METODE DE CLASIFICARE AUTOMATĂ ÎN ARHEOLOGIE

*Rezumat**

Aplicînd o serie de metode statistice de clasificare ierarhică pentru diferite situaţii arheologice, în funcţie de atribute sau criterii, grupate în matrici, autorii prezintă unele aplicaţii ale matricilor în arheologie. În lucrare sînt explicate principiile construcţiei dendrogramei (exprimarea grafică a ierarhiei de clasificare), analiza factorială şi a serierii prin folosirea circuitului hamiltonian**. Pe baza acestor metode de clasificare mormintele sau formele din Cimitirul A de la Iclod, din epocă neolitică (LAZAROVICI 1977, p. 220—221, fig. 7—8) sînt grupate prin dendrogramă sau analiza factorială. În acelaşi fel sînt analizate şi alte situaţii arheologice: complexe cu materiale aparţinînd fazei timpurii a grupului neolitic de la Pişcolt (LAZAROVICI — NÉMETI 1983, p. 28); analizele pedologice ale unor diferite probe din staţiuni neolitice din Transilvania (vezi PICIU — LAZAROVICI 1988).

Folosirea metodelor moderne de prelucrarea datelor arheologice cu ajutorul ordinatorului din baze de date, fie din tabele a fost aplicată experimental pe şantierele arheologice ale secţiei de preistorie a Muzeului de istorie al Transilvaniei (la Iclod, Parţa, Piatra Ilişovii, Tureni).

Datele arheologice de la Iclod, din Cimitirul A, grupate în matrici calitative, de prezenţă/absenţă (boolene) au fost analizate prin cîteva studii (de către LUPŞE—LAZAROVICI 1985; LAZAROVICI 1986) prezentîndu-se principii de clasificare folosite în literatura de specialitate (FERNANDES DE LA VEGA 1978; 1978a; TAGBY 1965, 1971; CLARKE 1978; DORAN — HODSON 1975; ORTON 1980).

Datele arheologice de la Piatra Ilişovii şi din alte părţi, aflate în baze de date au fost prelucrate cu calculatorul pe baza matricilor de abundenţă (KALMAR 1985; 1987). Statisticile făcute pe materialele din săpăturile arheologice, începînd cu anul 1973, au fost prelucrate pe calculator abia în ultima vreme, în curs fiind un studiu asupra unui sistem de prelucrare complexă a datelor în care cercetarea arheologică stă în centrul atenţiei.

---

* Folosim acesta pentru a-i rezuma cititorului esenţa articolului. Metodă folosită în revistele moderne.

** Diferiţi coeficienţi de similaritate între două obiecte sînt definiţi şi folosiţi pentru a transforma datele primare într-o matrice de similaritate.

	1	2	3	4	5	6	7 8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
22% A		•		•		•	•		•	•			•	•			•	•	•	•	•				•	•	•	•
17% B1	•	•	•			•		•	•	•			•	•	•		•			•	•				•	•	•	•
5% B3							•		•					•											•			•
1% B4																						•						
15% C1	•				•	•	•			•			•	•	•	•					•				•	•	•	
8% C2		•							•								•						•		•			•
3% D1			•														•											•
22% D2											•																	
13% D3		•		•		•				•			•	•	•	•		•									•	
4% D4														•												•		•
1% E												•			•													
1% F1																•												
3% F2										•	•										•							
2% G			•																								•	
1% H									•																			
1% I																									•			
J																								•				

Tabel 1. Matricea formelor (pe linie) și mormintelor (pe coloană) din Cimitirul A de la Ielod.

Graves Nr.	Cluster 1					Cluster 2					Cluster 3					Cluster 4					Cluster 5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	x	x	x	x	x																				
2						x	x	x	x						x										
3							x						x							x		x			x
4								x	x			x			x	x									
5					x	x						x	x							x					
6		x	x	x	x						x														
7/8													x	x		x	x			x					
9	x		x				x		x						x										
10			x			x	x			x				x											
11		x			x			x			x				x										
12																					x	x	x	x	x
13																x	x	x	x	x					
14		x	x	x	x											x									
15		x	x	x						x	x														
16	x	x	x	x	x																				
17	x															x	x	x	x	x					
18						x	x	x						x											x
19							x	x				x			x		x								
20							x		x		x		x				x								
21						x	x		x		x			x											
22		x	x	x					x		x														
23																				x	x	x	x	x	
24																					x	x	x	x	x
25																					x	x	x	x	x
26	x	x	x		x				x																
27		x	x	x						x	x														
28					x							x	x				x			x					
29	x	x	x						x																x

Tabel 2. Clasificarea mormintelor din cimitirul A de la Iclod în diferiţi clusteri folosind diferite formule de calcularea similarităţii sau corelaţiei.

