

# Minoan linguistic resources: The Linear A Digital Corpus

Tommaso Petrolito<sup>⊖</sup> Ruggero Petrolito<sup>⊖</sup> Grégoire Winterstein<sup>⊖</sup> and Francesco Perono Cacciafoco<sup>⊕⊖</sup>

<sup>⊖</sup> Filologia Letteratura e Linguistica, University of Pisa, Italy

<sup>⊖</sup> Linguistics and Modern Language Studies, The Hong Kong Institute of Education, Hong Kong

<sup>⊕</sup> Linguistics and Multilingual Studies, Nanyang Technological University, Singapore

tommasouni@gmail.com, ruggero.petrolito@gmail.com,

gregoire@ied.edu.hk, fcacciafoco@ntu.edu.sg

## Abstract

This paper describes the Linear A/Minoan digital corpus and the approaches we applied to develop it.

We aim to set up a suitable study resource for Linear A and Minoan.

Firstly we start by introducing Linear A and Minoan in order to make it clear why we should develop a digital marked up corpus of the existing Linear A transcriptions.

Secondly we list and describe some of the existing resources about Linear A: Linear A documents (seals, statuettes, vessels etc.), the traditional encoding systems (standard code numbers referring to distinct symbols), a Linear A font, and the newest (released on June 16th 2014) Unicode Standard Characters set for Linear A.

Thirdly we explain our choice concerning the data format: why we decided to digitize the Linear A resources; why we decided to convert all the transcriptions in standard Unicode characters; why we decided to use an XML format; why we decided to implement the TEI-EpiDoc DTD.

Lastly we describe: the developing process (from the data collection to the issues we faced and the solving strategies); a new font we developed (synchronized with the Unicode Characters Set) in order to make the data readable even on systems that are not updated. Finally, we discuss the corpus we developed in a Cultural Heritage preservation perspective and suggest some future works.

## 1 Introduction to Linear A and Minoan

Linear A is the script used by the Minoan Civilization (Cotterell, 1980) from 2500 to 1450 BC.

## Writing system

Cretan Hieroglyphic 2100 – 1700 BC

Linear A 2500 – 1450 BC

Linear B 1450 – 1200 BC

## Time span

Table 1: Time spans of Cretan Hieroglyphic, Linear A and Linear B.

The Minoan Civilization arose on the island of Crete in the Aegean Sea during the Bronze Age. Minoan ruins and artifacts have been found mainly in Crete but also in other Greek islands and in mainland Greece, in Bulgaria, in Turkey and in Israel.

Linear A is not used anymore and, even after decades of studies (it was discovered by Sir Arthur Evans around 1900 (Evans, 1909)), it still remains undeciphered.

All the assumptions and hypotheses made about Linear A and Minoan (its underlying language) are mainly based on the comparison with the well known Linear B, the famous child system originated by Linear A. In fact, Linear B was fully deciphered during the 1950s by Michael Ventris<sup>1</sup> and was found to encode an ancient Greek dialect used by the Mycenaean civilization.

Archaeologist Arthur Evans named the script ‘Linear’ because it consisted just of lines inscribed in clay (Robinson, 2009) while, in the same period (as shown in Table 1), Cretan hieroglyphs were more pictographic and three-dimensional.

Even if many symbols are shared by both Linear A and Linear B, it has not been possible to find intelligible words within inscriptions in Linear A by applying Linear B segmentation and phonemes.

Linear A consists of hundreds of symbols probably having syllabic, ideographic, and semantic values. Many of the Linear A symbols that are

<sup>1</sup><http://www.cam.ac.uk/research/news/cracking-the-code-the-decipherment-of-linear-b-60-years-on>

<b>Text</b>	‡𐀓𐀕
<b>Phonetic Value</b>	<i>KU-NI-SU</i>
<b>Possible Meaning</b>	<i>Knossos</i>

Table 2: Example of John G. Younger’s deciphering attempt.

found in Linear B (81 in total) are assumed to have syllabic values, while the remaining are assumed to be logograms.

There have been several attempts to decipher Linear A and the Minoan Language. We can divide the underlying hypotheses in six groups: Greek-like language (Nagy, 1963), distinct Indo-European branch (Owens, 1999), Anatolian language close to Luwian (Palmer, 1958), archaic form of Phoenician (Dietrich and Loretz, 2001), Indo-Iranian (Faure, 1998) and Etruscan-like language (Giulio M. Facchetti and Negri, 2003).

There is also an interesting attempt (Younger, 2000b) to decipher single words, specifically toponyms, by applying Linear B phonetic values to the symbols shared by both Linear A and Linear B and following the assumption that toponyms are much more likely to survive as loans in Mycenaean Greek (written in Linear B); we show an example of this approach in Table 2.

