CRMtex

A CIDOC CRM extension for modelling Ancient Textual Entities

CIDOC CRM SIG – February 2020 - Athens

Francesca Murano, Achille Felicetti
DILEF, University of Florence, Italy
VAST-LAB, PIN
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

CRMtex: a bit of history …

- 2015: EAGLE Project, Workshop on Epigraphy and CIDOC CRM in Cyprus
  - First idea of creating a CIDOC CRM extension for Epigraphy
- 2016: CRMepi, first attempt to define classes and properties for describing epigraphic entities
  - Inspired by "VBI ERAT LVPA" and "EAGLE" initiatives
- 2017: From CRMepi to CRMtex
  - Concept of "Ancient Text" valid for epigraphy but also for other fields (e.g. papyrology)
- 2017-2019: CRMtex presented to scholars
  - Various publications | Various SIG meeting presentations
  - Epigraphy.info initiative for integrating epigraphic data
  - ARIADNEplus project for "Inscriptions" application profile
- August 2019: Workshop in Plakias (Crete) for improving the model
  - Outcomes of the discussion presented here
- February 2020: version 1.0 of CRMtex available for review

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

What is a text

- Is a linguistic expression, independent from any medium, and intentionally used to communicate
  - A text is based on the correlation between an expression and a content

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Speaking and writing

Every speech can be transposed into an equivalent written message; the process is reversed when a reading occurs

«Spoken words are symbols of mental experience and written words are symbols of spoken words»

(Aristotle, On interpretation, I,1)

«A language and its written form constitute two separate systems of signs. The sole reason for the existence of the latter is to represent the former»

(Ferdinand de Saussure, CLG, 52)

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Speaking and writing

- Speech has a priority over writing, at least in four respects (according to J. Lyons):
  - *phylogenetic*: all languages are spoken but not necessarily written
  - *ontogenetic*: every human being learns to speak naturally and spontaneously; the ability to write comes only later and through specific training
  - *functional*: the spoken language is used in a wider and differentiated range of uses and functions
  - *structural:* writing originated as a representation of speech

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

CRMtex: an extesion for ancient texts

- Ancient Texts are documented only by written medium
  - We have no living speakers
  - We have various types of text: inscription, papyri, etc.
  - In respect to modern written text, an ancient texts is the product of manual work rather than a mechanised process (as in modern printing)
  - Special relationship between the text and its support
- **Uniqueness of the ancient text** → Even in case of texts written by same person on identical media and with an identical technique, the resulting copies are never identical
  - Mechanised process in antiquity: coins, medal stamps, seals
    - Close relation between text and support still important

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Texts and interpretation

- We consider a text as a **"semiotic feature"**, **i.e.** a number of marks physically traced on a support and intended to encode a linguistic expression
  - The intentionality is fundamental to define a text
  - Only in this case we speak of written communication
    - Accidental scratches or tool marks does not bring any semiotic information, thus they are not texts

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

The meaning of "text"

- **Physical manifestation**
  - A set of physical features (marks) on a given support, produced using specific techniques
    - Scribbled with ink, painted, engraved etc.
- **Abstract dimension**
  - Physical features to represent sets of concepts/ideas

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Written Text and Writing Activity

- **TX1 Written Text.** Subclass of *E25 Man-Made Feature*
  - Visible set of symbols (*glyphs*) intentionally traced (i.e. "written") on some kind of physical support by using specific techniques
  - Semiotic significance and the intentional purpose of conveying a specific message towards a given addressee or group of addressees

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Writing

- **"Writing"**
  - A sophisticated human technology allowing the encoding of a text in a specific language through series (or sequences) of **signs** specifically selected for this purpose

- **TX2 Writing**. Subclass of *E12 Production*
  - Activity of creating textual entities using various techniques (painting, sculpture, etc.) and by means of specific tools on a given physical support in a non-mechanical way

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

How does writing process work?

