22. Relational description, similarity and classification of complex archaeological entities

Costis Dallas
Benaki Museum, Athens, Greece.

22.1 Introduction

A number of alternative numerical classification and ordination methods have been used in the last few decades to address problems of archaeological typology. To my knowledge, these methods invariably depend on the manipulation of either a rectangular data matrix, consisting of rows of units characterised by columns of attributes (Doran & Hodson 1975:93–94), or of data structures derived from a data matrix, such as a matrix of similarity indices between units or a correlation matrix between attributes. In this paper, I argue that the rectangular data matrix is appropriate neither for the description nor for the analysis of an important category of complex archaeological entities. These entities, which comprise burial assemblages, house plans, design layouts and figured representations, exhibit internal structure which is not fixed, but variable. I propose instead that complex entities with variable internal structure may be appropriately described as attributed relational graphs. I also suggest that their taxonomic relationship may be investigated by a generalised *relational similarity coefficient* (*RSC*), extracted by means of inexact graph matching, a technique already used successfully in the field of image analysis and understanding.

The popularity of numerical classification in archaeology is linked with the acceptance of the polythetic approach to typology, advocated by Clarke (1978). According to Clarke, most archaeological entities may be seen as *aggregates* of entities of lower taxonomic rank. Thus, archaeological cultures consist of assemblages, which consist of types, which consist of artefacts, which are composed of specific traits (*ibid*:35–37). Artefacts, in particular, are defined as unstructured sets or lists of attributes, 'irreducible character(s) of two or more states, acting as independent variable(s) within a specific frame of reference', (*ibid*.:156); the similarity between two artefacts is seen as a function of the affinity between their respective global attribute lists (*ibid*.:158); artefact types, the *desiderata* of classification, are defined as 'homogeneous population(s) of artefacts which share a consistently recurrent range of attribute states within a given polythetic set', (*ibid*.:209).

Although not universally espoused (Whallon 1972; cf. Gordon 1981:69–74) the explicit polythetic approach was an important breakthrough in relation to traditional approaches to classification, usually combining monothetic theory with polythetic practice. Nevertheless, Clarke had intended his hierarchical definition of archaeological entities, including the attribute-artefact-type model, as a contribution to a 'healthy anarchy', which could 'provoke modifications and alternatives from which a satisfactory terminology may ultimately be selected', (Clarke 1978:416). In retrospect, it is remarkable how long the part of that definition concerning artefact morphology has remained unchallenged. Although the selection of an appropriate level of descriptive resolution for the definition of meaningful attributes was soon identified as a problem, the solution advanced was that alternative attribute sets should be used and the results of respective analyses compared, e.g. in the form of constellation analysis (Doran & Hodson 1975:205–209). While the realisation that attribute selection should be problem- or theory-driven is no doubt correct (cf. Shennan 1988:194), it obviously does not affect a fundamental premise of Clarke's hierarchical scheme: namely, that it is possible to describe archaeological artefacts by means of a list of global attributes.

Doran (1971) demonstrated how complex artefacts such as Iron Age *fibulae* can be successfully described and analysed using a global attribute list. *Fibulae* were decomposed to a number of constituent parts such as 'coil' and 'bow', and a set of pertinent attributes was defined for each part. A rectangular data matrix was formed by the union of attribute sets for object parts, and was used for numeric classification. Data description and analysis on the basis of a global attribute list is clearly possible for other complex artefacts, such as spearheads and pots (Orton 1980:38–42). It is no accident that the rectangular data matrix is the generic representation suggested for data analysis in the latest handbook on quantitative archaeology as well (Shennan 1988).

It would be rash to conclude, however, that all complex archaeological entities can happily reside in a rectangular data matrix. An example is provided by Shennan's rhetorical question:

> "...in studying a cemetery of inhumation burials containing grave goods and trying to make inferences about the social organisation of the community which deposited the burials, do you include information on the position of each of the grave goods in the grave as well as what they are? Perhaps the exact position of the limbs of the skeleton is significant in some way?" (Shennan 1988:10).

Although the issue of burial description is presented here in the guise of the old question of descriptive resolution (cf. Doran & Hodson 1975:101–102), it amounts to much more. Shennan suggests a pilot study of burials, 'using the full description' of skeletal and artefact position; such a study, however, would soon run up against a fundamental problem noted by Doran (1986:28), namely, that the description of burials cannot be reduced to a list of global attributes, since it involves a variable internal structure of parts, relationships and functions. In terms of internal structure, *fibulae* and arrowheads are just a special case

Part no.	Type
1	Outside

Relation no.	Width	Sides	Columns
1	1.5	1	0
2	2.5	1	0
3	2.5	1	0
4	2	1	0
5		1	0
6	1.5	1	0
7	1	1	0
8	3.5	1	0
9	1	1	0
10	1	1	0
11	22.5	3	10
12	3.5	1	1
13		1	0
14	1.5	1	0

Part no.	Type	Length	Width	Feature	Mosaic
2	Room	8	2.5	None	No
3	Room	17.5	12.5	None	No
4	Room	3.5	1.5	None	No
5	Room	10	5	Stair	No
6	Room	3.5	1.5	None	No
7	Room	3.5	1.5	None	No
8	Room	4.5	4.5	None	No
9	Room	3.5	3	None	No
10	Room	4	2	None	No
11	Room	8	3.5	Latrine	No
12	Court	9	8	Altar	No
13	Room	9	4.5	None	Yes
14	Room	5	4.5	None	No
15	Room	6.5	6	None	No

Table 22.1: Attributes and relations for the symbolic description of the ground floor of house II, E. from Delos.

of complex entities, composed of a fixed configuration of components. A general model of archaeological morphology should, however, provide for the adequate description and analysis of all complex entities, including those with variable internal structure.

22.2 Complex entities: current approaches

The importance of internal structure for archaeological data description has not escaped the notice of the French *logicistes*, as is apparent from the extensive discussion of configuration and syntax in the descriptive codes they devised for different classes of archaeological material (e.g. Gardin 1978:45–51; Salomé 1980:39–48, 100–108); in fact, several important database applications, especially in Classical archaeology, explicitly acknowledge that archaeological objects are composed by a nested structure of parts, and use a hierarchical decomposition scheme for their formal description and retrieval (Ginouvès & Guimier-Sorbets 1978; Eisner 1988; Guimier-Sorbets 1990). On the other hand, the relative lack of attention of quantitative archaeology to the internal spatial, topological and semantic structure of complex entities can only be explained by the impact of Clarke's unstructured set-based archaeological morphology model, and by the wide availability of analytical techniques operating on a rectangular data matrix.