In the next sections we describe the available existing resources concerning Linear A and the Linear A Digital Corpus: why and how we developed it.

## 2 Linear A available resources

Even if Linear A and Linear B were discovered more than one century ago, Linear A has not been deciphered yet. Nevertheless, many scholars worked on collecting and organizing all the available data in order to study and to decipher the script and the language.

Probably due to the fact that only historical linguists, philologists and archaeologists attempted to collect and organize all the existing data, nowadays a rich and well organized digital corpus is still not available.

In this section we describe all the available Linear A resources, including both physical documents and digital data.

<b>ID</b>	<b>Type of Support</b>
default	tablets (page, bars, lames)
Wa	noduli
Wb	sealings
Wc	roundels
Za	stone vessels
Zb	clay vessels
Zc	inked inscriptions
Zd	graffiti
Ze	architecture
Zf	metal objects
Zg	stone objects

Table 3: Indexed types of support (Younger, 2000e).

### 2.1 Linear A documents

Linear A was written on a variety of media, such as stone offering tables, gold and silver hair pins, and pots (inked and inscribed).

The clay documents consist of tablets, roundels, and sealings (one-hole, two-hole, and flat-based).

Roundels are related to a "conveyance of a commodity, either within the central administration or between the central administration and an external party" (Palmer, 1995; Schoep, 2002). The roundel is the record of this transaction that stays within the central administration as the commodity moves out of the transacting bureau (Hallager, 1996). Two-hole sealings probably dangled from commodities brought into the center; one-hole sealings apparently dangled from papyrus/parchment documents; flat-based sealings (themselves never inscribed) were pressed against the twine that secured papyrus/parchment documents (Younger, 2000g; Schoep, 2002) as shown by photographs (Müller, 1999), (Müller, 2002) of the imprints that survive on the underside of flat-based sealings.

There are 1,427 Linear A documents containing 7,362-7,396 signs, much less than the quantity of data we have for Linear B (more than 4,600 documents containing 57,398 signs) (Younger, 2000f).

### 2.2 Godart and Olivier’s Collection of Linear A Inscriptions

There is a complete and organized collection of Linear A documents on a paper corpus, the **GORILA** Louis Godart and Jean-Pierre Olivier, *Recueil des inscriptions en Linéaire A* (Godart and Olivier, 1976).

Unicode	GORILA	Pope&Raison	Syllable
10600	AB01	L30	DA
10601	AB02	L22	RO
10602	AB03	L2	PA

Table 4: Excerpt of John G. Younger’s transcription systems conversion table extended with the Unicode codes.

Godart and Olivier have indexed the documents by original location and type of support, following the Raison–Pope Index (Raison and Pope, 1971).

For example, the document **AP Za 1** is from **AP** = *Apodoulou* and the support type is **Za** = *stone vessels* as shown in Table 3.

Younger (2000h) provides a map with all the Cretan sites and one with all the Greek non-Cretan sites (Younger, 2000i).

Godart and Olivier also provide referential data about conservation places (mainly museums), and periodization (for example: **EM II** = *Second Early Minoan*).

Since 1976, this has been the main source of data and point of reference about Linear A documents and it has set up the basis for further studies. Even recent corpora, such as the *Corpus transnuméré du linéaire A* (Raison and Pope, 1994), always refer to GORILA precise volumes and pages describing each document.

### 2.3 John G. Younger’s website

Younger (2000j) has published a website that is the best digital resource available (there is another interesting project, never completed, on Yanis Deliyannis’s website<sup>2</sup>). It collects most of the existing inscriptions (taking GORILA as main source of data and point of reference) transcribed as Linear B phonetic values (like the *KU-NI-SU* transcription above).

The transcriptions are kept up to date and a complete restructuring in June 2015 has been announced (Younger, 2000j).

### 2.4 GORILA symbols catalogue

Many transcription systems have been defined.

The first one has been proposed by Raison and Pope (1971) and uses a string composed by one or two characters (*Lm*, *L* or *Lc* depending on the symbol, respectively metric, phonetic or compound) followed by a number, for example: *L2*.

<sup>2</sup><http://y.deliyannis.free.fr/linearA/>

This system has been widely used by many scholars such as David Woodley Packard (president of the Packard Humanities Institute<sup>3</sup>), Colin Renfrew and Richard Janko (Packard, 1974; Renfrew, 1977; Janko, 1982).

The second one, used in the GORILA collection (Godart and Olivier, 1976) and on John G. Younger’s website, consists of a string composed by one or two characters (*AB* if the symbol is shared by Linear A and Linear B, *A* if the symbol is only used in Linear A) followed by a number and eventually other alphabetical characters (due to *addenda* and *corrigenda* to earlier versions), for example: *AB03*.

