

J.-P. BENZÉCRI

**Actualité : l'analyse des données en  
archéologie préhistorique**

*Les cahiers de l'analyse des données*, tome 2, n° 2 (1977),  
p. 193-214

[http://www.numdam.org/item?id=CAD\\_1977\\_\\_2\\_2\\_193\\_0](http://www.numdam.org/item?id=CAD_1977__2_2_193_0)

© Les cahiers de l'analyse des données, Dunod, 1977, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

## ACTUALITÉ : L'ANALYSE DES DONNÉES EN ARCHÉOLOGIE PRÉHISTORIQUE

### [PRÉHISTOIRE]

par J.-P. Benzécri (1)

1. Le langage mathématique et l'outil informatique pénètrent aujourd'hui dans toutes les disciplines. Engouement sincère mais peu éclairé ; poudre au yeux ; progrès méthodologique... : il est impossible de tout trier et de tout juger ! D'ailleurs la parabole de l'ivraie et du bon grain nous avertit qu'il serait prématuré d'arracher du champ de la science, les herbes folles. Cependant entre les modèles mathématiques qui conservent les formules et l'enchaînement déductif des théories pures pour les ajuster non sans peine au réel, et les spectaculaires démonstrations de banques d'images appelées d'un trait de stylet sur un écran cathodique, on sait que nous avons choisi la voie de l'analyse des données : par un traitement (qui pour le mathématicien de profession est très simple et ne pose à l'informaticien aucun problème original, mais requiert l'intelligence de l'algèbre et de l'analyse multidimensionnelle et ne se conçoit pas hors de l'ordinateur) transformer un tableau de données (dont la compilation n'impose au spécialiste de la discipline - médecine, linguistique ou archéologie - que des contraintes tolérables auxquelles les érudits sont rompus depuis des siècles) en des cartes planes, des tableaux de nombres, des arborescences qui à l'observateur attentif manifestent des rapports, inspirent des idées, suggèrent des expériences (dont la critique et la validation n'ont pas l'absolue certitude du calcul algébrique ; mais ne requièrent que des rudiments de la science des figures et des nombres ; et donnent à l'induction une solidité à laquelle ne peut prétendre le simple enchaînement poétique des idées et des cas individuels).

Il semble aujourd'hui qu'après d'autres disciplines l'archéologie préhistorique a rencontré l'analyse des données. A la vérité il est impossible d'assigner une date à cette rencontre : dans le premier de ces cahiers est publié l'analyse par J. Robert et M. Roux d'un tableau compilé par Leroi-Gourhan (Préhistoire de l'Art Occidental Paris Mazenod 1965, 482 p.) et cette analyse remonte à 1970 ; d'ailleurs depuis 25 ans et plus, chaque méthode statistique nouvelle relevant de près ou de loin de ce que nous appelons analyse des données, n'a guère tardé à s'essayer à la préhistoire. Mais il est symptomatique qu'ait paru en 1975 un ouvrage britannique qui sous le titre de *Mathématiques et ordinateur en archéologie*, traite en fait pour plus de la moitié d'analyse des données ; cependant que dans un mémoire de maîtrise ès lettres soutenu à Paris en Juillet 1976 F. Djindjian, appliquant l'analyse des données à l'étude de l'outillage de pierre taillée, a obtenu des résultats auxquels le IX<sup>e</sup> Congrès de l'Union Internationale des Sciences Préhistoriques et Protohistoriques (Nice 13-18 Septembre 1976) a fait un chaleureux accueil ; accueil mérité selon nous par le progrès méthodologique décisif que constitue l'application conjuguée de l'analyse des correspondances, du codage sous forme disjonctive complète et de la classification par agrégation suivant la variance avec la distance du  $\chi^2$ .

Nous rendrons donc compte ici du traité britannique et du mémoire parisien ; en nous bornant toutefois à ce que peut saisir immédiatement un statisticien qu'intéresse la préhistoire ; tandis qu'un article de ce même cahier donne du chapitre le moins accessible de F. Djindjian un exposé détaillé illustré de nombreux graphiques grâce auxquels un lecteur attentif peut sans être spécialiste suivre au plus près l'archéologue.

2. Mathematics and Computers in Archeology : par J.E. Doran et F.R. Hodson ; Edinburgh University Press, 1975.

2.1 Le livre est dû à deux auteurs ; qui dès la préface annoncent non sans humour : "We have accepted that our views differ on a number of points and that our modes of expression are largely irreconcilable. In this situation it has seemed appropriate to state for each chapter the author whose primary responsibility it is, while stressing that criticisms of the co-author have always modified, sometimes drastically, its final content and form". D'après ce loyal partage des chapitres nous n'avons pas de peine à découvrir que nos vues sont à peu près celles de F.R. H., tandis que la lecture des pages dues à J.E. D. nous est rendue tolérable par ces "modifications radicales" que l'autre auteur y a introduites !

Après un chapitre d'introduction dû à F.R. H., J.E. D. expose dans la première partie (CC. 2 à 4) sous le titre d'*instruments fondamentaux de l'archéologie et des mathématiques (Basic Archeological and Mathematical Tools)* un abrégé de statistique mathématique où nous n'avons rien relevé qui soit propre à l'archéologie, sauf la datation des vestiges organiques par le <sup>14</sup>C radioactif (§ 3.19 ; pp 66-70). Graphes, probabilités, *statistical inference*, épreuves de validité, loi normale : autant de notions auxquelles est dévolue selon nous une place trop importante ; puisque, (abstraction faite de nos propres convictions manifestées de puis longtemps !) les auteurs eux-mêmes en offrent peu d'applications à des données concrètes. Mais aussi : géométrie et distance ; ajustement au sens des moindres carrés ; comparaison de la variance intra-classe (intérieure aux classes) à la variance interclasse (entre les classes) : notions dont l'analyse des données est redevable à la statistique classique. Et dans l'ensemble de la première partie, rien ne manque de ce qui, graphique ou exemple, peut éclairer un lecteur non mathématicien.

2.2.0 Avec cinq chapitres et 172 pages, la deuxième partie, signée F.R. H. et intitulée "*Datas analysis*" fait à elle seule la moitié du texte (indices exclus) ; à l'analyse des données se rattache encore le ch. 10 - "*Automatic Seriation*" - signé J.E. D.. En rendant compte de ces six chapitres - les seuls du livre qui nous intéressent directement, nous considérerons successivement les principes, les méthodes, les données.

2.2.1 Les principes : Ils diffèrent peu de ceux que nous préconisons. On lit p. 96 : "*On aura recours aux techniques statistiques (statistical procedures) pour décrire et condenser les données et pour suggérer les hypothèses, mais non pour estimer les paramètres d'une population parente supposée, ce qui est le contexte ordinaire de l'application des théories statistiques. On ne fera pas totalement fi des probabilités (will not be ignored completely)*" par exemple pour estimer l'erreur d'échantillonnage sur une fréquence ; apprécier la validité d'un résultat par une simulation ("*certain kinds of simulation*" ; nous en usons aussi : cf [Epr. Val.] TII B n° 8, pp. 210-230) ; mais dans l'immédiat, on préférera les méthodes heuristiques incomplètement formalisées ("*more informal, heuristic procedures*").

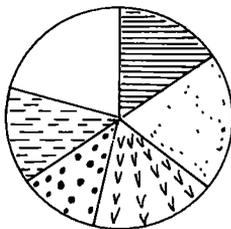
Les auteurs britanniques appellent "*attribute*" ce que nous appelons "variable" (e.g. couleur, matière, épaisseur) ; et "*attribute state*" ce que nous appelons "modalité d'une variable" (e.g. : marron ; en silex ; inférieur à 20 mm). Le problème de la description et du codage est ainsi clairement posé par eux (p. 99).

A Hill et Evans qui objectent que l'infinité des attributs rend impossible une description exhaustive ("*There is virtually an infinite number of attribute connexed with any item*"...) F.R. H. répond en substance (p. 101) : que d'abord l'infinité se réduit quand on considère non les modalités ("*attribute states*" : e.g. intervalles consécutifs en lesquels on peut diviser l'intervalle de variation d'une longueur) mais les variables proprement dites ("*attributes*") ; qu'ensuite, en évitant la redondance, en faisant confiance au bon sens pour écarter ce qui est non pertinent et obscur, on peut établir une description certes lourde ; mais apte à être réduite par l'analyse multidimensionnelle ; et suffisamment indépendante dans sa généralité, de toute hypothèse *a priori*, pour fonder des conclusions objectives (dans un domaine où les épreuves de validité de la statistique classique portent peu).

Et évoquant la perspicacité de Leroi-Gourhan et Brézillon reconnaissant les formes sur le terrain de fouille d'après la disposition soigneusement préservée des vestiges, F.R. H. (p. 151) note que nombre des méthodes discutées par lui ne visent qu'à transformer les données pour rendre à l'œil humain sous forme bidimensionnelle, une vue des descriptions multidimensionnelles initiales. Il admet que des méthodes très élaborées ("*very sophisticated numerical procedures*") sont indispensables pour améliorer cette appréciation visuelle ("*judgement by eye*"). Cherchons donc, dit-il, de telles méthodes, et gardons-nous de croire que puissent les remplacer des épreuves de validité stéréotypées ("*stereotyped tests of significance and the imposition of grids*") qui requièrent des hypothèses peu sûres ("*unwarranted*") quant aux données et à leur structure.

