METRON
MEASURING THE AEGEAN BRONZE AGE

Proceedings of the 9th International Aegean Conference/
9e Rencontre égéenne internationale
New Haven, Yale University, 18-21 April 2002

Edited by Karen Polinger FOSTER and Robert LAFFINEUR
LINEAR A AND MULTIDIMENSIONAL SCALING*

In the past, Linear A studies have suffered from a lack of archaeological and analytical context. The majority of Linear A documents come from the site of Haghia Triada, Crete, where the find-spots of most tablets and sealings are unclear, despite the heroic efforts of scholars attempting to reconstruct the Linear A administration there. Moreover, the larger context of the site is not well defined, since it does not neatly fit our typologies of Minoan sites. Analytically, Linear A is not well contextualized because studies of the script tend to focus on decipherment in an unsystematic way.

A number of scholars have attempted to decipher Linear A, identifying it with known languages such as Semitic, Luwian, and even Greek. These studies begin by attempting to etymologize a small number of individual words, largely ignoring overall context. No attempt at decipherment has yet met with acceptance. As Emmett Bennett has rightly noted, studies of Linear A which focus on decipherment are often flawed in that they begin by examining only a few well-known texts to make a series of conjectures about Linear A; these conjectures are then used to support further conjectures without any external confirmation. Bennett has characterized such studies as “a single column of dominoes, each resting on the one below, or it is like an inverted obelisk, four columns of dominoes, all resting on three which rest on two which rest on one domino at the base.”

We argue that only the total context of words and tablets can enable real progress in Linear A studies. Our only supposition is that the lexical items on any given tablet are related to each other in some way which is connected to the function of the tablet as an administrative document. Thus, we avoid the unproveable assumptions so common in studies which aim for decipherment, while providing a fuller context and understanding of individual words and tablets. To this end, we are applying statistical techniques to model the assemblage of words and tablets in a systematic and comprehensive way.

It remains the case that the language represented by the Linear A script is undeciphered. So when approaching the Linear A texts, one is confronted by a corpus of unintelligible words. In order to make significant progress in interpreting these texts, we must first learn how to identify whether any given word is a place name, personal name, divine name, commodity, or some other lexical type. Prior to the decipherment of Linear B, for example, the Linear B words for “boy” (ko-wo) and “girl” (ko-wa) had already been deduced based on contextual clues.

---

* This paper is the combined result of our research for a seminar on Linear A organized by Thomas G. Palaima in the spring of 2001 at the University of Texas at Austin, Department of Classics, Program in Aegean Scripts and Prehistory. We would like to thank all the members of that seminar, and in particular Tom Palaima for his help, guidance and encouragement.

---


5. BENNETT (supra n. 4), 53.

In contrast, studies which attempt to categorize Linear A words frequently begin with *a priori* assumptions about the nature of the texts. Often it is asserted that a list of words each followed by the numeral “1” is a list of personal names, or that a list of words and numbers without ideograms is a personnel list. These assumptions necessarily color any interpretation that is derived from them. However, because of the irregularity of Linear A texts, such rigid readings tend to cause more interpretive problems than they solve. There is no reason to believe that parallel entries are semantically equivalent; some Linear B tablets at Knossos (the Fs series), for example, list personal names and toponyms as parallel entries on the same tablet. Nor have these assumptions allowed us to make convincing readings of other Linear A texts. This is not to say that such interpretations are incorrect, but rather that they cannot be supported by internal examination of the Linear A tablets. Such assumptions have to be avoided for any progress to be achieved.

There are also problems in attempting to assess the context of Linear A words using traditional methods. So far, scholars have focused on a single word on one tablet, then expanding their study to the context of the same word on other tablets. This is too minimal a conception of context. One word is associated with several others on every tablet, potentially in several different contexts. If each tablet is assumed to list lexically identical items, then following the context of a word from one tablet to others would ultimately result in the conclusion that all words in Linear A are of a single lexical type. As Packard has rightly noted, “this method, if followed to its logical conclusion, will allow us to demonstrate that every sign-group is parallel to every other.”

