

Optimal Choice of Prototypes for Ceramic Typology

Uzy Smilansky¹, Itzhak Beit-Arieh⁴, Avshalom Karasik^{1,2}, Ilan Sharon² and Ayelet Gilboa³

¹Department of Physics of Complex Systems, The Weizmann Institute of Science, Rehovot, Israel.
{avshalom.karasik, uzy.smilansky}@weizmann.ac.il

²The Institute of Archaeology, The Hebrew University, Jerusalem, Israel.
Sharon@huji.ac.il

³The Zinman Institute of Archaeology, University of Haifa, Mount Carmel, Haifa, Israel.
gilboaaa@zahav.net.il

⁴Department of Archaeology, Tel-Aviv University, Tel-Aviv, Israel.
betarieh@post.tau.ac.il

Abstract. A crucial step in any typological analysis is the determination of the prototypes according to which the assemblage is to be classified. Two conflicting requirements dictate this choice: the number of prototypes should be minimal to allow an efficient classification. At the same time, the set of prototypes should be comprehensive so that the essential variability of the original assemblage is reproduced by the prototypes. This problem is especially complex when the assemblage consists of ceramic vessels of the same genre such as e.g., storage jars, cooking pots or drinking cups. Here, we present a computerized method to identify an optimal set of prototypes, which is based on the analysis of pottery profiles considered as planar curves. The profiles are clustered according to their correlations, and the correlation tree yields a preliminary set of types, whose number is much smaller than the number of profiles in the original assemblage, and which is based on the dominant but distinct features in each of the branches. The next step in the process is to find the optimal subset of types, which satisfies the conflicting requirements mentioned above. The optimal set of prototypes is the one which minimizes the number of types without affecting the quality of the description. The method will be illustrated by showing its application to an assemblage of a few hundreds of Early Bronze age holemouth vessels from Tel Arad (Southern Israel) and the Sinai peninsula (Egypt).

Keywords: pottery, typology, curvature analysis

1. Introduction

Archaeologists often encounter situations where an assemblage of objects is classified according to a set of predetermined types, so that the archaeological information carried by the assemblage could be deciphered. In addition, in sites reports of large-scale excavations, where extensive amount of ceramics are encountered, the construction of a typology is necessary in order to present the ceramic assemblage faithfully without illustrating every shard. Often, the typology is only determined by examining the assemblage at hand, without reference to any external, a-priori constructed typology. The problem addressed here is how to make an optimal choice of prototypes, for further typological analysis. Two conflicting requirements dictate this choice: the number of prototypes should be minimal to allow an efficient classification. At the same time, the set of prototypes should be comprehensive so that the essential variability of the original assemblage is reproduced by the prototypes. We concentrate in particular on the morphological analysis of ceramics, and use the information stored in the profiles of their cross-sections. The case which is analyzed here is a typical example: Rim shards of holemouth vessels were excavated in two sites: Tel Arad, a major Early Bronze Age town in Southern Israel, and Early Bronze age rural settlements in the Sinai peninsula (Egypt). The sites are approximately 300km apart. The pots from the two sites look quite similar, and this led to the suggestion that commercial

ties existed between them, in spite of their distance. Petrographical evidence supported this hypothesis, and the emerging picture is that a significant fraction of the vessels excavated in Tel Arad were produced from clays which probably originate from Sinai (Amiran, Beit-Arieh et al. 1973; Amiran, Paran et al. 1978; Porat 1989; Beit-Arieh 2003). Our aim here is to perform a detailed typological analysis based on published line drawings of Arad and Sinai excavation reports, and to assess independently to what extent the petrographic and the morphological classifications are compatible.

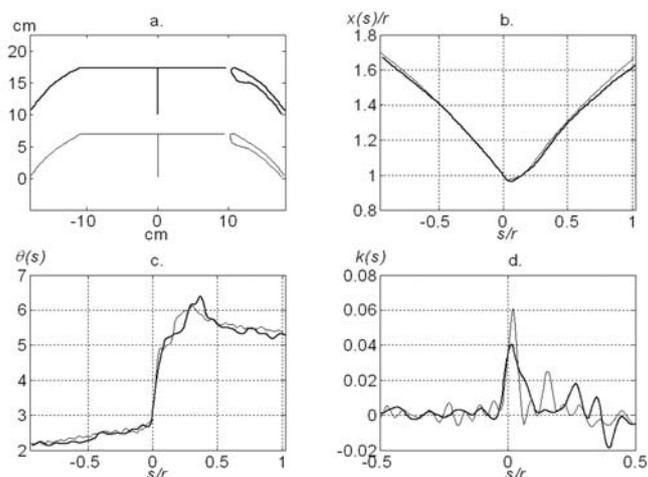


Fig. 1. Two profiles (a) and their Cartesian (b), tangent (c) and curvature (d) representations.

