Archeologia e Calcolatori
12, 2001, 7-45

EXPLORATORY ANALYSES OF STRUCTURED IMAGES: A TEST ON DIFFERENT CODING PROCEDURES AND ANALYSIS METHODS

1. Introduction

Archaeologists are often faced with the problem of studying large assemblages of finds bearing rather stereotyped figurative decorations. These objects – which may be pottery vessels, terracottas, stone or metal reliefs, seals, etc. – were produced by specialised craftsmen, often in large amounts. Their decoration was based on a more or less limited number of simple icons, which where combined in different ways – often involving fixed sequences of icons –, thus providing a larger variety of single images. Unlike simple decoration patterns consisting of a single icon repeated all over the available space (Fig. 1, a), such images are characterised by a hierarchical iconographical structure. In other words, the spatial relations (dimensions, closeness, order, etc.), according to which icons are associated, represent significant features of the resulting image (Fig. 1, b, c, d). In Fig. 1 example icons are simple geometric elements; however, in most cases, similarly structured images are composed by figurative elements, such as plants, animals, objects, and human beings.

Unlike major artworks, that may be profitably studied separately, since limited in number and showing a high individual variability, the study of objects of the type described above requires a different approach. In fact, in this case every specimen's full meaning may be revealed and explained only in the framework of the whole assemblage to which it belongs, so that this is the suitable context within which its individual features need to be evaluated. The main issues in the study of similar assemblages of finds may be the identification of chronological variation, of different centres of production, but also the connection of specific iconographic or compositional features either with these, or with the preferences of different categories of owners. On account of the large amounts of items involved, and of the limited (though

a b c d

Fig. 1 – Examples of different image patterns. a) simple pattern consisting of a single repeated icon; b) simple structured image (two different icons of the same size, alternating on the same line); c) simple structured image (two different icons of the same size, repeated on two superimposed lines); d) more complex structured image (repetition of a group formed by two different superimposed icons alternating with a single larger icon, of the same shape of the latter one).

sometimes significant) number of elements used – in different associations – in their decoration, a quantitative data analysis approach, based on mathematical and/or statistical tools, appears especially promising in order to obtain these results.

Given an assemblage of archaeological material of the type described above – namely a corpus of about 1000 images from Near Eastern cylinder seals of the late IV millennium B.C. – the study of its coding and of the consequent suitable data analysis tools has been the object of a long-lasting research project carried out by the authors [1]. Several different approaches were tested, considering the interest in analysing the contents of the engraved images: they ranged from the classical coding of presence/absence of iconographic elements (Camiz, Rova 1991, 1993b; Camiz 1994; Rova 1994, 1995), to a formalised text describing the image in full details, according to pre-defined criteria (Camiz, Rova 1991, 1992, 1993a, 1993b; Camiz 1994; Rova 1994, 1995; Camiz, Rova 1996), to symbolic strings describing the syntax of the image, i.e. its iconographic structure (Camiz 1994; Camiz, Tulli 1994; Rova 1996; Camiz *et al.* 1998).

The present work focuses on the analysis of the formalised texts used to describe seals images, in relation with the analysis performed on classical presence/absence data coding. Special attention will be devoted to methodological issues, and specifically to problems of both coding and calibration of the analysis methods, in order to optimise the obtained results. As a matter of fact, the main distinctive features of the analysed assemblage clearly emerged in all performed analyses, thus confirming both the general validity and the robustness of the used techniques. Nevertheless, the differences turned out not to be totally irrelevant, since different methods suited especially particular groups of images characterised by different structural features. This will be of particular interest for the study of different corpora, since it will allow to select *a priori* the most suitable techniques on the basis of the specific characteristics of the corpus, in order to quickly obtain the best and clearest results.

In order to ease the comprehension of the analyses results, the present contribution is based on a limited sample of 100 images belonging to the aforementioned corpus of seals (Rova 1994). Exploratory analyses were performed on this sample according to different coding procedures and different techniques. In this way we could identify the most important features of the data, as resulting from the different coding and the different analysis parameters, such as the particular coding, the selection of variables, the differences among quantitative vs. presence/absence analysis, etc. In the follow-

[1] Among the few applications available in literature of similar analysis methods to other assemblages, see Moscati 1986, 1995, concerning Etruscan mirrors and cinerary urns, respectively.

ing, the results of Textual Correspondence Analyses (Lebart, Salem 1988, 1994) are reported in detail and the discussion will focus on the found differences due to the different analysis techniques, say quantitative and qualitative, as well as the different results obtained by selecting each time different sets of lexical items. In addition, the results will be compared with those obtained through classical Multiple Correspondence Analysis (Lebart *et al.* 1985; Lebart *et al.* 1995; Bolasco 1999) on a presence/absence matrix of elements and attitudes.

2. Some preliminary remarks

The choice of using exploratory data analysis methods on a corpus of images decorating objects of archaeological origin involves some specific problems, which deserve being shortly discussed in advance.

The first problem concerns the relation between the analysed corpus and the original population to which it belongs. In our case, as for all archaeological material, this is unknown, and presumably not random at all, since it depends on the chances of discovery. For this reason, it was decided to rely on techniques belonging to the so-called *exploratory data analysis*.

The concept of exploratory data analysis may be understood in opposition to other mathematical and statistical methods, currently used in the scientific investigation. In quantitative frameworks, i.e. in the investigations involving mathematical techniques of any kind for the study of collected information, the investigation itself may be considered as a pathway leading from the definition of the framework where the specific study will be carried out, to the exact identification of a mathematical model, thoroughly describing the phenomenon under study. In particular, the pathway may be suitably described considering three steps (Camiz 1993, in press):

– an *exploratory phase*, where the analysis aims at searching structures and relations within the data that may suggest how to formulate further hypotheses;

– a *confirmatory phase* where the formulated hypotheses concerning structures and relations are tested and inferred to a reference population;

– a *modelling phase* where the mathematical model is built, based on the tested structures and relations, for a complete knowledge, that allows both simulation and forecasting.

It is very unlikely that archaeological data may be suitable to a treatment involving the whole above described pathway. In fact, in the confirmatory phase special attention should be devoted to the sampling, in order to allow the statistical inference of the results, but this contrasts with the non-randomness of the archaeological material. As well, it seems unreasonable to build a true mathematical model concerning the organisation or the variation of the structure of archaeological finds. Instead, no special requirements are necessary during the exploratory phase, so that exploratory data analysis has recently become the most used framework for a general investigation of collected data in any scientific environment, in particular where the construction of a mathematical model is very far from its conceivability.

In the exploratory phase, one defines the aims of the investigation and collects some data. These are then inspected, through exploratory data analysis, in order to understand their content, to reveal any existing data structure, able to give some general idea of the key-elements of the phenomenon. In particular, the scientist is interested in finding relations among the observed characters, in order to select them according to their ability to explain the phenomenon, and in classifying the considered units, in order to identify homogenous sub-populations. In order to inspect the data, rather than reading them all, a cumbersome work often impossible to carry on successfully, one may apply to some instruments able to reveal the information contained in the data in a synthesised way. The instruments should be able to reorganise the data, in order to reveal the structures that may exist, through a strong graphical synthesis, joined to several analytical tools, to be used as interpretative aids. In particular, this leads to describe which relations exist among the considered characters and which resemblance can be detected among the considered units. The study of the exploratory analyses results may be very long and complex, and several different analyses may be necessary to achieve a good comprehension of the collected data. Eventually, this knowledge allows the identification of data structures, that may lead the scholar to the formulation of hypotheses, concerning the relations among the elements composing the phenomenon. Far from being exhaustive of the complete pathway of the scientific investigation, exploratory tools proved to be sufficient for a very deep insight of analysed data, and some successful attempts are made to use them (although in a limited way) even as predictive tools in sensitive situations, such as medical investigations and diagnosis (Bacelar, Nicolau, Nicolau 1994; Benali *et al.* 1994).

The second question is related to the hierarchical structure of the images we are studying, that raises the problem of a suitable coding for the different images components and of suitable analysis techniques. Three different levels of image description deserve being considered (Fig. 2). The lowest, and simplest one, is the presence (or the frequency) of distinct icons, i.e. iconographic elements, like *human beings*, *animals*, *objects*, or *symbols*, which can appear in different attitudes, like *sitting*, *passing by*, *with open arms*, etc. (Fig. 2, a). The second level is the presence of small sets of elements, such as *a woman with open arms sitting on a bench carrying a vessel*, or *two rampant animals in front of each other with an object in-between*.

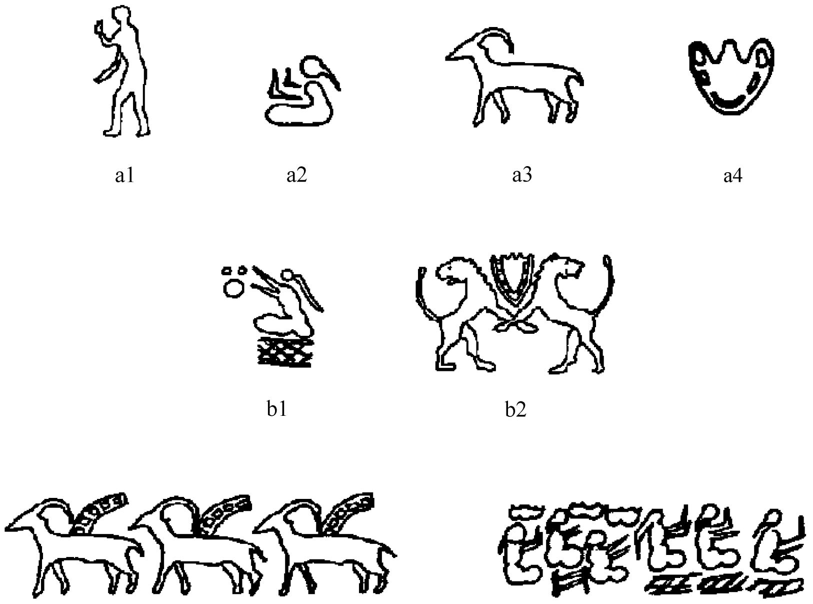


Fig. 2 – Examples taken from the analysed seal image corpus of: a) different iconographic elements: a1 = man with open arms, passing by; a2 = sitting woman, with parallel arms; a3 = passing caprid; a4 = vessel; b) fixed sub-patterns: b1 = a woman with open arms sitting on a bench carrying a vessel; b2 = two rampant lions in front of each other with an object in-between; c) overall image composition: c1 = image composed by three identical sub-patterns, each formed by two elements (Seal no. 4); c2 = image composed by two sub-patterns, each one subdivided into three smaller sub-patterns. Notice how the different elements and sub-patterns are associated to form a complex image (Seal no. 450).

These sets, which occur identical in different images (Fig. 2, b), represent small fixed sub-patterns, out of whose combination different images can be composed. The highest level is the image syntax, that is the overall composition resulting from the combination of sub-patterns, such as *image composed by two sub-patterns, the first subdivided into three* (Fig. 2, c). Although the different levels may require a separate coding and different analysis methods (in particular, for the analysis of the third level, which will not be discussed here, see ROVA 1996; CAMIZ *et al.* 1998), it is important to stress that only their integration can provide a complete picture of the studied corpus.

A third problem concerns the preliminary treatment (coding) necessary to transform the images in sets of characters to be analysed, since choices made at this stage of the work will heavily influence the results of the analysis, or, even worst, bias them. In this respect, it is clear that, for instance, the identification of iconographic elements and positions, etc., involves a certain degree of arbitrariness, which should be part of the scholar's (in this case, the archaeologist's) responsibility. Of course, the problem is not limited to the archaeological framework: every scholar, during his/her investigation, selects for his/her purposes a particular subset of a population of reference (the *sample*) and, for each sampled unit, he/she selects a particular subset of the information available, namely the *characters*, considered useful for his/her purposes. In our case, having used a textual coding, special attention must be drawn, in order not to include in the coding some uncontrolled bias, due to ambiguities in the meaning or differences in the literary styles used for the coding.

3. The images

The archaeological problems connected with the seal images have been thoroughly discussed elsewhere (Rova 1994). Suffice it to remind here that the original corpus we investigated consisted of 963 images, derived from cylinder seals and seal impressions of the so-called Uruk/Jemdet Nasr period (3300-2900 B.C. ca.), from several areas of the Near East. Some of their features made them especially suitable for our present purposes, in comparison with other sets of images. One of them was their large number, which allowed for a higher degree of reliability in the obtained results. However, even more important was the wide variety of the images. This involved both iconography and composition, and it appeared to have been obtained through relatively straightforward methods, that is the combination of a number of more or less fixed elements (human beings, animals and objects) and attitudes into some basic compositional patterns.

From the iconographic point of view, the images could be divided into groups according to their theme, such as scenes of war, manufacture and storage of goods, sacred ceremonies, complex compositions with animals, rows of animals and objects, etc. Some elements seemed specific to each theme, while some others could be shared by different themes or could occur, as sort of filling motifs, in various seals, regardless of their theme. Consider for instance, in Fig. 3, the icon represented by a *passing caprid*, which is a main element in seals no. 3 (a complex animalistic composition) and no. 4 (a simple row of animals) and a sort of "filling motif" in seal 767 (a sacred ceremony), or, in Fig. 4, the different role played by *human beings carrying objects* in seals nos. 19, 901, 750 and 930.