Instead of chasing a single word through the tablets, we must instead look at all words in context together, thereby understanding the relationship of these words to one another. Our approach is an attempt to visualize this web of associations, and to allow the words themselves to define these associations, rather than beginning with *a priori* assumptions. To create a visual representation of similarities and differences among words in Linear A, we have used multidimensional scaling (henceforth MDS), first used in Aegean prehistory by John Cherry to map Linear B toponyms in the Pylos polity. Given a matrix of numbers representing similarities between points, the MDS algorithm converts these numerical data into distances and displays the points spatially in a way that best conforms to those distances. For this analysis, we have used all words that occur more than once in the Hagia Triada texts to generate a web of word interactions.

The first step in this analysis is the creation of two matrices, one for tablets and another for words (see Pl. LXVIA). In the tablet matrix, tablet numbers are listed along the top row, and

---


9 PACKARD (supra n. 7) 64.

10 J.F. CHERRY, “Investigating the political geography of an early state by multidimensional scaling of Linear B tablet data,” *Mycenaean Geography* (1977) 76-82. Cherry’s use of MDS first made us aware of its potentially useful application to Linear A. It should be noted, however, that Cherry’s use of MDS is quite different from our own. Cherry employed MDS to conceptualize relationships between words all of which were already known to indicate places. The topographic function of these words in the Linear B tablets was a starting assumption of the project, and so at least in part, the actual topographical distances of these sites was at issue. In contrast, we are dealing with words that are certainly of several different types. Thus, we are utilizing MDS in order to reveal conceptual relationships between words, e.g., their function(s) in the tablets.

11 To compile occurrences of words, we have used the texts, readings, and indices of the standard publication of Linear A, namely L. GODART and J.-P. OLIVIER (eds), *Recueil des inscriptions en Linéaire A*, 5 vols. (1976-1985). Purely for the sake of clarity, we have transcribed Linear A signs according to their Linear B values.
then again along the first column. In each cell, the number of words in common on the two tablets is noted. For example, there are two words that both appear on HT 102 and HT 105. The MDS program then reads these numbers as similarities, with a greater gravity assigned to higher numbers. Accordingly, in the MDS output graph, tablets with high similarities are represented as closer than tablets with low similarities. The words matrix is laid out in the same format as the tablets matrix. In this matrix, each cell lists the number of tablets on which any two words appear together. The result is one 2-dimensional plot of tablets, and another of words.

A few things should be said about the function of MDS analysis. As noted, the numbers in the cells of the matrix correspond to the degree of similarity between two words or tablets. This can create only the best possible solution, not an absolutely accurate one. The tablet evidence is fragmentary, both in terms of how many tablets we have, and in terms of the preservation of individual tablets. We may be missing several associations that should be made. Nevertheless, the "stress" encountered by the computer software was remarkably low; here, stress refers to the degree of fit between the matrix of inputs and the resulting spatial representation. The algorithm was run 300 times to insure the best possible plot of words and tablets, resulting in a 0.13 stress level for the words, and a 0.08 stress level for the tablets. These low stress levels suggest that our body of data is sufficiently representative of the composition of the original Linear A archive. Finally, such MDS plots are the only way to see, in a systematic and comprehensive way, the complete context of any given tablet or word. We have a sufficient number of texts and words that are not unique that analysis yields interpretable results.

Finally, we should note that several well-known words were actually removed from our matrix, and hence, our analysis. The decision to remove words from our analysis was not made lightly, but was found in the end to be necessary for the analytical value of our method. Our initial matrices included all words, of which ku-ro and ki-ro were generating the most stress. These words have been convincingly interpreted as meaning "total" and "deficit" respectively, and occur quite often in the corpus. It was not difficult to see why they would generate stress: MDS treats all words in the tablets as comparable, and since words such as "total" and "deficit"

---

12 We have followed the standard notational system of GODART and OLIVIER (supra n. 11), in which tablets are cataloged by an abbreviation of the site where they were found and a number. All of our references are to tablets found at Hagia Triada, abbreviated HT.

