Czekanowski's Diagram
a Method of Multidimensional Clustering

Arkadiusz Soltysiak, Piotr Jaskulski.
Dept. for Historical Anthropology. Inst. of Archlology. Warsaw University.
Krakowskie Przedmiescie 26/28. 00-921. Warsaw 64. Poland.
E-mail:solar37@plearn.edu.pl

Abstract
At the beginning of this century, Polish anthropologist, Jan Czekanowski, proposed a simple and effective method of clustering objects, in a multidimensional space of characteristics. Originally, it was intended for examining the structure of a series of human remains, but also, other applications have been found (among them, for use in the biological sciences, ceramology and archeological typology).
This paper contains a description of Czekanowski's method from, input data set to an output diagram, representing a projection of multidimensional space, on two-dimensional plane. A new algorithm for clustering objects is also presented, together with a short characteristic of suitable software. The theoretical description is supplemented, by some practical remarks and a simple case study.

Introduction

The typology of various elements, of both, social and natural, human environment, is surely, one of the main issues for all empirical sciences. Astronomers perform a typology of stars, zoologists, for animals, linguists, for languages, rag-pickers, for rubbish, while archaeologists perform a typology for anything that falls into their hands. We can also typologize, the typology itself, and distinguish its two main types: deductive and inductive. Deductive typology occurs, when we employ a clear theoretical basis, when the classes are closed and exclusive, and their number fixed. On the contrary, in the case of inductive typology, the classes are empirical, and often unreliable; their number and ranges can be altered, with new classified objects. This typology is always used when a theoretical basis is insufficient, or doesn't even exist; so, in decidedly most cases.

From a methodological point of view, inductive typology is far more difficult and deceptive, than deductive typology. However, there is nothing to do about id but reconcile oneself to this situation, and try to find the best possible method, of distinguishing and defining classes, especially in a case, where the internal structure of data is completely unknown. Then, mathematical methods, of multidimensional clustering, can be applied. All of these methods are based on the assumption, that the characteristics of objects, inside one class (cluster), must have lower variability, than in all the series. Besides widely known algorithms of multidimensional clustering, one can mention, especially, tree clustering, which presents a series, in the form of a dendrogram and a k-mean clustering. Both are not perfect: the first one gives us hierarchic structure, which can be far from a factual distribution of objects, in variable space; the second one requires the number of clusters at input.

There also exists a third method of multidimensional clustering, which I would like to discuss here. It was proposed by Polish anthropologist, Jan Czekanowski, in 1909, and is used mainly by scholars, from the Polish School of Anthropology. For at least half of this century, it also has been adapted for use by archaeologists, biologists, and sociologists. Of course, Czekanowski's method is not perfect, and cannot be applied in all cases. Nevertheless, it seems worthy of notice, mainly because of two features: it can work with an incomplete data set, and the result – a two-dimensional diagram - is easy to interpret.

Czekanowski's Diagram

First, a short methodological description of Czekanowski's method should be presented. The algorithm is very simple, but it takes a lot of time, without the use of modern computers. One should make use of an input set of data, consisting of a series of objects of the same kind (that means, characterized, by the same variables), and after this, measure the distance between all possible pairs of objects. There are a lot of methods to measure distance, in multidimensional space, but most of them are not easy to use. The original formula, proposed by Czekanowski, was a simple City Block distance, divided by the number of variables, describing both confronted objects. This stipulation is necessary, if we take into account the possibility of examining a series, containing several objects, devoid of some measurements (e.g., broken bones). Here is the mathematical formula:

[1] $$DD = \frac{1}{n}\sum_{j=1}^{n}\left|M_{1j} - M_{2j}\right|$$

where DD indicates the average difference between two objects, resulting from elementary differences between their attributes; n, the number of variables (attributes), taken into account; M_{1j}, the value of j attribute, for the first object; M_{2j}, the value of j attribute, for the second object. The differences between the attributes of objects can be squared, to emphasize the lowest values of difference (i.e., most close objects), standardized and/or weighted, in order to make the variables, of various ranges, comparable.

Reflecting on Czekanowski's method, one can aknowledge that Euclidean distance might be more suitable. It is quite possible, but we must keep in mind, that this method came into being at the beginning of this century, when calculating Euclidean distance, for all pairs of objects from a series larger than ten to twenty items, and bearing an adequate number of variables, would have been too time consuming.

Moreover, Euclidean distance cannot be used, when some attributes, of some objects, are missing.

Thus, the first step is to calculate the distance values, for all pairs of objects, from an examined series. They can be formed into a square matrix, which afterwards, must be adequately arranged. The matrix represents a kind of projection, of multidimensional space in a two-dimensional plane. All values at the diagonal are equal zero, because each object is identical to itself. But, other cells in the matrix can have any zero or positive values. Now, the task is to reorder the matrix in a way so that the closest objects are neighbours, in the matrix. At first glance this would seem to be easy, but in reality, we must keep in mind that we are dealing with multidimensional space, and the arrangement of objects is not linear (unless all variables are strongly correlated). Thus, it can be appear that among three objects, one is very close to the two remaining objects, but that the others are far away from the first).