2.2.2 *Les méthodes* : Un livre destiné aux préhistoriens ne peut offrir un exposé mathématique complet, tel qu'on l'écrirait pour des statisticiens ; mais les idées sont clairement rendues ; les calculs se devinent pour qui les a déjà faits ; et des références bibliographiques minutieuses permettent de s'assurer du détail des méthodes.

2.2.2.0 Dans le Ch. 5, F.R. H. accorde une juste place aux tracés les plus élémentaires : histogrammes ; diagrammes plans où deux mesures primaires (e.g. longueur et largeur de la pièce) figurent en abscisse et ordonnée ; cercles partagés en secteurs ("*Pie charts*", *graphiques en tarte* dit l'auteur pour évoquer d'une locution imagée ces disques où des rayons séparent des parts inégales) pour représenter e.g. la part relative des divers types d'anses dans plusieurs lots de poterie.



Puis en passant au domaine proprement multidimensionnel, est posé au chapitre 6 le problème du choix d'une distance. En bref c'est sous une forme allégée et avec des références à l'archéologie, l'inventaire que nous avons nous-même emprunté à Sokal et Sneath (cf [Prob. Tax.] TI B n° 1 pp. 73-74 et M. Roux, ([Alpes I] TI C n° 2 pp. 360-374).

F.R. H. ne signale pas le rôle central de l'indice de Jaccard, démontré par M. Roux ([Alpes I] TI C n° 2 § 1.1 p. 362) ; il ne dit rien de la distance du  $\chi^2$  et du principe d'équivalence distributionnelle ; mais en géométrie, il affirme sa confiance en la formule euclidienne et sait qu'il est possible de plier celle-ci à un choix judicieux des axes et des coefficients (p. 156).

2.2.2.1 *Classification* . Le chapitre 7 est consacré à la classification automatique : il suffit de lire le § 4.5.1 pour s'assurer que son auteur en a reconnu les principaux algorithmes qu'il expose sans appareil mathématique, comme autant de manières dont un archéologue peut aborder la classification d'un important lot d'outils lithiques, par exemple 300 bifaces (handaxes), cas effectivement traité par F. Bordes; F.R.H. distingue donc :

1) La classification ascendante hiérarchique : "après un survol de l'ensemble, l'archéologue met ensemble les deux bifaces qui se ressemblent le plus ; ... puis il forme une deuxième paire, à moins qu'il n'adjoigne un nouvel individu à la première paire formée ; et ainsi de suite ... , des groupes s'édifient..."

2) La classification descendante : distinguer d'abord un attribut (e.g. taille, ou torsion...) d'après lequel le lot d'outils est partagé en deux ; chacun des groupes pouvant ensuite être divisé à son tour...

3) L'agrégation autour de centres variables : distinguer à titre provisoire, quelques spécimens, par exemple cinq ou dix qui serviront de type autour desquels toute la population sera groupée en classes ; puis au sein de chacune de celles-ci choisir au mieux un nouvel individu-type, bien central ; et recommencer autour des nouveaux types l'affinement de tout le lot à quelques classes ; etc.

Comme nous, F.R.H. se défie de la classification descendante: plus précisément de la définition des classes par dichotomie suivant un seul attribut (déf. monothétique) ; à la vérité, cette définition n'est possible qu'*a posteriori* : une fois le matériel parfaitement étudié suivant tous ses attributs, on peut généralement associer à la classification qu'on a édifiée par voie ascendante, une clef dichotomique (ou quasi dichotomique) ; requérant à chaque pas l'examen d'aussi peu de caractères que possible) suivant laquelle la détermination d'un individu se fait en descendant par une suite de questions qui se conditionnent mutuellement telles que : a-t-il des poils ? ; si oui sont-ils roux et longs ;

Quant à la classification ascendante, F.R.H. ne semble connaître que l'agrégation suivant le saut minimum (*single linkage*), l'agrégation suivant le diamètre (*complete linkage*) et l'agrégation suivant la distance moyenne (*average linkage*). Puisqu'il est reconnu que l'agrégation suivant le saut minimum conduit à un effet de chaîne (classes longues effilées cheminant d'individu en individu jusqu'à réunir des extrêmes très dissemblables entre eux) ; tandis que l'agrégation suivant la distance moyenne, qui donne généralement les meilleurs résultats, n'a pas la faveur des puristes (notamment parce qu'elle n'est pas définie par la seule donnée des inégalités entre les distances) ; F.R.H. (qui a examiné au passage le problème confus des classes empiétantes) conclut mélancoliquement : "c'est peut-être trop que d'attendre d'une hiérarchie rigide qu'elle soit à la fois utile et *"mathematically"* irréprochable".

Cependant à propos de l'agrégation autour de centres variables, F.R.H. signale le critère métrique essentiel : "rendre maximum le rapport de la variance interclasse à la variance intraclasse" (cf § 2.1). Or c'est ce critère qui selon nous fournit en l'agrégation suivant la variance une procédure rigoureuse pour édifier une hiérarchie ascendante qui s'accorde au mieux avec les représentations continues (cartes planes) objet de l'analyse factorielle.

La taxinomie doit tout aux sciences de la nature, toute la terminologie vient de la botanique et de la zoologie ; et les premiers algorithmes explicites sont l'oeuvre d'écologistes (F.R.H. cite Sørensen). Le terme de *type* utilisé par F.R.H. pour désigner le centre autour duquel s'agrège une classe a une longue histoire en taxinomie : un visiteur attentif du Muséum de Paris aura remarqué sur certains fossiles

l'étiquette "type" suivie parfois du nom de Cuvier et d'une date : ce type est seulement l'individu d'après lequel l'espèce a été décrite pour la première fois ; mais on a pu considérer idéalement le type comme un individu parfait présentant dans toute leur pureté les caractères de l'espèce. En bonne logique deux sortes de définitions extrêmes e.g. d'une espèce, sont possibles : 1°) en compréhension - comme un ensemble de qualités abstraites de toute référence individuelle ; et 2°) en extension - comme l'ensemble de tous les individus auxquels s'étend le nom de l'espèce. Le type, comme l'idée dans les dialogues de Platon, serait un individu unique relativement auquel se définirait la perfection de l'espèce. Plus pragmatiquement l'analyse des données devrait permettre de saisir sur une collection finie d'individus inégalement affectés d'un ensemble de qualités l'essence d'une classe potentiellement infinie : F.R. H. (cf e.g. p. 166) voit clairement le problème ; mais une définition, qui fait place à la compréhension et à l'extension et substitue au type unique un échantillon requiert selon nous la représentation simultanée de deux ensembles en correspondance et F.R. H. ne connaît pas cette technique.

F.R. H. dit cependant (p. 172) que les écologistes ont conçu diverses formes d'"inverse cluster analysis" pour trouver des agrégats de qualité (*attributes*) plutôt que d'individus. Il souligne avec justesse que chez les écologistes les attributs sont généralement la présence ou l'absence d'une espèce dans une zone, ce qui correspond à la recherche en archéologie des "assemblages" dont il a cité (p. 161), une explication assez lumineuse : selon G. Childe, une culture est "*a recurrent assemblage of archaeological types*" ; assemblage *recurrent* parce que explique F.R. H. d'après G. C., les types "*are found repeatedly associated together just because they result from the behaviour pattern standardised within one and the same society*". Quant aux "types" c'est l'école française de F. Bordes qui en a poussé l'étude : puisque depuis les temps acheuléens jusqu'à nos jours les outils sont standardisés... on peut en reconnaître des types e.g. les burins etc.. Ainsi l'analogie est claire : les types (burins, lames, grattoirs) du préhistorien correspondent aux espèces (centaurée, myrtille, myosotis...) du naturaliste. Les assemblages de types (tel outil avec tel autre et dans telles proportions ; mais sans un autre...) révélateurs selon G. C. de cultures (ou à tout le moins de comportement standard) sont les associations de l'écologiste (sur la terminologie de l'écologie, cf [Cl. Nat.] TI B n° 2). F.R. H. ne développe pas cette analogie qui s'impose toutefois à la lecture de son texte : nous croyons utile d'y insister parce que l'application des méthodes mathématiques en écologie nous étant bien connue d'après les travaux de M. Guinocet et de son école, la préhistoire pourrait bénéficier de l'expérience acquise ailleurs ; et c'est de ce point de vue que nous exposerons les travaux de F. Djindjian.

Soyons précis : plutôt que les variables ("*attributes*") ce sont leurs modalités ("*attributes states*" selon les auteurs britanniques) qu'il convient d'agrèger : F.R. H. cite un archéologue américain Tugby (cf p. 169) qui a effectivement construit la matrice de cooccurrence des modalités des variables : "*The first state in the analysis was to record in a symmetric matrix the number of times each trait occur with each other*". Faute d'une technique appropriée, l'analyse de cette matrice ne conduit à rien qui emporte l'assentiment de F.R. H., lequel se borne à des critiques ; critiques acceptables *a priori*, mais dont l'expérience acquise par F. Djindjian, (utilisant le programme STEAK et l'analyse des correspondances pour traiter un ensemble de variables mis sous forme disjonctive complète : cf Vol. II n° 1) montre qu'elles étaient trop sévères.

Dans le chapitre de la classification, sont également cités l'analyse discriminante et le calcul de la distance de Mahalanobis (qui requièrent une partition donnée *a priori* des individus en classes ; pour un aperçu de l'a. discriminante cf Histoire V : Cah. Vol. II n° 1 ; pour un exposé détaillé [Sep. Corr.] qui paraîtra dans ces cahiers).

