

Classification et sériation de tableaux présence-absence sur ordinateurs Apple et MS-Dos PC par le programme de Rozoy

R.R. NEWELL, J. BENNINK, Tr. CONSTANDSE-WESTERMANN et J.G. ROZOY

Résumé

La résolution correcte des problèmes archéologiques exige une adéquation des méthodes utilisées à la nature précise du problème en cause. Les tableaux d'incidences (non transformés) se prêtent à deux approches distinctes : la classification étudie les relations statiques à deux dimensions, et suppose des distributions gaussiennes des caractères et des unités, sans intervention d'une troisième cause de variation. L'algorithme basé sur le classement des barycentres des colonnes et des lignes la résoud par la diagonalisation du tableau. La sériation suppose l'intervention d'une troisième dimension (temps, espace ou autre) qui biaise fortement les distributions. La sériation repose sur le 5^e principe (de concentration des caractères) de Petrie. Le taux de concentration et la ligne de sériation, définis par les auteurs, sont des éléments objectifs qui permettent d'orienter le travail de sériation. L'algorithme de Rozoy, basé sur les "poids" des unités, et complété par une optimisation "manuelle" assistée par l'ordinateur, répond à ces nécessités en formant des tableaux triangulés.

Le programme de classification et de sériation de Rozoy est utilisable sur Apple II et Macintosh ou sur compatibles IBM en MS-DOS et Turbo-Pascal. Il comporte de multiples facilités permettant un contrôle constant de la machine et la visualisation ou impression du tableau. Son utilisation est illustrée par l'analyse approfondie du cimetière mésolithique danois de Bøgebakken, qui a nécessité l'emploi croisé de deux procédures de classification et de sériation, plus des "cluster analyses". Il faut employer plusieurs approches combinées pour analyser les causes de variation dans les données.

Abstract

Computerized techniques of pattern recognition in nominal state have been available to archaeologists for more than twenty years. Progress in the application of two such techniques, ordination and seriation, has pursued the most elegant programmed algorithm. By examining the underlying assumptions and operational goals of both techniques, we conclude that they are different, that they resolve different archaeological problems, and that the pursuit of one "optimal" algorithm is counter-productive. Therefore, we introduce a menu-driven program package, designed for Apple, Macintosh and MS-DOS personal computers, which includes both approaches plus innovations for making the analytical results mutually comparable. The effectiveness of the ROZOY NUMERICAL ORDINATION and SERIATION PACKAGE is illustrated by the integrative analysis of the Bøgebakken Mesolithic cemetery.

Key words : *ordination, seriation, Rozoy, Mesolithic, Bøgebakken.*

INTRODUCTION

On peut traiter bien des problèmes archéologiques en analysant, dans un ensemble cohérent d'unités (par exemple des tombes ou des sites), leurs caractères constitutifs, ainsi des dépôts funéraires ou des types ou caractères des céramiques. Ces *relations bidimensionnelles (Unités x Caractères)* peuvent permettre de résoudre des problèmes statiques, comme la classification et (ou) la mise en évidence de plus ou moins grandes discontinuités dans la structure des données. On peut les employer aussi pour découvrir ou évaluer la présence ou les effets d'une troisième dimension, de nature linéaire, qui peut y être inhérente, mais non immédiatement perceptible. En archéologie, la troisième dimension la plus fréquemment recherchée est le temps. On appelle ces deux approches qui concernent deux sortes différentes de problèmes archéologiques : classification et sériation (en anglais :

ordination et seriation).

La mise en ordre de tableaux bidimensionnels fut découverte assez tôt dans l'histoire de l'archéologie préhistorique et a connu depuis un développement considérable. Le plus gros en a été constitué par les comparaisons entre les efficacités respectives de quatre systèmes :

1. Présence-Absence,
2. Présence-Absence, transformées en indices ou en l'un des multiples coefficients de "distances" ou de similarité,
3. proportions,
4. proportions transformées de même.

Du fait de l'accès aux gros ordinateurs, ce sont les trois dernières approches qui ont d'abord le plus retenu l'attention des statisticiens et des programmeurs, et par suite des archéologues. Il semble qu'avec

la myriade de systèmes actuellement appliqués aux données archéologiques et les discussions sans fin quant à "la" meilleure méthode nous ayons perdu de vue la question plus fondamentale de l'appropriation des modèles mathématiques et des méthodes d'analyse aux données et problèmes archéologiques très divers. Une perception plus claire des différences entre les problèmes statiques, comme une classification, et dynamiques, comme les changements dans le temps et dans l'espace, nous impose de voir que les natures différentes des problèmes exigent la mise en oeuvre de non moins différents principes sous-jacents, présomptions et algorithmes.

Dans cet article nous ne traiterons que de l'usage des tableaux d'incidences (Présence-Absence) non transformés. Sur les inconvénients des coefficients divers, consulter Rozoy (1987, p. 367).

LA CLASSIFICATION

La classification des tableaux d'incidences utilise la forme de tableau développée pour la sériation à partir des travaux de Petrie (1899), mais avec des principes différents. Des algorithmes ont été établis par Goldmann (1974), Wilkinson (1974), Legoux (1980) et d'autres. Les principes les plus importants sont ceux de *symétrie des données* autour et au long de la diagonale X/Y du tableau (les caractères selon l'axe des X et les unités selon l'axe des Y). On présume que les apparitions, présences et disparitions des caractères fourniront, au moins en théorie, un tableau parfaitement ordonné ou "compact" qui ne peut toutefois être désigné comme "Pétrifié" (P-matrix de Kendall, 1971 b) puisqu'il ne suit pas les principes de Petrie. Le but de la classification est le rangement symétrique des caractères et des unités de telle façon que les occurrences des caractères présents soient *contiguës et rassemblées symétriquement le long et autour de la diagonale X/Y* . Les cases vides (données d'absence) sont groupées symétriquement autour du tableau compacté en haut à droite et en bas à gauche.

Les présomptions sous-jacentes à la classification sont que les occurrences de chaque caractère sont continues, que chaque unité contient des caractères représentant les critères prédominants des types étudiés, et que des processus d'une troisième

dimension, telle que le temps ou l'espace, sont absents ou subordonnés aux groupements bidimensionnels des caractères taxonomiques entre les unités en rapports. On présume encore qu'il n'y a pas de troisième dimension pour commander la présence ou l'absence des caractères dans les unités. Les cases vides de la partie remplie du tableau sont donc présumées ne représenter rien de plus que les écarts tolérables de la distribution continue dus à la conservation différentielle et (ou) à des aléas d'échantillonnage. Il est implicite dans cette méthode de présumer la tendance vers une courbe de distribution unimodale pour tous les caractères et une égale probabilité de leurs occurrences potentielles dans toutes les unités. En raison de cette présomption, il ne doit pas y avoir de grandes différences dans le nombre des caractères présents dans les unités. Enfin, on voit les unités comme des événements, et donc comme des moyennes entre les apports les plus anciens et les plus récents dans chaque unité (Dunnell, 1970).