- Languages are reduced to written form by codifying the expression and/or the content by means of a system of signs, i.e. a writing system
- *«In reducing a language to writing, that is, in making visible marks that evoke or recall linguistic performance, it would seem that each mark must represent a [linguistic unit, that is a] syntagmeme or a lexeme or a morpheme or a phoneme or whatever other kind of unit the inventor of the system may chose as his basis»* (E. Pulgram 1976)

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

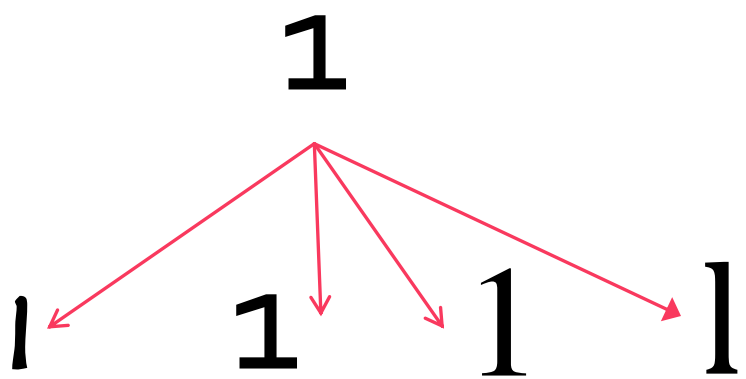
Writing Systems

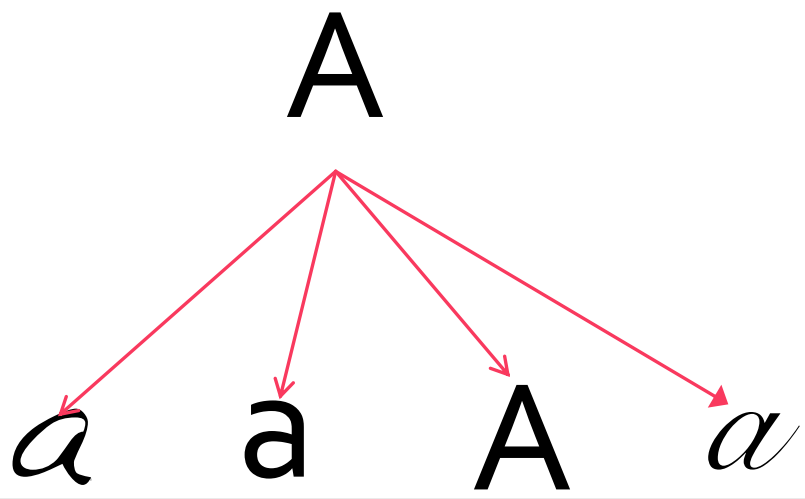
- **TX3 Writing System.** Subclass of E29 Design or Procedure.
  - Used to produce a TX1 Written Text during a TX2 Writing event
  - A conventional system consisting of a set of signs (graphemes) used to codify a natural language, by means of specific rules in the combination and phonological value assignment of the chosen graphemes
  - Phonetic associations result from historical processes and can vary throughout the history of the language for which they are used (see spelling rules)
  - As a code, a writing systems requires shared understanding between writers and readers, allowing the transmission of the message

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Glyphs and Graphemes

- Is important to distinguish the concrete level of the personal execution (i.e. the real act of tracing marks on a surface) from the abstract level which all the single occurrences must be took back to, on the basis of a sameness principle
- The signs used to encode / decode a message (e.g. the letters) are the concrete expressions (physical features - **glyphs**) of conceptual and conventional elements a linguistic community chose to encode its language (**graphemes**)





francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Glyphs and Graphemes

- **Graphemes**: the minimal functional distinctive units of a writing system
- **Glyphs**: physical signs composing a written text, constituting the material manifestations of writing system units, i.e. the graphemes
- In alphabetic writing systems, a grapheme corresponds to a letter
  - In a Latin inscription, single alphabet letters (glyphs) represent graphemes
- In Mycenaean Linear B inscriptions and in Old Persian cuneiform inscriptions glyphs represent syllabograms (graphemes representing a syllable, not a single sound)
- In an Egyptian hieroglyphic text, glyphs represent syllabic, alphabetic and also ideographic elements, i.e. elements standing for lexical/semantic units