Many scholars transcribe the symbols shared by Linear A and B with the assumed phonetical/syllabic transcription. This syllabic transcription is based on the corresponding Linear B phonetic values. Younger (2000a) provides a conversion table of Pope and Raison’s transcription system, GORILA’s transcription system and his own phonetic/syllabic transcription system.

Developing our corpus, we worked mainly on Younger’s syllabic and GORILA transcriptions, because the Unicode Linear A encoding is broadly based on the GORILA catalogue, which is also the basic set of characters used in decipherment efforts<sup>4</sup>. We provide an example of different transcriptions for the same symbol in Table 4. As can be noticed, the Unicode encoding is based on the GORILA transcription system.

### 2.5 Linear A Font

The best Linear A Font available is *LA.ttf*, released by D.W. Borgdorff<sup>5</sup> in 2004.

In this font some arbitrary Unicode positions for Latin characters are mapped to Linear A symbols.

On one hand this allows the user to type Linear A symbols directly by pressing the keys on the keyboard; on the other hand, only transliterations can be produced. The text eventually typed internally will be a series of Latin characters.

It should be remarked that this font would not be useful to make readable a Linear A corpus that is non-transliterated and encoded in Unicode.

<sup>3</sup><http://www.packhum.org/>

<sup>4</sup><http://www.unicode.org/versions/Unicode7.0.0/ch08.pdf>

<sup>5</sup><http://www.fontineed.com/author/D.W.%20Borgdorff>

## 2.6 Unicode Linear A Characters Set

On June 16th 2014, Version 7.0 of Unicode standard was released<sup>6</sup>, adding 2,834 new characters and including, finally, the Linear A character set.

Linear A block has been set in the range 10600–1077F and the order mainly follows GORILA's one<sup>7</sup>, as seen in Table 4.

This Unicode Set covers simple signs, vase shapes, complex signs, complex signs with vase shapes, fractions and compound fractions.

This is a resource that opens, for the first time, the possibility to develop a Linear A digital corpus not consisting of a transliteration or alternative transcription.

## 3 Corpus data format

Many scholars have faced the issues for data curation and considered various possibilities.

Among all the possible solutions, we chose to develop the Linear A Digital Corpus as a collection of TEI-EpiDoc XML documents.

In this section we explain why.

### 3.1 Why Digital?

Many epigraphic corpora have begun to be digitalized; there are many reasons to do so.

A digital corpus can include several representations of the inscriptions (Mahoney, 2007):

- pictures of the original document;
- pictures of drawings or transcriptions made by hand simplifying the document;
- diplomatic transcriptions;
- edited texts;
- translations;
- commentaries.

Building a database is enough to get much richer features than the ones a paper corpus would provide. The most visible feature of an epigraphic database is its utility as an *Index Universalis* (Gómez Pantoja and Álvarez, 2011); unlike hand-made indexes, there is no need to constrain the number of available search-keys.

Needless to say, the opportunity to have the data available also on the web is valuable.

<sup>6</sup><http://blog.unicode.org/2014/06/announcing-unicode-standard-version-70.html>

<sup>7</sup><http://www.unicode.org/versions/Unicode7.0.0/ch08.pdf>

## 3.2 Why Unicode?

Text processing must also take into account the writing systems represented in the corpus.

If the corpus consists of inscriptions written in the Latin alphabet, then the writing system of the inscriptions is the same as that of the Western European modern languages used for meta-data, translations, and commentaries.

In our case, unluckily, we have to deal with Linear A, so we need to find a way to represent our text.

Scholars objected to epigraphic databases on the ground of its poor graphic ability to represent non-Latin writing systems (García Barriocanal et al., 2011).

This led to the use of non-standard fonts in some databases which proved to be a bad move, compromising overall compatibility and system upgrading.

This approach is appealing because if the corpus needs to be printed, sooner or later fonts will be a need in all cases.

The font-based solution assumes that all the software involved can recognize font-change markers. Unluckily, some Database Management Systems (DMSs) do not allow changes of font within a text field and some export or interchange formats lose font information.

When the scripts of the corpus are all supported, which will be the case for any script still used by a living language, Unicode is a better approach. Despite Minoan not being a living language, Linear A is finally part of the Unicode 7.0 Character Code Charts<sup>8</sup> but some sign groups conventionally interpreted as numbers have no Unicode representation.

### 3.3 Why XML?

Until not so long ago, markup systems have always involved special typographical symbols in the text—brackets, underdots, and so on.

Some epigraphers see XML as a natural transformation of what they have always done, with all the additional benefits that come from standardization within the community.

There is a growing consensus that XML is the best way to encode text.

Some corpora may also use the typographical marks of the Leiden system, which has the advan-

<sup>8</sup><http://www.unicode.org/charts/>

```

<glyph xml:id="n5">
  <glyphName>
    Number 5
  </glyphName>
  <mapping type="standardized">
    5
  </mapping>
</glyph>

```

Figure 1: Example XML declaration of a glyph with no Unicode representation.

tage of being entirely familiar to the epigraphers who create and maintain the corpus.