Par un rangement des "barycentres" des caractères selon l'axe des X et une mise en ordre des "barycentres" des unités selon l'axe des Y , l'algorithme de Wilkinson (1974) aboutit à une classification fiable et efficace. Comme toutes les permutations et combinaisons possibles d'un tableau, même de taille modeste, ne peuvent être calculées pendant une vie humaine (Cowgill, 1972), il s'agit d'un procédé heuristique pour ranger les caractères (colonnes) et ordonner les unités (lignes) selon les 5 étapes suivantes :

1. calculer les moyennes des rangs ("barycentres") des caractères dans chaque colonne,
2. ranger les colonnes en fonction de ces moyennes ("barycentres"),
3. calculer les moyennes des positions des caractères dans chaque ligne,
4. ordonner les lignes en fonction de ces moyennes,
5. répéter les étapes 1 à 4 jusqu'à ce qu'aucune transposition de caractères ou d'unités ne soit plus nécessaire.

En traitant également les unités et les colonnes, la méthode des barycentres produit une assez bonne approximation d'un tableau parfait ou compacté. Celui-ci est symétrique et continu au long et autour de la diagonale X/Y (fig. 1).

Tableau de données brutes

	a	b	c	d	e	f	g	h
1	X						X	X
2		X		X	X			
3	X		X			X		
4				X			X	X
5		X			X			
6	X					X		X
7		X		X			X	
8			X		X			

Tableau ordonné

	c	f	a	h	g	d	b	e
8	X	X						
3	X	X	X					
6		X	X	X				
1			X	X	X			
4				X	X	X		
7				X	X	X		
2						X	X	X
5							X	X

Figure 1 : Classification d'un tableau d'incidences le long de la diagonale X/Y.

Numerically ordinated Unit/Attribute observations along the X/Y diagonal.

Si la classification fournit une bonne approximation d'un tableau compacté, l'absence de discontinuités dans les occurrences indique qu'il n'y a pas dans les données de groupements sous-jacents en entités taxonomiques statiques ou discrètes. Au contraire, ce résultat suggère au mieux une variation continue. Par contre, quand un tableau ordonné, symétrique autour de la diagonale X/Y, montre des discontinuités dans les ensembles d'occurrences au-dessus et

au-dessous de cette diagonale, les blocs formés par de telles discontinuités indiquent des groupements de caractères qui peuvent être interprétés comme des entités taxonomiques (fig. 2), si du moins ces anomalies ne sont pas dues à des insuffisances de conservation, de fouille ou d'analyse ou à la présence de groupes taxonomiques d'unités. En ce qui concerne la classification taxonomique, la permutableté des groupes d'unités dans la séquence est sans importance.

Tableau de données brutes

	a	b	c	d	e	f	g	h
1				X			X	X
2		X			X			
3	X		X			X		
4				X			X	X
5		X			X			
6	X		X			X		
7				X			X	X
8	X		X			X		

Tableau ordonné

	c	f	a	h	g	d	b	e
8	X	X	X					
3	X	X	X					
6	X	X	X					
1				X	X	X		
4				X	X	X		
7				X	X	X		
2							X	X
5							X	X

Figure 2 : Classification d'un tableau d'incidences le long de la diagonale X/Y.

Numerically ordinated Unit/Attribute observations along the X/Y diagonal.

Dans la pratique archéologique, même le tableau le plus parfaitement compacté, montrant une disposition symétrique continue ou discontinue, ne peut être considéré, *a priori*, comme complètement indépendant des effets de processus tridimensionnels tels que le temps ou l'espace. Le plus souvent l'archéologue doit se contenter d'un tableau dans lequel *la plus grande part* des structures peut être présumée provenir des relations statiques entre les caractères et les unités. *Il est donc toujours plus sage de tester pour la présence et le niveau d'influence de tels processus tridimensionnels.* Pour cela, il faut procéder à une sériation.

LA SÉRIATION

La sériation, quoique distincte de la classification, peut aussi être opérée sur un tableau d'occurrences pour identifier "une meilleure dimension *prédominante* qui explique *la plus grande partie* de la variation dans les unités" (Ester, 1981, p. 503, souligné par lui-même). La méthode est basée sur les deux premiers et les deux derniers des *5 principes de Petrie* (1899) :

1. superposition effective de tombeaux ou de sépultures (...),
2. développement ou dégradation des formes (...),
3. groupement statistique par ressemblance des proportions,
4. groupement des types similaires, et jugement par styles,
5. dispersion minimale de chaque type, *concentrant* les exemples extrêmes (souligné par nous).

Pour la sériation des tableaux d'incidences, il faut donner la priorité absolue au cinquième principe, en modifiant l'ordre des unités et d'elles seules.

Contrairement à la classification, la sériation doit opérer selon la présomption qu'il existe des différences dans les premières apparitions des caractères selon la dimension dominante et que *les distributions de ces caractères sont biaisées plus ou moins fortement par rapport à la structure d'ensemble du tableau*, par suite de l'action de la dimension dominante. La sériation est donc une mise en ordre des unités, considérées comme des événements de durée ou surface comparable, et dont la séquence est déterminée en premier lieu par les propriétés de leurs caractères selon la troisième dimension.

Les présomptions sous-jacentes à la sériation sont que les unités contiennent des caractères qui décrivent la troisième dimension dominante selon laquelle les unités seront ordonnées. De plus, chaque unité est conçue comme un ensemble clos (Worsaae, 1844), qui doit être daté par son dernier caractère (le plus récent s'il s'agit du temps). En outre, *les occurrences des*

caractères sont conçues comme distribuées de façon continue ou subcontinue et la sériation arrangerait les unités de telle façon que leur ordre s'approche aussi près que possible de la structure continue de la dimension dominante (Dunnell, 1970; Ester, 1981). En d'autres termes la sériation rend compte de *la plus grande part* de la variabilité entre les unités (Ester, 1981), toutes les déviations admissibles autour de cette structure, tout au moins pour les sites d'habitat, étant dues à la conservation différentielle et aux irrégularités d'échantillonnage.

Lors d'une sériation chronologique, les autres sources de variation dans l'ensemble d'unités doivent entraîner des différences de faible ampleur par rapport à celles révélant le plus petit intervalle de temps que l'on espère distinguer valablement (Cowgill, 1972). La finesse avec laquelle le temps peut être estimé dépend de la quantité de variabilité qui apparaît dans l'ensemble d'unités que l'on analyse, c'est-à-dire de la nature et de la qualité des données que l'on a su extraire du terrain et analyser.

Toutefois, Dunnell (1970) souligne que le passage de l'observation des résultats de la sériation à la détermination de la chronologie, de la variation spatiale, etc., est *par nature un raisonnement d'inférence*. En d'autres termes, toute sériation réussie n'est pas nécessairement une sériation chronologique. C'est plutôt un reflet de la meilleure mise en ordre des unités selon les propriétés tridimensionnelles des caractères existants dans les unités ou choisis par l'archéologue pour les représenter. Kuzara, Mead et Dixon (1966) ont démontré que la sériation peut arranger les unités selon différentes sortes de dimensions actives, par exemple les cadences de disparitions et remplacements de caractères dans l'espace. Un ordre issu d'une classification ou d'une sériation peut toujours être inversé, et la direction d'une sériation doit être établie par des moyens indépendants, tels que la stratigraphie, le ^{14}C , la cartographie, etc.

