

SERiations ET TRAITEMENTS DE DONNEES ARCHEOLOGIQUES AVEC LE LOGICIEL MICROSOFT EXCEL SUR MACINTOSH (OU COMPATIBLE IBM)

INTRODUCTION

Il y a près d'un siècle que les archéologues sont confrontés aux problèmes de sériation de leurs données. Depuis lors, différentes techniques ont été mises au point, à un rythme de progression encore plus soutenu depuis les années soixante (DJINDJIAN 1991, 183). Il semble toutefois que depuis que les mérites de l'analyse factorielle des correspondances ont été à juste titre soulignés, les progrès dans ce domaine marquent sensiblement le pas: comme si, au delà de ces analyses, plus aucun progrès ne pouvait être accompli. Cette impression risque d'être trompeuse: bien des efforts restent en effet à accomplir dans cette perspective. Il ne s'agit pas tant de mettre au point d'autres techniques, puisque celles dont on dispose sont satisfaisantes, que de rendre ces techniques toujours plus transparentes et accessibles aux utilisateurs, c'est-à-dire aux archéologues eux-mêmes.

Combien, en effet, peuvent se targuer de maîtriser parfaitement les techniques d'analyse factorielle au point d'être capable d'en réaliser soi-même une? La formation de l'archéologue est ainsi faite — même si d'un pays à l'autre ou à l'intérieur d'un même pays les situations peuvent varier grandement — que tout ce qui relève des traitements de données est laissé de côté comme s'il était possible d'en faire l'économie dans la pratique quotidienne de la discipline. Le risque, à terme, est évident: la création d'une caste d'initiés aux techniques d'analyses des données, caste qui s'imposerait comme *l'interface* obligée entre l'archéologue et les traitements statistiques, mathématiques — donc le plus souvent informatiques — des données archéologiques.

En corollaire, je soulignerai deux autres risques exactement antithétiques mais reposant sur les mêmes fondations (l'ignorance): le premier, c'est qu'il ne soit plus possible de critiquer un résultat issu d'une machine, ou de falsifier — au sens logiciste de *rendre faux* — un raisonnement s'appuyant sur ces techniques; je pense à tel exemple récent où une analyse factorielle suivie d'une classification ascendante hiérarchique suffit à expliquer les plus primaires des corrélations entre style céramique et organisation socio-politique; le second risque, c'est qu'on ne puisse profiter de la puissance de ces techniques pour avancer dans les analyses de données archéologiques au prétexte que certains les utili-

sent pour démontrer tout et n'importe quoi et se faire une renommée rapide et à bon compte.

En attendant que tous les archéologues soient aussi à l'aise dans la manipulation d'un chi-deux que dans celle d'une moyenne — et il faudra passer par là si l'on souhaite démythifier ces techniques et les faire progresser dans le champ de l'archéologie, si l'on souhaite aussi démystifier une certaine élite en cours de formation autour d'un savoir relatif mais très spécialisé — il est nécessaire de faire l'effort de développer les outils que tout archéologue pourra, dans l'environnement informatique qui lui est à présent le plus familier — celui du micro-ordinateur — utiliser avec le minimum de savoir technique préalable: déjà les bibliothèques de programmes d'analyses de données — statistiques multidimensionnelles, analyses factorielles, classifications ascendantes... — sont disponibles sur petites machines mais elles réclament un apprentissage qui est trop rarement dispensé dans les universités; une autre possibilité est de développer certains de ces outils à partir des applications moins puissantes mais plus largement répandues.

C'est cet aspect que je souhaiterais illustrer ici en présentant des procédures de traitement de tableaux de données — en particulier des sériations — développées sur un logiciel connu de la majorité des utilisateurs de micro-informatique, le tableur EXCEL de chez Microsoft. Les « macros » — sortes de petits programmes écrits dans un langage proche de la langue naturelle — détaillées dans les pages suivantes se regroupent en trois catégories: des macros de sériation, des modules de construction de tableau d'effectifs ou d'incidence et de matrices de co-occurrence (tableaux de contingence et certains tableaux de similarité) et un module permettant différents types de remplacements dans les matrices réalisées.

1. SÉRIATIONS AVEC EXCEL

1.1 *Module de sériation automatique par la méthode des barycentres*

Le premier module proposé permet de réaliser des sériations de matrices en utilisant la méthode des barycentres — autrement appelée méthode des moyennes réciproques ou en anglais « reciprocal averaging method » (DJINDJIAN 1985, 121).

Je commencerai par en rappeler le principe: le barycentre est l'autre appellation du centre de gravité. Une ligne (ou une colonne) comportant des 1 et des 0 peut être comparée au fléau d'une balance sur lequel seraient disposées des masses (les 1): le point d'équilibre est fonction de la distance entre les masses, du nombre des masses ainsi que de la valeur des masses. La balance « romaine » est la meilleure illustration concrète du principe. Dans le cas d'un tableau de

données, la masse est la valeur du contenu de chaque case et la distance de chaque masse est déterminée par la position de la case. Le barycentre d'une ligne ou d'une colonne est ainsi la somme des valeurs de chaque case (de la ligne ou de la colonne) multipliées chacune par le numéro de la case (dans la ligne ou la colonne), le tout divisé par la somme des valeurs de chaque case (de la ligne ou de la colonne).

La sériation du tableau consiste alors en l'ordonnancement itératif des lignes et des colonnes selon la valeur — croissante ou décroissante — de leur centre de gravité. Le traitement s'opère en quatre étapes (GRAHAM *et al.* 1976, 11: « four step method »): on commence par ordonner les lignes (ou les colonnes), on recalcule les barycentres, on ordonne les colonnes (ou les lignes), on recalcule... (LAGRANGE 1973, 64; FERNANDEZ DE LA VEGA 1971, 2-4; 1977, 158-160; IHM *et al.* 1978, 487 ss.; DJINDJIAN 1985a, 17 et 27-28; 1991, 175-176 et 186; RUBY 1990b, 3-4). Ces opérations sont répétées autant de fois que nécessaire, c'est-à-dire jusqu'au moment où les barycentres des lignes — si on a commencé par elles — ne sont plus déclassés par le tri des colonnes précédent. L'ordonnancement du tableau aboutit à un regroupement des valeurs non-nulles des lignes et des colonnes autour d'une diagonale théorique occupée par les centres de gravité, autrement appelée « diagonale chargée ». Le tableau est alors sérié: on parle d'un *scalogramme*.

On peut formuler deux remarques sur cette méthode. D'une part le document initial peut être un tableau d'effectifs, de fréquences ou encore un tableau d'incidence en présence/absence (0/1) (DJINDJIAN 1985a, 28; 1991, 186). D'autre part, le choix pour les barycentres entre l'ordre croissant et l'ordre décroissant selon lequel on va trier les lignes et les colonnes n'est pas important: il importe par contre que l'ordre soit le même pour les lignes et les colonnes.

Cette méthode, qui fait partie des techniques de sériations par rangement de matrice, a été mise au point en France à la fin des années soixante par F. Tonnellier (LAGRANGE 1973, 31; FERNANDEZ DE LA VEGA 1971, 2; 1977, 158). Mais il semble qu'elle ait été plusieurs fois — et indépendamment — redécouverte, en particulier par K. Goldmann (1972, 22 ss.; DJINDJIAN 1985a, 17 et 27-28; 1991, 175-176), Wilkinson (1974) et Legoux (1980). Le caractère itératif des tris et des calculs des barycentres a depuis le début incité les chercheurs à implémenter l'algorithme sur des calculateurs capables de prendre en charge ces actes répétitifs. C'est le même constat qui m'a poussé à adapter cette technique aux micro-ordinateurs en utilisant un simple langage de macro-programmation. La simplicité combinée à la puissance du logiciel Microsoft Excel a permis ce développement. Une première version de la macro a été proposée alors que la version 2.2 du logiciel sortait à peine. A présent que des versions ultérieures d'Excel sont proposées, il est possible d'actualiser cette macro de sériation.

Texte et commentaire de la macro

	B
1	<i>Macro Sériation Automatique, barycentres.</i>
2	<i>Version 3 US</i>
3	<i>© Pascal Ruby. Mai 1992.</i>
4	<i>Pour exécuter la macro-commande de sériation, taper Commande-Option-S</i>
5	
6	=OPEN(INPUT(" Nom du document à sérier ? ";2;" Document à traiter "))
7	=ERROR(1:B76)
8	
9	<i>Ecriture formules</i>
10	=SELECT("R1")
11	=INSERT()
12	=SELECT("C1")
13	=INSERT()
14	=SELECT.LAST.CELL()
15	=SET.NAME("Ligtab";GET.CELL(2))
16	=SET.NAME("Coltab";GET.CELL(3))
17	=SELECT(OFFSET(SELECTION();0;1-Coltab))
18	=FORMULA.ARRAY("=sum(column(RC[2]:RC"&Coltab&"")*(RC[2]:RC"&Coltab&"") /sum(RC[2]:RC"&Coltab&""))
19	=COPY()
20	=SELECT(!\$A\$3:ACTIVE.CELL())
21	=PASTE()
22	=SELECT.LAST.CELL()
23	=SELECT(OFFSET(SELECTION();1-Ligtab;0))
24	=FORMULA.ARRAY("=sum(row(R[2]C:R"&Ligtab&"C")*(R[2]C:R"&Ligtab&"C"))/su m(R[2]C:R"&Ligtab&"C"))
25	=COPY()
26	=SELECT(!\$C\$1:ACTIVE.CELL())
27	=PASTE()
28	=CALCULATE.NOW()
29	
30	<i>Tri sur barycentres</i>
31	=CALCULATION(3;;;)
32	=ECHO(0)
33	=SET.NAME("j";1)
34	=SET.NAME("b";0)
35	=COLUMN.WIDTH(5:1\$A\$1)
36	=SELECT(!\$A\$2)