I have identified, among quantitative archaeological studies, three approaches to the description and analysis of complex entities. The first is illustrated by Pader's structuralist analysis of Anglo-Saxon graves, a study recognising explicitly 'the symbolic nature of space use', and thus the importance of the internal spatial organisation of burials for communicating social meaning (Pader 1982:78–79). A global attribute list, representing both substantive and positional information about skeletal remains and grave offerings, is used in that study. Positional categories, 'e.g. extended and across, were determined after much preliminary data manipulation' (*ibid.*:81), and are not the same for different types of objects. The position of specific objects is defined by conditional attributes, used only if the objects were present; positional information is duplicated in a number of general categories, pooling all objects together. A similarity matrix is derived from the rectangular data matrix, according to a modified version of Gower's general similarity coefficient, and is used for principal coordinates analysis.

Pader is aware of the duplication of data pooled together in her data matrix, and in fact argues that the importance of positional information is a reason for their preferential weighting (*ibid.*:79). However, the coding of positional information is by necessity *ad hoc*, and there is no exact correspondence between object-specific and general positional categories; since the detail at which positions are recorded for different objects is unpredictably variable, one is not clear what to make of the statistical significance of the association of attributes with principal coordinates axes, forming the basis of Pader's descriptive account of burial variability. An important shortcoming is also that the global attribute list cannot account for spatial interrelationships between contained objects, unless they are anchored to a specific position with regard to skeletal parts. The inadequacy of the rectangular data matrix is illustrated by the fact that, in order to discuss the internal structure of burials, Pader uses an alternative set of important analytical categories — including congruence, addition and substitution of burial components — that cannot be derived directly from the original data representation (*ibid.*:113–126).

The second approach is illustrated by Ciolek-Torello's analysis of room function in Southwestern archaeology (Ciolek-Torello 1984). Instead of attempting to isolate

Part no.	Type
1	Outside

Relation no.	Width	Sides	Columns
1	2	1	0
2	1	1	0
3	1.5	1	0
4	3.5	1	0
5	13.5	4	8
6	4	1	2
7	2.2	1	0
8	1.5	1	0
9	1.5	1	0
10	1	1	0
11	1.5	1	0
12	1.5	1	0
13	1	1	0
14	1	1	0
15	1	1	0
16	1	1	0

Part no.	Type	Length	Width	Feature	Mosaic
2	Room	4.5	4.5	None	No
3	Room	4	3	None	No
4	Room	14.5	13.5	None	No
5	Room	3.5	1.5	Stair	No
6	Court	7	6.5	None	No
7	Room	11	7.5	None	Yes
8	Room	3	1	None	No
9	Room	4.5	3	None	No
10	Room	5	4	None	No
11	Room	7	5.5	None	No
12	Room	6	5.5	None	No
13	Room	2.5	2	None	No
14	Room	4	1	None	No
15	Room	2.5	2	Latrine	No
16	Room	4	3	None	No

Table 22.2: Attributes and relations for the symbolic description of the ground floor of house II, F from Delos.

habitation nuclei and derive a global attribute list for them, single rooms form the basic data unit in that study. It may be the case that connections between rooms are not known from the archaeological record; however, known spatial contiguity relations between rooms are not used in the analysis either. Compositional data about finds are employed to derive ordination scores and subsequently to classify rooms into six classes, corresponding to the domestic activities of manufacturing, storage and food processing, and their combinations (*ibid.*:143–147). The spatial disposition of different functional units is then examined *ex post facto*: 'different room types', it is noted, 'are not randomly distributed throughout the pueblo', but are organised around particular settlement areas (*ibid.*:148). However, despite the fact that households of different size are recognized, no attempt is made to infer the formal spatial structure of these households.

I followed a similar method of using the parts of entities as primary data in an iconographic problem, i.e., the identification of the primary deceased among figures represented on funerary reliefs of Classical Athens (Dallas 1987); I used available epigraphic evidence to generate a discriminant function involving iconographic attributes such as stance, figure size, costume and gestures, and then assigned figures to the 'primary deceased' and 'mourner' categories according to their iconography. While interesting associations were found, the results paid no respect to the fact that one and only one figure should be identified as the primary deceased for each grave-relief, and that the rôles of figures depend not only on their intrinsic attributes, but also on their spatial and topological configuration.

Even assuming that Ciolek-Torello's entity part description approach would be valid for Southwestern *pueblos*, it is clearly inadequate for the examination of ancient Greek house plans, where rooms with specific functions are demonstrably linked in a non-random spatial and topological configuration. As shown in the example of Classical grave reliefs, this is frequently the case with iconographic material as well. In sum, using entity parts as the fundamental descriptive units is an inadequate approach, since it makes no use of relational information which is available before the analysis and which may be essential for the determination of archaeological meaning.

The third approach to complex archaeological entities consists of the imposition of a specific theoretical framework on data prior to their description and analysis. Glassie's grammar of American folk house plans is a prime example of such a theory-laden descriptive and classificatory exercise, based on the operation of a set of production rules on a dictionary of elementary architectural forms (Glassie 1975). A descriptive system based on a case grammar, identifying domain-specific rôles such as primary and secondary subject, was introduced for the iconography of Greek vase-painting by Marie-Rose Salomé (1980); formal treatments of figurative art often depend on formalisation of a set of predetermined rules derived from accepted archaeological knowledge, sometimes in the form of an expert system (Gardin 1980:108–120; Lagrange & Renaud 1983; Lagrange & Renaud 1987). The symmetry analysis of design patterns may also be seen as an axiomatic production system, using geometric transformation rules to derive surface structures from a vocabulary of basic shapes (Washburn 1983). Rule-based representations are also applicable to figurative art, as is shown by the use of componential