La sériation comporte la présence inévitable de cases vides dans les colonnes, au-dessous de la diagonale X/Y. Ces discontinuités proviennent de choix culturels. Elles sont particulièrement marquées pour les tombes. En effet, chaque tombe ne contient qu'un nombre limité d'objets et de caractères extraits de ceux utilisés par la société vivante correspondante, et choisis tombe par tombe en fonction d'autres principes d'organisation que ceux qui commandent la sériation : distinctions entre sexes, rang social, symboles du lignage, appartenance à une bande ou à un clan, etc. Il s'y ajoute en moindre nombre les données manquantes dues à la conservation différentielle, aux défauts de fouilles et (ou) d'échantillonnage. Nous développerons plus loin les méthodes propres à réduire le nombre de cases vides à ce minimum incompressible.

L'ALGORITHME DE CONCENTRATION

Développant le travail de Kendall (1963, 1971 a-c), l'un de nous (Rozoy, 1987) a démontré qu'un *algorithme basé sur les poids des unités* selon la séquence qui suit procure l'application la plus satisfaisante du 5ème principe de Petrie.

1. Calculer pour chaque colonne la moyenne des rangs des occurrences du caractère.
2. Pour une ligne, *calculer le poids* de chaque occurrence comme la différence positive (au-dessus) ou négative (au-dessous) avec cette moyenne.
3. Le *poids de la ligne* est la somme de tous les poids des caractères présents dans la ligne. Pour éviter des déformations il doit être pondéré selon le nombre de caractères présents dans la ligne.
4. *Calculer les poids des deux lignes voisines et les comparer* avec celui de la ligne en question.
5. Si la valeur absolue du poids est supérieure à celle de la ligne inférieure, ou inférieure à celle de la ligne supérieure, *permuter* la ligne avec celle-ci.
6. Si une permutation a été effectuée, recalculer les moyennes des rangs pour les colonnes où des modifications ont eu lieu.

7. Répéter les étapes 2 à 6 pour chaque ligne de haut en bas, puis de bas en haut.
8. Répéter les étapes 1 à 7 jusqu'à ce que l'on ne trouve plus de permutation à effectuer.

Par suite de ce travail, *les données négatives (absences) sont groupées asymétriquement*, largement au-dessus de la diagonale X/Y et accessoirement au-dessous, mais ces dernières sont distribuées en partie sous les dernières occurrences et en partie insérées entre celles-ci. Le programme approche ainsi la concentration la plus économique des occurrences, essentiellement au-dessous de la diagonale X/Y , plaçant les premières occurrences de chaque caractère (qui sont aussi celles des derniers caractères des unités) le long d'une ligne oblique juste au-dessus de cette diagonale. Nous appelons cette ligne : *ligne de sériation*. Elle est l'expression synthétique de l'action de la troisième dimension dominante sur les rapports entre les deux dimensions du tableau. Sa forme permet de juger la valeur de la sériation obtenue. Nous exposerons plus loin les corrections à lui apporter "à la main" en fonction des particularités heuristiques de l'algorithme. La figure 3 présente un exemple artificiel très simplifié d'un tableau sérié.

Tableau de données brutes

	Y	a	b	c	d	e	f	g	h	
1	X							X	X	
2		X	X	X	X			X		
3	X		X				X			
4			X	X			X	X	X	
5			X		X					
6	X		X				X		X	
7	X	X		X						
8			X		X					X

Tableau sérié

	Y	c	f	a	h	g	d	b	e	
8	X	X								
3	X	X	X							
6	X	X	X	X						
1	.	X	X	X	X					
4	X	X	.	X	X	X				
7	.		X	.	.	X	X			
2	X		.	X	X	X	X	X		
5	X									X

Figure 3 : Sériation Unités - Caractères juste au-dessus et sous la diagonale X/Y .

Seriated Unit/Attribute observations just above and below the X/Y diagonal.

L'inconvénient évident de notre algorithme est celui de toutes les méthodes itératives : *il décide sur un plan local* (deux ou trois unités) sans pouvoir, malgré la référence à la moyenne, considérer le tableau dans son ensemble. Des obstacles (pièges) sont rencontrés, constitués de minima locaux que l'algorithme ne peut franchir ou contourner. Ces obstacles sont d'autant plus nombreux que le tableau est plus grand. La parade classique est de procéder à plusieurs essais avec des points de départ différents par tirage aléatoire des unités. Mais nous disposons d'un critère objectif de jugement basé sur le 5ème principe de Petrie : *le taux de concentration des caractères* (dont le calcul est exposé plus loin) permet de procéder par

optimisation "à la main". En effet, l'algorithme, puisqu'il juge au plan local, sacrifie parfois la concentration maximale (que l'on peut obtenir par appréciation d'une vue d'ensemble) à une amélioration purement locale. Non seulement il ne parvient pas toujours à la concentration maximale, mais en ce cas il la dégrade lorsqu'on la lui fournit. Des informaticiens de premier ordre travaillent actuellement à une méthode de vue d'ensemble utilisable par les ordinateurs électriques, mais on ignore encore si elle sera possible. Il faut donc, en attendant, mettre en action l'ordinateur biologique cérébral, infiniment plus discriminant, pour parfaire l'oeuvre de la machine.

L'OPTIMISATION "MANUELLE"

L'optimisation "à la main" est, en fait, réalisée par le cerveau humain aidé de la vision. L'ordinateur-machine exécute les translations de lignes et de colonnes décidées par l'opérateur, ce qui est une aide importante. Il calcule et affiche autant de fois que nécessaire les éléments de jugement. On peut ainsi concevoir, afficher à l'écran et imprimer plusieurs dizaines de variantes du tableau en une seule séance, et approcher peu à peu de la solution optimale. Lors de la sériation des tombes celtes du Mont Troté (130 lignes x 128 colonnes), l'un de nous (Rozoy, 1987) a commencé avec un taux de concentration de 7-8 % et a obtenu par un emploi judicieux du présent programme un maximum satisfaisant (pour ce site) de 23,64 %.

Le but recherché est de concentrer les caractères dans les colonnes et il est donc *nécessaire de disposer d'un critère* à ce sujet. Nous définissons comme "espace enclos" le total des nombres de cases incluses entre les premiers et derniers caractères (ceux-ci compris) dans les colonnes. Le *taux de concentration des caractères* est le rapport entre le nombre de cases occupées et l'espace enclos. Il constitue une mesure objective permettant de juger et de guider le travail, pour ne laisser dans l'espace enclos que les cases vides qui sont la conséquence des absences de caractères normalement attendues comme dit ci-dessus. Comme ce taux est défini en pourcentage, il est indépendant des dimensions du tableau et du nombre d'incidences et peut donc être comparé à ceux obtenus soit dans le même ensemble (éventuellement par classification ou en ayant éliminé ou ajouté des unités ou des caractères), soit dans des ensembles plus ou moins comparables. Ceci est une amélioration considérable des index de concentration de Kendall (1963) et de Doran (1971) qui ne prétaient pas à des comparaisons extérieures.