37	=FORMULA("Tri n°")
38	=ALIGNMENT(4)
39	=SELECT(!\$B\$2)
40	=ALIGNMENT(2)
41	=FORMULA(J+B)
42	=ECHO(1)
43	=ECHO(0)
44	=SELECT.LAST.CELL()
45	=SELECT(!\$C\$1:ACTIVE.CELL())
46	=SORT(2;!\$C\$1:1)
47	=CALCULATE.NOW()
48	=SELECT.LAST.CELL()
49	=SELECT(!\$A\$3:ACTIVE.CELL())
50	=SORT(1;!\$A\$3:1)
51	=CALCULATE.NOW()
52	=BEEP()
53	=SET.NAME("J";J+1)
54	=IF(J=11;GOTO(B55);GOTO(B58))
55	=IF(ALERT(" Dix tris ont déjà été effectués. Continuer ? ";1);GOTO(B56);GOTO(B65))
56	=SET.NAME("j";1)
57	=SET.NAME("b";b+10)
58	=GOTO(B59)
59	=SELECT(!\$C\$1)
60	=SET.NAME("a";DEREF(SELECTION()))
61	=SET.NAME("z";DEREF((OFFSET(SELECTION();0;1))))
62	=IF(z=0;GOTO(B65);IF(z=a;GOTO(B63);GOTO(B39)))
63	=SELECT(OFFSET(ACTIVE.CELL();0;1))
64	=GOTO(B60)
65	=SELECT("R2C1:R2C2")
66	=CLEAR(1)
67	=COLUMN.WIDTH(3;!\$A\$1)
68	=SELECT(!\$A\$1)
69	=BEEP()
70	=BEEP()
71	=SELECT("R1")
72	=EDIT.DELETE()
73	=SELECT("C1")
74	=EDIT.DELETE()
75	=ECHO(1)
76	=IF(ALERT(" Imprimer le résultat ou avoir un aperçu avant impression ? ";1);PRINT.PREVIEW();GOTO(B77))

77	=IF(ALERT(" Enregistrer le résultat? ";1);SAVE();GOTO(B78))
78	=RETURN()

La macro précédente comprend deux parties: la première (lignes 9 à 28) permet l'entrée automatique des formules des barycentres sur la matrice ouverte; la seconde (lignes 30 à 64) opère la sériation proprement dite. Bien que les deux étapes ne soient pas distinctes pour l'utilisateur — elles s'enchaînent sans intervention extérieure — il peut être utile de donner quelques explications supplémentaires sur le fonctionnement.

D'abord le document de départ: qu'il soit issu du traitement présenté plus loin ou qu'il ait une autre origine, il doit s'agir d'une matrice d'effectifs ou d'incidence (en 0/1); seules les cases non nulles doivent avoir un contenu puisqu'une absence de réponse (une case vide) est considérée par le logiciel comme contenant "0". La première ligne et la première colonne sont réservées à l'entrée des identifiants des objets et des propriétés (ou des individus et des caractères). Ce document aura préalablement été sauvegardé et devra — pour en faciliter le lancement — se trouver dans le même dossier (*directory*) que la feuille macro.

Au lancement de la macro par la frappe du raccourci-clavier défini lors de la « Définition du nom » (voir *infra*), le nom du document à ouvrir pour traitement est réclamé. Une fois le document ouvert, dans un premier temps une ligne et une colonne vierges sont insérées à l'emplacement de la première ligne et de la première colonne du tableau: c'est là que vont être inscrites les formules de barycentre¹. La macro "rédige" ensuite la formule des barycentres-lignes en actualisant les références relatives données dans le texte de la macro en B18 au moyen des références absolues de la dernière cellule. C'est là la principale force d'Excel: parce qu'il permet de sélectionner cette dernière cellule, les tableaux pourront être traités de la même façon quelle que soit leur taille. Dans le cas présent les formules des barycentres vont être entrées dans le tableau à partir de la dernière cellule de la première colonne et de la dernière cellule de la première ligne; ensuite ces formules sont copiées et collées dans les autres cellules de la première ligne et de la première colonne; pour chaque cellule, ces formules sont automatiquement actualisées en fonction des références de la cellule.

Cette entrée automatique des barycentres constitue la modification majeure de cette version de la macro ZMSA par rapport à la version précédente. Pour le reste, en effet, la macro opère comme auparavant. Les lignes 44 à 51 de la macro permettent un double tri (sur les colonnes puis, après un recalcul des barycentres, sur les lignes) qui boucle tant que deux tests n'obligent pas à en sortir:

— le premier — lignes 52 à 58 — n'a pour but que de sortir d'une sériation qui boucle, c'est-à-dire lorsque les modifications dues au tri des barycentres-lignes ne font qu'annuler les modifications du tri précédent sur les barycentres-colonnes et provoqueraient ainsi des tris sans fin; j'ai fixé ici le nombre à dix tris doubles au terme desquels on peut soit relancer dix nouveaux tris (dans le cas de grandes matrices, dix tris sont rarement suffisants) soit interrompre la procédure en considérant que le tableau ne peut pas être davantage sérié (ce qui ne veut pas dire que la sériation obtenue n'est pas satisfaisante pour autant);

— le second — lignes 59 à 64 — vérifie que le tri précédent sur les barycentres-lignes n'a pas modifié l'ordre croissant des barycentres-colonnes et dans le cas contraire relance la procédure pour un nouveau tri double.

¹ Je n'entre pas ici dans le détail du calcul du barycentre avec Excel puisque dans la version proposée ici, l'utilisateur n'a plus à se préoccuper de sa formulation. On peut toutefois se reporter aux explications que je donne dans RUBY 1990b, 6.

En m'appuyant sur les résultats de M.S. Lagrange, je n'ai pas envisagé de traitement particuliers pour les lignes ou les colonnes qui se trouvaient affectées d'une valeur identique de barycentre: les tests effectués par l'auteur ont montré en effet que les coefficients de dispersion autour de la diagonale variaient peu selon les différentes relations d'ordre choisies (LAGRANGE 1973, 65).

Par ailleurs, l'utilisateur est tenu informé de la progression des tris doubles par un compteur qui s'inscrit dans l'angle supérieur gauche de la matrice en cours de traitement: les lignes 30 à 41 programment ce compteur.

Enfin, les dernières lignes du module sont destinées à « nettoyer » la matrice une fois la sériation achevée: la ligne et la colonne réservées aux barycentres — à présent inutiles — sont supprimées; on peut ensuite visualiser le résultat — surtout s'il s'agit d'une matrice de grandes dimensions — en recourant à l'aperçu avant impression et en demandant une réduction; enfin, on a le choix entre sauvegarder ou pas le résultat (Fig. 1).

Sériation									
	A	B	C	D	E	F	G	H	
1		1a	7i	2g	1b	1n	3f	2s	
2	Tombe 1	1	1						
3	Tombe 3	1	1						
4	Tombe 8	1	1	1					
5	Tombe 2			1	1	1	1		
6	Tombe 6			1	1	1	1		
7	Tombe 10			1	1	1	1		
8	Tombe 4						1	1	
9	Tombe 7						1	1	
10	Tombe 9						1	1	
11	Tombe 5							1	
12									

Fig. 1

1.2 Sériation par l'algorithme de P. Ihm

La technique de sériation par les barycentres a rarement laissé indifférents les chercheurs travaillant sur ces problèmes d'ordination. On verra plus loin certaines des critiques formulées alors à l'encontre de cet algorithme "capricieux"². C'est à P. Ihm que l'on doit un puissant — et voisin — algorithme de substitution: il remarqua en effet qu'au prix de mineurs aménagements, la méthode développée par K. Goldmann (c'est-à-dire les moyennes réciproques) relevait des techniques d'analyses des correspondances (IHM 1983, 8, 13 et 17

²Je traduis ainsi le qualificatif « erratic » utilisé par HODSON, TYERS 1988, 33.

et bibliographie précédente). La procédure adoptée par P. Ihm consiste à envisager les coefficients ligne et colonne comme les coordonnées vectorielles d'individus dont on connaît les masses (0 et 1 dans un tableau en présence/absence). Il est donc possible de calculer un coefficient de corrélation entre les deux séries de valeurs que représentent les coordonnées — abscisses et ordonnées. L'algorithme retenu vise à maximiser ce coefficient de corrélation. La différence entre les coordonnées et les coefficients ligne et colonne bruts utilisés dans l'algorithme précédent est introduite par une pondération de ces coefficients. Cette pondération peut être calculée de deux façons:

— soit en soustrayant au coefficient de départ la valeur minimale de la série de coefficients et en divisant le tout par la différence entre la valeur maximale et la valeur minimale de la série; ainsi, pour une série de coefficients $x_1, x_2, \dots, x_n$, la pondération sera de la forme

$$\text{pour } i = 1 \text{ à } n, x_i = (x_i - x_{\min}) / (x_{\max} - x_{\min})$$

— soit en soustrayant au coefficient de départ la moyenne de la série et en divisant le tout par l'écart-type de la série de coefficients, soit la formule

$$\text{pour } k = 1 \text{ à } m, y_k = (y_k - \bar{y}) / \sigma_y$$

Par ailleurs, les coordonnées (ou les coefficients pondérés) sont calculées les unes en fonctions des autres: la coordonnée x_i d'une ligne sera égale à la somme des masses de chaque case (de la ligne) multipliées chacune par la coordonnée y_k de la colonne, le tout divisé par la somme des masses de chaque case de la ligne. Le calcul de la coordonnée y_k d'une colonne suit le même principe. On aura compris que ces références sont circulaires: les x_i sont calculés à partir des y_k et inversement les y_k sont calculés à partir des x_i . Cette circularité permet une itération du calcul des coordonnées jusqu'au moment où les valeurs de ces coordonnées se stabilisent: on peut alors procéder à l'ordonnancement des lignes et des colonnes en fonction des coordonnées finales pour obtenir un tableau sérié, un scalogramme. Le coefficient de corrélation entre les deux séries de coordonnées atteint un maximum³. Au moment de la publication, l'auteur de l'algorithme prit le soin de donner les lignes de programmation en BASIC de la procédure. Ces traitements peuvent être également effectués sur le logiciel Excel.