Et les auteurs concluent en affirmant comme il est d'usage que les procédures automatiques ont au moins deux avantages : 1°) de contraindre le chercheur à rendre explicites ses arguments ; 2°) de s'appliquer immédiatement à des domaines nouveaux et complexes (e.g. la classification des "assemblages" ; i.e. les inventaires par types d'outils trouvés dans une même fouille : problème écologique ; cf infra F. Djindjian) où l'archéologue est mal servi par son intuition.

2.2.2.2 *Analyse factorielle*. Sous le titre d' "Other Methods of Multivariate Analysis" le chapitre 8 traite principalement de l'analyse factorielle. F.R.H. voit clairement qu'il s'agit de rapporter un nuage de points à des axes bien choisis. Malgré son rôle classique, le modèle normal, pas plus qu'aucun autre modèle paramétrique, ne lui paraît convenir à la statistique. Comme il est d'usage dans le monde anglo-saxon (cf Histoire IV : Vol I n° 4 ; et aussi l'indice général de Hist. et Préhist. de l'A. des D. Vol II n° 1) les auteurs distinguent nettement l'analyse factorielle (rotations, saturations, communautés etc...) de la simple analyse en composantes principales (diagonalisation d'une matrice pour trouver des axes d'inertie) : bien que celle-ci ne soit vue par beaucoup que comme une première approximation préparant celle-là, F.R.H. conclut (p. 205) que seule vaut la réduction purement géométrique du multidimensionnel au plan (ou à quelques axes) : "*It would certainly make better sense to emphasize the "simplification" aspect of these studies, basically achieved by a principal components transformation, and to omit reference to "communalities", "specific" effects and the other perhaps misleading aspects of a model which is neither obviously relevant nor rigorously applied*".

Ici comme ailleurs, le manque d'une représentation simultanée, et l'absence de symétrie entre les rôles dévolus aux individus et aux variables (ou à leurs modalités), se font souvent sentir : on lit par exemple p. 197 : *it is important to remember that principal components analysis starts from covariances and corrélations between attributes, principal coordinates analysis from dissimilarities between units,...*. Inutile de préciser ce qu'on entend par "*principal coordinates analysis*" (terme attribué à Gower) : il suffit de noter que les auteurs ne voient pas comment traiter une matrice de corrélation et combiner linéairement des variables d'une part, ou traiter une matrice de distance et définir par des individus des axes d'autre part, peuvent n'être qu'une seule et même chose. (ici joue notamment le problème du choix des pondérations, auquel il est fait allusion p. 204 : "*this stage has effectively given equal weigh to each type whether frequent or rare, well or not so well delineated (e.g. to the classes "handaxes" and "other small tools")*" : à ce propos nous appliquerions le principe d'équivalence distributionnelle.).

D'autres techniques figurent sous le titre de "Non-metric Multidimensional Scaling" : ainsi l'analyse des proximités de R. N. Shepard ; et le programme INDSCAL de D. Carroll (dans la ligne de l'analyse des tableaux ternaires expliquée dans Histoire V : Vol II n° 1) : disons seulement que les auteurs britanniques ne songent pas à préférer ces techniques à la simple analyse en composantes principales. Comme au chapitre précédent, l'analyse discriminante est citée : car tout en visant au classement, elle est aussi une technique géométrique dont l'effet est de projeter le nuage des individus sur un espace de faible dimension où la discrimination soit au plus clair (p. 210). Enfin les auteurs semblent fonder de grands espoirs sur une analyse des constellations qu'on peut en bref décrire ainsi : plusieurs représentations euclidiennes d'un même ensemble d'individus ayant été obtenues à partir de différentes descriptions de ceux-ci (e.g. les mêmes assemblages, ou lots d'outils trouvés en un même lieu, cf *supra*, caractérisés *primo*, par l'inventaire des types de lames, *secundo*, par l'inventaire des types de burins...), s'efforcer d'amener par rotation toutes ces figures en coïncidences, pour apprécier la concordance des descriptions ou les rapports qu'elles ont entre elles... Il nous serait difficile de prouver que cette analyse des constellations ne doit pas survivre ! (nous-mêmes, nous nous sommes appliqués

à l'étude des déformations globales et locales d'une figure : cf TII B n° 12, § 2.2.3 ; TII C n° 2 note 3) ; mais les comparaisons qui intéressent le préhistorien s'obtiennent sans peine par le report en éléments supplémentaires des colonnes d'un tableau sur les axes issus de l'analyse d'un autre tableau.

Somme toute, ici encore, le livre dont nous rendons compte, témoigne d'une parfaite connaissance des problèmes ; et d'une grande sûreté de jugement : mais nous croyons qu'à l'"assemblage" il manque des outils essentiels.

2.2.2.3 *Sériation* : Aux méthodes de l'analyse des données est également consacré le chapitre 10, "Automatic seriation" que signe J.E. D.. Le problème est clairement posé par des exemples, tel celui-ci : D'après un ensemble d'objets manufacturés de même type - par exemple un ensemble de fibules, ou un ensemble d'épées - conjecturer l'ordre chronologique dans lequel ces objets ont été fabriqués ; ("*Given a set of related artifacts, for example fibulae or swords, determine the most likely chronological sequence of their manufacture*") ; ou encore : ordonner une série d'assemblages d'après les inventaires des objets qu'ils comprennent. Les auteurs considèrent ensuite ce que pourrait être un critère de comparabilité entre une sériation (ou permutation) des objets et ce que l'archéologie conjecture des lois de l'évolution temporelle des arts et des métiers : mais il est clair pour eux comme pour nous qu'il ne saurait être question de choisir e.g. entre les 3.628.000 permutations possibles de 10 unités (épées ou assemblages) par le calcul numérique explicite d'un critère (où l'on ne peut d'ailleurs guère songer à incorporer en une formule toutes les exigences de l'archéologue). D'où l'idée acceptée par J.E. D. qu'il convient de disposer les objets suivant leurs ressemblances non nécessairement sur un axe (présupposé chronologique) mais plutôt dans un plan, où pourront apparaître plusieurs tendances ("*trends*"), éventuellement associés au temps ("*which may or may not be chronological*") : c'est là un problème d'analyse des données ("*multivariate analysis*" ; "*one - or two-dimensional scaling*" dit le texte anglais). Comme presque toujours, on aboutit à l'analyse d'un tableau rectangulaire : tableau de contingence, pour des inventaires d'assemblages ; tableau de description sous forme disjonctive complète (ou éventuellement en vue de certaines techniques : matrices de similitude) pour des objets manufacturés. Et une fois de plus, l'absence de symétrie entre les rôles dévolus aux lignes et aux colonnes est cause d'embarras, témoin cette mention (p. 271) : "*In a recent study Cowgill (1972) has taken as the units in the scaling analysis not the seriation units but the attributes or types by which they are described. Thus it is the columns rather than the rows of the data matrix which are scaled*".

Parmi les méthodes les plus simples, une place est faite à celle généralement attribuée à Czekanowski (auteur non cité dans le livre anglais) : permuter les lignes et les colonnes d'un tableau jusqu'à en concentrer au mieux les cases pleines le long de la diagonale. Une suggestion de Meighan et Ascher (1959 ; cf p. 280) nous intéresse : sous le titre de seriation tripolaire, ces auteurs proposent en bref de se ramener par agrégation des données à un tableau de correspondance à trois colonnes : d'où un diagramme triangulaire pour l'ensemble des lignes, la ligne  $i$  étant figurée au barycentre des trois sommets  $j$  (représentant les colonnes) affectés chacun de la masse  $k(i, j)$ . Nous rappellerons ici un exemple traité par B. Escofier (TII C n° 1 : [Vas. Chin.] : reprenant une description de bronzes chinois d'où V. Elisseeff avait extrait un scalogramme par élimination de vases et de caractères, B. E. obtient sur le premier axe un ordre d'ensemble des vases et des caractères ; le deuxième axe isolant quelques caractères dont la suppression permet d'obtenir sur tous les vases un scalogramme parfait : ce qui nous laisse avec l'hypothèse que le premier facteur est chronologique ; et le deuxième orthogonal au temps, purement stylistique. On voit comment l'analyse des correspondances conjugue les idées citées ci-dessus : matrice en bande diagonale (Czekanowski ; scalogramme) et représentation barycentrique (Meighan et Ascher).

2.2.3 *Les données* : En vue notamment d'une confrontation avec le mémoire de F. Djindjian, recensons les exemples d'analyses proposés au Ch. 9 (47 p.).

2.2.3.1 : Trente fibules trouvées en Suisse (Müstringen, près de Berné) dans un cimetière remontant à l'âge du fer : Par la structure, la fibule est une épingle de nourrice ; par l'ornementation, c'est une broche. La matrice de description est très hétérogène, elle comporte 13 variables : angle, longueur, largeur, épaisseur, nombre de spires du ressort. Dans leur expérimentation sur les méthodes, les auteurs ont renoncé à traiter une population plus nombreuse (109 ou seulement 70 fibules), ou des descriptions plus minutieuses. Les arbres de classification obtenus par diverses voies arrêtaient l'oeil : car les auteurs ont placé à côté de chaque numéro d'ordre un dessin de l'objet ; les graphiques d'analyse factorielle (ou d'analyse discriminante) - simples nuages de points - au contraire sont muets (p. 228) : il serait certes plus difficile d'y placer des dessins du fait de l'inégale densité du nuage ; et la représentation simultanée des deux ensembles (individus et caractères) s'imposerait. En somme le texte conclut que "*a great deal of information about the interrelationships between the fibulae is obtained*" et parle de "*results of direct archeological interest*".