L'optimisation "manuelle" se base sur la perception de la ligne de sériation comme la meilleure vue d'ensemble de la combinaison des deux dimensions de la matrice et surtout de l'influence de la dimension dominante sériante sur ces deux éléments. *La ligne de sériation doit être sensiblement rectiligne, continue, sans indentations* et, à l'exception du plateau initial qui dérive de l'absence des tombes précédentes, *parallèle à la diagonale* et peu au-dessus d'elle. Les unités formant des indentations doivent donc être remontées jusqu'à ce que leur dernier caractère soit sur la ligne de sériation. Si la ligne de sériation présente des déformations, en particulier des bosses, il faut s'efforcer de les réduire en faisant remonter les unités inférieures et descendre celles qui forment la bosse. Enfin, des plages importantes de cases vides, surtout dans plusieurs colonnes et lignes consécutives, don-

nent l'indication de remonter fortement des groupes entiers de lignes inférieures (souvent celles contenant les dernières occurrences de caractères). Dans tous les cas on s'efforce de purger l'espace enclos du plus grand nombre possible de cases vides.

Rétablir l'ordre des premières apparitions des caractères (en déplaçant les colonnes) après certaines de ces modifications (en particulier après les remontées lointaines de groupes d'unités) ne change rien à la concentration ni aux calculs opérés par la machine, mais c'est nécessaire à une perception claire des ensembles par l'oeil et le cerveau humains, et donc à la poursuite du travail de sériation "manuelle".

Des anomalies importantes : segments verticaux ou horizontaux, si elles sont irréductibles, signalent *des discontinuités* ou des irrégularités dans la fourniture des données. A la limite, on parvient aux structures de blocs montrées par Legoux (1980) et par Leredde et Djindjian (1980), qui proviennent de la présence de deux ou plusieurs ensembles, si du moins il ne s'agit pas de manques accidentels : tableau trop pauvre ou gisement amputé. Les tableaux trop pauvres en unités, ou amputés, fournissent plutôt des segments horizontaux, ceux trop pauvres en caractères (site pauvre ou analyse insuffisante) donnent des segments verticaux. Les meilleurs résultats sont obtenus lorsque le nombre d'unités est au moins égal à celui des caractères.

La sériation automatique suivie de la sériation "manuelle" (cérébrale) assistée par l'ordination *combine les avantages des deux systèmes* : rapidité et examen ligne à ligne de la machine, vue d'ensemble assurée par le cerveau humain, application de principes directeurs fondés sur la théorie archéologique, contrôle constant par le calcul objectif des taux de concentration. Cela permet d'éviter à la fois toute subjectivité cérébrale et les pièges des sériations automatiques. On découvre ainsi, dans la "materia prima" des données brutes, la "materia secunda" des dimensions cachées, et tout particulièrement la chronologie des tombes.

LE TABLEAU SERIE

L'exécution réussie de notre algorithme fournit un tableau sérié comportant l'association de *quatre types de distribution des occurrences* :

1. caractères présents de façon (dis)continue du sommet au bas du tableau,
2. caractères présentant une distribution discontinue,
3. caractères présentant une distribution continue de leur première occurrence sur la ligne de sériation jusqu'à leur dernière occurrence, bien au-dessus de l'axe des X,
4. caractères qui paraissent continus, mais dont la

séquence est tronquée par la terminaison des données et du tableau (axe des X).

Le premier type correspond à des caractères qui sont présents tout au long de la dimension dominante révélée par la sériation (dans la figure 3, caractère c). Leur présence soit dépasse l'espace contenu dans ce groupe d'unités, soit témoigne de variations secondaires ou non-dimensionnelles. Ces caractères sont groupés à gauche du tableau sérié et on devrait les exclure.

Le second type de distribution des caractères (discontinue) indique, par la présence de cases vides, *un mélange de propriétés non-dimensionnelles* dans les données. Les plus importantes sont la richesse différente des unités (qui entraîne l'impossibilité d'utiliser des coefficients de distances ou de similarité) et l'inégale durée (ou espace) d'utilisation des caractères (qui entraîne la non-validité des algorithmes de classement par les barycentres; Rozoy, 1987). Il y a aussi les caractères spécifiques pour femmes ou hommes, les caractères fonctionnels exclusifs les uns des autres, etc., et encore la possible présence dans les données d'autres structures dimensionnelles de rang inférieur, la conservation différentielle et les irrégularités d'échantillonnage, pertes à la fouille etc. (dans la figure 3, caractères a, h, g).

Les caractères dont la limite inférieure des occurrences suit de près la ligne de sériation (souvent ceux de 3ème type, à distribution continue, ou presque) sont ceux qui sont *les plus sensibles à l'action de la 3ème dimension dominante* (dans la figure 3, les caractères f, d, b et e). Si la dimension dominante s'avère être le temps, ce sont les caractères les plus sensibles chronologiquement (parce qu'on les a employés moins longtemps). On devra évidemment exclure ces caractères dominants si l'on veut ensuite étudier des niveaux inférieurs de dimensionnalité dans le tableau. La plus ou moins grande asymétrie dans la disposition des cases vides est due à des propriétés différentes dans la 3ème dimension ou dans d'autres dimensions, ou à d'autres phénomènes culturels ou fonctionnels tels ceux cités ci-dessus. Ceci aussi doit être vérifié et les résultats utilisés pour une nouvelle série d'examen des données, de formulation d'hypothèses et de tests.

VERIFICATION ET FORMULATION DE NOUVELLES HYPOTHESES

Pour une présentation exhaustive de la théorie et des méthodes de cette sériation et d'autres, le lecteur voudra bien consulter Rozoy (1987) et Rozoy *et al.* (soumis). Le premier texte est un exemple typique du passage des considérations théoriques et méthodologiques générales à la programmation réussie de micro-ordinateurs par les archéologues eux-mêmes. Ceci était nécessaire pour qu'ils parviennent

à choisir ou créer l'approche analytique appropriée au problème archéologique en cause.

Les prémisses théoriques et les contraintes opérationnelles de la classification linéaire et de la sériation tridimensionnelle sont différentes. Pour cette raison nous avons inclus les deux algorithmes dans le programme. Mais nous devons souligner qu'*ils ne peuvent être utilisés de façon interchangeable*. Des sériations chronologiques, ou fonction d'une autre troisième dimension, ne peuvent être valablement obtenues qu'en attribuant la priorité absolue au 5ème principe de Petrie, ce que font les calculs et comparaisons des poids des unités. Nous avons démontré (Rozoy, 1987) que les résultats de cette méthode sont meilleurs que ceux de la classification simple, bidimensionnelle, de Mennessier (1973), Goldmann (1974), Wilkinson (1974) et Legoux (1980), peut-être plus esthétiques, mais reposant sur des bases malsaines, ou plus exactement appliquées à un domaine qui n'est pas le leur.

La classification par les barycentres produit une bonne approximation des principes de linéarité et symétrie autour de la diagonale X/Y . Mais *l'analyste sera bien avisé d'effectuer ensuite une ou plusieurs sériations sur le tableau classifié*, pour rechercher d'éventuels processus directionnels (dimensions) insoupçonnés dans son tableau présumé statique, bidimensionnel. Inversement, lors de la recherche de dimensions (temporelle ou autres) par la méthode de sériation par poids des unités, il sera avantageux *d'expérimenter un algorithme de classification sur le tableau final sérié*. On peut alors obtenir une idée du rôle quantitatif du processus directionnel principal que l'on étudie. Un tel essai fournira à l'analyste une représentation théorique idéale du maximum de compaction inhérent aux données de son analyse. Cela lui permettra aussi d'évaluer le résultat de sa sériation en termes comparables à d'autres. Rester sensible à la hiérarchie des multiples sources de variation inhérentes à tout tableau de données permet à l'analyste de découvrir et d'interpréter les groupements dominants dans une classification ou une sériation, de mesurer leur part de contribution dans le tableau total et donc de concevoir de nouvelles façons de tester et mesurer les niveaux inférieurs de variabilité dus à de nouveaux groupes.