³ Pour tous ces aspects, il est recommandé de se reporter à l'exposé original de la méthode dans IHM 1983.

Texte et commentaire de la macro

	E
1	<i>Sérialisation en analyse des correspondances. Algorithme de Peter IHM</i>
2	<i>Version 3 US</i>
3	<i>© Pascal Ruby, Mai 1992.</i>
4	<i>Pour exécuter la macro-commande de sérialisation, taper Commande-Option-1</i>
5	
6	=OPEN(INPUT(" Nom du document ?";2;" Document à traiter "))
7	=SELECT("R1:R3")
8	=INSERT()
9	=SELECT("C1:C3")
10	=INSERT()
11	=SELECT.LAST.CELL()
12	=SET.NAME("Ligtab";GET.CELL(2))
13	=SET.NAME("Coltab";GET.CELL(3))
14	=SELECT(OFFSET(SELECTION();-(Ligtab-3);0))
15	=SELECT(!\$E\$3:ACTIVE.CELL())
16	=FORMULA("1")
17	=DATA.SERIES(1;1;1;1)
18	=SELECT.LAST.CELL()
19	<i>calcul de x</i>
20	=SELECT(OFFSET(SELECTION();0;2-Coltab))
21	=SET.NAME("lignes";Ligtab)
22	=SET.NAME("colonnes";Coltab)
23	=FOR("compteur";1;Ligtab-4)
24	=FORMULA.ARRAY("=sum((RC[3]:RC["&COLONNES-2&"])*R["&Lignes+3&"]C[3]:R["&Lignes+3&"]C["&COLONNES-2&"])/sum(RC[3]:RC["&COLONNES-2&"])"
25	=SELECT(OFFSET(SELECTION();-1;0))
26	=SET.NAME("lignes";Lignes-1)
27	=NEXT()
28	=SELECT.LAST.CELL()
29	=ECHO(0)
30	<i>pondération x</i>
31	=SELECT(OFFSET(SELECTION();0;2-Coltab))
32	=FOR("compteur";1;Ligtab-4)
33	=DEFINE.NAME("cellule"&compteur&"";ACTIVE.CELL())
34	=SELECT(OFFSET(SELECTION();0;-1))
35	=SET.NAME("lignes";Ligtab)
36	=FORMULA("=(cellule"&compteur&"-AVERAGE(C[1]))/STDEVP(C[1])")

37	=SELECT(OFFSET(SELECTION();-1;1))
38	=NEXT()
39	=SELECT("C1")
40	=COPY()
41	=SELECT("C3")
42	=PASTE.SPECIAL(3;;;)
43	
44	<i>calcul de y</i>
45	=SELECT.LAST.CELL()
46	=SET.NAME("Lignes";GET.CELL(2))
47	=SELECT(OFFSET(SELECTION();-(Ligtab-2);0))
48	=FOR("compteur";1;Coltab-4)
49	=FORMULA.ARRAY("=sum((R[3]C:R["&Lignes-2&"]C)*(R[3]C["&- COLONNES+3&"];R["&Lignes-2&"]C["&- COLONNES+3&"])/sum(R[3]C:R["&Lignes-2&"]C))")
50	=SELECT(OFFSET(SELECTION();0;-1))
51	=SET.NAME("COLONNES";COLONNES-1)
52	=NEXT()
53	
54	<i>pondération y</i>
55	=SELECT.LAST.CELL()
56	=SET.NAME("Lignes";GET.CELL(2))
57	=SELECT(OFFSET(SELECTION();-(Ligtab-2);0))
58	=FOR("compteur";1;Coltab-4)
59	=DEFINE.NAME("cellule"&compteur&"";ACTIVE.CELL())
60	=SELECT(OFFSET(SELECTION();-1;0))
61	=FORMULA("=(cellule"&compteur&"-AVERAGE(R[1]))/STDEVP(R[1])")
62	=SELECT(OFFSET(SELECTION();1;-1))
63	=NEXT()
64	=SELECT("C1")
65	=CLEAR(3)
66	=SELECT("R1")
67	=COPY()
68	=SELECT("R3")
69	=PASTE.SPECIAL(3;;;)
70	=SELECT.LAST.CELL()
71	=SELECT(OFFSET(SELECTION();-(Ligtab-1);0))
72	=FORMULA("=ROUND(ABS(R[2]C);0)")
73	=COPY()
74	=SELECT(!\$E\$1:ACTIVE.CELL())
75	=PASTE()

76	=CALCULATE.NOW()
77	=SELECT("R1")
78	=COPY()
79	=INSERT()
80	=PASTE.SPECIAL(3;;;)
81	=SELECT("C2")
82	=COPY()
83	=SELECT("C3")
84	=PASTE.SPECIAL(3;1;FALSE;FALSE)
85	=CALCULATE.NOW()
86	=SELECT("R3")
87	=COPY()
88	=SELECT("R4")
89	=PASTE.SPECIAL(3;1;FALSE;FALSE)
90	=CALCULATE.NOW()
91	=SELECT.LAST.CELL()
92	=SELECT(OFFSET(SELECTION();-(Ligtab);0))
93	=FOR("compteur";1;Coltab-4)
94	=SET.NAME("arrondi";DEREF(SELECTION()))
95	=SELECT(OFFSET(SELECTION();1;0))
96	=SET.NAME("cellule";DEREF(SELECTION()))
97	=IF(cellule=arrondi;GOTO(E98);GOTO(E113))
98	=SELECT(OFFSET(SELECTION();-1;-1))
99	=NEXT()
100	=SELECT("C2")
101	=CLEAR()
102	=SELECT.LAST.CELL()
103	=SELECT(!\$E\$1:ACTIVE.CELL())
104	=SORT(2; !\$E\$3;1)
105	=SELECT.LAST.CELL()
106	=SELECT(!\$A\$6:ACTIVE.CELL())
107	=SORT(1; !\$C\$6;1)
108	=SELECT("R1:R3")
109	=EDIT.DELETE()
110	=SELECT("C1:C2")
111	=EDIT.DELETE()
112	=GOTO(E118)
113	=SELECT("R2")
114	=COPY()
115	=SELECT("R1")
116	=PASTE.SPECIAL(3;;;)

117	=GOTO(E81)
118	=BEEP()
119	=ECHO(1)
120	=IF(ALERT(" Enregistrer le résultat? ";1);SAVE();GOTO(E121))
121	=RETURN()

Le traitement comprend deux étapes principales: la première consiste en l'écriture des formules de calcul des coefficients et de leur pondération; la seconde se résume à un calcul itératif des coefficients avec une procédure de vérification de stabilisation. Le tableau de départ est le même que pour la procédure précédente.

Les lignes 6 à 18 ouvrent le document à traiter, ajoutent trois lignes et trois colonnes dans l'angle supérieur gauche du tableau et calculent un certain nombre de paramètres.

Les lignes 19 à 42 puis 44 à 69 introduisent le calcul des x et des y et leur pondération respective. Parce que le calcul des x est effectué en utilisant les y et celui des y est effectué en utilisant les valeurs de x , on est en présence de références circulaires que le logiciel Excel ne sait pas traiter: il faut alors utiliser une ligne et une colonne supplémentaires où sont reportées les valeurs de x et de y (et non plus les formules), valeurs qui sont utilisées dans le calcul des autres coordonnées; cela revient à interrompre la circularité. Les lignes 81 à 90 effectuent cette itération.

Les lignes 70 à 80 et 91 à 99 puis 113 à 117 introduisent le test qui permet de savoir qu'on a atteint la stabilité dans le calcul des x et des y . Ce test repose sur la comparaison des valeurs absolues et arrondies des valeurs de y précédentes avec les nouvelles (respectivement sur les lignes 1 et 2 du tableau): si ces valeurs sont les mêmes, le traitement s'interrompt. On peut jouer sur la valeur de l'arrondi en choisissant un certain nombre de chiffres après la virgule: dans la programmation présentée ici, on ne prend en compte aucun chiffre après la virgule mais il suffit de modifier le 0 de la formule ROUND(argument();0) pour obtenir des traitements plus longs, en particulier dans le cas de matrices difficiles à sérier.

Les dernières lignes 100 à 112 opèrent l'ordonnancement des lignes et des colonnes en fonction des ultimes valeurs atteintes par les x et les y ; les lignes et colonnes de formules, à présent inutiles sont éliminées.

1.3 Les mérites comparés des deux algorithmes

Ces deux algorithmes ont des qualités et des défauts communs mais diffèrent cependant sur plusieurs points. Il est utile de préciser les uns et les autres.

Dans ses recherches doctorales, M.S. Lagrange a pu comparer l'efficacité de la technique des barycentres par rapport à une autre technique mise au point par W. Fernandez de La Vega qui s'appuie sur deux matrices de similarité (LAGRANGE 1973, 30 et 60): la sériation de caractères portés par des églises cisterciennes a été possible et satisfaisant avec le premier procédé et point avec le second. Néanmoins, l'expérimentation a permis de souligner les lacunes de la méthode des barycentres.

La méthode des barycentres est en particulier très sensible à l'ordre de départ. Ainsi, un même tableau différemment ordonné — au moyen de plusieurs tirages aléatoires par exemple — donnera vraisemblablement plusieurs résultats différents, parfois très différents. R. Legoux avait déjà analysé cet

aspect et remarqué que s'il y avait, dans tous les cas, convergence, le résultat et la stabilité dépendaient de la « largeur de bande », c'est-à-dire la dispersion des cellules pleines autour de la diagonale chargée (LEGOUX 1980, 145). Graham *et al.*, en comparant différentes méthodes de sériation sur la nécropole de Münsingen, soulignaient les précautions que l'emploi du programme Axis, développé par Wilkinson à partir des moyennes réciproques, réclamaient (GRAHAM *et al.* 1976, 15). Quant à Hodson et Tyers, ils regrettent « les résultats approximatifs » de cette technique « capricieuse » et cet aspect leur fait préférer l'algorithme de P. Ihm (HODSON, TYERS 1988, 33).