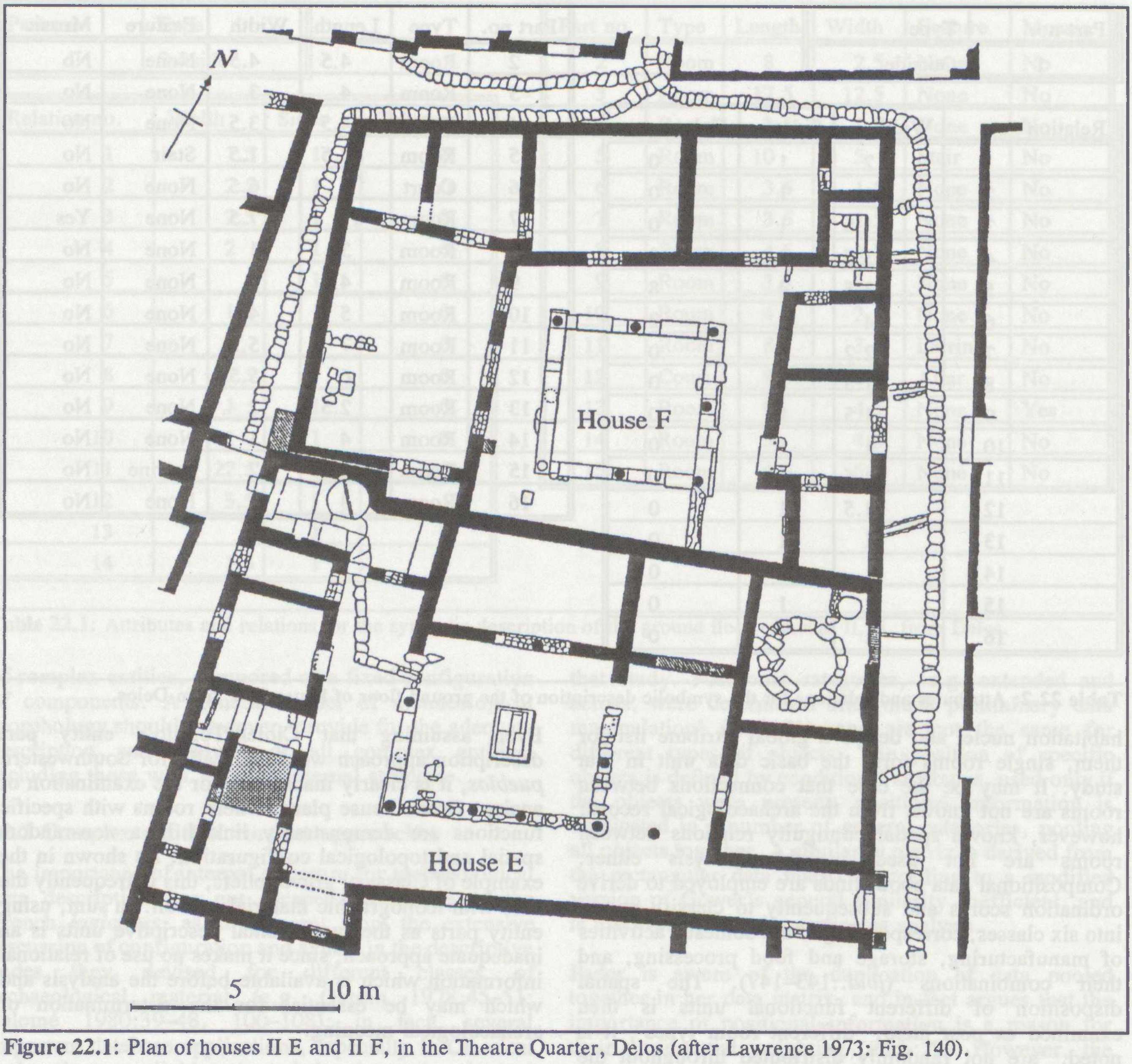

Figure 22.1: Plan of houses II E and II F, in the Theatre Quarter, Delos (after Lawrence 1973: Fig. 140).

trees, transition networks and grammatical phrase markers for the visualisation of pictorial syntactic relationships (Dallas forthcoming). The hierarchical decomposition of archaeological objects by means of trees akin to 'phrase markers' was found particularly useful in the analysis of ceramic decoration (Hodder 1982; Hardin 1983).

It is apparent that the third approach adopts richer data representations such as strings and trees, and takes into account the spatial, topological or semantic relations in the structure of the data. It depends, however, on a procedure of first constructing an axiomatic theory, such as a grammar, and then testing its premises on suitably structured empirical data. While this approach has merits for certain types of problems, it cannot be universally applied. Often, quantitative analysis of complex archaeological entities is used for data exploration, prior to the generation of structural hypotheses. In these cases, we should need a symbolic representation that is generic and objective, but also appropriate to the internal structure of complex entities.

22.3 Relational description

The goal of description of complex archaeological entities for numerical analysis is congruent with that of knowledge representation, i.e., the definition of symbolic representations of the *state of the world* amenable to machine manipulation (Levesque 1986). A symbolic representation for complex entities which display variable internal structure should allow effective data retrieval and comparison between entities, and should also 'supply the symbolic elements that go into the formulation of theories' (Gardin 1980:39–40); it should be general, in allowing the description of all possible facts in a given domain, and also neutral, in not depending on the adoption *a priori* of a particular theory on the structure of the data.

A general framework for the symbolic representation of archaeological objects is provided by relational graph structures. A graph is a richer data structure than the set or *aggregate* used in Clarke's model of archaeological morphology, since it can also represent not only set membership, but also relations between the