Unfortunately, the special brackets, underdots, and other typographical devices may not be supported by the character set of the computer system to be used.

A key incentive for using XML is the ability to exchange data with other projects.

It is convenient to be able to divide the information in many layers: cataloging, annotating, commenting and editing the inscriptions.

In some cases, merging different layers from different projects could be a need (for example when each of these projects is focused on a specific layer, for which provides the best quality), as a consequence the resulting data should be in compatible forms.

If the projects use the same Document Type Definition (DTD), in the same way, this is relatively easy.

While corpora that store their texts as word-processor files with Leiden markup can also share data, they must agree explicitly on the details of text layout, file formats, and character encodings.

With XML, it is possible to define either elements or entities for unsupported characters.

This feature is particularly interesting in our case, giving a solution for the numbers representation (Linear A numbers, except for fractions, have no Unicode representation). Suppose you want to mark up the sign group "𐤎", conventionally interpreted as the number 5, in the XML. As specified in the TEI DTD, this could be expressed as `<glyph ref="#n5"/>`, where the element `glyph` indicates a glyph, or a non-standard character and the attribute value points to the element `glyph`, which contains information about the specific glyph. An example is given in Figure 1.

Alternatively, the project might define an entity

to represent this character.

Either way, the XML text notes that there is the Linear A number 5, and the later rendering of the text for display or printing can substitute the appropriate character in a known font, a picture of the character, or even a numeral from a different system. Such approaches assume that tools are available for these conversions; some application, transformation, or stylesheet must have a way to know how to interpret the given element or entity.

The usage of XML provides two advantages: in first place, it makes possible the encoding of the characters that occur in the text (as shown above); in second place, it's really useful for encoding meta-information.

### 3.4 Why EpiDoc?

If a project decides to use XML, the most appropriate DTD (or schema) to be used needs to be chosen. As in every other humanities discipline, the basic question is whether to use a general DTD, like the TEI, or to write a project-specific one. Some projects need DTDs that are extremely specific to the types of inscriptions they are dealing with, instead other projects prefer to rely on existing, widely used DTDs.

Mahoney (2007) has deeply analyzed all the digitization issues, taking into account all the advantages and disadvantages of different approaches; her conclusion is that it's best to use **EpiDoc**<sup>9</sup> an XML encoding tool that could be also used to write structured documents compliant with the TEI standard<sup>10</sup>.

The EpiDoc DTD is the TEI, with a few epigraphically oriented customizations made using the standard TEI mechanisms. Rather than writing a DTD for epigraphy from scratch, the EpiDoc group uses the TEI because TEI has already addressed many of the taxonomic and semantic challenges faced by epigraphers, because the TEI-using community can provide a wide range of best-practice examples and guiding expertise, and because existing tooling built around TEI could easily lead to early and effective presentation and use of TEI-encoded epigraphic texts (Mahoney, 2007).

The TEI and EpiDoc approaches have already been adopted by several epigraphic projects (Bordard, 2009), such as the D emos project (Furman

<sup>9</sup><http://www.tei-c.org/Activities/Projects/ep01.xml>

<sup>10</sup><http://www.tei-c.org/index.xml>

University) and the corpus of Macedonian and Thracian inscriptions being compiled at KERA, the Research Center for Greek and Roman Antiquity at Athens (Mahoney, 2007).

Also other scholars evaluate EpiDoc as a suitable choice. Felle (2011) compares the EAGLE (Electronic Archive of Greek and Latin Epigraphy<sup>11</sup>) project with the EpiDoc existing resources, viewing these resources as different but complementary.

Álvarez et al. (2010) and Gómez Pantoja and Álvarez (2011) discuss the possibility of sharing Epigraphic Information as EpiDoc-based Linked Data and describe how they implemented a relational-to-linked data solution for the Hispania Epigraphica database.

Cayless (2003) evaluates EpiDoc as a relevant digital tool for Epigraphy allowing for a uniform representation of epigraphic metadata.

The EpiDoc guidelines are emerging as one standard for digital epigraphy with the TEI.

EpiDoc is not the only possible way to use the TEI for epigraphic texts but the tools, documentation, and examples<sup>12</sup> make it a good environment for new digitization projects as ours.

### 3.5 EpiDoc structure

An EpiDoc document is structured as a standard TEI document with the `teiHeader` element including some initial `Desc` sections (`fileDesc`, `encodingDesc`, `profileDesc`, `revisionDesc`, etc) containing metadata, general information and descriptions (here we annotated place, period, kind of support and specific objects/fragments IDs). An interesting use of `encodingDesc` is shown in Figure 1 above: the `glyph` element has to be defined inside its parent element `charDecl` and its grandparent element `encodingDesc`.

