

E. HENRI RIARD

M. TENSAOUT

Construction d'un schéma d'implication entre variables issues d'une analyse des correspondances multiples. Développements. Application

Mathématiques et sciences humaines, tome 135 (1996), p. 19-31

http://www.numdam.org/item?id=MSH_1996__135__19_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1996, tous droits réservés.

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

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

CONSTRUCTION D'UN SCHEMA D'IMPLICATION ENTRE VARIABLES
ISSUES D'UNE ANALYSE DES CORRESPONDANCES MULTIPLES.
DÉVELOPPEMENTS. APPLICATION

E. Henri RIARD¹, M. TENSAOUT²

RÉSUMÉ — *L'une des limites de l'analyse des correspondances multiples appliquée à de grands tableaux de données qualitatives est la difficulté d'analyse et d'interprétation des structures de relations entre variables.*

Afin de dépasser la frontière descriptive, il est proposé une méthodologie de recherche de schémas d'implication reposant sur les fréquences conditionnelles données par les tableaux de Burt.

L'analyse des correspondances multiples y est utilisée comme filtre principal de variables à partir desquelles, et dans le cadre des méthodes explicatives, sont construits des graphes orientés. Ils sont accompagnés d'indices (fonction, direction d'influence, force motrice, outre celui de masse) qui apportent une plus grande lisibilité des variables impliquées et ouvrent sur d'autres analyses statistiques avancées.

SUMMARY — Construction of an involvement scheme among variables issued from a multiple correspondence analysis. Developments. Application.

One of the limits of multiple correspondence analysis applied on large tables of qualitative items, is the difficulty of analysis and interpretation on the structure of relations among variables.

In order to outpass the descriptive frontier, we propose a research methodology on involvement schemes which is based on conditional frequencies given by Burt's tables.

The multiple correspondence analysis is used as a principal filter for variables from which (and as part of explanatory methods) oriented graphs are built-up. They are accompanied by some indicators (function, influence direction, driving force, besides one of the mass), which provide a greater legibility of the variables involved, and open the way to other advanced statistical analysis.

1. INTRODUCTION

La richesse de l'analyse des correspondances multiples appliquée à de grands tableaux de données d'enquête (de forme $I \times J$) est cependant amenuisée par les difficultés à expliquer la structure des relations entre variables ou leur proximité sur le plan factoriel et à les utiliser pleinement dans l'interprétation de l'analyse. Ces restrictions tiennent au caractère symétrique attribué aux relations entre ces variables que rien n'interdit (par exemple Worsley, 1987) de considérer *a priori* comme asymétriques.

¹ Docteur ès Lettres, Laboratoire de Psychologie, Université Paris XIII, Avenue J.B. Clément, F 93430 Villetaneuse.

² Docteur en Économie, Laboratoire Gama, Université Paris X, 200 avenue de la République, F 92001 Nanterre Cedex.

Partant de ce point de vue, il est proposé une méthodologie de la recherche de schémas d'implication basée sur la comparaison de fréquences conditionnelles qui donne lieu à la construction de graphes. La procédure ne concerne que les variables estimées influentes à partir des critères de l'analyse des correspondances multiples. Son originalité consiste en effet à opérer à partir des résultats de cette dernière qui paraît plus adaptée pour réduire le nombre de variables initiales³. La théorie combinatoire permet elle aussi de rechercher des schémas d'implication entre variables mais elle nécessite un nombre restreint de variables pour qu'une interprétation soit possible (Van Buggenhaut, 1987, p.17).

Nous présentons [section II] les différentes approches des schémas d'implication (implications statistiques, approche booléenne, théorie des graphes), puis [section III] nous proposons la procédure d'analyse de fréquences conditionnelles développée sur un court exemple à partir duquel [section IV] sont avancés d'autres indices, qui concourent, eux-aussi, à enrichir la lecture des données de l'analyse des correspondances multiples.

2. MÉTHODES D'ANALYSE D'UN TABLEAU DE CONTINGENCE (I × J)

2.1. Présentation

Seuls les tableaux de contingence de grande taille, construits à partir des questionnaires ou des épreuves projectives présentent un intérêt véritable pour l'analyse des données⁴. Leur analyse est le plus souvent réalisée sur la base d'une réduction de données, la recherche de facteurs, la construction de classes, la sélection de variables qui résument au mieux l'information du tableau initial. Les variables ainsi sélectionnées permettent d'élaborer des schémas explicatifs ou prédictifs et décisionnels.

Aucune méthode d'analyse des données ne répond simultanément à ces exigences.

2.2. Pour répondre à ces objectifs, de multiples méthodes peuvent être retenues

2.2.1. *L'analyse des correspondances multiples*

L'analyse des correspondances multiples a une efficacité reconnue pour réduire les variables du fait qu'il s'agit d'une méthode descriptive exploratoire. Mais comparativement aux modèles graphiques et aux analyses combinatoires (cf. par exemple Guénoche, Monjardet, 1987), elle ne peut répondre aux autres exigences.