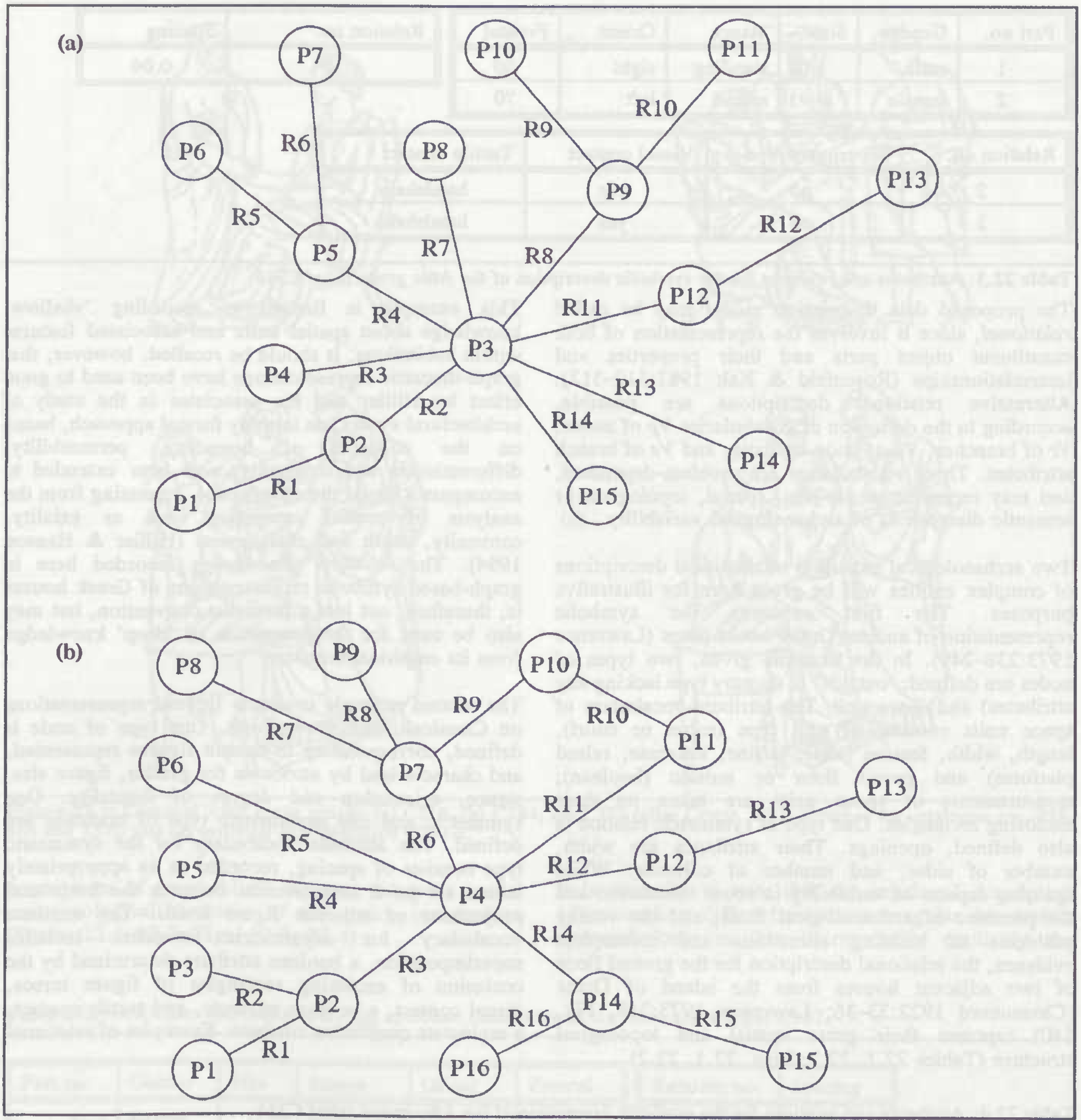

Figure 22.2: Relational description graphs for the ground floor of Delos houses: (a) II E, and (b) II F.

set's constituent elements (Doran & Hodson 1975:13). As it was noted in image analysis, pictures, a typical class of complex entities, 'can generally be represented by a graph structure in which the parts correspond to nodes, labelled with lists of property values … and the arcs are labelled with lists of relationship values' (Rosenfeld & Kak 1982:304–305); this is also the case with other complex entities characterised by variable internal structure, such as burials, architectural plans and designs. If all relations between object parts are symmetric, the representation is called a *digraph*; a *multigraph* is necessary for the representation of ordered relations, since more than one branch may be required between a given pair of nodes.

A general symbolic representation for the description of complex archaeological entities may, therefore, be formally defined as an attributed relational graph (ARG), of the form

$$ARG = (P, R, A, E)$$

where

P	is a finite set of nodes, selected from V, a vocabulary of object part types,
R	is a finite set of ordered and/or symmetric branches, selected from Vr, a vocabulary of relation types between pairs of object parts,
A	is a finite set of attribute values, selected from Va, a vocabulary of attributes for object part types, and
E	is a finite set of attribute values, selected from Ve, a vocabulary of attributes of object part relation types.

Part no.	Gender	Size	Stance	Orient	Frontal
1	male	1.00	standing	right	60
2	female	0.91	seated	left	70

Relation no.	Spacing
1	0.04

Relation no.	Superimposition	Visual contact	Tactile contact
2	no	yes	handshake
3	no	yes	handshake

Table 22.3: Attributes and relations for the symbolic description of the Attic grave relief C216.

The proposed data description model may be called *relational*, since it involves the representation of both constituent object parts and their properties and interrelationships (Rosenfeld & Kak 1982:310–312). Alternative relational descriptions are possible, according to the definition of vocabularies *Vp* of nodes, *Vr* of branches, *Va* of node attributes and *Ve* of branch attributes. These vocabularies are problem-dependent, and may represent whole-part, spatial, topological or semantic dimensions of archaeological variability.

Two archaeological examples of relational descriptions of complex entities will be given here for illustrative purposes. The first concerns the symbolic representation of ancient Greek house plans (Lawrence 1973:238–249). In the example given, two types of nodes are defined: 'outside' (a dummy type lacking any attributes) and space unit. The attribute vocabulary of space units consists of unit type (room or court), length, width, feature (altar, latrine, staircase, raised platform) and paved floor or mosaic (boolean); measurements of space units are taken on their enclosing rectangles. One type of symmetric relation is also defined, openings. Their attributes are width, number of sides, and number of columns. While ignoring aspects of variability in room orientation and the presence of archaeological finds, and the vexing problems of building alterations and incomplete evidence, the relational description for the ground floor of two adjacent houses from the island of Delos (Chamonard 1922:33–36; Lawrence 1973:248, Fig. 140) captures their gross spatial and topological structure (Tables 22.1, 22.2; Figs. 22.1, 22.2).

This example is limited to recording 'shallow' knowledge about spatial units and associated features within habitations. It should be recalled, however, that graph-theoretic representations have been used to good effect by Hillier and his associates in the study of architectural syntax; an initially formal approach, based on the concepts of boundary, permeability, differentiation and contiguity, was later extended to encompass a social theory of space, stemming from the analysis of spatial categories such as axiality, convexity, depth and shallowness (Hillier & Hanson 1984). The 'shallow' knowledge recorded here in graph-based symbolic representations of Greek houses is, therefore, not just a formalist convention, but may also be used for the generation of 'deep' knowledge from its empirical domain.

The second example concerns figured representations on Classical Attic gravestones. One type of node is defined, corresponding to human figures represented, and characterised by attributes for gender, figure size, stance, orientation and degree of frontality. One symmetric and one asymmetric type of branches are defined. The attribute vocabulary for the symmetric type consists of spacing, recorded as an appropriately scaled set point measurement between the horizontal projections of adjacent figure heads. The attribute vocabulary for asymmetric branches includes superimposition, a boolean attribute determined by the occlusion of enclosing rectangles of figure torsos, visual contact, a boolean attribute, and tactile contact, a multistate qualitative attribute. Examples of relational

Table 22.4: Attributes and relations for the symbolic description of the Attic grave relief C384.

Part no.	Gender	Size	Stance	Orient	Frontal
1	female	0.82	seated	right	80
2	male	1.00	standing	right	45
3	male	0.98	standing	left	30

Relation no.	Spacing
1	0.24
2	0.33

Relation no.	Superimposition	Visual contact	Tactile contact
3	yes	no	no
4	no	no	no
5	no	yes	handshake
6	no	yes	handshake
7	no	yes	no
8	no	no	no