2. Method

The information about the assemblage is provided by digitized profiles obtained by scanning traditional line drawings, produced manually (for the shortcomings of such a drawings see (Gilboa, Karasik et al. 2004)). Two profiles of rim shards from the combined Arad – Sinai assemblage can be seen in Fig. 1a. Our subsequent analysis proceeds by treating the profiles as curves in the plane which should be sorted out into similarity classes. Similarity, in turn, can be defined in various ways. Here we use three distinct, mathematically legitimate ways, which differ in the aspects of the compared profiles they emphasize.

2.1 The Description of Profiles as Planar Curves

The mathematical description of a curve proceeds as follows: Let s be the arc length measured along the profile, and L the total length of the profile. In the present case it is convenient to set the point $s = 0$ at the top of the rim. By increasing s , the profile is traversed in the positive mathematical sense. The contour is completely specified when any of the following representation functions is known:

$x(s)$ – the Cartesian distance of the point s on the profile from the axis of rotation of the vessel (Wilcock and Shennan 1975; Wilcock and Shennan 1975).

$\theta(s)$ – the angle of the tangent at the point s on the profile measured relative to the x-axis (Leese and Main 1983; Main 1986).

$\kappa(s)$ – the curvature at the point s on the profile (Leymarie and Levine 1988; Mokhtarian and Bober 2003).

Two rather similar rims, described by their three functions are shown in Figure [1,b–d]. The function $x(s)$ is the smoothest, as it follows the overall outline of the profile. $\theta(s)$ and $\kappa(s)$ are obtained from $x(s)$ by taking successive derivatives, and therefore they are progressively less smooth. The position of the rim is accentuated as a steep jump in $\theta(s)$, and a sharp spike in $\kappa(s)$. $\theta(s)$ is intermediate between the two extreme functions, but it certainly gives more weight to the inflections in the rim area than . It is clear that in the present example, the best resolution between the profiles is achieved in the curvature representation. In order to classify and compare the profiles, we standardize the description of the assemblage by measuring distances along the profile in units of its rim radius r which can be extracted from the drawings. The lengths of the rim fragments are not the same, and in order to compare, we truncate them at prescribed distances from the rim (Gilboa, Karasik et al. 2004; Karasik, Sharon et al. 2004).

We can now measure exactly the similarity or dissimilarity between two profiles i and j . This is achieved by defining a distance (metric) $d(i,j)$ between the profiles. The definition of a distance depends on representation, and the choice of representation will enhance different features as explained above. Explicitly,

$$d_x(i,j) = \sqrt{\int_0^1 (x_i(s) - x_j(s))^2 ds}, \quad d_\theta(i,j) = \sqrt{\int_0^1 (\theta_i(s) - \theta_j(s))^2 ds},$$

$$d_\kappa(i,j) = \sqrt{\int_0^1 (\kappa_i(s) - \kappa_j(s))^2 ds}. \quad (1)$$

The suffixes x , θ and κ distinguish between the different representations. Another useful definition is the scalar product: $\langle ij \rangle_x = \int x_i(s)x_j(s)ds$, (and similarly for $\langle ij \rangle_\theta$ and $\langle ij \rangle_\kappa$). The profiles correlation is

$$C_{i,j} \equiv \langle ij \rangle / \sqrt{\langle i|i \rangle \langle j|j \rangle}$$

which again can be computed in three different ways, depending on the function used to define the metric. If, alternately i and j are run vs. all the profiles in the assemblage we obtain a correlation matrix whose entries vary in the range $-1 \leq C_{i,j} \leq 1$. $C_{i,j}$ attains its maximal value when the profiles are identical, and it stores the entire information needed for the subsequent typology.