Composition patterns were influenced, and limited, by the same shape of the seals: these are little stone cylinders, on whose side the image was engraved. This was then impressed on a plastic surface (normally clay) by rolling the seal. In this way, a continuous frieze in the shape of an elongated band was produced. This certainly drove engravers to prefer rather simple patterns, usually consisting of rows of elements of the same height, sometimes alternating with two or three smaller superimposed elements, on more

a

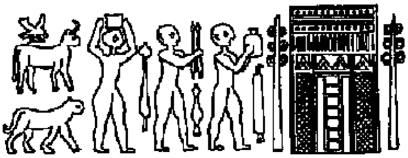
b

c

Fig. 3 – Seals of different theme and style portraying passing caprids. a) Seal no. 3; b) Seal no. 4; c) Seal no. 767.

a

b



c

d

Fig. 4 – Seals of different theme and style portraying human beings carrying objects. a) Seal no. 19; b) Seal no. 901; c) Seal no. 750; d) Seal no. 930.

complex compositions. Even within these compositions on a single row, association of different iconographic elements allowed to obtain significant variety, ranging from the simplest (the same element repeated several times, Fig. 5, a) through more complex ones (small fixed sequences of different elements repeated, sometimes with minor variation, Fig. 5, b, c), to the most complex and irregular ones (several different elements, to which secondary attributes could be added, with no repeated sequence, Fig. 5, d).

To sum up, the association of different elements and "themes" with different composition patterns produced a significant number of recognisable image groups and sub-groups. As we already pointed out, an especially

Fig. 5 – Examples of different composition patterns on single-row seal images. a) Seal no. 455; b) Seal no. 51; c) Seal no. 430; d) Seal no. 387.

remarkable feature was the occurrence, on different groups of seals, not only of the same elements and positions, but also of some small fixed sequences of elements and composition sub-patterns.

The tests we are dealing with in the present study were not performed on the whole corpus of images, but on a sample of 100 selected images (more than 10% of the total amount), in order to make their interpretation faster and more straightforward. We also wanted to obtain a rough test on the results variation as a function of the different considered data set. During the choice, incomplete or ambiguous images, as well as some *unica*, were excluded from the selection; among the remaining ones, special care was taken to ensure that all most significant seal groups be represented within the selected, so that the main features of the seal corpus were reproduced in the analysed sample.

To make the following discussion easier to understand for the non-specialised reader, the main groups of images represented in the analysed sample will be shortly described. "Naturalistic" seals are characterised by their elaborate style and varied composition. They include *war* scenes (Fig. 6, a), *sacred ceremonies* (Fig. 6, b), *storage* and *craftsman* scenes (Fig. 6, c), and *complex compositions* with *animals* (Fig. 6, d). The so-called "schematic" seals, on the contrary, show simple, repetitious patterns, quickly made with the use of mechanical tools (such as a drill). They include simple *rows* of objects – *vessels* (Fig. 6, e) etc. – and animals – like *caprids* (Fig. 6, f) or *spiders* (Fig. 6, g) –, and more complex compositions. The latter show *sitting women* – most probably craftswomen – manipulating *vessels* (Fig. 6, h) or processions of *standing women carrying standards* and other objects (Fig. 6, i). Finally, some seals – which represent human beings, animals and objects in more or less complex compositions (Fig. 6, j) – show intermediate features between these two main groups.

Fig. 6 – Examples of different groups of seal images: "naturalistic" (nos. 1-4) and "schematic" (nos. 6-9); no. 9 shows intermediate features. a) Seal no. 566 (war scene); b) Seal no. 750 (sacred ceremony); c) Seal no. 351 (storage scene); d) Seal no. 50 (complex animalistic composition); e) Seal no. 774 (row of vessels); f) Seal no. 833 (row of caprids); g) Seal no. 455 (row of spiders); h) Seal no. 796 (sitting women manipulating vessels); i) Seal no. 429 (standing women carrying objects); j) Seal no. 335 (human beings alternating with animals and objects).

4. The coding procedures and the analysis methods

4.1 *The coding*

The first stage of our work involved a simple presence/absence coding of elements (e.g. *king*, *woman craftsman*, *caprid*, *lion*, *vase type-1*, etc.) and positions (e.g. *passing with open arms, rampant with parallel paws*, etc.). It aimed at analysing the lowest image level (presence and mutual association of single icons) [2]. Some variables were also added, which marked the presence/absence of more general categories, including different similar elements or positions (like *human being*, *female human being*, *animal*, *with open arms*, etc.). This was interesting in order to plot on the representation space these super-categories, allowing an interpretation of the results on different levels of generalisation.

To proceed to the study of the second level, we applied to the possibilities offered by a textual description of the image contents, a more flexible tool than the first coding. In fact, once very strict rules are defined for the construction of a formalised text describing the image content, the information transferred in this way is sufficiently complete to understand the content of the image. In particular, special care must be devoted to constantly using the same form for the description of the same element or the same attitude. Furthermore, all terms should not be inflected according to grammatical rules, otherwise losing the 1-1 correspondence between descriptors and described objects.

Several advantages are attributed to such coding: with a good practice, it may be easier to proceed to this coding than to the previous one; instead, with particular care, the previous one may be included in, or at least automatically extracted from the latter, via a computer program. Furthermore, the text may be inspected not only for the identification of the occurrences of a single form, corresponding to an element or an attitude, but as well for either *repeated segments*, i.e. sequences of forms that appear exactly in the same way in different texts, or *nearly-segments*, i.e. sequences of forms differing from each other only for one or two forms (Bécue, Haeusler 1995). Segments and nearly-segments are very important for our descriptive task: in fact, there are objects and/or attitudes that may be described only through *polyforms*, i.e. sequences of forms having a unique meaning; in addition, the

[2] The definition of some of the iconographic elements involved a certain degree of interpretation, since they were defined both by formal features and by function inferred from the image context. A purely descriptive codification, as advocated, e.g., by Suter (1999, 49-51), in which figures and objects are identified in terms of their postures, gestures and formal features alone, would certainly be less subjective, but would result in an excessive number of elements, many of which would not be useful for the specific analysis aims (see, e.g., the type of coding proposed for seal images by Digard *et al.* 1975). The definition of positions, on the other hand, followed more strict formal criteria, so that these may be used, if necessary, as a cross-check for the iconographic elements.

association among elements and attitudes is as well described through polyforms, i.e. segments. Furthermore, image sub-patterns involving two or more elements may be described thoroughly through segments, or at least through nearly-segments, when only minor differences exist among them.

The preliminary coding of the seal images involved transforming each image into a formal text, according to a set of fixed rules [3]. Starting from the top left of the image, continuing rightwards and from the top to the bottom, each icon was described by means of a sequence of lexical forms, defining, in this order, the iconographical element, its position, the position of arms and/or paws (according to the same criteria used in the classical coding) and its orientation (*right*, or *left*). Furthermore, additional lexical forms were added to record specific attitudes (e.g. the presence of animals with turned head). Different elements were connected through relation markers (*and*, *plus*, *on*, *intertwined with, inside, above-below/alongside*), while different sub-patterns were divided through punctuation marks: *comma*, *semicolon*, *period*. A detailed description of the original coding procedure is given in Rova 1994.

4.2 *The analysis methods*

The primary tool that we used in all our analyses was *Correspondence Analysis* (*CA*: Benzécri 1982; Lebart *et al.* 1985; Lebart *et al.* 1995). Starting from a contingency table crossing two characters, the aim of *CA* is to represent the different nominal values of both characters on common factors spaces. The analysis takes into account both table's rows and columns profiles, i.e. the frequency distributions of each character's values, according to each value of the other character, and compares them, by computing the so-called *chi-square distance*, a special kind of Euclidean distance, that takes into account the *weights*, i.e. the marginal frequencies of each character's value. These distances are interpreted as weighted distances among points scattered on a representation space. The quest for orthogonal directions of maximum scattering (evaluated as the points *inertia*) leads to the *eigenanalysis* of the derived *scalar products matrix*. As a result, *eigenvectors* are obtained, identifying the searched *axes* of maximum scattering directions, considered as *factors* of the variability. Since the values of both characters may be plotted on the same space, each eigenvector corresponding *eigenvalue* represents a measure of the agreement between the views of the factors, according to either the rows or the columns profiles. In fact, their square root is considered as *canonical correlation* between these two views.

The axes are independent directions of the values scattering in the considered spaces, and the position of each value of a character in relation

[3] The textual coding was carried out in Italian. For sake of comprehension, in the examples presented in the following, the texts have been translated into English.

with the other set of values positions may be used to understand the reciprocal associations. In particular, values of the same character close to each other may have similar profile; values set in orthogonal directions in respect to the scattering centroid (the origin of the axes) should have profiles independent from each other, and values opposed on the same line passing through the origin have alternate behaviour, say they are present in opposite conditions. Considering values of different characters, their positions should be studied according to the whole set of the other character values. So, if they are close to each other, this means that their position in relation to the set of all other character values is similar and may depend on the same factors, so that there could be some kind of association among them. If they are opposite no common presence should be expected, and if they are along orthogonal directions no relation may be seen between them.

Sometimes a causal meaning may be attributed to factors, so that the diversity of the represented objects (detectable through their different position on the representation space) may be attributed to the different value of the factor itself, if measurable. This is actually one of the aims of the analysis itself, and it may be described as the *contribution* of a factor to the scattering itself and indeed one may think that a factor value may cause the presence of some particular characters' values rather than others.

When several characters should be taken into account, *Multiple Correspondence Analysis* (*MCA*) is the technique to be used, with analogous properties. In this case several characters values are represented, so that the mathematical formulation is somehow different. Nevertheless, the characters' values are plotted on graphical spaces in the same way with the same interpretation of the reciprocal positions. As an addendum, in *MCA* the representation of the units is also possible, on corresponding spaces, so that the units position is interpreted in terms of their corresponding factors values.

A special mention should be devoted to the use of *supplementary elements*: other character's values, not taken into account in the factors quest, or, in the case of *MCA*, other measures taken on units as well as other units on which the same characters are observed, may be represented on the factor spaces, in a position corresponding to the relation they have with all other elements. In this way, not all elements play the same role in the analysis, thus introducing some asymmetry that may be used both for factors interpretation through exogenous characters (i.e. the supplementary ones) and for the identification of possible causal relations (at least, in statistical terms) among the elements playing different roles.

A recent specialisation of *CA* can be used when dealing with textual material: it is the case of the study of several different written texts, ranging from the comparison of literary texts of the same or different authors, to the analysis of *open-ended questions* in the surveys. In this case, interest is drawn to the distribution of the *lexical forms* in the different texts. The *lexical table* is the contingency table crossing lexical forms with texts, so that each entry in the table is the frequency (or the simple presence) of a particular form in a particular text. *Textual Correspondence Analysis* (*TCA*: LEBART, SALEM 1988, 1994) is thus the *CA* of such lexical table. As a result, those texts that have similar profiles, i.e. such that the forms have similar frequencies in both, should result close to each other on graphical spaces, and texts where the terms used are totally different should result far away. Analogously, proximity among forms should mean that they are present in the same texts with nearly the same profiles, etc.

Textual Correspondence Analysis (*TCA*) is the specific tool for the *CA* of textual material. Its use is analogous to the use of *CA*, with the advantage that it makes it possible to deal either with the whole texts or with a selection of forms, segments, and/or nearly-segments. This flexibility allows to tailor the analysis to the scholar's specific needs. Both quantitative and presence/absence analyses are possible, taking into account the frequency of a form in a text or its simple presence, respectively, thus adding further possibilities to the investigation.

4.3 *The analyses*

Classical *MCA* was performed, with satisfactory results, on the contingency tables derived from the first coding (see CAMIZ, ROVA 1991, 1992; and, in more detail, ROVA 1994, 1996). In particular, some exogenous characters – such as the geographical origin of the seal – were represented as supplementary, in order to check to what extent they could be considered to influence the presence of elements or attitudes in the images.

The texts corresponding to the corpus of 963 seal images were then submitted to a first *TCA* (for its results and their implications for the archaeological problem represented by the Uruk/Jemdet Nasr seals, see CAMIZ, ROVA 1992, 1993a, 1993b; ROVA 1994, 1995).

We soon became aware, however, that the effectiveness of textual analysis (especially as far as repeated segments were concerned) was reduced by a number of problems, connected with the chosen coding rules, the analysis procedure and major software limitations. For this reason, these were repeatedly altered in some details, in order to obtain more satisfactory results (see e.g. CAMIZ, ROVA 1996).

As for the changes in coding rules, in Fig. 7 three different codings are reported for three seals. Let us consider, for instance, the position of the lexical forms defining the orientation: originally located between the element and its position (*woman craftsman sitting left with open arms*), they were transferred to the end of the sequence (*woman craftsman sitting with open arms left*) in order not to interrupt the continuity between the element and its position. It may also be noted that some polyforms were progressively tied through underscores, in order to transform them into forms. In this way, some lexical forms, that had very limited or confusing meaning or interest when left alone, were correctly tied to the main element. It is the case of the forms *type_1, type_2,* etc., used to distinguish among different types of *vessels*, *standards*, etc., that otherwise could be confusing. Other polyforms were abbreviated, only due to program results readability.

On the other hand, some polyforms were joined to a single longer lexical forms (e.g. *woman craftsman sitting with open arms left* became *woman_craftsman sitting with_open_arms left*, and *woman_craftsman sitting_with_open_arms left*): in this way they could be considered by the program as lexical forms and included in the analysis as such. This also allowed to obtain longer and more significant segments.