In order to better visualize the matrix of distances, Czekanowski proposed replacing numbers with simple graphic characters, e.g., white squares, filled with black rectangles, or black dots, of differentiated diameter. The lowest distances are designated by purely black squares (the largest dot), the greatest, by purely white squares (no dot). By changing the ranges of attributing characters, and the number of characters, we can easily emphasize, or blur, some tendencies. Moreover, the diagram can be assymetrical: ranges of attributing, particular characters can be different, above and below the diagonal. Czekanowski's diagram, with black and white symbols, allows us to imagine the real arrangement of objects, much better than with a dendroparallelogram.

The only problem, as we have already said, is setting the objects best in the best possible order. Taking into account the diagram of Czekanowski, we can formulate the optimal arrangement as follows: the closer diagonal, the more black; the farther the diagonal, the more white. This rule may be transgressed, only in a case where, e.g., one object is slightly similar to another object in all of the series; but this second object is already arrranged, among many more similar objects. Then, the first object should not disturb more compact groups, and must be transferred.

Until now, the only manner of ordering diagrams was manual, consisting of the gradual fitting and grouping of objects. This process was very strenuous, and it limited the length of an examined series. The largest, manually ordered diagram, to be found in the archive of Warsaw University, contains 108 objects, and surely, it took, at least, one full month (if not longer), to be counted up and ordered.

Advantages and disadvantages.

Using Czekanowski's method, one must keep in mind several things. First of all, the variables, describing objects, must be of the same sort: interval, ordinal, or nominal dychotomic (binary). In the case of variables of various sorts, some of them must be transcoded. Also, when all variables are of an interval type, it is useful to transcode them, into ordinal ones. Then, the problem with standarization disappears. But, of course, with that operation, a part of the information, about the actual distribution is lost.

The algorithm can work with an incomplete data set. But, if an object has only a few fixed variables, it will be assumed, to be close to many, very different objects, or groups of objects. Thus, the examiner must determine, before the analysis, how many variables can be maximally omitted. From our experience, it appears that the limit of 50%, should be acknowledged as the limit, not to be surpassed.

The most difficult part of the analysis is to suitably arrange objects in the diagram, in order to obtain as clear a picture, of eventual clusters, as possible. It demands a lot of experience and intuition from the examiner. Also, inferences made, based on the arranged diagram of Czekanowski, should be very careful. The illustration of real proprieties, of the discussed method, is presented below (fig. 1). The left part represents Czekanowski's diagram, for a series of Mesopotamian cylinder seals, examined from the point of view of height and width; the right part represents the histogram of these two variables. Both pictures are two-dimensional, and so, comparable.

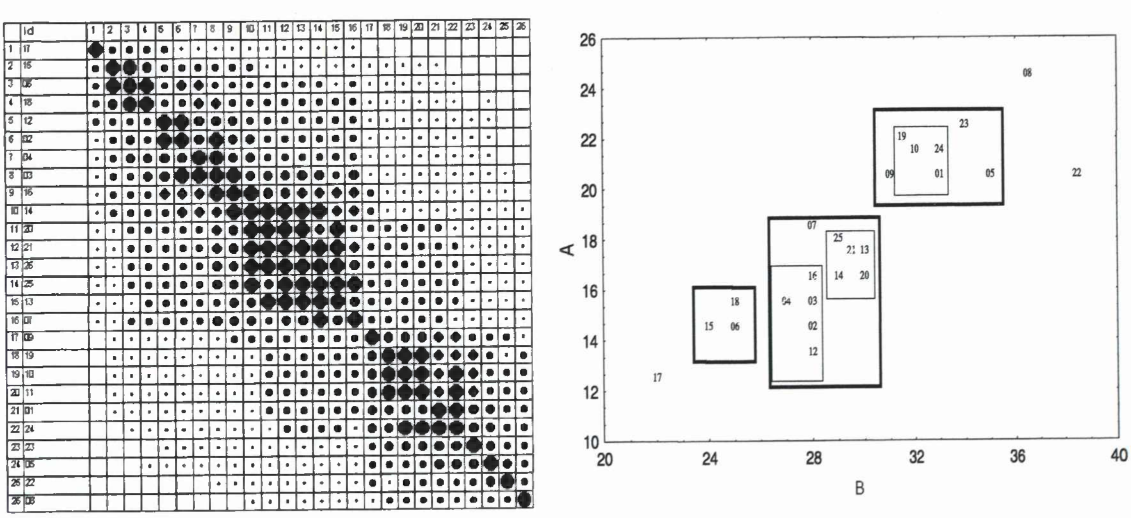


Figure 1. Czekanowski's diagram of (a) and the histogram (b), for the sample series of objects, in two-dimensional space.

