

**A METHOD OF STRUCTURAL ANALYSIS  
WITH AN APPLICATION  
TO  
“LES LIAISONS DANGEREUSES,”**

As the text itself becomes more and more important in literary criticism, the concept of structure becomes more and more popular among critics. They are all interested in the structure of texts but there are few precise definitions of structure and even fewer successful attempts at finding the structure of a text. It seems to us that mathematics can help in this domain. Up to now, with a few notable exceptions, only elementary mathematical tools have been applied to the study of literature, mainly in the counting of the frequencies of words to develop concordances.

We present herein a mathematical method for discovering the structure of literary works. Though it can be applied to any kind of literary work it is especially well suited to the study of plays or novels. We also present an illustration of the method by applying it to an epistolary novel, *Les Liaisons Dangereuses*, it is however to be understood that the method is not restricted to novels only.

A novel contains such elements as plot, character, description, situation. Following the usual mathematical terminology we will call all these elements variables. If we consider a novel as a succession of periods unfolding through time, it is evident that, as the variables change throughout the novel, the degree of similarity between periods varies. Some periods are similar because they have the same variables, others are not. We will

call structure the pattern of similarity - dissimilarity between periods throughout the novel. It is to be understood that one can find several structures in a literary work, depending on the variables being studied. We are aware that it is possible to define the structure of a novel in many other ways, but this definition has the great advantage of being operational. It allows a precise and objective description of a structure and it makes it easier to compare the structures of different literary works.

Because a novel is characterized by many variables, the methods for finding the structure are usually impressionistic. The critic surveys the novel and abstracts his own impression of the structure, but since he cannot keep in mind all the changes occurring in all the variables he is bound to resort to a subjective analysis underestimating certain points and overestimating others. But if the human mind cannot deal successfully with all the variables at once, the computer can. After analyzing all the variables the computer can decide which of the periods are similar and which are not. Furthermore, it can plot the results on a graph and show a pictorial representation of the structure.

Before presenting an illustration from *Les Liaisons Dangereuses*, we will give a brief outline of the method. It involves five successive steps.

#### *1. Selecting the variables.*

The choice of the variables to be included in the analysis is determined by the nature of the novel being studied, and the possibility of measuring these variables. To give an example, since *Les Liaisons Dangereuses* is an epistolary novel the number of letters exchanged by the main characters could be of some interest to the critic. The number of letters, then, would be one of the chosen variables.

### 2. *Partitioning the novel.*

The novel must be divided into successive periods from beginning to end. This partitioning can correspond to the divisions made by the author, himself, such as chapters of a novel or acts of a play ; but the critic is free to divide the work anyway he chooses.

### 3. *Measuring the variables.*

The critic assigns a number to each of the variables, within every period. To take the same example as above, if one of the variables is the exchange of letters between Mme de Merteuil and Valmont in *Les Liaisons Dangereuses*, and if in the first period they exchange 6 letters, then the value of this variable in the first period is equal to 6. The same operation is performed for all the variables in all the periods. Thus, if there are 30 variables, each period of the novel will be characterized by 30 numbers.

### 4. *Finding the structure.*

Since the structure of the novel has been defined as the pattern of similarity - dissimilarity among the periods, it is necessary to determine the degree of similarity between any two periods. This degree is indicated by a number called coefficient of similarity. It is computed from the values assigned to the variables in step 3. The values of this coefficient throughout the novel indicates the overall pattern of similarity between the periods that is to say the structure.

### 5. *Graphical representation of the structure.*

The periods of the novel can be represented by points on a graph and the distance between these points can be made "proportional" to the degree of similarity between the corresponding periods (see mathematical appendix). If two periods are very dissimilar in terms of the variables, they will be represented on the graph by points very far apart. If there is

little change from one period to another, then the two points will be close.

### STRUCTURE OF *LES LIAISONS DANGEREUSES*

For convenience we will follow the steps indicated in the outline.

#### *1. Selecting the variables.*

The structure of an epistolary novel is determined by the number of letters exchanged by the characters. This fact has led us to emphasize these variables. We have also studied other variables such as the relationship between the different plots and the feelings of like-dislike between the main characters. For lack of space, however, only the results concerning the exchange of letters are given here.