In the present contribution we discuss the new analyses performed on the said sample of 100 selected images. All computations have been performed through both SPAD.N (Lebart *et al.* 1991) and SPAD.T (Lebart *et al.* 1994) computer programs, specially modified to suit our purposes. The textual analyses were done on both simple presence/absence of lexical form and segments and on their frequency, and considering as active variables forms having different minimum frequencies. It thus turned out that different groups of images, characterised by different features (number of present elements, relation of these with each other, simple versus complex syntax, etc.), reacted differently to these variations, each of them being particularly enlightened by a specific coding and analysis procedure. For this reason, these will be discussed in detail in the following paragraph.

5. Discussion of the results

5.1 *Multiple Correspondence Analysis*

The results of classical *MCA* on the 100 selected images are very similar to those obtained on the whole image corpus, described in detail in Rova 1994 [4]. The first four axes are accounted for over 31% of total variation. In short, on the first axis *animalistic* compositions are separated from the various types of images representing *human beings*. Within the latter, the second axis distinguishes *complex naturalistic* (especially *war* and *sacred*) scenes from *schematic* representations; the third distinguishes *sacred* from *profane* scenes, and the fourth separates those (mainly schematic seals) in which *female be-*

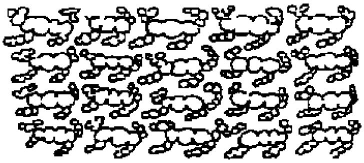
[4] The relative positions of the characters values on the first four axes are actually nearly identical, while some differences emerge only starting from the fifth axis. It is worth to emphasise that this agreement represents a good test of the general validity of this analysis.

Seal no. 50

Original coding
vessel type_1, above-below/alongside caprid rampant left parallel paws with_turned_head intertwined_with caprid rampant right parallel paws with_turned_head; lion crouching left with_turned_body on bird passing left on lion crouching right with_turned_body. vessel type_1, above-below/alongside caprid rampant left parallel paws with_turned_head intertwined_with caprid rampant right parallel paws with_turned_head; lion crouching left with_turned_body on bird passing left on lion crouching right with_turned_body.
Coding no. 1
vessel type_1, above-below/alongside caprid rampant parallel_paws turned_head left intertwined caprid rampant parallel_paws turned_head right; lion crouching turned_body left on bird passing left on lion crouching turned_body right. vessel type_1, above-below/alongside caprid rampant parallel_paws turned_head left intertwined caprid rampant parallel_paws turned_head right; lion crouching turned_body left on bird passing left on lion crouching turned_body right.
Coding no. 2
vessel_1, above-below/alongside caprid rampant_parallel_p turned_head left intertwined caprid rampant_parallel_paws turned_head right; lion crouching turned_body left on bird passing left on lion crouching turned_body right. vessel_1, above-below/alongside caprid rampant_parallel_p turned_head left intertwined caprid rampant_parallel_p turned_head right; lion crouching turned_body left on bird passing left on lion crouching turned_body right.

Seal no. 51

Original coding
vessel type_3 on vessel type_3 upside_down, woman craftsman sitting left parallel arms on bench; vessel type_3 on vessel type_3 upside_down, woman craftsman sitting left parallel arms on bench; vessel type_3 on vessel type_3 upside_down; man craftsman sitting left parallel arms on bench;
Coding no. 1
vessel type_3 on vessel type_3 upside_down, woman craftsman sitting parallel_arms left on bench; vessel type_3 on vessel type_3 upside_down, woman craftsman sitting parallel_arms left on bench; vessel type_3 on vessel type_3 upside_down; man craftsman sitting parallel_arms left on bench.
Coding no. 2
vessel_3 on vessel_3 upside_down, w_craftsman sitting_parallel_a left on bench; vessel_3 on vessel_3 upside_down, w_craftsman sitting_parallel_a left on bench; vessel_3 on vessel_3 upside_down; m_craftsman sitting_parallel_a left on bench.



Seal no. 788

Original coding
undefinable animal passing left and undefinable animal passing left and undefinable animal passing left and undefinable animal passing left and undefinable animal passing left and undefinable animal passing left, on , undefinable animal passing left and undefinable animal passing left and undefinable animal passing left and undefinable animal passing left and undefinable animal passing left, on undefinable animal passing left and undefinable animal passing left and undefinable animal passing left and undefinable animal passing left and undefinable animal passing left.
Coding no 1, coding no 2
undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left, on undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left, on undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left and undefinable_animal passing left and undefinable_ animal passing left.

Fig. 7 – Three seal images (nos. 50, 51, 778) with corresponding descriptive texts (original coding; codings nos. 1, 2 as used for the here discussed analyses).

ings are especially frequent, from the remaining ones. On the same axis, a clear distinction appears between *schematic* seals representing *spiders*, and *complex animalistic* compositions, as well.

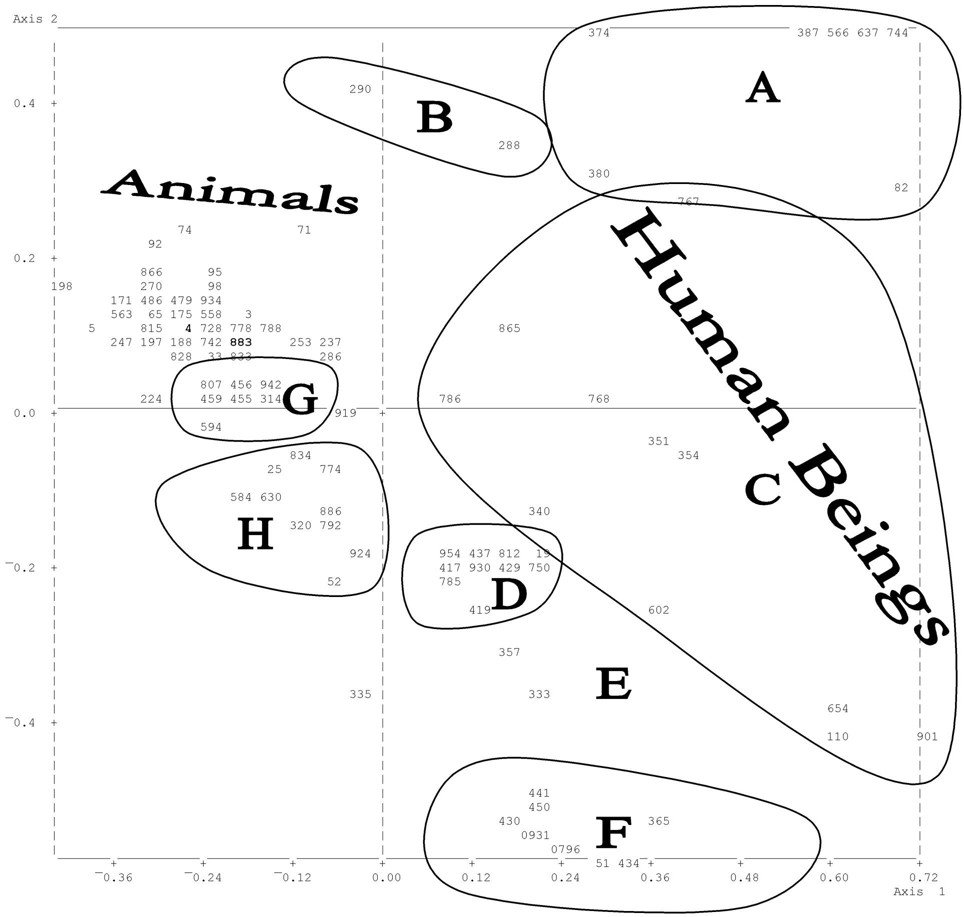


Fig. 8 – Representation of the images on the plane spanned by *MCA* axes 1 and 2. Seals mentioned in the text are emphasised. A: War scenes; B: Hunting scenes; C: Naturalistic sacred scenes; D: Schematic sacred scenes (processions of women carrying standards); E: Naturalistic storage and craftsman scenes; F: Schematic craftsman scenes (sitting women manipulating vessels); G: Spiders; H: Vessels.

In Fig. 8, the images are projected on the plane spanned by axes 1 and 2. The main types of seals described above are clearly distinguished. It can be noticed, however, that some groups (*naturalistic sacred*, *war* and *hunting* scenes, and different types of *schematic* seals) stand out better than others. In particular, *animalistic* seals appear to be rather undistinguished, with the significant exception of *schematic* seals representing *rows* of *spiders*. In general, thematic distinctions are emphasised, each "theme" being represented by a number of distinctive icons which are often associated to each other.

Images containing a large number of different, distinctive elements and positions, even if in very irregular composition patterns, as normally *complex naturalistic* seals are, stand out most clearly. Therefore, for them, and for comparable images, this kind of analysis appears fully adequate. It is not totally satisfactory, however, for other seals, like for instance those representing simple *rows* of *caprids* (e.g. no. 833, Fig. 6, f) or alternating animals and objects (e.g. no. 4, Fig. 3, b), which are constantly located near the axes origin. These images actually show opposite features in comparison with the previous ones: presence of few, or of only one, repeated elements – especially if these tend to appear in different groups of seals – and very standardised syntactical structure. Clearly, a simple presence/absence analysis of elements does not represent an optimal tool for their distinction.

5.2 *Textual Correspondence Analyses*

As we said, the use of *TCA* allowed a more flexible approach to the images sample. Different analyses were performed on both selected lexical forms and repeated segments, in order to enlighten specific aspects of the various groups of seals. In particular we searched for connections concerning both the presence and association of different icons (first level) and with the presence of fixed sub-patterns or sequences of icons (second level).

5.2.1 Coding no. 1

The first version of the textual description of the 100 seal images was specifically aimed at exploring the first (lowest) level of image description, as a possible substitute for the classical *MCA* discussed above. Care was taken for every element, body, and arms/paws position to correspond to one specific lexical form. Some lexical forms corresponding to the more general categories used in the first coding were also added, but this was not systematically done, to avoid texts to become too long. Therefore a subjective choice was made between what was considered more interesting (for instance, *man* and *woman*, *hybrid*) and what could be omitted (for instance, *human being*, *animal*, *object*). This involved a lower degree of control by the researcher on the interplay between these different types of variables. For this reason, the results of these analyses are not totally comparable to those of classical *MCA* discussed above. If textual coding should be used, in future studies, as a substitute of the classical one, special care should be devoted to these specific aspects during the coding procedure. Some examples of coding no. 1 texts are given in Fig. 7.

On coding no. 1 texts the following analyses were performed: a) qualitative (presence/absence) analysis on all lexical forms, and on lexical forms excluding those corresponding to syntactical connective particles (*and*, *plus*, *on*, *intertwined with*, etc.) and element orientation; b) quantitative (frequency) analysis on all lexical forms, on lexical forms excluding those corresponding to syntactical connective particles and element orientation, and on the latter only. In all cases, lexical forms showing a minimum frequency of 2 (in two different images) were considered as active variables.

A) Qualitative (presence/absence) analysis on lexical forms

As expected, the results of presence/absence analysis on lexical forms replicate to a large extent those obtained through *MCA*. Here, however, the first four axes are accounted for only over 23% of total variation. There are as well some significant differences in emphasis on various image features. The difference between *complex naturalistic* and *schematic* seals – represented by the first axis – is here in the foreground in comparison with the difference between scenes with *human beings* and *animalistic* compositions, which emerges only with the second axis. The third and fourth axes distinguish different themes within the seals with *human beings* and *animals* respectively, again with a special emphasis on the difference between *naturalistic* and *schematic* seals. Fig. 9 shows the projection of the images on the first factor plane. Comparison with Fig. 8 enlightens the clear-cut discrimination operated by *TCA* between *naturalistic* and *schematic* seals, and the better differentiation of animalistic compositions. The latter is better developed by the fourth axis, and involves distinction between simple rows of animals, rows of animals alternating with objects (*ladders*, *vessels*), and compositions with *hybrids* and special animals (*snakes*, *lions* and *birds of prey*) in elaborate decorative schemes (complex intertwinings, symmetric patterns, etc.). While lexical forms defining element orientation were constantly gravitating around the origin of the axes, those defining syntactical joins appeared to be significantly correlated with the different seals groups.

The presence/absence analysis was repeated, excluding all lexical forms pertaining to orientation and syntax from the active variables, in order to test the influence of the latter on the analysis results. The new results, however, showed only minor deviations from the previous ones. As a consequence, the better distinction operated on *animalistic* and *schematic* seals by *TCA* in comparison with *MCA* is only partially due to their inclusion. These differences are to be attributed, for the remaining part at least, to two different factors. The first is the inclusion in the textual coding of lexical forms defining specific attitudes, which especially affect animalistic images. The second is the different role played in the two codings by the variables defining single elements and positions versus those defining more general categories. The absence, in the textual coding, of many of these, contributed to better enlighten groups of images (as in our case both *schematic* and *animalistic* seals) characterised by a low number of specific icons. Their systematic presence, on the contrary, contributed to the clearer distinction of different "themes" (*animal* versus *human* images, *war* versus *sacred* and *craftsmanship* scenes) encompassing both *schematic* and *naturalistic* seals operated by *MCA*. Only for groups of images characterised by a large number of distinctive icons, however, it did not result in a contemporary loss of specificity for each single group.