The `teiHeader` element is followed by the `text` element including the body element composed by a series of unnumbered `<div>`s, distinguished by their `type` attributes (we show an example of the `Epidoc <div>` element in Figure 2).

Typical divisions might include:

- `text` itself (`type="edition"`);
- `translation` (`type="translation"`);

<sup>11</sup><http://www.eagle-eagle.it/>

<sup>12</sup>[http://wiki.tei-c.org/index.php/Samples\\_of\\_TEI\\_texts](http://wiki.tei-c.org/index.php/Samples_of_TEI_texts)

- `description` (`type="description"`);
- `commentary` (`type="commentary"`);
- `historical information` (`type="history"`);
- `bibliography` (`type="bibliography"`).

```
<div lang="minoan"
  n="text"
  type="edition"
  part="N"
  sample="complete"
  org="uniform">
<head lang="eng">Edition</head>
<cb rend="front" n="HM 1673"/>
<ab part="N">
  <lb n="1"/>
  <w part="N">ϣϣ</w>
  <space dim="horizontal"
    extent="1em"
    unit="character"/>
  <w part="N">ϣ</w>
  <lb n="2"/>
  <w part="N">ϣ</w>
  <g ref="#n5"/>
  <w part="N">ϣϣ</w>
  <lb n="3"/>
  <w part="N">ϣ</w>
  <g ref="#n12"/>
  <w part="N">ϣϣ</w>
  <lb n="4"/>
  <w part="N">ϣϣ</w>
  <g ref="#n6"/>
  <lb n="5"/>
  <w part="N">ϣϣ</w>
  <lb n="6"/>
  <g ref="#n4"/>
  <w part="N">ϣ</w>
  <supplied reason="damage">
    ↓
  </supplied>
  <gap extent="2em"
    reason="lost"
    unit="character"
    dim="right"/>
</ab>
</div>
```

Figure 2: Example of the `Epidoc <div>` element with `type="edition"`.

The EpiDoc DTD introduces a finite set of possible values for the `type` of a `<div>`, so that there

is a clear distinction between sections covering different aspects, such as the commentary, the description or the archaeological history.

One advantage of structured markup is that editors can encode more information about how certain a particular feature is. The date of an inscription, for example, can be encoded as a range of possible dates. EpiDoc includes the TEI `<certainty>` element and the `cert` attribute to encourage editors to say whether or not they are completely confident of a given reading. After some discussion, the EpiDoc community (Mahoney, 2007) decided that certainty should be expressed as a yes-or-no value: either the editor is certain of the reading, or not. Gradual certainty is too complicated to manage and is best explained in the commentary.

## 4 Developing the Linear A Corpus

The hope that computational approaches could help decipher Linear A, along with the evident lack of rich digital resources in this field, led us to develop this new resource. In this section we describe which issues we faced and which solving strategies we used.

### 4.1 Data Collection

Luckily the existence of Younger's website and GORILA volumes, together with the Raison–Pope Index, made possible a semi-automatic collection process, starting from syllabic transcriptions taken from Younger's website (with his permission), converting them in Unicode strings through Python scripts and acquiring all the metadata provided in Younger's transcriptions (location and support IDs, conservation place, periodization etc.).

Younger's resources on his website consist of two HTML pages, one containing inscriptions from Haghia Triada (that is the richest location in terms of documents found there) (Younger, 2000k) and the other containing documents from all the other locations (Younger, 2000l).

Younger's transcriptions are well enriched with metadata. The metadata convey the same information found in GORILA, including the Raison–Pope Index, plus some additional description of the support (this was not necessary in GORILA volumes, where the transcriptions are shown just next to the documents pictures) and the reference to the specific GORILA volume and pages.

## 4.2 Segmentation Issues

When working on ancient writing systems, segmentation issues are expected to come up. John G. Younger explains (Younger, 2000c) that in Linear A separation is mainly indicated in two ways: first, by associating sign groups with numbers or logograms, thereby implying a separation; second, by placing a dot between two sign groups, thereby explicitly separating the sign groups or between a sign group and some other sign like a transaction sign or a logogram. Younger also explains that in texts that employ a string of sign groups, dots are used to separate them and this practice is most notable on non-bureaucratic texts and especially in religious texts.

On his website, Younger also covers the hyphenization issue (Younger, 2000d), explaining that in some cases we find a split across lines and the reason may involve separating prefixes from base words (the root of a sign group) or base words from their suffixes. As Younger points out, this hypothesis would require evidence showing that affixes are involved. The hyphenization issue is more complex to solve because a 'neutral' resource should avoid transcriptions implying a well known segmentation for Linear A sign groups. In Younger's transcriptions, split sign groups are reunified in order to make it clearer when a known sign group is there. Instead, our digital collection keeps the text as it is on the document, all the information about interpretations of such kind can be stored separately.