#### *2. Partitioning the novel.*

Except for a few limitations that will be discussed later, the critic is free to divide the novel the way he wants. In *Les Liaisons Dangereuses*, the letters are dated, spanning twenty weeks. But using the week as the unit of division would have broken the exchange of letters in a piecemeal way, so 10 two-week periods were chosen. Furthermore, the last letter of the novel was not considered because it comes almost one month after the previous letter ; and its uniqueness makes its place in the total correspondence easy to see.

#### *3. Measuring the variables.*

For each period, we have to determine who writes to whom and how many times. The number of letters within each period is shown in table 1.

TABLE 1.

PERIODS	1	2	3	4	5	6	7	8	9	10
Merteuil to Danceny						1			1	
Merteuil to Cécile	1				1					
Merteuil to Valmont	3	3	3	3	1	1	2	1	4	
Merteuil to Volanges				1	1					
Valmont to Danceny				1					1	
Valmont to Cécile				2						
Valmont to Merteuil	3	7	6	2	4	1	3	2	5	
Tourvel to Valmont		3	3	2				1		
Valmont to Tourvel		5	3	3			1			
Tourvel to Rosemonde					2	2	2	2	1	
Tourvel to Volanges	2	3								
Rosemonde to Danceny										1
Rosemonde to Tourvel					1	3	2			
Rosemonde to Volanges										1
Volanges to Danceny			1							
Volanges to Merteuil					1					
	9	21	16	14	11	8	10	6	12	2

TABLE 1 (2nd part).

PERIODS	1	2	3	4	5	6	7	8	9	10
	9	21	16	14	11	8	10	6	12	2
Volanges to Tourvel	1	1								
Volanges to Rosemonde									4	4
Danceney to Cécile		4	2	2		1				
Danceney to Merteuil						1			2	
Danceney to Valmont			1	1					2	
Danceney to Rosemonde										2
Danceney to Volanges			1							
Cécile to Danceney		2	2	1	1	1			1	
Cécile to Merteuil	1	1			2					
Cécile to Valmont				2						
Cécile to Sophie	4	4	2	1						
Letters exchanged between secondary pers.				1	3	2			1	4
TOTAL	15	33	24	22	17	13	10	6	22	12

Table 1 : Number of letters exchanged during the 10 periods  
(example : Merteuil sends Valmont 3 letters during period 1)

The 10 columns of the table correspond to the 10 periods of the novel. Each row shows the exchange of letters between two characters. It can be seen that during the first period Mme de Merteuil writes Valmont 3 letters and receives 3 from him, Mme de Tourvel sends Mme de Volanges 2 letters and so forth for a total of 15 letters. The few letters sent to, or received by secondary characters have been regrouped into a single variable. Since table 1 has 28 rows, 28 different variables must be considered in the analysis.

#### *4. Finding the structure.*

We must now compare the 10 periods that is to say the 10 columns of table 1, and determine their degree of similarity, taking into account the 28 variables. It is easy to follow the exchange of letters between only two characters throughout the novel, but it is very difficult to do so for 28 variables simultaneously. We must find a way to summarize the many changes occurring between any two periods. This can be done by computing a coefficient of similarity between the two corresponding tables. We have used a coefficient  $s$ , whose exact formula is given in the mathematical appendix. Roughly speaking the coefficient of similarity  $s$  between two periods is the average difference between the corresponding two tables.

The coefficient of similarity should not reflect the difference between the total number of letters in each period, but it should rather show the difference between the patterns of exchange. Therefore the percentage of letters within each period has been used instead of the number of letters.

Table 2 indicates the values of the coefficients of similarity between any two periods of the novel.

TABLE 2.

		PERIODS									
		1	2	3	4	5	6	7	8	9	10
PERIODS	1										
	2	.44									
	3	.59	.25								
	4	.73	.41	.35							
	5	.62	.64	.65	.71						
	6	.85	.71	.69	.68	.47					
	7	.60	.60	.52	.67	.53	.49				
	8	.63	.61	.50	.68	.59	.69	.33			
	9	.62	.65	.56	.59	.58	.59	.54	.56		
	10	1.00	1.00	1.00	.95	.82	.85	1.00	1.00	.77	

Table 2 : table of coefficients of similarity between all pairs of periods.  
 (Example : the coefficient of similarity between period 2 and period 3 is .25)