B) Quantitative (frequency) analysis on lexical forms

The first four axes of this analysis are accounted for over 24% of total variation. As it could be expected, the use of quantitative analysis results in a higher emphasis on repetitious images – that is images which contain more than an example of the same icons. This clearly emerges by comparing the projection of the images on

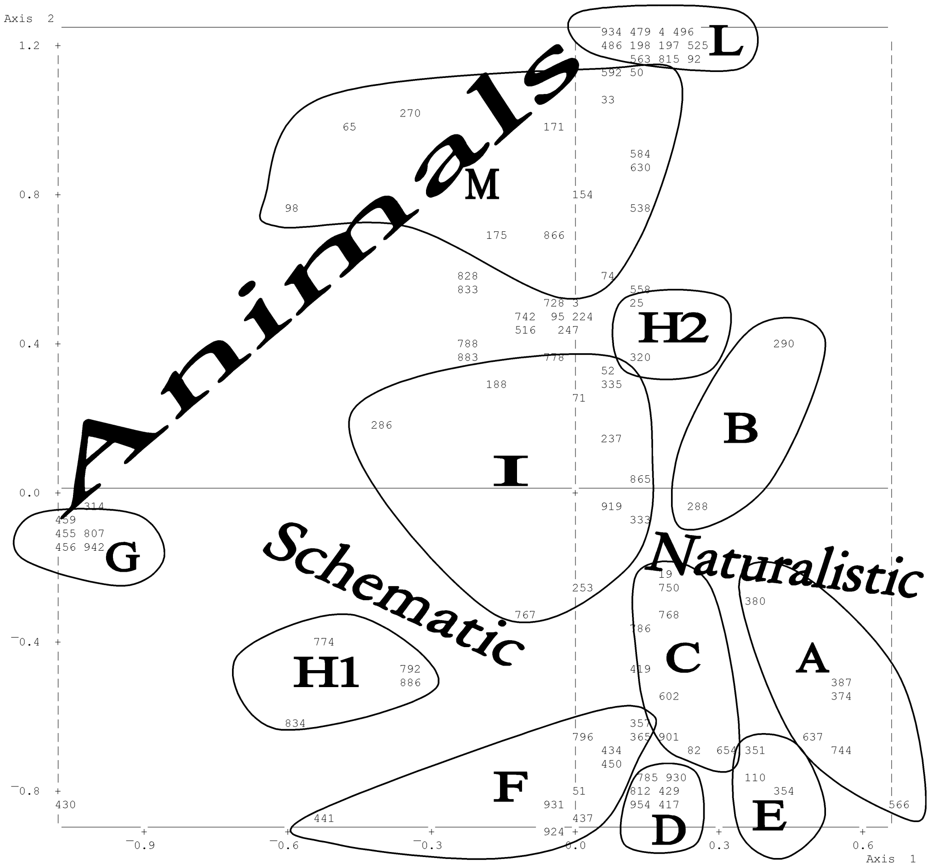


Fig. 9 – Qualitative (presence/absence) *TCA* on coding no. 1. Representation of the images on the plane spanned by axes 1 and 2. A: War scenes; B: Hunting scenes; C: Naturalistic sacred scenes; D: Schematic sacred scenes (processions of women carrying standards); E: Naturalistic storage and craftsman scenes; F: Schematic craftsman scenes (sitting women manipulating vessels); G: Spiders; H1: Schematic vessels; H2: Non schematic vessels; I: Scenes with humans, animals and objects; L: Animals alternating with "ladders"; M: Hybrids and "special animals".

the plane spanned by the first two axes (Fig. 10) with that resulting from the presence/absence analysis (Fig. 9). In both cases, on the first factor *schematic* seals are opposed to *naturalistic* ones, on the second axis *animalistic* are distinguished from *human* images, and lexical forms and images having the highest co-ordinates are nearly the same: *rows* of *spiders* opposed to *war* scenes on the first axis, *rows* of *vessels* and *women manipulating vessels* opposed to *hybrids, snakes, birds-of-prey*, and *rampant* animals on the second one. In the quantitative analysis, however, most of the remaining images are less well distinguished, since they are somehow compressed in a band around the origin. The second axis is better balanced than the first

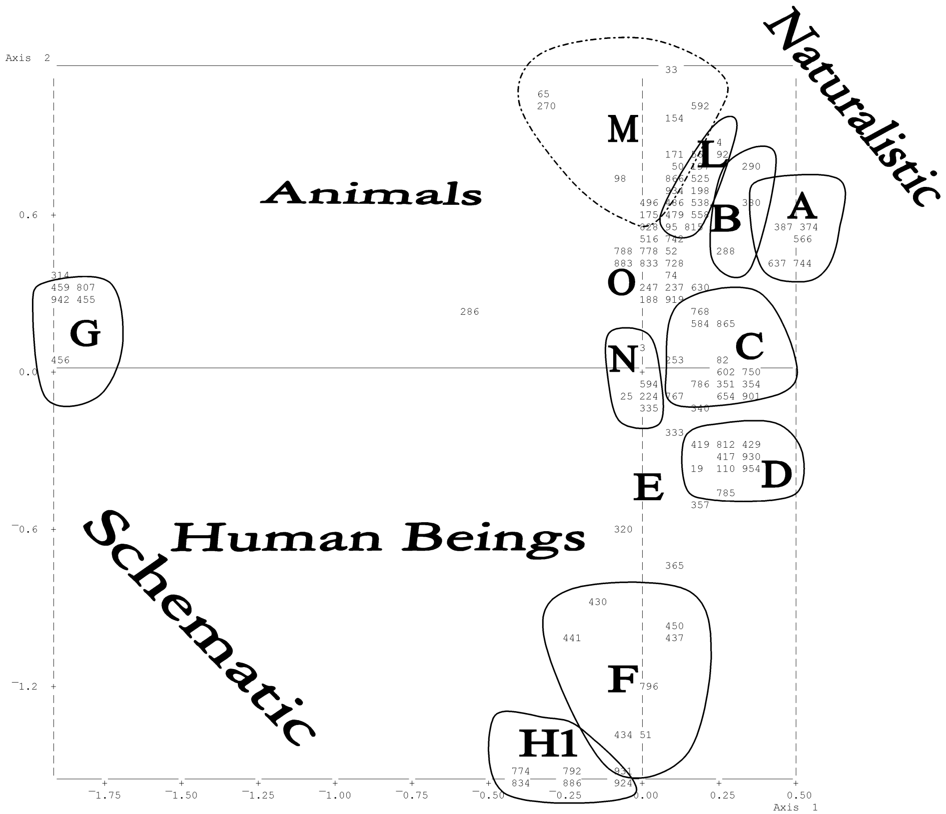


Fig. 10 – Quantitative (frequency) *TCA* on coding no. 1. Representation of the images on the plane spanned by axes 1 and 2. A: War scenes; B: Hunting scenes; C: Naturalistic sacred scenes; D: Schematic sacred scenes (processions of women carrying standards); E: Naturalistic storage and craftsman scenes; F: Schematic craftsman scenes (sitting women manipulating vessels); G: Spiders; H1: Schematic vessels; L: Animals alternating with "ladders"; M: Hybrids and "special animals"; N: Animals alternating with vessels; O: Rows of animals.

one, and clearly identifies various groups of images, but it should be noticed that the distinction between *naturalistic* and *schematic* images heavily interferes with that between animal and human scenes. Differences between the two analyses increase if the following axes are considered: in the quantitative analysis, the third axis distinguishes among *naturalistic* images the *human* from the *animal* compositions, the fourth – within the seals with human beings – distinguishes *schematic* from *naturalistic* scenes.

To sum up, the first four axes provide in both cases a reliable and rather complete picture of the analysed images, but qualitative analysis gives a better balanced synthetic picture when only the first two axes are considered, whereas in quantitative analysis a few groups of images are excessively enlightened to the detriment of the remaining ones. For future studies, quantitative analysis should therefore be preferred for corpora exclusively formed by very repetitive images made up of very few recurring elements. It may be used as well for composite corpora like the present one, provided that the presence/absence analysis is associated, to better enlighten different image groups.

As in the case of the presence/absence analysis, a second quantitative analysis was performed after excluding lexical forms pertaining to orientation and syntax. Here, too, results of the two analyses were in general similar, although in this case the syntactical forms played a greater role [5], as it was to be expected, since the occurrence of the same syntactical link connecting different elements is a typical feature of the most repetitive groups of images.

To further test the discriminating power of syntactical forms on the studied images, a third quantitative analysis was performed, in which these acted as the only active variables. It should be stressed that the aim of this analysis was not, as for the previous ones, to get a synthetic picture of the whole studied sample. Accordingly, albeit the first two axes summarize over 50% of total variation, the projection of images on these axes (Fig. 11) does not reflect the whole range of their variability and complexity. Nevertheless, some of the resulting clusters of images do correspond well to specific groups of seals as emerging from both visual inspection and from all preceding analyses: this is the case of the simple *rows* of *spiders*, *vessels* and other *animals*, on the left side of the graph, of the more *complex animalistic compositions*, in the upper part, and of *schematic* images with *women manipulating vessels*, in the lower part, to the right.

What is more important, this analysis can represent a preliminary step to reach the third level of analysis (that is that of overall image composition). This is actually much too complex to be reduced to the simple frequency of different syntactical links and needs, as we already said, specific tools to be properly analysed. Nevertheless, the results of the present analysis offer a rough, but interesting approximation to the composition structure of the studied images sample. The first axis sets the simplest paratactic structures, characterised by the exclusive presence of the syntactical link *and* (*rows* of *objects* and *animals*, mostly but not exclusively belonging to *schematic* seals, e.g. Fig. 6, e-g) apart from all the remaining images (see Fig. 11). On the second axis, on the negative side we find *schematic* seals characterised by small repeated sequences of two or more superimposed elements or elements connected with each other, characterised by the syntactical links *on* and, to a lesser extent, *plus* (e.g. *rows* of *sitting women manipulating vessels* – Figs. 5, b-c; 6, h –, *animals* connected by *ladders* – Fig. 3, b –, and processions of *standing women carrying standards* – Figs. 4, d; 6, i). On the positive side there are, on the contrary, highly decorative animalistic compositions, associated to the syntactical link *intertwined with* (e.g. Fig. 6, d). *Naturalistic* scenes featuring *human beings* (*sacred* ceremonies, *war* scenes) show more complex composition patterns, which appear to be characterised by the co-presence of different syntactical links (though with a certain preference for *plus*, that is for the presence of numerous attributes to the main image elements, Figs. 3, c; 4, b-c; 5, d; 6, a-b), and therefore cluster around the origin on both axes 1 and 2.

[5] Orientation was, on the contrary, unimportant in all cases.

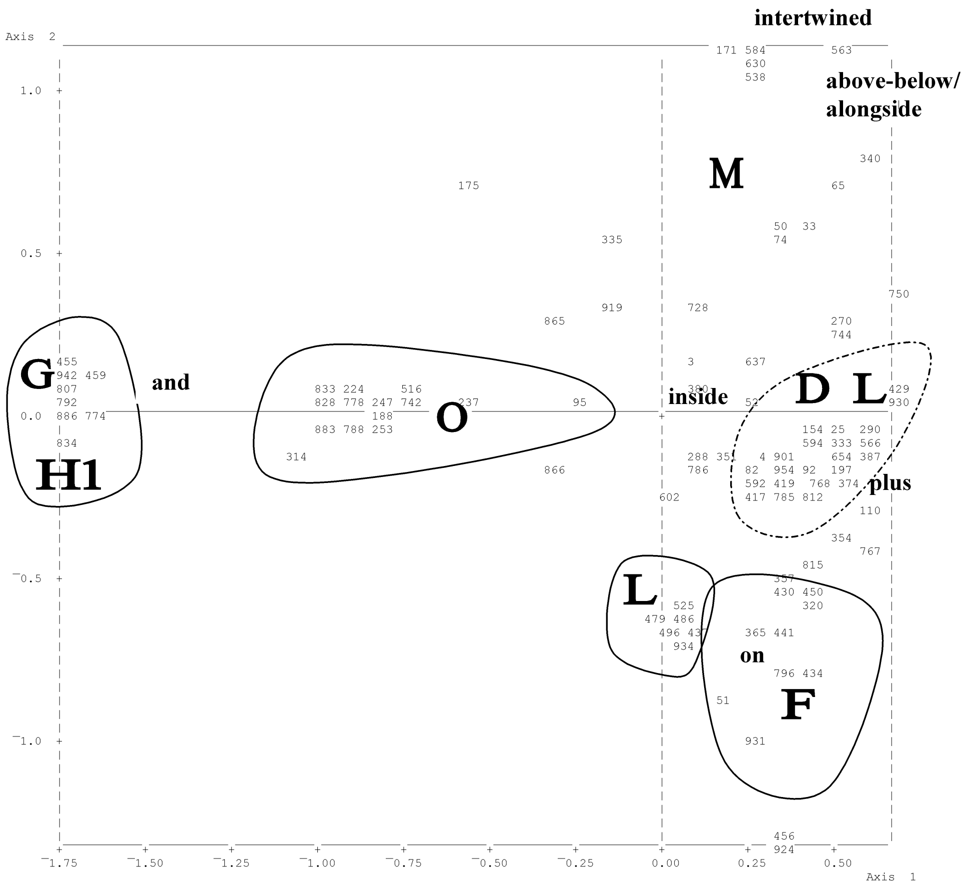


Fig. 11 – Quantitative (frequency) *TCA* on coding no. 1 (syntactical links as only active lexical forms). Representation of the syntactical links and of the images on the plane spanned by axes 1 and 2. D: Schematic sacred scenes (processions of women carrying standards); F: Schematic craftsman scenes (sitting women manipulating vessels); G: Spiders; H1: Schematic vessels; L: Animals alternating with "ladders"; M: Hybrids and "special animals"; O: Rows of animals.