METODELE

Dendrograma este o construcţie arborescentă ce exprimă ierarhia de clasificare a atributelor în funcţie de situaţii luate două cîte două sau invers, prin intercorelarea atributelor care definesc asemănările sau deosebirile dintre obiecte.

Pentru construcţia dendrogramei se pleacă de la matricea de date (fie că este vorba de date calitative sau cantitative) care se transformă

într-o matrice de similaritate, o matrice numerică (mai zisă de tip Robinson: TAGBY 1971, 635 şi urm; 1966, p. 6—7; LUPŞE — LAZAROVICI 1985, p. 112), pe care matrice se fac calcule statistice. Din matricea de similaritate se calculează coeficienţii de similaritate sau indicele de corelaţie (sînt diferite posibilităţi de a calcula diferit coeficienţii de similaritate: LUPŞE — LAZAROVICI 1985, 115; DORAN — HODSON 1975, p. 139—140, FRENŢIU 1985, 30); ORTON 1980, p. 43; WILKINSON 1974, 24—26) pentru obiectele sau situaţiile de clasificat.

Intre oricare 2 obiecte O_i şi O_j se defineşte un raport (coeficient) de similaritate r_{ij}, Pentru cele m obiecte sau situaţii există m (m—1) coeficienţi de similaritate diferiţi care definesc matricea de similaritate

$$R = \begin{matrix} r_{11}, & r_{12}, & \dots & r_{1m} \\ r_{21}, & r_{22}, & \dots & r_{2m} \\ r_{m1}, & r_{m2}, & \dots & r_{mm} \end{matrix}$$

unde $r_{ii} = 1$ şi r_{ij} 1. 1.

Evident este faptul că cele mai legate între ele sînt acele obiecte i şi j pentru care coeficientul de similaritate este cel mai mare. Pornind de la acest criteriu dendrograma grupează cele două obiecte, grupe de obiecte sau situaţii arheologice obţinîndu-se o ierarhie de clasificare sub formă arborescentă (de unde numele de dendrogramă) (pentru detalii vezi FRENŢIU 1985; LAZAROVICI — LUPŞE 1985; ORTON 1980, 47—50).

REZULTATELE

Pentru calcularea indicelui de similaritate în prima situaţie a fost aplicată formula $S_{jk} = \frac{N_{jk}}{n}$ unde n reprezintă frecvenţa relativă şi N frecvenţa absolută (LUPŞE — LAZAROVICI 1985, p. 115 formula 1). În acest fel prin programele scrise de M. Frenţiu a fost obţinută dendrograma 1 reprezentînd ierarhia de clasificare a mormintelor de la Iclod în funcţie de formele ceramice. În cadrul ierarhiei de clasificare s-au obţinut patru grupe mari: în prima intră mormintele 1, 16, 9, 17, 5; în grupa a doua intră mormintele 2, 10, 26, 29, 18, în cea de-a treia intră 6, 11, 14, 15, 27, 22, iar în cea de-a patra M18, 28, 7—8, 20 (vezi *fig. 1)* avînd similarităţi diferite.