The interpretation of Czekanowski's diagram is transferred to the histogram, and presented in the form of frames. Two variables, describing objects, are correlated and do not form very distinct groups. But in the diagram, we can distinguish three, fairly clear clusters: (6+15+18), ((2+3+4+12+16) + (13+14+20+21+25) + 7) and (9+(1+10+19+24)+5+23), as well as three singular objects, one, very small (17) and two, very large (8 and 22). In the second cluster, two sub-clusters appear; in the third, there is a singular concentration.

This illustration is clear and expressive, but we must keep in mind that the more variables (dimensions), the more complicated the spatial relation between objects. The number of non-correlated variables, exceeding ten – twelve, can make analysis extremely difficult, or even impossible.

As already said, Czekanowski's diagram has been used, mainly by anthropologists, for examining series of human skulls. In archaeology, it can be used for clustering pottery or flint tools, and even, for iconographical analyses. Generally, all series of objects, characterized by quantitative variables, may be taken into account. As an illustration, one can consider the graphic representation of Arabic numerals. We can distinguish three variables: the number of lines, number of curves, and number of loops. Czekanowski's diagram, arranged for a series of ten digits, looks as follows:

		1	2	3	4	5	6	7	8	9	10
1	**Digit 4**	◆	◆	◆	♦	•					
2	**Digit 7**	◆	◆	◆	♦	•					
3	**Digit 1**	◆	◆	◆	◆	♦	•	•	•	•	
4	*Digit 5*	♦	♦	◆	◆	◆	•	•	•	•	•
5	*Digit 2*	•	•	♦	◆	◆	♦	♦	♦	♦	•
6	**Digit 3**			•	•	♦	◆	♦	♦	♦	•
7	*Digit 6*			•	•	♦	♦	◆	◆	◆	◆
8	*Digit 9*			•	•	♦	♦	◆	◆	◆	◆
9	*Digit 0*			•	•	♦	♦	◆	◆	◆	◆
10	*Digit 8*				•	•	•	◆	◆	◆	◆

Figure 2. Diagram created for the classification of Arabic numerals. Distance values have been replaced with graphic characters, in the ranges (0 and 0.3), (0.7), (1.0), (1.3), (1.7 and 2.0).

The series of numerals has been divided into four clusters: (4+7+1), (2+5), (3), (6+9+0+8). The first contains the numerals, characterized only by lines; the second, by lines and curves; the third, only by curves; and the last, by curves and loops. Such a interpretation points directly, to a more general conclusion, that lines never accompany loops, and curves can occur, with both remaining elements, or solitarily. This example is very simple, but it illustrates well, that the method of Czekanowski can be helpful, in very varied problems.

Computer optimalization

Before the age of microcomputers, the process of counting and arranging a diagram was very time-consuming. Modern machines can not only accelerate the process, but they also give us new possibilities, which were inconceivable in Czekanowski's times. New formulas of distance may be used, and the ranges of numbers replaced by specified, graphic characters which can be fixed automatically. The list of possible improvements is not limited to these two examples. Much more important is that, thanks to computers, the last stage of analysis (the ordering of objects and definition of clusters), can be partially automatized.

The simplest, ordering algorithm consists of the transformation of a series, by the selection of one object, a second object, closest to the first, and a third, to the second, and so on, till the last free item. This method is also very easy to use without computers, but it fails in cases with bigger series or a more complicated distribution of objects. Much more advanced is the algorithm, based on the optimalization of cell values, in the whole matrix. It demands the definition of the optimalization function, as well as the definition of operations, to which a matrix will be subjected, in order to attain optimal distribution. In 1997, A. Soltysiak proposed such a method, using an evolutional or genetic algorithm, and the following formula for the function of optimalization:

[2] $$U_m = \frac{2}{n^2} \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \frac{(i-j)^2}{W_{ij}+1},$$

where $\mathbf{U_m}$ indicates the factor for matrix arrangement; **n**, the number of objects in a matrix; **i**, the column of the matrix; **j**, the row of the matrix, **Wij**, the value of a specified cell. The function of optimalization, defined this way, takes the lowest value, if the lowest values of the cells are closest to the diagonal (i.e., the most similar objects adjoining one another). The distribution, close to diagonal, is more emphasized then at the far corners. In contrast to the "first closest" method, in this method, the series are taken into account, simultaneously.

The function of optimalization represents the average ratio of the square distance of a cell, from the diagonal, to the value of this cell, increased by one, in order to avoid division by zero. Thus, it can be used only for the comparison of matrices, in which the number of objects and the sum of the values of all cells are identical. Practically exclusively, the variants of the same matrix can be taken into account.