Le même problème existe aussi, mais de façon moindre, avec l'algorithme de P. Ihm: avec cette méthode, les résultats sont beaucoup plus stables et si l'on peut assister à certaines permutations de blocs entiers, ou seulement à l'intérieur de blocs, les grandes partitions restent sensiblement les mêmes⁴. Cette stabilité plus grande s'accompagne d'une très faible dispersion autour de la diagonale: c'est l'effet recherché par la maximisation du coefficient de corrélation. On pourra comparer les résultats bruts des sériations réalisées, avec les deux algorithmes précédents, de la nécropole de Münsingen telle qu'utilisée — également à des fins d'expérimentation — dans le travail de D.G. Kendall⁵ et mesurer les déplacements de la tombe 170 (Figg. 7, 8). Pour schématiser un peu, on peut dire que la méthode des barycentres est plus à l'aise dans la production de partitions que dans celle de sériations.

Pour la méthode des barycentres classique, la dispersion des points par rapport à la diagonale avait été appréhendée de façon assez proche par Tonnellier et par Goldmann: il peut s'agir de « la distance quadratique moyenne des croix à la diagonale principale » (LAGRANGE 1973, 65), ou du « Gütemaß » *i.e.* la somme des différences des points extrêmes sur les lignes et les colonnes (GOLDMANN 1972, 23); mais dans l'un et l'autre cas, il ne s'agit pas vraiment d'une mesure de la qualité de la sériation et encore moins d'un test permettant de décider s'il faut continuer le traitement.

Un dernier point mérite d'être mentionné: les deux procédures, comme l'analyse factorielle des correspondances, ne donne pas le sens de lecture de la sériation; les individus et caractères les plus récents peuvent être positionnés dans la partie supérieure gauche du tableau. C'est à l'archéologue de décider dans quel sens il doit lire le scalogramme.

Sans considérer les qualités et défauts propres aux deux méthodes que l'on

⁴ Parce qu'on ne peut utiliser des individus supplémentaires comme dans les analyses factorielles, la procédure est assez sensible aux individus et aux caractères anormaux: la tombe 170 du cimetière de Münsingen provoque de fortes modifications d'une sériation à une autre.

⁵ KENDALL 1971. Pour l'expérimentation présente, les lignes et les colonnes de la matrice de départ avaient été ordonnées selon un coefficient de rang aléatoire.

vient d'évoquer, les avantages respectifs des deux différentes procédures implémentées sur Excel se mesurent en terme de temps de traitement: pour un petit tableau, il semble que la méthode classique des barycentres peut donner un résultat plus rapide; pour des matrices de plus grandes dimensions, la méthode basée sur l'algorithme de Ihm montrera une nette supériorité. Car la donnée essentielle des traitements sur Excel est la lenteur d'exécution qui tient à un langage de macro-programmation qui ne peut être comparé à un vrai langage de programmation: il n'est qu'à comparer les quelques lignes de BASIC que peut se contenter d'indiquer P. Ihm et la centaine d'instructions nécessaires pour Excel! Je reste toutefois persuadé que le problème du temps de traitement est relatif dans ce cas et qu'il n'est pas nécessairement pénalisant. Il faut enfin souligner que les deux algorithmes peuvent traiter des matrices de toutes dimensions.

1.4 *Sériations automatisées: quelques remarques méthodologiques générales*

On met parfois en avant la capacité de traitement *objectif* de l'ordinateur en comparaison des capacités *subjectives* de l'archéologue. Cette course à l'objectivité est vaine: elle l'est également dans le cas des sériations. L'informatisation des procédures de sériations — même lorsqu'il ne s'agit comme ici que d'automatisation — ne permet pas d'approcher un résultat objectif. Quelle que soit la procédure adoptée, la sériation proprement dite n'est — comme l'a bien montré F. Djindjian — qu'une étape dans un long raisonnement, depuis la question initiale "Pourquoi sérier?" jusqu'à sa réalisation en passant par la définition du corpus et le choix des variables⁶. A chacune de ces étapes, la *subjectivité* de l'archéologue intervient et elle interviendra également ensuite, même après l'application d'une technique d'ordonnancement mathématique: il sera souvent nécessaire de procéder à un "lissage" de diagonale qui permettra de passer de l'ordonnancement logique à un ordonnancement archéologique, ou si l'on préfère, d'une ordination à une construction-explication; dans tous les cas également, l'archéologue devra décider lui-même du sens de lecture de la sériation puisque ni les algorithmes présentés ici ni l'analyse factorielle des correspondances ne le renseignera sur ce point.

Le problème de l'éventuelle objectivité des procédures informatisées est en fait le résultat d'une confusion: on se réfère à une opposition entre objectif et subjectif alors que plus pertinemment c'est celle entre *implicite* et *explicite* qui devrait être prise en compte. La formalisation du raisonnement en archéologie permet d'en rendre explicites les étapes et reproductibles les procédures,

⁶ DJINDJIAN 1991. De nombreuses tentatives de reprendre des sériations anciennes ont avorté devant l'impossibilité de redéfinir les types utilisés: dans ce cas, aucune technique de sériation ne sera plus efficace qu'une autre; elles seront, au sens propre du terme, sans objet.

d'apprécier le poids des prémisses dans les hypothèses conclusives. L'information, par la formalisation qu'elle requiert, implique cette exigence de "rendre explicite". Il faut une fois de plus revenir au travail de Gardin et à ses efforts de formalisation: comme toute construction-explication, une sériation — ou une partition — peut avoir un ou des sens selon différents plans: le lieu L, le temps T et le groupe humain H. La seule façon de maîtriser des variations qui interviennent dans trois plans différents est de fixer, à la manière des résolutions d'équations à plusieurs inconnues, les plans de variations les uns après les autres: pour produire du sens chronologique et transformer ce qui n'est qu'une ordination en une *construction-explication*, il faudra s'assurer que la variabilité dans les deux autres plans est contrôlée, ce qui se fera en travaillant sur une combinaison LH (lieu x groupe humain) homogène (sur ces questions, GARDIN 1979, 158 ss.).

Pour illustrer ce propos, je prendrais deux exemples de problèmes de sériation rencontrés dans la protohistoire italienne. Au risque de créer des partitions sexuelles interprétées comme chronologiques⁷, on a parfois préféré opérer des sériations distinctes, pour les tombes masculines d'un côté, les tombes féminines de l'autre. Au risque initial, on en a substitué un autre bien plus grand: celui de ne plus maîtriser ni les partitions sexuelles ni les partitions chronologiques.

Dans un article publié en 1962 portant sur la nécropole du Premier Âge du Fer de Sala Consilina, K. Kilian publie deux matrices dans lesquelles sont regroupées, en l'absence de détermination anthropologique, les tombes à arme — les tombes d'hommes — et les tombes sans armes — les tombes de femmes. Le nombre total de mobiliers pris en considération dans ces matrices est de 206 tombes réparties comme suit: 159 tombes sans armes et 47 tombes à armes. Apparemment, K. Kilian a débuté sa partition chronologique sur les tombes sans armes et mis ainsi en évidence deux grandes phases (I et II) elles-mêmes découpées en trois sous-phases (IA, IB, IC et IIA, IIB, IIC). L'auteur a ensuite cherché à mettre en évidence ce découpage sur les tombes à armes et y est parvenu. Malheureusement, ce procédé occultait un phénomène particulièrement déterminant dans ce cas: les tombes ne contiennent plus d'armes après la sous-phase IIA et, à la même époque, disparaissent de nombreux marqueurs sexuels si bien que, sur les critères habituels, la matrice des tombes sans armes regroupe, pour la sous-phase IIB, des sépultures qui peuvent être aussi bien masculines que féminines (RUBY 1990a). La partition « a priori » des mobiliers

⁷ Voir les remarques formulées par B. D'Agostino à propos de la sériation proposée sur des tombes de Daunie par TINÈ BERTOCCHI (1975, 360) et l'interprétation de la partition effectuée, qui est sexuelle plutôt que chronologique; de même, dans son compte-rendu de l'ouvrage de Hencken sur Tarquinia, J. Close-Brooks estime que la phase IC de l'auteur est un groupe de tombes féminines (CLOSE-BROOKS 1970, 238).

en tombes masculines et tombes féminines, lorsqu'elle est fondée sur la composition de ces mobiliers ou la présence d'une catégorie d'objets particuliers, expose l'archéologue à ne pas percevoir les changements intervenant dans la sphère de l'idéologie funéraire en fonction du temps.

F. Djindjian a proposé une solution pour éviter ces deux risques. Il faut effectuer une sériation unique où les variables sont séparées en trois groupes: un premier groupe de variables chronologiques puis deux autres qui rassemblent les variables sexuelles (DJINDJIAN 1991, 170).

Un autre biais peut-être introduit lorsque la sériation est opérée sur une certaine catégorie d'objets. Ainsi, pour l'Age du Fer italien, la chronologie des fibules est à présent claire et bien connue (voir par exemple PERONI 1979). Dans certaines nécropoles où les tombes sont suffisamment riches de mobiliers de bronze, les sériations ne sont parfois opérées que sur ceux-là ou presque⁸: le résultat ne correspond plus à la périodisation interne de la nécropole c'est-à-dire qu'on ne produit plus des *phases* (propres à une nécropole) mais directement des *étapes* (définies sur une région) ou des *horizons* (définis sur des ensembles de régions), ce qui n'est pas exactement la même chose.

De son côté, P. Ihm souligne les risques d'interpréter des différences régionales comme des partitions chronologiques (IHM 1983). Cela donne l'occasion de rappeler que les ensembles sériés doivent être géographiquement homogènes: il doit s'agir d'un même noyau de nécropole, ou à défaut d'une seule nécropole.