2.2.2. *Les classifications*

L'objectif est de réduire les observations en classes homogènes. Mais cette méthode est plus adaptée à la réduction des lignes (individus, objets) qu'à celle des colonnes (variables), et par ailleurs, le choix des indices de similarité se révèle assez délicat (Volle, 1993, p.272).

Les modèles graphiques et la méthode combinatoire sont surtout explicatives et permettent d'affiner les résultats obtenus par les méthodes précédentes.

2.3. Les modèles graphiques

Les modèles graphiques permettent de sélectionner un ensemble de variables clé fondées sur certaines des propriétés des modèles graphiques (Whittaker, 1990 ; Fine, 1992 ; Falguerolles et

³ La réduction du nombre de variables peut être obtenue par d'autres méthodes que celle de l'analyse des correspondances multiples des données. Par exemple, par la typologie qui porte cependant davantage sur les individus que sur les variables.

⁴ Les tableaux de petite taille ont des méthodes spécifiques d'analyse, par exemple le khi carré.

Jmel, 1993). Elles diffèrent de l'analyse des correspondances multiples⁵ du fait que la sélection des variables obéit à des critères statistiques issus de la théorie des probabilités (par exemple modèles log-linéaires, théorie de l'information). Mais deux inconvénients majeurs limitent le recours aux seuls graphes.

2.3.1. Leur lisibilité réduite lorsqu'ils sont utilisés directement sur de grands tableaux, ce qui aboutit à ne retenir généralement qu'une dizaine de variables (Falguerolles et Jmel, 1993).

2.3.2. L'impossibilité de comprendre la dynamique inhérente à la structure des variables d'un graphe (leurs relations restent symétriques), du fait qu'ils ne sont pas orientés (Falguerolles et Jmel, 1993).

Les schémas d'implication, dont les plus représentatifs sont l'approche statistique et l'approche booléenne, permettent de dépasser ces limites.

2.4. Les schémas d'implication : si A alors B

Dans un sens strict, toute relation d'implication se traduit par $A \Rightarrow B$ (si A alors B). Elle doit être systématiquement vérifiée sur l'ensemble de l'échantillon. L'implication "non stricte", ou "quasi implication", est une relation dans laquelle l'observation de l'implication d'une variable A sur une variable B est vérifiée uniquement sur une proportion (non précisée) de l'échantillon (Van Buggenhaut, 1987). L'introduction des limites d'un intervalle de confiance à un seuil de probabilité fixé est une autre façon de traiter des implications non strictes.

2.4.1. *Approche booléenne*

L'approche booléenne est classée dans les méthodes combinatoires. Ces dernières relèvent des mathématiques discrètes des structures finies (ordres, graphes, hypergraphes), alors que l'analyse multifactorielle des correspondances repose sur l'algèbre linéaire et la géométrie euclidienne. Dans cet article on s'intéresse principalement aux analyses booléennes des questionnaires (Flament, 1976 ; Van Buggenhaut, 1987).

Il est montré (Flament, 1976) qu'il est possible de caractériser un protocole et d'extraire des schémas d'implication pour la recherche d'un idéal d'une algèbre de Boole et de son algèbre quotient (Van Buggenhaut, 1987). Dans ce cas l'implication stricte est déterminée à l'aide du P.C.U.⁶ et concerne la totalité ou une proportion de la population seulement (Van Buggenhaut, 1987, p.14). Cependant, on ne retient que les P.C.U. dont la longueur est au plus égale à 2. Au-delà, le schéma d'implication est complexe et devient difficilement interprétable.

2.4.2. *Implication statistique*

L'implication statistique repose sur des indices construits à partir des probabilités conditionnelles. Elle est appréciée différemment selon les auteurs.

- Elle est dite "presque satisfaite" lorsque l'indice $H(a, b)$ appartenant à l'intervalle $]-\alpha + 1]$ est proche de 1. En exemple, le tableau suivant croisant la distribution des variables a et b .

⁵ L'analyse des correspondances multiples utilise des critères empiriques pour la sélection de variables à partir de plans factoriels, tels les contributions relatives (COR), ou absolues (CTA), les valeurs propres, les valeurs test. Ces variables sont dites "influentes".

⁶ P.C.U. Projections Canoniques Ultimes

a \ b	1	0	Total
1	n_{11}	n_{10}	n_a
0	n_{01}	n_{00}	$n_{\bar{a}}$
Total	n_b	$n_{\bar{b}}$	N

Tableau 1.

n_{10} est équivalent à $n_{a \wedge \bar{b}}$

L'indice de quasi-implication est donné par la formule suivante :

$$H(a,b) = 1 - n \frac{n_{a \wedge \bar{b}}}{n_a n_{\bar{b}}}$$

(Loevinger, 1947, in Gras et Larher, 1993).

Mais selon Gras et Larher (*idem*) cet indice est invariant⁷ et il ne repose pas sur une échelle de probabilité.