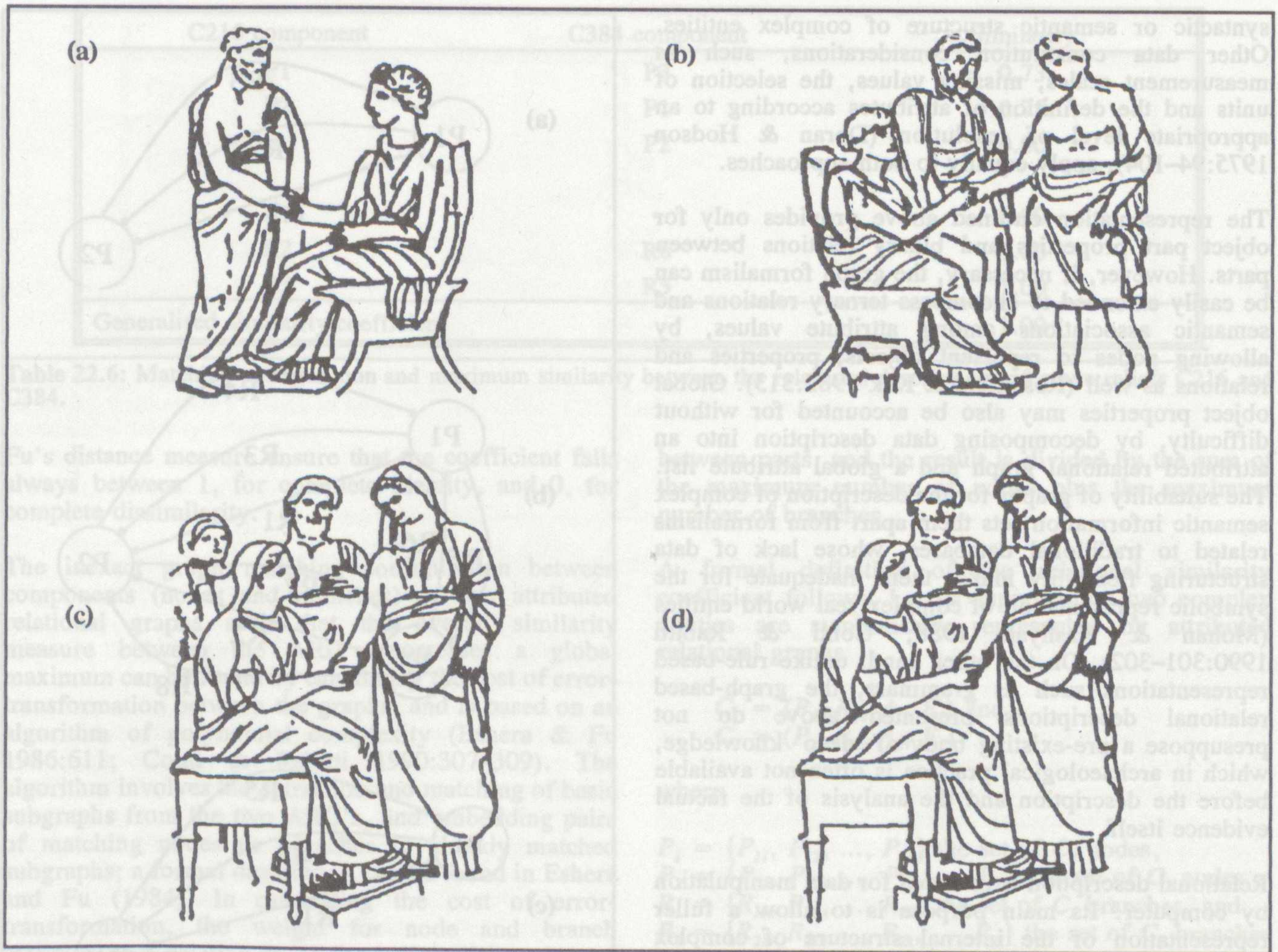

Figure 22.3: Classical Attic grave reliefs: (a) C216, (b) C384, (c) C337 (after Conze 1893: sketches for catalogue nos. 216, 384 and 337), and (d) artificial representation C337-a.

descriptions of grave-reliefs illustrate the gross spatial and topological structure of the entire composition (Tables 22.3–22.5; Figs. 22.3, 22.4).

Table 22.5: Attributes and relations for the symbolic description of the Attic grave relief C337.

Part no.	Gender	Size	Stance	Orient	Frontal
1	female	0.90	standing	right	30
2	female	0.82	seated	right	70
3	male	0.96	standing	right	10
4	female	1.00	standing	left	30

Relation no.	Spacing
1	0.17
2	0.10
3	0.21

Relation no.	Superimposition	Visual contact	Tactile contact
4	no	yes	no
5	yes	no	no
6	yes	no	no
7	no	no	no
8	no	yes	amplified
9	no	yes	handshake
10	no	yes	no
11	yes	no	no

It is essential to note that the relational description advocated here differs from the rectangular data matrix approach only in that it represents explicitly the syntactic or semantic structure of complex entities. Other data constitution considerations, such as measurement scales, missing values, the selection of units and the definition of attributes according to an appropriate level of resolution (Doran & Hodson 1975:94–104), apply equally to both approaches.

The representation outlined above provides only for object part properties and binary relations between parts. However, if necessary, the graph formalism can be easily extended to encompass ternary relations and semantic associations among attribute values, by allowing nodes to represent objects, properties and relations as well (Rosenfeld & Kak 1982:313). Global object properties may also be accounted for without difficulty, by decomposing data description into an attributed relational graph and a global attribute list. The suitability of graphs for the description of complex semantic information sets them apart from formalisms related to traditional databases, whose lack of data structuring flexibility makes them inadequate for the symbolic representation of complex real world entities (Mohan & Kashyap 1988; Conti & Rabitti 1990:301–302). On the other hand, unlike rule-based representations such as grammars, the graph-based relational descriptions presented above do not presuppose a pre-existing body of 'deep' knowledge, which in archaeological practice is often not available before the description and the analysis of the factual evidence itself.

Relational description is intended for data manipulation by computer. Its main purpose is to allow a fuller representation of the internal structure of complex archaeological entities, both for data retrieval and quantitative analysis. If queries are expressed in the form of a graph representation, data retrieval becomes a matching operation, whereby stored relational descriptions of data are searched in order to identify units isomorphic with the symbolic representation of the query. In practice, however, the complexity of entities makes it necessary to allow a degree of tolerance, or error, in the degree of isomorphism sought. Since matching and retrieval of relational descriptions of complex entities generally depends on a sufficient degree of affinity, rather than on absolute identity, the ability of using relational descriptions as an effective means of data retrieval depends on the definition of an appropriate measure of similarity (Conti & Rabitti 1990:306–307).

22.4 A generalised similarity coefficient

As suggested by Eshera and Fu (1984; 1986), a suitable measure of distance between relational descriptions of images may be derived by the method of inexact graph matching. Images are described in the form of attributed relational graphs (ARGs), with image parts defined as nodes, relations as branches, and geometric traits of parts and relations as numeric attributes. The distance between two images is derived by calculating the cumulative cost for error-transformation between their ARG symbolic representations. It is suggested that costs for specific transformations, such as node substitution, branch insertion or deletion, be calculated according to domain-specific weights. The method is advanced as a noise-tolerant method for high-level image understanding, but has since been successfully used for low-level image analysis as well (Conti & Rabitti 1990).