2.2. Prototypes from the Correlation Matrix

The present subsection implements the ideas described above regarding the Arad – Sinai assemblage. The analysis is conducted in terms of the curvature representation. A cluster tree (see Fig. 2.), obtained from the correlation matrix, is used for a preliminary sorting of the profiles (represented by circles).

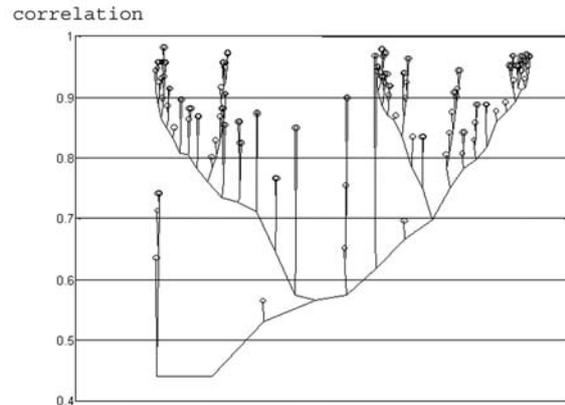


Fig. 2. Correlation tree based on curvature analysis. Each circle represents a real profile.

The lines indicate affinities and hierarchical order. The tree is generated recursively. In the first step, the $M \times M$ correlation matrix is computed ($M=216$ in the present case), and the pair with the highest correlation, say (i,j) , is identified. Two vertices denoted by circles are drawn at the height $y = C_{i,j}$. The horizontal distance between the points is arbitrarily fixed to a value Δx , and the pair is positioned at the middle of the x interval. The mean profile, defined as the mean of the two representing functions, is computed, and added to the list of profiles, while the two original profiles are eliminated. Thus a virtual profile replaces its two real ‘parents’, in the correlation tree this virtual profile is indicated by the junction point between the two lines. This ends the first step, and now the effective assemblage consists of only $M-1$ profiles. In the next step, the correlation matrix is computed, and the pair with the $(M-1) \times (M-1)$ highest correlation, say (k,l) , is identified. If k and l stand for real profiles, the procedure described previously is repeated, and the pair of vertices is drawn at a horizontal position x' which is not yet occupied by a higher pair in the tree. If one of the profiles is virtual, or both are, the

junction point of their ‘parents’ is drawn at height $y = C_{k,l}$. Finally, the vertex is connected to the ‘parents’ vertices by lines. The other member of the pair is drawn in a distance Δx away. Again, the ‘parent’ pair is replaced by their average, and the number of effective profiles is reduced by 1. This procedure is repeated until the assemblage is exhausted when it consists of 2 profiles.

Branches are the collection of real profiles which emerge from a common root. The correlations between profiles in a branch are higher than the correlation value of the root. The mean profile of each branch defines a “type” which we call a “branch profile” and the group of all branch profiles (Fig. 3) displays the variability of shapes in the assemblage.

In the present example we considered only branches with more than 3 real profiles and with root correlation larger than 0.8. We emphasize once again that the branch profiles are virtual in the sense that they do not stand for real objects. However, they are very useful since they represent the typological variety of the assemblage in a concise way. Our typological analysis is based on the representation of the profiles as vectors in an abstract linear space (see Gilboa, Karasik et al. 2004 for details). The prototypes in this sense are the *minimal* set of independent vectors which spans with *maximal fidelity* the profiles in the assemblage.

This minimal set is constructed by identifying the linearly independent vectors which span the set of branch profiles. In practice this is done by diagonalizing the branch profiles correlation matrix, and keeping the eigenvectors with non vanishing eigenvalues. In the present example, the eigenvalues are 0.0060, 0.0291, 0.1272, 0.2206, 4.6172. It is clear that the profiles generated from the eigenvectors corresponding to the two largest eigenvalues suffice for the representation of the entire assemblage. They are shown in

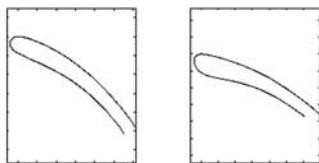
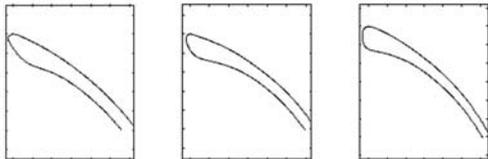


Fig.3. The ‘branch profiles’ from the branches of the correlation tree in Fig. 2.