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Glyphs and Graphemes

- Phonographic writing systems represent phonological units of different size
- The 1:1 correspondence between sound (phoneme, syllables, etc.) and sign (grapheme) is lost in diachrony
  - Spelling conventions and phonetic changes to which linguistic systems are subjected in history
- Examples of spelling discrepancies between writing and reading in English:
  - the <i> grapheme stands for various phonemes: /ɪ/ (as in him), /ʌɪ/ (as in time), /i/ (as in police), /a/ (as in timbre);
  - the /f/ phoneme can be represented by <f> (as in film), <ph> (as in philology), <gh> (as in enough) …

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Ferdinand de Saussure

- "**Writing obscures language**; it is not a guise for language but a disguise … Whoever says that a certain letter must be pronounced a certain way is mistaking the written image of a sound for the sound itself …
- For French *oi* to be pronounced *wa*, this spelling would have to exist independently; actually *wa* is written *oi.* To attribute the oddity to an exceptional pronunciation of *o* and *i* is also misleading, for this implies that language depends on its written form and that certain liberties maybe taken in writing, as if the graphic symbols were the norm" (FdS, CLG, 52)

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Glyphs

- **TX9 Glyph.** Subclass of E25 Man-Made Feature
- To represent the concrete manifestation of single signs traced by the writer while codifying a linguistic expression
- Glyphs are typically observed by the scholar during a reading activity (TX5) carried out to decode and recognise the graphemes (TX8) they represent
- Examples:
  - The S-shaped feature engraved on the second line of the South inscription on the Arch of Constantine, representing the letter (grapheme) "S" of the Latin writing system used to render the phoneme /s/ of Latin language
  - The feature  engraved on the first line of Darius I's inscription (TX1) in Bagistan, representing the ideal syllabogram 'da' of the ancient Persian syllabary, used to render the /da/ syllable of ancient Persian language

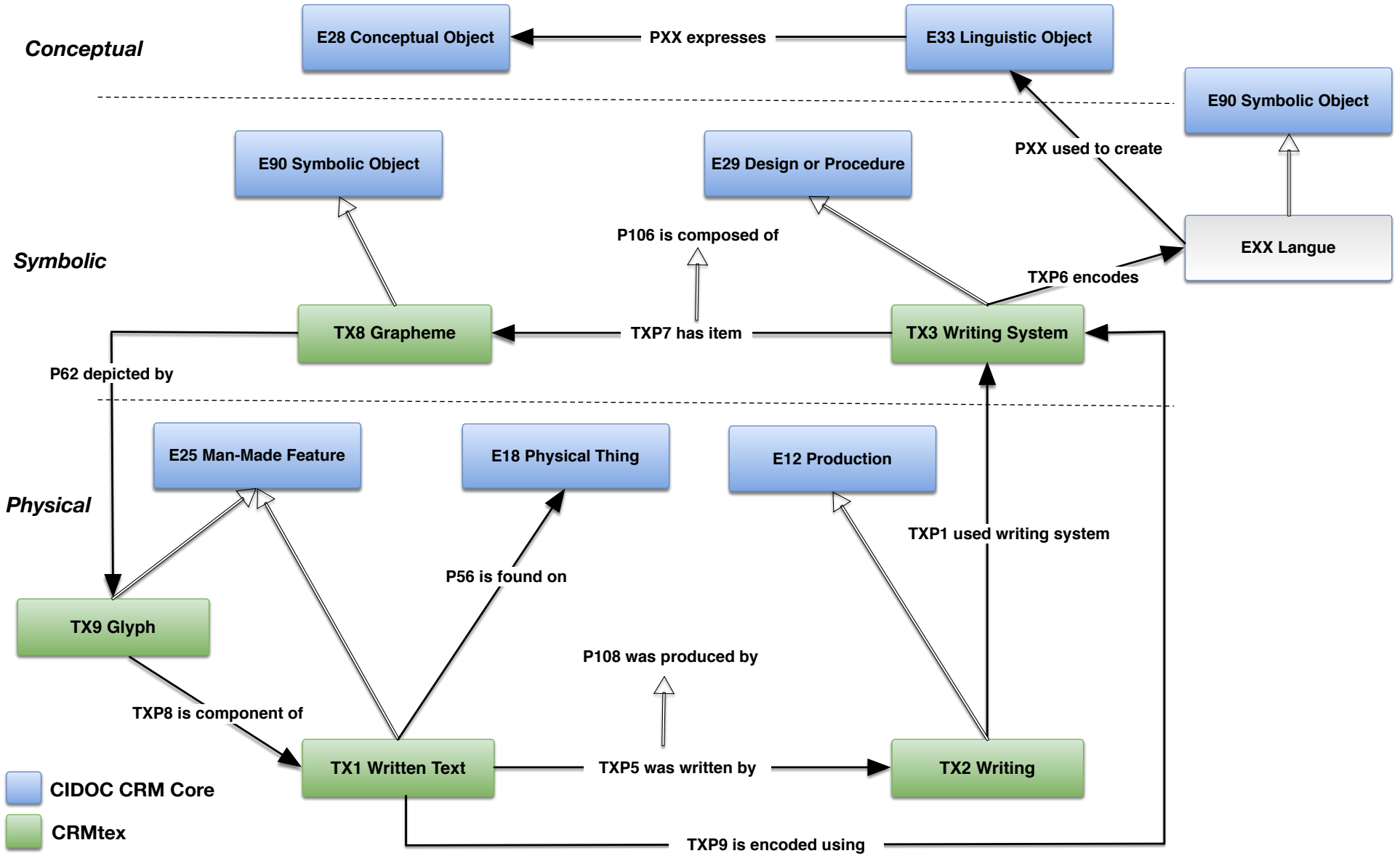
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Graphemes

