



**HAL**  
open science

# Traitement de données structurées par un ordre temporel ou hiérarchique : utilisation dans l'analyse de l'activité

Claudette Mariné, Marie-Françoise Valax, Max Reinert

## ► To cite this version:

Claudette Mariné, Marie-Françoise Valax, Max Reinert. Traitement de données structurées par un ordre temporel ou hiérarchique : utilisation dans l'analyse de l'activité. *Le travail humain*, 1990, 53 (1), pp.79-89. hal-02464384

**HAL Id: hal-02464384**

**<https://hal.science/hal-02464384>**

Submitted on 7 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# TRAITEMENT DE DONNEES STRUCTUREES PAR UN ORDRE TEMPOREL OU HIERARCHIQUE : UTILISATION DANS L'ANALYSE DE L'ACTIVITE

VALAX Marie-Françoise, MARINE Claudette, REINERT Max\*

\*URA-CNRS 259, Univ. Toulouse-le Mirail, 5 Allées Antonio Machado, 31058  
Toulouse Cedex

*Abstract : Processing of data framed by a temporal or hierarchical order :  
application to behaviour analysis.*

This paper presents a data analysis method that allows to draw from behaviour modalities of mental activity structures, by comparing typical action or knowledge structures between different groups of subjects. The interest of the Hierarchical Descendant Classification (HDC) is shown by two studies, the first one relating to temporal order observed in procedures, the other one to hierarchical order on decision making

*Key words : Cognitive organization, Data analysis, Graph, Temporal order,  
Hierarchical order*

*Résumé :*

Cette note présente une méthode d'analyse des données permettant, à partir d'observables comportementaux, de dégager des modalités d'organisation de la conduite, et de comparer des structures d'actions et/ou de connaissances caractéristiques de différents groupes de sujets. Les intérêts de la méthode de Classification Descendante Hiérarchique (CDH) sont illustrés à travers deux études, l'une relative à l'organisation temporelle de l'action, l'autre à l'organisation hiérarchique de critères de décision.

*Mots clés : Organisation cognitive, Analyse des données, Graphe, Ordre temporel,  
Ordre hiérarchique*

Dès que l'on s'intéresse à des conduites de travail cognitivement complexes, l'accent est mis sur l'activité de structuration qui organise des unités de connaissances. Ainsi, de Montmollin (1984) définit la compétence comme des "ensembles stabilisés de savoirs et de savoir-faire, de conduites types, de procédures standards, de types de raisonnements, ..." (p.122). On retrouve cette idée dans le modèle de Rasmussen (1983) où apparaissent trois niveaux de structuration qui, différenciés par la complexité et la flexibilité, sont déterminés par le niveau de familiarité du sujet avec la tâche (habiletés, règles, connaissances). Ainsi, la compréhension des conduites de travail doit passer par l'analyse de l'architecture et du contenu de ces unités organisées de connaissances, de leur construction et de leur actualisation dans la réalisation de la tâche (Leplat, 1988).

Cependant, si ce principe est largement accepté -tant en psychologie du travail qu'en psychologie cognitive-, le problème de leur extraction reste posé. En se référant au modèle de l'action proposé par von Cranach (1982), on peut dire que l'objectif consiste à rechercher, derrière le foisonnement des observables, des régularités dans les comportements. Ces dernières permettent d'inférer l'organisation des buts et sous-buts, les plans (Hoc, 1987), les représentations du problème (Richard, 1988), les stratégies de résolution (George, 1988) qui guident et contrôlent le choix et le déroulement de l'action. Ces régularités doivent se manifester par un agencement typique de comportements (prises d'information, actions, communications, décisions) qu'un sujet ou un groupe de sujets itère au cours de l'exécution d'une tâche et/ou transfère d'une tâche à une autre. De même, dans une approche comparative, on doit pouvoir identifier des régularités différenciant des groupes de sujets sur une ou plusieurs dimensions comportementales.

Etant donné la diversité des conduites de travail, la saisie de ces régularités impose généralement des analyses complexes sur les observables recueillis, analyses difficiles à réaliser avec les outils actuellement disponibles. Dans cette perspective, cet article présente une méthode d'"analyse des données" illustrée à travers deux applications : la première dégage des règles fonctionnelles par une analyse de la succession observable des actions dans le temps, la seconde met en évidence une hiérarchie interne de critères inférée à partir d'un ensemble de choix.