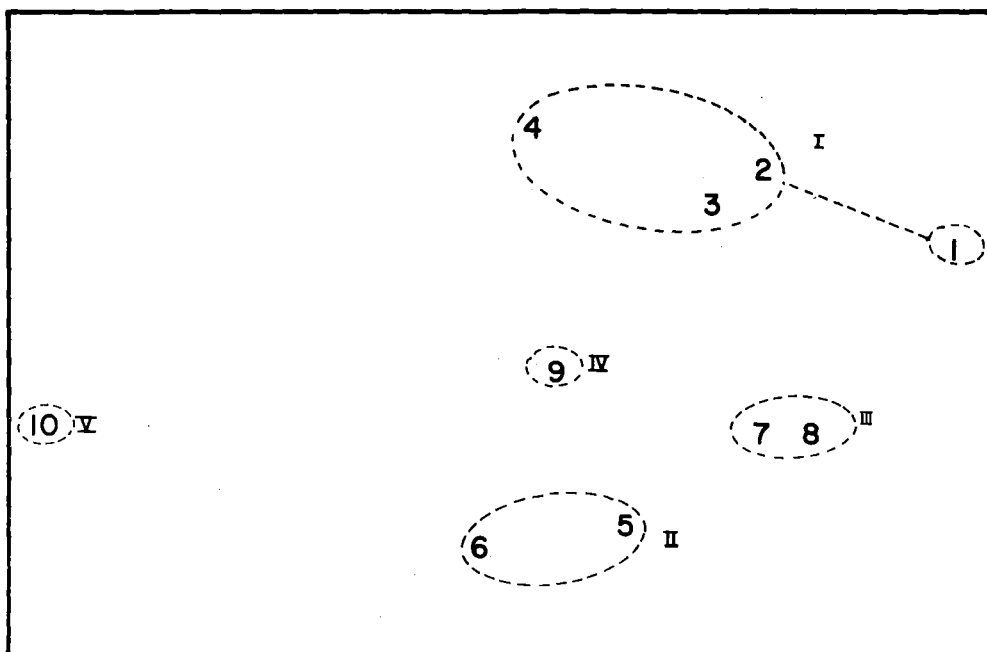
Since the coefficient of similarity  $s$  represents the average difference between two periods, the smaller the value of  $s$ , the more similar the periods, as far as the exchange of letters is concerned.  $s=0$  indicates that the two periods are identical ;  $s=1$  indicates the maximum difference in the novel. It can be seen that  $s=.44$  for periods 1 and 2, and  $s=.25$  for periods 2 and 3 ; these figures show that the patterns of exchange of letters are more similar between periods 2 and 3 than between periods 1 and 2. Notice that table 2 is symmetrical since the value of  $s$  between periods 1 and 2 is the same as the value of  $s$  between periods 2 and 1. Therefore only one half of the table is given. Table 2 can be considered as a mathematical description of the structure of the novel, but most people would find it difficult to interpret, especially when the novel has been divided into many periods. Fortunately it is possible to represent the table on a graph that is easy to comprehend.

##### *5. Graphic representation of the structure.*

We can obtain a graph on which the 10 periods are represented by 10 points in such a way that the distance between any two points varies with the value of the coefficient of similarity between these points. If the periods are similar the corresponding points are close ; if the periods are dissimilar the points are far apart. Such a graph appears as a pictorial representation of the structure.

Graph 1 shows the structure of *Les Liaisons Dangereuses*.

GRAPH 1.



The 10 periods of the novel are represented by 10 points which we indicate by numerals from 1 to 10. The computer has placed the points so that distances between them show the degree of similarity between periods. The dotted lines have been added to show how the 10 periods form 5 groups corresponding to the 5 "acts" indicated by roman numerals.

A first look at the graph shows five groups of different moments in *Les Liaisons Dangereuses*, as if the novel unfolded like a five act play. The first act would include period 1 and periods 2, 3, 4. The fact that period 1 stands apart from the second group, indicates that important changes occur between period 1 and periods 2, 3, 4. The second act consists of periods 5 and 6, situated in a different part of the graph, far from periods 2, 3, 4. Periods 7 and 8 constitute the third act. Period 9 stands by itself and can be said to be the fourth act. Period 10 is more apart on the graph than any other period, which means that important changes occur in what can be considered the fifth act. These five acts are characteristic of five important aspects of the structure of *Les Liaisons Dangereuses*. As has been said before, these results obtained by an objective method define the structure for a set of variables, the exchange of letters, which is representative of the form of the novel. Yet it will be seen that these five acts do correspond to five characteristic aspects of the content of the novel. We will now examine the relationship between the form, as shown by the graph, and the content.