In the table below, three variants of the same simple matrix are presented, with the $\mathbf{U_m}$ factor: the first one is not ordered, the second ordered by the "first closest" method, and the last one, ordered with the lowest value of U_m. The actual structure - one cluster A+D, with C, slightly similar to A, and B, slightly similar to D - is revealed properly, in the last diagram. The "first closest" method failed, even in such a simple example.

	A	B	C	D
A	0	2	1	0
B	2	0	2	1
C	1	2	0	2
D	0	1	2	0

Um=1.75

	A	D	B	C
A	0	0	2	1
D	0	0	1	2
B	2	1	0	2
C	1	2	2	0

Um=1.12

	C	A	D	B
C	0	1	2	2
A	1	0	0	2
D	2	0	0	1
B	2	2	1	0

Um=1.02

Figure 3. The value of the Um factor, for three variants of the same matrix: not ordered, ordered using the "first closest" method, and ordered with the lowest value of Um.

The proper function of optimalization solves only half of the problem. It allows us to ascertain which variant of a matrix is better ordered, but does not give an answer to the question of how we can obtain such a sequence of objects, for which the U_m factor will be as low as possible. This question relates to data of an unknown structure, with a defined input state and the function of optimalization - so it can be solved by the use of a simple evolutional (genetic) algorithm. Its input data set is formed by an object sequence, subjected to random changes of various types (e.g., shift, interchange, diversion of a fragment, etc.). The sequence, from before the change, is compared to this changed sequence and the worst results (with greater U_m values) are rejected. This process can be repeated many times. Because the changes are random, it is impossible to anticipate, after how many repetitions, the optimal sequence may be attained. Thus, one must determine after which repetition, without a decrease in U_m values the process of mutation/selection must be interrupted, with the assumption that the local optimum has been found. In search of a general optimum, this work should be repeated, with random variants in the input data set.

In the graph shown below (fig. 4), a sample process of optimalization is presented, as a curve of decreasing U_m value, in the course of more than 700 consecutive mutations/selections. The function of fitness has a logarithmic course, which is self-evident, taking into account the nature of transformation. However, because the changes are random, the deviation can be considerable, in some cases.

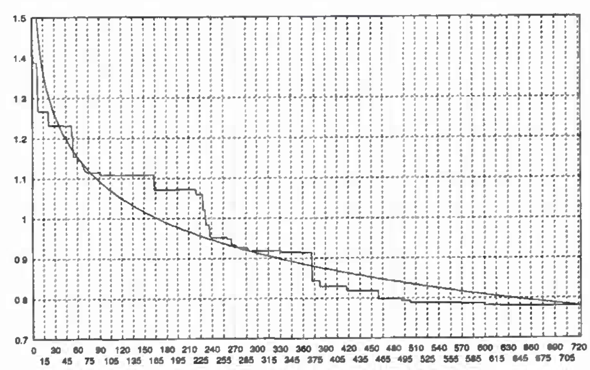
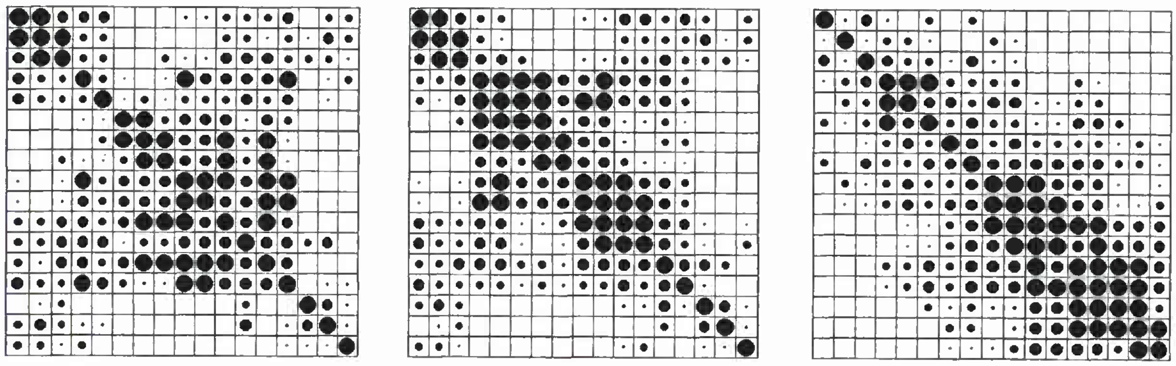


Figure 4. A graph representing the value of Um factor, in the course of the following steps for ordering a matrix, with the use of an evolutional algorithm.

The next illustration present three diagrams, arranged with the use of various algorithms. Distinctly, this one arranged by evolutional algorithm contains most compact and clear groups of objects.



Non-arranged diagram

Diagram arranged with the "first closest" algorithm

Diagram arranged with an evolutional algorithm

Figure 5. Three diagrams presenting variants of the same matrix, arranged with the use of various algorithms.

