".

C_{21},C_{22}: "parmi des arguments donnés, choisir et ordonner ceux des arguments nécessaires à une démonstration".

C_{30},C_{31}: "même type d'exercice que C_{21} et C_{22} dans une situation plus complexe de par notamment la profusion des arguments dont certains sont superflus et où il y a plusieurs solutions possibles".

C_{40},C_{41}: "complétion de figures commencées dont le centre de symétrie est précisé (C_{40}) ou à déterminer (C_{41}) avec une contrainte de traçage".

D_{10}: "dans une liste de nombres donnés, choisir différents couples de nombres (x,y) tels que la moyenne des deux arguments du couple appartienne aussi à cette liste".

D_{15},D_{16}: "règle d'association à découvrir et à mettre en oeuvre à partir de deux suites de couples de nombres qui se correspondent terme à terme".

D_{20}: "démonstration de l'existence d'un centre de symétrie d'une figure dont la construction est évoquée physiquement".

D_{30},D_{31}: "reconnaissance de l'opération de moyenne sur des arguments représentés par des nombres ou des lettres".

D_{32}: "reconnaissance générale de la parité de la symétrie centrale".

D_{40},D_{41}: "construction d'images symétriques avec une contrainte géométrique dans la construction".

D_{42},D_{43},D_{44}: "appréhension de l'opération quotient sous la forme de l'application de l'ensemble D^+ des décimaux positifs dans lui-même : $x \mapsto y$ tel que $xy = k$ donné ; composition de telles applications".

E_{10}: "reconnaissance d'une implication logique liée à l'existence d'un centre de symétrie pour un polygone".

E_{20},E_{21}: "remplissage d'un "grand" parallélogramme ABCD au moyen d'un "petit" MNPQ à l'aide de symétries centrales ; reconnaissance de l'impossibilité de ce remplissage en raison de l'orientation de MNPQ (E20), caractérisation d'un autre MNPQ pour lequel l'opération est possible et doit être décrite avec précision (E21)".

E_{30},E_{31}: "reconnaissance de figures invariantes par la composée de deux symétries centrales de centres distincts (E_{30}) ou de trois symétries centrales de centres distincts (E_{31})".

On tente, sur l'ensemble des 401 élèves qui ont passé le test, de contrôler l'effet de certaines variables exogènes en étudiant la distribution de certains paramètres descriptifs tels que l'âge, le sexe, la catégorie socio-professionnelle des parents, le type de la classe (T_1,T_2 ou T_3 ordonnés pratiquement selon le niveau intellectuel des élèves, le type 1 étant considéré le meilleur). Mais, on introduit également trois paramètres pédagogiques : la distance temporelle à l'apprentissage (de 2 à 5 mois), la méthode d'apprentissage utilisée (approche plus ou moins formelle et approche manipulatoire et active), enfin, l'ordre dans lequel sont présentées les questions. De plus, afin d'analyser l'effet du contenu, des items de même niveau taxinomique relèveront d'une part d'une représentation géométrique et d'autre part, d'une représentation algébrique du concept de symétrie centrale (e.g. moyenne arithmétique ou géométrique).

Nous renvoyons à la thèse de R. Gras (cf. [6]) pour l'analyse statistique des effets respectifs de ces différentes variables sur la réussite au test. Signalons toutefois que sur notre population de 401 élèves on constate que

. le sexe apparaît discriminant (moyenne du pourcentage de réussite pour les différents items du questionnaire : 46 % chez les garçons et 40 % chez les filles)

. le type de la classe a une importance certaine (moyenne du pourcentage de réussite : 46 % pour le type T_1, 40 % pour T_2 et 30 % pour T_3 qui est ce qu'on appelle le type T_2 aménagé)

. l'ordre de déroulement des items n'intervient que très localement où alors s'effectue une auto-compréhension par filiation ou voisinage de questions.

. la méthode d'apprentissage a des effets sensibles (moyenne du pourcentage de réussite : 47 % sur l'ensemble des sujets ayant effectué des manipulations et 42 % sur l'ensemble des sujets n'ayant pas manipulé

. l'ordre dans la fréquence de réussite est conforme à l'échelle sur la suite des classes d'objectifs cognitifs ; ainsi, les fréquences moyennes de réussite pour, respectivement les différentes classes taxinomiques, sont : 75 % pour A, 55 % pour B, 35 % pour C, 22 % pour D et 8 % pour E. Toutefois, on note pour un sous-ensemble non négligeable d'élèves des inversions entre l'ordre taxinomique tel que nous l'avons perçu et l'ordre dans la fréquence de succès ; par exemple, la réussite en A domine au sens large celle en B dans 83 % des cas, celle en C dans 96 % des cas, celle en D dans 99 % des cas et celle en E dans 100 % des cas. En particulier, l'ordre taxinomique est rigoureusement vérifié pour 42 % des sujets.

FREQUENCE DE REUSSITE PAR ITEM SUR LA POPULATION DES 401 ELEVES

N° item	Réussite	N° item	Réussite
A 10	96 %	C 31	6 %
11	68 %	40	75 %
12	93 %	41	21 %
13	92 %	D 10	40 %
14	96 %	15	27 %
20*	43 %	16	24 %
30*	41 %	20	28 %
40	77 %	30	14 %
41	72 %	31	5 %
B 10	84 %	32*	10 %
11*	22 %	40	52 %
20	52 %	41	7 %
26	55 %	42	18 %
27	58 %	43	33 %
30	54 %	44	1 %
31	61 %	E 10	7 %
$C_{10}=(C_{10},C_{11},C_{12})$	51 %	20	24 %
21	62 %	21	3 %
22	16 %	30	3 %
30	16 %	31	2 %

Rappelons que les items marqués d'un * prennent appui explicitement ou implicitement sur un matériel de manipulation.

IV.3. Construction d'un graphe d'implication

La donnée des deux indices d'implication poissonniens $q_p(a,\bar{b})$ et $q_p(\bar{a},b)$, définis dans le paragraphe précédent, conduit à celle des 2 matrices carrées: $M(a,\bar{b})$ et $M(\bar{a},b)$ des intensités d'implication $\psi[q_p(a,\bar{b})]$ et $\psi[q_p(\bar{a},b)]$, entre les comportements de réussite de la population de 401 élèves. Si l'on fixe un seuil de 5 %, les deux matrices engendrent deux nouveaux tableaux carrés d'élément générique :

$$D(a,b) \text{ (resp. } D(\bar{a},b)) = \begin{cases} 1 \text{ si } \psi[q_p(a,\bar{b})] (\text{resp.} \psi\,[q_p(\bar{a},b)]) \geqslant 0{,}95 \\ 0 \text{ sinon} \end{cases}$$

Plus opérationnels que les précédents, ces tableaux servent à construire le graphe d'implication. On n'exclut pas la référence aux premiers qui permettent de trancher, dans tous les cas, sur l'acceptabilité ou non d'un arc, eu égard à la satisfaction de la relation de transitivité.

Plus précisément, l'algorithme de construction du graphe se présente ainsi :

1) On part d'un item i (plus exactement d'un comportement de réussite à l'item i) qui devient un noeud source du graphe, n'admettant pas d'antécédent. Ceci signifie, selon la théorie des graphes que le noeud i admet un degré intérieur nul et un degré extérieur positif. Autrement dit encore, le nombre d'arcs implicatifs d'extrémité i (resp. d'origine i) est nul (resp. différent de 0) :

$(\forall j)\ D(j,i) = 0$ et $(\exists j / j \neq i)\ D(i,j) = 1$

Les noeuds j tels que D(i,j)=1 sont appelés "successeurs naturels" de i.

2) Les noeuds successeurs de i étant supposés ordonnés dans l'ordre croissant des pourcentages de réussite, on choisit celui, j, dont le pourcentage est minimum. On relie i à j par un arc : i —> j.

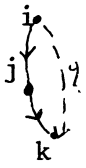
3) Puis, k ayant maintenant le pourcentage de réussite le plus faible parmi les successeurs naturels de i ou de j :

a) - ou bien D(i,k)= D(j,k)=1 (k est successeur naturel de i et j) et on enchaîne k à j : i → j → k.

b) - ou bien D(i,k)=1 et D(j,k)=0. La satisfaction simultanée de deux critères permet d'accepter k comme "successeur de liaison" de j :

b.1) l'intensité d'implication de j sur k est de l'ordre de 0,9 au moins

b.2) le nombre de successeurs naturels communs à i et à j est relativement important



c) - ou bien D(i,k)= 0 et D(j,k)=1.

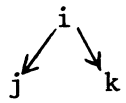
On distingue dans ce cas, deux situations

c.1) l'intensité d'implication de i vers k est supérieure à 0,5 ; on enchaîne alors

c.2) sinon, on retient seulement celle des deux implications i => j ou j => k, pour laquelle la condition de type b.2 est la mieux respectée.

Pour l'instant, seule une combinaison optimale empiriste, mais non exempte de rationalité didactique, de ces critères, permet de prendre la décision. On rencontre donc quelquefois des arcs comme celui-ci par exemple :

i → j ---0,9---> k (l'intensité de j => k est 0,9)

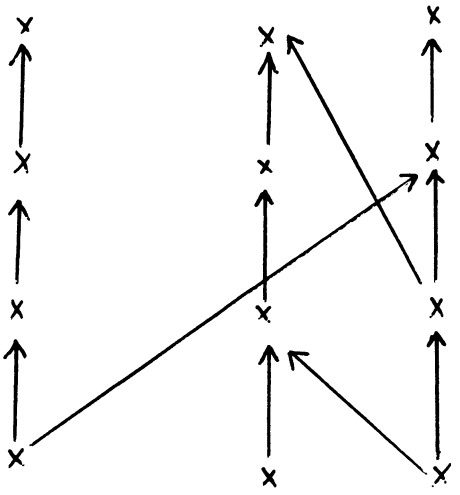
Ou, dans un cas de refus, celui-ci : 

4) 1 admettant le pourcentage de réussite minimum parmi tous les successeurs naturels ou de liaison de i, j ou k, on enchaînera ou non 1 à k après examen des intensités d'implication de i,j et k sur 1 et des familles de successeurs naturels communs de ces mêmes noeuds.

On arrête le chemin de source i lorsque tous ses successeurs naturels sont placés, de même que les propres successeurs de ceux-ci. Ainsi, le long du chemin de source i, tous les arcs . —>. respectent le seuil de 10 % (à 15 % suivant les cas) et toutes les fermetures transitives respectent le seuil de 50 %.

5) On part d'une autre source i' et on construit comme précédemment le chemin d'origine i'. Il arrive que figurent, parmi les successeurs naturels ou de liaison de i', des mêmes successeurs de i. Les tests aux mêmes seuils que précédemment restent valides.

6) Les autres chemins étant construits, on tient à garder une structure de représentation où les noeuds sont les seuls points doubles admissibles, ceci afin de faciliter l'interprétation du graphe en des termes d'indépendance. Ainsi, on réorganise le graphe pour supprimer des imbrications comme celle-ci :



Eventuellement, pour permettre cette structure éclatée, des successeurs de liaison sont ajoutés ou supprimés. La construction à la main est assez longue : nous cherchons à l'automatiser, mais il va de soi que la notion de tolérance, empreinte d'une certaine subjectivité, freinera la modélisation de l'algorithme.

Dans la thèse de R. Gras, on trouve le graphe G (cf. page suivante) construit à l'aide de q_b (indice binomial), graphe qu'il analyse ensuite en des termes didactiques.