• Gras et Larher proposent quant à eux un indice $[\varphi(a, \bar{b})]$ qui porte sur des variables binaires :

$$\varphi(a, \bar{b}) = \frac{1}{\sqrt{2\Pi}} \int_{q(a, \bar{b})}^{\infty} e^{-t^2/2} dt$$

$$q(a, \bar{b}) = \frac{n_{a \wedge \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

$$\varphi(a, \bar{b}) \geq 0,95 \Leftrightarrow q(a, \bar{b}) \leq -1,65$$

$$\varphi(a, \bar{b}) \geq 0,5 \Leftrightarrow q(a, \bar{b}) \leq 0$$

• L'implication est dite "admissible" lorsque l'indice $q(a, \bar{b})$ est $\leq -1,65$ (pour $\alpha = 0,05$) ou, autrement, lorsque dans une expérience, le nombre d'individus contredisant l'implication est "invraisemblablement" petit par rapport à celui "attendu" dans une hypothèse d'absence de lien implicatif. (Gras et Larher, 1992, p.7)⁸. Une extension de cette implication statistique entre variables numériques ou modales est également possible (Bailleul et Gras, 1994).

En résumé, les limites techniques des études actuelles sur les indices d'implication entre les variables d'un graphe sont les suivantes :

⁷ Indice indépendant de toute dilatation des ensembles de départ.

⁸ La littérature concernant la construction d'indices est abondante. Il n'est pas dans notre propos d'en faire une présentation exhaustive (*cf.* Ganascia, 1991). En résumé, pour des échantillons importants, il est possible d'utiliser l'un ou l'autre de ces indices.

- Lorsque le nombre de variables est élevé, ce qui est très fréquent dans les études empiriques menées dans les sciences humaines, l'examen des variables deux à deux, nécessaire à la construction des schémas d'implication, est difficilement applicable.
- Le caractère multidimensionnel des données⁹ n'est pas pris en compte et il s'avère impossible de construire une configuration.
- Les hypothèses fortes posées sur les lois.

Tout ceci réduit la portée de l'analyse, d'autant qu'aucune méthode ne répond à elle-seule à ces différents critères, ce qui rend le choix exclusif de l'une d'entre elles d'autant plus discutable.

Partant des plans factoriels dans lesquels les proximités entre les variables influentes peuvent être interprétées comme des relations symétriques ($A \Leftrightarrow B$)¹⁰, notre méthode consiste à définir et à construire des schémas d'implication à partir de ces variables. Les proximités entre variables peuvent être recherchées à partir de leur contribution à l'inertie d'un axe factoriel ou bien à partir d'une forme classique de nuages de point (la forme circulaire, la forme tétraèdre, l'effet Guttman, l'effet de taille...). On parle alors de "régionnement du plan".

D'autres méthodes sont possibles pour cette opération. Par exemple la recherche de classes homogènes issues d'une classification automatique des variables.

3. PROPOSITIONS POUR LA RECHERCHE DE SCHÉMAS D'IMPLICATION

3.1. Les étapes

3.1.1. Cette recherche est réalisée sur la base des variables dites " influentes ". Il s'agit des variables qui résument au mieux l'information initiale. Le choix de ces variables est réalisé sur la base des valeurs propres (qui permettent de retenir les axes), et à l'aide de leur contribution relative et de leur contribution absolue à ces axes.

Cette étape préliminaire qui filtre les variables les plus pertinentes résout pour partie la question du choix des variables à examiner, ce que ne peuvent réaliser les méthodes combinatoires. En effet, si celles-ci peuvent, en théorie, tout à fait permettre la sélection des variables, en pratique, elles présentent un certain nombre de limites sur les tableaux de type $I \times J$, lorsque les variables sont numériques¹¹ ; et pour l'interprétation des résultats (cf. 2.4.1).

Ces variables constituent le référentiel dans la recherche d'un schéma d'implication. Elles sont issues d'une même région et sont mises en relation deux à deux à partir des fréquences conditionnelles du tableau de Burt. Ceci correspond à un sous-tableau de Burt apuré.

⁹ Il traite simultanément de l'ensemble des variables, ce qui n'est pris en compte que partiellement dans les analyses réalisées par paire de variables.

¹⁰ En analyse des correspondances multiples, après avoir tenu compte des erreurs de perspective, la proximité observée entre deux variables résulte de profils voisins ou peut être due à la présence d'une ou plusieurs autres variables.

¹¹ Pour le traitement des tableaux de type $K \times K$ (variables ordinales ou quantitatives), les méthodes combinatoires sont les plus efficaces.

3.1.2. Parmi ces variables, seules sont retenues celles dont la fréquence dépasse un certain seuil (cf. exemple 3.2). La comparaison des fréquences conditionnelles est réalisée à partir du test de Mc Nemar sur échantillon apparié¹².

3.1.3. La recherche d'une relation de prédominance ou d'asymétrie est effectuée sur ces variables prises deux à deux. Cette relation est réflexive mais non transitive. Mais sous certaines conditions, il est possible de définir une relation de pré-ordre partiel (Bailleul et Gras, 1994, p. 43).