5.2.2 Coding no. 2

The second version of the textual description (coding no. 2) was specifically developed from the first one for the analysis of repeated segments. In order to get a larger number of fixed sequences of lexical forms spanning more than a single iconographic element, it was decided to join some lexical forms through an underscore, so that a single form could identify the element completely (i.e. *woman craftsman* became *woman_craftsman*, etc.). As a result, in the second coding every icon is described by a maximum of three lexical forms, describing the element, its attitude (including both body and arms/paws positions), and its orientation. No lexical forms describing general element or attitude categories (like *woman*, *sitting*, etc.) are present any more. Even if this implied a certain loss of information, on the whole this coding is more objective and consistent than the previous one. Comparison between the analyses performed on both codings can therefore also act as a good test of how much results may vary as a consequence of slight inconsistencies in the coding procedure. Some examples of coding no. 2 texts can be found in Fig. 7.

The 100 texts derived from coding no. 2 were submitted, under the same conditions, to the same set of analyses as those derived from the first coding. In addition, on the second coding we ran both qualitative and quantitative analyses of selected repeated segments and those based on pooled lexical forms and selected segments.

A) Qualitative (presence/absence) analysis on lexical forms

The general pattern of the studied images (groups of images and their mutual relations) resulting from this analysis is similar to that obtained from qualitative analysis performed on the texts derived from the first coding. This clearly emerges from comparison between Fig. 9 and Fig. 12, which show the projection of images on the first planes arising from the two analyses. The variation explained by the first factor plane is a little reduced here (11.5%), in comparison with the previous one (13.4%). In both cases, the same sequence can be followed, from *schematic* seals featuring *rows* of *spiders*, of *vessels*, of *sitting women manipulating vessels*, of *standing women holding standards*, through less regular *craftsmanship* scenes and complex *naturalistic*, *sacred*, and *war* scenes, to *hunting* scenes, to various *animals* scenes, and finally to special *animalistic* compositions with *hybrids, snakes,* etc. It is thus possible to state that in this – and therefore presumably in all similar cases – differences in coding procedure only slightly alter the analysis results, and in any case never to such an extent that they can obscure the main factors responsible of the studied corpus variability and the internal clustering of its members.

Parenthetically, the same analysis performed, using coding no 2, on the whole seal images corpus yielded very similar results, too [6]. This proves that the resulting factors are sufficiently robust, since they are not altered by the selection of a corpus sample, done with the only aim to ensure the presence of seals belonging to all main iconographic groups, without attention either to their proportion or to the specific contained iconographic elements.

On the other hand, the small differences detected by comparison between the various analyses can be used to enhance our knowledge of various distinct features of the images corpus. In our specific case, the analysis on the second coding gives a better distinction of the different seal groups, since it emphasizes the specific traits of each of them, whereas the analysis on the first coding rather underlines their common features. This does not involve a major loss of information in the case of very distinctive seal types (e.g. *rows* of *spiders* for the *schematic* seals, or *war* scenes for the *naturalistic* ones), which stand out clearly enough in both analyses. For less char-

[6] This is only mentioned here, since the analysis of the whole corpus of the Uruk/ Jemdet Nasr seals will be the object of a different contribution by the same authors.

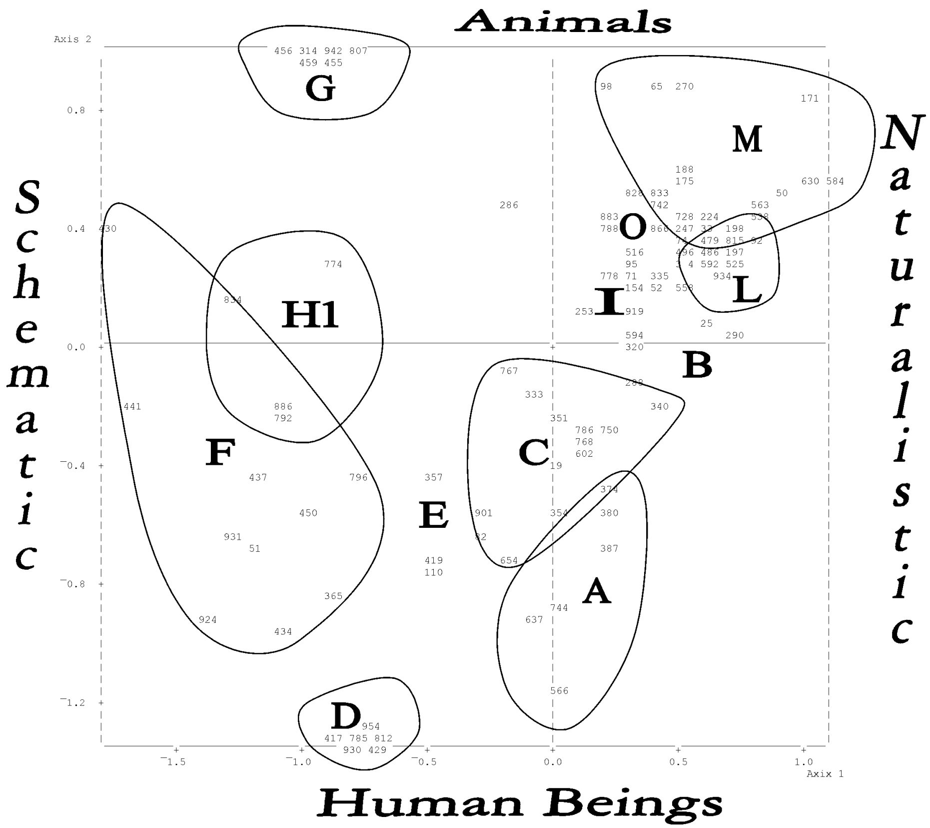


Fig. 12 – Qualitative (presence/absence) *TCA* on coding no. 2. Representation of the images on the plane spanned by axes 1 and 2. A: War scenes; B: Hunting scenes; C: Naturalistic sacred scenes; D: Schematic sacred scenes (processions of women carrying standards); E: Naturalistic storage and craftsman scenes; F: Schematic craftsman scenes (sitting women manipulating vessels); G: Spiders; H1: Schematic vessels; I: Scenes with humans, animals and objects; L: Animals alternating with "ladders"; M: Hybrids and "special animals"; O: Rows of animals.

acterised images, however, like in our case for the different groups of *animalistic* seals and for *sacred* scenes, the second coding clearly appears more effective in order to obtain a significant picture (see the distribution of all these images groups in Figs. 9 and 12). From a different point of view, the analysis performed on the first coding better enlightens the presence in the seal imagery of wide thematic distinctions (e.g. between human and animalistic images) encompassing various groups of seals, but is less able to distinguish among the latter, and vice versa.

Once again, how this happens can be seen by comparing the position of images and variables on the axes obtained from the two analysis. In spite of their general similarity – in both cases the first axis distinguishes *schematic* from *naturalistic* seals, the second *human* from *animalistic* seals –, these show a clear shift in the relative importance of the same few elements (special, very distinctive seals and icons). In the coding no. 1 analysis, for instance, the first factor emphasises, at both axis extremes, two special groups of images (*rows* of *spiders* and *war* scenes), while all other images (especially *animal* compositions) occupy more indistinct positions. In the coding no. 2 analysis, on the contrary, images are distributed along a continuum pattern, where as well it is possible to distinguish the many different groups of the corpus (from *schematic* seals, through *naturalistic human* scenes, to different types of *animal* scenes). These same aspects, however, emerge in the first analysis on the second axis, so that the global results can be considered almost equivalent for the researcher. As in the previous cases, differences between the two analyses tend to increase if the following factors are examined, but this is not so important, since, considering our present aims, in all cases the first two factors provide a clear enough picture of the relations among the studied objects.

Like in the case of the first coding, results of the presence/absence analysis performed after excluding all forms defining syntax and orientation were nearly identical to those obtained for all the lexical forms. The only differences, minor changes in the position of single images and groups thereof, do not deserve further discussion.

B) Quantitative (frequency) analysis on lexical forms

As already remarked for the first coding, quantitative analysis results into a higher importance attributed to the most repetitious groups of images. The graphs of the first factor planes derived from the two quantitative analyses on both codings (Figs. 10 and 13: in this case summarizing over 12% of total variation) – are actually quite similar to each other, though the last one shows a slightly more balanced distribution of both variables and images on the first axis. The effect of the different coding of elements becomes negligible in comparison with the increased discriminating power of repeated versus not repeated elements, and can only be perceived in small details, like, for instance, the better distinction of the group of *schematic* seals featuring processions of *women carrying standards* (Figs. 4, b; 6, i).

Quantitative analysis on coding no. 2 lexical forms was carried out on the whole seal corpus, as well. Results were on the whole similar, except for one detail, which is nevertheless worth mentioning: the excessive importance gained in this analysis (contrary to the qualitative one) by a few *unica* (e.g. images made up of a single otherwise very rare icon repeated several times), which completely dominate the third axis. This aspect did not come into evidence in the test on 100 selected images, since most *unica* had been removed from it, but should be considered, in the future, when reflecting on the possible use of quantitative (frequency) analyses.

This is also the only analysis in which the exclusion of syntax and orientation lexical forms significantly alters the results, probably because these forms have a very high frequency in comparison with most of the remaining ones. In fact, the re-coding caused the disappearance of forms defining general categories of elements and attitudes. At first sight, the distribution of images on the plane of the first two axes of this new analysis (Fig. 14: summarizing nearly 11% of total variation) looks rather different from that of the previous analysis (Fig. 13). Under a more careful scrutiny, however, the main difference lies in the exchange between the first and the second

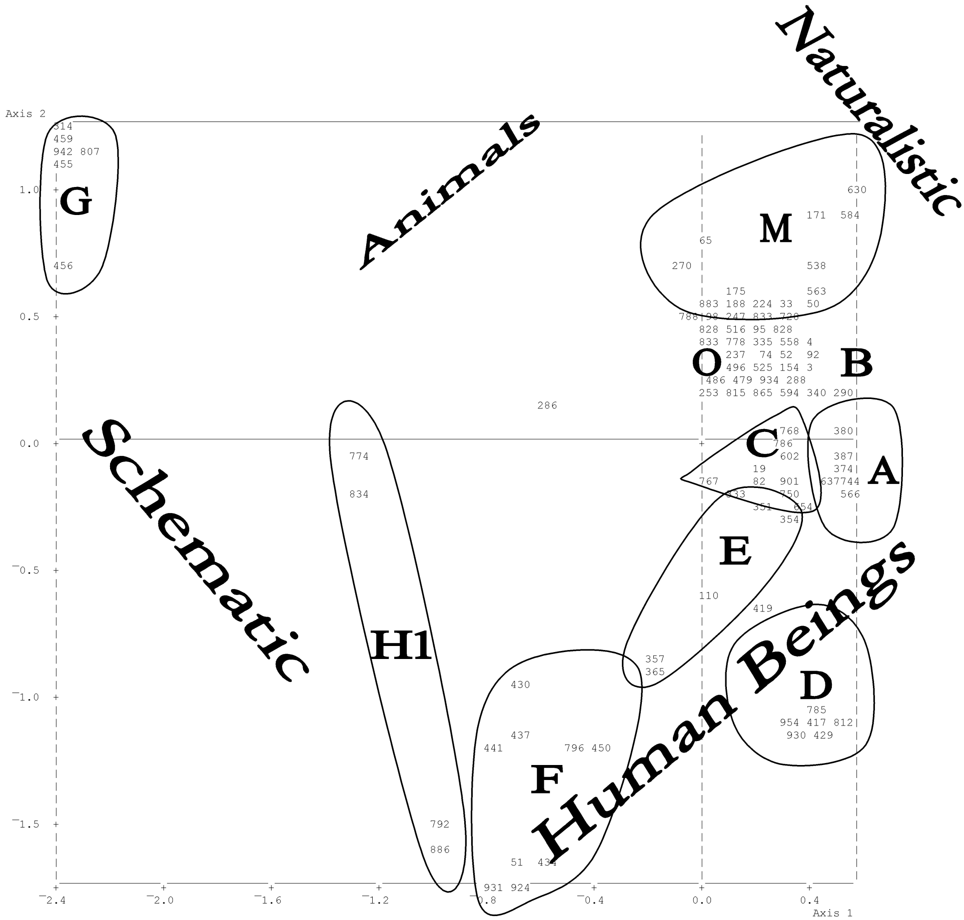


Fig. 13 – Quantitative (frequency) *TCA* on coding no. 2. Representation of the images on the plane spanned by axes 1 and 2. A: War scenes; B: Hunting scenes; C: Naturalistic sacred scenes; D: Schematic sacred scenes (processions of women carrying standards); E: Naturalistic storage and craftsman scenes; F: Schematic craftsman scenes (sitting women manipulating vessels); G: Spiders; H1: Schematic vessels; M: Hybrids and "special animals"; O: Rows of animals.

axis, that is in the relative importance of the first two factors, which remain in both cases roughly the same.