M. Roux à l'instigation de F. Djindjian, a repris ces données en appliquant comme il se doit, le codage sous forme disjonctive complète. Les résultats sont satisfaisants ; autant qu'ils peuvent l'être avec un aussi faible échantillon. La méthode, quant à elle, accepterait sans difficulté des données bien plus amples ; et se prête à des essais tels que mise en éléments supplémentaires de sous échantillons (provenant e.g. d'un site écarté) ou de groupes de variables (e.g. variables de décoration ; la description technique donne seule les éléments principaux).

2.2.3.2 : Vingt six épées (toutes en bronze sauf une, en fer) ; 19 variables, dont 16 mensurations diverses, et 3 dénombrements de trous de rivets effectués en diverses places de l'arme. Cet exemple est traité bien moins en détail que le précédent (4 pages aux épées ; 19 pages aux fibules) : les auteurs portent notamment des jugements favorables à la classification et à l'analyse discriminante.

2.2.3.3 : Une population de 488 bifaces (handaxes) extraits d'un lot de 4800, provenant de 38 sites importants de Grande-Bretagne et mesurés par Roé ; cinq variables seulement : la longueur et quatre rapports de mensurations. Comme pour les épées il s'agit d'une étude préliminaire ; dans l'ensemble la situation est confuse ; au passage les auteurs remarquent sur l'histogramme du rapport largeur/épaisseur, une coupure possible proche du niveau 2,35 déjà remarqué par Bordes sur ses collections de bifaces de France.

2.2.3.4 : Analyses chimiques, comportant la recherche d'éléments à l'état de trace ; d'une part sur des bronzes ; d'autre part sur des poteries provenant de toute l'Europe ainsi que d'Egypte (poteries seulement). Le texte note (p. 250) combien est embarrassante l'expression numérique des teneurs en traces (e.g. en Bi dans le bronze) et propose une transformation logarithmique après décalage d'échelle ; selon nous le codage par classe après examen de l'histogramme s'impose. Quant aux bronzes la conclusion est : "*far more cluster analyses using far more efficient algorithms are required for the vast sample of trace-element analyses that have already been published in Europe*". Nous suggérons donc au lecteur de s'attaquer à l'imposant corpus de données rassemblé principalement par le groupe de Stuttgart (e.g. Junghans et al. : des livres publiés à Berlin ; ed. Mann ; 1960 et 1968). Les données de poterie semblent aux auteurs britanniques d'une interprétation plus facile que celles des bronzes ; mais elles pourraient aussi être revues.

2.2.3.5 : Les assemblages paléolithiques de Ksar Akil (près Beyrouth) : pour 14 niveaux datés approximativement au  $^{14}\text{C}$  entre 40.000 et 30.000 ans avant l'incarnation, on a relevé au total plus de 15.000 objets répartis en 128 types suivant le système de Bordes ( cf *supra* § 2.2.2.1). Un tableau de contingence donc que nous aimerions à traiter. Les auteurs quant à eux, impressionnés par la masse des informations en jeu ("*such a large body of paleolithic data*") estiment que quant à la simplicité et à l'efficacité, les traitements informatiques déjà effectués l'emportent sur les autres "*approaches*" qu'on pourrait proposer.

2.2.4 Au delà de l'analyse des données : "*Beyond data analysis. Problems and Prospects*". Tel est le titre de la troisième partie, signée J. E. D. , dont nous avons déjà extrait le chapitre sériation (ch. 10) qui pour nous relève de l'analyse des données (cf § 2.2.2.3). Au delà de cette analyse, il ne nous appartient pas d'aller ! Bornons-nous à signaler un chapitre 11, sur les modèles mathématiques ; où brille (p. 291) cette équation différentielle proposée par Ammerman et Cavalli-Sforza (1973) pour décrire l'extension dans l'espace de l'agriculture européenne :

$$\delta \psi / \delta t = \alpha \psi (1 - \psi) + \beta \nabla^2 \psi ;$$

un tel modèle évoque les équations de la lutte pour la vie, formulées et étudiées par Vito Volterra au début de ce siècle (en bref : oscillations au cours du temps des populations de loup, chèvre et choux ; la destruction des chèvres par les loups aboutissant e.g. à une famine chez ceux-ci ; et à une prolifération des choux libres des dents des chèvres et engraisés par les cadavres des loups ; d'où prolifération des chèvres survivantes etc.). De semblables équations on peut attendre une meilleure intelligence qualitative de l'homéostasie (équilibre) des systèmes naturels ; non un ajustement quantitatif au réel. Au chapitre 12 consacré aux banques de données, l'image de la tour de Babel est en filigrane : mais brille cette maxime toute britannique : "*... perfection is unnecessary ; all that is required is something a little better than has gone before*". Tel est aussi l'esprit du dernier chapitre (ch. 13) qui sans hasarder de prophétie se borne à affirmer que les mathématiques et l'ordinateur interviendront de plus en plus en archéologie non pour mortifier cette science, mais pour en bigarrer encore la palette déjà vive !

Le livre assurément, qui nous a donné beaucoup à dire, et souvent à redire, est riche et invite à poursuivre.

3. Contribution de l'analyse des données à l'étude de l'outillage de pierre taillée ; par F. Djindjian ; mémoire de maîtrise ès-lettres soutenu devant l'Université de Paris I en Juin 1976.

Ce mémoire a les proportions d'une thèse ; et croyons-nous l'originalité et l'érudition requises pour une belle thèse. Quant à la perfection, on peut critiquer la composition d'ensemble ; et lors de la soutenance, H. Delporte s'est divertit à relever les incertitudes de la pagination ; témoin non ambigu que plusieurs travaux sont là superposés sans qu'en soit achevée la synthèse.

L'auteur distingue quatre parties qu'il numérote modestement comme des paragraphes.

§ 1 L'analyse des données

§ 2 L'a. des d. en typologie et en étude morpho-technique de l'outillage lithique : applications aux burins Noailliens de l'abri du facteur à Tursac.

§ 3 Un essai de reconnaissance des formes sur un lot de deux cents bifaces moustériens de Tabaterie (Dordogne).

#### § 4 L'Aurignacien de la Ferrassie : ses rapports avec l'Aurignacien du Périgord.

Inutile dans une revue spécialisée sur l'analyse des données, de suivre le § 1 dont l'une des divisions a, e.g. pour titre : l'analyse des correspondances : petit mode d'emploi sans formules à l'usage des préhistoriens. Souhaitons seulement que l'exposé de F. Djindjian en qui statisticiens et préhistoriens reconnaissent l'un des leurs, soit bien reçu par ces derniers. Quant aux trois études particulières, objet des §§ 2, 3 et 4, nous aurons plaisir à les replacer dans le cadre inspiré de l'écologie déjà esquissé au § 2.2.2.1.

#### 3.1 Le point de vue écologique en archéologie préhistorique :

Sous le nom d'écologie l'inquiétude de l'homme moderne, évoque l'équilibre dans la biosphère, des espèces végétales et animales et singulièrement de la plus redoutable de celles-ci : des fils d'Adam. Cependant depuis trois quarts de siècle peut-être, des savants se sont spécialisés dans l'étude des associations entre des espèces et leur milieu. Devant le volume considérable des données dont une telle étude requiert le collationnement, ces savants ont dû créer des algorithmes explicites ; ils ont fondé la taxinomie mathématique (cependant que l'analyse factorielle naissait de la biométrie et de la psychométrie). Nous ne suivrons pas l'histoire de ces recherches : nous rappellerons seulement quels types de tableaux l'écologiste (l'écologie végétale nous suffira), soumet aujourd'hui au statisticien ; afin qu'apparaisse l'analogie des problèmes avec ceux de l'archéologie préhistorique.

3.1.1 Format des données écologiques : L'unité d'observation en écologie, est l'individu d'association : ou surface floristiquement homogène que le spécialiste distingue des surfaces contiguës non seulement d'après sa physionomie (une forêt où prédominent les hêtres ; une prairie etc..) mais d'après le relevé attentif de toutes les espèces qui y poussent (pour une définition précise empruntée à M. Guinochet cf [Cl. Nat.] TI B n° 2 § 3.3). L'individu d'association peut être décrit par deux types de données ; d'une part les caractères végétaux : présence ou absence de telle espèce (éventuellement son abondance) ; d'autre part des variables physiques, chimiques, voire climatologiques mesurées sur le sol ou sur le milieu. Si l'on borne l'analyse aux données du premier type (lesquelles selon M. Guinochet résument tout à elles seules) on a fait non de l'écologie *lato sensu*, mais proprement de la phytosociologie ou coenologie dont il s'agira uniquement ici sous le nom d'écologie.

De façon précise, soit I un ensemble de parcelles de sol (chacune représentant par son homogénéité acceptable, un individu d'association) ; J un ensemble d'espèces végétales ; qui seront soit toutes celles relevées dans l'une au moins des parcelles ; soit seulement une partie. On construit le tableau de relevés  $k_{IJ}$  : avec  $k(i,j)$  une information qui peut être :

- présence-absence : 1 si j se rencontre en i ; zéro sinon.
- dénombrement : le nombre d'individus de l'espèce j trouvés en i
- abondance : un indice numérique (prenant, e.g., 5 valeurs de 0 à 4)

Sans discuter les mérites divers pour le botaniste des analyses fondées sur ces diverses informations, rappelons que les nombres bruts (b) sont à la fois difficiles à évaluer et cause de perturbations pour les très fortes valeurs ; inconvénients dont sont exemptes les données a et c ; la préférence de M. Guinochet allant aux tableaux en présence-absence (cf M. Roux [Alpes I] TI C n° 2 § 3). Pour le préhistorien il en peut être autrement :

en particulier l'intérêt des dénombrements analysés tels quels semble établi.

