Coding to Decipher Linear A

Niki Cassandra Min Eu
Linguistics and Multilingual Studies
Nanyang Technological University
Singapore
NEU001@e.ntu.edu.sg

Duo Duo Xu
Department of Chinese Studies
National University of Singapore
Singapore
chsxd@nus.edu.sg

Francesco Perono Cacciafoco
Linguistics and Multilingual Studies
Nanyang Technological University
Singapore
fcacciafoco@ntu.edu.sg

***Abstract*- This paper discusses the program logic for an attempt at using coding to aid in the decipherment of Linear A, the writing system of the Ancient Minoan Civilization. Using Python, a system can be created to compare Linear A strings to lexical lists and dictionaries of languages from a compatible time period.**

Keywords—Linear A, Python, Programming, Historical Linguistics

I. Introduction

Linear A is the writing system of the Ancient Minoan Civilization, a Bronze Age Aegean civilization that flourished in Crete and several other Aegean islands, and was used between 1700-1450BCE before being replaced [1]. In-between the use of Linear B by the Mycenaeans for Mycenaean-Greek, there was also a period of Cypro-Minoan, used by the pre-Greek people of Cyprus [2]. Discovered alongside Linear B samples by Sir Arthur Evans in 1886, Linear A samples have been found in a variety of locations including Cyprus, Aegean Islands like Kea, Kythera, Melos and Thera [3], and mainland Greece and Turkey [4].

Since its discovery, many researchers have tried attributing a language family relation to Linear A, or have tried deciphering the language of the writing system, with limited success. Various languages (and language families) have been attributed to Linear A, but a large-scale attempt has not yet been made to process all the samples against dictionaries and lexical lists of various languages at a time. In order to vastly expand the search for a potential language family, this paper proposes the use of a program coded in Python to carry out the search faster, and help narrow down the list of potential candidates for in-depth analysis.

II. Literature Review

Linear A has around 90 signs/ symbols in regular use, 80% of which are unique when compared to Linear B and have been found to be used as individual signs as in combination [1]. Found on a variety of artefacts, it is generally agreed that a majority of the inscriptions found on tablets, roundels and seals denote economic transactions or were used for a stocktaking purpose. This conclusion was reached based on two sets of evidence: first, internal analysis found that a large number of tablets bore logograms in addition to regular signs, denoting commodities such as figs or olives, and preceded numbers. This, Linear A inscriptions have been found on stone vases (some inscribed, others painted), on stucco architectural features, libation tables, metal objects and other items. A vast number of the samples in Linear A are made up of roundels (see Figure 1: Roundel KH We 2057, from GORILA Vol. 3), a clay disc with one or more impressions. They were used as the "conveyance of a commodity, either within the central administration or between the central administration and an external party" [5].

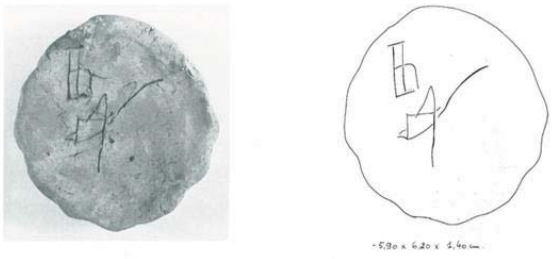

Figure 1: Roundel KH We 2057, from GORILA Vol. 3

The difficulty in deciphering Linear A begins with the number of samples available for analysis: The corpus of Linear A is, as of now, very small. There are 1427 artefacts with Linear A inscriptions, with signs appearing around 7400 times [5]. This may seem sufficient, but it is a small amount when compared to the Linear B corpus which appears on more than 4600 artefacts with signs occurring around 57000 times.

Most decipherment attempts begin by provisionally assigning Linear B phonetic values to Linear A signs which are graphically similar. This is done because Linear B, deciphered by Michael Ventris in 1952, was modelled on Linear A- a conclusion which can be drawn due to the shared signs between the writing systems. By doing this, we are able to 'read' the signs of Linear A, but the reliability of doing so is compromised. For one, the time difference between the use of the two writing systems is very vast, and Linear B encodes Mycenean Greek. Attempts to link Linear A to Greek have been thus far inconclusive, producing meaningless 'words'[6], with an additional issue being that 80% of Linear A signs are unique [7] and thus have no clear phonetic equivalent. These unsuccessful attempts to link Linear A to Greek have also affected its likelihood of being an Indo-European language. The 'Minoan' that the writing system encodes appears unrelated to any language, allowing a vast number of languages to be proposed as possible relations. Among the proposed language relations have been Greek [8], Etruscan [9], Sanskrit [10], and various Semitic Languages [11], [12].