*A. Act 1 : periods 1, 2, 3, 4.*

Period 1 cannot really be close to any others of the novel since it has two specific features ; first it contains the introduction to the novel, second, the main characters do not write to each other. If we except the letters that Valmont and Mme de Merteuil send to each other, and that will be found regularly throughout the novel, there are 3 letters between Mme de Tourvel and Mme de Volanges, and 4 between Cécile and Sophie. Mme de Volanges and Sophie can be considered as confidants since they do not play an active part in the novel, and soon disappear when Mme de Tourvel and Cécile do not need them anymore.

The apparition of confidants in the exposition of the novel shows that the epistolary novel has some connection with a theatre play : the characters express life while they are actually living it, and there is also a need for confidants to inform the reader, at the beginning of the first act.

The introductory scene apart, the first act groups periods 2, 3, 4 together. Valmont and Mme de Merteuil go on writing to each other, while organizing the plot. Three important events occur through the correspondence of the characters involved : Mme de Tourvel's love for Valmont, Cécile's love for Danceny, and Danceny's love for Cécile. At the same time, the number of letters to the confidants either decreases (Cécile to Sophie), or disappears altogether (Mme de Tourvel to Mme de Volanges, in periods 3 and 4). Therefore, periods 2, 3, 4 appear to be, as it is usual in the first act of a play, periods in which the action develops, and in which all the main characters appear.

*B. Act 2 : periods 5, 6.*

Periods 5 and 6 are different from periods 2, 3, 4, as can be seen from the distance between those two sets of periods. The amount of letters decreases and the pattern of correspondence is quite different. The correspondence between Valmont and Mme de Merteuil goes on, but new elements appear. Mme de Tourvel and Mme de Rosemonde start writing to each other, Cécile sends two letters to Mme de Merteuil, Danceny writes his last 2 letters to Cécile and 5 letters (a relatively significant number : 5 out of a total of 30), are related to secondary characters, such as Azolan, le comte de Gercourt and le Père Anselme. These new elements constitute what we will call the second act. It is an act where nothing extremely important happens (Cécile's seduction by Valmont is not important in itself, except for what it reveals as far as Valmont's feelings are concerned), but where everything is prepared. A net is woven around Mme de Tourvel, who be-

comes Valmont's prey : she is in love with him and he carefully prepares her fall by gaining the confidence of her maid and confessor (this explains Valmont's letter to Azolan and le Père Anselme.) A net is also woven around Danceny, since not only does Valmont take Cécile from him, but Mme de Merteuil entwines him in the comedy of love. Therefore during this second act, everything is being organized and the reader expects that something important will be occurring soon.

*C. Act 3 : Periods 7 and 8.*

The number of letters continues to decrease : periods 7 and 8 have only 10 and 6 letters, respectively. This number is very small in view of the number of letters that appear in other periods (33 in period 2, or 23 in period 3, for example). This is because all the correspondence stops except for the one between Valmont and Mme de Merteuil, Mme de Rosemonde and Mme de Tourvel and Valmont and Mme de Tourvel. This correspondence is concerned only with the affair between Valmont and Mme de Tourvel. It is as if everything else had disappeared from the stage, Cécile's love for Danceny, Valmont's affair with Cécile, even Valmont's interest in Mme de Merteuil, so that the attention is focused on what appears to be the central event in the novel : the relationship between Mme de Tourvel and Valmont. The net around Mme de Tourvel closes during the third act. The graph shows that period 7 and 8 are situated at a change of direction, reflecting a key turn in the structure of the novel, at the very moment when the relationship between Mme de Tourvel and Valmont reaches a crucial point.

*D. Act 4 : period 9.*

Period 9 stands by itself, which means that it is different from any other period in the book. It differs from period 8 in that it groups a large a-

mount of letters (22 as opposed to 6) and presents a new correspondence between Mme de Volanges and Mme de Rosemonde. Period 9 is rich in events as well as in letters and can be considered as the fourth act, in which the intrigues come to a culminating point, without reaching their final resolution yet. It is truly the climax of the novel.