- **TX8 Grapheme.** Subclass *E90 Symbolic Object*
- Used to represent the abstract units with distinctive value in a given writing system
  - In an alphabetic system, these units are essentially the abstract letters of the alphabet in question
- Examples
  - The ideal letter "S" of the Latin alphabet, used to represent the /s/ sound, rendered by the specific S-shaped feature engraved on Latin inscriptions
  - The ideal 'da' syllabogram of the ancient Persian syllabary, used to represent the /da/ syllable rendered by the cuneiform sign  engraved on the first line of Darius I's inscription in Bagistan

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

CRMtex: Text and Writing



francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

New class and property proposals

- Current class E56 Language -> subclass of E55 Type
- Proposal: **EXX Langue** -> subclass di E90 Symbolic Object
- Ideas for a tentative scope note:
  - a) A structured system of communication, based on a process of semiosis relating signs to particular meanings
    - Languages are codes (systems of signs) allowing the expression of the thinking

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

New class and property proposals

- b) Human languages are specific and different actualized manifestations of the human language faculty
  - As such, human languages are historical-natural, i.e. not artificially created by humans. They are part of the life and growth of the individual and the community, inevitably subject to evolve and diversify over time (W. von Humboldt)
- c) Human languages are historically determined systems, through which members of a community express themselves and communicate with each other. The passage of information implies shared knowledge of the code among the participants in the communication.

- **PXX expresses**: to state that a specific E33 Linguistic Object is used to express a certain E28 Conceptual Object

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Physical arrangement of text

- **TX4 Writing Field.** Subclass of *E25 Man-Made Feature*
- The surface or portion of the physical carrier reserved, delimited and arranged for the purpose of accommodating a written text
- Used to highlight and isolate text from the other parts of the object to which it belongs, to enhance and guarantee its readability
- Fundamental in Epigraphy (*Epigraphic field*)
- Distinction between area(s) containing written text and empty parts of the support (margins, *intercolumnia*, etc.)
- Geometric properties of text (ductus, disposition, geometric distances): to be further investigated