DESCRIPTION OPERATOIRE

Le programme de classification et de sériation de Rozoy consiste en trois blocs de programmes liés :

1. Entrée et édition des données.
2. Traitement du tableau.
3. Mesure et présentation des résultats.

Les disquettes contiennent aussi un programme de description et le mode d'emploi. L'exécution de cet ensemble est illustrée par un diagramme reproduit à la figure 4.

Commençant avec un taux de concentration de 53,15 %, la concentration a porté ce taux à 65,53 %. Une optimisation "à la main" a élevé cette mesure à un maximum de 69,64 %, livrant le tableau sérié final (fig. 6).

La figure 6 peut s'analyser comme suit : le caractère 1, sis à la gauche de la sériation, se retrouve de façon continue à travers tout le tableau. De ce fait, il ne nous apprend rien sur la 3ème dimension. Les caractères 2 à 10 et 24 à 28 offrent quelque groupement bidimensionnel, mais leur ligne de terminaison (de l'unité 17 à l'unité 15) n'est pas parallèle à la ligne de sériation. Leur groupement représente donc un mélange des effets de propriétés taxonomiques et peut-être la présence d'un processus additionnel d'une troisième dimension. Les caractères 11 à 19, qui n'apparaissent que dans l'unité 3 (tombe 19c), forment une ligne horizontale non diagnostique. Enfin,

la ligne inférieure de l'espace enclos des caractères 20 à 23 et 29 à 56 est parallèle à la ligne de sériation, indiquant que ces caractères sont les plus sensibles à la dimension dominante qui détermine la sériation.

Comparer l'ordre des tombes aux six relations de superposition stratigraphique observées montre que la dimension dominante n'est pas la chronologie. Il n'y a pas non plus de corrélation avec les dimensions sociales que sont l'âge, le sexe ou le statut social, ni avec les orientations des tombes. L'ordre des tombes est par contre en corrélation significative avec leur disposition topographique dans le cimetière par rapport à celle placée en premier par la sériation : la tombe 3. Les distances linéaires de tête à tête produisent un coefficient de rang de Spearman (Siegel, 1956) de 0,699, les distances entre milieux des tombes, de 0,750 (pour les deux coefficients, $p < 0,020$). Malgré ce résultat satisfaisant, on ne doit pas perdre de vue

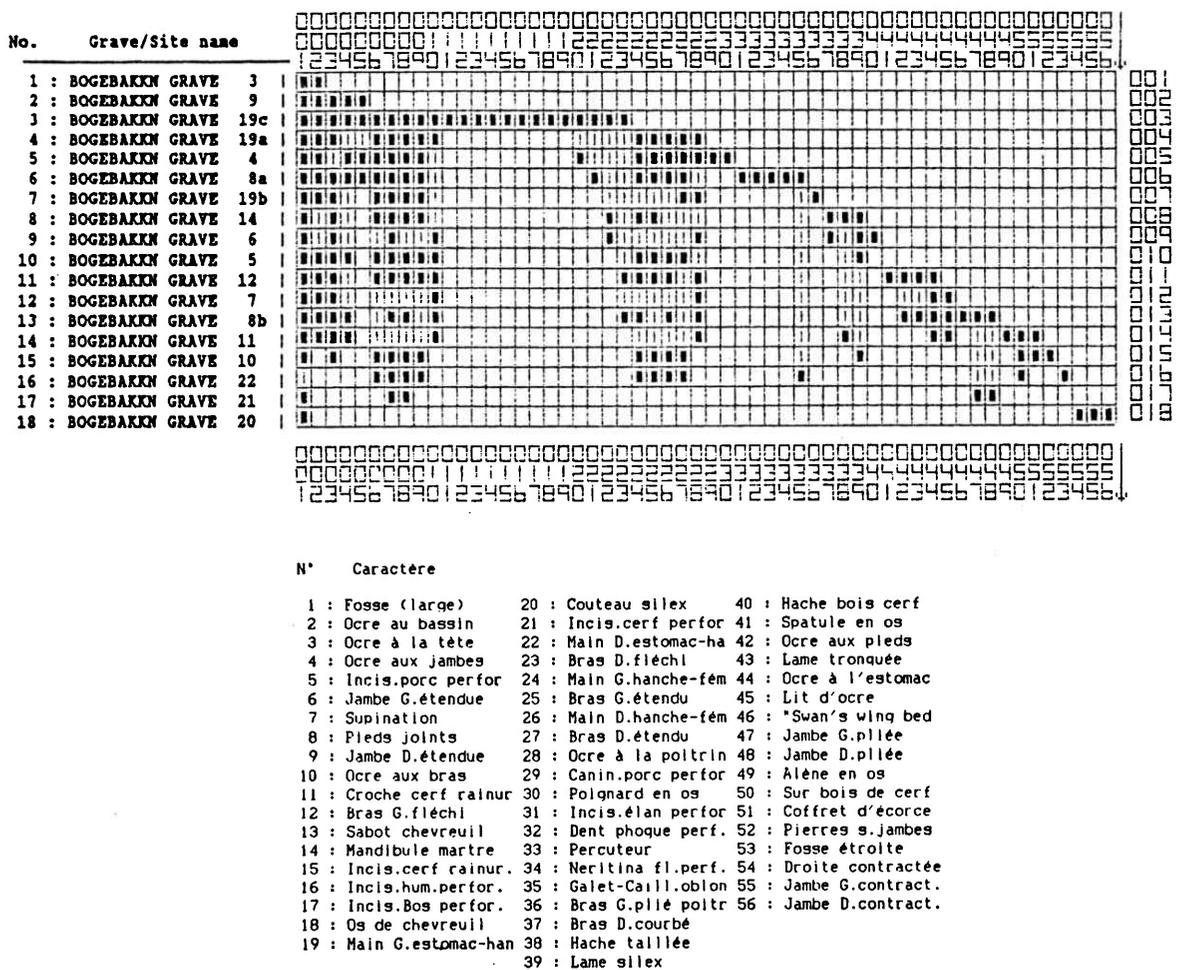


Figure 6 : Tableau sérié final du cimetière mésolithique de Bøgebakken.
Final seriated data-matrix of Bøgebakken mesolithic cemetery.

que la résolution quantitative de la sériation est basse. Si nous soustrayons du taux de concentration final celui des données brutes, la différence est de 16,49 %, contribution très modeste à la variabilité des groupes. Au vu de ces résultats et interprétations, il est clair qu'une analyse de classification des relations statistiques, bidimensionnelles, qui apparaissent dans la partie moyenne de la sériation, est nécessaire.