Ces remarques insistent sur la nécessité d'une réflexion épistémologique encore plus précise de la part des archéologues. Si l'informatisation des traitements des données en archéologie aboutit à une spécialisation des rôles encore plus importante, le risque est grand de n'avoir plus les moyens de mener à son terme ce type de réflexion. Le "Pourquoi" d'un traitement, la signification d'un résultat ne peuvent pas être mesurés de façon interne: la validation des procédures et des résultats ne peut être — J.C. Gardin nous l'a appris — qu'externe, ce qui signifie que l'archéologue doit garder une pleine maîtrise sur les traitements mis en œuvre.

2. CONSTRUCTION DE TABLEAUX DE DONNÉES: INCIDENCE OU EFFECTIFS ET CO-OCCURRENCE

2.1 *Les tableaux d'effectifs ou d'incidence*

Quelle que soit la procédure de sériation retenue, la première opération à effectuer en vue d'un traitement de ce type est la réalisation d'un tableau de

⁸ Par exemple GASTALDI 1979. Cette remarque ne remet pas en cause la cohérence interne ni la validité de la partition.

présence/absence ou un tableau d'effectifs: les lignes pourront, par exemple, correspondre à des tombes; les colonnes à des types mis en évidence précédemment. On peut, pour cela, utiliser le premier module de la macro. Il suffit de partir d'une simple liste des individus et des caractères réalisée sur une feuille de calcul d'Excel. Il est également possible d'importer une liste réalisée sur un document texte ou sur un autre fichier. Le document doit être organisé de la façon suivante:

- 1ère colonne: nom des individus (ou des objets), par exemple « tombe 1 », « tombe 2 », etc.
- 2e colonne et suivantes: nom des caractères (des propriétés), par exemple des types présents dans la tombe, à raison d'un caractère par colonne; les caractères n'ont pas à être classés ni ordonnés;
- les colonnes ne doivent pas être nommées; la première ligne correspond au premier individu;
- le document devra pour des raisons pratiques porter un titre commençant par la lettre A (il ne s'agit que d'un problème d'ordre des différents documents ouverts).

En résumé, le document de départ doit ressembler à la liste suivante (Fig. 2). Cette liste, traitée avec le recours du module suivant de la macro, donne la matrice de la Fig. 3.

Il n'existe aucune limite au nombre d'individus, ni au nombre total de caractères différents, ni au nombre de caractères différents par tombe. La même macro peut produire indifféremment un tableau d'effectifs (le même caractère peut être présent, et comptabilisé, plusieurs fois) ou un tableau d'incidence (les

ADIS					
	A	B	C	D	E
1	Tombe 1	1a	7i		
2	Tombe 2	3f	2g	1b	1n
3	Tombe 3	1a	7i		
4	Tombe 4	3f	2s		
5	Tombe 5	2s			
6	Tombe 6	3f	2g	1b	1n
7	Tombe 7	3f	2s		
8	Tombe 8	1a	2g	7i	
9	Tombe 9	3f	2s		
10	Tombe 10	3f	2g	1b	1n
11					

Fig. 2

	Worksheet6							
	A	B	C	D	E	F	G	H
1		1a	7i	3f	2g	1b	1n	2a
2	Tombe 1	1	1					
3	Tombe 2			1	1	1	1	
4	Tombe 3	1	1					
5	Tombe 4			1				1
6	Tombe 5							1
7	Tombe 6			1	1	1	1	
8	Tombe 7			1				1
9	Tombe 8	1	1		1			
10	Tombe 9			1				1
11	Tombe 10			1	1	1	1	

Fig. 3

caractères sont considérés en présence/absence, c'est-à-dire en 1/0). La différence pourra intervenir:

- soit en amont du traitement, au moment de la constitution de la liste de départ; on décidera alors de garder ou non les caractères multiples;
- soit en aval; on pourra utiliser le module utilitaire de remplacement (voir *infra*) pour uniformiser le contenu des cellules non vides en "1".

Un traitement ultérieur (module de remplacement) pourra remplacer le contenu des cellules par des fréquences. Quelle que soit la solution retenue, on se souviendra que, pour les sériations ultérieures, le choix d'un tableau d'effectifs, d'incidence ou de fréquence n'a pas d'importance.

Texte et commentaire de la macro

	A
1	Macro disjonctif complet
2	Version 3 US
3	© Pascal Ruby. Mai 1992.
4	Pour exécuter la macro-commande taper Commande-Option-D
5	
6	=ECHO(0)
7	=NEW(1)
8	=OPEN(INPUT(" Nom du document à traiter? ";2;" Document à traiter "))
9	=SELECT(!\$A\$1)
10	=SET.NAME("w";DEREF(SELECTION()))
11	=IF(w=0;GOTO(A47);GOTO(A12))

12	=SELECT("R1")
13	=COPY()
14	=ACTIVATE(INDEX(DOCUMENTS();2))
15	=SELECT("R1")
16	=PASTE()
17	=SELECT(!\$A\$1)
18	=ECHO(1)
19	=ECHO(0)
20	=CUT()
21	=SELECT(!\$A\$2)
22	=SET.NAME("n";2)
23	=SELECT(OFFSET(SELECTION();1;0))
24	=SET.NAME("t";DEREF(SELECTION()))
25	=SET.NAME("n";n+1)
26	=IF(t=0;PASTE();GOTO(A23))
27	=SET.NAME("h";1)
28	=SELECT(!\$A\$1)
29	=SELECT(OFFSET(SELECTION();0;h))
30	=COPY()
31	=SET.NAME("q";DEREF(SELECTION()))
32	=IF(q=0;GOTO(A33);GOTO(A37))
33	=ACTIVATE(INDEX(DOCUMENTS();1))
34	=SELECT("R1")
35	=EDIT.DELETE()
36	=GOTO(A9)
37	=SELECT(!\$A\$2)
38	=SET.NAME("f";2)
39	=SELECT(OFFSET(SELECTION();0;1))
40	=SET.NAME("k";DEREF(SELECTION()))
41	=IF(q=k;GOTO(A42);IF(k=0;PASTE();GOTO(A39)))
42	=SELECT(OFFSET(SELECTION();1;0))
43	=SET.NAME("f";f+1)
44	=IF(f=n;GOTO(A45);GOTO(A42))
45	=SET.NAME("cellule";DEREF(SELECTION()))
46	=FORMULA.FILL(cellule+1)
47	=SET.NAME("h";h+1)
48	=GOTO(A28)
49	=ECHO(0)
50	=ACTIVATE(INDEX(DOCUMENTS();2))
51	=SELECT("R1")
52	=EDIT.DELETE()

53	=SELECT.LAST.CELL()
54	=SELECT(!\$A\$1:ACTIVE.CELL())
55	=COLUMN.WIDTH(3)
56	=SELECT(!\$A\$1)
57	=ACTIVATE(INDEX(DOCUMENTS();1))
58	=CLOSE(FALSE)
59	=ECHO(1)
60	=BEEP()
61	=BEEP()
62	=IF(ALERT(" Imprimer le résultat ou en avoir un aperçu avant impression ? ";1);PRINT?();GOTO(A663))
63	=IF(ALERT("Enregistrer le résultat? ";1);GOTO(A64);GOTO(A65))
64	=SAVE.AS(INPUT(" Nom sous lequel le document va être enregistré ? ";2;" Enregistrer sous le nom ");1::)
65	=RETURN()

Le principe de la macro est de lire le document initial ligne après ligne pour constituer un nouveau document sur lequel va être organisé le tableau d'effectifs ou d'incidence. Les lignes 12 à 17 de la macro gèrent cette lecture ligne à ligne. Ensuite, l'identifiant de l'objet est inscrit dans la première colonne (macro — lignes 19 à 27) tandis que les désignations des caractères viennent remplir la première ligne (macro — lignes 28 à 32); les lignes suivantes (33 à 36) permettent de boucler ce traitement et de passer à une nouvelle ligne du document initial.

Le remplissage des cellules — dénombrement des effectifs ou présence/absence — est pris en charge par la seconde partie de la macro: il s'agit d'abord de bien positionner la cellule à l'intersection d'une ligne et d'une colonne précises (lignes 37 à 44); ensuite d'inscrire dans la cellule le contenu non-nul — soit un "1" pour un codage en présence/absence soit le nombre réel du même caractère porté par l'individu en question (lignes 45 à 47) car la non-réponse est interprétée comme une valeur nulle. La fin de la macro est destinée à nettoyer et mettre en forme le tableau final (lignes 50 à 56), à fermer le document de base dans sa forme initiale (ligne 57 à 59), à visualiser, imprimer et sauvegarder la matrice destinée à être sériée grâce au second module.

2.2 Module tableau de co-occurrence

Le quatrième module permet la construction d'une matrice de co-occurrence: à partir d'un tableau d'incidence en présence/absence où sont croisés les individus et les caractères qu'ils portent, on crée un tableau où sont croisés les individus ou leurs caractères entre eux; dans le premier cas, on indiquera dans chaque cellule le nombre de caractères communs à deux individus, dans le second le nombre d'individus possédant à la fois deux caractères particuliers (LAGRANGE 1973, 27-29). Cette matrice peut être considérée comme un tableau de similarité à occurrence simple.

Il n'y a pas grande utilité à traiter un tableau de co-occurrence qui n'est pas

sérié, à moins que cela ne constitue la première étape de traitements ultérieurs qui viseront à établir une sériation⁹. En effet, un des intérêts de la matrice de co-occurrence est qu'elle renforce les partitions apparues dans une sériation précédente et qu'elle permet de mieux en apprécier la cohérence interne. Ces partitions peuvent être chronologiques: les matrices de co-occurrence réalisées à partir de matrices en présence/absence sériées soulignent alors les associations récurrentes des types qui ont une signification temporelle¹⁰. Les matrices sont également utiles sur des partitions d'une autre nature: elles conviennent par exemple aux partitions sexuelles, comme le montrent les analyses de M. Gebühr qui propose une partition sexuelle à partir de la co-occurrence des classes fonctionnelles dans des mobiliers de tombes germaniques d'Allemagne du Nord et du Danemark des premiers siècles de notre ère (GEBÜHR, KUNOW 1976).