## 1 - L'OUTIL D'ANALYSE

Il est usuel de contrôler l'existence de liens hypothétiques entre variables dépendantes et indépendantes à l'aide d'outils statistiques. Cependant, lorsque la variable dépendante n'est pas directement observable et que l'on doit l'appréhender à l'aide d'un ensemble d'indicateurs, l'analyse de données peut s'avérer utile pour réduire cet ensemble d'indicateurs à quelques traits caractéristiques (facteurs ou classes). Une fois cette opération effectuée on peut étudier les relations entre ces traits et les modalités des variables indépendantes en recourant à des analyses classiques.

Dans le cas présent, les données sont sous forme de couples ordonnés, soit par une succession temporelle, soit par une hiérarchie des choix associés à chaque individu. L'analyse consiste à rechercher des groupements de sujets associés à des graphes typiques.

L'étude des graphes est réalisée généralement (Belley et al., 1981) à partir des matrices de transition entre items (par exemple les actions successives) calculées sur la population totale. Cependant, exprimant des moyennes, ces transitions posent des problèmes lorsque les données sont issues de populations hétérogènes. La solution proposée consiste à analyser non pas les matrices de transitions, mais les couples d'items relevés pour chaque sujet à partir desquels peuvent être construites les familles de graphes. Cela revient à analyser un tableau ayant en colonne les couples observés et en ligne les sujets (Reinert, 1986). Ce tableau a la particularité de comporter de nombreux zéros. Pour son analyse nous utilisons un programme spécifique de Classification Descendante Hiérarchique (CDH) fondée sur la métrique du  $\chi^2$ . Il permet d'obtenir une hiérarchie de partitions sur les individus. Chaque classe obtenue est ensuite décrite sur la base des couples les plus significatifs de la classe. Il est alors possible de reconstruire le graphe exprimant la structure séquentielle des états typiques de ces classes (voir annexe).

## 2 - APPLICATION SUR DES ORDRES TEMPORELS

Nous présentons dans ce paragraphe une analyse de la structure temporelle de l'action réalisée dans le cadre plus général d'une étude conjointe de la représentation et de l'organisation de l'action en fonction de l'expertise (Bernard et Parise, 1987).

## 2.1 - Situation empirique et protocole

L'étude a porté sur la comparaison de sculpteurs novices (5 sujets en apprentissage) et experts (5 professeurs de sculpture). La tâche consistait à reproduire un visage en terre par modelage.

Au plan de la technique de sculpture, la réalisation du modèle implique un travail cyclique progressif et homogène sur la matière. Le fait que les experts ont intériorisé cette condition devrait se manifester par l'existence de séquences d'actions structurées. Ces séquences désignent un ensemble d'actions ordonnées entre les différentes parties du modèle, qui apparaissent fréquemment au cours de la réalisation de la tâche.

Pour codifier le recueil des données, les différentes parties du modèle ont été désignées par une lettre (fig. 1). L'observateur affectait un numéro d'ordre à chaque partie travaillée afin de relever la chronologie des actions.

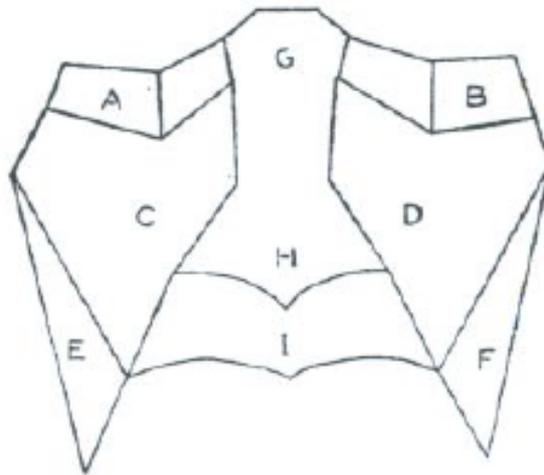


Figure 1- Représentation schématique du modèle à reproduire

*Schematic format of the model*

Pour le traitement, les données ont été saisies par couples d'actions successives (une action sur la partie A du modèle suivie d'une action sur la partie B est codée AB), en découpant chaque protocole en séquences de 10 mn. Le choix de ce découpage correspond à notre objectif de faire émerger des structures d'actions localisées dans le temps permettant de faire apparaître leur évolution éventuelle en cours d'exécution.