L'ensemble de cette procédure permet de construire un graphe fini orienté dont les sommets représentent les variables et les arêtes fléchées les relations d'implication entre variables. Celles-ci peuvent être testées avec un critère empirique ou statistique.

3.1.4. Le réseau des relations ainsi construit permet de comprendre la dynamique de l'ensemble des variables à partir de la proximité de celles d'une région et d'expliquer aussi la proximité de deux variables par la présence d'une troisième. En effet, la démarche classique dans ce dernier cas consiste en un retour sur les données initiales (examen des distributions des fréquences des variables). Dépassant cette approche descriptive, l'analyse des fréquences conditionnelles est fondée sur des tests et des critères statistiques qui apportent une réponse plus rigoureuse à la question des proximités entre variables.

3.2. Développement de la méthodologie

Soit deux variables X et Y à deux modalités (respectivement A, B et $1, 2$), retenues selon les critères de l'analyse des correspondances multiples.

Le tableau de Burt de configuration générale extrait du tableau initial est le suivant :

		X		Y	
		A	B	1	2
X	A	1	0	$P_{(1/A)}$	$P_{(2/A)}$
	B	0	1	$P_{(1/B)}$	$P_{(2/B)}$
Y	1	$P_{(A/1)}$	$P_{(B/1)}$	1	0
	2	$P_{(A/2)}$	$P_{(B/2)}$	0	1

Tableau 2. Tableau des profils horizontaux de Burt

Ce tableau peut se lire en lignes ou en colonnes. Il indique les fréquences ou les effectifs entre deux variables croisées.

Il s'agit de proportions conditionnelles.

Lecture : $p_{(A/1)}$ signifie la probabilité d'avoir la modalité A à la question X sachant qu'il a été répondu la modalité une de la question Y ¹³.

¹² Test des hypothèses $H_0 : p_1 = p_2$; et $H_1 : p_1 \neq p_2$. La règle de décision au risque $\alpha = 0,05$. Si le Khi carré de Mc Nemar est supérieur à 3,83, l'hypothèse H_1 est retenue. Sinon H_0 . Le khi carré calculé est donné par la formule $\chi^2 = (n_1 - n_2)^2 / (n_1 + n_2)$ dans laquelle n_1 et n_2 (effectifs) sont définis à partir de la formule du tableau 1.

¹³ La fréquence conditionnelle $p_{j/i} = n_{ij} / n_i$ représente la proportion d'individus qui présentent la modalité x_j sachant y_i .

Exemple

- Si $p_{(A/I)} = 80/100$, cela signifie que I entraîne la présence de A dans 80 cas sur 100.
- Si par ailleurs $p_{(I/A)} = 10/100$, cela signifie que A entraîne la présence de I dans 10 cas sur 100.

LA PRÉSOMPTION D'IMPLICATION VA DONC DANS LE SENS $1 \Rightarrow A$

Cette règle de décision est empirique. Pour des probabilités moins triviales, il est nécessaire de recourir à des tests statistiques. Une illustration en est proposée dans l'exemple suivant.

3.3. Exemple

3.3.1. On considère les données recueillies par questionnaire auprès de 210 sujets par la méthode de l'échantillonnage au hasard stratifié. Sept questions "ouvertes" ou multi-items ont été posées. Leur traitement par l'analyse des correspondances multiples a permis de retenir les modalités significatives qui ont donné lieu à une recodification sous forme de variables dichotomiques. Ici est considéré le sous-ensemble des filles (effectif = 100). Le régionnement des variables sur le plan factoriel a donné quatre régions principales. Les variables présentées sont issues de l'une de ces régions, les variables non influentes (celles dont la CTR^{14} est inférieure à la CTR moyenne - soit $1/\text{nombre de variables}$ -) ne sont pas reportées.

Les variables dichotomiques retenues pour cet exemple sont les suivantes :

- A Projet précis et argumenté (Oui/ Non)
- B Présence d'autres projets par le passé (Oui/ Non)
- C Recherche personnelle d'informations sur le projet professionnel (Oui/ Non)
- D Mariage envisagé (Oui/ Non)
- E Connaissance des conditions matérielles de vie de son père dans sa jeunesse (Oui/ Non)
- F Connaissance d'éléments de la vie affective de sa mère dans sa jeunesse (Oui/ Non)
- G Orientation donnée par la mère au projet de la fille (Oui/ Non)
- H Discrimination des rôles dans la famille (Oui/ Non)

3.3.2. La représentation graphique usuelle permet de relever l'existence de deux sous-groupes (A, B, D et C, E, F, H) et d'une variable plus isolée (G), variables rassemblées par une dimension latente (cf. Figure 1).