Comme 25 des 56 caractères ne sont présents qu'une seule fois, ce qui ne permet aucune classification, il faut procéder à un examen critique des données. Une analyse de fréquence du nombre de tombes dans lesquelles un caractère défini est présent produit une distribution trimodale, qui est confirmée

par un test du χ^2 simple. Les modes sont 1-2 ($n = 35$), 3-13 ($n = 20$) et 17 ($n = 1$). Le dernier mode représente la fosse de tombe, qui comme nous l'avons vu traverse tout le tableau et est donc totalement non diagnostique. Sur ces bases, les caractères des modes 1 et 3 ont été retirés de la liste pour la classification à deux dimensions. Ce pas de l'analyse est justifié parce que 30 des 35 caractères du mode 1 déterminaient les propriétés tridimensionnelles de la sériation. La tombe 3, qui ne possède plus qu'un seul caractère, doit aussi être retirée. Le tableau restant de 16 x 20 a été classifié et a livré le résultat présenté dans la figure 7, avec un taux de concentration de 74,37 %.

No.	Grave/Site name	
1	: BOGEBAKKN GRAVE 9	
2	: BOGEBAKKN GRAVE 7	
3	: BOGEBAKKN GRAVE 11	
4	: BOGEBAKKN GRAVE 8b	
5	: BOGEBAKKN GRAVE 12	
6	: BOGEBAKKN GRAVE 19c	
7	: BOGEBAKKN GRAVE 6	
8	: BOGEBAKKN GRAVE 19a	
9	: BOGEBAKKN GRAVE 5	
10	: BOGEBAKKN GRAVE 4	
11	: BOGEBAKKN GRAVE 19b	
12	: BOGEBAKKN GRAVE 8a	
13	: BOGEBAKKN GRAVE 10	
14	: BOGEBAKKN GRAVE 22	
15	: BOGEBAKKN GRAVE 14	
16	: BOGEBAKKN GRAVE 21	

N°	Caractère
1	: Lit d'ocre
2	: Ocre à l'estomac
3	: Ocre aux bras
4	: Ocre à la poitrine
5	: Ocre aux jambes
6	: Ocre au bassin
7	: Ocre à la tête
8	: Sur bois de cerf
9	: Main G.hanche-fémur
10	: Main D.hanche-fémur
11	: Bras D.plié
12	: Bras G. étendu
13	: Jambe G.étendue
14	: Jambe D.étendue
15	: Supination
16	: Pieds joints
17	: Lame silex
18	: Bras D.étendu

Figure 7 : Tableau classifié du cimetière mésolithique de Bøgebakken.

Ordinated data-matrix of Bøgebakken mesolithic cemetery.

A partir de la classification discontinue montrée par la figure 7, il semble que la plus grande part de la variabilité présente dans les relations unités x caractères réside en leur état statique, bidimensionnel, et que les parties distinguées représentent des entités taxonomiques ou des divisions sociales au sein de la société de Bøgebakken. Ces parties sont les unités 1-5 (représentées par les caractères 1-17), les unités 6-9 (représentées par les caractères 3-20), les unités 10-12 (représentées par les caractères 4-19) et les

unités 13-16 (représentées par les caractères 8-20). Si nous examinons les rangs des unités (tombes) en fonction de la stratigraphie observée, il est clair que la classification est indépendante de la chronologie, bien que sa concordance avec la chronologie relative des superpositions apparaisse un peu meilleure que celle apportée par la sériation ci-dessus. En testant les groupes pour des dimensions sociales indépendantes (âge, sexe et rang social), nous pouvons obtenir une diagnose inférentielle de chacun. Ces tests probabilis-

tes, confirmés par des “cluster analyses”, ont révélé que les compositions des groupes sont toutes indépendantes de l’âge et (ou) du sexe, qui sont distribués également entre les groupes. Les résultats de ces “cluster analyses”, comparés aux séquences des tombes produites par la classification et par la sériation, montrent que la séquence de la sériation est pleinement conforme aux résultats des “cluster analyses”. La concordance de la séquence de la classification est pour la plus grande part bonne, mais manifeste quelques divergences, figurées en caractères italiques dans le tableau 1 :

	Cluster analyses	Sériation	Classification
Tombe	3	3	–
”	9	9	9
”	19c	19c	7
”	19a	19a	11
”	4	4	8b
”	8a	8a	12
”	19b	19b	19c
”	14	14	19a
”	6	6	6
”	5	5	5
”	12	12	4
”	7	7	19b
”	8b	8b	8a
”	11	11	10
”	10	10	22
”	22	22	14
”	21	21	21
”	20	20	–

Tableau 1 : Bøgebakken : comparaison des méthodes d’analyse.

Comme il sera précisé ci-après, toutes ces divergences proviennent de données incomplètes (documentation manquante) et d’irrégularités d’échantillonnage, menant à des associations fallacieuses. Pour le moment, la conclusion de ce qui précède est que les unités distinguées par la sériation et les “cluster analyses” et largement confirmées par la classification présentent un groupement et un ordre de rangs en termes de statut social (cf. Constandse-Westermann et Newell, 1989; Newell *et al.*, en prép.). Il semble que l’on puisse déduire de cette partition que l’ordre de la sériation suit la disposition concentrique des unités taxonomiques, partiellement retrouvée par la classification, dans le plan du cimetière de Bøgebakken.

Comparant les analyses de sériation et de classification, nous trouvons que la sériation améliore la mesure de concentration des données brutes de 16,49 %. La classification, partant d’un niveau presque identique de 56,49 %, aboutit à 74,37 %, montrant une amélioration du tableau de 17,88 %. Ce résultat implique que 25,63 % (100 % – 74,37 %) demeure inexpliqué. En soustrayant le résultat final de la sériation de celui de la classification, nous pouvons établir

le surplus de variation expliqué par la classification par rapport à la sériation, soit 4,73 %. Partant de valeurs presque identiques (53,15 % pour la sériation et 56,49 % pour la classification), la sériation produit une amélioration de 16,49 % contre 17,88 % pour la classification. *La classification est la plus efficace des deux approches*, l’emportant de 1,39 %. Ces résultats sont-ils indépendants ou reflètent-ils plutôt, et jusqu’à quel point, deux expressions des mêmes dimensions et propriétés reliées? Ceci doit être établi ensuite.

Pour déterminer la dépendance ou indépendance entre les résultats de la classification et de la sériation, on doit partir de l’approche la plus efficace, en l’occurrence la classification, et réordonner les unités sériées en conformité avec la séquence la plus réussie de la classification. C’est là une condition pour rendre comparables les deux tableaux et valider l’analyse évaluative suivante. En divisant la mesure de concentration de la nouvelle sériation par la mesure de concentration supérieure de la classification parallèle, on obtient une approximation de la proportion de dépendance et d’isomorphisme des deux épreuves. Dans cette analyse de Bøgebakken, le rapport est de $65,29/74,37 = 87,79\%$, indiquant que le résultat de la sériation est répété et réexprimé par 87,79 % du groupement taxonomique révélé par la classification. Donc les 12,21 % (100 % – 87,79 %) de variation restante dans la classification doivent être expliqués par d’autres moyens. Pour démontrer la nature subordonnée et dépendante des résultats de la sériation, nous avons retourné l’expérience précédente en changeant l’ordre des unités dans la meilleure classification pour le conformer à celui de la sériation finale. Le résultat a produit une mesure de concentration de 69,16 %. Divisant la mesure de la sériation finale par cette mesure heuristique, soit $69,16\% / 69,94 = 99,31\%$, nous devons conclure que *la sériation est isomorphique à toute la variabilité résolue par la classification dominante, sauf 0,69 %*. Donc la sériation, subordonnée, dévie de la classification, dominante, de 0,69 %.