Following the graphic similarity to Linear B, many scholars have postulated different hypotheses to identify the language. The first attempt to decipher Linear A was proposed by Vladimir I. Georgiev in 1957 to bear a Greek relation, and Gregory Nagy would continue this line of inquiry in 1963, presenting arguments on the phonetic-graphemic, lexical, and semantic levels. Nagy shows evidence of "varying worth" [8] of these Greek-like, or Indo-European, elements. In his paper, he claims consonant clusters such as I-JA-TE (which he had identified in Graffito II 12 of Phaistos, as *ne-ma i-ja-te*), and can be observed in Greek, are of "clearly Indo-European origin", and he speculated later in his paper (209) that Luwian, a language belonging to the now-extinct Anatolian branch of the Indo-European family, was the language in question. Despite proposing several examples that show this Indo-European connection, Nagy's claim is weakened because similar findings have been highlighted in other language families, a point that will be discussed later. Adopting Youngers' stance on this method of determining language family relationships, the weakness of Nagy's method of decipherment lies in the use of vocabulary to identify a language. This is because vocabulary is prone to being borrowed and the examples given by Nagy may not actually be from Linear A / Minoan [7]. Towards the end of his paper, Nagy speculates that there could be a relation between Linear A and Luwian, an Indo-European language. This would then be developed by later scholars.

Based on the Linear B phonetic values, Leonard R. Palmer theorized that Linear A could be the writing of an Anatolian language, possibly Luwian, or a Cretan variant of Luwian. Palmer posited this mostly because he believed that "Greece and Crete were twice invaded by Indo-European people during the second millennium BC" [13], an event which would have sparked a mass migration to Crete. This theory was based on two elements: the presence of "Minyan" ware in Beycesultan (Western Anatolia), and a Linear A inscription that he interpreted as "Mount Parnassos" and, according to him, was based on Luwian. In Luwian, 'Parnassos' means '(place) of the temple'. He also based this theory on the inscriptions found on "vessels of Minyan shapes",and claims to have recognized Luwian deity names in Linear A on, for example, the Pylos tablets. One such example is Pylos FR1227, where Palmer claims to read *wa-na-so-i pa-se-da-o-ne*, believing it to mean 'The two Queens and Poseidon'. This Indo-European link found another supporter in was Gareth Alun Owens, who released a collection of essays entitled *Kritika Daidalika* [14], and suggested a similar relation to Luwian but as an archaic relative. Using Linear B phonetic values, he detailed 50 words of the Minoan language which he 'deciphered' [10] Owens attributed, to two Linear A inscriptions, the values *ja-di-ki-te* and *i-da*. These he linked to the two holy mountains in Crete (Dikte and Ida). He stated that these words had "good Indo-European etymology". *I-da*, in particular, he proposed, was very similar to *i-na-/i-ja-*, and suggested that these words "(come) from the same root and indicate 'holy'" – a root which he connected to *ieros* in Greek and *isirah* in Sanskrit, 'proving' that Linear A must encode an Indo-European language. Owens postulated that Linear A represents a language from the *Satem* branch of the Indo-European family with "closer lexicographical characteristics with Greek and Sanskrit, more than with Hittite".