This test, and also others, carried out on various data sets, have revealed that the method presented here proves correct in practice. But, one must keep in mind that the optimal sequence may not always be consistent with our expectations. In the last diagram, in figure 5, the two first objects should be interchanged, because the second one is more similar to farther objects (except the closest to the first object); the Um value is lower, although taking into account a specific fragment of the sequence, the first object should be connected to the third. Thus, one must take into account the necessity for some small corrections, also in cases where the smallest possible value of the function of optimalization has been attained.

Software

In the Department of Historical Anthropology, at Warsaw University, Czekanowski's method has been implemented as a computer programme, called MaCzek. The first version, created in 1995, by the author of this paper, was a simple batch language interpreter, for MS-DOS. It could process up to 255 objects and variables, of all three types. Originally, only the "first closest" algorithm of arrangement was included; but, in 1997, a new utility, UmCzek was added, containing the evolutional algorithm.

At the beginning of 1998, Mr Piotr Jaskulski started to work on version 2.00 of MaCzek, transferred under the operational system, MS-Windows. It contained a lot of new tools and powers, e.g., eight optional measures of distance, with possibility for a user's own definition; two kinds of evolutional algorithms, one based on the method described above, and one simplified; as well as various options, facilitating interpretation.

First case study

The goal of the study, done by M.Karczewska and P.Jaskulski, was to check the possibility of specifying the principles of classification for pottery dated from the I to V century A.D., from the North-East region of Poland, for the use of automatic classification, performed by computer programs. This kind of analysis could be useful to archaeologists for interpretations of broken objects, and the further building of reaserch syntheses. The first step in our study was to perform a cluster analysis, because we decided to omit the results of previous research on this pottery, to check how it could be compared with statistical methods (which, besides their limitations, in our opinion, should be more objective). The present stage of our research is concentrated on distinguishing proper pottery categories, with the help of cluster analyses.

The pottery, which was used for this study, consisted of 24 vessels from 16 cremational graves. The pottery originated both from urns and from the bowls, which were used to cover the urns (15 urns, 6 bowls, 3 others). This was a small part of the whole series, which was excavated at Paprotki Kolonia 12 - the only part, which is now attainable for statistical analysis. One can ask if this material is representative of a whole series, a question which can be answered, only after finishing the field work, and describing the rest of the excavated material.

Based on the description card for the pottery, 16 diagnostic attributes were distinguished: 5 describing the structure of the pottery, 8 connected to the method of finishing interior and exterior surfaces, and 3 describing the form of the essels. This is a list of attributes, with corresponding codes:

SECTION B - SURFACE FINISHING

interior surface finishing

B1	polished =	(1) present, (0) absent
B2	burnished =	(1) present, (0) absent
B3	coarsed =	(1) present, (0) absent
B4	others =	(1) present, (0) absent

exterior surface finishing

B5	polished =	(1) present, (0) absent
B6	burnished =	(1) present, (0) absent
B7	coarsed =	(1) present, (0) absent
B8	others =	(1) present, (0) absent

SECTION C - STRUCTURE

C1	minimal wall thickness (in milimeters)
C2	maximal wall thickness (in milimeters)
C3	type of inclusion (1) sand, (2) gravel
C4	grain size of mineral inclusion (1)small, (2) average, (3) big
C5	amount of grains

SECTION E - SHAPE

E1	index A	- quotient of rim diameter to maximal diameter of the vessel
E2	index B	- quotient of high of vessel to maximal diameter of the vessel
E3	index C	- quotient of average wall thickness and average bottom thickness

Because of the great number of attributes, not all of them could be included in the analysis, at once - including too many attribute dimensions would have the effect of dividing the whole series into individual objects. That is why it was necessery to divide the analysis into parts, or to choose a few of the most distinctive attributes. We chose both methods.

The first analysis was devoted to the "B section" - connected with surface finishing. From all 8 attributes we, rejected one -B3 (coarsing of interior surface), because it didn't appear once. The B1 attribute (polishing of interior surface) was no more distinguishing, because it appeared in most of the cases (95.8%).

We used MaCzek 2.0 to draw Czekanowski's diagram, on the basis of described data. We chose distance, marked DfDD (Distance for Descriptive Data), which relied on calculating percentages of the attributes, different for any chosen pair of vessels. Analysis showed the existence of 6 different groups, consisting of 2 to 4 vessels each and several vessels, outside the group, which had intermediate position, or were quite different from all of the group. For checking if distinguished groups and zones had any archaeological significance, we studied the chronology and shape of the pottery (urn or bowl). This showed that no conclusion can be drawn from a relationship between surface finishing and the chronology of a vessel; there were vesstels from both eather and later periods in the group, so it was difficult to determine any regularity.