Since the re-coding procedure left lexical forms pertaining to syntax and orientation unaltered, the analysis performed on these forms alone was, of course, identical to the analysis performed on the coding no. 1 texts (Fig. 11), and does not need to be discussed again. However, the projection of the two sets of lexical forms on the first plane (Figs. 15, 16) offers the opportunity for a few comments on the ways elements and attitudes are associated with each other in the different image groups. One can compare, for instance, the position of the forms *woman_craftsman* and

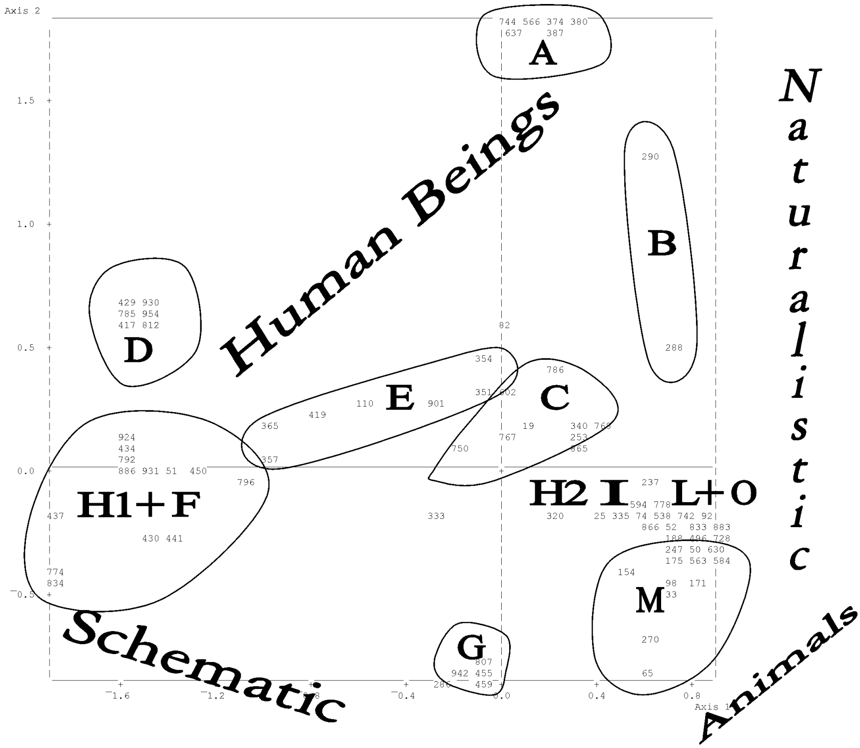


Fig. 14 – Quantitative (frequency) *TCA* on coding no. 2 (syntactical links and orientation excluded from active lexical forms). Representation of the images on the plane spanned by axes 1 and 2. A: War scenes; B: Hunting scenes; C: Naturalistic sacred scenes; D: Schematic sacred scenes (processions of women carrying standards); E: Naturalistic storage and craftsman scenes; F: Schematic craftsman scenes (sitting women manipulating vessels); G: Spiders; H1: Schematic vessels; H2: Non schematic vessels; I: Scenes with humans, animals and objects; L: Animals alternating with "ladders"; M: Hybrids and "special animals"; O: Rows of animals.

sitting_with_parallel_arms, which immediately identify *schematic* seals with *sitting women manipulating vessels*, with the position of the forms *woman*, *craftsman*, *sitting* and *with_parallel_arm*, which they share with other seal groups. In the same way, the different positions of *man*, *woman*, *carrier*, *craftsman* on one side, and of *man_carrier*, *woman_carrier*, *man_craftsman* and *woman_ craftsman* on the other can be compared, and the association of each of these specific icons with different objects, attitudes etc. can be better enlightened.

5.3 *Repeated Segments Analysis*

For the reasons outlined above, *Repeated Segments Analysis* (*RSA*) was performed only on the second text coding. Though its main aim was to investigate the occurrence in the corpus of small fixed sub-patterns (in other

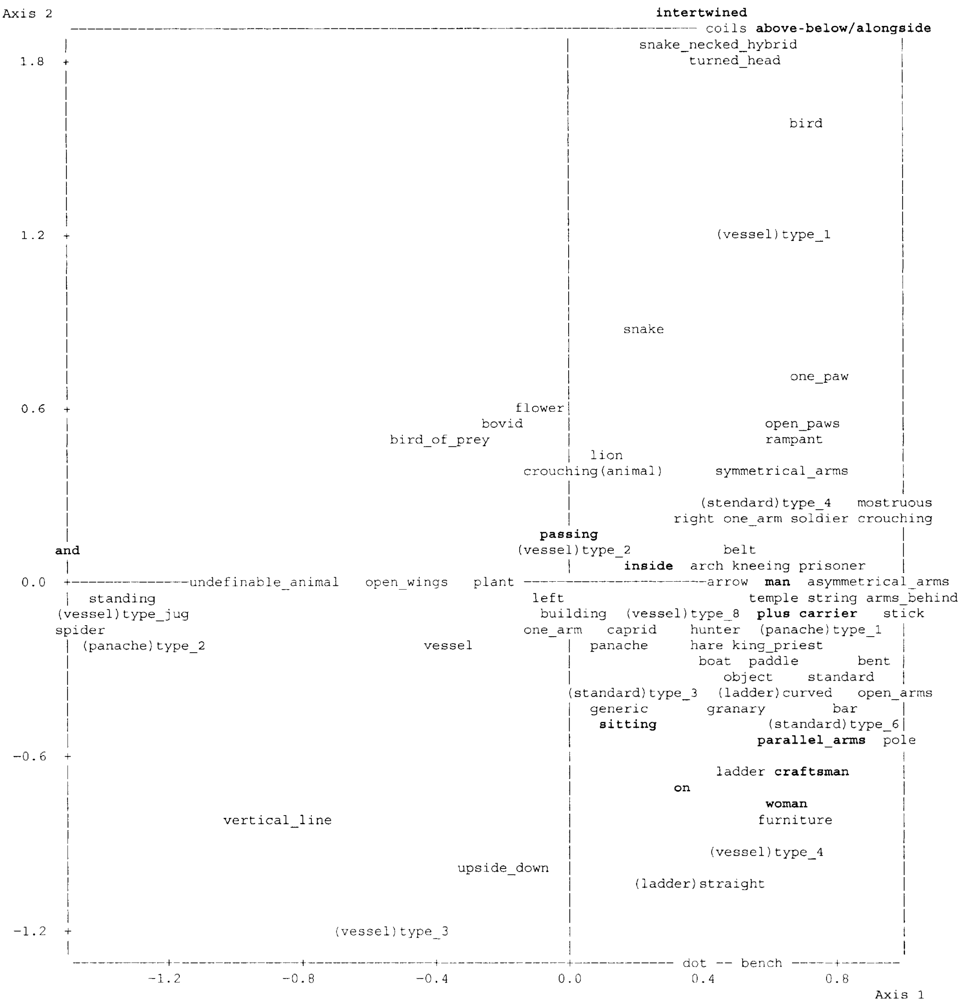


Fig. 15 – Quantitative (frequency) *TCA* on coding no. 1 (syntactical links as only active lexical forms). Projection of the lexical forms on the plane spanned by axes 1 and 2. Syntactical links and lexical forms mentioned in the text are emphasised.

words of the second image level), it turned out to be a useful tool for the study of specific aspects of the lowest (element) level, as well.

Repeated segments occurring in the texts and showing a minimum frequency of two have been automatically selected and listed by the program, resulting into 584 different segments. A further selection, manually performed by the researcher, was required in order to eliminate meaningless or useless sequences: thus, all segments occurring in a single image, and all segments which represented parts of a single icon were excluded from the following

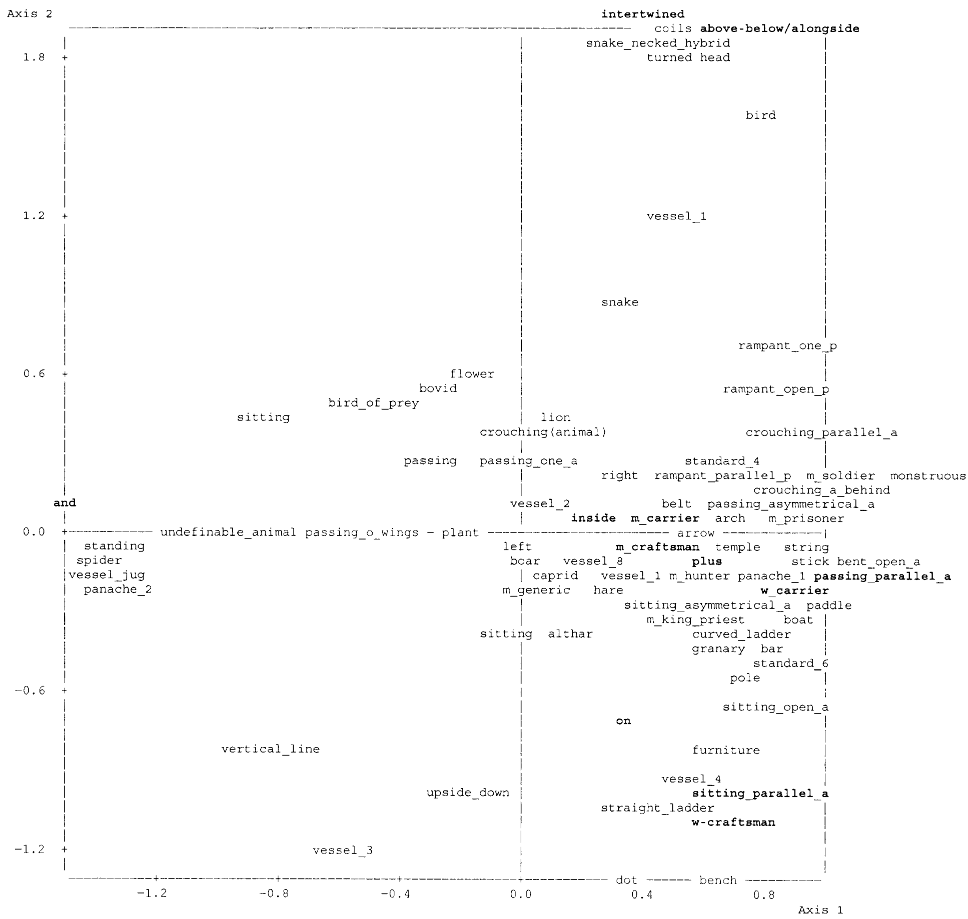


Fig. 16 – Quantitative (frequency) *TCA* on coding no. 2 (syntactical links as only active lexical forms). Projection of the lexical forms on the plane spanned by axes 1 and 2. Syntactical links and lexical forms mentioned in the text are emphasised.

analyses. The remaining segments, some examples of which are shown in Fig. 17, together with the corresponding icons, belonged to two different categories. The first ones (type *a*) were sequences composed by one element, its attitude, and – if applicable – its orientation, which corresponded to a single icon, and therefore still belonged to the lowest image level. The other segments (type *b*), sequences of two (or more) elements joined by a syntactical link, or at least of an element and the following/preceding syntactical link, represented true image sub-patterns of different kinds. Therefore, we decided to start analysing them separately.

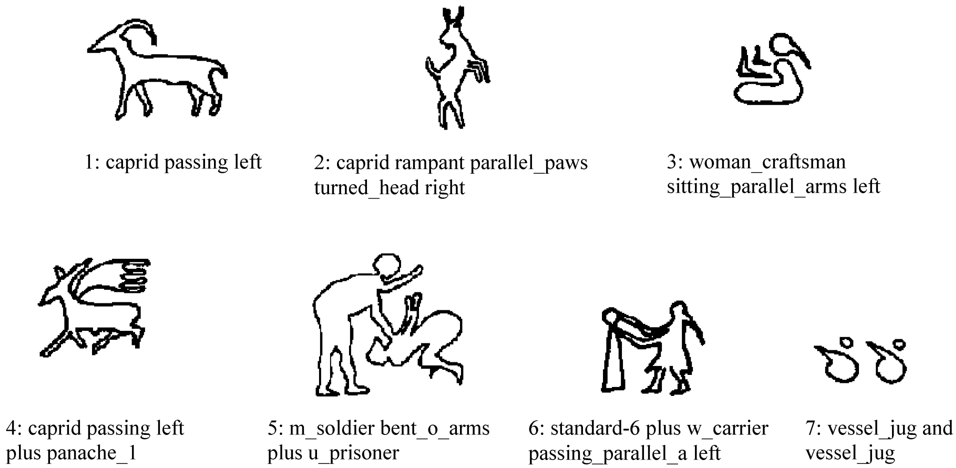


Fig. 17 – Examples of selected repeated segments, with corresponding icons. Nos. 1-3 type a; nos. 4-7 type b.