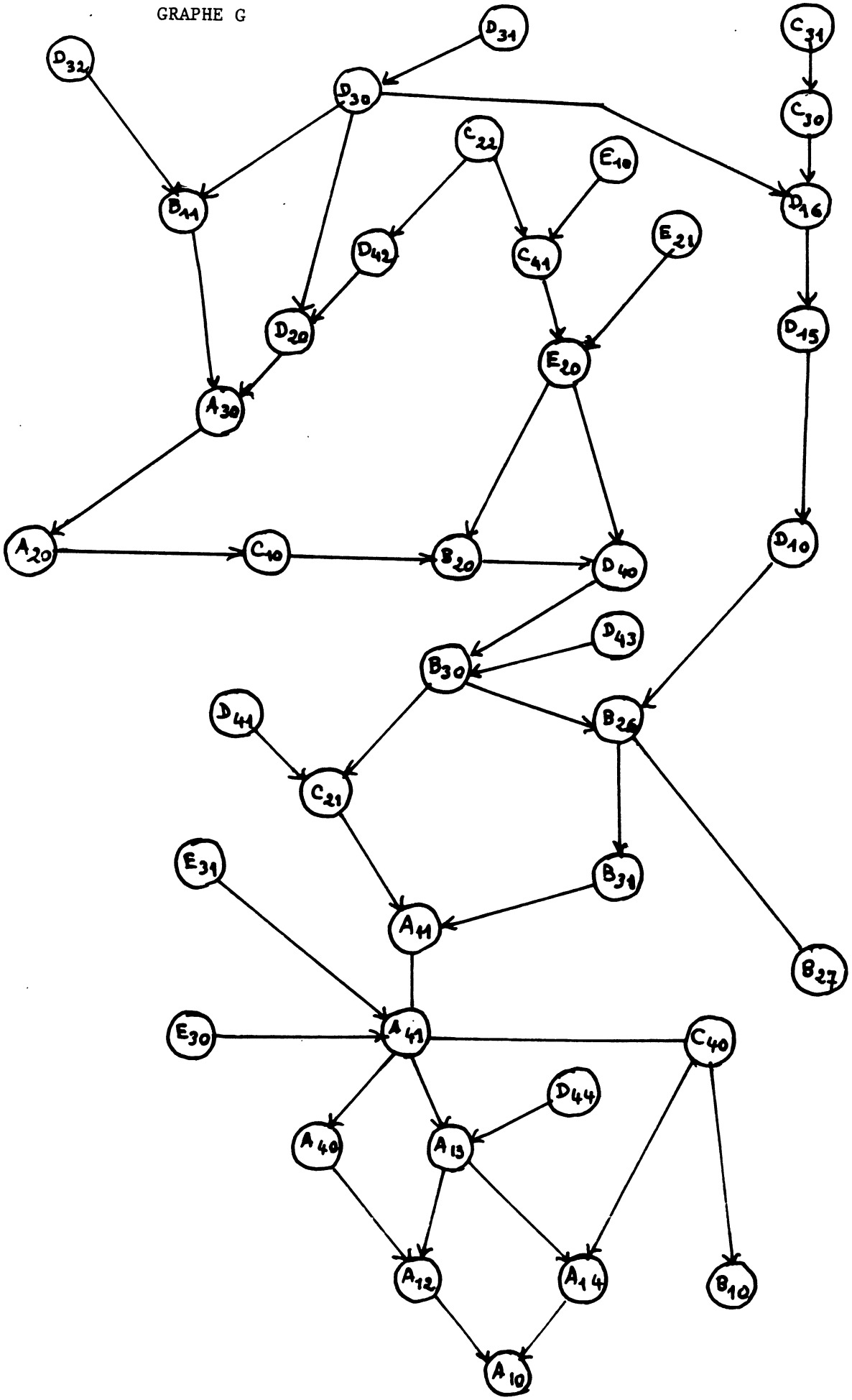
Remarquons que le graphe G, comme les suivants, peut se lire et donc s'interpréter dans les deux sens. Si le comportement de réussite à a implique le comportement de réussite à b, alors, avec la même intensité, le comportement d'échec à b implique le comportement d'échec à a. En effet :

$$q_p(a,\bar{b}) = \frac{s(a,\bar{b}) - \frac{n(a)n(\bar{b})}{n}}{\sqrt{\frac{n(a)n(\bar{b})}{n}}} = q_p(\bar{b},a)$$

et de même : $q_p(\bar{a},b) = q_p(b,\bar{a})$

Cette remarque permet de distinguer l'analyse de données par le graphe d'implication, d'autres méthodes comme les analyses hiérarchiques ou factorielles. L'analyse y est parfaitement symétrique en réussite-échec, ce qui n'enlève en rien l'avantage à tirer d'une situation asymétrique.

GRAPHE G


D31
C31
D32
D30
C30
C22
E10
B11
D16
D42
C41
E21
D20
D15
E20
A30
A20
C10
B20
D40
D10
D43
B30
D41
B26
C21
E31
B31
A11
B27
E30
A41
C40
D44
A40
A13
A12
A14
B10
A10

IV.4. Analyse de graphes d'implication

Nous proposons ici de comparer les 2 graphes obtenus par étude, à l'aide de $q_p(a,\bar{b})$ et $q_p(\bar{a},b)$, de la sous-population de 229 élèves (sur 401), ayant manipulé avant l'évaluation de leur appropriation du concept de symétrie centrale (cf. tableau des réussites décroissantes).

Effectifs décroissants des réussites aux items du test dans la sous-population de 229 élèves

Items	Effectifs réussites	Items	Effectifs réussites
$A_{10} = A_{14}$	219	D_{10}	78
B_{10}	216	D_{43}	77
A_{12}	212	B_{11}	75
A_{13}	209	C_{41}	64
C_{40}	193	D_{15}	59
A_{40}	182	D_{16}	55
A_{41}	167	C_{22}	51
B_{20}	164	D_{42}	44
A_{11}	162	D_{32}	35
C_{21}	159	D_{30}	34
C_{10}	154	C_{30}	26
$B_{31} = D_{40}$	146	D_{41}	21
B_{30}	145	E_{10}	17
B_{27}	141	D_{31}	14
A_{20}	136	C_{31}	12
B_{26}	134	E_{21}	10
A_{30}	120	$E_{30} = E_{31}$	7
D_{20}	81	D_{44}	2
E_{20}	80		

L'indice $q_p(a,\bar{b})$, non centré et réduit, conduit au graphe G_1 ; l'indice $q_p(\bar{a},b)$, après centrage et réduction, conduit au graphe G_2. Ces graphes se distinguent donc du précédent, cité en IV.3, par le référentiel des individus et par la définition de l'indice.

a) Etude de G_1

Le tableau T_1 ci-dessous des indices $D(a,\bar{b})$ montre que la taxinomie est mieux "respectée" ici qu'elle ne l'était dans l'étude précédente. En effet, à peine 9 % seulement des 180 implications admissibles au seuil de 0,05, contre 15 % dans l'autre étude, s'opposent à la conjecture taxinomique. Ceci nous semble l'effet conjugué :

. de l'usage d'un indice mieux adapté semble-t-il au problème lui-même, pour les raisons indiquées en II.5

. de la plus grande homogénéité de la population testée puisqu'elle a subi un même type d'apprentissage de la notion par manipulation.

La construction de G_1 à l'aide de T_1 se fait comme indiqué plus haut. On n'y placera pas l'implication : $D_{44} \Rightarrow C_{10}$, et on associera à certains noeuds des successeurs de liaison :

D_{42} ---> A_{30}(intensité: 0,87) D_{42} ---> D_{20}(intensité 0,89)

D_{31} ---> D_{10}(intensité: 0,92) D_{32} ---> B_{26}(intensité:0,90)
---> B_{30}

ceci afin d'optimiser l'information donnée par le fonctionnement transitif du graphe et afin de ne pas faire se couper 2 arcs ailleurs qu'en des noeuds.

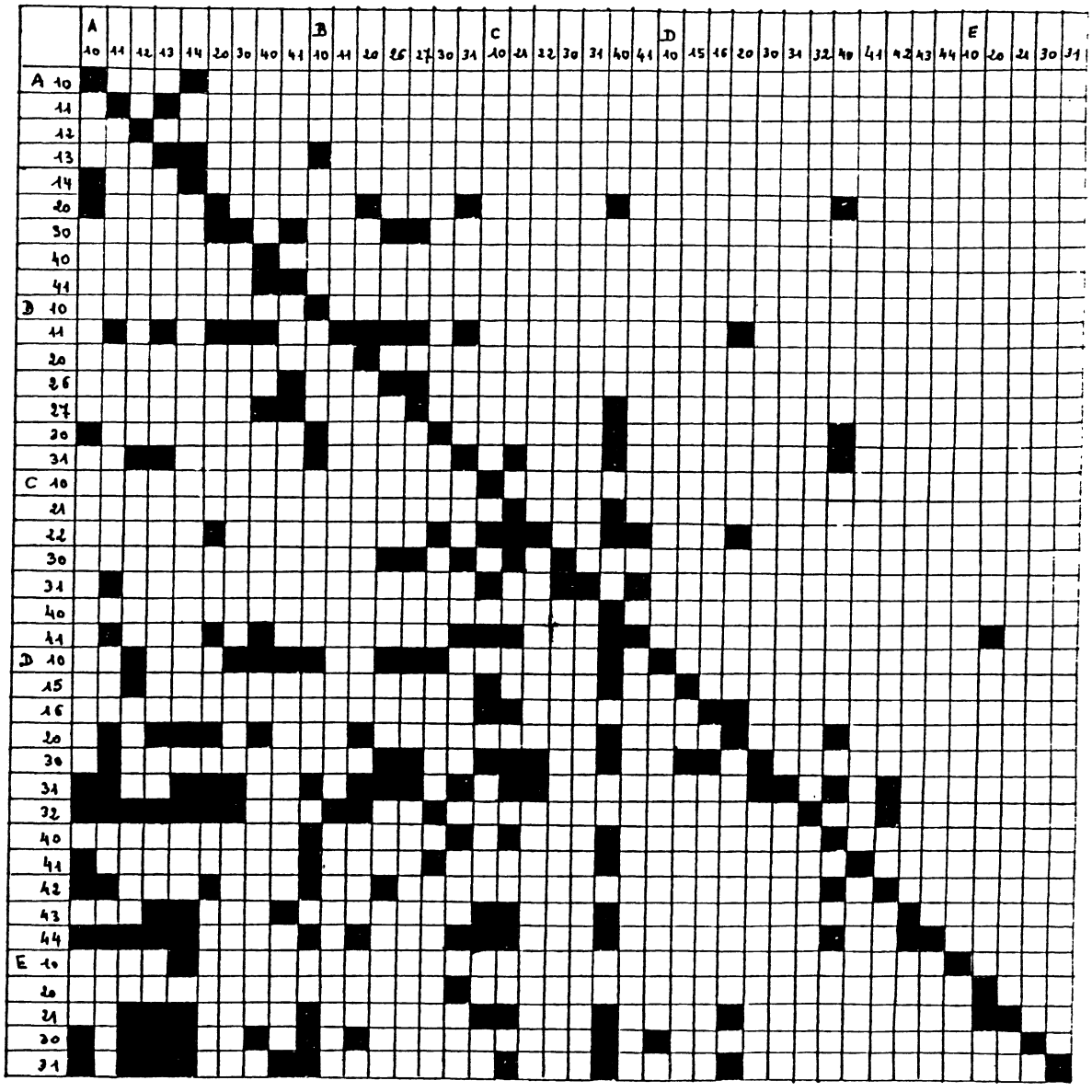
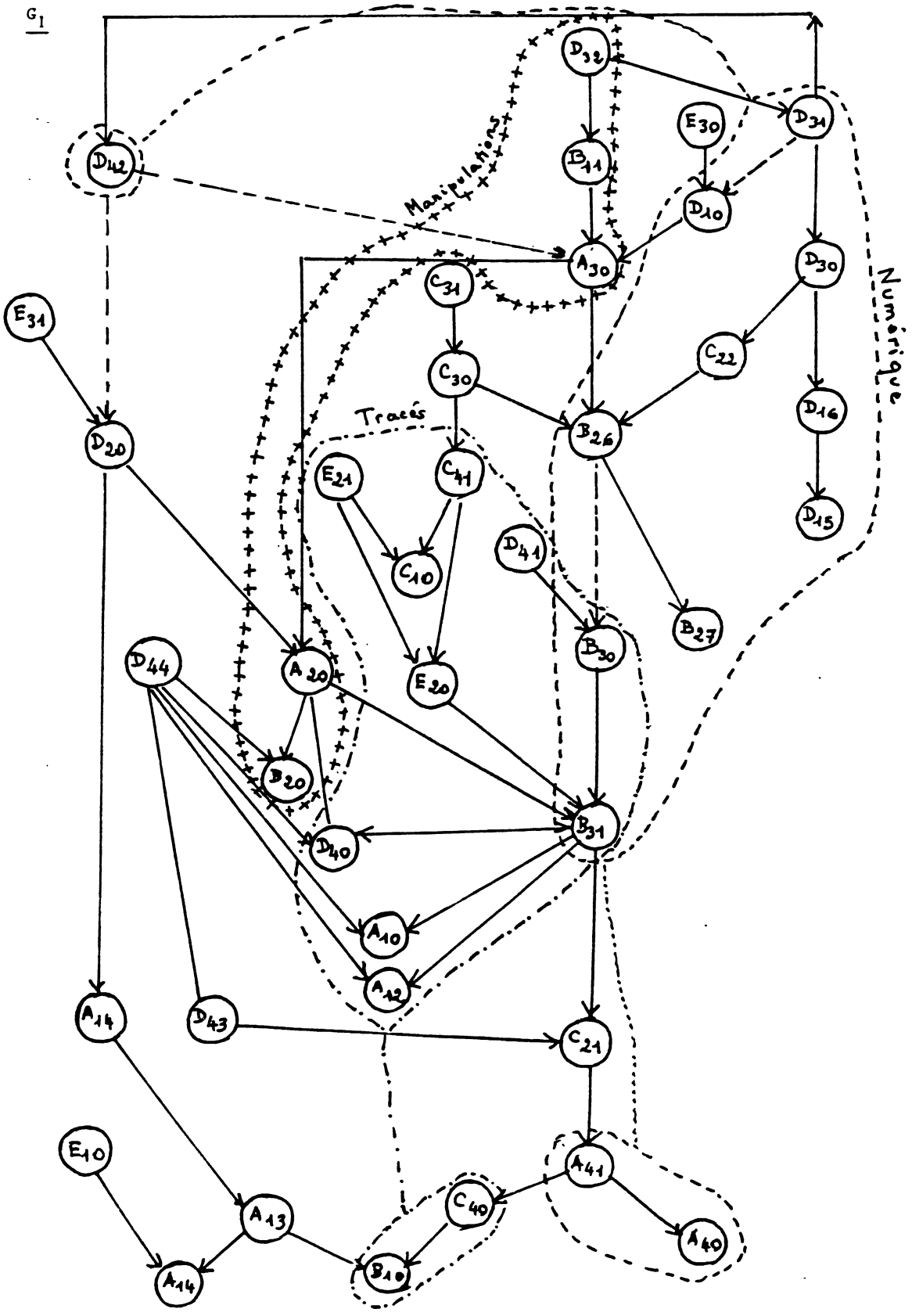