Texte et commentaire de la macro

	C
1	Réalisation d'une matrice de co-occurrence.
2	Version 3 US
3	© Pascal Ruby. Mai 1992.
4	Pour exécuter la macro-commande, taper Commande-Option-c
5	
6	=NEW(1)
7	=SAVE.AS("aaaaa";,;)
8	=OPEN(INPUT(" Nom du document ? ";2;" Document à traiter "))
9	=ECHO(0)
10	=SELECT.LAST.CELL()
11	=SET.NAME("Ligtab";GET.CELL(2)-1)
12	=SET.NAME("Coltab";GET.CELL(3)-1)
13	=SELECT("R1")
14	=COPY()
15	=ACTIVATE.NEXT()
16	=SELECT("R1")
17	=PASTE()
18	=ACTIVATE.PREV()

⁹ Par exemple si l'on souhaite utiliser la méthode du « Scalogramme privilégié de W. Eli-seef » ou celle de Brainerd et Robinson. Sur un panorama de ces techniques, voir DJINDJIAN 1991; également LAGRANGE 1973 et DJINDJIAN 1985a, 14 sq. Il peut s'agir d'un mode de sériation particulier: voir GEBÜHR, KUNOW 1976 et GUBÜHR, KAMPFFMEYER 1980.

¹⁰ Voir par exemple les matrices de co-occurrence réalisées par BRUN 1986 à partir de ses sériations de nécropoles de l'Âge du Bronze.

19	=SET.NAME("ligne";0)
20	=SET.NAME("col";0)
21	=SET.NAME("ligne";ligne+1)
22	=SET.NAME("col";col+1)
23	=SET.NAME("refcol";0)
24	=SET.NAME("compteur";0)
25	=SET.NAME("Tombes";0)
26	=IF(col=Coltab;GOTO(C51);GOTO(C27))
27	=SELECT(OFFSET(!\$A\$1;1;col))
28	=SET.NAME("refcol";refcol+1)
29	=IF(refcol>Coltab-col;GOTO(C21);GOTO(C30))
30	=SET.NAME("LigCel";0)
31	=IF(LigCel=Ligtab;GOTO(C41);GOTO(C32))
32	=DEFINE.NAME("cellule";SELECTION();)
33	=IF(DEREF(SELECTION())=1;GOTO(C34);GOTO(C37))
34	=SELECT(OFFSET(SELECTION();0;refcol))
35	=IF(DEREF(SELECTION())=1;SET.NAME("Tombes";Tombes+1);SET.NAME("Tombes";Tombes+0))
36	=FORMULA.GOTO("cellule")
37	=SELECT(OFFSET(SELECTION();1;0))
38	=SET.NAME("LigCel";LigCel+1)
39	=GOTO(C31)
40	
41	<i>Ecriture matrice</i>
42	=SET.NAME("compteur";compteur+1)
43	=IF(Tombes=0;GOTO(C27);)
44	=ACTIVATE.NEXT()
45	=SELECT(!\$A\$1)
46	=SELECT(OFFSET(SELECTION();ligne;compteur+col))
47	=FORMULA.FILL(Tombes)
48	=ACTIVATE.PREV()
49	=GOTO(C25)
50	
51	<i>Transposer par rapport à la diagonale</i>
52	=ACTIVATE.NEXT()
53	=SELECT.LAST.CELL()
54	=SET.NAME("Colonnes";GET.CELL(3))
55	=SET.NAME("ligne";0)
56	=SET.NAME("col";1)
57	=SELECT(OFFSET(!\$A\$1;ligne;col))
58	=SELECT("RC:RC["&Colonnes-2-ligne&"]")

59	=COPY()
60	=SELECT(OFFSET(ACTIVE.CELL();1;-1))
61	=PASTE.SPECIAL(2;1;0;1)
62	=SET.NAME("col";col+1)
63	=IF(col=Colonnes;GOTO(C66))
64	=SET.NAME("ligne";ligne+1)
65	=GOTO(C57)
66	=SELECT.LAST.CELL()
67	=SELECT(!\$B\$1:ACTIVE.CELL())
68	=COLUMN.WIDTH(3)
69	
70	<i>Remplissage de la diagonale</i>
71	=SET.NAME("ligtab";ligtab+2)
72	=IF(ALERT(" Remplir la diagonale ? ";1);GOTO(C73);GOTO(C101))
73	=ACTIVATE(INDEX(DOCUMENTS();2))
74	=SET.NAME("cptlig";0)
75	=SET.NAME("cptcol";0)
76	=SELECT("R1")
77	=INSERT()
78	=SELECT.LAST.CELL()
79	=SELECT(OFFSET(SELECTION();-(ligtab-1);0))
80	=FORMULA("=sum(R{2}C:R"&ligtab&"C)")
81	=COPY()
82	=SELECT(!\$B\$1:ACTIVE.CELL())
83	=PASTE()
84	=CALCULATE.NOW()
85	=SELECT.LAST.CELL()
86	=SELECT(OFFSET(SELECTION();-(ligtab-1);0))
87	=COPY()
88	=ACTIVATE(INDEX(DOCUMENTS();1))
89	=SELECT.LAST.CELL()
90	=SELECT(OFFSET(SELECTION();-cptlig;-cptcol))
91	=PASTE.SPECIAL(3;1;;)
92	=SET.NAME("cptlig";cptlig+1)
93	=SET.NAME("cptcol";cptcol+1)
94	=ACTIVATE(INDEX(DOCUMENTS();2))
95	=IF(cptcol=Coltab;GOTO(C98);GOTO(C96))
96	=SELECT(OFFSET(SELECTION();0;-1))
97	=GOTO(C87)
98	=SELECT("R1")
99	=EDIT.DELETE()

100	=ACTIVATE(INDEX(DOCUMENTS();1))
101	=ECHO(1)
102	
103	<i>Dessin graphique 3D</i>
104	=IF(ALERT(" Editer le graphique ? ";1);GOTO(C105);GOTO(C111))
105	=SELECT.LAST.CELL()
106	=SELECT(!\$A\$1:ACTIVE.CELL())
107	=ECHO(0)
108	=NEW(2;2)
109	=GALLERY.3D.COLUMN(6)
110	=FULL(TRUE)
111	=ECHO(1)
112	=BEEP()
113	=RETURN()

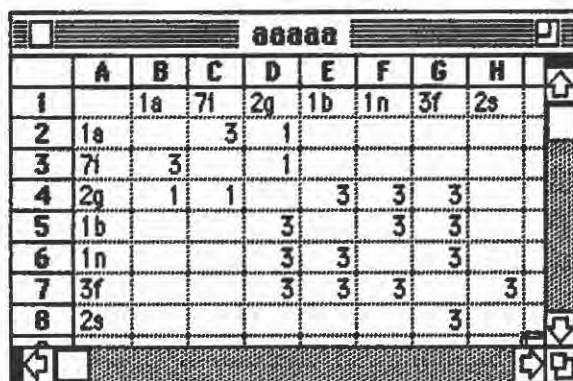
Cette macro est articulée en cinq étapes. Dans un premier temps (lignes 18 à 49), chaque colonne de la matrice de départ est comparée aux colonnes suivantes. Auparavant, un nouveau document aura été créé qui contient, dans la première ligne et dans la première colonne, les mêmes identifiants, c'est-à-dire la première ligne du tableau de départ. La comparaison est effectuée au moyen d'un certain nombre de boucles imbriquées (il n'est qu'à apprécier le nombre de compteurs nécessaires) et de lectures directes du contenu des cellules: la comparaison n'est cherchée avec l'autre colonne que lorsque la cellule de la colonne de référence est non-nulle.

Au terme de chaque comparaison de colonne à colonne, le total des présences communes est inscrit dans la matrice de co-occurrence. Cette écriture constitue la deuxième étape (lignes 41 à 49). On repart ensuite pour une autre comparaison, soit avec la même colonne de référence et une autre colonne de comparaison, soit avec une autre colonne de référence si la précédente n'a été comparée aux n-1 colonnes suivantes (on aura donc n! comparaisons). Il est à noter qu'à cette étape, l'écriture ne porte que sur une moitié de la matrice car elle est parfaitement symétrique par rapport à sa diagonale.

Ce n'est que dans un troisième temps, une fois que toutes les comparaisons ont été établies et que la moitié de la matrice a été écrite, que la seconde moitié est à son tour remplie par une simple transposition de la première de l'autre côté de la diagonale (lignes 51 à 69).

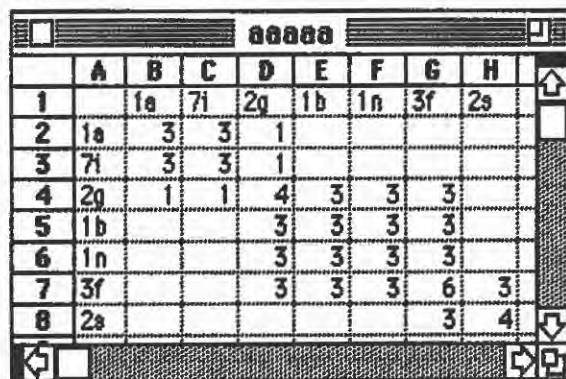
La quatrième étape peut être évitée: elle consiste en effet à remplir la diagonale, jusque là demeurée vide, par le nombre d'occurrences du caractère unique. Le choix du remplissage de cette diagonale dépend de la finalité de la matrice de co-occurrence qu'on réalise: s'il s'agit de souligner les partitions en évaluant les associations préférentielles de types ou de classes fonctionnelles, il peut être utile de remplir la diagonale qui permettra d'avoir une idée de la proportion des objets impliqués dans des associations par rapport au nombre d'objets de chaque type ou de chaque classe (Fig. 5). Par contre, si des traitements ultérieurs doivent être effectués, il convient de s'abstenir de remplir cette diagonale (Fig. 4).

L'ultime étape — qu'on peut, comme la précédente, éviter — consiste à éditer un graphique en trois dimensions de la matrice de co-occurrence, parfois plus lisible, pour les grandes lignes, qu'un tableau de chiffres.