The theories of a Luwian (or Indo-European) connection were proposed various times over the years, but never gained consensus among the academic community. Palmer is first criticized for the heavy reliance on his interpretation of the tablets, which can have varying interpretations because of an incomplete understanding of the orthography. Immerwahr's 1963 critique of Palmers' work also touches on the issue of the Minyan ware, expressing how "few prehistoric archaeologists will accept this premise that the Minyans were Luwians and that the Indo-European migration that marked the end of the Early Helladic was not yet a Greek migration", also citing a shortage of archaeological evidence. Mylonas (1962) also challenges the Luwian theory on various grounds, echoing Immerwahr's view that there is a large amount of doubts about whether the Beycesultan people were Luwians, citing that Palmer's evidence of the Minyan Ware was not sufficiently qualified [15]. Minyan Ware is identified based on characteristics such as colour and form features and as a result of this, establishing the development of the pottery is difficult. Palmer's theory also relies on an invasion in 1700BC, which coincides with the use of Linear A and resulted in the naming of a mountain, *Parnassos*. However, there is still a lack of concrete evidence of when Mount Parnassos was named and, additionally, no archaeological evidence that the Luwians used the area at all as a place of worship (recall that Parnassos is supposed to mean '(place) of the temple'). A mountain could not have possibly be named after a temple that did not yet exist, which further bolsters the lack of physical evidence to support Palmer's linguistic evidence. There is, all in all, very little evidence pointing to anything other than trade contact between Luwians and Minoans and, therefore, it is unlikely that their languages would be related. Additional reasons for the rejection of the connection include the small states along the Western coast of Asia Minor that would have been natural barriers to the contact between the Luwians and Minoan Crete, and no remarkable resemblance between Minoan and Luwian morphology [16].

The second major language family of interest for Minoan is the Semitic language family, first proposed to be connected with Linear A by Cyrus H. Gordon (1966). Like most scholars working on the topic, Gordon applied the phonetic values of Linear B to the Linear A samples and found give words identified by Ventris and Chadwick [17]: *su-po*, *ka-ro-pa*, *pa-pa*, *su-pa-ra*, *pa-ta-qe* (all accompanied with pot signs), as well as the commonly found *ku-ro* at the end of administrative tablets. Gordon, who had extensive knowledge of the Semitic languages and worked specifically with Ugaritic, recognized that three of these vessel words show consonantal roots that exist in Ugaritic: *sp*, *krpn*, and *spl* (matching the first, second, and fourth words listed previously). Following this success, Gordon would continue to identify words in Linear A that were recognizable in various languages belonging to the Semitic language family, like Akkadian and Hebrew, eventually believing Linear A was connected specifically to West Semitic. Western Semitic is a good candidate for relation, as dialects of it were spoken along the Mediterranean seaboard, an area which is geographically close to Crete. In a lecture based on Gordon's initial findings, Maurice Pope (1958) gave a lecture that bolstered the possibility of Semitic as the language not only by corroborating some of the words that Gordon had identified, but also pointing out certain Semitic grammatical features, such as the presence of a copula on tablets 117a, and 122a & b, where *u-* can be found at the beginning of the second word consistently. This is important because in Akkadian and ancient Hebrew where 'and' is denoted by *u* and *waw*, showing a possible connection to Semitic. Of course, this is by no means conclusive; however, the presence of grammatical inflection [18] on top of this identification seemed to only further promote the connection. The word *kuro* is also commonly raised as an indicator of, at a basic level, some Semitic influence – it is the only word in Linear A whose meaning is the most probable under the Semitic theory, meaning 'total'. Present archaeological evidence does not rule out Semitic influence, but, at the same time, it does not fully support Semitic influence either. Jan Best (1972) would continue Gordon's attempts, presenting a controversial paper promoting Linear A as the script of a Semitic language, closely related to Ugaritic [19].

Language contact with Semitic is, at the very least, a possibility. Minoans traded all over the Eastern Mediterranean, and there has been evidence of cultural contact in places like Cyprus, Canaan (located in present-day Lebanon, Syria, Jordan, and Israel), and the Levantine Coast. Minoan-Style wall paintings were also discovered in 2009 in Tel Kabri in Israel. In Tel Kabri, remains of a Canaanite city from the Middle Bronze age (2000-1550 BC) coincide with the time Linear A was in use, and Canaan is a Semitic-speaking region. Kamares Ware (a distinct type of Minoan Pottery that reached its peak in popularity around MMIIB, about 1750 BC) has also been found in many Egyptian sites including the Delta, Middle Egypt, and Aswan in Upper Egypt [20]. Evidence of Middle Minoan pottery (dating 2100BC to 1500BC) can also be found in the Aegean Islands, the Near East (the countries of the Arabian peninsula), Mesopotamia, and Anatolia, showing how much the Minoans traded in the surrounding regions, increasing the possibility for language contact.