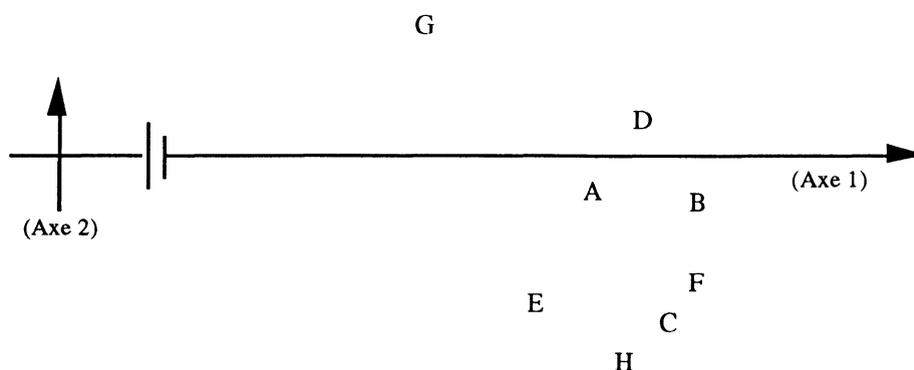


Figure 1. Extrait du plan factoriel

¹⁴ "C'est la contribution d'une variable à l'inertie d'un axe. Elle permet de classer des variables par ordre décroissant selon leur rôle dans l'analyse".

3.3.3. Recherche d'un schéma d'implication entre variables

Soient les variables A, B, D .

	Précision du projet (var. A)		Projets antérieurs (var. B)		Mariage envisagé (var. D)	
	Oui	Non	Oui	Non	Oui	Non
Précision du projet (A)	Oui 60	Non 0	Oui 38	Non 22	Oui 18	Non 42
Projets antérieurs (B)	Oui 38	Non 12	Oui 50	Non 0	Oui 19	Non 31
Mariage envisagé (D)	Oui 18	Non 02	Oui 19	Non 01	Oui 20	Non 0
	Oui 42	Non 38	Oui 31	Non 49	Oui 0	Non 80

Tableau 3. Tableau de Burt (effectifs)

Lecture : cf tableau 1. N.B. Les effectifs observés sont les suivants :
var. A : précision projet de vie : Oui : 60 ; Non : 40.

Recherche des implications entre variables.

1 - Entre var. A et var. B

- Le test de Mc Nemar est égal à : 13,5¹⁵
- Le test de Gras et Lahrer est égal à : - 1,78¹⁶

On en déduit que B implique A

2 - Entre var. D et var. B

- Le test de Mc Nemar est égal à : 2,98
- Le test de Gras et Lahrer est égal à : - 2,84

On en déduit que D implique B .

3 - Entre var. D et var. A

- Le test de Mc Nemar est égal à : 9,6
- Le test de Gras et Lahrer est égal à : - 2,12

On en déduit que D implique A .

La concordance des résultats obtenus par l'indice de Gras et Lahrer et le khi carré permet d'affirmer l'existence d'une relation d'implication entre les variables A et D , et A et B dans laquelle on observe $D \Rightarrow A$ et $B \Rightarrow A$, ce que résume le graphe suivant :

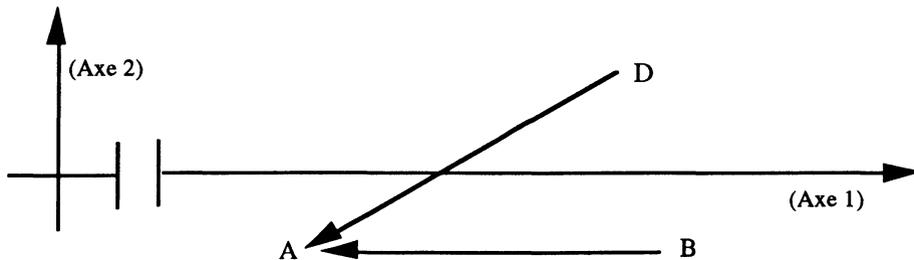


Figure 2

¹⁵ La formule numérique développée est : $[(38 - 12)^2 / 50]$.

¹⁶ La formule numérique développée est $[(12 - (50 \times 40) / 100) / 4,47]$. Cette implication est admissible au niveau de confiance 96,2%.

Cette procédure appliquée à l'ensemble des variables choisies du plan d'une région considérée permet de construire le graphe suivant.