### 4.3 Obtaining Unicode transcriptions

We managed to obtain Unicode encoded transcriptions by automatically converting Younger's phonetic transcriptions to GORILA transcriptions (manually checked against GORILA volumes) and then by automatically converting GORILA transcriptions to Unicode codes and printing them as Unicode characters (UTF-8 encoding). In order to create the syllables-to-GORILA and the GORILA-to-Unicode dictionaries, we took into account Younger's conversion table mentioned in Subsection 2.4 and the official Unicode documentation (containing explicit Unicode-to-GORILA mapping information). All these processing steps have been implemented through Python scripts.

#### 4.4 XML annotation

Once collected the whole corpus encoded in Unicode, we automatically added part of the XML annotation through a python script. These documents have been later manually corrected and completed, checking against GORILA volumes.

#### 4.5 A new Linear A font

Before the Unicode 7.0 release, there was no way to visualize Unicode characters in the range 10600–1077F. Even now, systems that are not updated may have trouble to visualize those characters. Some implementations for Unicode support in certain contexts (for example for L<sup>A</sup>T<sub>E</sub>X's output) are not always up-to-date, so it is not obvious that the fonts for the most recent characters sets are available. We decided to develop a new Linear A font, solving the main issue found in LA.ttf (wrong Unicode positions). Starting from the official Unicode documentation, we created a set of symbols graphically similar to the official ones and aligned them to the right Unicode positions. We decided to name the font John\_Younger.ttf to show our appreciation for Younger's work. He made the results of GORILA available to a wider public on digital media; this is the same goal we want to pursue by developing and distributing this font. We released the font file at the following URL: <http://openfontlibrary.org/en/font/john-younger>.

### 5 The Linear A Digital Corpus as cultural resource

As stated by European Commission (2015) and UNESCO (2003), the meaning of the notion of *cultural heritage* does not apply just to material objects and works of art, but also to 'intangible cultural heritage', as traditions and creative expressions. In this perspective, linguistic corpora fit perfectly this definition; in fact, they contain information about tradition, knowledge and lifestyle of a certain culture.

Despite the fact that the Minoan language has not yet been deciphered, we know that the Linear A corpus provides interesting information concerning economy, commerce and religion.

As mentioned in Subsection 2.1, Schoep (2002) made a critical assessment of the Linear A tablets and their role in the administrative process, studying the physical supports.

Ruth Palmer (1995) made a deep study of commodities distributions (listing precise quantities and places) among Minoan centers, even without a full understanding of documents contents. As Palmer points out, 'the ideograms for basic commodities, and the formats of the Linear A texts are similar enough to their Linear B counterparts to allow valid comparison of the types and amounts of commodities which appear in specific contexts'. So, it's possible to have 'an idea of the focus of the economy' and of 'the scale and complexity of the transactions'. From the linear A tablets, we can infer information about the resources management and administration system of Minoan centers.

Van den Kerkhof and Rem (2007) analyzed the Minoan libation formulas: religious inscriptions on cups, ladles and tables that were used in the offerings of oil and other powerful drinks at dawn. The priestesses that carried out the Minoan libation ritual used all kinds of utensils, and they often inscribed their sacred formulas onto these objects. Around thirty of these texts have survived (whole or in part) on libation tables, ladles and vases, written in various kinds of handwriting. Transcripts of these religious inscriptions are available from Consani et al. (1999) and from John G. Younger (2000m) on his website. As noticed by Duhoux (1989) the Minoan libation formulas have a fixed structure with variable elements. In fact, some studies (Davis, 2014) about Minoan syntax have been made by observing the sign groups order found in these regular formulas. More importantly, the presence of olive-like ideograms could tell us that the Minoans used olive oil for libation (Van den Kerkhof and Rem, 2007). Beyond all these parts of the Minoan cultural heritage already available, a huge part is preserved there too: the Minoan language, with its hidden stories reflecting the life of a civilization. We hope that our contribution can be useful to the community and that the Minoan, in its digital form, may finally be deciphered through computational approaches.

### 6 Future Work

We are working on XSL style sheets in order to create suitable HTML pages. All the data will be freely available and published at the following URL: <http://ling.ied.edu.HK/~gregoire/lineara>. A further step will be developing a web interface to annotate, and dynamically enrich the corpus information.