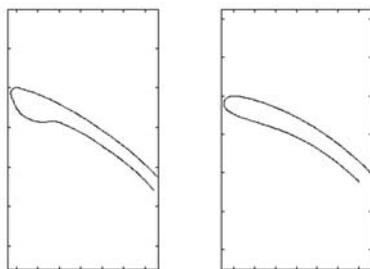


Fig. 4. The prototype profiles.

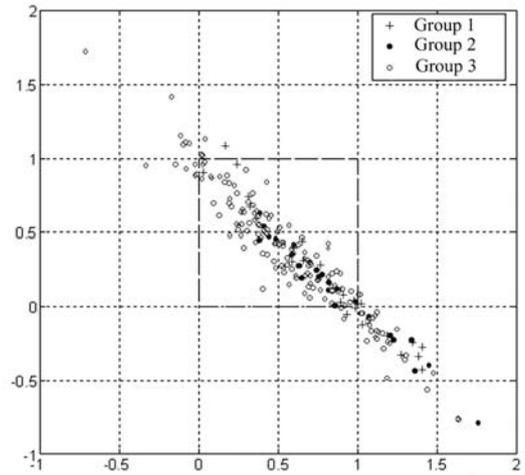


Fig. 5. The distribution of the profiles in the prototypes space.

Fig 4. Once the prototypes are determined, the classification proceeds by using the method proposed in (Gilboa, Karasik et al. 2004): Each profile is expressed as a linear combination of the prototype profiles, and the resulting distribution is shown in Fig. 5.

The points concentrate within a narrow strip along a line. This reflects the fact that the second eigenvalue is approximately 20 times smaller than the first. The vertical projection of a point onto the line determines its distance from each of the two prototypes which are indicated by the points (1,0) and (0,1). This distance, denoted below μ_x , provides the entire typological information in terms of a single parameter. This procedure entails a number of approximations (such as, e.g., the elimination of the eigenvectors with small, yet non vanishing eigenvalues from the prototype basis).

The quality of the analysis can be evaluated by computing the extent by which each of the profiles deviates from the space spanned by the prototypes. This parameter was defined in (Gilboa, Karasik et al. 2004) and its distribution for the Arad – Sinai assemblage is shown in Fig. 6. Its mean is approximately 0.17, which ensures a high reliability. With this check the task of defining the optimal set of prototypes is completed. The archaeological consequences of the analysis of the Arad – Sinai assemblage are discussed below.

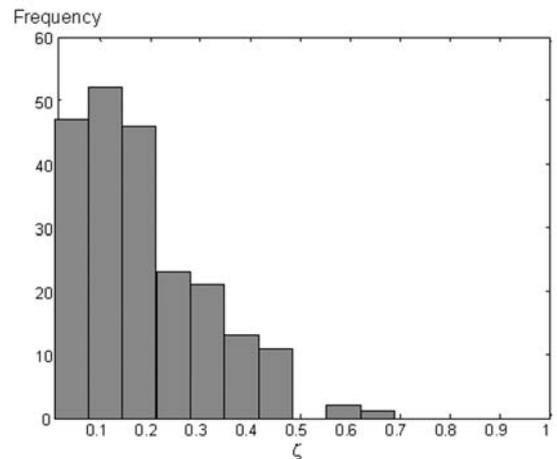


Fig. 6. The distribution of the quality measure ζ (Gilboa, Karasik et al. 2004).

3. The Typology of the Arad – Sinai Assemblage

Based on petrographic analysis the Arad – Sinai holemouth vessels assemblage were divided by the archaeologists into three groups.

- Holemouth cooking pots excavate in the Sinai sites, of Sinai petrographic composition.
- Holemouth cooking pots excavate in Tel Arad, of Sinai petrographic composition.
- Holemouth cooking jars excavate in Tel Arad, of Arad petrographic composition.

Out of the 450 drawings of these vessels in the various archaeological reports, only 216 display sufficiently large part of the profile to warrant reliable analysis. (31 of group 1, 27 of group 2 and 158 of group 3). For each profile the 3 representing functions were extracted, and the correlation matrix, branch types and prototypes were computed for each representation. Subsequently, for each representation the data were sorted employing the method described above. The only difference between the results is that the resolution, and hence the statistical significance, is superior for the curvature representation. The distributions of the μ_x values for the three subgroups of the assemblage are plotted in Fig. 7.