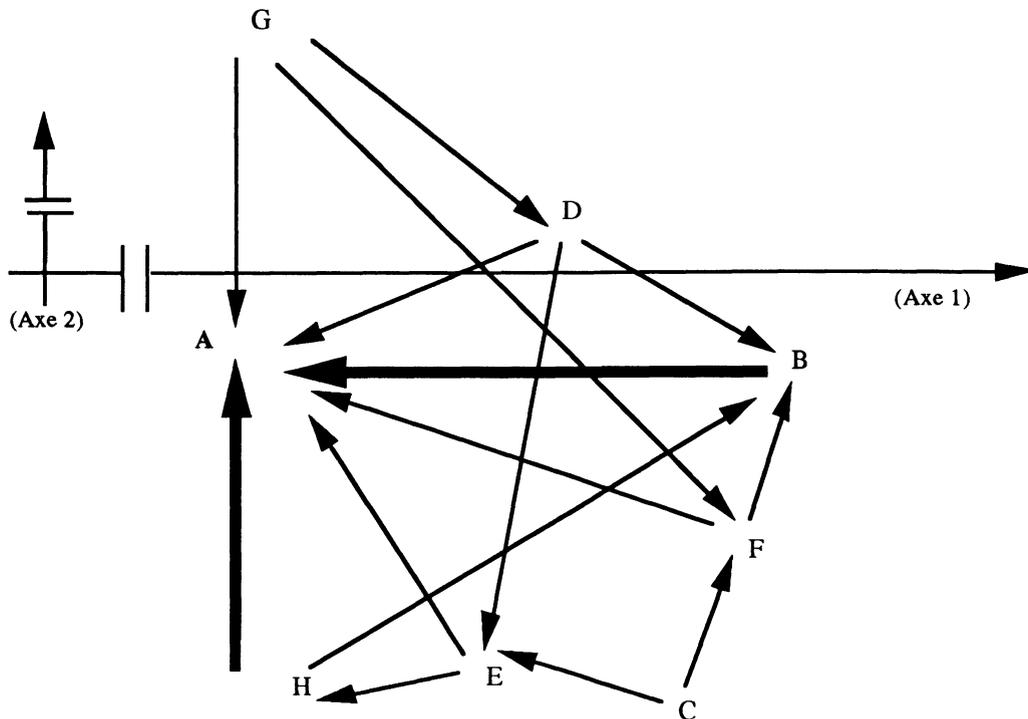


Figure 3. Les variables considérées sont dichotomiques.
(Ne sont reportées que les variables qui satisfont aux critères *COR* et *CTR*).

3.3.4. De cette représentation graphique il résulte que la précision du projet (A) résulte plutôt des influences exercées par l'ensemble des autres variables et non pas uniquement de celles du sous-groupe de A (soit B et D sur le plan factoriel (cf. figures 1 et 2)).

Elle permet donc, tout en conservant tous les avantages, de dépasser la lecture usuelle basée sur les seuls critères d'aide à l'interprétation (*COR*¹⁷, *CTR*, axes factoriels, valeurs test...).

3.3.5. Le réseau de relations entre variables peut être soumis à une analyse factorielle confirmatoire de type "méthode causale" (par exemple, *in* Bagozzi, 1980).

4. INTERPRÉTATION DU SCHÉMA D'IMPLICATION

La lecture du schéma d'implication permet d'apporter une réponse plus précise à la question des rapports que des éléments psychiques peuvent entretenir. Les relations entre variables de cet ordre sont sous-tendues par la présence de l'énergie psychique mobilisée par cette situation. Elle est analysable selon trois indices (direction d'influence, fonction et intensité) issus directement des schémas d'implication, et d'un quatrième (masse) déjà présent dans l'analyse des correspondances multiples.

¹⁷ C'est le cosinus carré α , dans lequel α est l'angle formé par une variable avec un axe. C'est un indicateur de la qualité de la représentation d'une variable par sa projection sur l'axe (la qualité de la représentation d'une variable est considérée comme "bonne" lorsque *COR* est proche de 1).

4.1. La direction des influences

Les arêtes du graphe donnent les grandes voies de la circulation de l'énergie. Cette lecture directe fournit une information précieuse qui donne du sens à des dimensions psychologiques d'autant qu'elles s'inscrivent dans des mouvements de changement, de transformation. Cette lecture peut être facilitée par l'indice développé au paragraphe suivant.

4.2. La fonction¹⁸ d'une variable dans un graphe

Trois fonctions directement issues des schémas d'implication sont proposées : émetteur, synapse, récepteur.

L'attribution d'une fonction à une variable clarifie le sens à donner aux rapports de double détermination qu'entretiennent les composantes de la vie psychique. Deux de ces fonctions émanent directement de la théorie de la communication : celle d'*émetteur*, définie comme la qualité d'une variable impliquant au moins la présence d'une autre variable sans être impliquée par quelque variable que ce soit (dans l'exemple, les variables *G* et *C*) ; celle de *récepteur*, définie comme la qualité d'une variable impliquée par une ou plusieurs autres variables sans qu'elle n'implique elle-même la présence d'une ou plusieurs autres variables (dans l'exemple, la variable *A*).