Certes, nous venons d'en voir un exemple, il est impossible d'identifier la méthode du préhistorien à celle de l'écologiste ; mais il vaut d'en détailler l'analogie.

**3.1.2 Ecologie et systématique** : Le spécialiste de l'écologie végétale est nécessairement un botaniste, instruit au préalable de la systématique du règne végétal : ayant recueilli une plante, il pourra (parfois aidé d'un spécialiste) en trouver le nom : e.g. *Lolium perenne* pour telle espèce du genre ivraie (*Lolium*). (Et si la taxinomie végétale discipline vénérable mais que d'aucuns trouvent vétuste dans sa minutie vit toujours au siècle de la biologie moléculaire, c'est entre autres causes, qu'hors de la poussière des herbiers, elle est indispensable à l'écologiste de plein champ !). En contrepartie, de quoi le préhistorien dispose-t-il ? de l'ensemble des types établis par Bordes ; qu'on peut appeler le Linné de l'outillage lithique (encore que Bordes ait eu des prédécesseurs : e.g. G. de Mortillet au XIX<sup>e</sup> siècle). L'analogie est donc : pour les plantes "l'espèce" ; pour les outils le "type". Mais tandis que les plantes, êtres vivants étroitement définis dans leur organisation ont fait l'objet de très nombreuses études qui en ont établi la taxinomie avec une grande certitude (même si tout n'est pas résolu ; les techniques génétiques nouvelles aidant à la définition des espèces etc.) les outils lithiques préhistoriques, qui ne portent sur eux que les traces d'un façonnage humain destiné à les rendre propres à une fonction inconnue de nous (un préhistorien d'aujourd'hui(\*) répugne à conjecturer publiquement à quoi servait tel fragment de silex que les savants du XIX<sup>e</sup> siècle ont appelé burin, par analogie avec tel outil contemporain) ne sont qu'imparfaitement rangés sous des types. D'où une première voie de recherche proposée au préhistorien armé de l'analyse des données : rendre plus certaine, plus fine, plus rigoureuse, la taxinomie de l'outillage lithique. C'est à quoi contribuent les §§ 2 et 3 du mémoire de F. Djindjian. Disons tout de suite que de telles recherches sont très difficiles car elles requièrent une description des formes ; laquelle pour être exhaustive peut s'étendre à l'infini (cf § 2.2.1 : l'objection de Hill et Evans). Nous y reviendrons à propos des analyses de F. Djindjian.

**3.1.3 L'analyse des tableaux de relevés** : S'il accepte au moins à titre provisoire, la taxinomie des outils lithiques telle qu'elle est e.g. dans la nomenclature de Bordes, le préhistorien peut constituer un tableau rectangulaire  $k_{IJ}$  tout analogue au tableau de relevés de l'écologiste. Certes il en coûte un grand labeur (plus de 15.000 objets étiquetés à Ksar Akil ; cf § 2.2.3.5) ; mais rien ne prouve qu'il en coûte plus ici que là. Quant à la définition du champ à relever : l'analogie entre écologie et préhistoire est frappante : nous avons rappelé au § 3.1.1 la définition de l'individu d'association comme floristiquement homogène : de même le préhistorien qui inventorie un assemblage, conservé par exemple dans une masse de sédiments meubles, doit se garder de mêler deux lots contigus. Ici comme là, il est impossible de rien faire sans le sens du terrain qui est un privilège du spécialiste ; et les préhistoriens contemporains notamment ceux que guide Leroi-Gourhan ont acquis une acuité de regard qui ne leur laisse même rien négliger de la position relative des objets ; significative elle aussi, de la vie qui les utilisait après les avoir fabriqués.

A la différence des descriptions de formes (formes d'outils ou formes d'êtres vivants), les tableaux de relevés sont pour le statisticien les données les plus régulières et les plus simples : ce sont des tableaux de contingence auxquels l'analyse des correspondances s'applique immédiatement, sans codage préalable (e.g. par les programmes tels que

(\*) Même s'il est - tel Jean Tixier - un virtuose de la taille du silex ; ou-tel A. Leroi-Gourhan - un ethnologue averti de toutes les technologies contemporaines héritées de l'âge de pierre : cf *L'homme et la matière* ; Albin Michel éd. 1971 (1943).

STEAK et DIBOUDI). On sait que l'analyse de tels tableaux fournit des cartes planes, où figurent à la fois l'ensemble I des individus (i. d'association en écologie ; assemblages en préhistoire) et celui J des caractères (espèces végétales en écologie ; types d'outils en préhistoire) ; ces cartes pouvant être complétées par des classifications automatiques. Le spécialiste jouant suivant des règles maintenant bien établies, des calculs de contribution et de la mise en éléments supplémentaires, parvient alors à reconnaître dans le domaine de ses recherches soit des gradations continues (le long des axes factoriels), soit des oppositions franches (visibles sur les cartes ; et confirmées par les dichotomies de la classification).

Reste à interpréter gradations et oppositions. En écologie végétale, il est d'usage de ne traiter sur un même tableau que des relevés provenant d'un même empire floristique : e.g. on ne cherche pas à comparer directement une parcelle de forêt congolaise à une parcelle de forêt amazonienne, alors que nombre d'espèces parfaitement aptes à prospérer dans l'une et l'autre forêt se trouvent confinées à une seule par l'immensité difficile à franchir, de l'Océan. Il est même exceptionnel que figurent sur un même tableau les relevés de parcelles séparées par des distances de plusieurs centaines de kilomètres, on se restreindra plutôt à une vallée ou à un versant. (Ce n'est que depuis peu que R. Salanon et J. P. Briane ont entrepris l'analyse de tableaux de quelque mille relevés, mettant en jeu des centaines d'espèces). Dans ces conditions, on peut affirmer qu'aux fluctuations près (entraînant des absences occasionnelles ; sans influence sur les résultats de l'analyse ; l'outil statistique agissant comme un filtre qui par sommation élimine le bruit), sont présentes sur chaque parcelle (d'étendue raisonnable) toutes les espèces de la flore régionale susceptibles de s'y établir. C'est par quoi le relevé écologique exprime toutes les potentialités du milieu ; en entendant par milieu non seulement le climat et la composition du sol nu, mais aussi les cycles physiques ou chimiques (évaporation ou rétention de l'eau ; catabolisme des déchets organiques) où la vie (notamment celle des microorganismes) joue le premier rôle. Un tel cadre confère à l'interprétation son unité.

3.1.4 *Le primat du temps* : Qu'en peut-il être en préhistoire ? Aux yeux de maint chercheur la dimension chronologique revêt une importance exclusive : c'est elle seule qu'il cherchera à comprendre dans une sériation de ses relevés (§ 2.2.2.3) ; surtout si comme à Ksar Akil (§2.2.3.5) ceux-ci proviennent de strates dont la superposition régulière donne de façon certaine la succession, précisée encore par la mesure de la radioactivité du  $^{14}\text{C}$ . Il est cependant vraisemblable *a priori* que le facteur temps n'est pas seul en jeu pour expliquer la diversité des assemblages : car en même temps, et dans des espaces voisins peuvent coexister des assemblages qui diffèrent soit parce que les hommes qui les ont produits appartiennent à des civilisations différentes, soit parce qu'au sein d'une même société ils remplissent des fonctions différentes. L'assemblage ne peut-il pas être aussi bien l'analogie soit de la caisse à outil d'un artisan, soit du dépôt d'un fabricant d'outils spécialisé, soit d'une décharge publique... Bien plus, le facteur temps ne joue pas toujours en un même lieu de façon univoque : parfois se superposent dans des conditions climatiques et culturelles stables, les vestiges d'une civilisation qui progresse régulièrement : perfectionnant les techniques, créant ou abandonnant graduellement tel outil. Mais d'ordinaire l'histoire est jalonnée de crises : un peuple en chasse un autre ; une épidémie fait le vide ; le changement de climat impose à l'homme de lutter pour la vie avec de tout autres armes. Ainsi le retour de conditions semblables pourra reproduire en un même lieu des formes de vie, donc d'outillage, qui en avaient disparu ; non parce qu'elle étaient intrinsèquement surannées et que le perfectionnement de la civilisation les avait dépassées ; mais parce qu'elles ne répondaient plus pour un temps aux contraintes extérieures à l'homme. Même si faute du sens de la préhistoire, nous avons choisi maladroitement nos images, aucun préhistorien ne verra croyons-nous dans nos réflexions plus qu'une vérité banale

et admise : toutefois c'est souvent faute de lever les yeux sur l'évidence que le chercheur pris entre les données et les graphiques s'égare, emporté par une hypothèse implicite.

Assurément, bien que le cadre des relevés d'une même étude écologique telle que nous l'avons définie ci-dessus soit la synchronie sur une aire géographiquement circonscrite, la dimension temporelle n'est pas étrangère à la botanique (non plus qu'à la zoologie). La paléobotanique suit l'évolution des formes végétales sur une échelle temporelle où toute la durée de la préhistoire humaine figure comme une subdivision d'ordre inférieur. Or la définition des étages géologiques - citons pêle-mêle des termes comme dévonien, luthécien, jurassique - relève de la paléoécologie : si l'on peut parler de dévonien hors du Devonshire, de luthécien fort loin de Paris, de jurassique là où la chaîne du Jura ne parvient pas, c'est que les caractéristiques de ces étages sont certains fossiles (animaux ou végétaux ; particulièrement des coquilles, des microfossiles...), qui principalement par le milieu marin, ont à des époques successives, prospéré sur presque toute la planète.