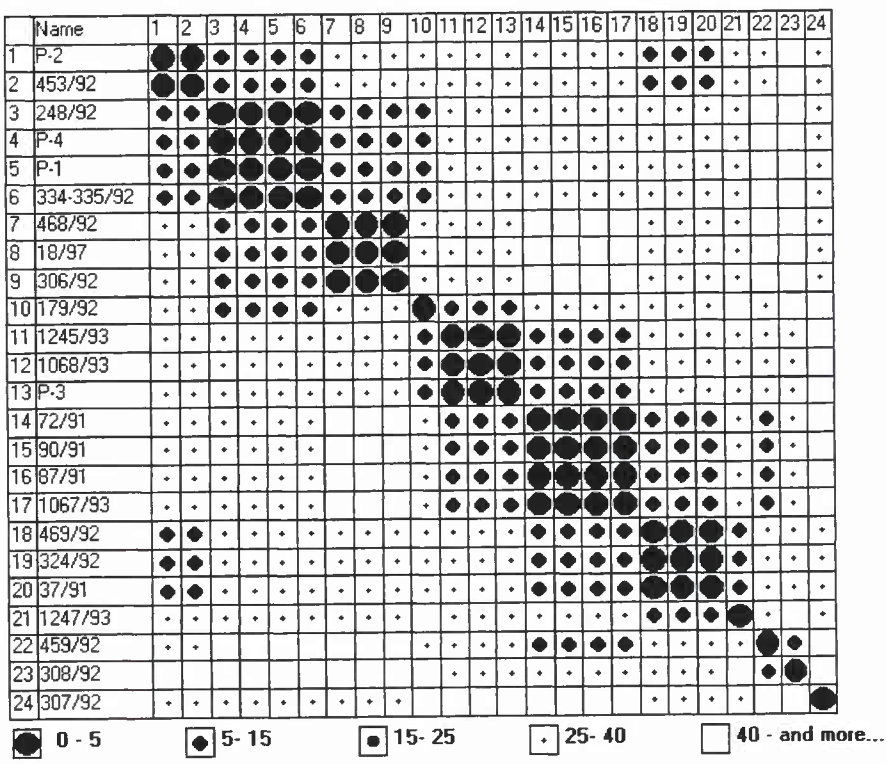


Figure 6. Diagram presenting the results of the first analysis.

The next stage was devoted to vessel structures (C section). We took four attributes: minimal & maximal wall thickness, grain size of mineral inclusion, and the amount of grains. The type of inclusion was rejected, because, for most of the vessels, it appeared to be gravel inclusion. Only in one case, was it "sand and gravel". To obtain more appropiate results, we performed an initial standardization of data.

We chose, then, to do analysis on the basis of the measurments of the vessels. In this case, only the urns were analysed. From the sixteen vessels, used in this text, two are missing, due to their bad condition. Parts of the vessels have only fragmental data, also due to destruction. Rejection of these parts would result in rapid decrease of the number of vessels, to 10; so we decided to use, instead the function of distances, used in anthropological research in Poland - Czekanowski's MCD, which allows the comparision of objects without a full set of data. Below, there is a data table (measurements in millimeters).

Number - type	WYS	SW	MWB	SD	GD	GS	W-A	W-B	W-D
gr. 2 - 1	240	245	310	156	15	7	0.79	0.77	0.47
gr. 3 - 1	122	160	174	90	12	8	0.91	0.70	0.66
gr. 5B - 1	148	174	208	80	14	6	0.83	0.71	0.43
gr. 6A - 4	178	248	248	120	10	6	1.00	0.71	0.60
gr. 8 - 1	198	244	278	140	16	8	0.87	0.71	0.50
gr. 23 - ?	-	-	-	142	10	8	-	-	0.76
gr. 27 - 1	205	273	315	130	18	10	0.86	0.65	0.55
gr. 46 - 2	250	131	260	-	13	7	0.50	0.96	0.54
gr. 52 - 1	-	-	243	-	-	6	-	-	-
gr. 58 - 3	208	110	200	100	10	8	0.55	1.04	0.80
gr. 64 - 2	309	146	334	134	11	8	0.43	0.92	0.73
gr. 66 - 2	320	170	387	-	-	10	0.43	0.82	-
gr. 72 - 2	395	190	386	160	-	-	0.49	1.02	-
gr. 102 - 2	280	110	235	100	7	6	0.46	1.19	0.86
gr. 113 - 1	195	133	195	62	8	7	0.68	1.00	0.87

SW - rim diameter, MWB - maximal diameter of vessel, SD - bottom diameter, GS - average wall thickness, GD - average bottom thickness, TYP - type of vessel (1 biconical vessel with neck and rim, 2 spherical vessels with neck and rim, 3 biconical bulging vessels, 4 tumbler vessels). The direct use of metric attributes is usually not the best method for describing the object or phenomenon. One can use instead, indices, describing proportions of the vessels, not directly connected to the actual measurements.