Ce schéma de lecture des graphes a été complété par l'adjonction d'une troisième fonction appuyée sur la notion de *synapse* dont la caractéristique majeure est de posséder la double fonction d'émetteur et de récepteur (dans l'exemple les variables *D*, *B*, *F*, *E*, *H* ont la fonction synapse). Outre leur fonction de "relais" d'énergie psychique entre deux (ou plus) autres variables, ces variables traitent de l'"intermédiaire" et pourraient avoir pour propriété de transformer l'énergie psychique afin de la rendre opératoire auprès de certaines autres variables. Ainsi dans l'exemple, la présence d'un projet précis (*A*) résulte de l'influence exercée directement par toutes les variables (sauf *C*). Mais l'influence de certaines d'entre elles n'est pas uniquement directe et est renforcée par l'intermédiaire d'autres variables (par exemple l'influence de *D* sur *A* est renforcée par l'intermédiaire de *E* et de *B*). Les fonctions attachées aux variables permettent également de moduler le sens à donner à un jeu d'influences s'exerçant sur une variable (ayant elle-même fonction de récepteur ou de synapse), selon la qualité de fonction des variables qui l'influencent (émetteur ou synapse) et le nombre de variables ayant ce rôle. Il est en effet probable que l'implication ne peut avoir la même signification selon qu'elle est le fait d'une ou de plusieurs variables, que dans ce dernier cas il y ait pluralité ou non de fonctions (émetteur ou synapse, compte-tenu du point précédent). Les fonctions facilitent la lecture du graphe. Par exemple, le point de départ peut être constitué par les variables "émetteur", puis se poursuivre par celles de "synapse", enfin par celle(s) de "récepteur" (dans l'exemple : $G \rightarrow A$; $G \rightarrow D \rightarrow A$; ou $\rightarrow B$; $G \rightarrow F \dots$).

Ces fonctions peuvent être reportées sur le graphe selon un jeu de symboles.

Cette procédure peut être appliquée à n'importe quelle variable dichotomique d'un graphe construit dans ces circonstances et est enrichie par l'indice développé au paragraphe suivant.

4.3. L'intensité de l'implication (ou pouvoir moteur)

L'intensité de l'implication est donnée par le Khi carré de Mc Nemar et l'indice de Gras et Larher.

¹⁸ Le terme de "fonction" n'est pas à prendre dans un sens mathématique. Il s'agit du mode de relation d'une variable avec l'ensemble des autres variables d'un même graphe.

Elle donne une visibilité à la puissance de la réalisation d'un événement. Par exemple, lorsque le *Khi carré* de Mc Nemar est plus élevé pour une variable comparativement à une autre, cela signifie que sa présence est davantage "expliquée" par cette variable que par celle des autres. Un jeu de traits (pointillés, ou traits pleins par exemple) reportés sur le graphe peut traduire cette intensité. Dans l'exemple développé, la puissance des variables *B* et *H* s'exerçant sur *A* est supérieure à celle des autres, ce qui module l'affirmation selon laquelle toutes les variables (à l'exception de *C*) influencent la présence de *A*. En effet, les variables *D*, *E*, *F*, *G* ont une puissance motrice sensiblement égale entre elles mais inférieure à celle de *B* et *H*. Donc, une hiérarchisation du jeu des influences est possible et peut être combinée avec les autres indices. Dans l'exemple, elle s'établit comme suit : *B*, *H* puis *D*, *E*, *F*, *G*.

4.4. Le poids de chaque variable

Le poids de chaque variable a la même signification que dans l'analyse des correspondances multiples qui le fournit (hiérarchisation des variables). Il est donc proportionnel à l'effectif des sujets qui, dans l'exemple, ont répondu positivement aux questions. Il peut être d'autant mieux utilisé pour la compréhension de l'ensemble qu'il est alors rattaché à une "fonction". Celle-ci sera donc d'autant plus importante que son inertie sera élevée (et inversement). Ce qui peut être noté sur le graphique par un jeu de symboles et guider lui-aussi la lecture, en la hiérarchisant selon l'importance de l'émetteur par exemple, et - ou - du récepteur.

5. CONCLUSION

La méthodologie proposée qui combine les avantages de différentes approches permet de répondre pour partie à la problématique de l'analyse des structures des relations entre variables. Ses avantages sont très nettement situés : - dans le choix des variables à partir de l'analyse des correspondances multiples ; - dans les cartes causales établies ; et - dans les indices qui les accompagnent. Les relations entre des variables deux à deux doivent être interprétées comme des hypothèses sur la direction *dominante* des influences entre elles, compte-tenu des précautions courantes.

Des extensions de cette méthodologie sont possibles à des tableaux binaires de type $P_{ijk} = Pr(X = i, Y = j, Z = k)$ lorsque les effectifs des cases le permettent ($n_{ijk} > 5$), et sur des variables numériques et modales.

La méthodologie proposée rend lisibles diverses manifestations de l'énergie psychique entre variables nominales. Le relief évident qui est ainsi restitué aux phénomènes psychiques ouvre sur des perspectives intéressantes découlant des fonctions et des critères subséquents qui peuvent sans doute être généralisées à de nombreuses autres situations. Mais ces nouvelles pistes offertes par l'outil statistique ne constituent qu'une aide à la décision que seuls le chercheur et le praticien sont à même de prendre du fait de leur expérience, et dont ils ne doivent pas se désapproprier.