## 2.2 - Présentation des résultats

### 2.2.1- Partition des structures de données

Sur un total de 40 séquences novices et expertes, on obtient les partitions suivantes :

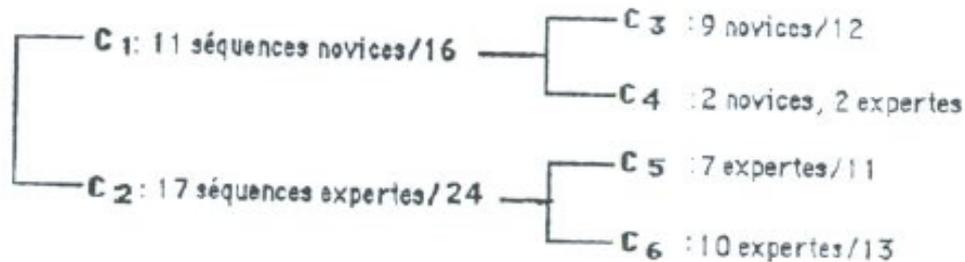


Fig. 2 - Partitions effectuées par la CDH

*Partitions made by HDC*

La première partition oppose des séquences issues majoritairement des protocoles des novices (sur 16 séquences de la classe 1, 11 sont des séquences novices) à des séquences plutôt expertes (sur 24 séquences de la classe 2, 17 sont des séquences expertes).

La seconde partition fait apparaître deux classes non spécifiques d'un niveau d'expertise (classes 4 et 5) et deux classes spécifiques (classes 3 et 6).

### Graphe des relations

Tableau 1 — Couples obtenus dans la classe 1

Identification des couples	Effectif dans la classe 1	Effectif total du couple	Pourcentage	$\chi^2$
AC	8	14	57,14	2,64
BH	4	4	100	6,67
CA	9	13	69,23	6,86
CE	11	18	61,11	6,06
DE	5	6	83,33	5,52
EC	13	14	92,86	5,07
GD	6	8	75,00	5,10
Débutants	11	18	61,11	6,08

Pour décrire le contenu des différentes classes de séquences, on examine les couples les plus significativement présents dans ces classes. De même, il est possible de savoir d'où viennent ces séquences : ainsi, parmi les 18 séquences débutants répertoriées au départ, 11 sont incluses dans la classe 1 (tableau 1).

Pour illustrer l'obtention du  $\chi^2$ , prenons l'exemple du couple AC :

		Classe 1		
		Présent	Absent	
AC	Présent	<b>8</b>	6	14
	Absent	8	<b>18</b>	26
			16	24

**En gras : valeur donnée**

Standard : Valeurs déduites

$\chi^2 = 2,64$

Notons qu'ici le seuil de signification du  $\chi^2$  est bas, ceci est lié d'une part à la taille restreinte de l'effectif, et d'autre part, au fait que, dans ce type d'analyse, ce n'est pas la signification d'un couple unique qui est intéressante, mais la signification de la structure d'ensemble des couples.