Aplicînd alte trei formule pentru calcularea coeficientului de similaritate sau a indicelui de similaritate generalizat folosit în statistică (pentru acesta vezi FRENŢIU 1985) au fost construite alte patru dendrograme *(fig. 2—4)* folosindu-se formulele:

$$2)\ S_{(j,k)} = \frac{N_{jk}}{N_j - N_k - N_{jk}}$$

$$3)\ S_{(j,k)} = \frac{N_{jk}}{2} \frac{1}{N_j} \frac{1}{N_k}$$

$$4)\ S_{(j,k)} = \frac{N_{jk}}{N_j \cdot N_k}$$

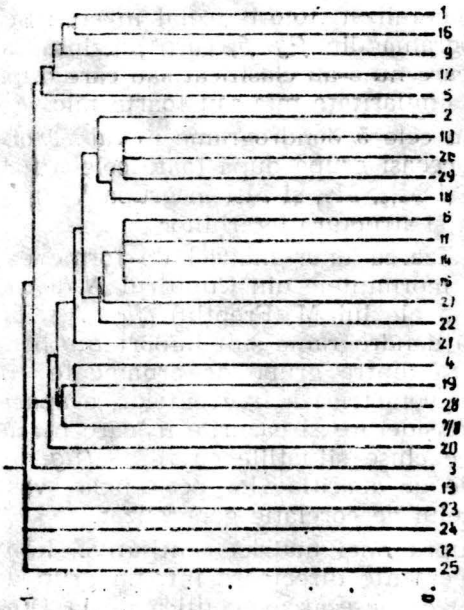

Fig. 1 formula 1

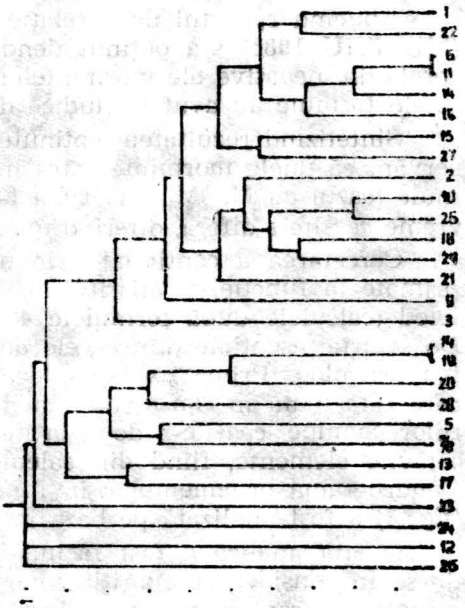

Fig. 2 formula 2

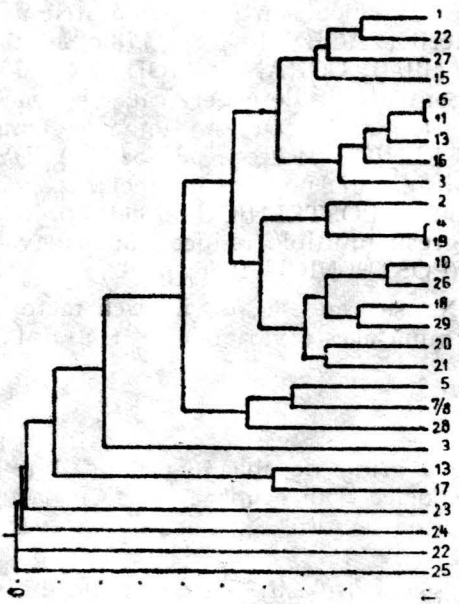

Fig. 3 formula 3

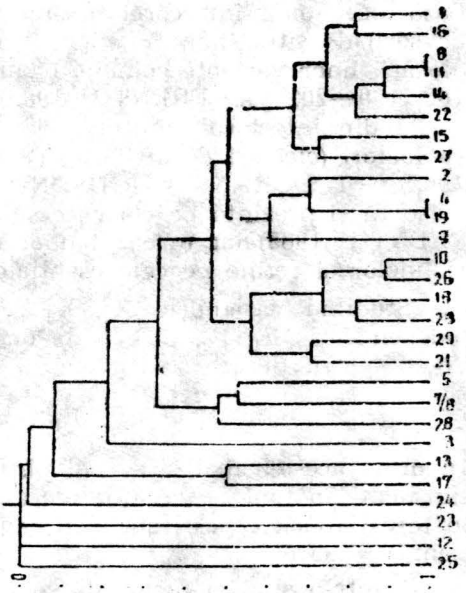

Fig. 4 formula 4

Dendrograma mormintelor

Aplicînd raportul de corelaţie generalizat, folosit uzual în statistică (FRENŢIU 1985) s-a obţinut dendrograma din *fig. 5*, care prezintă şi corelaţiile negative ale mormintelor care nu s-au clasificat sau care după unele formule au avut un indice de similaritate mic sau foarte mic.