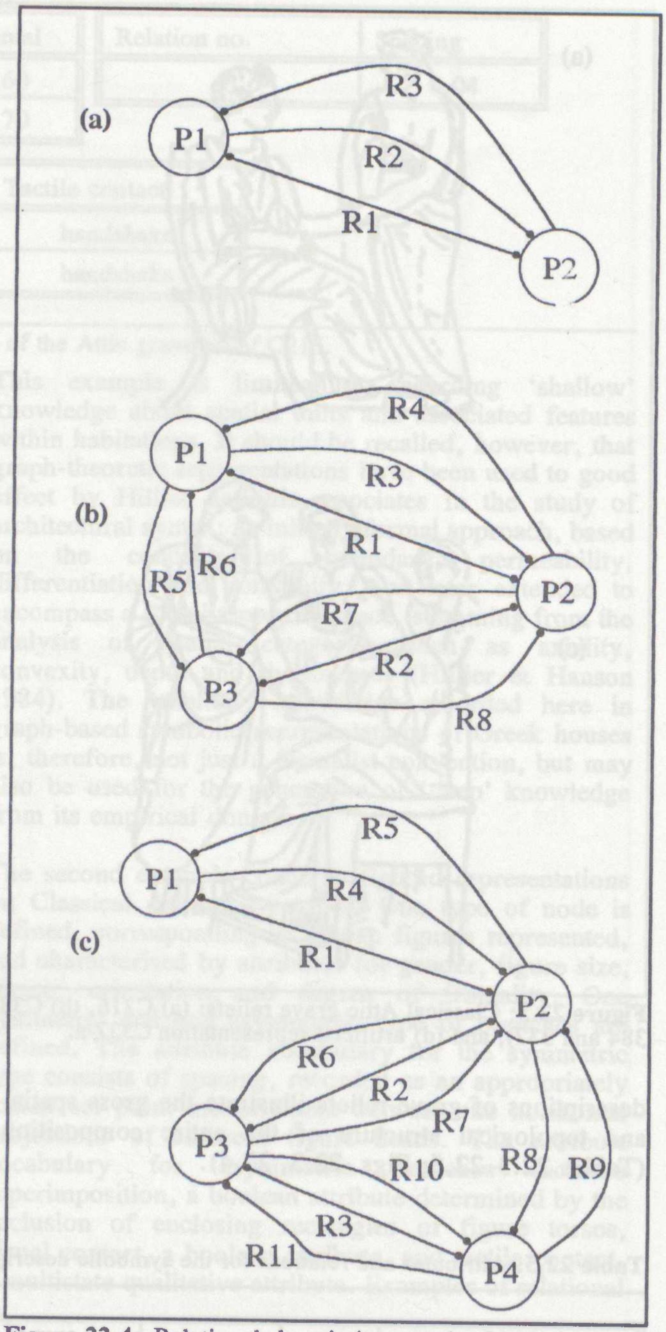

Figure 22.4: Relational description graphs for Attic grave reliefs: (a) C216, (b) C384, (c) C337.

In this paper, I suggest that the classification of complex archaeological entities can be achieved by the use of a generalised *relational similarity coefficient*, derived by means of inexact graph matching. The relational similarity coefficient described here is a generalisation for multi-part relational complex entities of a measure of similarity between units. On account of its generality and its popularity in archaeological applications, the Gower similarity coefficient (Doran & Hodson 1975:142–143; Gordon 1981:23) is used here to illustrate the procedure; however, other measures, such as the simple matching or Jaccard coefficients, could be equally applied. Modifications to Eshera and Fu's distance measure ensure that the coefficient falls

C216 component	C384 component	Similarity
P1	P3	0.73
P2	P1	0.75
-	P2	0.00
R1	-	0.00
R2	R6	1.00
R3	R5	1.00
Generalised similarity coefficient		1.00

Table 22.6: Matching configuration and maximum similarity between the relational descriptions of grave reliefs C216 and C384.

Fu's distance measure ensure that the coefficient falls always between 1, for complete identity, and 0, for complete dissimilarity.

The inexact graph matching configuration between components (nodes and branches) of two attributed relational graphs such that the overall similarity measure between the ARG's possesses a global maximum can be found by calculating the cost of error-transformation between the graphs, and is based on an algorithm of polynomial complexity (Eshera & Fu 1986:611; Conti & Rabitti 1990:307–309). The algorithm involves the extraction and matching of basic subgraphs from the two ARG's, and embedding pairs of matching nodes or branches to already matched subgraphs; a formal description can be found in Eshera and Fu (1984). In calculating the cost of error-transformation, the weight for node and branch insertion and deletion is 1; the weight for node and branch substitution is the distance measure between respective pairs of node or branch attributes.

The proposed similarity measure for archaeological entities is defined as the complement of the appropriately scaled cost of error-transformation between graphs representing two complex entities. To calculate the relational similarity coefficient between two complex entities, the Gower similarity coefficients of matched pairs of nodes, representing object parts, are summed together with the similarity coefficients of matched pairs of branches, representing relationships between parts, and the result is divided by the sum of the maximum number of nodes plus the maximum number of branches.

A formal definition of the relational similarity coefficient follows. Let us suppose that two complex entities are symbolically represented by attributed relational graphs

$C_1 = (P_1, R_1, A_1, E_1)$ and
$C_2 = (P_2, R_2, A_2, E_2)$

where

$P_1 = \{P_{11}, P_{12}, ..., P_{1p}\}$ the set of C_1 nodes,
$P_2 = \{P_{21}, P_{22}, ..., P_{2p}, ..., P_{2q}\}$ the set of C_2 nodes,
$R_1 = \{R_{11}, R_{12}, ..., R_{1k}\}$ the set of C_1 branches, and
$R_2 = \{R_{21}, R_{22}, ..., R_{2k}, ..., R_{2l}\}$ the set of C_2 branches

such that the best inexact matching between C_1 and C_2 is

$M = \{P_{11}P_{21}, P_{12}P_{22}, ... P_{1p}P_{2p}, R_{11}R_{21}, R_{12}R_{22} ...R_{1k}R_{2k}\}$.

The relational similarity coefficient RSC_{12} between C_1 and C_2 is defined as

$$RSC_{12} = \left(\sum_{i=1}^{p} G_i + \sum_{j=1}^{k} G_j \right) / (q + l)$$

where Gi is the Gower similarity coefficient for the best matched pair of nodes $P_{1i}P_{2i}$ and G_j that for the pair of

Table 22.7: Matching configuration and maximum similarity between the relational descriptions of grave reliefs C216 and C337.