Tableau T_1 des $D(i,\bar{j})$ (1 carreau noir pour $D(i,\bar{j})=1$)

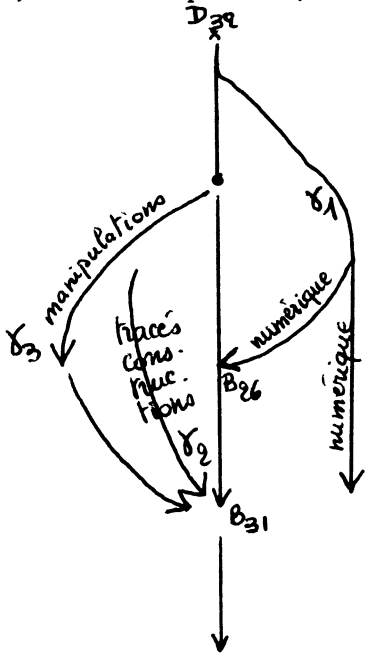
GRAPHE IMPLICATION DANS POPULATION 'Manipulation'



Une analyse sommaire de G_1 montre que :

1) Les sources (resp. les puits) correspondent, à l'exception de C_{31} (resp. D_{15}), à des niveaux supérieurs (resp. inférieurs) de la taxinomie, soit D ou E (resp. A ou B). La taxinomie est généralement décrite dans l'ordre décroissant le long de chaque chemin de G_1.

2) Schématiquement, trois chemins γ_1, γ_2 et γ_3 se développent de façon indépendante :

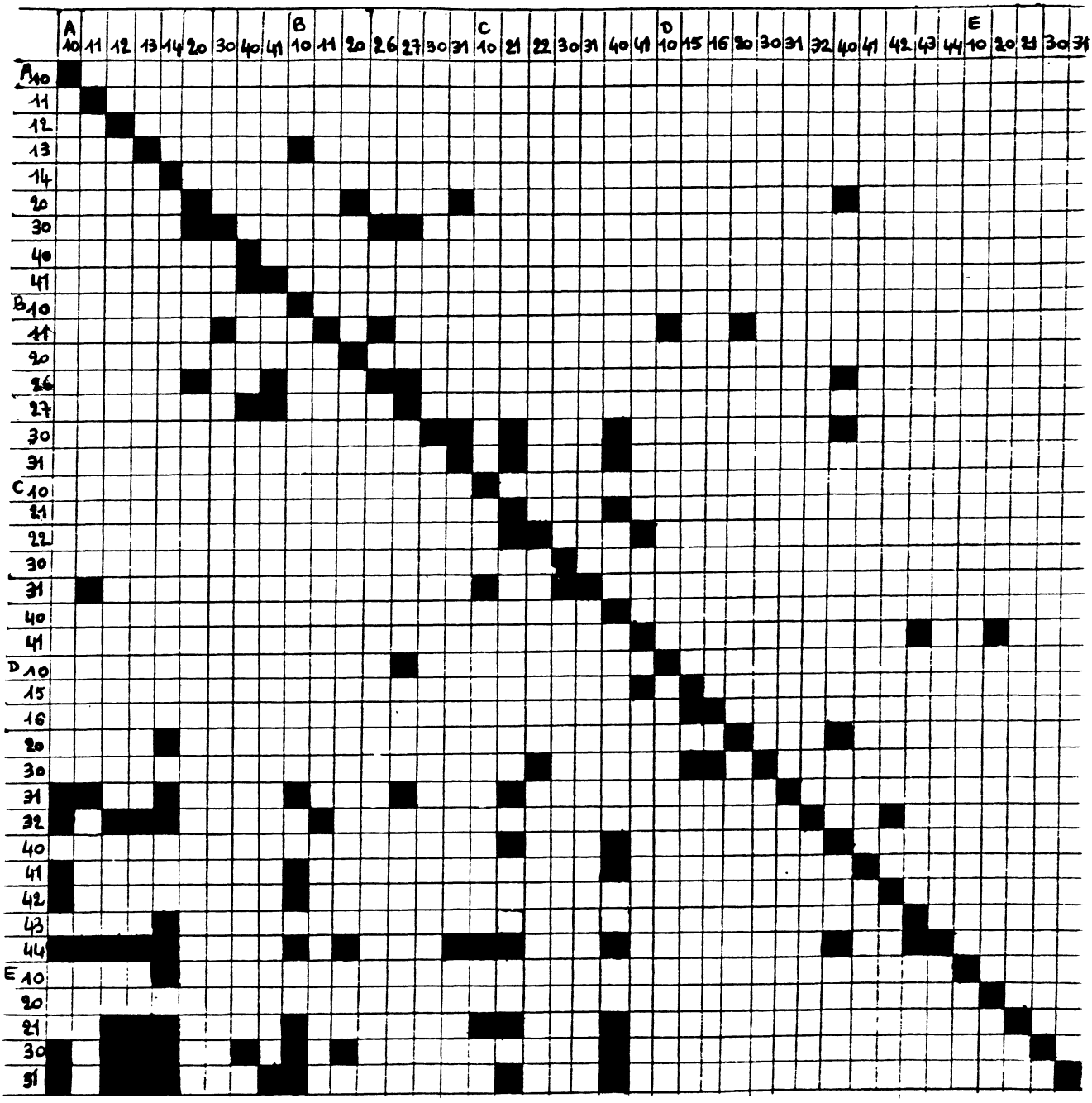


γ_1 correspond à des tâches de nature numérique
γ_2 à des tâches de construction
γ_3 à des tâches prenant appui sur la manipulation.
La jonction s'effectue en l'item B_{31} dont la résolution fait appel à des tâches de nature algorithmique (comme le long de γ_1) et géométrique (comme le long de γ_2 et γ_3).

Ainsi, G_1 restitue convenablement la taxinomie, ce qui réciproquement renforce sa crédibilité, tout en mettant en évidence l'éclatement des opérations mentales en 2 catégories suivant qu'elles se nourrissent de contenus algébriques ou géométriques. C'est une dimension dont nous avions parlé plus haut en citant la taxinomie de Guilford. Le simple examen des résultats bruts au test n'aurait pas permis de la dégager, même si les analyses factorielle et classificatoire la laissaient présager.

b) Etude de G_2

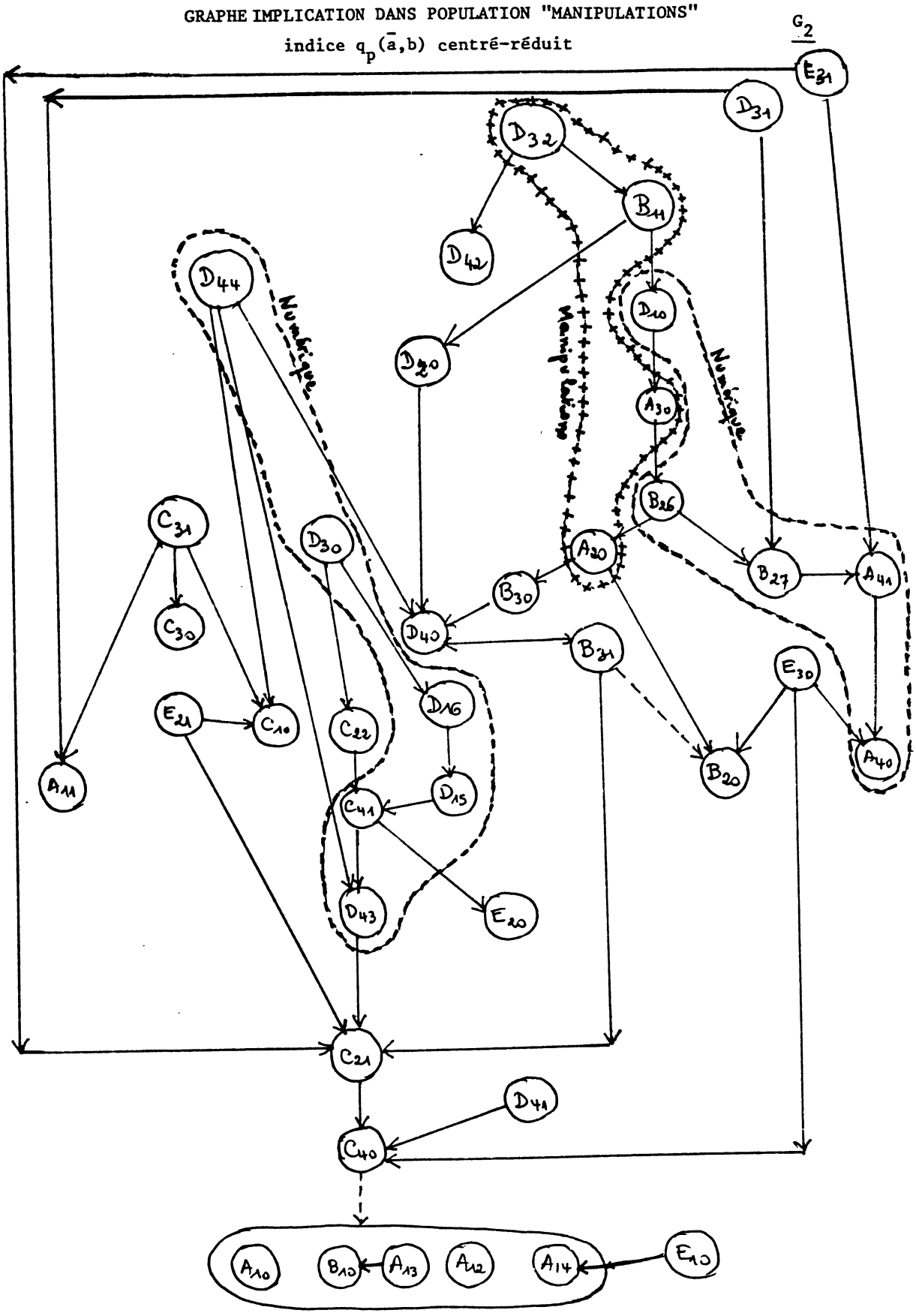
Le tableau T_2 montre cette fois une meilleure adéquation de l'implication à la taxinomie, puisque sur 96 implications par $q_p(\bar{a},b)$ centré-réduit, on relève seulement 5 d'entre elles qui s'opposent au caractère préordinal de la taxinomie. Ce tableau ne fournit pas des informations sur les implications, redondantes avec T_1, contrairement à ce qui se passait avant le centrage et la réduction où $q_p(\bar{a},b)$ affaiblissait seulement $q_p(a,\bar{b})$. A priori, son étude doit donc nous fournir des éléments nouveaux.