*E. Act 5 : period 10.*

The distance between period 9 and period 10 is the largest between two consecutive periods. Therefore these last two periods are very different from each other. The difference is not surprising since period 10 constitutes the last period in the novel, and therefore contains both the last act and the conclusion (if we keep apart letter 175 which would constitute in itself period 11). Since Valmont died at the very beginning of period 10, the correspondence between Valmont and Mme de Merteuil, which has been a constant element throughout the novel, comes to an end. The reader is kept informed of the events thanks to the letters exchanged between Mme de Rosemonde and Danceny, and Mme de Volanges and Mme de Rosemonde. This last period appears to be symmetrical to the first one, not only in the number of letters, but also in the part played by confidants, such as Sophie in period 1, and Mme de Rosemonde in period 10. Letter 175 dated almost one month after letter 174, comes as an epilogue. It is not indicated on the graph, but it would stand very far from all the other periods.

The structure of the pattern of letters in *Les Liaisons Dangereuses* brings to light two important matters : first, that the novel unfolds like a five act play, in which each act is carefully prepared by the one which precedes ; second that the first and the last act are symmetrical in the main part played by confidants. The graphical representation of the structure of *Les Liaisons Dangereuses*, confirms Jean-Luc Seylaz' definition of this work

as a "géométrie sensible" (1).

### METHODOLOGICAL ASPECT OF THE ANALYSIS

In this paragraph we will examine two topics : how to choose and measure variables, and the consequences of dividing the novel into different sets of periods.

#### *The Selection and measurement of variables.*

A variable is any element that the critic is studying in the novel, such as plot, description, situation, or as in *Les Liaisons Dangereuses*, the correspondence between characters. In order to determine the structure, the variables must be measured in some way. This has probably been the greatest obstacle to the use of mathematical methods in the study of literature. Most critics will say that it is impossible or meaningless to translate a text into numbers. We think that there is a great deal of dogmatism in this assertion, and that it is mainly due to a lack of familiarity with numbers. It all depends upon the kind of variables which is being measured. We will consider two types of variables.

#### *1. Variables whose measure is perfectly objective.*

The number of exchanges (either epistolary or oral) between the characters of a work fits into this category. It can be exactly counted in at least two types of literary work : epistolary novels and plays, since it is indicated by the author himself. In many novels it can also be easily determined by the critic.

Many dichotomous variables can be objectively measured too. A dichotomous variable is a variable which can only take two values : for example, the absence or presence of a characteristic. This type of variable appears very frequently in the analysis of a literary work. To give a few examples : a given situation is either present or absent, a type of plot is either present or absent, a character either or does not have a certain trait, he either plays a certain role or does not, two given characters either interact together or do not, or they either interact in a certain way or do not. If the characteristic is present a value of one is given to the variable, if the characteristic is absent, a value of zero is attributed to the variable. Thus if we measure whether or not a character plays a specific role in each period, this variable is measured throughout the novel by a sequence of ones and zeroes : one, if he plays this role, zero if he does not.

*2. Variables whose measure involves a judgment of the critic.*

In this category falls the case of a characteristic whose presence or absence is not absolutely sure. Another frequent example is the rating scale. Here the critic attributes a number of points according to the degree to which a situation reflects a characteristic. An example of rating scale would be the study of the feelings of like-dislike between characters. It is often believed that this type of measurement is worthless because it is purely subjective. The degree of objectivity of such a judgment can be evaluated easily by having several people rate the same object independently. If their ratings disagree then the measure is not reliable enough to warrant further treatment. If they agree then the measure can be considered as objective since one definition of an objective measure is that it is unaffected by the idiosyncrasies of the critic performing the measure. This method of evaluating the reliability of a scale has been used in psychology for a long time.



*Partitioning the novel into periods.*

The critic is free to divide the novel any way he wants to, but there are a few rules that must be followed. The length of any period should be small enough to be homogeneous, but large enough to avoid having too many periods. The division into periods should naturally fit the novel or the play ; for instance, one should not put in the same period two "scènes" of a play, if they are very unlike. Finally if one wants to compare the structure of several works, their divisions must be in some way comparable.

The usefulness of the method can be readily appreciated if one compares the table of exchange of letters (table 1) with the graphical representation of the structure (graph 1). A critic who wants to find the structure of exchange of letters must necessarily start from table 1. It is obvious that he will have great difficulty in abstracting the structure from this table. On the other hand the graph shows him very clearly how the different periods of the novel are related.