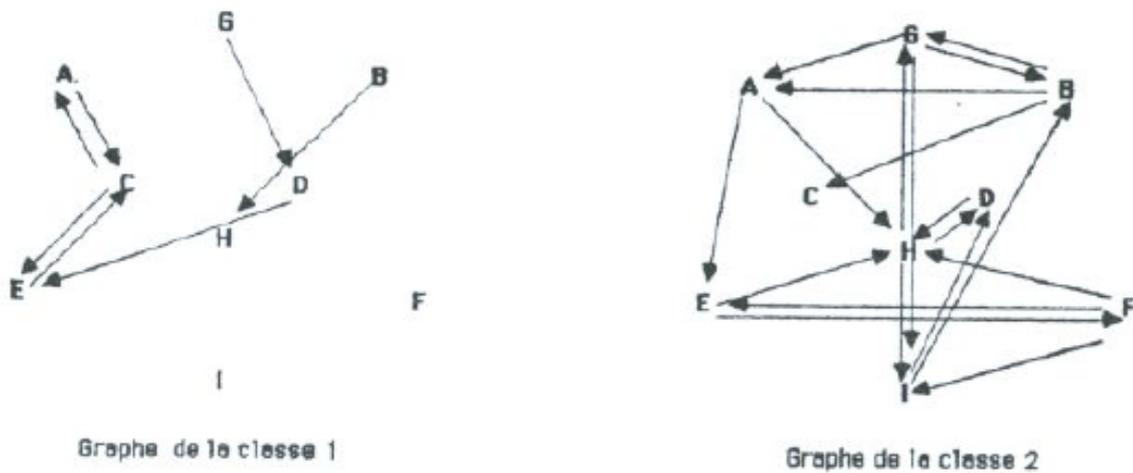


Fig. 3 - Graphes des relations sous la première partition.

*Graphs of relations under the first partition*

A partir de ces données, on peut représenter graphiquement les couples d'action significatifs sur la base de la structure du modèle initial à effectuer par les sujets. La correspondance spatiale de la représentation graphique et de la structure d'action à analyser facilite une lecture psychologiquement signifiante de l'image.

Ainsi, en comparant les graphes correspondant aux classes 1 et 2 (fig. 3), on peut saisir plusieurs différences quant à la structuration de l'action en fonction de l'expertise.

L'aspect le plus manifeste est relatif à la richesse des liaisons observées. Alors que le graphe 1 comporte très peu de relations systématiques et des parties isolées, le graphe 2 se caractérise par de nombreuses relations entre toutes les parties du modèle. (L'absence de relations entre parties ne signifie pas que celles-ci ne sont pas travaillées, mais que les enchaînements d'une partie à l'autre du modèle ne sont pas itérés au cours de la réalisation).

Contrairement au graphe 1, le graphe 2 reflète une structure de l'action organisée autour d'un axe central G-H-I -axe de symétrie du modèle- avec H comme point central. On observe également un travail alterné entre les parties droite et gauche du modèle, soit directement (B-A, B-C, E-F, F-E), soit par l'intermédiaire de l'axe central.

Enfin, le caractère cyclique de l'organisation de l'action se manifeste par l'existence, dans le graphe 2, de très nombreuses boucles prenant appui sur l'axe central (G-B-G, H-D-H, E-F-E, G-A-H-G, D-H-I-D, G-I-B-G, G-A-E-H-G), alors qu'on n'en observe que deux (A-C-A et E-C-E) sur la partie gauche du graphe 1.

### **3 - APPLICATION SUR DES ORDRES HIERARCHIQUES**

Dans un travail visant à étudier la fonctionnalité des structures de décision (Valax et al, 1989), nous nous sommes intéressés à la variation des hiérarchies de critères en fonction de la place de l'opérateur dans le système de décision. L'étude qui porte sur une tâche d'ordonnancement concerne des opérateurs issus de deux services complémentaires de l'entreprise :

- le planning qui gère le lancement et le suivi de fabrication,
- la fabrication qui réalise les opérations prescrites par le service planning.

#### **3.1 - Dispositif expérimental**

Les sujets ont individuellement classé 14 critères d'ordonnancement apparus lors d'entretiens et de réunions effectués dans les deux services. Ces critères sont référés à deux objectifs distincts : la gestion des délais (10 items, tels que : Durée des opérations, Délais de livraison, Diminution des temps d'attentes, ...), la gestion des ressources (4 items : Occupation des machines, Occupation des hommes,

Couplage homme/machine, Qualité des pièces). Ces items sont notés D1 à D10, pour la gestion des délais, et R1 à R4 pour la gestion des ressources.

Pour éviter les biais dus à une présentation globale (effets de mémorisation, de perception, etc...) le classement des critères est déduit d'une épreuve de comparaisons par paires dans laquelle chaque paire est présentée deux fois au sujet en inversant l'ordre. A chaque paire, le sujet devait indiquer le critère considéré comme plus important pour l'ordonnement. Chaque opérateur a ainsi jugé 182 paires.