13 MDS can either treat these similarity values as absolute - in which case the degree of similarity must correspond directly to distance between points; that is, the distance between two points with a similarity of 2 will be twice as long as the distance between two points with a similarity of 4 - or as interval, in which the degree of similarity implies a hierarchy of measurement, but not an absolute distance; that is, a the distance between two points with a similarity of 4 must be smaller than that between two points with a similarity of 2 or 3, but otherwise the exact length is unrestricted. Absolute measurements are generally used only for topographic analysis when real distances are at issue. Otherwise, interval measurement (or a variant of it) is to be preferred.

14 For a useful summary of MDS and its applications, see J.B. KRUSKAL and M. WISH, Multidimensional Scaling (1985).

15 Or to put it another way, MDS will generate a plot of points that is most consistent with the input data.

16 That is to say, a real data set will never exactly conform to the computer's attempt to render it in 2-dimensional form. Stress is therefore an indication of the amount of difficulty the computer has converting the vast array of data into a simple "map." For example, if only 3 data points exist in our matrix, each with a dissimilarity value of 1, the result is a perfect equilateral triangle, with no stress whatsoever. Nevertheless, one can easily imagine a scenario where the dissimilarity values would not be able to form a geometric triangle; for example, one cannot have a triangle in which one side is longer than the sum of the length of the other two sides. In this case, the computer will attempt to find a solution that is most faithful to the data which it is given. The more the computer must compensate for inconsistencies in the data, the higher the resulting stress.

17 As a general rule of thumb, anything under 0.15 is considered an acceptable stress level, although data can be considered useful if the stress level is below 0.4. See KRUSKAL and WISH (supra n. 14) for fuller discussion.
are so common and widely applied, one would not expect their use to correlate with the uses of the other words in any given tablet, or to be restricted in their associations to a single lexical category. 18 Their presence in the matrix was significantly distorting analysis and obscuring more subtle relationships. 19 One other word, a-du, was also removed from our matrices. Like ku-ro and ki-ro, it is a very common word, and therefore exerts a strong influence over the entire analysis. Furthermore, it is clear from its use that it is a word whose function is quite different from other Linear A words. 20 As MDS is most useful when evaluating similarities and differences between words whose uses are roughly equivalent, we felt justified in removing it from our matrices. 21 This also allowed us to evaluate its use and context independently of its influence on the MDS chart.

In the chart of tablets, there are five separate groupings of tablets, arranged roughly in a circle (see Pl. LXVIb). Such a clustering suggests that the tablets of each group share a fairly consistent and distinct vocabulary, and can be treated as meaningful and discrete units. This hypothesis is supported by correlating these tablet groups with their associated ideograms. It should be emphasized that the ideograms were not part of the original data set, yet they show the same consistency in clustering as the tablets. This substantiates our claim that each group of tablets can be treated as a meaningful unit, and shows that there is a consistency between word and ideogram usage in the Linear A corpus.

On the MDS chart, tablets with agricultural staples cluster at the top, tablets with "I"s at the bottom of the chart, wine tablets at the bottom-right of the chart, and tablets with "man" ideograms cluster along the left side of the chart. 22 In these groups we see a means by which we may divide the data into rough quadrants, separated by two main axes (see Pl. LXVIb). 23 These quadrants may also be treated as significant units, since in MDS graphs, every distance is meaningful. So in addition to treating clusters of tablets as discrete units, the tablets on either side of a major axis may be treated as distinct groups. The significance of these axes is not readily apparent; they might separate tablets consisting primarily of personal names from tablets with place names, or religious names and places from secular ones, or even local concerns from palatial concerns.