No theory comes without controversy- much like the Luwian hypothesis, many scholars reject Semitic as a possibility for the language of Linear A. Gordon found approximately 50 words, and the reliability of these matches is compromised because all the words identified are vocabulary items. As mentioned previously, vocabulary items are not considered a reliable means of identifying a possible family connection. Packard (1974) also points out the difficulty in connecting the five words (*su-po*, *ka-ro-pa*, *pa-pa*, *su-pa-ra*, *pa-ta-qe*) with Ugaritic names because of how vowels are ambiguous in Semitic writing [21]. Additionally, because trade was so prominent, these word strings, found on administrative tablets, could have just been loanwords from the surrounding regions. Chadwick also rejected the Semitic theory, stating that "if the vowels are ignored we are leaving out half the information presented by the script". This is because, in Semitic languages, vowels could be considered 'semi-vowels' with a specific 'colour'. The common criticism of Gordon's work also stems from the fact that he linked various elements to not one Semitic language, but several – Canaanite, some Aramaic, some Akkadian, and so on. This apparent lack of any specific Semitic language prompted the view by many scholars that Gordon's work was not successful in establishing a Semitic link.

More recent decipherment attempts have turned to algorithmic approaches, in the hope that computerized, automated efforts would be more efficient in generating more matches [22]. Revesez (2017) proposed that the language of Linear A was connected to the Uralic family, and unlike previous attempts, presented an algorithm which would "find the syllabic values of the Linear A symbols". Taking these values, Revesez then uses the proposed Linear A values to build a Uralic-Minoan dictionary which then is used to 'translate' twenty-eight Linear A documents from GORILA. This novel new approach allowed Revesez to 'read' close to 30 sets of inscriptions and propose a dictionary. However, the problem of biased interpretation may remain for two reasons. Firstly, Revesz explicitly set out to prove the hypothesis that the Minoan language could be linked to the Uralic family. The determination of the Syllabic values of Linear A were carried out specifically with other proposed languages of the family and were based entirely on the graphic similarity. Cross-family

comparisons were not made to evaluate the relative likelihoods of the Minoan's language relation to one family over another's. Next, words seem to fit in the most plausible positions, but this has been done without any consideration of the provenance of the artefacts that contain the clusters. Such an approach is incomplete as many contextual clues which can help evaluate the relative validity of interpretations, and hence debunk certain interpretations which might at first seem tenable, were not considered. It is also interesting to note that the examples of translation he had used in his paper were restricted to the libation tables and objects, and no attempt was made using the economic tablets that can be found in GORILA volumes 1 and 3, which have a known and agreed upon context.

III. Methodology

A. Document Preparation

Documents for comparison needed to be created and sorted for input into the program. Two major lists were created: one that contained all usable Linear A samples and another for the dictionary or lexical list it was being compared to. Samples were drawn from Godart and Olivier volumes 1, 3. Volume 2 was excluded because it included mostly individual signs, which could not be formed into strings for comparison. Volume 4 was left out because it records libation tables and could use a ritual language, a version of the language that would otherwise not be used outside of its purpose [23]. Various dictionaries and lexical lists also needed to be converted into a digital format in order to be used by the program.

B. Intended Program Logic

There are various elements and variables that need to be considered when designing the program and the logic it runs on in order to give us lists that can then be used for a manual translation attempt. In a previous paper, I attempt a manual translation and matching via root comparison: the shortfalls of this method have since been accounted for. The considerations and details for each step have been included in the methodology.

Two excel files are prepared: one containing samples of Linear A from GORILA 1, and 3, and another with words pulled from the various dictionaries listed above in 3A.

The program then draws from Linear A master list that contains the samples, and **splits word strings into 2, 3 and 4 phone long chunks.** Phones are defined in Linear A according to the individual symbol and its Linear B phonetic equivalent. For example, IO ZA 1 from GORILA 4 was initially transcribed with the Linear B phones as per below, with dashes separating the phones of individual symbols, and x's demarcating places where symbols were missing or unclear, due to the age of the sample:

A-TA-I-A301-WA-JA x JA-DI-KI-TU x JA-SA-SA-RA-[x x x]-SI x I-PI-NA-MA-x

In order to make it suitable for the program, we reformat the string into something like this:

A-TA-I-[A301]-WA-JA[]JA-DI-KI-TU[]JA-SA-SA-RA-[]-SI[]I-PI-NA-MA

Gaps between 'words' or missing signs are demarcated with []. One of the biggest problems on applying this method to Linear A is the presence of symbols that do not have a phonetic equivalent in Linear B. Ideally, the program would be able to apply the search and consider any dictionary entry which matches other elements of the string a positive match.