En définitive, on ne peut pas chercher l’interprétation de la variation inexpliquée restante (au maximum 25,63 %) dans l’intervention indépendante d’une troisième dimension qui aurait été révélée par la sériation. Donc, nous poursuivons notre explication en confrontant nos résultats avec les données et en recherchant perturbations du site, conservation différentielle, pertes éventuelles à la fouille, comptes rendus incomplets et hasards d’échantillonnage. Le réexamen des unités de la meilleure classification montre que 14 des 51 cases vides, qui contribuent aux 25,63 % de variabilité inexpliquée, sont dues aux perturbations, à des problèmes de conservation ou (et) de publication incomplète. Les tombes constituant ce

biais sont les suivantes, chacune avec son nombre (N) de données manquantes : tombe 6 (7), tombe 19b (4), tombe 7 (2) et tombe 4 (1). Donc 14 / 51 de la variabilité inexplicée par la classification est due à ces facteurs, soit 7,04 %. Cela laisse 18,59 % de variabilité restant à expliquer en termes culturels.

CONCLUSIONS

La conscience grandissante des différences entre les formulations statiques et dynamiques des problèmes archéologiques rend logique que l'on développe des approches analytiques qui reflètent un équilibre optimal entre la nature et la qualité des données d'une part, et la mise en oeuvre des présomptions et principes sous-jacents les plus adaptés de l'autre. Notre sensibilité grandissante à la nature polythétique et systématique de la variabilité culturelle doit nous prévenir contre la tendance à trop nous reposer sur le simple choix binaire d'une approche ou d'une autre, dans l'espoir qu'à elle seule l'approche choisie fournira une résolution complète ou même maximale des structures inhérentes aux données. Donc, pour explorer la hiérarchie potentielle des causes de la variabilité culturelle observée, nous insistons pour que le modèle et la méthode, soit de classification, soit de sériation, qui paraissent les plus appropriés archéologiquement, soient étudiés et appliqués en premier et qu'on leur donne le plus grand poids lors de l'interprétation des résultats. Toutefois, il nous faut aussi recommander que ces résultats soient contrôlés par l'autre modèle et l'autre méthode pour obtenir un contrôle analytique sur d'éventuelles dimensions et résidus d'ordre inférieur, qui est une aide à la formulation de nouvelles questions archéologiques. C'est pour cela que nous avons inclus les deux algorithmes sur les disquettes du programme. Enfin, les auteurs sont fermement convaincus qu'avec la conclusion réussie de l'analyse, comme nous l'avons testée expérimentalement, la structuration des données observées aura procuré de meilleurs aperçus sur la diagnose et l'interprétation des données archéologiques existantes.

ANNEXE : LE PROGRAMME ROZOY DE CLASSIFICATION ET SERIATION

Les disquettes

Les disquettes du programme Rozoy de classification et sériation permettent d'enregistrer, trier, puis classer et (ou) sérier des unités analytiques et leurs caractères, disposés dans un tableau de données "Présence-Absence" à deux dimensions. Écrit en Apple Pascal U.C.S.D. pour les séries Macintosh et Apple II et en Turbo-Pascal pour les machines compatibles MS-DOS, le programme peut traiter des ta-

bleaux allant jusqu'à 140 lignes pour les unités sur les ordonnées et 140 colonnes pour les caractères sur les abscisses. Vu l'étendue des applications possibles, les unités (en ordonnées) sont désignées comme "Tombe/Site", tombe désignant les éléments liés à une seule personne.

Pour la classification comme pour la sériation, les données peuvent être utilisées à l'état brut, ou après mélange aléatoire, ou après préclassement. Ce préclassement est opéré soit par le programme de classification, soit archéologiquement "à la main", et il existe pour la sériation une possibilité de fixer arbitrairement les n premières et dernières lignes pour orienter le travail de la machine et limiter le piégeage dans les minima locaux. L'établissement d'une bonne concentration linéaire (sériation) par les comparaisons des "poids" des unités est opéré par la procédure "Concentrer" qui toutefois, pour les raisons heuristiques exposées ci-dessus, ne parvient généralement pas au résultat maximal. Il faut alors appliquer une optimisation "manuelle" utilisant les options qui permettent de déplacer à volonté les lignes ou (et) les colonnes. Le programme permet à l'analyste de voir son tableau classifié ou sérié interactivement sur l'écran et de connaître la mesure du tableau. Celle-ci indique l'espace enclos, les nombres de cases vides et occupées dans cet espace et le taux de concentration des caractères.

Les options de transposition ou permutation des unités ou des caractères, de visualisation, de mesure et d'impression du tableau facilitent à l'analyste la recherche de la forme idéale de celui-ci. Le programme offre aussi la possibilité de lister, corriger, modifier, transposer, ajouter ou retrancher aux listes des caractères ou (et) des unités. L'option Entrées-Sorties permet de sauvegarder ou changer de nom et de description les tableaux ainsi modifiés. Ce robuste programme comporte aussi une série de contrôles internes et interactifs qui assurent toute conformité entre la machine et l'analyste.

Le programme peut produire les sorties suivantes sur imprimante :

1. la mesure du tableau, décrite ci-dessus,
2. le tableau (ordonné ou sérié), dans l'état du moment, les occurrences des caractères étant figurées par un pavé (sur MS-DOS) ou par une lettre (sur Apple). On peut aussi figurer les cases vides dans l'espace enclos par un trait vertical, qui facilite la perception de la continuité,
3. une liste alphabétique ou par numéros des caractères,

4. une liste alphabétique ou par numéros des unités,
5. une liste des unités (tombes/sites) avec les caractères qu'elles contiennent,
6. une liste des caractères avec les unités qui les contiennent.

Disponibilité du système

- Ecrit à l'origine (en 1981) en Basic Applesoft, avec le texte en français, par le Dr Rozoy, le programme fut ensuite transcrit par lui-même (en 1982-1983) en Apple Pascal U.C.S.D. pour l'Apple II. Il est aussi utilisable sur Macintosh. Il est disponible dans ces versions (améliorées depuis) en français et en anglais. Une version MS-DOS avec texte en anglais a été réalisée par R. Newell et Tr. Constandse-Westermann, puis convertie en Turbo-Pascal 4.0 par J. Bennink, qui améliora les modes de calcul et le graphisme, ajouta les menus, les écrans d'aide et l'affichage du tableau, et simplifia les procédures d'entrées et sorties, ce qui les rend plus souples. La version MS-DOS Turbo-Pascal du programme de classification et sériation de Rozoy est disponible en français et en anglais comme MS-DOS EXE file et comme Turbo-Pascal objet file. Les auteurs les fournissent sur disquettes 5.25 ou 3.5 pouces formatées à 360 Kb ou 720 Kb.