A
B
C
D
E
10 11 12 13 14 20 30 40 41
10 11 20 26 27 30 31
10 21 22 30 31 40 41
10 15 16 20 30 31 32 40 41 42 43 44
10 20 21 30 31

T_2

GRAPHE IMPLICATION DANS POPULATION "MANIPULATIONS"

indice $q_p(\bar{a},b)$ centré-réduit

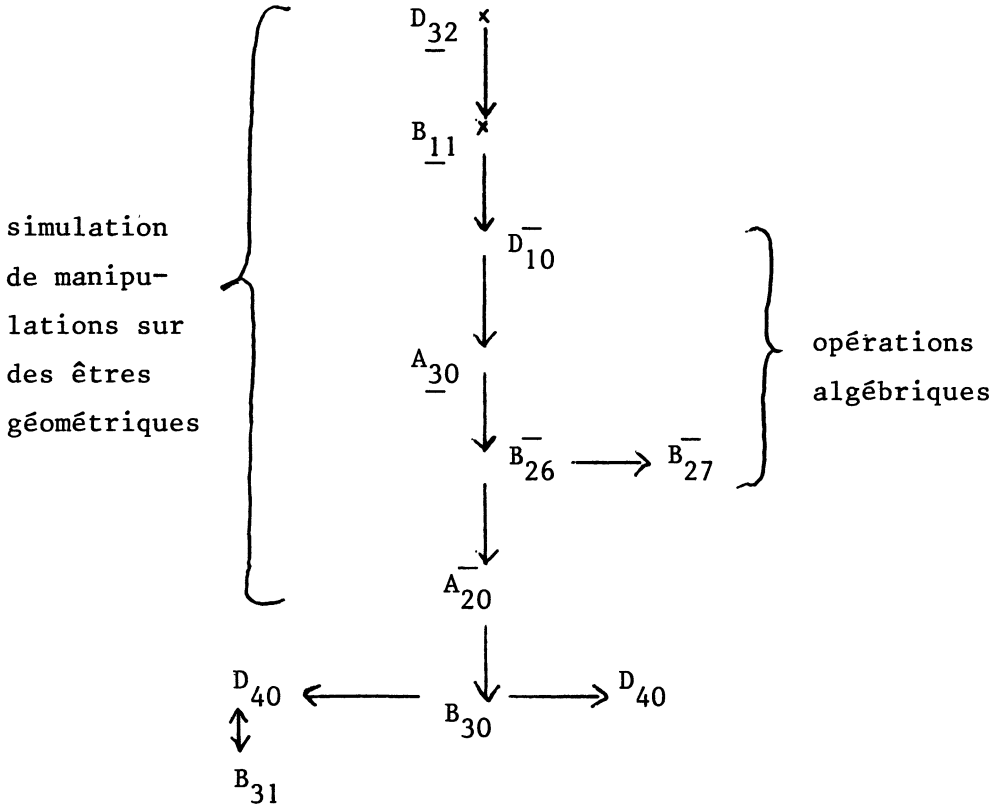


La construction de G_2 se trouve facilitée par un nombre plus modeste d'implications, sans que soient exclus, grâce à la fermeture transitive, les arcs permettant de conférer une structure intéressante à ce graphe.

Les remarques ci-dessous visent à comparer et à différencier G_1 et G_2:

1) comme dans G_1, les sources (resp. les puits) sont constituées des items relevant des niveaux supérieurs (resp. inférieurs) de la taxinomie, à l'exception encore de C_{31}, mais également de D_{42}, C_{10} et C_{30}.

2) l'ordre taxinomique n'est pas plus infirmé dans G_2 que dans G_1 et les grandes sous-structures révélées par γ_1, γ_2 et γ_3 demeurent en place. Cependant, ces chemins se trouvent interrompus par d'autres items comme c'est le cas de γ_3 par D_{10} et B_{26}. Ces deux items relatifs à des opérations algébriques, présentent des fréquences de réussite s'insérant précisément dans l'ordre des fréquences des noeuds de γ_3 :



3) de G_1 à G_2, on note la disparition de quelques chemins dont les plus intéressants nous semblent ceux-ci :

$$\alpha)\ (C_{31} \longrightarrow C_{30}) \longrightarrow (B_{30} \longrightarrow B_{31})$$

où l'opération algorithmique de la démonstration s'associait, non contre nature, à l'opération algorithmique d'une construction géométrique.

$\beta)$ $D_{31} \longrightarrow D_{30}$. Or, ces deux items ont des affinités manifestement très fortes. D_{31} exige l'acquisition de la réversibilité de l'opération : $(x,y) \longrightarrow \frac{x+y}{2}$, acquisition qui domine et suit, génétiquement parlant, celle de l'opération directe.

Formulons alors une hypothèse qui mérite de rechercher sa validation à travers d'autres études comme celles-ci. Reprenant les 2 cas i) et ii) énoncés dans III.1, nous conjecturons que :

. $q_p(a,\bar{b})$ semble être un indicateur de i): en effet, s'il reste par construction sensible à la complexité, c'est la nature des tâches et, par suite, les démarches intellectuelles, les processus de réalisations à travers les contenus qui infléchissent la structure du graphe que nous lui associons.

. $q_p(\bar{a},b)$ rend davantage compte de ii) en se présentant comme un très bon indicateur des faibles variations de complexité, donc de la finesse d'emboitement de grandes classes de complexité .

La complexité taxinomique ne serait-elle pas le facteur principal de ressemblance entre les deux graphes ?

Nous allons nous attacher dans l'avenir à éprouver cette hypothèse qui, si elle se confirmait, donnerait un statut spécifique aux deux formes d'indices dans l'analyse didactique et justifierait, s'il en était besoin, leur coexistence.

V - EVALUATION DE LA RELATION ENTRE LE GRAPHE D'IMPLICATION ET LA TAXINOMIE

V.1. Introduction

Nous avons, au paragraphe précédent, donné les grandes lignes de l'interprétation des graphes discrets d'implication obtenus à partir de $q_p(a,\bar{b})$ et de $q_p(\bar{a},b)$ par rapport à la taxinomie d'objectifs cognitifs et à travers une analyse des tâches. Une telle approche, nous l'avons vu, suppose l'affectation des différents items du questionnaire aux différentes classes d'objectifs cognitifs. Une telle affectation est, répétons-le, une opération très délicate où l'appréciation subjective du didacticien intervient. En d'autres termes, l'échelle d'appropriation d'un même concept, définie par la taxinomie, peut être sans reproches , alors que l'affectation des items aux classes de la taxinomie peut être problématique. Ce n'est sans doute pas la seule raison qui a fait que dans l'exemple qui nous intéresse ici, l'ordre taxinomique des items n'a pas toujours été celui des fréquences de réussite.

De toute façon, de la taxinomie nous retiendrons le préordre total qu'elle définit sur l'ensemble des items du test et nous chercherons à déterminer dans quelle mesure ce préordre total s'ajuste d'une part, au graphe d'implication pondéré et d'autre part, au graphe d'implication discret. Une "bonne" valeur de l'ajustement justifierait a posteriori la démarche entreprise pour l'interprétation des graphes et validerait du même coup la taxinomie à travers son opérationalisation dans le cadre du concept étudié de symétrie centrale. On appréciera d'autant plus une "bonne" valeur de l'adéquation qu'au départ, le graphe d'implication est compatible avec la relation de préordre total sur l'ensemble des items définie par la fréquence de réussite , alors que le préordre taxinomique ne l'est pas.

Relativement à un ensemble $A = \{a_i / 1 \leqslant i \leqslant n\}$ de comportements ordonnés (par complexité décroissante, par exemple) où l'indexation est établie conformément à l'ordre :
$(\forall 1 \leqslant i,j \leqslant n)$, $i<j$ si et seulement si $a_i \Rightarrow a_j$, on propose l'indice suivant, également de J. Loevinger,

$$\mathcal{H} = 1 - \frac{\Sigma\{p(\bar{a}_i \wedge a_j / i<j\}}{\Sigma\{p(\bar{a}_i)p(a_j)/i<j\}} \qquad (1)$$

pour "mesurer" le degré de pertinence de l'échelle ainsi définie sur A. Cet indice présente la même faiblesse, mise en évidence au paragraphe III.1 ci-

dessus, que celui H(a,b) (cf. formules (5) et (6) § III.1) qui en est l'origine.

Nous aurions certes pu associer au préordre total défini à partir de la fréquence de réussite ou bien, au graphe discret d'implication, un indice de même inspiration que (1), mais basé sur l'un ou l'autre des deux indices $\psi\,[q_p(\bar{a},b)]$ et $\psi[q_p(a,\bar{b})]$, où nous aurions cherché à nous situer par rapport à une hypothèse d'absence de lien définie au niveau global de l'ensemble des couples d'items. Notre démarche sera différente ; en effet, la donnée du graphe pondéré par $\psi[q_p(\bar{a},b)]$ (resp. $\psi[q_p(a,\bar{b})]$) est une structure descriptive du comportement trop riche pour qu'on cherche à la valider par un indice statistique. C'est plutôt cette structure qui permettra dans sa confrontation avec le préordre total défini par la taxinomie de valider localement cette taxinomie et de tester du même coup la manière dont nous l'avons opérationalisée.

L'évaluation de cette relation se fait à partir d'un indice très général de comparaison entre deux relations quelconques pondérées ou non sur un même ensemble fini. Dans notre cas précis, l'ensemble considéré sera celui, A des 40 items du test et les deux relations seront respectivement le préordre total défini par la taxinomie, que nous appellerons dans la suite plus brièvement "la taxinomie" et le graphe discret (resp. pondéré par l'intensité d'implication).