5.3.1 Analysis of lexical forms plus type *a* selected repeated segments

Type *a* segments were used together with lexical forms for a new, comprehensive analysis of the lowest image level. The inclusion of the new variables rose their total number from 94 to 145. Like in the previous cases, both qualitative and quantitative analyses were performed. As for the first one, the first two axes remained almost unchanged in comparison with qualitative analysis on lexical forms alone, with slight changes becoming perceptible from the third axis onward. The projections of images on the first factor planes are therefore very similar to each other (see Fig. 12). On the other hand, the projection of variables on the same plane (Fig. 18) [7] offers the researcher the possibility to investigate how elements combine with different attitudes to create specific icons of the different seal groups. Consider, for instance, the different locations on the plane of the forms *caprid* and *lion*, in comparison with the segments *passing caprid* and *rampant caprid* (or the segments *passing lion* and *rampant lion*): the *passing* animals tend to be associated with simple *rows* of *animals*, whereas the *rampant* ones appear within more complex and decorative *anima*l composition, etc. In some cases, an almost fixed association between elements and attitudes can be observed: see for instance the position of *woman_craftsman*, *sitting_parallel_arms*, and of the segment *woman_craftsman sitting_parallel_arms*. This immediately identifies what we called *schematic craftsman scenes*, which are further identified by the presence of *benches* and different types of *vessels*. On the contrary, male workers' figures appear in less stereotyped positions and associations (see the positions of *man_craftsman*, *man_craftsman sitting_parallel_arms*, *man_ craftsman passing_parallel_arms*). This possibilities could be better exploited, in future studies, so that a complete analysis of

[7] For the sake of comprehension, one should consider that in Fig. 18 the positive side of the second axis corresponds to the negative in Fig. 12, and vice versa.

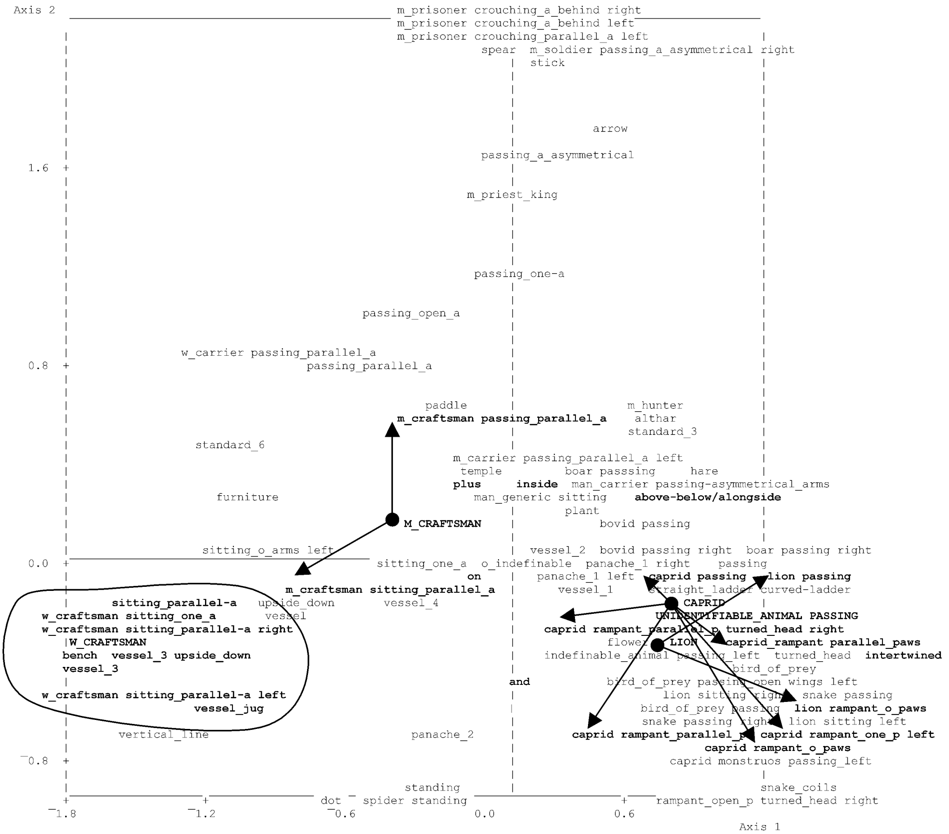


Fig. 18 – Qualitative (presence/absence) *TCA* of lexical forms plus type a selected repeated segments (coding no. 2). Projection of some lexical forms and segments on the plane spanned by axes 1 and 2. Syntactical links and lexical forms and repeated segments mentioned in the text are emphasised.

the first level (including elements, categories of elements, attitudes and their combination into icons) could be obtained from a single textual coding. In our case, for instance, this aim could be reached directly from coding no. 1, through a single analysis combining lexical forms and carefully selected repeated segments.

The results of quantitative analysis show a larger deviation from those of the same analysis on lexical forms alone. In general, it can be said that the importance of small groups of repetitive images tends to become excessive, so that the discriminating power on the remaining images is strongly diminished. There are, however, some groups of seals whose peculiarities emerge only in this type of analysis. Within our study sample, this is the case of repetitive *rows* of *passing unidentifiable animals* (a highly undistinctive element), which are clearly identified through their association in the segment *unidentifiable-animal passing*, which shows a high contribution on both the first two axes.

5.3.2 Analysis of type *b* selected repeated segments

The selection of type *b* repeated segments, to be used as active variables, involved some further problems. In theory, only sequences formed by two or more icons with syntactical links in-between should be considered. In the present 100 images sample, however, these turned out to be very few (49 segments), and furthermore limited for the most part to the same, repetitive groups of images, to be useful for an overall analysis [8]. We tried, therefore, to include in the analysis also sequences formed by an icon and a syntactical link (or vice versa), which should at least provide some information on how different icons tend to associate into fixed sub-patterns. A first analysis, performed on all segments thus selected, did not yield satisfactory results, so that we decided to abandon this procedure.

We preferred to concentrate on the original selection of 49 segments instead. A further selection of segments was deemed necessary to avoid the excessive importance gained by repetitious images: all segments including more than two icons were excluded: these were actually all repetitions of shorter segments, belonging to rows of identical icons (see e.g. the three segments: *vessel_type3 and vessel_type3*; *vessel_type3 and vessel_type3 and vessel_type3*, and *vessel_type3 and vessel_type3 and vessel_type3 and vessel_type3*). Active repeated segments were thus reduced from 49 to 31.

Both qualitative and quantitative *RSA*, performed on these selected segments, had to be limited to the seals (51 out of 100) in which at least one of these was present. Each axis contributes to enlighten single segments (or associations of segments) which are characteristic of separated groups of seals, probably due to their limited number. This is reflected by the first 13 axes eigenvalues (all equal to 1). As a consequence, the 13-dimensional space of representation has no "main" directions, but all have the same importance, and the ordering of the 13 resulting eigenvectors is totally random. In addition, the distribution of both segments and image groups on the plane of any two factors shows only the closeness among image groups and repeated segments specific of the groups themselves. No overall relation among seal groups can be detected from the plane observation. Thus, the usual distinctions between *schematic* and *naturalistic*, *human* and *animalistic* seals, etc., are not visible. Only as an example, we show the plane spanned by the first two factors (Fig. 19).

In spite of the said problems, the results of this analysis are interesting and, above all, different enough from those of all the previous analyses to be shortly discussed. Interestingly enough, the presence of specific repeated segments is not only limited to the most repetitive, *schematic* images: as a matter of facts also groups of *complex naturalistic* seals (e.g. *war* – Figs. 5, d; 6, a – but also *sacred* – Figs. 3, c; 4, b, c; 6, b – and *craftsmanship* scenes) have their own distinctive fixed sequences of icons. Furthermore, the presence of fixed sub-patterns appears to be the only way to

[8] Automatic selection of repeated segments on the whole corpus of Uruk/Jemdet Nasr seals, on the other hand, resulted into a much more diversified set of type *b* segments, which better reflects the variety of its images. This shows how fixed sub-patterns can be detected even in the case of less repetitive compositions, if a sample of images large enough is analysed.

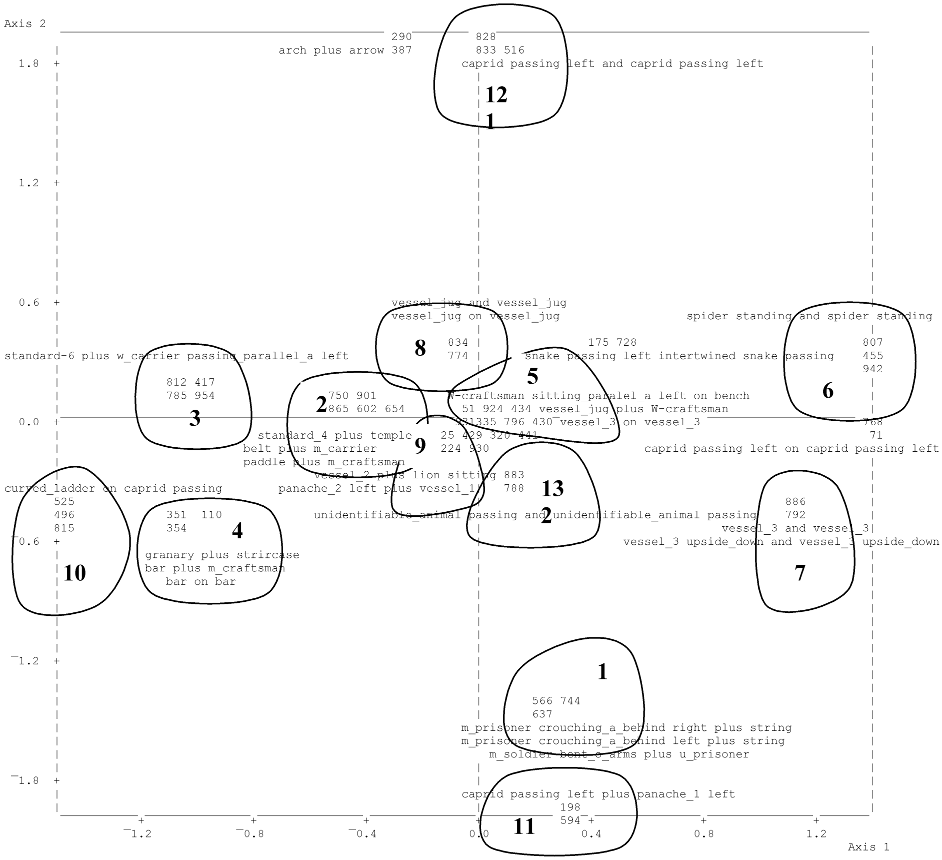


Fig. 19 – Qualitative (presence/absence) *TCA* of type b selected repeated segments (coding no. 2). Representation of the images and of the repeated segments on the plane spanned by axes 1 and 2. 1: War scenes; 2: Naturalistic sacred scenes; 3: Schematic sacred scenes (processions of women carrying standards); 4: Naturalistic storage and craftsman scenes; 5: Schematic craftsman scenes (sitting women manipulating vessels); 6: Spiders; 7: Schematic vessels (rows of vessels type 3); 8: Schematic vessels (rows of jugs); 9: Non schematic vessels; 10: Caprids alternating with "ladders"; 11: Caprids alternating with "panaches"; 12: Rows of caprids; 13: Rows of unidentifiable animals.

characterise some groups of seals (e.g. different compositions of animals and objects, like the so-called *panaches*) which failed to be distinguished by any of the previous analyses (probably because they share single elements which are otherwise typical of different groups of seals).

Summarising, this analysis gives a quite clear distinction of most groups of images, with no example of significant overlapping between groups. Contrary to most of the previous analyses, it could certainly represent an interesting starting point for a classification of the seals.

6. Conclusions

A number of interesting conclusions concerning the suitability of the proposed analysis methods for the study of large corpora of rather stereotyped images can be drawn from the evidence discussed in the previous chapters. These concern first of all the general utility of *CA* – in its several forms – for this type of material, if the significant amount of time and preliminary effort required for its coding are evaluated against the results which can be obtained. As a matter of facts, although *CA* and other exploratory techniques of data analysis are relatively well known and have been often used by archaeologists, they have only sporadically been attempted in the specific field of iconographic studies.

A second question relates to the specific advantages, if any, of the textual coding versus the classical one. The analysis of formalised texts, created by the researcher in order to describe his study object (in this specific case consisting of images) actually represents an attempt at a new, previously unattested application of textual analysis. In fact, this method was originally developed for quite different aims: the study of literary texts, of open-ended questions of surveys, etc., in any case aiming at the analysis of the raw text itself (Lebart, Salem 1988, 1994). In our case, instead, the text is a special coding of the items under study, so that it gets an instrumental character that forces the researcher to be particularly careful in its use.

A last issue concerns the possibility of optimising the results, through the calibration of the whole study procedure, on the basis of the specific features of the analysed corpus, and of the specific aims pursued by the researcher. In other words, we tried to proceed beyond the stage of the more or less casual experimentation of various methods, each time on different assemblages, which has hitherto characterised many applications of mathematical methods to humanities, in favour of a more systematic approach.

For our purposes, the used computer programs, SPAD.N and SPAD.T, have been modified in order to better suit our needs, in particular concerning the selection of items to be taken into account, the transformation of sequences of forms into segments, and the merging of both forms and segments for an overall study.

As for the first question, the tests on the 100 Uruk/Jemdet Nasr seals presented above showed beyond doubt that, regardless of the method used, results were satisfactory enough to amply justify the use of *CA*. The pattern of the studied image sample and the clusters of images suggested by all performed analyses were sufficiently similar to each other to be considered stable, and they well corresponded to the main image groups, as perceived by the archaeologist. In all cases the graphs deriving from the first two axes provide a clear, synthetic overview of the studied corpus. Furthermore, the projection of the variables (or lexical forms and repeated segments) on the same factor planes (see, for instance, Figs. 15, 16, 18, 19) gives a suggestive image of the main factors responsible of the corpus diversity. The main advantage of this kind of analysis is, however, the ability to project on factor planes different sets of variables (like for instance, the provenance, dating etc. of the analysed images, etc.) in order to investigate to which degree these are connected with the different images groups. This easily allows to test different working hypotheses about the studied images (for similar applications to the corpus of Uruk/Jemdet Nasr seals, see Rova 1994).