L'épreuve concerne 30 opérateurs, 14 sont issus du service planning, 16 du service fabrication. Dans ce dernier service, nous avons considéré deux ateliers distincts quant au nombre de pièces à gérer par unité de temps : 6 opérateurs viennent de l'atelier montrant le taux le plus faible, 10 de l'atelier montrant le taux le plus fort.

Les données ont été saisies sous forme de couples. Pour chacun d'entre-eux, ont été considérés la nature des critères et leur ordre hiérarchique. Par ailleurs, dans l'objectif de tester la cohérence globale des réponses par rapport à une structure d'ordre hypothétique à révéler, nous avons construit une population fictive "d'anti-sujets". (Etant donné un sujet  $i$ , nous avons créé son contraire  $i'$ : si un sujet  $i$  répond BA à la paire A-B, on crée un anti-sujet  $i'$  (fictif) qui répond AB à cette même paire). Une forte cohérence des réponses doit se manifester dans les résultats de la CDH par une différenciation marquée entre sujets et anti-sujets. La construction de ces anti-sujets est donc un moyen de tester cette cohérence, sans faire d'hypothèses sur la nature explicite de la structure d'ordre.

## **3.2 - Analyse des résultats**

### ***3.2.1-Partition des structures de données***

Une première partition oppose sujets et anti-sujets, une seconde les services planning et fabrication. La variable caractéristique de l'atelier ne détermine pas de structures différenciées.

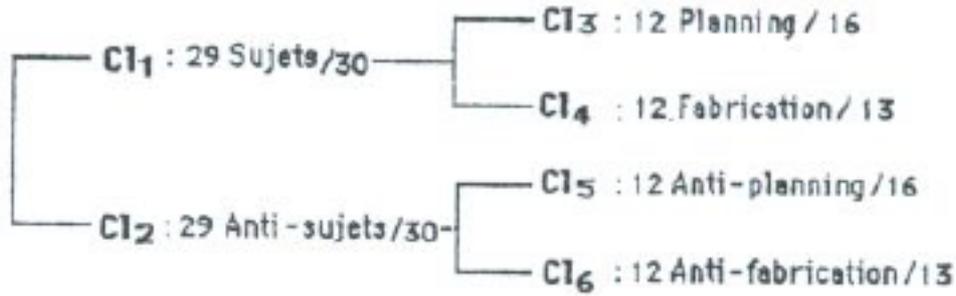


Fig.4 - Partitions effectuées par la CDH

*Partitions made by the HDC*

Les structures de données des sujets (C1) et anti-sujets (C2) s'opposent très nettement. Un seul sujet (appartenant à la fonction planning) est classé "anti-sujet".

La deuxième partition oppose planning et fabrication. La classe 3 comprend 16 sujets, dont 12 appartiennent au service planning. La classe 4 regroupe 13 sujets, dont 12 sont issus du service fabrication.

Ces résultats indiquent que la population globale présente une unité dans ses choix, et par extension qu'il existe, dans l'entreprise considérée, une structure décisionnelle d'ordonnancement type. Cependant, cette dernière peut se diviser en deux sous-structures, chacune typique d'un service particulier.

### **3.2.2-Grappe des relations**

Le graphe est construit dans ses grandes lignes comme le précédent, le changement essentiel vient de la position relative des items. Cette position, qui est définie dans la première étude par les relations spatiales entre les éléments du visage, est déterminée ici de façon plus arbitraire. Sur l'axe horizontal la position d'un item est fonction du nombre de liens qu'il génère. Sur l'axe vertical, cette position dépend uniquement de la lisibilité du graphe.