Les problèmes posés à propos des assemblages d'outils se retrouvent ici. Certaines espèces sont à un moment donné ubiquistes (partout présentes sur le globe ; au moins dans leur habitat propre : eau douce pour l'une, eau salée pour l'autre...) ; et de plus naissent, évoluent et disparaissent dans le temps à un rythme qui les rend propres à servir d'horloge ; c'est sur elle que l'intuition du paléontologiste fonde sa reconstitution de l'histoire de la vie. Mais quelle que soit la confiance que mérite l'intuition de savants dont les dons innés animent une laborieuse érudition nous suggérons l'analyse de vastes et nombreux tableaux de relevés recensant avec les espèces généralement utilisées comme caractéristiques des centaines d'autres moins en vue. Il est peu vraisemblable que tous les schémas reçus survivent intacts à de telles analyses ; que des axes insoupçonnés n'apparaissent pas aussi importants que ceux déjà reconnus. De ce point de vue, il est avantageux à l'archéologie préhistorique que son écologie et sa taxinomie s'édifient plus tard que celle de la paléontologie ; en un siècle où le traitement automatique des données est possible ; selon des règles posées par l'esprit humain mais dont le jeu explicite et autonome se déploie dans l'ordinateur en dehors de la vie psychique.

Enfin, entre paléontologie et archéologie préhistorique les rapports ne sont pas seulement d'analogie mais de contiguïté. Leroi-Gourhan (1969) p. 8 (\*) écrit : "La sédimentologie, la paléontologie humaine et animale et la préhistoire typologique sont unies par un puissant lien de méthode qui est la stratigraphie. [...] . Elles procèdent conjointement depuis un siècle et demi, dans le quaternaire, sur la lancée de la géologie stratigraphique qui naquit dans la seconde moitié du XVIII<sup>e</sup> siècle. [...] de nombreux chercheurs ont allié les sédiments, les témoins zoologiques et les oeuvres de l'homme, [...]." Aujourd'hui encore, bien que la diversité des techniques perfectionnées impose une spécialisation plus étroite, le préhistorien doit être instruit de la géologie ; le géologue du quaternaire ne peut considérer la terre, sa flore et sa faune, sans l'homme et ses arts.

### 3.2 Les analyses morphologiques de F. Djindjian

3.2.1 La description des formes : une analogie naturelle entre les domaines, autant que notre expérience plus grande de la statistique appliquée aux formes vivantes qu'aux outils lithiques, nous incite à continuer de parler de botanique ou de zoologie à propos de l'archéologie préhistorique. On trouve dans le traité de l'Analyse des Données plusieurs études de formes animales (TI C n<sup>os</sup> 4-5-6-7). Qu'il s'agisse de rongeurs ou de lémuriens, ces études concernent toutes des crânes ; et ceux-ci sont décrits par des mensurations prises soit en ligne droite

(\*) A. Leroi-Gourhan [1969] *Leçon inaugurale au Collège de France*, 32 p.

soit en épousant la courbure de l'os, entre des repères précis qui sont des sutures d'os, des saillies, des insertions de dents... Le choix des mesures requiert une bonne connaissance de l'anatomie ; il peut être guidé par des principes statistiques généraux (e.g. prendre sur quelques lignes des mesures de segments consécutifs plutôt que des mesures empiétantes) ; et l'expérience de plusieurs analyses apporte des conclusions stables : dont la plus simple est que les dents jugales (ou molaires) ne croissent plus une fois écloses elles ont chez l'animal jeune une taille relative plus grande que chez l'adulte dont la croissance osseuse est achevée, d'où un facteur d'âge (1° axe des analyses mêlant des sujets de tous âges) opposant la longueur des séries de dents, à celle des intervalles dépourvus de dents (diastème séparant les incisives des molaires chez les rongeurs ou les équidés). Aujourd'hui la méthode semble au point : quelque 20 mesures prises au pied à coulisse par un agent technique non instruit de la zoologie et seulement entraîné par un zoologiste, permettent e.g. dans le domaine des microinsectivores (étudiés par P. Botros sous la direction de L. Bellier) des déterminations automatiques aussi sûres que celles des spécialistes - lesquels sont fort rares (un ou deux à Paris). Quant aux plantes, la plasticité de leur forme (qui tolère des variations dont l'amplitude relative est bien plus grande que pour un crâne) pose des problèmes plus complexes. Tandis que les mensurations crâniométriques peuvent être soumises telles quelles (sans codage par classe) au programme d'analyse des correspondances ou de classification automatique, la description d'une plante requiert généralement à la fois des variables qualitatives et quantitatives, le codage de l'ensemble s'effectuant comme il est expliqué par M. Roux dans ces *Cahiers* (cf *Erodium* Vol II n° 1) ; de plus le recueil des données (description d'un spécimen) nous paraît beaucoup plus difficile que pour un crâne : témoins nos essais de repérages des éléments caractéristiques d'un pied de *Myosotis* rapportés en TI C n° 1.

Ce contexte étant rappelé, voyons comment F. Djindjian décrit les outils de pierre.

3.2.2 Les bifaces moustériens de Tabaterie. Expliquons successivement la nature de ces outils lithiques ; la provenance du lot étudié ; la description numérique, très accessible au non-spécialiste, adoptées par F. Djindjian ; et les conclusions de ses analyses.

"Les plus anciens des outils trouvés et identifiés jusqu'ici, l'on dans [Grahmann, Payot 1955] p. 168, sont constitués par des galets allongés d'un format bien en main ou par des rognons de silex [concrétions siliceuses de forme arrondie que l'on trouve e.g. dans les terrains crayeux] ayant naturellement une forme commode dont le plus souvent une seule extrémité a été retouchée, pour y former un tranchant ou une pointe. [...] En poursuivant le travail de retouche, un galet finissait par fournir un outil taillé plus ou moins de tous les côtés, le coup de poing. [...] Pendant des milliers de siècles on en a sans cesse fabriqué..."

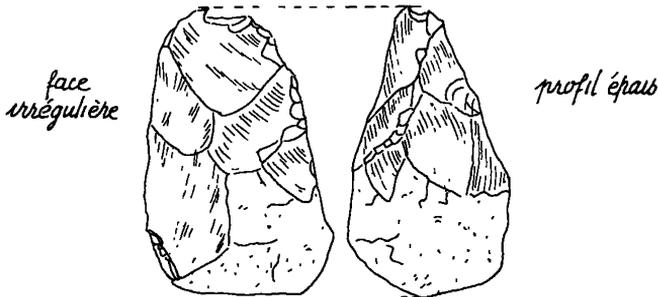


Figure 1 coup de poing abbevillien (1<sup>re</sup> interglaciaire)  
d'après [R. Grahmann, 1950] p. 194.  
"gros, volumineux, atteignant souvent la dimension de la main."

Conjointement, se développe l'industrie des éclats : un éclat détaché par percussion d'un bloc originel a généralement une forme aplatie et présente donc deux faces ; dont l'une, le long de laquelle s'est effectué le détachement, lisse, bien reconnaissable est appelée face d'éclatement. En général un outil sur éclat est retouché sur une seule de ses faces. Or dès l'Acheuléen, (les civilisations acheuléennes ainsi nommées d'après le site français de Saint Acheul se développent en Europe sur une longue période du Paléolithique inférieur comprise entre 600.000 et 100.000 ans avant nous) l'analogue du coup de poing de notre figure 1 est façonné non seulement sur rognon de silex mais aussi sur éclat ; mais à la différence de la plupart des types d'outils sur éclat, il est retouché sur les deux faces. C'est pourquoi les préhistoriens, soucieux de donner aux outils des noms descriptifs et répondant à une définition précise ont abandonné le terme de coup-de-poing pour celui de biface, nom qui selon Leroi-Gourhan ([1969] page 12) fut proposé en 1920 par Vayson de Pradene. F. Bordes (1961, cité d'après M. Brézillon [1971] page 151 (\*)) donne la définition suivante : "les bifaces sont des outils de types variés, généralement taillés à partir de rognons de silex, mais aussi à partir de gros éclats de silex [...]. Leur caractéristique commune est d'être taillés sur les deux faces, par retouche totale plus ou moins envahissante."

La série des deux cents bifaces étudiée par F. Djindjian provient de deux gisements de Tabaterie (Dordogne) : l'abri Brouillaud et le site de plein air de Sandougne distants entre eux de quelques mètres. Ces outils appartiennent à l'époque moustérienne : les civilisations moustériennes, ainsi nommées d'après le site français du Moustier, se développent en Europe durant le Paléolithique moyen, longue période d'environ 50.000 ans qui fait suite au Paléolithique inférieur et se termine quelque 40.000 ans avant nous. Selon F. Djindjian, "la représentativité de l'échantillon est fortement altérée par les conditions de fouilles (couches épaisses, tris) et de conservation..." Couches épaisses : les pièces ont été relevées sans le détail de leur étagement, qui permettrait de mettre en relation avec la chronologie d'éventuelles variations de forme. Tris : les chercheurs n'ont conservé que les spécimens qui leur paraissaient les meilleurs ; le statisticien veut l'exhaustivité. Avec cela, F. Dj. affirme sans gêne que "l'échantillon n'est d'ailleurs ni pire ni meilleur que tous ceux qui s'empoussièrent dans les musées ou collections particulières". Si l'analyse en a été tentée c'est pour le progrès des méthodes descriptives ; non avec l'espoir d'atteindre immédiatement des nouveautés intéressantes la préhistoire.