The suitability of textual coding for the study of images through their transformation into formalised texts has been proved by the rough equivalence of its results to those obtained with classical coding. In comparison with the latter, textual coding requires a special attention, which is however counterbalanced by the wider range of analysis possibilities it offers (choice between quantitative and qualitative analysis, *RSA*, etc.). One should also not underestimate the possibility offered by analysis on textual coding to achieve the analysis of different sets of variables (icon features, icons, sequences of icons) in a faster and so-to-say semi-automatic way, through the interplay between slight modification of the formalised texts, selection of different active variables, and choice between *TCA* and *RSA*. In the discussed case study, for instance, lexical terms, like *woman*, etc. defining general element categories, have been either selected for the analysis or excluded from it, by joining them with the following word to create a new lexical form. On the other hand, the study of icons as a whole, formed by the association of a specific element and attitude, has been achieved through *RSA*.

There are, nevertheless, some limits to these advantages. As a tool for the investigation of images, *TCA* turned out to be particularly appropriate for simple, repetitive images composed of relatively few elements in stereotyped attitudes and simple compositional patterns. In these cases, all the possibilities offered by the method can be exploited with good results. On the contrary, corpora formed by complex syntax images, composed of a large number of elements – especially if these differ from each other in a number of small but significant details – definitely exceed the method's capabilities [9]. As a matter of facts, the mere coding of such images in textual form implies a number of complex, nearly insoluble problems, concerning not only the selection of significant features and the choice of appropriate terms to describe them, but especially their arranging into a fixed order within the text. For such types of images, *MCA* based on classical coding through a presence/

[9] A further problem, especially for quantitative and repeated segments analysis, can be caused by the presence of a few *unica* showing repeated sequences of otherwise rare icons, which tend to play an excessive role on the first axes.

absence matrix of iconographic elements, preliminary selected by the researcher, seems to represent the most suitable analysis tool. Beside other advantages, classical coding allows to take into account a large number of potentially interesting features (like details common to different iconographic elements, etc.) whose coding in textual form would be virtually impossible.

In the case of image corpora – like the Uruk/Jemdet Nasr seals investigated above – made up of both simple, repetitive, and more complex but still to some extent stereotyped images, analysis of textual coding can still provide more than satisfactory results. Nevertheless, we may suggest a careful selection of active textual items and a tailored choice of suitable analysis modes. According to the needs, one may choose either presence/absence or quantitative analysis, either forms or repeated segments, different frequency thresholds for the selection of active forms, etc. If it is felt necessary, the results can be associated to those obtained through *MCA* of classical coding, or other analysis methods, in order to better enlighten different features of the various image groups.

We thus come to the third and last of the problems mentioned above. As we just said, analyses of both classically and textually coded data can be associated, to clarify different aspects (repetitive versus complex irregular images, specific versus common features of the single groups of images, etc.). This procedure involves a considerable preliminary work, since it requires a double image coding and the use of two different analysis programs. Approximately the same results can be obtained, however, through *TCA*, which in this respect shows considerable advantages in comparison with the classical method. The choice of qualitative or quantitative analysis allows to emphasise different groups of complex and respectively repetitious images. Furthermore, already at the level of lexical form analysis, it is possible, through simple, semi-automatic transformations of the basic texts, to investigate in detail either the common features of the different image groups (by analysing the simplest lexical forms) or their specific features (e.g. by joining different lexical forms together), and compare their distribution within the image corpus. Similar results can be obtained by performing *RSA* on selected sequences of lexical forms, either alone or in association with lexical forms.

Analysis on selected sequences of repeated segments also proved an effective tool to proceed toward the second level of analysis of the images, that is the study of small sub-patterns formed by elements connected through specific syntactic links, a level that can not be reached from the classical coding. The studied case showed how this approach is especially effective for the most stereotyped and repetitive groups of images; but undoubtedly it can prove useful also in the case of the more diversified ones, if analysed assemblages are large enough. In such cases, special care must be devoted to the preliminary selection of segments, which can involve a certain degree of subjective choice to avoid the simplest images to gain overwhelming importance on all others. Parenthetically, the significant differences observed between the results of the analyses performed on lexical forms and on segments corresponding to small sub-patterns confirm our perception of the importance of the second level of investigation for an overall picture of image corpora. As a matter of facts, in the present study case, some groups of images turned out to be especially characterised, rather than by the presence/ absence of specific elements, by that of fixed sequences of elements, which a classical analysis may not be able to grasp.

To a certain extent, and especially in the case of the simplest composition patterns (like rows of a single element) textual coding can also offer a tool for the preliminary study of the third analysis level – that of overall image composition –, which can be accomplished both through quantitative analysis of syntactical connectors, and through selection of specific repeated segments.

Briefly, *CA* based on the textual coding provides the best results for the study of images in the case of simple, repetitive images, which also need a limited effort to be transformed into formal texts. For more complex images, it can still be used with satisfactory results, though with some special caution, as a substitute of the classical methods. For corpora composed by both complex, irregular and simple, repetitive images, it makes it possible to use an integrated approach, which can provide a balanced picture where each image group is effectively distinguished. Qualitative analysis on lexical forms alone, or on lexical forms coupled with selected repeated segments, turned out to provide the best synthetic picture of the studied corpus. On the other hand, the remaining analyses which have been performed revealed their specific utility in the analytical study of different groups of seals and variables. Furthermore, textual methods proved able to tackle the study of aspects of the images (fixed sub-patterns, etc.) which exceed the simple iconography and are rather related to composition. To sum up, the interplay with different possibilities offered by textual methods provides the researcher with an adaptable tool to investigate different aspects of an image corpus in detail, and thus allows him/her to test various working hypotheses about its structure, as well as various coding modes. This makes it a worth using method, in spite of the special care needed, and of some still unsolved difficulties, both in the phases of coding and in the choice of active variables.

Sergio Camiz
Dipartimento di Matematica "Guido Castelnuovo"
Università di Roma "La Sapienza"

Elena Rova
Dipartimento di Scienze dell'Antichità
e del Vicino Oriente
Università "Ca' Foscari", Venezia

Acknowledgement

This work has been carried out thanks to research grants nos. 98.00337.CT12 and 99.00580.CT12 of Consiglio Nazionale delle Ricerche. Modifications to SPAD.T program for our specific use were performed by Prof. Monique Bécue during her stay at Roma University "La Sapienza" as invited professor. Both institutions grants and Prof. Bécue's help are gratefully acknowledged.

REFERENCES

Bacelar-Nicolau H., Nicolau F.C. 1994, *Exploratory and Confirmatory Discrete Multivariate Analysis in a probabilistic approach for studying the regional distribution of Aids in Angola*, in Diday *et al*. 1994, 610-618.

Bécue M., Haeusler L. 1995, *Vers une post-codification automatique*, in S. Bolasco, L. Lebart, A. Salem (eds.), *JADT 1995 – III Giornate Internazionali di Analisi Statistica dei Dati Testuali*, Roma, CISU, 1, 35-42.

Benali H., Buvat I., Frouin F., Bazin J.P., Chabriais J., Di Paola R. 1994, *Factor Analysis of Medical Image Sequences (FAMIS): Fundamental principles and applications*, in Diday *et al*. 1994, 619-627.

Bénzécri J.-P. (ed.) 1982, *L'analyse des données*, I-II, Paris, Dunod.

Bolasco S. 1999, *Analisi multidimensionale dei dati*, Roma, Carocci.

Camiz S. 1993, *STATIS Ordinations vs. the Juhász-Nagy Models: the Predictability of an Exploratory Tool,* «Abstracta Botanica», 17 (1-2), 29-36.

Camiz S. 1994, *Metodi di codifica ed analisi dei dati per lo studio di sigilli sumerici,* in *Atti della XXXVII Riunione Scientifica della Società Italiana di Statistica*, Roma, CISU, 2, 105-112.

Camiz S. in press, *Exploratory 2- and 3-way Data Analysis and Applications,* «Lecture Notes of *TICMI*», Tbilisi International Centre of Mathematics and Informatics, 2 (2000).

Camiz S., Rova E. 1991, *Analysis of ancient Near-Eastern cylinder seals (late fourth millenium B.C.)*, in E. Diday, Y. Lechevallier (eds.), *Symbolic-Numeric Data Analysis and Learning*, Commack (NY), Nova Science Publisher, 131-142.

Camiz S., Rova E. 1992, *Analyse des sceaux cylindriques du Proche Orient ancien de la période de Uruk/Jemdet Nasr*, in M. Bécue, L. Lebart, N. Rajadell (eds.), *Jornades Internacionals d'Anàlisi de Dades Textuals 1990*, Barcelona, Universitat Politecnica de Catalunya, 213-216.

Camiz S., Rova E. 1993a, *L'analyse des segments repétés pour l'étude des sceaux du Proche-Orient*, in *Secondes journées internationales de analyse statistique des données textuelles*, Montpellier, Telecom Paris, 437-445.

Camiz S., Rova E. 1993b, *Multivariate Analysis Techniques applied to ancient Near Eastern seals of the late fourth millennium B.C.*, in J. Pavuk (ed.), *Actes du XII^e Congrès International des Sciences Préhistoriques et Protohistoriques (Bratislava 1991)*, Bratislava, 133-138.

Camiz S., Rova E. 1996, *Metodi di analisi per lo studio di un gruppo di sigilli cilindrici vicino-orientali e di altre immagini strutturate*, «Archeologia e Calcolatori», 7, 647-659.

Camiz S., Rova E., Tulli V. 1998, *Exploratory analysis of images engraved on ancient Near-Eastern seals based on a distance among strings*, «Statistica», 58, 670-689.

Camiz S., Tulli V. 1994, *Distance indexes among hierarchic strings*, in R. Dutter, W. Grossmann (eds.), *Compstat 1994, Proceedings in Computational Statistics, Short Communications*, Vienna, Vienna Institutes for Statistics, University of Technology, University of Vienna, 45-46.

Diday E. *et al.* (eds.) 1994, *New Approaches in Classification and Data Analysis*, Berlin, Springer.

Digard F. *et al.* 1975, *Répertoire analytique des cylindres orientaux publiés dans les sources bibliographiques éparses*, I-III, Paris, Editions du Centre National de la Recherche Scientifique.

Lebart L., Morineau A., Bécue M., Haeusler L. 1994, *Spad.t – version 1.5 – Manuel de référence*, Paris, Cisia.

Lebart L., Morineau A., Lambert T., Pleuvret P. 1991, *Spad.n – version 2 – Manuel de référence,* Paris, Cisia.

Lebart L., Morineau A., Piron M. 1995, *Statistique exploratoire multidimensionelle*, Paris, Dunod.

Lebart L., Morineau A., Warwick K. 1985, *Multivariate Descriptive Statistical Analysis*, New York, J. Wiley and Sons.

Lebart L., Salem A. 1988, *Analyse statistique des données textuelles*, Paris, Dunod.

Lebart L., Salem A. 1994, *Statistique textuelle*, Paris, Dunod.

Moscati P. 1986, *Analisi statistiche multivariate sugli specchi etruschi*, Contributi del Centro Linceo Interdisciplinare 74, Roma, Accademia Nazionale dei Lincei.

Moscati P. 1995, *Quantitative analysis of Etruscan cinerary urns*, in J. Huggett, N. Ryan (eds.), *Computer Applications and Quantitative Methods in Archaeology*, Oxford, Tempus Reparatum, 101-104.

Rova E. 1994, *Ricerche sui sigilli a cilindro vicino-orientali del periodo di Uruk/Jemdet Nasr*, Orientis Antiqui Collectio 20, Roma, Istituto per l'Oriente "C. Nallino".

Rova E. 1995, *Metodi di codifica e analisi per lo studio di immagini strutturate: un'applicazione ai sigilli cilindrici del periodo di Uruk/Jemdet Nasr*, «Archeologia e Calcolatori», 6, 7-34.

Rova E. 1996, *A method for analyzing the composition structure of images: a test on Near-Eastern cylinder seals*, *Proceedings of the XIIIth Congress of the International Union of Prehistoric and Protohistoric Sciences*, Forlì, MAC, 1, 27-34.

Suter C.E. 1999, *Review of Rova 1994*, «Journal of Near Eastern Studies», 58, 47-53.

ABSTRACT

In order to test the ability of textual coding to depict the features of structured images, a corpus of images of Near-Eastern seals of the late IVth millennium B.C. was studied through different exploratory analysis techniques. Two different coding systems were considered: the classical presence/absence coding of iconographical elements present in the images and a new textual coding, based on a formalised text describing the image. These were submitted to Multiple and Textual Correspondence Analyses, respectively. The textual analyses were performed according to two different coding systems, and several choices of the items involved. The results of the different analyses are discussed and compared. In particular, textual analysis proved effective in substituting the classical coding in the description of the iconographic elements appearing in the images. In addition, it allowed us to broaden the investigation to include aspects of the images (occurrence of fixed sub-patterns and composition) which are beyond the capacities of classical coding. The ability of textual coding to select particular elements, and/or element sequences, to be taken into account in the analyses, was also considered an interesting feature for a fine-tuning of the analyses to the particular characters of specific corpora. Thus, the use of a formalised text as an intermediate between images and analysis tools proved to be a method worth using, in spite of the special care needed, and some still unsolved difficulties.