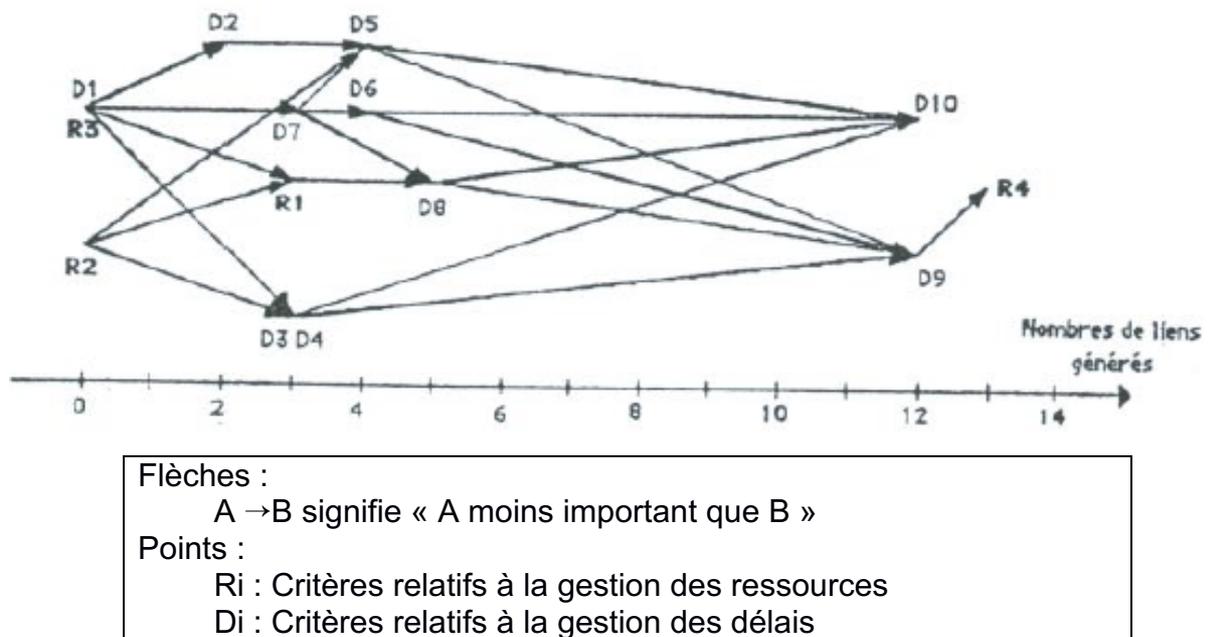


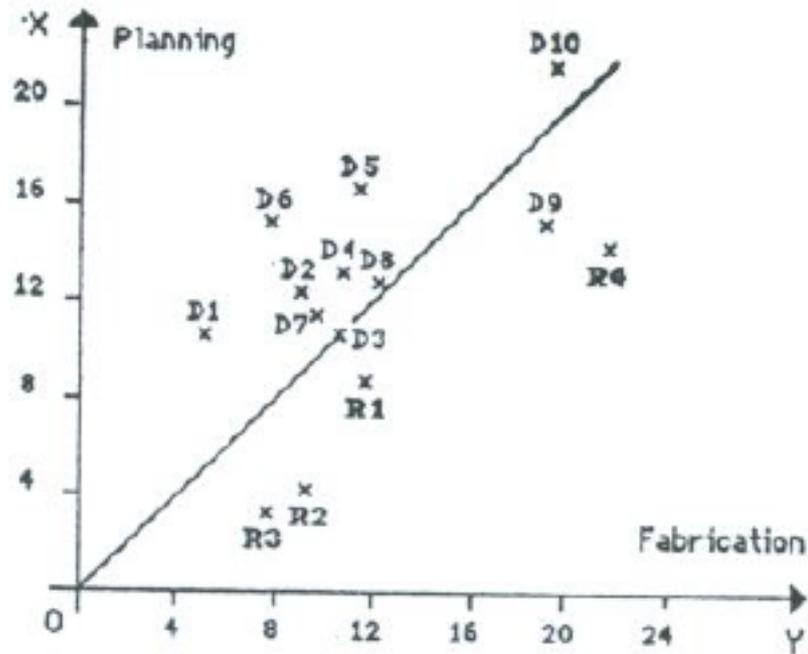
Figure 5 - Graphe représentant les hiérarchies alternatives types de l'échantillon global  
*Graph mapping the standard alternative hierarchies of the overall sample.*

Le graphe (fig 5) présente de multiples branches, mais aucune boucle. Ainsi, on peut dire que si la structure décisionnelle type n'est pas strictement linéaire (diverses alternatives) la transitivité des choix est respectée.