On the basis of the diagram, one can distinguish 2 zones. The first one, includes verses 1 to 5 and the second one, 7 to 12; verse 6 (vessel from grave 6A) has an intermediate position, between these two zones. The extreme objects, in verses 13 and 14, are connected to the second zone. It seems that indices A,B, and C are good enough attributes of distinction, when considering the shape of the vessel, and based on the diagram, one can distinguish two main vessel types:

1. Alpha type of vessel
    - index A over 0.70 (rim diameter "mid-smaller" from maximal vessel diameter)
    - index B to 0.75 (vessel wider than higher; "mid-flat")
    - index C to 0.75 (bottom thicker than wall, on the average).

2. Beta type of vessel
    - index A to 0.60 (rim diameter clearly smaller than maximal vessel diameter)
    - index B over 0.90 (vessel with comparable diameter and height)
    - index C over 0.70 (bottom thickness comparable to average wall thickness)

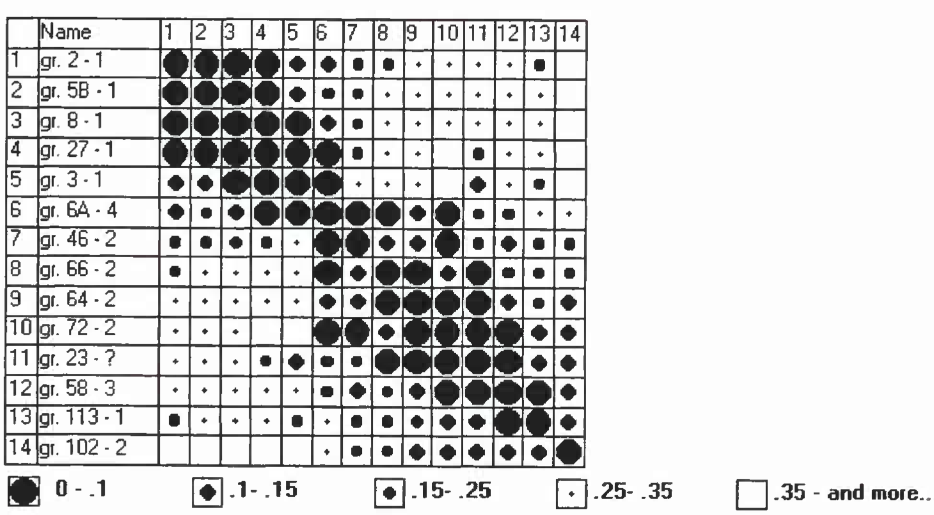


Figure 7. Diagram presenting analysis, on the basis of the measurments of vessels.

Second case study

In most cases, Czekanowski's method is used for the analysis of the objects, described by quantitative variables. But it can be helpful, also, in studies on objects, having binary characteristics (e.g., a series of pictures either containing or not containing, specific, iconographic elements). It is worthwile to present here, a sample, cluster analysis of a series, with such qualitative specifications.

The series contained twenty-eight Hittite, cuneiform tablets, enumerating the names of Anunake, the gods of the Underworld [*A. Archi, The Names of the Primeval Gods, "Orientalia", nr 59, 1990, fasc. 2, ss. 114-129*]. Exactly the first half of the series was composed of royal official letters, and the second, of miscellaneous texts, including magical and mythological ones. All together, on all the inscriptions, the names of 34 gods appeared. Thus, the specific names would be the objects, characterised by the presence/absence, at succeeding inscriptions. Of course, in the case of binary characteristics, objects can become variables and vice versa, but the clustering of the names of the gods seemed much more interesting, than the clustering of tablets, and even more so, as we disposed their deductive typology (the series of tablets was divided into two classes: royal inscriptions and those which remained).

Seven tablets were broken *in loci*, where the names of Anunake were enumerated. For that reason, all names except those which were readable received indeterminate values for seven suitable variables. The following diagram presents the results of analysis, which were very clear and easy to interpret.