Sintetizînd rezultatele obţinute în cele 5 dendrograme în *tab. 2* observăm că unele morminte intră în aceeaşi grupă după toate cele 5 formule (cazul cu M_1, M_{16} în grupa I-a şi M_{12}, M_{24} şi M_{25} în grupa V-a), în vreme ce altele diferă, diferind uneori şi structura ierarhiilor.

Cercetarea a continuat prin analizarea unor matrici de forme exprimate în funcţie de apariţia lor în mormintele din Cimitirul A de la Iclod (calculele după formulele 4 şi 5 ale lui M. Frenţiu) *(fig. 7 şi 8)*. Deosebirile esenţiale dintre cele două dendrograme sînt importante pentru arheolog. Prima exprimă legătura dintre grupe de ornamente în care este luată în considerare media geometrică a frecvenţelor, a apariţiilor comune, cea de-a doua ia în considerare şi cazurile în care nu au comune elemente, fiind din calcule excluse situaţiile cu zero *(fig. 8)*. Dendrograma ornamentelor în funcţie de apariţia lor pe unele vase *(fig. 8)* a fost analizată pe baza indicelui de corelaţie generalizat.

O altă analiză a fost făcută asupra unei matrici cantitative, luîndu-se în considerare apariţia unor elemente directoare într-un grup de complexe (gropi menajere şi bordeie) din aşezarea neolitică de la Dindeşti aparţinînd fazei timpurii a grupului Pişcolt pornindu-se de la tabelul apariţiilor ornamentelor fiind raportată corelaţia între complexe *(fig. 10* — datele preluate de la LAZAROVICI — NÉMETI 1983, p. 28).

Analiza factorială este un criteriu mai puternic de clasificare, fiind o metodă de analiză statistică (POSTELNICU 1966, p. 53 ş.a.) avînd la bază ideea unor intercorelaţii empirice de obiecte în funcţie de atribute, de serii de situaţii în funcţie de criterii (TAGBY 1965, p. 778). Soluţia alocării unor varianţe comune (comunalităţi: DORAN — HODSON 1975, 182, 199—201 ş.a.; FRENŢIU 1985) pentru a le descoperi cauzele, legăturile din interiorul mulţimii şi exprimarea lor cu un număr restrîns de factori (clasteri de atribute) (POSTELNICU 1966, p. 53—54; TAGBY 1965, p. 7: DORAN — HODSON 1975, p. 197) pe care îi precizează, îi măsoară şi prezintă pe cei care se opun (POSTELNICU 1966, p. 53; FRENŢIU 1985) iar în cazul unor regresii multiple reduce numărul variabilelor şi reţine pe cele esenţiale (POSTELNICU 1966, p. 62).

Studiind variabilele $X_1, X_2, \ldots X_n$ se cere să se găsească factorii independenţi: $F_1, F_2, \ldots F_k$ care determină aceste variabile ($K \Theta n$). Mai exact pentru fiecare i să avem

$$X_1 = a_{i1} \cdot F_1 + a_{i2} \cdot F_2 + a_{ik} \cdot F_k$$

În urma acestei analize se pot obţine k grupe de obiecte (sau atribute) exprimate în funcţie de situaţiile arheologice. Din grupa i vor face parte toate variabilele X_k care sînt determinate de factorul F_i, adică de un coeficient a_{ki} mare.

Printr-o analiză factorială se ajunge la o clasificare mai reală decît cea dată de dendrogramă întrucît un obiect poate face parte din două clase diferite cu alte grade de apartenenţă.