## References

- Fernando Luis Álvarez, Elena García Barriocanal, and Joaquín L. Gómez Pantoja. 2010. Sharing Epigraphic Information as Linked Data. In *Metadata and Semantic Research*, pages 222–234. Springer.
- Gabriel Bodard. 2009. EpiDoc: Epigraphic documents in XML for publication and interchange. *Latin on Stone: Epigraphic Research and Electronic Archives, Roman Studies: Interdisciplinary Approaches*, forthcoming.
- Hugh Cayless. 2003. Tools for Digital Epigraphy. *Proc. of the Association for Computing in the Humanities/Association for Literary and Linguistic Computing*, Athens GA.
- Carlo Consani, Mario Negri, Francesco Aspesi, and Cristina Lembo. 1999. Testi minoici trascritti. *Rome: CNR–Istituto per gli studi micenei ed egeo-anatolici*.
- Arthur Cotterell. 1980. *The Minoan World*. Charles Scribner's Sons.
- Brent Davis. 2014. Syntax in Linear A: The Word-Order of the 'Libation Formula'. *Kadmos*, 52, Issue 1.
- Manfried Dietrich and Oswald Loretz. 2001. *In memoriam Cyrus H. Gordon*. Ugarit-Forschungen. Ugarit-Verlag. <https://books.google.it/books?id=tlrvMQEACAAJ>.
- Yves Duhoux. 1989. Le Linéaire A: problèmes de Déchiffrement in Problems in Decipherment. *Bibliothèque des Cahiers de l'Institut de Linguistique de Louvain*, 49:59–119.
- European Commission, 2015. *Supporting cultural heritage - European Commission*. [http://ec.europa.eu/culture/policy/culture-policies/cultural-heritage\\_en.htm](http://ec.europa.eu/culture/policy/culture-policies/cultural-heritage_en.htm).
- Arthur John Evans. 1909. *Scripta minoia*, vol. I. Oxford: Clarendon Press, I.
- Paul Faure. 1998. 8. La Marle (Hubert), Linéaire A. La première écriture syllabique de Crète. Vol. 1: Essai de lecture; Vol. 2: éléments de grammaire. *Revue des Études Grecques*, 111(1):339–340.
- Antonio Enrico Felle. 2011. Esperienze diverse e complementari nel trattamento digitale delle fonti epigrafiche: il caso di EAGLE ed EpiDoc. *Diritto romano e scienze antichistiche nell'era digitale. Convegno di studio (Firenze, 12-13 settembre 2011)*, pages 47–54.
- Elena García Barriocanal, Zeynel Cebeci, Aydın Öztürk, and Mehmet C. Okur. 2011. *Metadata and Semantic Research: 5th International Conference, MTSR 2011, Izmir, Turkey, October 12-14, 2011. Proceedings*. Communications in Computer and Information Science. Springer. [https://books.google.it/books/about/Metadata\\_and\\_Semantic\\_Research.html?id=ydHP9izRsdoC&hl=en](https://books.google.it/books/about/Metadata_and_Semantic_Research.html?id=ydHP9izRsdoC&hl=en).
- Giulio M. Facchetti and Mario Negri. 2003. *Creta minoica: sulle tracce delle più antiche scritture d'Europa*. LS Olschki.
- Louis Godart and Jean Pierre Olivier. 1976. *Recueil des inscriptions en linéaire A*. Librairie Orientaliste Paul Geuthner.
- Joaquín L. Gómez Pantoja and Fernando Luis Álvarez. 2011. From relational databases to Linked Data in Epigraphy: Hispania Epigraphica Online, Elena García Barriocanal, Zeynel Cebeci, Mehmet C. Okur and Aydın Öztürk (eds.), *Metadata and Semantic Research (Proceedings of the 5th International Conference, MTSR 2011, Izmir, Turkey, October 12-14, 2011)*. *Communications in Computer and Information Science*, 240.
- Erik Hallager. 1996. *The Minoan roundel and other sealed documents in the neopalatial Linear A administration*, volume 14. Université de Liège, Histoire de l'art et archéologie de la Grèce antique.
- Richard Janko. 1982. A stone object inscribed in Linear A from Ayios Stephanos, Laconia. *Kadmos*, 21(2):97–100.
- Anne Mahoney. 2007. Electronic Textual Editing: Epigraphy. *TEI-Text Encoding Initiatives*. [http://www.tei-c.org/About/Archive\\_new/ETE/Preview/mahoney.xml](http://www.tei-c.org/About/Archive_new/ETE/Preview/mahoney.xml).
- Walter Müller. 1999. Die Tonplomben und andere gestempelte Tonobjekte. I. Pini (ed.), *Iraklion Archäologisches Museum, Teil 6, Die Siegelabrücke von Aj. Triada und anderen Zentral- und Ostkretischen Fundorten. CMS (Corpus der Minoischen und Mykenischen Siegel) II,6*. Berlin, pages 339–400.
- Walter Müller. 2002. Untersuchungen zur Typologie, Funktion und Verbreitung der Tonplomben von Knossos. I. Pini (ed.), *Iraklion Archäologisches Museum, Teil 8, Die Siegelabrücke von Knossos. CMS (Corpus der Minoischen und Mykenischen Siegel) II,8*. Berlin, pages 24–93.
- Gregory Nagy. 1963. Greek-like Elements in Linear A. *Greek, Roman, and Byzantine Studies*, 4(4):181–211.
- Gareth Owens. 1999. The structure of the Minoan language. *Journal of Indo-European Studies*, 27(1-2):15–55.
- David W. Packard. 1974. *Minoan Linear A*. University of California Press. <https://books.google.it/books?id=vax3kwoscWQC>.
- Leonard Robert Palmer. 1958. Luvian and Linear A. *Transactions of the Philological Society*, 57(1):75–100.