As has been said before different variables will lead to different structures. It is possible to study variables representative of the form and variables characteristic of the content. Comparing the two types of structure provides a method to determine the relationship between form and content. Although we have chosen to illustrate the method with a set of variables representative of the form only, it has been possible to show the interaction between form and content.

This method is not limited to novels. It can also be used to describe the structure of plays and other literary works. It might be interesting to compare structures within genres and authors. This method checks on subjective interpretations and brings to their right proportions elements that

an impressionistic appreciation is bound to distort. Since the method gives coefficients of similarity between periods, the literary critic may use these coefficients to group periods according to their similarity and therefore to divide the work into nonarbitrary, homogeneous parts. The same coefficients can be examined so as to determine how fast changes occur throughout different periods. This is an objective way to appreciate the rhythm or tempo of literary work. This method uncovers structures which are not evident to the reading eye, but nevertheless exist, and are worth being investigated.

Christiane and Claude ALLAIS  
University of California

## MATHEMATICAL APPENDIX

This section deals with methodological problems and it could be skipped by readers not interested in the mathematical details of the method.

From a mathematical viewpoint, the novel is represented by a  $n \times p$  matrix, where  $n$  is the number of variables and  $p$  is the number of periods. The structure is defined by the  $p(p-1)$  similarity coefficients between the  $p$  periods. In order to facilitate its interpretation, the table of similarity coefficients is submitted to the non metric analysis program SS A1 (Smallest Space Analysis 1). (2) This program, devised by L. Guttman and J. Lingoes, represents the  $p$  periods by  $p$  points placed into a space of the smallest possible dimension. (3) It requires that the distances between the points be a monotonic function of the value of the corresponding coefficients. A monotonic function means that the order of the magnitudes of the distance is the same as the order of the magnitudes of the coefficients. Therefore it is not quite true to say, as we did before in order to simplify, that distances between points are proportional to the coefficients. Let us illustrate this point by an example : the similarity coefficient between periods 1 and 32 is .44, and the coefficient between periods 2 and 3 is .25 ; instead of plotting the points such as that the distances 1-2 and 2-3 are exactly in the ratio .44/.25, it is just required that distance 1-2 be larger than distance 2-3. The advantage of requiring conservation of order rather than strict proportionality is that it allows the program to put the  $p$  points in a space of fewer dimensions. In the example given the space has only two dimensions. This allows a very simple representation of the structure in a plane.

The program puts the points in a space of the smallest possible dimension. If the dimensionality is 1 or 2, the graphical representation is likely to be understood by everybody, but if it is 3 or more, most people without mathematical training are unable to visualize the graph and the main advantage of a graphical representation is lost. Two solutions are possible. Either one sets up the program so as to yield fewer dimensions, in which case the graph will only approximate the table of coefficients, or one does again the same analysis after reducing the number of periods. Since the number of periods is arbitrary, the last method is usually to be preferred. In the example, we have compromised between the two : in order to keep the number of periods equal to 10 we have settled for an approximation of the table of coefficients ( $\Phi = .007$ , coefficient of alienation = .12).

A few words should be said about the coefficient of similarity. Its mathematical definition is :

$$S_{jk} = \frac{\sum_{i=1}^n |V_{ij} - V_{ik}|}{\sum_{max}}$$

where

$$\begin{aligned} V_{ij} &= \text{value of the variable } i \text{ in period } j \\ i &= 1, 2, 3, \dots, n \\ j, k &= 1, 2, 3, \dots, p \\ \sum_{max} &= \text{maximum value of } \sum_{i=1}^n |V_{ij} - V_{ik}| \end{aligned}$$

in the novel. In other words the coefficient of similarity between two periods  $j, k$ , is the sum of the absolute differences between the value of the variables. This sum is divided by its maximum value in order to keep the value of  $s$  between 0 and 1.

### Footnotes

1. Seylaz, Jean-Luc, *Les Liaisons Dangereuses et la création romanesque chez Laclos*, Paris, 1958, p. 27.
2. Lingoes, J.C., *An I.B.M. 7090 program for Guttman-Lingoes smallest space analysis*, I. Behav. Sci., 1965, 10, 183-184.
3. Lingoes J.C., *New computer development in pattern analysis and non-metric techniques*, in *Uses of Computers in Psychological Research*, Gauthier - Villars, Paris, 1966, 1-22.