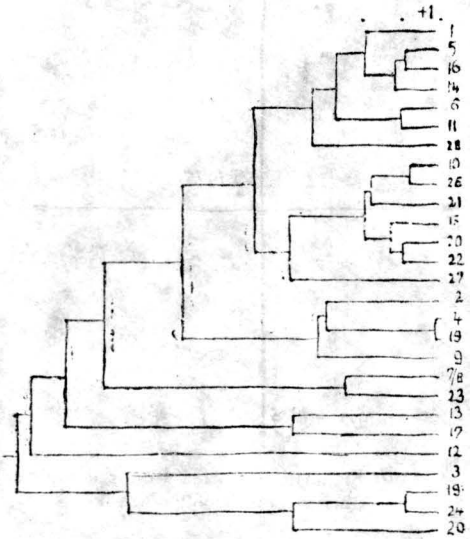

Fig. 5 formula calculată după indicele de corelaţie generalizat

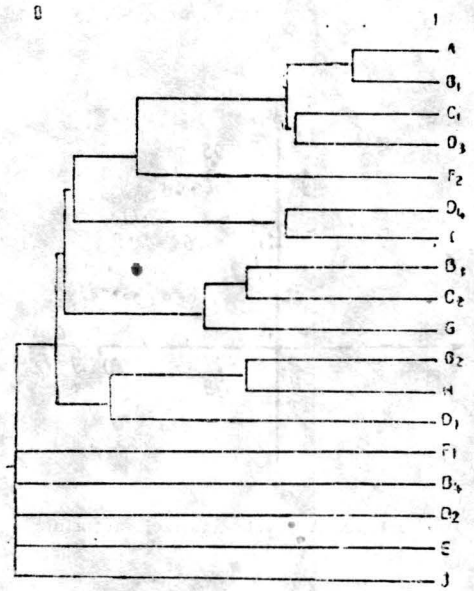

Fig. 6 Sinteza grupelor în funcţie de formula 3

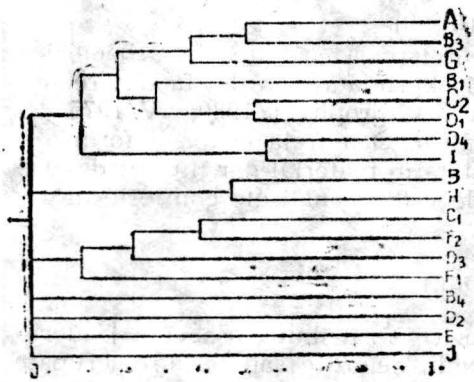

Fig. 7 Dendrograma formelor în raport de morminte (formula 4)

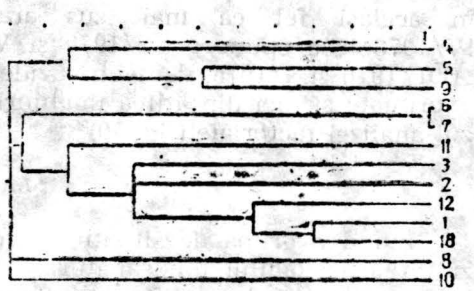

Fig. 8 Iclod A, dendrograma formelor (formula 5)

Prelucrări cu ajutorul calculatorului s-au făcut fiind aplicat în literatura de specialitate programul VARIMAX (DORAN — HODSON 1975, p. 200—201; IMREH 1984, p. 35; FRENŢIU 1985).

REZULTATELE

În urma rotaţiei ortogonale efectuată asupra matricei de corelaţie conţinînd corelaţia mormintelor în funcţie de formele de vase (datele de intrare de la LAZAROVICI 1977, p. 220 fig. 8) pentru un număr de 7 factori au dat următoarele situaţii: factorul 1 şi 2 explică un număr mare