BIBLIOGRAPHIE

- AMARGER S., DUBOIS D., PRADE H., *Imprecise quantifiers and conditional probabilities*, in : *Symbolic and quantitative approaches to uncertainty* (Kruse R., Siegel P.), Springer-Verlag, 33-37.
- ARABIE P., BOORMAN A., LEVITT P., "Constructing block models : how and why", *Journal of Mathematical Psychology*, 17, 1978, 21-63.
- BACHER F., *Les enquêtes en psychologie*, T. 2, Lille, Presses Universitaires de Lille, 1982.
- BAGOZZI R.P., *Causal models in marketing*, New York, Wiley, 1980.
- BAILLEUL M., GRAS R., "L'implication statistique entre variables modales", *Mathématiques, Informatique et Sciences humaines*, 1994, 128, 41-57.
- BENZECRI J.P., *Correspondence Analysis Handbook*, New York, Dether, 1992.
- CAZES P., BAUMEDER A., BONNEFOUS S., PAGES J-P., "Codage et analyse de tableaux logiques ; introduction à la pratique des variables qualitatives", *Cahiers du BURU, Série Recherche*, 1977, 27.
- CHANDON J.-L., PINSON S., *Analyse typologique: théories et applications*, Paris, Masson, 1981.
- DER MEGREDITCHIAN G., *Le traitement statistique des données multidimensionnelles*, T.1, Paris, École Nationale de Météorologie, 1988, (in P. Pottier, 1994).
- FAVERGE J.M., *Méthodes statistiques en psychologie appliquée*, Paris, Presses Universitaires de France, 1965.
- FALGUEROLLES A. de, JMEL S., "Un modèle graphique pour la sélection de variables qualitatives", *Revue de Statistique appliquée*, 1993, XLI, 1, 23-41.
- FALGUEROLLES A. de, JMEL S., "Modèles graphiques gaussiens et analyse en composantes principales : complémentarité et choix de variables", *Publication du Laboratoire de statistique et probabilités*, Toulouse, 1992, 3, 1-25.
- FINE J., *Modèles graphiques d'association*, dans : *modèles pour l'analyse des données multidimensionnelles*, Paris, Economica, 1992, 273-313.
- FLAMENT C., *L'analyse booléenne de questionnaires*, Paris, Mouton, 1976.
- FREUD S., *Métapsychologie*, Paris, Gallimard, 1968.
- GANASCIA J.G., "CHARADE : Apprentissages de bases de connaissances," in *Introduction symbolique-numérique à partir de données*, Cepadues, 1991.
- GIBBONS J.D., *Nonparametric Measures of Associations*, London, New Delhi, Newberry Park, University paper, 1993.
- GOODMAN L.A., KRUSKAL W.H., *Measures of association for cross classifications*, New York. SpringerVerlag, 1979.
- GRAS R., LARHER A., "L'implication statistique, une nouvelle méthode d'analyse de données", *Mathématiques, Informatique et Sciences humaines*, 1992, 120, 5-31.
- GUÉNOCHE A., MONJARDET B., *Méthodes ordinales et combinatoires en analyse des données*, *Mathématiques et Sciences humaines*, 1987, 100, 5-47.
- HERMAN J., *Analyse des données qualitatives : traitement d'enquêtes, modèles multivariés*, T. 2, Paris, Masson, 1986.
- LERMAN I.C., GRAS R., ROSTAM H., "Elaboration et évaluation d'un indice d'implication pour des données binaires, I et II", *Mathématiques et Sciences humaines*, 1981, 74, 5-35 et 75, 5-47.
- LONG J.S., *Covariance structure models: an introduction to LISRE*, Sage University paper 34.
- POTTIER P., "Mesure de la liaison entre deux variables qualitatives: relations entre un coefficient de corrélation généralisé et le khi carré", *Revue de Statistique appliquée*, 1994, XLII, 1, 41-62.
- REYNOLDS H.T., *Analysis of nominal data*, London, New Dehli, Beverly Hills, Stages publications, 1984.

- RIARD E.H., *L'adolescent et sa famille: la genèse du projet de vie* (étude à partir d'une population scolarisée des deux sexes de 14-15 ans), Thèse de Doctorat d'Etat ès Lettres et Sciences humaines, Université Paris-X Nanterre, 1991.
- SAPORTA G., *Probabilités, analyse de données et statistique*, Paris, Éditions Technip, 1990.
- SPRENT P., *Applied non-parametric Statistical methods*, London, New York, Tokyo, Melbourne, Madras, Chapman and Hall, 1990.
- VAN BUGGENHAUT, J., "Questionnaires booléens : schémas d'implications et degrés de cohésion". *Mathématiques et Sciences humaines*, 1987, 98, 9-20.
- VOLLE M., *Analyse des données*, Paris, Economica, 1993.
- WHITTAKER J., *Graphical models in applied multivariate statistics*, New York, Wiley, 1990.
- WORSLEY K. J., "Un exemple d'identification d'un modèle log-linéaire grâce à une analyse des correspondances", *Revue de Statistique appliquée*, 1987, 35, 13-20.