Déjà soigneusement étudiée par F. Bordes, la description des bifaces semblera facile au non-spécialiste, car elle ne requiert la perception d'aucun détail structurel inaperçu du profane. Après avoir un moment essayé mais sans succès, le codage le plus brutal : appliquer la pièce sur une grille rectangulaire où elle s'inscrit et noter les carrés recouverts (d'où une description en 0,1 : cf Rétine et Perceptron in Histoire IV Vol 1 n° 4 p. 366), F. Djindjian a adopté un système classique de mensurations épaisseur, largeur etc.. celles-ci étant soit soumises telles quelles à l'analyse des correspondances (tableau  $k(i, j)$  = valeur de la  $j$ -ième mesure prise sur l'individu  $i$  : comme pour les crânes), soit traduites en rapports eux-mêmes codés par classes sous forme disjonctive complète ( $k(i, j) = 1$  si pour le spécimen  $i$  possède la propriété  $j$  qui peut être rapport épaisseur/largeur inf à 0,3... ; ce qui rappelle des codages adoptés pour les plantes). De façon précise, la figure 2 ci-dessous donne le système de mensurations adopté par F. Dj., système directement issu de celui de F. Bordes (cf [M. Brézillon] p. 152. Pour mesurer un biface F. Dj. le pose à plat et place au jugé un axe de symétrie approximatif  $z z'$  orienté de la base vers la pointe. Il prend de part et d'autre de cet axe trois séries de mesures transversales à l'axe :

(\*) Michel N. Brézillon. La dénomination des objets de pierre taillée IV° supplément à "Gallia préhistoire" C N R S 1971, 423 p.

(L,M), (N,P) (Q,R) respectivement au premier, second et troisième quart de l'outil à partir de sa pointe. F. Dj. note aussi la dimension longitudinale Z, la largeur maximale I et la hauteur J, comptée à partir de la base, du niveau où l'outil atteint sa largeur maximale. Quant aux mesures relatives au profil, nous nous bornerons à signaler que F. Dj. reconnaît à chaque outil une face plus plate qui est celle à partir de laquelle sont comptées les mesures S, U, W. De toutes les mesures de notre figure 2, F. Bordes considère seulement celles qui, dans les notations de F. Djindjian, s'écrivent : Z, I, J, (N+P), K. Il est clair que pour imparfaites que soient ces mesures, elles suffisent à distinguer un objet rond d'un objet plat ou allongé ; un triangle ou un cœur d'une ellipse ; toutes distinctions qui sont à la base de la terminologie de F. Bordes (Biface triangulaire, cordiforme, amygdaloïde, etc...)

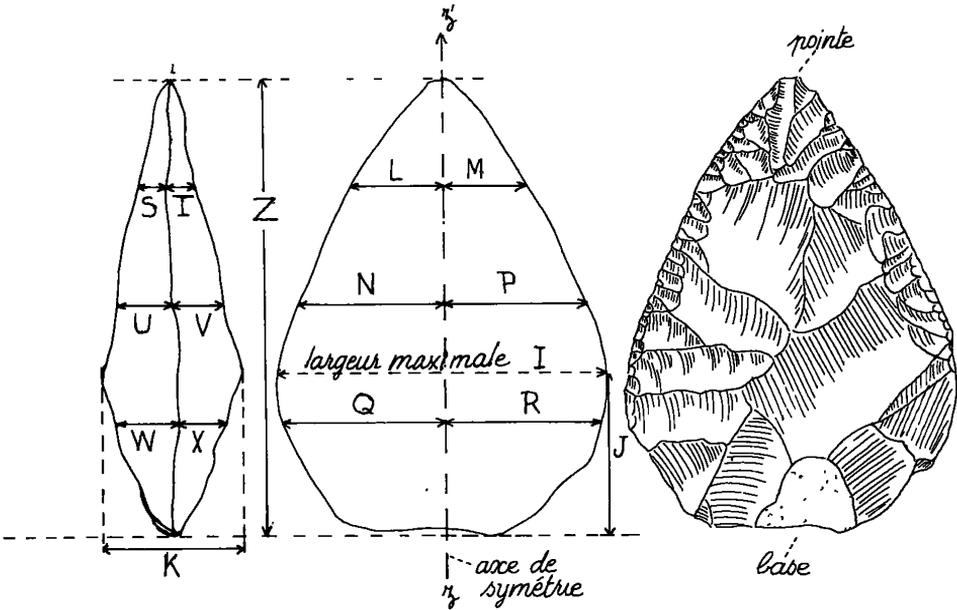
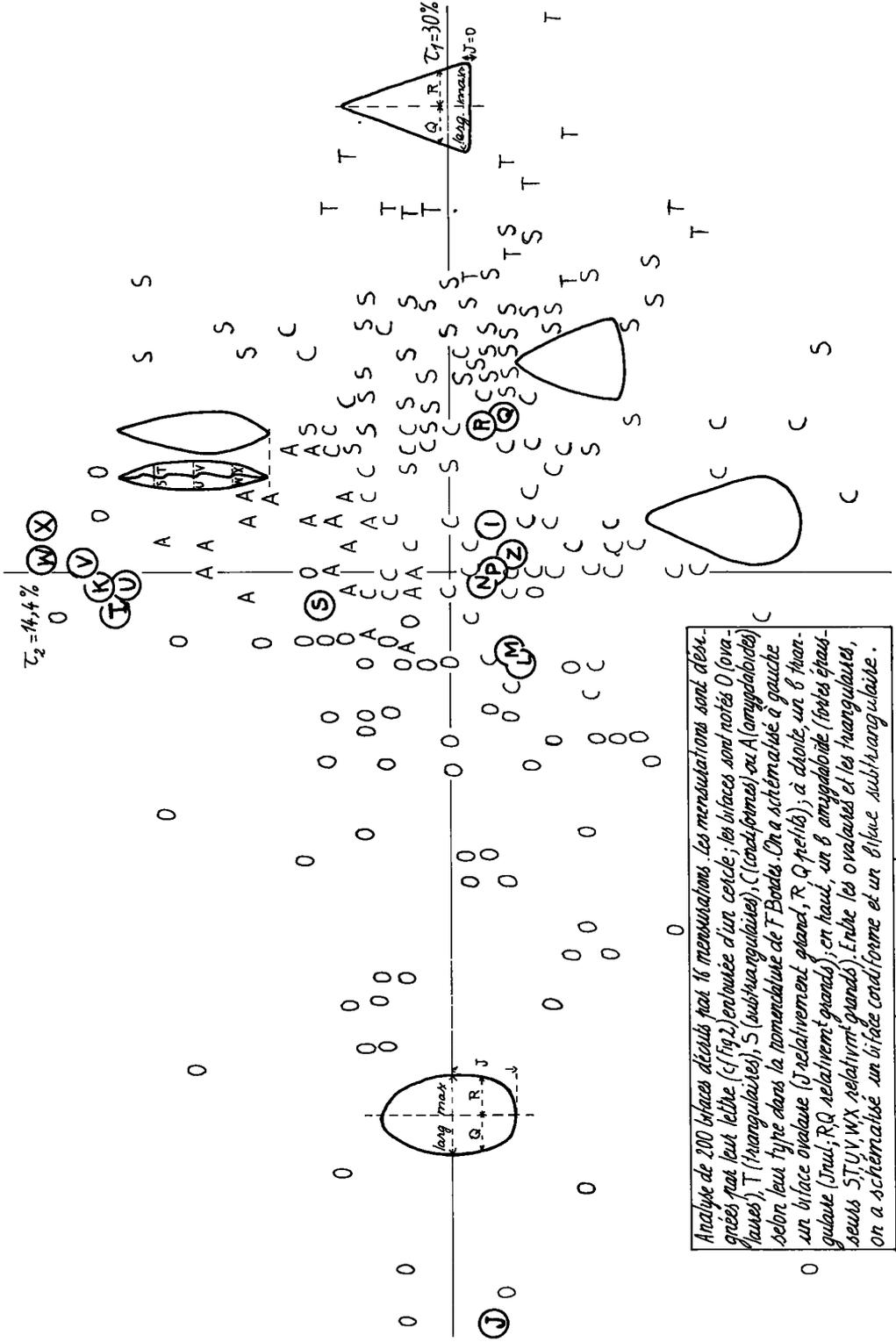


Figure 2. mensurations de F. Djindjian sur un biface moustérien

De l'analyse factorielle, F. Djindjian conclut que les deux cents outils forment un nuage continu : aux types définis par F. Bordes, tels que subtriangulaires, cordiformes (en cœur), amygdaloïdes (en amande), ovulaire..., correspondent des zones contiguës du nuage. On voit sur la fig. 3 que la représentation simultanée des individus (bifaces) et des variables rend compte immédiatement de la définition de ces types de forme. Mais il n'y a pas de discontinuité entre ces types. Si parfois la classification automatique produit des groupes rejoignant les types de Bordes, ce n'est qu'un résultat instable, qu'une variation de codage modifiera. D'où cette conclusion "La typologie ... est artificielle. Elle ne correspond vraisemblablement pas à une intention de taille des hommes préhistoriques".