Pour arriver à notre but, il nous faut répondre aux trois questions suivantes :

. qu'est-ce qu'une relation pondérée et comment la représente-t-on ?

. comment définit-on l'indice de comparaison entre deux relations pondérées ?

. comment un tel indice de comparaison peut servir pour l'évaluation du graphe d'implication par rapport à la taxinomie ?

La réponse à ces questions, constitue l'objet des trois points discutés dans ce paragraphe.

V.2. Relation pondérée sur un ensemble A

Une relation pondérée sur A , que l'on appelle aussi variable "mesure" ou variable "pondération" sur A x A , se définit par la donnée d'une application $\mathfrak{F}$ de A x A dans $\mathbb{R}$:

$$\xi : A \times A \to \mathbb{R}$$
$$(a_i, a_j) \mapsto \xi(a_i, a_j)$$

où $\xi(a_i,a_j)$ est la "mesure" affectée à (a_i,a_j). Si l'on indexe A par $I = \{1,\ldots,n\}$ où $n = \text{card}(A)$, on notera alors ξ_{ij} la "mesure" affectée à (a_i,a_j) ou à(i,j). Dans ces conditions, une telle variable "mesure" peut être représentée au moyen de la matrice $(\xi_{ij})_{(ij)\in IxI}$. Remarquons que la variable "mesure" est une variable très générale qui contient comme cas particulier toutes les variables rencontrées dans la science de l'Homme et de la Nature.

V.3. Indice de comparaison entre deux variables "pondération"

Le choix d'un indice de comparaison ou indice de similarité entre deux variables est un problème crucial, complexe et ancien. De nombreux indices ont été proposés pour des cas particuliers ; en ce qui nous concerne le choix d'indice de similarité entre deux variables, quelle qu'en soit la nature, se situe dans le cadre de la démarche de I.C. Lerman où l'indice se réfère à une hypothèse d'absence de lien (cf. § II).

Soient donc ξ et η deux variables "pondération" sur $A \times A$, représentées par les matrices carrées (ξ_{ij}) et (η_{ij}). De façon analogue au § II.1, l'élaboration d'un indice de similarité entre deux structures finies de même type α et β, en particulier entre ξ et η, se fait de la façon suivante :

On part d'un indice "brut" soit $s(\alpha,\beta)$

On associe à $s(\alpha,\beta)$ une variable aléatoire $S(\alpha,\beta)$ par l'introduction d'une hypothèse N d'absence de lien.

On prend comme indice définitif $P(\alpha,\beta) = \text{Prob}\,[S< s/N]$.

Nous partons donc d'un indice "brut" observé, soit

$$s(\xi,\eta) = \sum_{(i,j)\in I_2} \xi_{ij}\,\eta_{ij} \quad \text{où} \quad I_2 = \{(i,j)\in I \times I / i \neq j\} \qquad (2)$$

Dans le cadre de l'hypothèse N, on associe deux variables aléatoires duales à $s(\xi,\eta)$ soient :

$$S_1(\xi,\eta) = \sum_{(i,j)\in I_2} \xi_{ij}\,\eta_{\sigma(i)\sigma(j)} \quad \text{et} \quad S_2(\xi,\eta) = \sum_{(i,j)\in I_2} \xi_{\tau(i)\tau(j)}\,\eta_{ij} \qquad (3)$$

où σ et τ sont des permutations aléatoires dans l'ensemble G_n, de toutes les permutations sur (1,...,n), muni d'une probabilité uniformément répartie.

Nous allons, avant de reprendre certaines expressions calculs, rappeler l'historique de l'introduction de cet indice. S'inspirant d'un ancien article de H.E. Daniels (cf.[3]), G. Lecalvé (cf.[12]) a eu l'idée de généraliser l'indice de proximité de I.C. Lerman entre variables d'un même type (parmi les plus couramment rencontrées dans un questionnaire) (cf.[15]) à la situation de la comparaison entre deux pondérations sur ExE, où E est l'ensemble des individus. La technique utilisée par G. Lecalvé était géométrique et correspondait en fait à celle de H.E. Daniels. I.C. Lerman a repris cette idée de façon combinatoire, ce qui l'a conduit à une expression plus précise des moments de la distribution de la v.a. S et à une meilleure justification de la tendance asymptotique normale de la loi de S(cf.[14]).

Il s'avère, nous venons de l'apprendre, que cette approche a des affinités avec celle de L. Hubert et al. (cf.[9],[10] et [21]). Toutefois, dans la publication principale de ces chercheurs (cf.[21]), qui n'est pas antérieure à celles qui nous concernent (cf. [12] et [14]), l'origine de l'idée est attribuée à N.Mantel (1967) (cf. [19]), alors qu'il aurait fallu remonter jusqu'à H.E.Daniels (1944). D'autre part, on ne trouve dans [21] que l'expression des deux premiers moments de la v.a. S. Enfin, l'optique développée par ces auteurs reste celle, classique, des tests d'hypothèse et non, comme c'est le cas pour nous, de l'évaluation d'une proximité au moyen d'un indice que nous appelons de "vraisemblance du lien", qui se réfère à une échelle [0,1] de probabilité et qui est donné par le complément à 1 du seuil critique tel qu'il est défini dans les tests.

L'élément aléatoire étant σ de G_n (cardinal $G_n = n!$), on a :

a - Espérance de S :

$$\mathcal{E}(S) = \frac{1}{n!} \sum_{\sigma \in G_n} \left(\sum_{(i,j) \in I_2} \xi_{ij} \eta_{\sigma(i)\sigma(j)} \right)$$

$$= \sum_{(i,j) \in I_2} \xi_{ij} \left(\frac{1}{n!} \sum_{\sigma \in G_n} \eta_{\sigma(i)\sigma(j)} \right)$$

On a $\sum_{\sigma \in G_n} \eta_{\sigma(i)\sigma(j)} = (n-2)! \sum_{(k,h) \in I_2} \eta_{kh}$

car le nombre de permutations pour lesquelles $\sigma(i)=k$ et $\sigma(j)=h$, $k \neq h$, est $(n-2)!$; $\eta_{\sigma(i)\sigma(j)}$ prend donc $(n-2)!$ fois la valeur η_{kh}. D'où

$$\mathcal{E}(S) = \frac{1}{n(n-1)} \left(\sum_{(i,j) \in I_2} \xi_{ij} \right) \left(\sum_{(k,h) \in I_2} \eta_{kh} \right) = n(n-1)\, \bar{\xi}\, \bar{\eta} \quad (4)$$

où $\bar{\xi} = \frac{1}{n(n-1)} \sum_{(i,j) \in I_2} \xi_{ij}$ et $\bar{\eta} = \frac{1}{n(n-1)} \sum_{(k,h) \in I_2} \eta_{kh}$

où $I_2 = \{(i,j)/(i,j) \in I \times I \text{ et } i \neq j\}$; on a $\text{card}(I_2) = n(n-1)$.

b - <u>Variance de S</u> :

$V(S) = \mathcal{M}_2(S) - \mathcal{E}^2(S)$. $\mathcal{M}_2$ étant le moment d'ordre deux de S. On a :

$\mathcal{M}_2(S) = \mathcal{E}\left[\sum_{(i,j) \in I_2} \xi_{ij} \eta_{\sigma(i)\sigma(j)} \right]^2$. En développant le carré, on a :

$$\mathcal{M}_2(S) = \mathcal{E}\left[\sum_{(i,j) \in I_2} \xi_{ij}^2 \eta_{\sigma(i)\sigma(j)}^2 \right] + \mathcal{E}\left[\sum_{((i,j),(h,k)) \in I_4} \xi_{ij} \xi_{hk} \eta_{\sigma(i)\sigma(j)} \eta_{\sigma(h)\sigma(k)} \right]$$

avec $I_4 = \{((i,j),(h,k)) \in I_2 \times I_2 / (i,j) \neq (h,k)\}$ (5)

Pour calculer les sommations, on décompose I_4 comme suit :

$I_4 = I \cup G1 \cup G'1 \cup G2 \cup G'2 \cup H$ où

$I = \{((i,j),(j,i))\}$

$G1 = \{((i,j),(i,k))\}$

$G'1 = \{((i,j),(k,i))\}$ des lettres distinctes désignent

$G2 = \{((i,j),(k,j))\}$ des indices distincts.

$G'2 = \{((i,j),(j,k))\}$

$H = \{((i,j),(k,l))\}$

L'intérêt de cette partition est que, d'une part les permutations $\sigma \in G_n$ laissent invariants ces différents sous-ensembles et que d'autre part le nombre de permutations qui, appliquées à un point de I_4, donnent la même image est $(n-4)!$ si le point appartient à H, $(n-3)!$ si le point appartient à G1, G'1, G2, G'2, et $(n-2)!$ si le point appartient à I ou I', où on note pour des raisons d'homogénéité $I' = \{((i,j),(i,j))\}$. On a alors :

$$\eta_2(S) = \mathcal{E}[\ \underset{I'}{\Sigma\bullet} + \underset{I}{\Sigma\bullet} + \underset{G1}{\Sigma\bullet} + \underset{G'1}{\Sigma\bullet} + \underset{G2}{\Sigma\bullet} + \underset{G'2}{\Sigma\bullet} + \underset{H}{\Sigma\bullet}\ ]$$

où par exemple on a $\underset{G1}{\Sigma\bullet} = \underset{((i,j),(1,k))\in G1}{\Sigma} \xi_{ij}\,\xi_{ik}\,\eta_{\sigma(i)\sigma(j)}\eta_{\sigma(i)\sigma(k)}$. et

$\mathcal{E}[\underset{G1}{\Sigma}] = \underset{G1}{\Sigma}\ \xi_{ij}\ \xi_{ik}\ \left(\frac{1}{n!}\ \underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\ \sigma(j)}\ \eta_{\sigma(i)\sigma(k)}\right)$. Pour les différents sous-ensembles, on a les résultats intermédiaires suivants, où le dernier membre emprunte des notations du programme de calcul :

$$\underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\sigma(j)}\,\eta_{\sigma(i)\sigma(j)} = (n-2)!\ \underset{((k,1),(k,1)\in I'}{\Sigma}\ \eta^2_{k1} = (n-2)!\quad \text{SV2IP}$$

$$\underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\sigma(j)}\,\eta_{\sigma(j)\sigma(i)} = (n-2)!\ \underset{I}{\Sigma}\ \eta_{ij}\,\eta_{ji} = (n-2)!\ \text{SV2I}$$

$$\underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\sigma(j)}\,\eta_{\sigma(i)\sigma(k)} = (n-3)!\ \underset{G1}{\Sigma}\ \eta_{ij}\,\eta_{ik} = (n-3)!\ \text{SV2G1}$$

$$\underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\sigma(j)}\,\eta_{\sigma(k)\sigma(i)} = (n-3)!\ \underset{G'1}{\Sigma}\ \eta_{ij}\,\eta_{ki} = (n-3)!\ \text{SV2GP1}$$

$$\underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\sigma(j)}\,\eta_{\sigma(k)\sigma(j)} = (n-3)!\ \underset{G2}{\Sigma}\ \eta_{ij}\,\eta_{kj} = (n-3)!\ \text{SV2G2}$$

$$\underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\sigma(j)}\,\eta_{\sigma(j)\sigma(k)} = (n-3)!\ \underset{G'2}{\Sigma}\ \eta_{ij}\,\eta_{jk} = (n-3)!\ \text{SV2GP2}$$

$$\underset{\sigma\in G_n}{\Sigma}\ \eta_{\sigma(i)\sigma(j)}\,\eta_{\sigma(k)\sigma(1)} = (n-4)!\ \underset{H}{\Sigma}\ \eta_{ij}\,\eta_{k1} = (n-4)!\ \text{SV2H} \qquad (6)$$

Si on note de façon analogue les sommations sur la variable ξ par SV1IP, SV1I, SV1G1, SV1GP1, SV1G2, SV1GP2, SV1H où par exemple :
SV1GP1 = $\sum_{((i,j),(k,i)) \in G1} \xi_{ij}\, \xi_{ki}$, le moment d'ordre deux devient alors :

$$\eta_2(S) = \frac{SV1I * SV2I + SV1IP * SV2IP}{n(n-1)} + \frac{SV1H * SV2H}{n(n-1)(n-2)(n-3)}$$

$$+ \frac{SV1G1 * SV2G1 + SV1GP1 * SV2GP1 + SV1G2 * SV2G2 + SV1GP2 * SV2GP2}{n(n-1)(n-2)} \quad (7)$$

Ainsi, on calcule la variance $V(S) = \eta_2(S) - \gamma^2(S)$.