We are not suggesting that tablets of one cluster list only semantically identical terms, as past studies have done. Rather these graphs allow for the identification of problematic terms, thus avoiding the methodological pitfalls mentioned earlier. This point is well demonstrated by the tablets listing the word da-ri-da. This word occurs three times in cluster II (on HT 10, 85, and 122), and once in cluster I (on HT 93). This distribution suggests that da-ri-da is most closely associated with the vocabulary of cluster II tablets, and that it is uncharacteristic on HT 93. If we look at its function on these four tablets, we see that in the cluster II tablets, da-ri-da is listed twice with a whole number and no ideogram, and once on a tablet recording men.

18 Using a modern analogy, we would not expect the definite article ("the") to significantly correlate with lexical items in a given sample of English texts treated as analytical units.
19 That is to say, ku-ro and ki-ro were exerting a significant "gravitational" force over words which otherwise had little to nothing in common, because they inevitably had high similarity ratings with a great number of words, which MDS attempted to represent as spatially close to ku-ro and ki-ro. By so doing, subtle differences were obliterated by the sheer force of only two words.
20 It is clear that a-du is unique, but its semantic value and function remain obscure, despite many attempts to understand it. Recently, a-du has been identified as the 3rd person plural imperative of the verb "to do" by A. UCHTEL and M. FINKELBERG, "Some Possible Identifications in the Headings of the Linear A Archives," SMEA 36 (1995) 29-36.
21 a-du generated a great deal of stress in our initial simulations, which is what one would expect of a word whose use is unlike that of the other words in our analysis. Including a-du in our analysis forced several groups of words to be proximal to each other, when otherwise they would have remained quite separate. That is to say, the variation within the group of words linked to a-du is greater than the similarity between them.
22 By "agricultural staples" we refer to the ideograms that represent barley, wheat, olives, olive oil and figs.
23 The axes were inserted along natural breaks in the chart, where points tended to cluster on either side of a large gap. These gaps will usually separate points that are in binary opposition to one another.
On HT 93, a tablet recording grain, it is followed by a vase ideogram (406\textsuperscript{va}s). This entry is unique on HT 93. The other entries on this tablet list larger amounts of grain, never with a vase ideogram. From these instances we may infer that on HT 93, da-di-da is, in fact, not semantically parallel to the other words. MDS has therefore allowed for the identification of semantic distinctions within a single tablet, even though we cannot specify the nature of that distinction. This pattern, in which a single lexical item is located chiefly in one tablet group, but is also found once in another, occurs several other times. MDS allows us to differentiate between uses of the same word based on the instances in which it is used uncharacteristically, and so allows us to escape the circular logic which assumes uniformity in the meanings and uses of Linear A words.

Similarly, a long-standing crux in Linear A studies is the status and nature of “headers,” which are lexical items that are not directly associated with the recording of commodities, but rather stand at the top of a list of paired lexical items and amounts of goods.\textsuperscript{24} Headers generally consist of a single word followed by entries; very few consist of two or three word sequences.\textsuperscript{25} We may suppose that headers, based on their use in Linear B, may provide information about location, month, occasion, context, or conditions under which a transaction takes place. However, Linear A headers, in contrast to Linear B headers, are short. Moreover, they defy consistent interpretation, since most also occur as entries. Traditional interpretations generally make unwarranted assumptions about their nature, such as that they are predominantly place-names.

Headers occur on tablets in every class which we have defined. Moreover, headers which occur more than once generally appear in more than one cluster of tablets, indicating that there is a fluid relationship between the semantic value of the header and the accompanying vocabulary on the tablet. For example, a-du, the most prominent and most studied header, occurs on five tablets in four different groups (this is also true when a-du is included in our matrix).\textsuperscript{26} The fact that a-du occurs in four discrete clusters of tablets suggests that it does not exert a significant restrictive force over its accompanying vocabulary.\textsuperscript{27} For example, a-du shows up as a header on HT 85 and 95; these tablets do not share vocabulary and are in different groups in our tablet graph.\textsuperscript{28} Significantly, the vocabulary groups associated with these tablets are also differentiated in the word chart (see Pl. LXVII).