- Ruth Palmer. 1995. Linear A commodities: a comparison of resources. *POLITEIA. Society and State in the Aegean Bronze Age, Aegaeum*, 13:133–156.
- Jacques Raison and Maurice Pope. 1971. *Index du linéaire A.*, volume 41 of *Incunabula Graeca*. Edizioni dell'Ateneo. <https://books.google.it/books?id=N10eYAAACAAJ>.
- Jacques Raison and Maurice Pope. 1994. *Corpus transnuméré du linéaire A.* BCILL (Louvain-la-Neuve). Isd. <https://books.google.it/books?id=TbgcAQAAIAAJ>.
- Colin Renfrew. 1977. A Linear A tablet fragment from Phylakopi in Melos.(with a note on the inscription by William C. Brice). *Kadmos*, 16(2):111–119.
- Andrew Robinson. 2009. *Writing and script: a very short introduction*. Oxford University Press.
- Ilse Schoep. 2002. The administration of neopalatial Crete: a critical assessment of the Linear A tablets and their role in the administrative process. *Minos: Revista de filología egea*, (17):1–230.
- UNESCO, 2003. *What is meant by "cultural heritage"?* <http://www.unesco.org/culture/ich/index.php?lg=EN&pg=00022>.
- Astrid Van den Kerkhof and Peter Rem. 2007. The Minoan libation formulas.
- John G. Younger. 2000a. Linear A texts in phonetic Transcription: 1. List of Linked Files, Concordance: Raison-Pope-GORILA signs (a Microsoft Word document). [http://www.people.ku.edu/~jyounger/LinearA/RAISON-GORILA\\_SIGNS.doc](http://www.people.ku.edu/~jyounger/LinearA/RAISON-GORILA_SIGNS.doc).
- John G. Younger. 2000b. Linear A texts in Phonetic Transcription: 10c. Place Names. <http://www.people.ku.edu/~jyounger/LinearA/#10c>.
- John G. Younger. 2000c. Linear A texts in Phonetic Transcription: 16. Word Separation. <http://people.ku.edu/~jyounger/LinearA/#16>.
- John G. Younger. 2000d. Linear A texts in Phonetic Transcription: 17. Hyphenization. <http://people.ku.edu/~jyounger/LinearA/#17>.
- John G. Younger. 2000e. Linear A texts in phonetic Transcription: 4. Conventions (bibliographical, epigraphical). <http://people.ku.edu/~jyounger/LinearA/#4>.
- John G. Younger. 2000f. Linear A texts in Phonetic Transcription: 5. Basic Statistics. <http://www.people.ku.edu/~jyounger/LinearA/#5>.
- John G. Younger. 2000g. Linear A texts in Phonetic Transcription: 7c. The Documents. <http://people.ku.edu/~jyounger/LinearA/#7>.
- John G. Younger. 2000h. Linear A Texts in Phonetic Transcription: A map of Crete showing the location of the cretan sites. [http://people.ku.edu/~jyounger/pix/LinearA\\_Crete.jpg](http://people.ku.edu/~jyounger/pix/LinearA_Crete.jpg).
- John G. Younger. 2000i. Linear A Texts in Phonetic Transcription: A map of Greece showing the location of the greek sites outside Crete. [http://people.ku.edu/~jyounger/pix/LinearA\\_Greece.jpg](http://people.ku.edu/~jyounger/pix/LinearA_Greece.jpg).
- John G. Younger. 2000j. Linear A texts in phonetic Transcription: Homepage. <http://www.people.ku.edu/~jyounger/LinearA/>.
- John G. Younger. 2000k. Linear A texts in phonetic Transcription: HT (Haghia Triada). <http://people.ku.edu/~jyounger/LinearA/HTtexts.html>.
- John G. Younger. 2000l. Linear A texts in phonetic Transcription: Other Texts (not Haghia Triada). <http://people.ku.edu/~jyounger/LinearA/misctexts.html>.
- John G. Younger. 2000m. Linear A Texts in Phonetic Transcription: Presumably Religious Inscriptions. <http://www.people.ku.edu/~jyounger/LinearA/religioustexts.html>.