The distributions of group (1) and group (2) are rather similar, and both are shifted with respect to the distribution of group (3). This observation is supported by a t-test analysis which can be summarized as follows: a. groups (1) and (2) belong to the same distribution with probability 66%; and b. the probability that either group (1) or group (2) are similar to group (3) is resp. 0.73% and 0.28%. The significant agreement between the typological and the petrographical classification enhances the credibility of the original surprising and important assertion that active commerce existed between Arad and the Southern Sinai region.

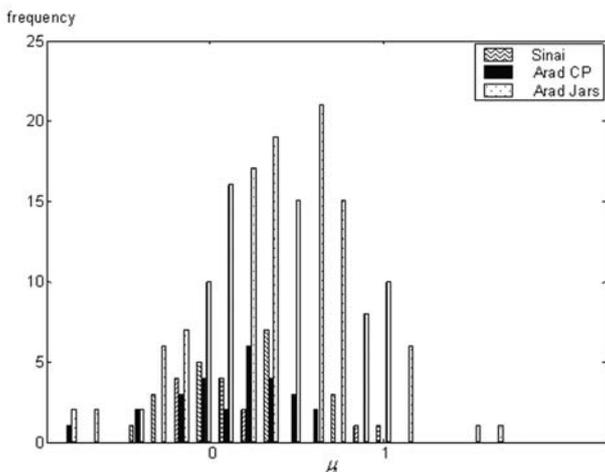


Fig. 7. The distribution of the interpolation parameter μ_x for the three groups of the Arad – Sinai assemblage.

Acknowledgements

This research was supported by a grant from the BIKURA program of the Israel Science Foundation and by the Kimmel Center for Archaeological Sciences at the Weizmann Institute. We are indebted to Dr U. Avner for sharing with us his ideas.

References

- Amiran, R., Beit-Arieh, I. et al., 1973. The Interrelationship Between Arad and Sites in Southern Sinai in the Early Bronze Age II (preliminary report) *Israel Exploration Journal* 23, 193–197.
- Amiran, R., Paran, U. et al., 1978. *Early Arad I: The Chalcolithic Settlement and Early Bronze City, first-fifth seasons of Excavations, 1962–1966*. Jerusalem, Israel Exploration Society.
- Beit-Arieh, I., 2003. *Archaeology of Sinai: The Ophir Expedition*. Tel-Aviv, Emery and Claire Yass Publications in Archaeology.
- Gilboa, A., Karasik, A., et al., 2004. Towards computerized typology and classification of ceramics. *Journal of Archaeological Science* 31(6), 681–694.
- Karasik, A., Sharon, I., et al., 2004. Typology and classification of ceramics based on curvature analysis. *Computers Applications and Quantitative Methods in Archaeology – CAA 2003*. Vienna.
- Leese, M. N. and Main, P. L., 1983. An approach to the assessment of artefact dimension as descriptors of shape. *Computer Applications in Archaeology 1983*, Bradford University of Bradford, School of Archaeological sciences.
- Leymarie, F. and Levine, M. D., 1988. *Curvature morphology. Computer Vision and Robotics Laboratory*. Montreal, Quebec, McGill University, Canada.
- Main, P., 1986. Accessing outline shape information efficiently within a large database. *Computer Applications in Archaeology 1986*, Birmingham, Computer Center, University of Birmingham.
- Mokhtarian, F. and Bober, M., 2003. *Curvature Scale Space representation, and MPEG-7 standardization*. Dordrecht, The Netherlands, Kluwer academic Publishers.
- Porat, N., 1989. *Petrography of Pottery from Southern Israel and Sinai. L'urbanisation de la Palestine a L'age du Bronze Ancien*. P. de Miroschedji. Oxford. I, 169–188.
- Wilcock, J. D. and Shennan, S. J., 1975. Computer Analysis of Pottery shapes. *Computer Applications in Archaeology, Birmingham*.
- Wilcock, J. D. and Shennan, S. J., 1975. Shape and style variation in Central German Bell Beakers. a computer assisted study. *Science and Archaeology* 15, 17–31.