Sur l'axe horizontal, qui représente l'ordre hiérarchique des critères, deux groupes de critères se distinguent très nettement. Sans entrer dans les détails, on constate l'importance accordée à la gestion des délais (D1 à D10) par rapport à la gestion des ressources (R1 à R4).

Cependant, la complexité de ce type de graphe (hiérarchies alternatives) rend difficile des comparaisons similaires à celles du paragraphe 2. Ainsi, pour faciliter ce type d'analyse, il peut être utile de recourir à des représentations linéaires, fondées sur les seuls niveaux hiérarchiques des critères -représentations qu'autorise l'absence de boucle dans les graphes complexes-.

Dans cette perspective, nous avons travaillé sur les fréquences moyennes des couples initiés par A ou B ou C ou etc. Notons que chaque paire étant présentée deux fois à un sujet donné, pour ce sujet, la fréquence d'un critère varie entre 0 et 26.



<p><b>Axes</b>                  Sur OX : Fréquence moyenne des choix du service planning                  Sur OY : Fréquence moyenne des choix du service fabrication</p> <p><b>Points :</b>                  Ri : Critères relatifs à la gestion des ressources                  Di : Critères relatifs à la gestion des délais</p>
--

Figure 6 -Comparaison de la fréquence moyenne des choix entre planning et fabrication.

*Comparison of the average frequency of choices between planning and manufacturing.*

Ainsi, sur la figure 6, qui met en évidence les valeurs relatives accordées aux délais (D) ou aux ressources (R) par chacun des services, on remarque le privilège accordé par le planning à la gestion des délais, et réciproquement le privilège accordé par la fabrication à la gestion des ressources.

*En conclusion*, si cette méthode n'implique pas des prédictions quant à la nature de la structure recherchée, elle suppose néanmoins des hypothèses pour le guidage du recueil d'observables pertinents. Par rapport aux statistiques traditionnelles, elle permet, d'une part, de prendre en compte simultanément un ensemble d'observables relativement important, et d'autre part, de révéler le poids

d'indicateurs qui pourrait être sous-estimé dans une approche strictement déductive. Elle s'avère particulièrement pertinente dans l'étude exploratoire du fonctionnement cognitif (Vermersch, 1984) dans la mesure où elle permet de traiter le caractère ordonné de la connaissance et de l'action. De plus, la possibilité de représentation graphique, en visualisant un ensemble de relations complexes, facilite l'interprétation des observables.

Les résultats des études comparatives présentées ici, laissent entrevoir une utilisation plus large de ce type d'outils, en particulier pour l'analyse de l'évolution des conduites complexes que l'on peut rencontrer dans les situations de travail et de formation.

## BIBLIOGRAPHIE

- BELLEY, R., POULI-DUBOIS, D., RICARD, M. (1981) - Rapport sur différentes méthodes d'analyse de séquences de comportements. Polycopié, Université de Montréal, 42p., Juin 1981.
- BENZECRI, J.P., et coll (1973) - L'analyse des données. Tome 2 : L'analyse des correspondances . Paris, Dunod, 616p.
- BERNARD, N., PARISE, B. (1987)- Le rôle de l'expertise dans la représentation mentale, Mémoire de maîtrise, Université Toulouse-le Mirail, 96p.
- von CRANACH, M. (1982). - The psychological study of goal-directed action: basic issues, in M. von Cranach, R. Harré (Eds), The analysis of action, Londres, Cambridge University Press, 35-73.
- GEORGE, C. (1988). - Interactions entre les connaissances déclaratives et procédurales, in P. Perruchet (Ed), Les automatismes cognitifs, Liège, Mardaga, 101-137.
- HILL, M.O. (1979) - TWINSPAN: A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Ecology and Systematics, Cornell University, Ithaca, New-York, USA
- HOC, J-M. (1987). - Psychologie cognitive de la planification, Grenoble, P.U.G., 180 pages.
- LEPLAT, J. (1988) - Les habiletés cognitives dans le travail, in P. Perruchet (Ed), Les automatismes cognitifs, Liège, Mardaga, 139-172.
- MONTMOLLIN, M. de (1984) - L'intelligence de la tâche. Eléments d'ergonomie cognitive. Berne, Peter Lang, 139p.
- RASMUSSEN, J. (1983) - Skills, Rules, and knowledge : Signals, Signs, and Symbols and others distinctions in human performance, IEEE Transactions on Systems, Man, and Cybernetics, Vol 3 MC 13, 3, 257-266.