	A	B	C	D	E	F	G	H
1		1a	7i	2g	1b	1n	3f	2s
2	1a		3	1				
3	7i	3		1				
4	2g	1	1		3	3	3	
5	1b			3		3	3	
6	1n			3	3		3	
7	3f			3	3	3		3
8	2s						3	

Fig. 4



	A	B	C	D	E	F	G	H
1		1a	7i	2g	1b	1n	3f	2s
2	1a	3	3	1				
3	7i	3	3	1				
4	2g	1	1	4	3	3	3	
5	1b			3	3	3	3	
6	1n			3	3	3	3	
7	3f			3	3	3	6	3
8	2s						3	4

Fig. 5

2.3 Module utilitaire de remplacement

Pour accroître la lisibilité d'une matrice sériée, en particulier dans le cas de tableaux de grandes dimensions qu'on veut visualiser en réduction au moyen de l' "Aperçu avant impression" mais aussi dans le cas d'une impression, on peut souhaiter remplacer les "1" du codage en présence/absence par des signes plus perceptibles: des points · obtenus par la frappe des touches option-shift-point (Fig. 6). La macro suivante est un utilitaire qui permet ce remplacement rapide et systématique.

Elle peut toutefois être utilisée pour d'autres types de remplacement: des cases vides par des "0", des valeurs différentes de "1" par des "1" pour passer d'un tableau d'effectifs à un tableau d'incidence. Il est en effet possible d'ajouter, après la ligne 9, des lignes rédigées de la même façon (= Formula.Replace) en changeant seulement les "1" et "." par les valeurs qu'on souhaite substituer.

Je conseillerais néanmoins de ne pas oublier la macro de remplacement pro-

	A	B	C	D	E	F	G	H
1		1a	7i	2q	1b	1n	3f	2a
2	Tombe 1	•	•					
3	Tombe 3	•	•					
4	Tombe 8	•	•	•				
5	Tombe 2			•	•	•	•	
6	Tombe 6			•	•	•	•	
7	Tombe 10			•	•	•	•	
8	Tombe 4						•	•
9	Tombe 7						•	•
10	Tombe 9						•	•
11	Tombe 5							•

Fig. 6

posée dans la version précédente (RUBY 1990b, 16): elle est certes beaucoup plus lente car elle opère cellule après cellule (la fonction "Formula.Replace", ou en français "Remplacer.Cellule" n'a été introduite que dans la version 2.2 d'Excel) mais pour cette raison elle permet l'introduction dans chaque cellule de formules de calcul, par exemple des pourcentages par ligne ou par colonne à la place du contenu si l'on veut transformer un tableau d'occurrence simple en tableau de contingence basé sur les pourcentages.

Texte de la macro

	D
1	Module Utilitaire de Remplacement. Remplacement des "1" par "•" dans matrice sériée
2	Version 3 US
3	© Pascal Ruby. Mai 1992.
4	Pour exécuter la macro-commande taper Commande-Option-P
5	
6	=OPEN(INPUT(" Nom du document sur lequel remplacer les 1 par • ? ";2;" Document à traiter "))
7	=SELECT.LAST.CELL()
8	=SELECT(!B\$2:ACTIVE.CELL())
9	=FORMULA.REPLACE("1";"1";1;2;0)
10	=FORMAT.FONT("zapf dingbats";12;1;0;0;0;0;0)
11	=ALIGNMENT(3)
12	=SELECT(!\$A\$1)
13	=BEEP()
14	=BEEP()
15	=BEEP()
16	=RETURN()

3. QUELQUES REMARQUES MÉTHODOLOGIQUES SUR LE LOGICIEL EXCEL ET LA RÉDACTION DE MACROS

Le texte des différentes macros indiquées précédemment doit être rédigé dans les cinq premières colonnes d'une feuille macro sauvegardée sous le nom de ZMSA. A part la production de graphique en 3 dimensions qui ne fonctionne qu'à partir de la version 3 du logiciel, l'ensemble des macros est compatible avec la version 2.2 du logiciel.

A la différence de certains autres logiciels-tableurs, Excel n'est doté que d'un langage de macro-programmation et point d'un langage de programmation véritable. Il y a des avantages et d'inévitables inconvénients à cette situation: les avantages tiennent surtout à la syntaxe proche de celle d'un langage naturel qui rendent son apprentissage facile; les inconvénients ont les mêmes causes, il s'agit d'un langage lent et chaque version (US, italienne, française, ...) a son propre langage. Cette dernière remarque doit être toutefois corrigée: à partir de la version 3 du logiciel, une macro écrite avec une version française peut être sauvegardée en version française et en version US; de même si on lance cette macro française avec une version US — ou inversement — on s'apercevra que tous les termes des fonctions sont automatiquement traduits dans la langue de la nouvelle version; "`= Remplacer.cellule()`" deviendra ainsi "`= Formula.replace()`".

Toutefois la traduction ne porte que sur les fonctions *actives* de la macro et point sur l'ensemble du texte de la programmation. Ainsi, une formule écrite entre guillemets, car destinée à être introduite dans des cellules du tableau, restera inchangée et la traduction devra être faite manuellement. De même, les références aux cellules devront être révisées car si, en français comme en italien, une cellule peut être définie par son numéro de ligne et son numéro de colonne — L3C9 par exemple — en version US cela deviendra R3C9 (ligne se dit « row »). Plus subtil encore, les références relatives données en nombre de lignes et nombre de colonnes, pour des déplacements à l'intérieur d'un tableau, devront eux-aussi être revus: au lieu du français L(3)C(5) qui signifie « la cellule placée 3 lignes au dessous et 5 colonnes sur la droite de la cellule de départ », la version US exigera une rédaction sous la forme R[3]C[5] avec des crochets au lieu des parenthèses classiques ().

Dans les macros présentées ici, j'ai adopté la version US du langage de macro-programmation — et donc du logiciel tout court. Ce n'est pas par souci de compliquer les choses, bien au contraire. L'expérience m'a montré qu'il était plus facile de traduire de l'américain vers une autre langue — français, italien, etc. — que par exemple du français vers l'italien: l'abondante bibliographie existant à propos d'Excel dans chaque langue permettra de trouver les équiva-

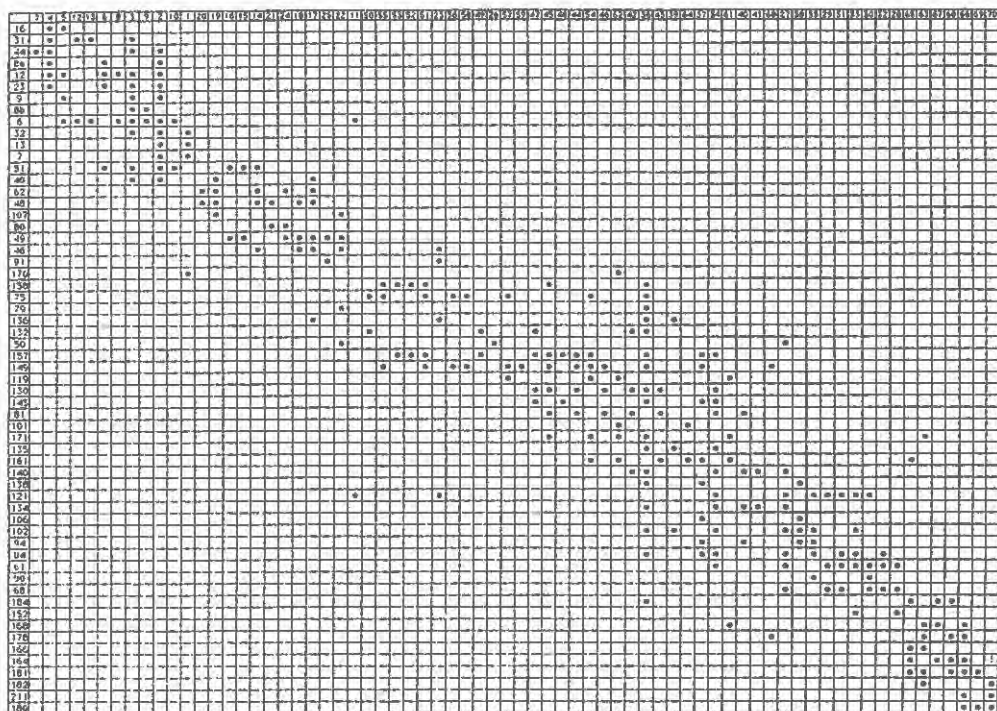


Fig. 7 — Münsingen sériée selon la méthode des barycentres.

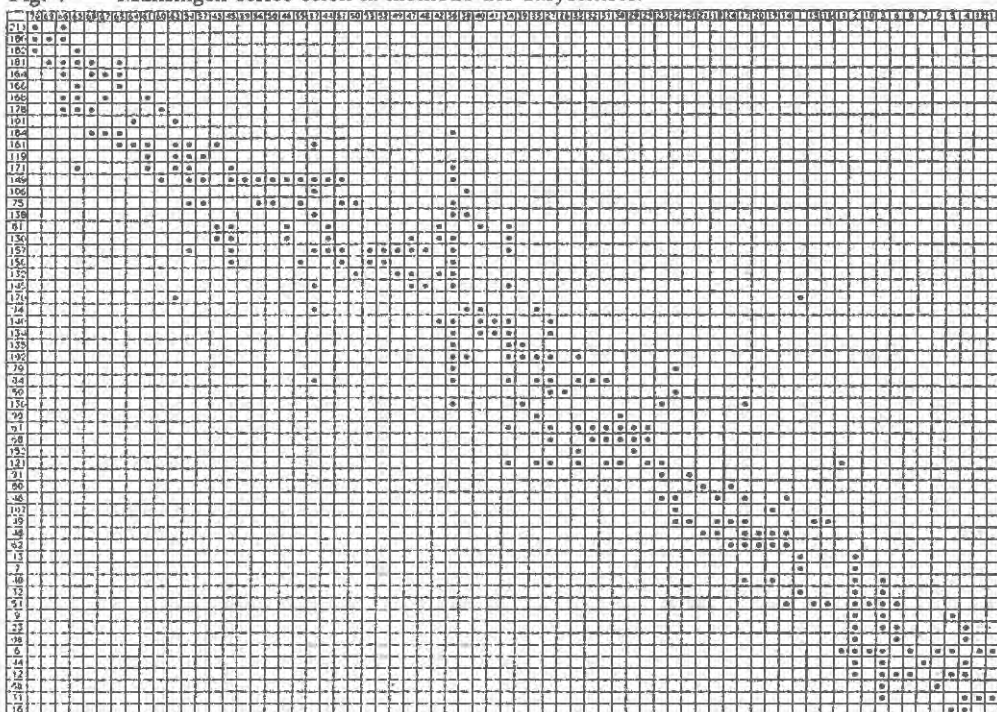


Fig. 8 — Münsingen sériée selon l'algorithme de P. Ihm.

lents à partir de la version US plutôt qu'entre deux versions non US. C'est d'autant plus vrai que la version 3 du logiciel permet ce passage immédiat entre la version US et les versions non US.