Soit

$$Q(\xi,\eta) = \frac{s(\xi,\eta) - \gamma(S)}{\sqrt{V(S)}} \quad (8)$$

On a alors

$$P(\xi,\eta) = Pr^{N}[S < s] = Pr^{N}\left[\frac{S - \gamma(S)}{\sqrt{V(S)}} < Q(\xi,\eta)\right]$$

En faisant référence à la loi normale, on a en définitive

$P(\xi,\eta) = \phi[Q(\xi,\eta)]$ (9) où ϕ est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$.

Remarques

. L'indice de similarité entre ξ et η est donc $P(\xi,\eta)$; mais si $Q(\xi,\eta)$ atteint de grandes valeurs et que l'on a à comparer plusieurs indices de similarité entre eux, il est intéressant de prendre comme indice Q et non pas P. Par exemple si on a $Q(\xi_1,\eta) = 5$ et $Q(\xi_2,\eta) = 6$, avec l'indice P on ne peut pas apprécier la différence de similarité ; on aura dans les deux cas P=1. Soulignons d'autre part, que le point de vue développé n'est nullement celui des tests d'hypothèses où il s'agirait juste de mettre à l'épreuve l'hypothèse d'indépendance entre les deux relations. Cette dernière hypothèse constitue pour nous un référentiel. Pour se fixer les idées, disons qu'une valeur 7 de l'indice $Q(\xi,\eta)$ serait autrement appréciée qu'une valeur 3 ; alors que ces deux valeurs de $Q(\xi,\eta)$ auraient conduit à la même attitude dans le cadre de la démarche des tests d'hypothèses.

. D'un point de vue opérationnel, il faut calculer les différentes sommes partielles SV1I, SV2I, SV1G1, SV2G1, SV1H, SV2H ... . On peut établir à partir d'une certaine forme de ces sommes, que l'algorithme de calcul est en n^2, dans le cas très général de la comparaison de deux variables "pondération" sur IxI. D'autre part, dans le cas où l'une des variables est d'un type particulier tel que définissant une partition, un ordre total ou un préordre total, on effectuera le calcul de façon plus directe. Un exemple sera donné au paragraphe V.4 où l'une des variables est préordinale[1].

V.4. Utilisation de l'indice précédent pour l'évaluation du graphe d'implication par rapport à la taxinomie

La première variable "pondération" ou relation pondérée sera la taxinomie. D'une façon formelle, la taxinomie est un préordre total à cinq classes sur A. Ce préordre peut être considéré comme une relation pondérée en posant :

$$\xi : A \times A \longrightarrow \mathbb{R}$$

$$\xi(a,b) = \begin{cases} 1 \text{ si a précède strictement b au sens de la taxinomie} \\ 0 \text{ sinon} \end{cases} \qquad (10)$$

Pour fixer les idées on a $\xi(a,b) = 1$ si la question a est considérée plus "difficile" que la question b.

On représente la variable par la matrice bloc triangulaire (ξ_{ij}) suivante :

	A	B	C	D	E
A					
B	1				
C	1	1			
D	1	1	1		
E	1	1	1	1	

(1) Les différents programmes sous-tendant cette étude peuvent être demandés auprès de H. ROSTAM à l'adresse indiquée au bas de la première page de cet article.

$\xi_{ij} = 0$ si i et j appartiennent à une même classe ou i appartient à une classe qui suit (i.e. dans notre contexte, considérée plus "facile" que) la classe de j.

$\xi_{ij} = 1$ si i appartient à une classe considérée plus "difficile" que la classe de j.

Pour calculer les sommes partielles, intervenant dans l'indice, concernant la taxinomie, il suffit d'avoir les cardinaux des classes de la taxinomie. Soit n_i, i=1,... N_c ces cardinaux où N_c désigne le nombre de classes. On a $\sum_{1 \leq i \leq N_c} n_i = n = \text{card}(A)$.

Pour des raisons de facilité d'écriture et opérationnelles, on note

$$n_i^c = \sum_{1 \leq j \leq i \leq N_c} n_j = (n_1 + n_2 + \ldots + n_{i-1})$$

$$n_i^f = \sum_{j>i} n_j = (n_{i+1} + n_{i+2} + \ldots + n_{N_c})$$

Dans ces conditions, on a directement :

$$SV1I = \sum_{I} \xi_{i'j'} \xi_{j'i'} = 0$$

$$SV1IP = \sum_{I'} \xi_{i'j'} = \sum_{I2} \xi_{i'j'} = \sum_{i<j} n_i \times n_j = \sum_{i=1}^{N_c-1} n_i n_i^f$$

$$SV1G1 = \sum_{G1} \xi_{i'j'} \xi_{i'k'} = \sum_{j=2}^{N_c} n_j \; n_j^c (n_j^c - 1)$$

$$SV1GP1 = SV1GP2 = \sum_{G'1} \xi_{i'j'} \xi_{k'i'} = \sum_{G'2} \xi_{i'j'} \xi_{j'k'} = \sum_{i=2}^{N_c-1} n_i \; n_i^c \; n_i^f$$

$$SV1G2 = \sum_{G2} \xi_{i'j'} \xi_{k'j'} = \sum_{j=1}^{N_c-1} n_j \; n_j^f \; (n_j^f - 1)$$

$$SV1H = \sum_{H} \xi_{i'j'} \xi_{k'l'} = \sum_{i<j} n_i \; n_j \left[\sum_{k=1}^{N_c-1} n_k n_k^f - 2n + n_i + n_j + 1 \right] \qquad (11)$$

où un indice de la forme "i" est celui d'une classe et de la forme "i'" d'un élément.

La deuxième variable "pondération" sera soit le graphe pondéré "d'intensité d'implication", soit le graphe discret "d'implication" associé à ce dernier par le choix d'un seuil.

Le graphe "d'intensité d'implication" est, rappelons-le, obtenu en affectant à tout arc du graphe de la relation de préordre total définie par la fréquence de réussite (cf. § III.3 (γ)) la valuation définie par $\psi[q_p(\bar{a},b)]$ (resp. $\psi\,[q_p(a,\bar{b})]$) ou d'ailleurs $\psi\,[q'_p(\bar{a},b)]$ (resp. $\psi\,[q'_p(a,\bar{b})]$). (cf. formule (31) § III.3). Désignons par $\eta(a,b)$ variable "pondération", elle se trouve définie comme suit :

$$\eta : A \times A \longrightarrow \mathbb{R}_+$$

$$(\forall (a,b) \in A \times A), n(a,b) = \begin{cases} \psi[q_p(a,\bar{b})] \text{ (resp. } \psi[q_p(\bar{a},b)]) \\ \text{si } \operatorname{card}(E(a)) < \operatorname{card}(E(b)) \\ 0 \text{ sinon} \end{cases} \tag{12}$$

Dans le cas du graphe discret d'implication, une telle variable sera définie par

$$(\forall (a,b) \in A \times A),\ \eta(a,b) = \begin{cases} 1 \text{ si } \operatorname{card}(E(a)) < \operatorname{card}(E(b)) \text{ et si} \\ \psi\,[q_p(a,\bar{b})] \geq \psi_o \text{ (resp. } \psi[q_p(\bar{a},b)] \geq \psi_o) \\ 0 \text{ sinon} \end{cases} \tag{13}$$

ψ_o étant le seuil à partir duquel (dans une optique cette fois-ci des tests d'hypothèses), on admet l'implication.

En indexant A par $I = \{1,2,\ldots,n\}$, cette dernière variable sera représentée par la matrice carrée $\{\eta_{ij} / (i,j) \in I \times I\}$ (14)

Pour calculer les sommes partielles concernant la deuxième variable η c'est la méthode la plus générale de calcul qui doit être utilisée pour évaluer les sommes (11).

Une fois les sommes partielles déterminées, on en déduit la variance $V[S(\xi,\eta)]$. L'espérance $\mathcal{E}[S(\xi,\eta)]$ se calcule sans difficulté. L'indice brut de similarité $s(\xi,\eta)$ s'obtient seulement par référence à la matrice (η_{ij}) ; (ξ_{ij}) sert simplement comme tableau d'adressage pour les éléments "utiles" de (η_{ij}).

Ainsi, on obtient $P(\xi,\eta)$ et $Q(\xi,\eta)$ qui mesurent la conformité du graphe "d'implication" ou graphe "d'intensité d'implication" avec la "taxinomie".

Nous terminons ce paragraphe en donnant les résultats des différents traitements que l'on a effectués dans le cadre de l'expérience réelle décrite dans l'introduction.

Une première série de traitements est effectuée sur la base de l'échantillon de 401 élèves :

. Indice de similarité entre la "taxinomie" et le graphe "d'intensité d'implication" ; on a obtenu :

$Q(\xi,\eta) = 5,12$ où η est définie par $\eta(a,b) = \psi[q_p(a,\bar{b})]$

$Q(\xi,\eta) = 4,98$ avec $\eta(a,b) = \psi[q_p(\bar{a},b)]$

. Indice de similarité entre la "taxinomie" et le graphe "d'implication" ; on a dans ce cas :

$$Q(\xi,\eta) = 3,91 \quad \text{avec } \eta(a,b) = \begin{cases} 1 \text{ si } \psi[q_p(a,\bar{b})] \geq \psi_\circ \\ 0 \text{ sinon} \end{cases}$$

$$Q(\xi,\eta) = 3,16 \quad \text{avec } \eta(a,b) = \begin{cases} 1 \text{ si } \psi[q_p(\bar{a},b)] \geq \psi_\circ \\ 0 \text{ sinon} \end{cases}$$

Une deuxième série de traitements est effectuée sur la base de 229 élèves. Parmi les 401 élèves qui ont subi le test il y en avait 229 qui avaient manipulé préalablement un certain matériel dont l'évocation facilitait la résolution de certaines questions du test.

. Indice de similarité entre la "taxinomie" et le graphe "d'intensité d'implication". On a les résultats suivants :

$Q(\xi,\eta) = 4.83$ avec $\eta(a,b) = \psi[q_p(a,\bar{b})]$

$Q(\xi,\eta) = 4.77$ avec $\eta(a,b) = \psi[q_p(\bar{a},b)]$

. Indice de similarité entre la "taxinomie" et le graphe "d'intensité d'implication" obtenu à l'aide des indices $q'_p(a,\bar{b})$ et $q'_p(\bar{a},b)$. On a

$Q(\xi,\eta) = 4.54$ où $\eta(a,b) = \psi[q'_p(a,\bar{b})]$

$Q(\xi,\eta) = 4.27$ où $\eta(a,b) = \psi[q'_p(\bar{a},b)]$

Ces résultats prouvent l'excellente conformité du graphe discret "d'implication" ou pondéré "d'intensité d'implication" à la taxinomie.