- REINERT, M. (1986) - Classification Descendante Hiérarchique : Un algorithme pour le traitement des tableaux logiques de grandes dimensions, in *Data analysis and informatics*, North-Holland, Amsterdam, 23-28.
- REINERT, M. (1987) - Un logiciel d'analyse lexicale (ALCESTE). *Cahiers de l'analyse des données*, 4, 1987, 471-484.
- RICHARD, J-F., POITRENAUD, S. (1988). - Problématique de l'analyse des protocoles individuels d'observations comportementales, in J-P. Caverni, C. Bastien, P. Mendelsohn, G. Tiberghien (Eds), *Psychologie cognitive. Modèles et méthodes*, Grenoble, P.U.G., 405-426.
- VALAX, M.F., THURIOT, C., CELLIER, J.M., ERSCHLER, J. (1989) - Planification et contrôle de l'exécution des activités dans les systèmes de production : Une analyse des fonctions planification ordonnancement. Rapport final PIRTTEM n°870004, 1989, 34p.
- VERMERSCH, P. (1984) - L'observation systématique dans l'étude du fonctionnement cognitif, *Psychologie Française*, 29, 3/4, 297-302.

#### **ANNEXE : PRINCIPE ET CONDITIONS D'APPLICATION DE LA CDH**

La méthode de classification proposée a été mise au point il y a déjà près de dix ans d'après une première idée de Benzecri (1973) et d'une méthodologie proposée par Lacoste et Roux (in Benzecri, op. cité). Une méthode similaire a été publiée aux Etats-Unis vers la même époque (Hill, 1979), mais cette approche a été très peu utilisée en France. Le programme tel qu'il existe aujourd'hui a spécialement été étudié pour le traitement de grands tableaux logiques provenant de données textuelles (Reinert, 1987). Il comprend de nombreux "zéros" (1000 vocables par 5000 unités textuelles maximum, si le nombre de "un" est inférieur à 40000).

Les tableaux construits à partir de données séquentielles ayant ces mêmes caractéristiques, nous avons été amenés à élargir l'application de ce programme à ce type de données. Ce parallélisme implique des conditions dans le codage des données recueillies. On doit être en mesure, à un instant donné, de coder les comportements observés dans au moins une catégorie (exhaustivité) et une seule (exclusivité) et ce, sous forme de couples.

La procédure est de type itératif. On part d'une seule classe comportant toute les séquences. A chaque pas, on segmente en deux classes la plus grande classe de séquences de la partition obtenue au pas précédent. L'analyse d'une classe s'arrête lorsque le nombre de séquences classées est trop petit. Une fois ces classes obtenues, le programme calcule les listes de couples les plus spécifiquement utilisés

dans chaque classe (à l'aide d'un  $\text{Khi}_2$ , dans un tableau à quatre cases croisant la présence ou l'absence d'une séquence dans une classe, avec la présence ou l'absence du couple dans cette séquence).

Le problème est de trouver, à chaque pas, la partition en deux classes maximisant l'inertie inter-classe des deux sous-nuages correspondants dans l'espace  $R_j$  muni de la métrique du  $\text{Khi}_2$ . Le calcul effectué ne donne qu'une solution approchée : On recherche d'abord le premier facteur de l'A.F.C.; ensuite, on calcule la partition en deux classes, compatible avec l'ordre des séquences dans le facteur, à l'aide d'un algorithme de type "segmentation"; enfin, un algorithme d'échange améliore la partition obtenue jusqu'à un maximum local.

Ce programme permet ainsi d'obtenir une hiérarchie de partitions sur les séquences. Chaque classe de séquences est ensuite décrite à l'aide des couples les plus significativement présents. Il est alors possible, à l'aide du contenu de ces couples, de reconstruire des graphes exprimant la structure séquentielle des états spécifiques de ces classes.