Les italiques sont utilisées dans le texte des macros pour indiquer ce qui pouvait être remplacé par n'importe quoi et en n'importe quelle langue. Il aurait pu en être ainsi des noms donnés aux « compteurs » — par exemple "ligne" dans une expression = Set.Name("ligne";0) —: ils n'apparaissent toutefois à aucun moment lors de l'utilisation et peuvent demeurer tels que je les ai arbitrairement fixés. Pour ce qui est du nom de la feuille macro (ZMSA) comme des autres noms des documents fixés arbitrairement, il faut veiller, si l'on veut les changer, à conserver la cohérence dans l'ordre alphabétique des différents documents: lorsqu'une macro manipule plusieurs documents, elle repère l'un par rapport à l'autre selon l'ordre alphabétique de leur nom.

Les lignes d'en-tête et, d'une manière générale, toutes celles qui ne commencent pas par le signe "=", peuvent être supprimées — il faut faire attention alors aux références internes à la macro — de même que les alertes sonores (les BEEP). On peut également supprimer les lignes ECHO(0) et ECHO(1) qui gèrent les rafraîchissements de l'écran: on gagne toutefois beaucoup de temps si on limite les affichages des déplacements dans de grandes matrices. Enfin, il est nécessaire d'attirer l'attention sur deux points:

— d'une part la syntaxe d'Excel est exigeante: on ne doit pas substituer, par exemple, un ";" à un ":"; le texte de la macro peut par contre être écrit indifféremment en majuscules ou en minuscules; il n'y a jamais d'espace entre les signes;

— d'autre part, avant de lancer les macros, il sera nécessaire de définir un nom pour chacun des cinq modules; c'est à cette occasion que seront définis les raccourcis-clavier permettant le lancement des macros; avec la version 3 du logiciel, on pourra aussi créer un "bouton" auquel sera affecté chaque module de macro et qui permettra son lancement.

Je ne reviendrai pas sur le problème de la lenteur des traitements opérés avec Excel. Cette lenteur est relative et point pénalisante: le traitement immobilise une machine mais pas son utilisateur; ces macros ont été conçues et testées sur un Macintosh Plus (machine construite autour du processeur 68000 cadencé à 8MHz) et tournent donc sur l'ensemble de la gamme Macintosh d'Apple; enfin, le résultat obtenu est un résultat sans erreur, ce qui n'est pas toujours le cas lors de manipulations manuelles de grandes matrices.

La diffusion des macros présentées plus haut — aussi modestes que soient les procédures qu'elles mettent en œuvre et les traitements qu'elles autorisent — doit être comprise comme une contribution à la maîtrise donnée aux archéologues sur les outils informatiques de traitements des données archéologiques.

Les procédures présentées ici ne sauraient se substituer à l'utilisation de gros systèmes ni aux techniques plus performantes: elles ne doivent pas pour autant être oubliées.

PASCAL R. RUBY
Ecole Française de Rome

BIBLIOGRAPHIE

- BRUN P. 1986, *La civilisation des Champs d'Urnes. Etude critique dans le Bassin Parisien*, Paris.
- CLOSE-BROOKS J. 1970, Compte-rendu de Hencken 1968, « Journal of Roman Studies » LX, 238-240.
- DJINDJIAN F. 1985a, *La sériation en archéologie: un état de l'art méthodes et applications*, in *Panorama des traitements des données en archéologie*, Juan les Pins, 9-46.
- DJINDJIAN F. 1985b, *Sériation and toposériation by correspondance analysis*, in A. VOORRIPS, S.H. LOVING (edd.), *To Pattern the Past*, « PACT », 11, 119-135.
- DJINDJIAN F. 1991, *Méthodes pour l'archéologie*, Paris, A. Colin.
- FERNANDEZ DE LA VEGA W. 1971, *Sur deux techniques de sériation*, CNRS, Centre d'analyse documentaire pour l'archéologie, Institut d'art et d'archéologie, 1970-71, 17 pp.
- FERNANDEZ DE LA VEGA W. 1977, *Deux algorithmes de sériation*, in *Raisonnement et méthodes mathématiques en archéologie*, éditions du CNRS, Paris, 157-170.
- GARDIN J.C. 1979, *Une archéologie théorique*, Paris, Hachetté.
- GASTALDI P. 1979, *Le necropoli protostoriche della valle del Sarno: proposta per una suddivisione in fasi*, « Annali del Seminario di Studi del Mondo Classico. Istituto Universitario Orientale di Napoli Archeologia e Storia Antica », 1, 13-55.
- GEBÜHR M., KAMPPFMEYER U. 1980, *Einsatz von Kleinrechnern in der Ur- und Frühgeschichtsforschung*, « Acta praehistorica et archeologica », 11/12, 3-20.
- GEBÜHR M., KUNOW J. 1976, *Der Urnenfriedhof von Kemnitz, Kr. Potsdam-Land. Untersuchungen zur anthropologischen Bestimmung, Fibeltracht, sozialen Gliederung und "Depot" sitte*, « Zeitschrift für Archäologie », 10, 185-222.
- GOLDMANN K. 1972, *Zwei Methoden chronologischer Gruppierung*, « Acta praehistorica et archeologica », 3, 1-34.
- GRAHAM I., GALLOWAY P., SCOLLAR I. 1976, *Model Studies in Computer Seriation*, « Journal of Archaeological Sciences », 3, 1-30.
- HODSON F.R., TYERS P. 1988, *Data analysis for archaeologists: the Institute of Archaeology packages*, in RAHTZ S.P.Q. (ed.), *Computer and Quantitative Methods in Archaeology*, Vol. I, Oxford, BAR, 31-61.
- IHM P. 1983, *Korrespondenzanalyse und Seriation*, « Archäologische Informationen », 6, Heft 1, 8-21.
- IHM P., (unter Mitwirkung von) LÜNING J., ZIMMERMANN A. 1978, *Statistik in der Archäologie. Probleme der Anwendung, allgemeine Methoden Seriation und Klassifikation*, « Archaeo-Physica », 9, Bonn.
- KENDALL D.G. 1971, *Seriation from abundance matrices*, in F.R. HODSON, D.G. KENDALL, P. TAUTU, *Mathematics in the Archaeology and Historical Sciences*, Mamaia 1970, Edinburg, Edinburg University Press, 229-264.

- LAGRANGE M.S. 1973, *Analyse sémiologique et histoire de l'Art. Examen critique d'une classification*, Thèse de Doctorat EPHE Paris 1970. T.A. Documents 3, Paris.
- LEGOUX R. 1980, *Le recours à l'informatique: la chronologie relative par permutation matérielle automatique*, in P. PERIN, *La datation des tombes mérovingiennes. Historique. Méthodes. Applications*, Genève, 138-149.
- PERONI R. 1979, *Osservazioni sulla cronologia della Prima Età del Ferro nell'Italia continentale*, in V. BIANCO PERONI, *I rasoi nell'Italia continentale*, « Prähistorische Bronze Funde », VIII, 2, München, 192-200.
- RUBY P. 1990a, *Le crépuscule des marges. Etude chronologique et sociale de la nécropole du Premier Age du Fer de Sala Consilina (Italie)*, thèse de Doctorat, Université Paris I — Panthéon - Sorbonne, Paris mai 1990, 3 volumes. A paraître aux Editions du Centre J. Bérard de Naples.
- RUBY P. 1990b, *Sériation par la méthode des moyennes réciproques sur Macintosh avec Microsoft Excel*, « Archéologues et Ordinateurs », 15, Centre de Recherches technologiques de Valbonne, 3-18.
- TINÈ-BERTOCCHI F. 1975, *Formazione della civiltà dauna dal X al VI sec.*, in *Atti del Colloquio Internazionale di Preistoria e Protostoria della Daunia*, Foggia 1973, Firenze, Istituto Italiano di preistoria e protostoria, 271-285.
- WILKINSON E.M. 1974, *Techniques of data Analysis — Seriation theory*, in E.M. WILKINSON, C.E. MULLINS, A. TABBAGH, *Technische und Naturwissenschaftliche Beiträge zur Feldarchäologie*, « Archaeo-Physika », 5, Bonn, 7-142.

ABSTRACT

The article describes the transformation of a number of procedures for archaeological data processing into Excel (Macintosh and MS Dos versions) macro programming language. Automatic seriation is dealt with first. The general principles of the two best known algorithms — the reciprocal averaging method and P. Ihm's correspondence analysis — are discussed. The Excel programme text is then presented and commented. After detailed analysis of the comparative merits of the two algorithms, a number of general remarks of a methodological nature are made about problems posed by seriation and particularly by automatic seriation techniques.

The article then turns to the construction of incidence (presence/absence) tables and co-occurrence matrices. Here again, general principles are discussed before presentation of the programme text and comments on the macro. The last macro presented carries out replacements of cell contents. The conclusion contains some general methodological remarks about Excel and the writing of macro-programmes in Excel's macro-programming language.