Matériel nécessaire

La version Apple II exige une machine avec deux lecteurs de disquettes, un minimum de mémoire de 64 Kb et une imprimante en ligne, avec au moins quatre fonts internes, et des caractères de 9 à 17 cpi. Les logiciels consistent en deux fichiers de programme et un fichier documentaire, comprenant ensemble 22 000 octets, sur un disque, et un fichier "System Library" de 40 000 octets contenant la bibliothèque de quatre fichiers, à placer sur la disquette Apple 1:

La version MS-DOS exige une machine XT ou AT compatible IBM avec au moins un lecteur de disquettes et un minimum de mémoire de 256 Kb, une imprimante en ligne avec au moins quatre fonts internes et des caractères de 9 à 17 cpi. Les logiciels consistent en un fichier de programme, un fichier d'aide, un fichier Config, un fichier documentaire, un fichier d'enregistrement et un fichier automatique INSTALL, le tout comprenant 170 000 octets sur un seul disque. Enfin, le programme a été adapté pour l'emploi optionnel d'une souris compatible Microsoft. Le programme est fourni comme un seul fichier compilé MS-DOS EXE et peut être utilisé par les MS-DOS 3.0 et plus élevés, sans exiger l'usage de facilités Turbo-Pascal.

REMERCIEMENTS

Les auteurs tiennent à exprimer leur gratitude au Dr Kielman, diplômé de l'Université de Groningen,

membre du Projet collectif des systèmes AVEBE/AKZO, qui a mis en rapports le premier et le second auteur. Sans la connexion AVEBE/AKZO, la réalisation de la version MS-DOS Turbo-Pascal de notre programme ne serait demeurée qu'un rêve.

Toutefois, les auteurs reconnaissent leur entière responsabilité pour les programmes et la documentation les accompagnant. Si nous avons fait de cet algorithme un cauchemar pour l'analyste, nous comptons et sommes toujours prêts à procurer conseils, consolations, compassion ou convivialité.

Bibliographie

- ALBRETHSEN, S.E. et BRINCH PETERSEN, E., 1975. Gravene på Bøgebakken, Vedbaek. *Søllerødbo-gen* : 9-60.
- ALBRETHSEN, S.E. et BRINCH PETERSEN, E., 1976. Excavation of a Mesolithic cemetery at Vedbaek, Denmark. Copenhagen, *Acta Archaeologica*, 47 : 1-28.
- ALBRETHSEN, S.E., BRINCH PETERSEN, E., BALSLEV JØRGENSEN, J. et ALEXANDERSEN, V., 1976. De levende og døde... for 7 000 år siden. *Fra Nationalsmuseets Arbejdsmark* : 5-23.
- CONSTANDSE-WESTERMANN, T.S. et NEWELL, R.R., 1989. Limb lateralization and social stratification in Western European Mesolithic societies. In : I. Hershkowitz (éd.) : *People and Culture in Change. Proceedings of the second Symposium on Upper Palaeolithic, Mesolithic and Neolithic Populations of Europe and the Mediterranean Basin*. Oxford, B.A.R. International Series, 508 : 405-433.
- COWGILL, G.L., 1972. Models, methods and techniques for seriation. In : D. Clarke (éd.) : *Models in Archaeology*. London, Methuen : 381-424.
- DORAN, J.E., 1971. Computer analysis of data from the La Tène cemetery at Münsingen-Rain. In : F.R. Hodson, D.G. Kendall et P. Tautu (éd.) : *Mathematics in the Archaeological and Historical Sciences*. Edinburgh, Edinburgh University Press : 422-431.
- DUNNEL, R.C., 1970. Seriation method and its evaluation. *American Antiquity*, 35 : 496-512.
- ESTER, M., 1981. A column-wise approach to seriation. *American Antiquity*, 46 : 496-512.
- GOLDMANN, K., 1974. Die zeitliche Ordnung prähistorischer Funde durch Seriation. *Archäologisches Korrespondenzblatt*, 4 : 898-894.
- KENDALL, D.G., 1963. A statistical approach to Flinders Petrie's sequence dating. La Haye, *Bulletin de l'Institut international de Statistique*, 40 : 657-680.

- KENDALL, D.G., 1971a. Seriation from abundance matrices. In : F.R. Hodson, D.G. Kendall et P. Tautu (éd.) : *Mathematics in the Archaeological and Historical Sciences*. Edinburgh, Edinburgh University Press : 233-252.
- KENDALL, D.G., 1971b. Abundance matrices and seriation in archaeology. *Zeitschrift für Wahrscheinlichkeitstheorie*, **17** : 103-122.
- KENDALL, D.G., 1971c. A mathematical approach to seriation. *Philosophical Transactions of the Royal Society of London, A*, **269** : 125-134.
- KUZARA, R.G., MEAD, G.R. et DIXON, K.A., 1966. Seriation of anthropological data : A computer program for matrix ordering. *American Anthropologist*, **68** : 1442-1445.
- LEGOUX, R., 1980. Le recours à l'informatique. La chronologie relative par permutation matricielle. In : P. Périn (éd.) : *La datation des tombes mérovingiennes : Historique, méthodes, application*. Genève, Droz : 138-155.
- LEREDDE, H. et DJINDJIAN, F., 1980. Traitement automatique des données en archéologie. *Dossiers de l'Archéologie*, **42** : 52-69.
- MENESSIER, C., 1973. Tombes gallo-romaines du Limousin. Traitement graphique de l'information. *Recherches d'Archéologie celtique et gallo-romaine*, **83**.
- NEWELL, R.R., CONSTANDSE-WESTERMANN, T.S. et DE LANGEN, G.J. (en prép.). The Mortuary Practices of the Mesolithic Societies in Western Europe.
- PETRIE, W.M.F., 1899. Sequences in Prehistoric Remains. *Journal of anthropological Institute*, **29** : 295-301.
- ROZOY, J.G., 1987. *Les Celtes en Champagne. Les Ardennes au second Age du fer. Le Mont Troté, les Rouliers*. Mémoires de la Société Archéologique champenoise, **4**. Charleville, Rozoy éd., 2 vol., 750 p.
- ROZOY, J.G., BENNINK, J., CONSTANDSE-WESTERMANN, T.S. and NEWELL, R., Ordination and seriation of nominal data matrices on Apple and MF-DOS Personal Computers : the Rozoy numerical ordination and seriation program. *Archaeometry* (soumis).
- SIEGEL, S., 1956. *Nonparametric Statistics for Behavioral Sciences*. London, Mc Graw-Hill, 312 p.
- WILKINSON, E.M., 1974. Techniques of data analysis seriation theory. *Technische und naturwissenschaftliche Beiträge zur Feldarchäologie*. Köln, Rheinland Verlag : 1-142.
- WORSAAE, J.J.A., 1844. *Dänemarks Vorzeit durch Alterthümer und Grabhügel beleuchtet*. Kobenhaven.

Adresse des Auteurs : R.R. NEWELL
Biologisch Archaeologisch Instituut
Poststraat, 6
9712 ER Groningen (Pays-Bas)
et Department of Anthropology
S.U.N.Y. Binghampton,
Binghampton, New-York, (U.S.A)

J. BENNINK
Brahmslaan, 6
9603 CB Hoogezand (Pays-Bas)

Tr. CONSTANDSE-WESTERMANN
Instituut voor Prae- en Protohistorie
Singel, 453
1012 WP Amsterdam (Pays-Bas)

Dr J.G. ROZOY
26, rue du Petit Bois
F-08 000 Charleville-Mézières (France)