However, these separate vocabulary groups are linked by HT 10, a tablet whose header is a word from HT 95 (kunisu), and whose entries are largely drawn from the vocabulary group of HT 85. With MDS, we can recognize that HT 10 contains words that come from very different vocabulary groups, allowing us to interpret it more fully. The entries on HT 10 are drawn largely from words associated with tablet group II. Most of these tablets have the man ideogram or consist of entries with no ideogram and “1”s or low numbers. The header kunisu, on the other hand, occurs elsewhere on tablets recording large amounts of grain.\textsuperscript{29} It is therefore tempting to interpret kunisu as a toponym and the entries as personal names.\textsuperscript{30}

\textsuperscript{24} A clear definition of Linear A headers based on formal grounds is missing from the literature, despite general agreement on which words are headers, conveniently identified in the publication of GODART-OLIVIER (supra n. 11). Although it is not crucial to our argument, we define ‘header’ as one or more lexical items which is NOT followed by (1) a numeral, (2) an ideogram + numeral, or (3) a transactional item + (1) or (2). Transactional items cannot be defined on formal grounds, only on use. For a discussion of transaction signs, see I. SCHÖEP, “Some notes on Linear A ‘Transaction Signs,’” Minos 29-30 (1994-1995) 57-76.

\textsuperscript{25} There are only three occurrences of two-word headers at Haghia Triada (HT 28b, 34, 87) and two occurrences of headers with three words (HT 96a, 117a).

\textsuperscript{26} a-du occurs in the following tablet groups, with the individual tablets in parentheses: I (HT 88), II (HT 85), III (HT 99) and IV (HT 86 and 95).

\textsuperscript{27} Even when a-du does occur on tablets with the same vocabulary, it does not occur with precisely the same words. For example, on HT 86, a-du heads da-me and mi-ne-te, while on HT 95 it heads not only da-me and mi-ne-te, but also saru, kunisu, di-deru, and qe-ra-u, which are headed by a-karu on HT 86.

\textsuperscript{28} HT 85 is in our group II, HT 95 is in our group IV.

\textsuperscript{29} kunisu also occurs on HT 86 and 95.

\textsuperscript{30} This is merely a working hypothesis. To further test it, one would want to examine the role of words associated with tablet group II in all contexts.
A more traditional analysis would not have noticed this contrast, but it is clearly shown by the MDS graph. So by comparing the uses of individual words and groupings of the tablets, we move closer to a fully contextualized analysis.

MDS also allows us to answer questions about the Haghia Triada corpus as a whole. The find-spots of one set of tablets, those from the Casa del Labete, have been documented most fully.31 All other texts from Haghia Triada were published with little mention of their find-spots, although we know that they were found in the main villa. This has led to disagreement as to whether the Casa del Labete constitutes an archive distinct from the rest of the Haghia Triada texts. It has even been argued that the tablets from room 9 of the Casa del Labete constitute an earlier deposit, separate from the rest of the Casa del Labete material found in room 7.32 Yet both the Casa del Labete tablets taken as a whole and the tablets from room 9 are spread evenly throughout the MDS chart, suggesting that they are not distinct from other Haghia Triada tablets, and consequently do not constitute a specialized or chronologically distinct archive.33

The word matrix provides a necessary complement to the tablets matrix by supplying the context for individual words (see Pl. LXVII). The chart displays the relationship of any word to all others in the HT corpus, showing associated words as spatially proximate, while non-associated words are distant.34 The chart of words can be usefully divided into five clusters, some of which overlap.35 This configuration contrasts with the structure of the tablets chart, which presents us with relatively discrete groupings. It is worth noting that some of these clusters overlap in the words graph, but the tablets on which they occur are kept separate in the tablets graph. That is to say, there is not a simple one-to-one correspondence between the two matrices. While the tablets matrix regards the tablet as the unit of analysis and treats the words themselves on the tablet in an undifferentiated manner, the words matrix treats each word as the unit of analysis. Thus, the unique associations of each word are revealed, showing the complexities of word use in Linear A and the difficulties involved in interpreting individual lexical items.