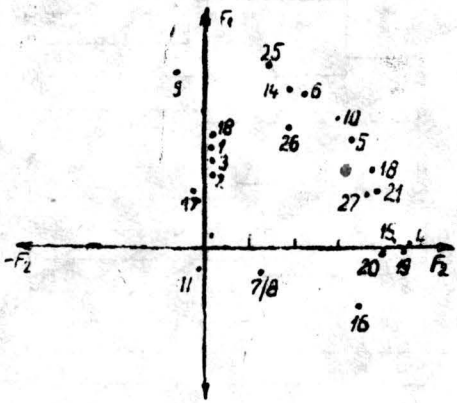

Fig. 9 Iclod A, axele analizei factoriale

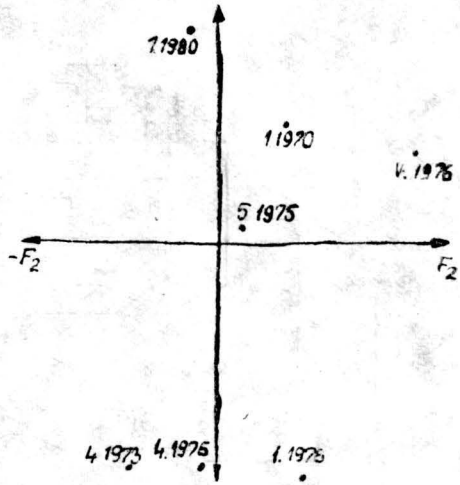

Fig. 10 Situaţia statistică a unor complexe din grupul Pişcolt, faza I

de morminte în vreme ce factorul 3 se opune factorului 1 în M 1, 3, 5, 18, 19 şi 20, factorul 2 în M 7—8 şi 16 iar factorilor 1 şi 2 în M 10 şi 13. Dacă am reprezenta valorile factorului 1 şi 2 în coordonate cartesiene am obţine cîteva grupări de morminte (fig. 9).

Analiza factorială a matricii de corelaţie a complexelor arheologice din grupul de complexe de la Dindeşti exprimată prin doi factori redată în acelaşi fel ca mai sus arată că gropii I/1980, V/1975 şi IV/1973 i se opune gr. 2/1970 şi V/1976. O a treia grupă o formează gr. 1/1976 şi 4/1976, deosebiri evidenţiate în material dar nu atît de clar exprimate şi nici din atîtea unghiuri de vedere analizate cum este în cazul analizei factoriale (fig. 10).

SERIEREA

Luînd o grupă de situaţii arheologice şi atribute, aşa cum i-au rezultat arheologului din săpătură, colecţii, depozite sau tezaure, grupate într-o matrice, ele pot fi ordonate operaţiile constînd din permutări de linii şi coloane, în mod succesiv, pînă la obţinerea unei grupări cît mai compacte. Natura, forma şi structura acestei grupări poate lua forme diferite, poate avea mai multe soluţii. Greutatea unei bune aranjări este determinată şi de numărul mare de operaţii în cazul unor matrici de mari dimensiuni. Asemenea operaţii s-au făcut începînd cu anul 1904, de către Petrie, de unde denumirea operaţiei şi a matricii: matricea Petrie sau matricea de date, în care se opera direct cu datele arheologice (MEIGHAM 1959, p. 203, WILKINSON 1974, p. 14, 22).

Începînd cu anii 1951 pentru a se scurta operaţiile, pentru a aplica calcule statistice iar mai tîrziu cu calculatorul, — s-au făcut noi încercări, matricea de date fiind înlocuită cu matricea numerică, cu matrici de similaritate, matrici de tip Robinson, Robinson — Breiner sau Belous

(TAGBY 1965, p. 3; DORAN — HODSON 1975, p. 276, 278; TAGBY 1971, p. 636; WILKINSON 1974, p. 24; ORTON 1980, p. 84).

Aceste matrici au fost analizate şi dezvoltate de Kendal în mai multe rînduri care a elaborat şi programe pentru prelucrarea cu calculatorul (apud WILKINSON 1974, p. 20—21, 24, 28—29). Soluţiile pentru obţinerea unei serii cît mai bune au avut ca termene definite prin „concentraţia principală", „curba de popularitate", măsurarea similarităţii, a disimilarităţii (ORTON 1980, p. 84). În acelaşi timp au fost aduse critici unora dintre soluţii: datele cronologice sînt preferate celor de similaritate; (WILKINSON 1974, p. 29); faţă de tipologie cronologia este mai importantă (IBIDEM).

Pentru calcularea similarităţii au fost elaboraţi diferiţi algoritmi, la fel pentru seriere: calcularea seriilor prin măsurarea similarităţii, a disimilarităţii (ORTON 1980, p. 83; DORAN—HODSON 1975, p. 135—138; 142, 276; ş.a.) precum şi calcularea seriilor prin diferenţă (KELLY — BUCCELLATI — ELSTER 1973, p. 195 şi urm.).

O metodă pe care am luat-o în studiu a fost aceea a alcătuirii seriilor prin calcularea circuitului Hamiltonian (WILKINSON 1974, p. 31, 38 şi urm.).