C216 component	C337 component	Similarity
P1	P4	0.53
P2	P2	0.77
-	P1	0.00
-	P3	0.00
R1	-	0.00
R2	R9	1.00
R3	R8	0.67
Generalised similarity coefficient		0.42

C384 component	C337 component	Similarity
P1	P2	0.98
P2	P3	0.91
P3	P4	0.79
-	P1	0.00
R1	R2	0.72
R2	R3	0.76
R3	R6	1.00
R4	R7	1.00
R5	R8	0.67
R6	R9	1.00
R7	R10	1.00
R8	R11	0.67
Overall similarity		0.79

Table 22.8: Matching configuration and maximum similarity between the relational descriptions of grave reliefs C384 and C337.

branches $R_{1i}R_{2i}$, computed in the usual manner (Gordon 1981:23).

The Classical Attic grave-relief images represented above as attributed relational graphs may be used to illustrate the use of the relational similarity coefficient. The similarity between all three pairs of examples given above is computed. For each pair of images, the best inexact matching between pairs of image parts and relations is first found; the relational similarity coefficient is then calculated according to the formula given above (Tables 22.6, 22.7 and 22.8). Of the three pairs, the highest similarity coefficient, 0.63, is yielded by C337 and C384; the two images share a similar compositional structure, except for the addition of the background girl in the left side of C337, the gender of the standing figure in the right, and details of figure size and degree of frontality. The images of C216 and C384 were assigned a moderately low similarity coefficient of 0.32; their similarity lies in the configuration of seated female in handshake with standing male, who, however, are in reversed positions in the two images, and, in C384, they are separated by an intermediate background figure. Finally, C216 and C337 are characterised by a low similarity coefficient of 0.20, corresponding to their important differences in directionality, exact form of contact and number of figures.

It should be noted that the proposed formula for the derivation of the relational similarity coefficient takes into account both relationships between existing matched pairs of object parts, and also relationships of parts of one object that have no counterpart in the other object. An alternative approach is provided by the modified formula:

$$RSC\text{mod}_{12} = \left(\sum_{i=1}^{p} G_i + \sum_{j=1}^{k} G_j \right) / (q + k)$$

whereby the Gower similarity coefficients of matched pairs of nodes are summed together with the similarity coefficients of matched pairs of branches and the result is divided by the sum of the maximum number of nodes plus the *minimum* number of branches.

A small experiment was conducted in order to examine the practical differences between the two formulae. Apart from the three figured representations, C216, C337 and C384, an artificially constructed representation, C337-a, derived from C337 through the deletion of the leftmost figure, was also used (Fig. 22.3d). The similarity between the four images was calculated according to both the original and the modified version of the relational similarity coefficient (Table 22.9). While the original relational similarity coefficients are in general lower than their modified counterparts, the similarity matrices produced by the two formulae are monotonic, so far as the three Attic stelae C216, C384 and C337 are concerned.

An interesting difference is, however, apparent in the similarity coefficients between the stelae C384 and C337 and the artificial representation C337-a (Figs. 22.3b, 22.3c and 22.3d). According to the original formula, C337-a was found to be very similar to C384 (*RSC* = 0.90), with which it shares the same compositional structure (seated and standing foreground figures shaking hands, background figure in the middle) despite the differences in gender and other traits of the individual figures. The similarity between C337-a and C337, identical representations in all respects apart from the omission of the leftmost background figure in the former, was found to be only 0.73. The original formula accounts more for the overall structural similarity between complex entities, not privileging similarities between individual parts.

The modified formula, on the other hand, produced for C337-a a slightly higher similarity coefficient with C337 (0.92) than with C384 (0.90). The fact that C337-a and C384 have the same number of figures, arranged in the same structural positions, is given less importance by the modified formula than the similarity between individual image parts, characterising C337 and its subset C337-a.

	C216	C384	C337	C337-a
C216	1.00 (1.00)			
C384	0.32 (0.58)	1.00 (1.00)		
C337	0.20 (0.42)	0.63 (0.79)	1.00 (1.00)	
C337-a	0.27 (0.50)	0.90 (0.90)	0.72 (0.92)	1.00 (1.00)

Table 22.9: Similarity matrix of images C216, C384, C337 and C337-a, according to original and modified versions of the relational similarity coefficient.

While the effect produced by the modified relational similarity coefficient may be desirable in some circumstances, it should be noted that the effect is related to implicitly privileging object part similarity over similarity of overall structure. However, if it is desired that a specific aspect of complex entities should be given more emphasis in the calculation of similarity, this is best achieved explicitly at the data definition stage, by creating graph structures that reflect well the perceived importance of different aspects of the data. Otherwise, important decisions of data definition will be hidden within the mechanics of similarity coefficient calculation. For this reason, it is recommended that the original formula of calculating the relational similarity coefficient should be preferred for general use.

22.5 Discussion

The relational similarity coefficient presented here for use with complex archaeological entities is also a city-block metric, and may be seen as a straight forward conceptual generalisation of the popular Gower similarity coefficient. In a simplified manner, it may be described as the maximal Gower similarity coefficient that can be computed between two complex entities after all possible pairings of their nodes have been tried. As a city-block metric, the relational similarity coefficient is amenable to a wide range of classification and ordination analyses operating on a triangular similarity matrix. Thus, it may provide an effective means of classifying complex entities according to a variety of cluster analysis techniques, or enable the placement of entities in a low-dimensional space by means of principal coordinates analysis or non-metric scaling (Gordon 1981).

There is, however, a more interesting dimension to the procedure outlined above. The best inexact matching between components of complex entities does not simply allow the extraction of a numerical measure of similarity, but also constitutes a definition of their internal structure. Each class in a classification of complex archaeological objects, derived from this relational similarity coefficient, will include sets of object parts or relations, comprising the best matched pairs of components. The centroids of these sets, generalised descriptions for typical entity components, are also combined into attributed relational graphs. These graphs represent a model for each class, i.e., its structural description, in the same manner as a polythetic set is used as a class summary in rectangular data matrix classification.

Seen from the perspective of class models, inexact graph matching is equivalent to a parsing or interpretation of a given relational description. Unlike global attribute list representations, which impose a specific rôle to object parts prior to their description and analysis, and data analyses based directly on object parts, which ignore the importance of configuration and interdependence, relational descriptions allow the direct derivation of the structural position of parts from their intrinsic attributes and the relations in which they participate. They are, therefore, appropriate for a wide range of archaeological problems, such as the identification of the deceased in Attic funerary reliefs, and the determination of room function in Classical house plans.