VI - ALGORITHME D'ECHANGE ET RECHERCHE D'UN PREORDRE S'AJUSTANT AU "MIEUX" AU GRAPHE D'IMPLICATION

L'indice de comparaison entre deux relations pondérées vient d'être défini au § V. Cet indice permet d'apprécier le degré de "ressemblance" de deux relations pondérées vis-à-vis d'une troisième. Ainsi, l'indice peut jouer le rôle d'un certain critère de choix.

La première question que l'on se pose est la suivante : peut-on améliorer la taxinomie ? Cette question, posée de façon aussi brutale, n'a pas de sens. En effet, répétons-le une deuxième fois, il faut distinguer trois niveaux dans les concepts manipulés :

Au premier niveau, on a la taxinomie d'objectifs cognitifs proposé par R. Gras (voir le tableau d'objectifs cognitifs § IV).

Au deuxième niveau, il y a l'ensemble A des 40 items élaboré par R. Gras "conforme" à la taxinomie et qui est censé représenter cette dernière dans le cadre d'un concept donné. Savoir dans quelle mesure le préordre sur A représente la taxinomie n'est pas, ici, de notre préoccupation.

Enfin au niveau trois, on a les graphes d'implication ou les graphes d'intensité d'implication, qui sont construits à l'aide de différents indices d'implication proposés au § III et dont les sommets sont les éléments de A .

Pour préciser la question précédente, signalons que notre étude se situe entre le niveau deux et le niveau trois. Au § V, lors de l'évaluation du graphe d'implication, par rapport à la taxinomie, on a implicitement plus ou moins confondu le niveau un et le niveau deux ; en d'autres termes on a supposé que le préordre sur A représentait, à travers le concept de la symétrie centrale, la taxinomie.

Ici, puisque l'on parle de l'amélioration, la question ne peut plus se poser en terme de taxinomie. En effet, si une amélioration est possible, cela ne peut être qu'au niveau du préordre sur A . D'où notre nouvelle préoccupation qui se résume de la façon suivante :

On cherche, au moyen d'un algorithme, à substituer au préordre initial sur A (qui était censé représenter la taxinomie), un autre préordre sur A qui s'ajuste "au mieux" au graphe d'implication.

Le but final est de trouver un "meilleur" préordre sur A sans poser de contraintes a priori. A l'état actuel de notre recherche, nous avons mis au point un algorithme d'échange qui pose comme contrainte la fixité des cardinaux des classes du préordre. On envisage d'éliminer cette contrainte par la mise au point d'un algorithme de transfert.

On présente donc, en premier lieu, l'algorithme d'échange; ensuite, on explicitera l'amélioration de l'indice de similarité entre le nouveau préordre et le graphe d'implication.

VI.1. Algorithme d'échange

Soient $\xi^{(1)}$ un préordre sur A de N_c classes de cardinaux respectifs $n_1, n_2, \ldots, n_{Nc}$ et η un graphe d'intensité d'implication sur A. Soit $Q(\xi^{(1)},\eta) = \dfrac{s(\xi^{(1)},\eta) - \mathcal{E}[S(\xi^{(1)},\eta)]}{\sqrt{V[S(\xi^{(1)},\eta)]}}$, l'indice de similarité entre $\xi^{(1)}$ et η.

L'algorithme d'échange se propose d'améliorer Q en remplaçant $\xi^{(i)}$ par $\xi^{(i+1)}$ déduit de $\xi^{(i)}$ par échange entre deux éléments appartenant à deux classes consécutives de ce dernier. Plus précisément à chaque étape, on fait le "meilleur" échange parmi tous les échanges possibles entre éléments de tous les couples de classes successives.

Il y a $\sum_{1 \leqslant i \leqslant N_c - 1} n_i \, n_{i+1}$ échanges possibles.

Dans l'échange, les cardinaux des classes restent inchangés, il est facile de voir dans ces conditions les égalités suivantes :

$$\mathcal{E}[S(\xi^{(i)},\eta)] = \mathcal{E}[S(\xi^{(i+1)},\eta)]$$

$$V[S(\xi^{(i)},\eta)] = V[S(\xi^{(i+1)},\eta)] \quad \text{pour tout } i.$$

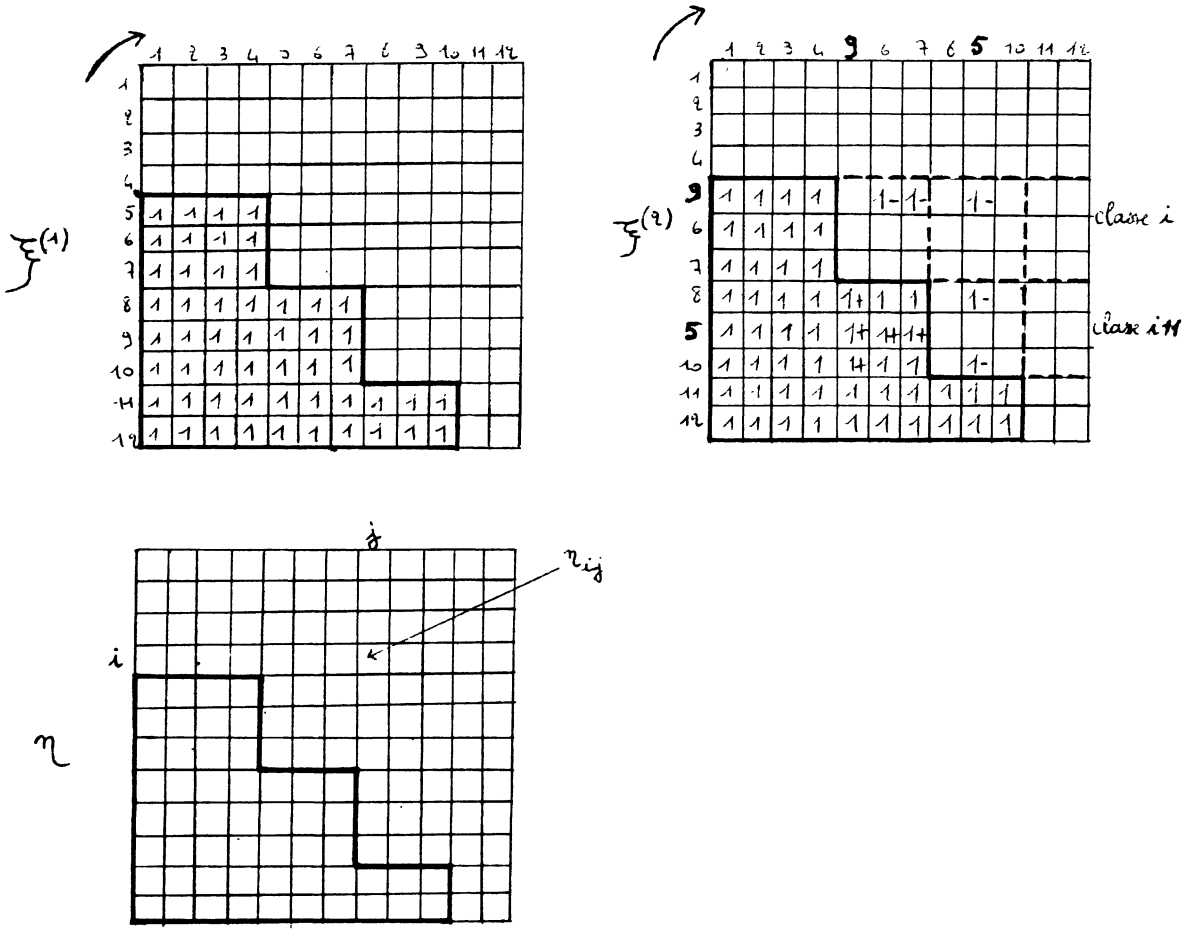
En effet les sommes partielles intervenant dans $\mathcal{E}$ et V sont invariantes par un tel échange .

On a vu au § V.4 que les sommes partielles, concernant ξ, ne dépendent que des cardinaux des classes. Or, ceux-ci ne changent pas.

Seul l'indice brut change. Ce dernier se calcule par référence à la matrice (η_{ij}). Illustrons sur un exemple l'effet d'un échange sur l'indice brut.

Soient A l'ensemble à 12 éléments notés de 1 à 12, $\xi^{(1)}$ un préordre à 4 classes de cardinaux respectifs 4,3,3,2. Soit η un graphe d'implication "sur A". η sera représenté par la matrice (η_{ij}). Pour fixer les idées, on peut avoir $\eta_{ij} = \psi\,[q_p(i,\bar{j})]$.

$\xi^{(1)}$ de façon formelle, sera représenté par la matrice $\{\xi^{(1)}_{ij}/(i,j) \in I \times I\}$ bloc triangulaire suivante, où les cases vides sont remplies de zéros.



On a $s(\xi_1,\eta) = \sum_{(i,j)\in I_2} \xi^1_{ij}\,\eta_{ij} = \sum_{(i,j)\in \text{partie hachurée}} \eta_{ij}$

On voit que dans le calcul de l'indice brut, $\xi^{(1)}$ joue le rôle d'un tableau d'adressage pour les éléments de η. Supposons un échange entre l'élément 5 de la classe 2 et l'élément 9 de la classe 3. On a le nouvel indice brut :

$$s(\xi_2,\eta) = s(\xi_1,\eta) + \eta_{(8,9)} + \eta_{(5,9)} + \eta_{(10,9)} + \eta_{(5,6)} + \eta_{(5,7)} - \eta_{(9,6)} - \eta_{(9,7)} - \eta_{(9,5)} - \eta_{(8,5)} - \eta_{(10,5)}.$$

D'une façon plus générale, si on échange un élément 1 de la classe i avec un élément m de la classe(i+1):

. on retranche de l'indice brut initial tous les éléments de η correspondant aux implications suivantes :

- $m \Rightarrow x$ où $x \in$ classe i
- $n \Rightarrow 1$ où $n \in$ classe (i+1) et $n \neq m$

. On ajoute à l'indice brut initial tous les éléments de η correspondant aux implications suivantes :

- $1 \Rightarrow m$
- $\bar{1} \Rightarrow x$ où $x \in$ classe i et $x \neq 1$
- $n \Rightarrow m$ où $n \in$ classe(i+1) et $n \neq m$.

A chaque étape, il faut avoir les $\sum_{1 \leqslant i \leqslant N_c - 1} n_i n_{i+1}$ valeurs d'indice brut, correspondant au même nombre d'échanges. On fera l'échange qui augmente le plus la valeur de l'indice brut.

Une fois l'échange effectué, pour l'efficacité de l'algorithme, on met à jour le tableau η_{ij}.

L'algorithme permet d'arrêter les calculs lorsque la variation de l'indice d'une étape à l'autre devient négligeable.

Remarque : En associant à chaque couple de classes successives le meilleur échange et la valeur de l'accroissement de l'indice brut qu'il entraîne , on peut noter que si à une étape J, le meilleur échange a eu lieu entre la classe i et la classe (i+1) , il ne restera plus à l'étape (J+1) qu'à repérer le meilleur échange entre d'une part, les classes (i-1) et i, i et (i+1), enfin (i+1) et (i+2). Cette circonstance permet d'économiser très sensiblement le nombre d'évaluations.