A-TA-I-A301
A-TA-I
A-TA
TA-I-A301-WA
TA-I-A301
TA-I
I-A301-WA-JA
I-A301-WA
I-A301
A301-WA-JA
A301-WA

All the above should generate as a part of the splitting of word strings for the string A-TA-I-A301-WA-JA. The program should not loop to the first syllable and should do this for all of the words In the list.

Next, the program then needs to make adjustments in the Dictionary List so that it can be compared properly. After accounting for variables, a new dictionary list is generated for comparison. For example, in the Hamito Semitic Dictionary, capital letter V stands for a variable vowel. This means: abVnan □ abanan □ abenan □ abonan etc.

For the book which suggests Basque as Proto-Indo-European, C is for a variable consonant.

The unique variables of each dictionary or language need to be accounted for. Some languages also feature use of 'special characters' (i.e. θ or ð), and so the program must be able to read those. That said, we have encountered no dental fricatives in the data. An important consideration for each individual test language are the C and V's that we use as variables- for example, not all consonants of the English language are valid consonants in other languages, and as such, these differences need to be accounted for by the program.

The program then compares items from (1) with items from (2) and outputs a file which shows all the matches. These matches can then be taken and manually processed based on their probability, obtained from the number of matches we got from the program.

The basic structure of the program is implemented to compare two spreadsheets for similarities. One spreadsheet contains Linear A transcriptions, while the other one the entries from dictionaries. In Python context, the module "pandas" is often imported for this purpose. The module helps to check if the two dataframes have the same shape and elements. Then the module "numpy" should be imported to find out the index of the cells where the value is "True". Alternatively, the module "pandas.DataFrame.equals" helps to find out the elements with exactly the same values.

C. *Frequency Analysis Based on Online Corpus of Linear A*

In the recent months, the Linear A corpus has been digitalized online by Robert Hogan, called the Linear A Explorer. Recording basic information of some of the Linear A samples, it also features commentary (where available) by John Younger. The explorer is able to provide a frequency analysis by matching recurring clusters, and is an incredibly useful tool that can now be used to cut down the time it takes to identify recurrences- hovering over a word cluster informs the user of any matches throughout the rest of the corpus. While useful, it is important to note that its largest limitation is that clusters with any kind of variation are ignored by its search program. For example, in GORILA Volume 4, a common cluster amongst the libation tables is string A-TA-I-A301-WA-JA. On the explorer, there are a recorded 7 instances of this exact string. However, manual analysis produces additional results such as A-TA-I-A301-U-JA and A-TA-I-A301-WA-E that have similar if not identical preceding strings to their A-TA-1-A301-WA-JA counterpart. This does not mean the frequencies showed are unreliable, rather, they must be selected with care. Samples from GORILA 1 and 3, as mentioned previously, are record the transaction of commodities. As such, these vocabulary items are more easily identified for their length (by cluster) and are likely to vary less. As such, strings with higher frequencies that we see on the explorer can be reliably used for experiment with the program and to double check the results generated by the program itself.

IV. Conclusion

The application of this program to Linear A represents a key move away from previous, philological attempts. A large majority of studies so far have relied on outward resemblances with words from other languages. It attempts a method similar to that of Revesez (2017), while using the phonetic values from Linear B. The current program, while simple, allows Linear A to be compared to a variety of languages from the Semitic family, expanding the matching process beyond word recognition. In addition, changes to the program can be made to accommodate comparison with other language families. The program offers the opportunity to narrow the candidate for a language family through a larger, statistics-based process instead, taking into account variables and the full corpus of comparison language. This, of course, is not perfect- to quote Yves Duhoux from March 1998, "The conclusion must be that even if one can find casual resemblances between words in both languages (remember this MUST statistically happen) …they are probably structurally different." The overall phonological and more importantly, morphological system must be resolved before a complete conclusion can be drawn. The program can be eventually adjusted to make matching other language-families possible. This program aims to hasten the process by 'brute force', but would aid in future research efforts.