In the present paper, an attempt has been made to propose an alternative model for complex archaeological data description, similarity and classification. It was shown that quantitative analyses based on a rectangular data matrix are not well-suited to the analysis of complex archaeological objects, since they either ignore, or preclude their internal structure. A relational description approach is proposed, based on attributed relational graphs, whose nodes represent object parts and branches represent the spatial, topological or semantic relations between parts. Inexact graph matching, a technique used in image analysis, is advanced as a means for finding the best correspondence between components of complex entities. A relational similarity coefficient for complex entities is defined as a generalisation of the Gower similarity coefficient, by computing the cost of error-transformation between the attributed relational graph representations of the entities. The approach appears promising, leading not only to the derivation of a quantitative summary of similarity, but also to the possible extraction of structural information about the data. It should be noted, however, that despite the method's theoretical merits, little practical experience has been accumulated with both inexact graph matching and the relational similarity coefficient introduced above. Further research will be necessary, to allow the comparative examination of the procedure outlined here against data matrix-based approaches, using a suitable body of archaeological data.

Acknowledgments

I wish to thank P. Constantopoulos and P. Conti for useful comments on inexact graph matching; errors in this paper remain, naturally, my own.

References

CHAMONARD, J. 1922. *Le Quartier du Théatre*, Délos VII, Vol. 1.

CHARNIAK, E. & D. MCDERMOTT 1985. *Introduction to Artificial Intelligence*, Reading (MA), Addison-Wesley.

CIOLEK-TORRELLO, R. 1984. "An alternative model of room function from Grasshopper Pueblo, Arizona", in H.J. Hietala (ed.), *Intrasite Spatial Analysis in Archaeology*, Cambridge, Cambridge University Press: 127–153.

CLARKE, D.L. 1978. *Analytical Archaeology*, 2nd Edition, revised by B. Chapman, London, Methuen.

CONKEY, M.W. 1989. "Structural analysis of Palaeolithic art", in C.C. Lamberg-Karlovsky (ed.), *Archaeological Thought in America*, Cambridge, Cambridge University Press: 135–154.

CONTI, P. & F. RABITTI 1990. "Image retrieval by semantic content", in C. Thanos (ed.), *Multimedia Office Filing: the MULTOS approach*, Amsterdam, Elsevier North-Holland: 331–349.

CONZE, A. 1893. *Die attischen Grabreliefs*, Berlin, Walter de Gruyter.

DALLAS, C.J. 1987. "The significance of costume on Classical Attic grave *stelae*: a statistical analysis", unpublished D.Phil. thesis, University of Oxford.

DALLAS, C.J. forthcoming. "Syntax and semantics of figurative art: a formal approach", in P. Reilly & S.P.Q. Rahtz (eds.), *Archaeology in the Information Age*, London, Routledge.

DORAN, J.E. 1971. "Computer analysis of data from the La Tène cemetery at Münsingen-Rain", in F.R. Hodson, D.G. Kendall & P. Tautu (eds.), *Mathematics in the Archaeological and Historical Sciences*, Edinburgh, Edinburgh University Press: 422–431.

DORAN, J.E. 1986. "Formal methods and archaeological theory: a perspective", *World Archaeology*, **18**(1): 21–37.

DORAN, J.E. & F.R. HODSON 1975. *Mathematics and Computers in Archaeology*, Edinburgh, Edinburgh University Press.

EISNER, M. 1988. "Arbor: un approccio sistematico ai reperti di struttura complessa", in *Archeologia e Informatica*, Rome: 21–26.

ESHERA, M.A. & K-S. FU 1984. "A graph distance measure for image analysis", *IEEE Trans. Syst., Man, Cybern.*, 14(3): 398–408.

ESHERA, M.A. & K-S. FU 1986. "An image understanding system using attributed symbolic representation and inexact graph-matching", *IEEE Trans. Pattern Anal. Machine Intel.*, 8(5): 604–618.

GARDIN, J-C. 1978. *Code pour l'Analyse des Ornements*, Paris, CNRS.

GARDIN, J-C. 1980. *Archaeological Constructs: an aspect of Theoretical Archaeology*, Cambridge, Cambridge University Press.

GINOUVÈS, R. & A-M. GUIMIER-SORBETS 1978. *La Constitution des Données en Archéologie Classique*, Paris, CNRS.

GLASSIE, H. 1975. *Folk Housing in Middle Virginia: a Structural Analysis of Historical Artifacts*, Knoxville, University of Tennessee Press.

GORDON, A.D. 1981. *Classification. Methods for the Exploratory Analysis of Multivariate Data*, London, Chapman & Hall.

GUIMIER-SORBETS, A.M. 1990. *Les Bases de Données en Archéologie: Conception et Mise en Oeuvre*, Paris, CNRS.

HARDIN, M.A. 1983. "The structure of Tarascan pottery painting", in D.K. Washburn (ed.) *Structure and Cognition in Art*, Cambridge, Cambridge University Press: 8–24.

HILLIER, B. & J. HANSON 1984. *The Social Logic of Space*, Cambridge, Cambridge University Press.

HODDER, I. 1982. "Sequences of structural change in the Dutch neolithic", in I. Hodder (ed.), *Symbolic and Structural Archaeology*, Cambridge, Cambridge University Press: 162–177.

LAGRANGE, M-S. & M. RENAUD 1983. "L'interpretation des documents figurés en archéologie et histoire de l'art: essai de simulation sur ordinateur", in F. Lissarague & F. Thélamon (eds.), *Image et Ceramique Grecque*, Rouen, Université de Rouen: 43–66.

LAGRANGE, M-S. & M. RENAUD 1987. "Superikon, essai de cumul de six expertises en iconographie", in Gardin, J-C. *et al.* (eds.), *Systèmes Experts et Sciences Humaines: le cas de l'Archéologie*, Paris, Eyrolles.

LAWRENCE, A.W. 1973. *Greek Architecture*, Harmondsworth, Penguin.

LEVESQUE, H.J. 1986. "Knowledge representation and reasoning", *Annual Review of Computer Science*, 1: 225–287.

MOHAN, L. & R.L. KASHYAP 1988. "An object-oriented knowledge representation for spatial information", *IEEE Trans. Soft. Eng.*, 14(5): 675–681.

ORTON, C. 1980. *Mathematics in Archaeology*, Cambridge, Cambridge University Press.

PADER, E-J. 1982. *Symbolism, Social Relations and the Interpretation of Mortuary Remains*, British Archaeological Reports (International Series) **130**, Oxford, British Archaeological Reports.

ROSENFELD, A. & A.C. KAK 1982. *Digital Picture Processing*, Vol. 2, New York, Academic Press.

SALOMÉ, M-R. 1980 *Code pour l'Analyse des Représentations Figurées sur les Vases Grecs*, Paris, CNRS.

SHENNAN, S. 1988. *Quantifying Archaeology*, Edinburgh, Edinburgh University Press.

WASHBURN, D.K. 1983. "Toward a theory of structural style in art", in D.K. Washburn (ed.) *Structure and Cognition in Art*,Cambridge, Cambridge University Press: 1–7.

WHALLON, R. 1972. "A new approach to pottery typology", *American Antiquity*, 37: 13–33.