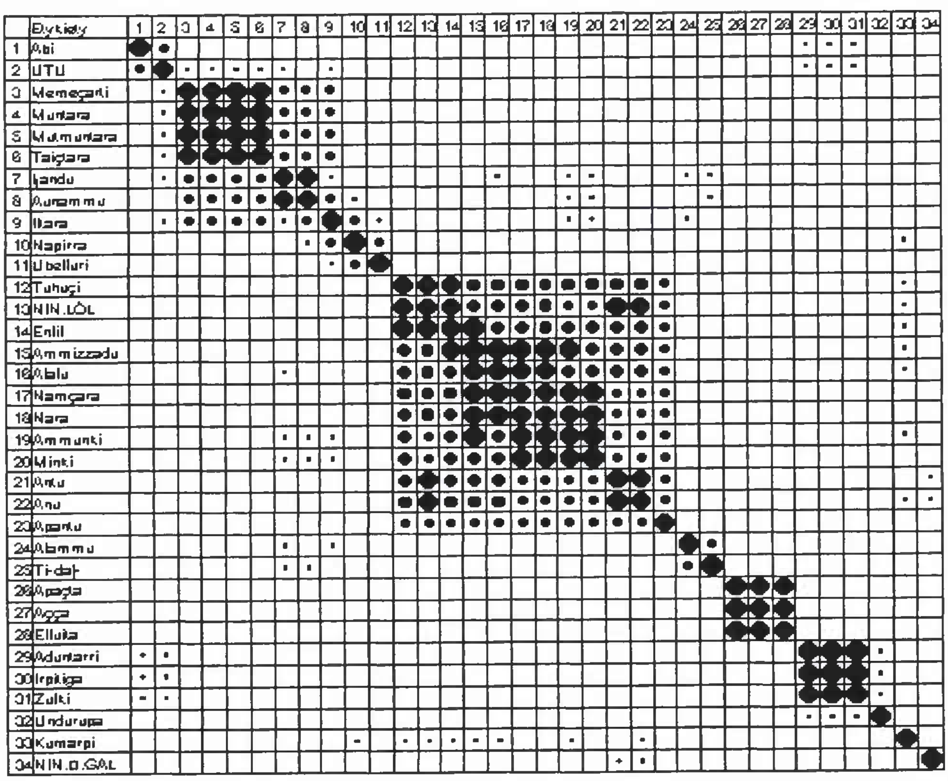


Figure 8. Diagram presenting the relation between Anunake from Hittite cuneiform texts.

he clusters visible on the diagram are very distinct, which proves that there existed some separate traditions, in which specified groups of gods were considered to be more important. We can distinguish eight such groups, and two separate names (see the table below). The largest group, V, contains the official names, used in royal inscriptions, and also in a great part of the other ones. The remaining groups represent various local traditions, usually weakly documented. Two gods are separated: Kumarpi, who played an important role in Hittite mythology, and Sumerian NIN.É.GAL who appears only in two royal inscriptions.

Abi	1000000000?0??? 100000000??00	I
UTU	1000000000?0??? 100010000??00	
Memešarti	1000000000?0??? 000110000??00	II
Muntara	1000000000?0??? 000110000??00	
Mutmuntara	1000000000?0??? 000110000??00	
Taištara	1000000000?0??? 000110000??00	
Ijandu	1000000000?0??? 001110010??00	III
Aunammu	1000000000?0??? 0011100111??00	
Iltara	1000000000?0??? 00011000 11??10	IV
Napirra	1000000000?0??? 0000000011??00	
Ubelluri	1000000000?0??? 0000000001??00	
Tuhuši	1111111111?111? 0000000011??11	V
NIN.LÍL	111111111111??? 000000000??01	
Enlil	111111111111?1? 001000000??01	
Ammizzadu	1111111111?1??? 011110000?1101	
Alalu	1111111111?11?1 011110000??100	
Namšara	1111111111?1111 111110100?1?11	
Nara	1111111111?1111 111110100?1?11	
Ammunki	1111111111?11?? 1111100111?11	
Minki	1111111110?111? 1111100111?11	
Antu	1111111110?11?? 000000000??00	
Anu	1111111110?11?1 000000000??100	
Apantu	1100111000?11?? 001000000??01	
Alammu	1000000000?0??? 000000010??10	VI
Ti-da[-	1000000000?0??? 000000010??00	
Apašta	1000000000?0??? 000000100??00	VII
Aššа	1000000000?0??? 000000100??00	
Elluita	1000000000?0??? 000000100??00	
Aduntarri	1000000000?0??? 110001000?1?00	VIII
Irpitiga	1000000000?0??? 110001000?1?00	
Zulki	1000000000?0??? 110001000?1?00	
Undurupa	1000000000?0??? 000000000?1?00	
Kumarpi	1000000111?0??? 000000001??100	IX
NIN.É.GAL I	011000000?0??? 000000000??00	X

Figure 9. The result of clustering the names of Anunake.

Further interpretation, of the results of this analysis, should be left to experts in Hittite religion and mythology. We can only point out that the method, discussed in this paper, proved correct in this case.

Concluding remarks

The diagraphical Czekanowski's method is pretty old, but with the use of modern, digital machines it appears to be very effective. Of course, it needs more study. For example, until recently, the final most important part of analysis, distinguishing clusters, was realized by an investigator, equipped only with intuition and experience. Now, the partial automatization of this stage seems to be possible.

Finally, it is enough to say that Czekanowski's method is one of the more simple and reliable tools, for clustering objects in a multidimensional space of characteristics.

References

ARCHI A. (1990), "The names of the Primeval gods", *Orientalia*, 59, fasc. 2, pp. 114-129.

WIERCINSKI A. (1988), "El análisi multivariado de asociaciones entre las deidades mexicanas y los signos de los días Tonalpohualli, y vice versa", *Archeologia, scienza e società nell'America Precolombina. Atti del convegno internazionale*, Brescia, 17-18-19 Giugno 1998, pp. 121-128.