Since the input data for both the tablet and word matrix are the same, with only a shift in the unit of analysis, it is not surprising that in many cases they provide much the same information. For example, HT 86 and HT 95 are quite close together on the tablet chart; they share five words, all of which also appear in a well-defined cluster at the top of the words matrix.36 In this case, the words matrix has not revealed anything that could not have been discerned from the analysis of the tablets chart, since the words appear to group in roughly the same manner as their associated tablets do.

On the other hand, there are cases in which the words matrix provides significant new information. For example, four words occur on tablets from our tablet group V (particularly HT 9, 17 and 19); these same words also turn up on HT 122, in our tablet group II (see Pl. LXVIIb). Group II is not proximal to group V, indicating that in general there is not much lexical overlap between the two. While the group V tablets deal with wine, HT 122 has the ideogram for "man" on one side (designated "b") and no ideogram with low round numbers...

---

31 The tablets found in the Casa del Labete are HT 83-123 and 129-131. For detailed syntheses, see MILITELLO (supra n. 1); I. SCHÖPE, "Context and Chronology of Linear A Administrative Documents," Aegean Archaeology 2 (1995) 29-65; E. HALLAGER, The Minoan Rounded and Other Sealed Documents in the Neopalatial Linear A Administration (1996) 41-45.
32 HT 114-123 and 129-131 are from room 9 of the Casa del Labete, which is argued to be of an earlier date by MILITELLO (supra n. 1) 242-243 and SCHÖPE (supra n. 7) 23-24, but not by HALLAGER (supra n. 31) 45.
33 It is worth noting that had the archives been chronologically or functionally distinct, one would have expected relatively high stress levels, since MDS analysis assumes that all items included in the analysis are alike. So for example, in our initial simulations, we included words from other sites, such as Kato Zakro, Kniania, etc.; in these simulations, we encountered much higher stress levels.
34 Note that we have only accounted for words that occur more than once in the Haghia Triada corpus.
35 We have attempted to be relatively conservative in our definition of these groups, meaning that we have taken into account on which tablets words occur as well as the spatial display of the MDS charts.
36 Six, if one accepts that qa\textsuperscript{2}\text{-}ja\textsuperscript{2}n (HT 95) represents the same word as qa\textsuperscript{2}\text{-}ja\textsuperscript{2}wa (HT 86).
on the other (designated "a"). When we consider where these four words occur on tablet 122, we notice that they are all on the "a" side of the tablet, while the words on the "b" side tend to cluster to the right of the HT 122 word group, away from the wine tablet words. Thus, the associations of the words, as represented by the MDS graph, reveal subtle differences in their use, which is reflected in the format of the tablet. This analysis reveals the value of examining words which occur in more than one grouping of tablets, since we can use the contextual data from MDS as a critical point of entry. That is to say, we can begin by asking why some words are associated with very different groups of vocabulary. Again, it is worth noting that this is only possible by a full contextual analysis of each word, taking into account its associations with every other word in the available corpus.

The advantages of MDS are considerable: the approach provides much-needed methodological rigor to Linear A studies, and requires only the most basic of assumptions, namely that in Linear A, context matters. We have attempted to demonstrate the viability of this approach and to illustrate some of the many means by which the data may be interpreted. It is important to remember that the MDS charts are not ends in themselves; rather, they are interpretational devices. The results from MDS can be used as a critical point of entry to begin asking questions of the Linear A corpus or to supplement other studies by providing context for individual words and tablets. When studying Linear A, one should first allow the contents of the tablets to speak for themselves as a corpus; using MDS allows for such an approach.

Dimitri NAKASSIS and Kevin PLUTA
| Pl. LXVIa | Sample input MDS matrices for Linear A tablets and words. |
| Pl. LXVIb | MDS chart of Linear A tablets at Haghia Triada. |
| Pl. LXVII | MDS chart of Linear A words at Haghia Triada. |