Analyse de 200 bifaces découps par 16 mensurations. Les mensurations sont désignées par leur lettre (cf Fig. 2) entourée d'un cercle; les bifaces sont notés O (ovales), T (triangulaires), S (sub-triangulaires), C (convexes) ou A (amygdaloïdes) selon leur type dans la nomenclature de T. Bordes. On a schématisé à gauche un biface ovale (J relativement grand, R, Q petit); à droite, un biface triangulaire (J, R, Q relativement grands); en haut, un b amygdaloïde (fibres épaissies ST, UV, WX relativement grands). Entre les ovales et les triangulaires, on a schématisé un biface convexe et un biface sub-triangulaire.

FIGURE 3

Satisfait des techniques de description des formes, F. Dj. ne nie pas qu'on les puisse améliorer. Il souligne que l'analyse des correspondances révèle au mieux l'influence du choix des mesures et de leur codage ; et par ses représentations continues "met l'accent sur les risques d'un emploi trop confiant des classifications automatiques", lesquelles taillent toujours dans tout ensemble des classes fussent-elles contiguës et instables...

Reste enfin le souhait que "la fouille d'une séquence longue de niveaux riches en bifaces" permette de suivre au cours du temps l'évolution sans doute continue de la distribution des formes(\*)

3.2.3 *Les burins de l'abri du facteur à Tursac* : Une définition empruntée à F. Dj. montrera que la description d'un burin, ou seulement sa perception requiert un esprit aiguisé :

"Biseau d'un burin : figure géométrique créée par l'intersection d'au moins quatre surfaces : la face inférieure, la face supérieure du support, le plan de frappe, et le ou les négatifs des chutes de burin".

Quelque six termes techniques, d'ailleurs indispensables car ils correspondent à des gestes de l'outilleur, dont la succession s'inscrit de façon certaine sur le silex ; mais ces traces sont aussi incompréhensibles au profane que le sont les bruits du cœur pour qui n'a pas appris à ausculter.

Devant les variables relevées par F. Dj. sur les burins de silex, le statisticien se demande comment on peut être assuré qu'une description comprenne ce qui est essentiel à l'objet : ce qui en définit l'ordre ou la fonction. Telle description macroscopique, assez semblable à celle des bifaces, pourra suffire à caractériser un muscle (sinon le tissu musculaire lui-même) ; la même analyse appliquée au cerveau ne saisit rien. Le burin est inintelligible si l'on n'en a pas reconnu le biseau : or cette explication ne tombe pas sous le sens. Ici comme souvent lorsqu'on réfléchit sur les méthodes - or la statistique est avant tout une méthode - on doit reconnaître qu'aucun filet ne peut par lui-même saisir l'essence des choses ; et pourtant nous avons confiance que la critique méthodique des découvertes conduit l'esprit à la certitude objective.

Doutant, *a priori*, si le lot qu'il étudie comprend des outils d'un type particulier dit "Burin de Noailles", F. Djindjian, assigne pour but à son étude "d'abord de mettre en évidence l'existence de deux types puis de les séparer avec une précision de l'ordre de 1%. Dans un deuxième temps il s'agit de définir un ensemble minimal de critères... susceptibles de discriminer le plus efficacement possible les deux types...." Or il nous paraît que par l'analyse multidimensionnelle (qui seule permet d'avoir une vue globale de l'ensemble des données et de leurs rapports), F. Dj. atteint les objectifs qu'il s'est assignés. Ce succès incitera préhistoriens et statisticiens à coopérer au difficile progrès de la taxinomie (typologie) de l'outillage lithique. C'est pourquoi paraît ailleurs dans ce cahier un exposé de l'étude des burins qui vise à permettre au statisticien de comprendre le préhistorien ; et à celui-ci d'apprécier la stabilité des résultats obtenus.

(\*) Un exposé complet des analyses de F. Djindjian sur les deux cents bifaces moustériens de Tabaterie est publié dans les Cahiers du centre de recherche de préhistoire, n° 5, 1976, 3 rue Michelet, Paris 6°, cf : F. Djindjian et E. de Croiset : Un essai de reconnaissance des formes sur une série de deux cents bifaces moustériens de Tabaterie (Dordogne) par l'analyse des données.

3.3 *Analyses coenologiques de F. Djindjian* : Par ce titre dont l'explication est au § 3.1.1, nous désignons l'analyse de tableaux de contingence  $k_{IJ}$ . Chaque élément  $i$  de  $I$  correspond à un niveau de fouille (couche présumée homogène d'un site bien circonscrit) où l'on a trouvé un assemblage d'outils lithiques dont l'inventaire est fait suivant une liste type  $J$ ,  $k(i,j)$  est le nombre d'outils relevant du type  $j$  trouvés dans l'assemblage  $i$ . C'est le format des données de Krar Akil (§ 2.2.3.5), format classique en écologie, et objet depuis quelque 10 ans (Binford 1966) d'analyses en composantes principales chez les préhistoriens.

F. Djindjian étudie d'abord 18 niveaux de La Ferrassie (niveaux provenant d'un même site où ils se superposent suivant un ordre chronologique incertain) décrits par 33 types d'outils choisis dans la liste type de D. de Sonneville Bordes et J. Perrot. Puis il étend son étude à l'ensemble de l'Aurignacien du Périgord. (Les civilisations aurignaciennes, ainsi nommées d'après le site français d'Aurignac, se développent en Europe occidentale durant le Paléolithique supérieur période longue d'environ 30.000 ans qui fait suite au Paléolithique moyen et se termine environ 9.000 ans avant nous). Nous nous bornerons à publier ici un graphique commenté.



La figure 4 représente le plan des axes factoriels 1 et 2 de l'analyse croisant 17 niveaux de La Ferrassie désignés chacun par une lettre entourée d'une ellipse e.g. F et 29 types d'outils (désignés par leur numéro dans la liste type), un niveau et 4 types d'outils ayant été éliminés à la suite d'une première analyse.

Voici la liste des niveaux :

ELSB ; F ; GF ; GSN ; GSS ; H ; IS ; I1F ; I2F ; J ; K2 ; K3A ; K3B ; K3C ; K4 ; K5 ; K6 dans l'ordre de leur succession verticale descendante sur le terrain.

Dans le plan des axes 1 et 2, F. Djindjian reconnaît trois groupes de niveaux ; l'Aurignacien 1 : K6 et K5, avec lames retouchées, racloirs et pièces esquillés ; l'Aurignacien 2 : K4 à K2, GSS, avec appauvrissement des types de l'Aurignacien 1 et abondance de grattoirs à museau, burins sur troncature et grattoirs carénés ; l'Aurignacien 3, plus diffus, avec les burins busqués. Ces mêmes groupes, caractérisés par les mêmes types d'outils, se retrouvent dans l'étude du Périgord dans son ensemble.

F. Dj. examine également les plans 2 x 3 de ces analyses, et conclut que "l'Aurignacien du Périgord présente trois stades lithiques de lignée évolutive unique. Le dernier stade représenté seulement à Caminade et à la Ferrassie demande à être confirmé" ; et il suggère d'adjoindre à l'étude les sites récemment fouillés dans le Périgord et dans le Lot ; puis d'autres sites répartis sur toute l'Europe occidentale et centrale (F. Djindjian poursuit présentement ces recherches).

Mais un incident de la soutenance montrera combien l'unidimensionnalité (primat du temps § 3.1.4) doit être regardée avec réserve : F. Dj. ayant mis en doute la représentativité de certains assemblages qui apparaissent sur les graphiques hors de la lignée évolutive conjecturée et admise, H. Delporte affirma qu'ayant conduit les fouilles, il se portait garant que les assemblages suspectés (et certes hors ligne) n'avaient pas été relevés avec moins de science ni de conscience que les autres.

#### 4. Conclusion

La lecture du livre de Doran et Hodson (§ 2) nous a convaincu que des préhistoriens savent que nombre de problèmes de leur discipline seraient résolus par une analyse multidimensionnelle appropriée : par quoi nous entendons le traitement simultané, seul permis à l'ordinateur, des données d'un grand tableau. Cependant les expériences dont ils rendent compte leur laissent désirer certains perfectionnements aux méthodes.

Le mémoire de F. Djindjian (§ 3) nous paraît témoigner que les obstacles opportunément signalés par D. & H. peuvent être évités ; d'autant plus qu'en butte à des difficultés tout analogues (§ 3.1) l'écologie est parvenue à utiliser avec sûreté l'outil multidimensionnel.

Alors qu'un engouement (§ 2) en lui-même peu conforme au sérieux de la science rend acceptable tout ce qu'impriment les ordinateurs, il est à craindre que le verbiage informatico-statistique ne déçoive et n'écarte les préhistoriens. Mais aussi l'occasion s'offre de fonder sur des travaux sains une collaboration fructueuse.

Bibliographie complémentaire

Sur le point d'envoyer le présent cahier à l'impression nous recevons du laboratoire d'informatique pour les Sciences de l'homme un ouvrage des éditions du C N R S dont la matière est comprise dans le thème du présent article ; nous en donnons ici le titre et la table.

Raisonnement et méthodes mathématiques en archéologie ; C N R S Paris 1977 textes recueillis par M. Borillo, W. Fernandez de la Vega, A. Guenoche.

Introduction (M. Borillo).

Chapitre I : Introduction à la statistique et à l'analyse multivariée (P. Ihm et R. Sibson).

Chapitre II : Classification et sériation (I.C. Lerman, A. Guenoche P. Ihm).

Chapitre III : Algorithmes de sériation (S. Regnier, W. Fernandez de la Vega).

Chapitre IV : Le rôle des mathématiques dans le raisonnement en archéologie (J. Doran, C.A. Moberg, R. Whallon).