References

[1] G. Cadogan, *Palaces of Minoan Crete*. Routledge, 1976.

[2] B. Davis, "Introduction to the Aegan Pre-Alphabetic Scripts," University of Melbourne, Australia, Melbourne, Australia, Paper 1, 2010.

[3] F. Perono Cacciafoco, "Linear A and Minoan The Riddle of Unknown Origins," Nanyang Technological University, 17-Jan-2014.

[4] Anadolu Agency, "Mycenean artifacts found in Bodrum," *Hurriyet Daily News*, Turkey, 10-Nov-2014.

[5] I. Schoep, "Social and Political Organisation on Crete in the Proto-palatial Period: The Case of Middle Minoan II Malia," *J. Mediterr. Archaeol.*, vol. 15, Jun. 2002.

[6] L. Godart, "Du Lineaire A au Lineaire B," in *Aux origines de l'hellénisme: La Crète et la Grèce.*, vol. 15, Paris: Publications de la Sorbonne, 1984, pp. 121–128.

[7] J. Younger, "Linear A Texts: Homepage," *Linear A Texts and Inscriptions in Phonetic Transcription & Commentary*, 30-Nov-2000. [Online]. Available: http://people.ku.edu/~jyounger/LinearA/#7c. [Accessed: 14-Dec-2017].

[8] G. Nagy, "Greek-like elements in Linear A," *Greek Roman Byzantine SStudies*, vol. 4, no. 4, pp. 181–211, 1963.

[9] G. M. Facchetti and M. Negri, *Creta minoica: sulle tracce delle più antiche scritture d'Europa*. L.S. Olschki, 2003.

[10] G. A. Owens, "The Structure of the Minoan Language," *J. Indo-Eur. Stud.*, vol. 27, pp. 15–56, 1999.

[11] C. H. Gordon, *Evidence for the Minoan Language*. Ventor Pub, 1966.

[12] J. G. P. Best, *Some preliminary remarks on the decipherment of Linear A*. Amsterdam : Hakkert, 1972.

[13] L. R. Palmer, *Mycenaeans and Minoans; Aegean Prehistory in the Light of the Linear B Tablets*. Faber & Faber, 1961.
[14] G. A. Owens, *"Kritika Daidalika": Evidence for the Minoan Language: Selected Essays in Memory of James Hooker on the archaeology , epigraphy and philology of Minoan and Mycenaean Crete*. Hakkert, 1997.
[15] G. E. Mylonas, "The Luvian Invasions of Greece," *Hesperia J. Am. Sch. Class. Stud. Athens*, vol. 31, no. 3, pp. 284–309, 1962.
[16] F. Perono Cacciafoco, "Linear A and Minoan: Some New Old questions," *Analele Univ. Din Craiova Ser. Ştiinţe Filol. Lingvistică*, no. 1–2, pp. 154–170, 2017.
[17] M. Ventris and J. Chadwick, *Documents in Mycenaean Greek*, 2nd ed. Cambridge [Eng.]: University Press, 1973.
[18] M. Pope, "On the Language of Linear A," University of Capetown, 1958.
[19] G. A. Rendsburg, "'Someone Will Succeed in Deciphering Minoan': Cyrus H. Gordon and Minoan Linear A," *Biblic. Archaeol.*, vol. 59, no. 1, p. 36, Mar. 1996.
[20] P. Bradley, *The Ancient World Transformed*. Cambridge University Press, 2014.
[21] D. W. Packard, *Minoan Linear A*. University of California Press, 1974.
[22] P. Revesz, "Establishing the West-Ugric Language Family with Minoan, Hattic and Hungarian by a Decipherment of Linear A," *WSEAS Trans. Inf. Sci. Appl.*, vol. 14, pp. 306–335, Nov. 2017.
[23] W. T. Wheelock, "The Problem of Ritual Language: From Information to Situation," *J. Am. Acad. Relig.*, vol. 50, no. 1, pp. 49–71, 1982.