VI.2. Amélioration de l'indice

Soient $\xi^{(i)}$ le préordre à l'étape i, $\xi^{(i+1)}$ le préordre à l'étape i+1. $\Delta_i = s(\xi^{(i+1)},\eta) - s(\xi^{(i)},\eta)$ est l'amélioration de l'indice brut de l'étape i à l'étape i+1. L'indice Q centré réduit à l'étape i+1 sera

$$Q(\xi^{(i+1)},\eta) = \frac{s(\xi^{(i+1)},\eta) - \mathcal{E}(S)}{\sqrt{V(S)}}$$

Résultats concrets dans le cadre de l'expérience réelle :

On part des résultats des traitements déjà effectués dans V.4. Dans le cadre de l'indice de similarité entre le préordre initial et le graphe d'implication on avait :

$$Q(\xi,\eta_1) = 5,12 \text{ pour } \eta_1(a,b) = \psi\left[q_p(a,\bar{b})\right]$$

$$Q(\xi,\eta_2) = 4,98 \text{ pour } \eta_2(a,b) = \psi\left[q_p(\bar{a},b)\right]$$

Les indices bruts étant respectivement

$s(\xi,\eta_1) = 422.43$

$s(\xi,\eta_2) = 335.14$

Après passage de l'algorithme d'échange, on obtient

$Q^*(\xi,\eta_1) = 7.596$

$Q^*(\xi,\eta_2) = 7.546$

Soit des indices bruts : $s^*(\xi,\eta_1) = 505.089$

$s^*(\xi,\eta_2) = 426.500$

L'amélioration est nette, elle représente 50 % de l'indice du départ. Il restera à préciser le contenu du nouveau préordre, qui semble respecter de près celui défini à partir de la fréquence de réussite.

Arrêtons ici l'introduction de cet aspect de la recherche qui doit se poursuivre et se concrétiser dans le cadre d'une thèse de 3ème cycle de H. Rostam où on s'attaquera par ailleurs au problème de la représentation automatique du graphe discret d'implication.

VII - INDICE D'IMPLICATION ENTRE DEUX CLASSES D'ATTRIBUTS

A un couple (a,b) d'attributs descriptifs, on associe le croisement entre les deux classifications dichotomiques "nettes" $\{E(a),E(\bar{a})\}$ et $\{E(b),E(\bar{b})\}$ et le tableau (10) (§ II.2) est celui des cardinaux des classes de ce croisement. Soit (A,B), un couple d'ensembles disjoints ($A \cap B = \phi$) d'attributs orientés (i.e. $a \in A$(resp. $b \in B$) si et seulement si $\bar{a} \notin A$ (resp. $\bar{b} \notin B$)) ; I.C. Lerman a dans [12] généralisé les données du croisement entre deux classifications dichotomiques "nettes", au croisement entre deux classifications dichotomiques "floues", respectivement associées à A et à B, par l'introduction d'un degré d'appartenance relatif d'un individu à une même classe d'attributs. On développe dans ces conditions l'analyse du croisement de deux classifications "floues" par la construction d'indices affectés aux cases du croisement qui généralisent formellement ceux du cas "net" et qui sont conformes à la statistique du χ^2.

Désignons par c(A) et c(B) les cardinaux des classes A et B ; d'autre part, par $\bar{A}$(resp. $\bar{B}$) l'ensemble des attributs directement opposés à ceux de A(resp. B) :

$$\bar{A} = \{\bar{a}/a \in A\}, \quad \bar{B} = \{\bar{b}/b \in B\} \quad . \quad (1)$$

On a bien entendu

$$c(\bar{A}) = c(A) \text{ et } c(\bar{B}) = c(B) \quad . \quad (2)$$

On montre dans [12] que les nombres correspondant à ceux n(a), $n(\bar{a})$, n(b) et $n(\bar{b})$ du cas "net" matérialisé par le tableau (10) (§ II.2), sont respectivement

$$n(A) = \frac{1}{c(A)} \sum \{n(a)/a \in A\}, \quad n(\bar{A}) = \frac{1}{c(A)} \sum \{n(\bar{a})/a \in A\}$$

$$n(B) = \frac{1}{c(B)} \sum \{n(b)/b \in B\}, \quad n(\bar{B}) = \frac{1}{c(B)} \sum \{n(\bar{b})/b \in B\} \quad (3)$$

où on a la propriété, assurant la cohérence de nos formules

$$n(A) + n(\bar{A}) = n(B) + n(\bar{B}) = n \quad (4)$$

On montre également que les nombres correspondants à ceux $n(a \wedge b)$, $n(a \wedge \bar{b})$, $n(\bar{a} \wedge b)$ et $n(\bar{a} \wedge \bar{b})$ du cas "net" sont respectivement :

$$n(A \wedge B) = \frac{1}{c(A)c(B)} \Sigma \{card(E(a) \cap E(b))/(a,b) \in A \times B\}$$

$$n(A \wedge \bar{B}) = \frac{1}{c(A)c(B)} \Sigma \{card(E(a) \cap E(\bar{b}))/(a,b) \in A \times B\}$$

$$n(\bar{A} \wedge B) = \frac{1}{c(A)c(B)} \Sigma \{card(E(\bar{a}) \cap E(b))/(a,b) \in A \times B\}$$

$$n(\bar{A} \wedge \bar{B}) = \frac{1}{c(A)c(B)} \Sigma \{card(E(\bar{a}) \cap E(\bar{b}))/(a,b) \in A \times B\} \qquad (5)$$

où on a également les formules de cohérence

$$\begin{aligned} n(A \wedge B) + n(A \wedge \bar{B}) &= n(A) \\ n(\bar{A} \wedge B) + n(\bar{A} \wedge \bar{B}) &= n(\bar{A}) \\ n(A \wedge B) + n(\bar{A} \wedge B) &= n(B) \\ n(A \wedge \bar{B}) + n(\bar{A} \wedge \bar{B}) &= n(\bar{B}) \end{aligned} \qquad (6)$$

Représentons, dans le tableau 2 x 2 de croisement, A(resp. B) par l'indice 1 et $\bar{A}$(resp. $\bar{B}$) par l'indice 0. Ici encore, compte tenu des relations (6) qu'on vient d'écrire, la connaissance du contenu de l'une des cases du tableau 2 x 2 des $n(r \wedge s)$ $(0 \leqslant r,s \leqslant 1)$ détermine, pour n(A) et n(B) donnés, la connaissance du contenu des autres cases.

Dans ces conditions, le nombre $\chi(r,s)$ $(0 \leqslant r,s \leqslant 1)$ qui correspond à la contribution "orientée" de la case (r,s) à la statistique du χ^2 du cas "net" (cf. formule (12) § II.2) se trouve défini par

$$\chi(r,s) = (n(r \wedge s) - n(r)n(s)/n) / \sqrt{n(r)n(s)/n} \qquad (7)$$

$0 \leqslant r,s \leqslant 1$.

Ici encore, on démontre que la somme des carrés des statistiques $\chi(r,s)$, $0 \leqslant r,s \leqslant 1$, est dans l'hypothèse d'absence de lien à caractère hypergéométrique, la réalisation d'une v.a. du χ^2 à 1 degré de liberté (cf [12]).

Par conséquent χ définit une mesure de l'intensité de lien entre classes d'attributs qui généralise directement l'indice noté q_p (se référant au modèle poissonnien) du cas "net" du croisement entre deux attributs. De façon explicite

$$\begin{cases} \pi(0,0) = \pi(A,B) \\ \pi(0,1) = \pi(A,\bar{B}) \\ \pi(1,0) = \pi(\bar{A},B) \\ \pi(1,1) = \pi(\bar{A},\bar{B}) \end{cases} \qquad (8)$$

Ainsi, les deux indices d'implication entre comportements, $q_p(\bar{a},b)$ et $q_p(a,\bar{b})$ se généralisent naturellement par les indices d'implication entre classes de comportements $\pi(\bar{A},B)$ et $\pi(A,\bar{B})$.

De même que les indices d'implication $q_p(\bar{a},b)$ et $q_p(a,\bar{b})$ n'ont de sens d'être considérés que si $n(a) \leqslant n(b)$, nous ne regarderons $\pi(\bar{A},B)$ et $\pi(A,\bar{B})$ que si $n(A) \leqslant n(B)$ (cf. formules (3)).

Un tel indice ($\pi(\bar{A},B)$ ou $\pi(A,\bar{B})$) peut par exemple permettre, relativement à l'opérationalisation de la taxinomie à travers un test et autour d'un concept, d'évaluer relativement aux différentes classes du préordre déterminé par la taxinomie, l'intensité d'implication pour les couples de classes consécutives. Ce qui permettra de reconnaître les sauts les plus importants entre une classe et la suivante.

BIBLIOGRAPHIE

[1] BLOOM B. et collaborateurs, *Taxinomie des objectifs pédagogiques*, Tome 1 : *domaine cognitif*, Montréal, Edition Nouvelle, 1969.

[2] BROUSSEAU G., "Evaluation et théorie de l'apprentissage en situation scolaire", *Conférence de Campinas (Brésil), février 1979*, texte ronéoté, I.R.E.M. de Bordeaux.

[3] DANIELS H.E., "The relation between measures of correlation in the universe of sample permutations", *Biometrika*, vol. 33 (1974).

[4] DEGENNE A., *Techniques ordinales en analyse des données statistiques*, Tome II, Paris, Hachette, 1972.

[5] FELLER W., *An introduction to probability theory and its applications*, Volume I, second edition, New York, John Wiley, 1964.

[6] FLAMENT C., DEGENNE A., VERGES P., "Analyse de similitude ordinale" *in Actes du Colloque International DGRST-CNRS, Marseille 11-13 décembre 1975, Informatique et Sciences Humaines*, n°40/41, Mars/juin 1979.

[7] GRAS R., *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'état, Université de Rennes-I, 1979.

[8] GUTTMAN L., STOUFFER S.A., SUCHMAN E.A., LAZARSFELD P.F., *The American Soldier, Vol.4 : Measurement and Prediction*, Princeton, University Press, 1950.

[9] HUBERT L., "Generalized Concordance", *Psychometrika*, 44, 2, 1979, 135-142.

[10] HUBERT L.J., BAKER F.B., "Evaluating the conformity of sociometric measurements", *Psychometrika*, 43, 1, 1978, 31-41.

[11] HUBERT L.J., SCHULTZ J.V., "Quadratic assignment as a general data analysis strategy", *British Journal of Mathematical and Statistical Psychology*, 1976, 29, 190-241.

[12] LECALVE G., *Problèmes d'analyse des données*, Thèse d'état (2ème partie), Université de Rennes-I, 1976.

[13] LERMAN I.C., "Combinatorial analysis in the statistical treatment of behavioral data", *Quality and Quantity*, 14 (1980), 431-469.

[14] LERMAN I.C., "Formal analysis of a general notion of proximity between variables" *in Actes Partiels du Colloque 'Congrès Européen des statisticiens' Grenoble, Septembre 1976*, Amsterdam, North Holland, 1977.

[15] LERMAN I.C., "Etude distributionnelle de statistiques de proximité entre structures finies de même type. Application à la classification automatique", *Cahiers du B.U.R.O.*, n°19, Paris, 1973.

[16] LERMAN I.C., "Croisement de classifications floues", *Publ. Inst. Stat. Univ. Paris*, 1979, XXIV, fasc. 1-2, 13-46.

[17] LERMAN I.C., "Analyse ordinale d'une classe d'échelles", *in Analyse des Données*, tome I, Paris, Publications A.P.M.E.P., 1980.

[18] LOEVIGNER J., "A systematic approach to the construction and evaluation of tests of ability", *Psychological Monographs*, 61, n°4, 1947.

[19] MANTEL N., "The detection of disease clustering and a generalized regression approach", *Cancer Research*, 1967, 27, 209-220.

[20] MATALON B., *L'analyse hiérarchique*, Paris, Mouton-Gauthier-Villars, 1965.

[21] SCHULTZ J.V., HUBERT L.J., "A non parametric test for the correspondance between two proximity matrices", *Journal of Educational Statistics*, 1976, 1, 59-67.

[22] TOURNEUR Y., "Classification des questions d'évaluation en mathématique" et "Taxonomie des objectifs cognitifs en mathématique : étude du modèle N.L.S.M.A.", *Mathematica et Paedagogia*, n°56 et n°57, 1972.

[23] VERGNAUD G., "Activité et connaissance opératoire", *Bulletin A.P.M.E.P.*, n°307, février 1977.