Presupunînd că există o ordonare a mormintelor se observă că două linii consecutive într-o serie au distanţa Hamming mai mică decît două linii care nu sînt vecine. Considerînd fiecare linie un punct în spaţiul R^n şi distanţa dintre puncte fiind distanţa Hamming dintre linii, ordonarea liniilor şi coloanelor are proprietatea că realizează circuitul hamiltonian de lungime minimă. În consecinţă, serierea revine prin efectuarea unor permutări de linii şi coloane la găsirea circuitului hamiltonian de lungime minimă. Cunoscînd din serie unele date relative (stratigrafia comparată) sau absolute (datare cu C_{14}, termoluminiscenţă, dendrocronologie etc.) se poate obţine o orientare cronologică a seriei şi implicit datarea seriei. In cadrul seriei respective se vor forma clasteri care vor defini fazele culturale ale civilizaţiei respective, vor fi precizate fazele de trecere sau etapele cronologice.

MILITON FRENŢIU — GHEORGHE LAZAROVICI

PRESCURTĂRI

CLARK 1978 — D. Clark, *Analithical Archaeology*, (1968), 1978.

DORAN—HUDSON 1975 — J. E. Doran — F. R. Hodson, *Matematics and Computers in Archaeology*, Edinburg, 1975.

FERNANDES DE LA VEGA 1978 — *Test d'une hypothése uniforme contre une hypothése unimodale et aplication en Archaeology*, în *Archaeology et Calcul*, 7, 1978, p. 133—135.

1978a — *Methodes de classification automatique en archaeologie*, în *Archaeology et Calcul*, 7, 1978, p. 147—165.

FRENŢIU 1985 — M. Frenţiu, *Program pentru prelucrarea datelor geologice*, în *Sesiunea de comunicări*, Universitatea din Cluj-Napoca, 1985, p. 29—34.

KALMAR 1986 — *Prelucrarea datelor arheologice de la Piatra Ilişovii cu ajutorul calculatorului*, Ziridava 18, 1986.

KALMAR—BAGOTSKY—LAZAROVICI 1987 — Z. Kalmar, Cl. Bagotsky, Gh. Lazarovici, *Cercetările de la Piatra Ilişovii, Banatica*, 9, 1987—1988.

KELLY-BUCELATI — ELSTER 1973 — M. Kelly-Bucelati, E. Elster, *Statistics in Archaeology and its Application to the Near Easter Data*, în *Orientalia*, 42, 1—2, 1973, p. 195—211.

LAZAROVICI 1977 — Gh. Lazarovici, *Sfîrşitul culturii Vinča—Turdaş în Cîmpia Transilvaniei*, în *Studii şi comunicări Caransebeş*, II, 1977, p. 211—230.

LAZAROVICI 1986 — Gh. Lazarovici, *Aplicaţii ale matricilor în arheologie, Sesiunea de Comunicări*, 1986, Universitatea din Cluj-Napoca.

LAZAROVICI — NÉMETI 1983 — Gh. Lazarovici — I. Németi, *Neoliticul dezvoltat în nord-vestul României*, în *ActaMP* VII, 1983, p. 17—60.

LUPŞE — LAZAROVICI 1985 — I. Lupşe — Gh. Lazarovici, *Aplicarea unor metode moderne de clasificare, prelucrare şi interpretarea unor date arheologice*, în *Sesiunea de comunicări*, Universitatea din Cluj-Napoca, 1985, p. 109—126.

MEIGHAM 1959 — C, W. Meigham, *A new Method for the Seriation of Archaeological Collection*, în *American Antiquity*, 25, 1959, Pc „O/!"LLc.

ORTON 1980 — Clive Orton, *Matematics in Archaeology*, London 1980.

POSTELNICU 1966 — V. Postelnicu, *Analiza factorială şi posibilităţile ei de aplicare*, în *Revista de statistică*, 10, 1966, p. 53—62.

TUGBY 1965 — J. D. Tugby, *Archaeological objectives and Statistical Methods. A frontier in Archaeology*, în *American Antiquity*, 31, 1, 1965, p. 1—16.

1971 — *Archaeology and Statistics*, în *Science in Archaeology*, London, 1971, p. 635—648.

WILKINSON 1974 — E. Wilkinson, *Techniques of Data Analizis Seriation*, în *Arheo Phisica*, 5, 1974, p. 1—134.