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Message decoding

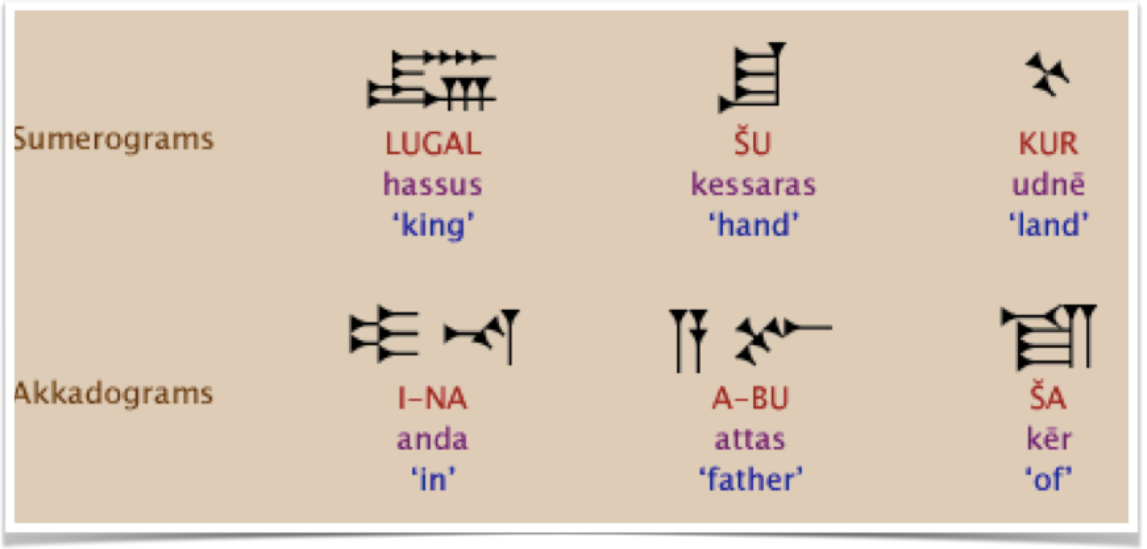
- Message retrieval:
  - reading the written message presupposes the ability to read the language of the writer
  - diachronic activity for ancient texts
- The decoders (readers, including scholars) will attribute to each sign or group of signs the adequate phonetic value
  - on the basis of the knowledge of the linguistic system
    - on the basis of spelling conventions in place at a given historical moment
  - Also in case of numbers and other symbols, readers assign phonetic values in the language is currently using

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Inheriting from other languages

Sumerograms and Akkadograms in Hittite inscriptions

- The Hittite writing system is a syllabary (every cuneiform sign represents a syllable)
- There are cuneiform signs inherited from Akkadic and Sumerian, called Akkadograms and Sumerograms
- These signs work more or less as Arabic numbers in our writing systems
- They consist of Sumerian or Akkadian words, written according to their writing systems, used to express Hittite words
- "Graphic loans" not linguistic loans



francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Reading

- **TX5 Reading.** Subclass of the CRMsci *S4 Observation*
  - Semiotic procedure of decoding (and therefore understanding) a written text
  - A specific observation (S4) in which the decoding of the signs is performed, i.e. the linguistic value is recognised and the message is understood.
  - In cases in which decoding does not happen (e.g., the observer is able to describe the signs but not to assign a specific linguist value to them), the S4 class could be used as it is.
  - Carried out for scientific purposes, in order to analyse and study the text according to different disciplinary perspectives.
  - For study purposes, the reading procedure requires a scientific autoptic examination of the text
    - An accurate analysis of the surface and the signs and prescribes the use of specific tools and procedures, for establishing as faithfully as possible the exact value of each sign drawn on the physical feature.

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Transcription

- **TX6 Transcription.** Subclass of *E7 Activity*
- Activity of re-writing the text conducted by an editor.
- Could involves a writing system (TX3) different from that of the original text, implying a transposition of the sounds of a language from a writing system to another one (=re-encoding)
  - E.g. Latin letters to render the Linear B (Mycenaean syllabary)
- For scientific purposes, this operation often properly consists in a 'transliteration'
  - Implies a 1:1 relation between the signs of the two writing systems

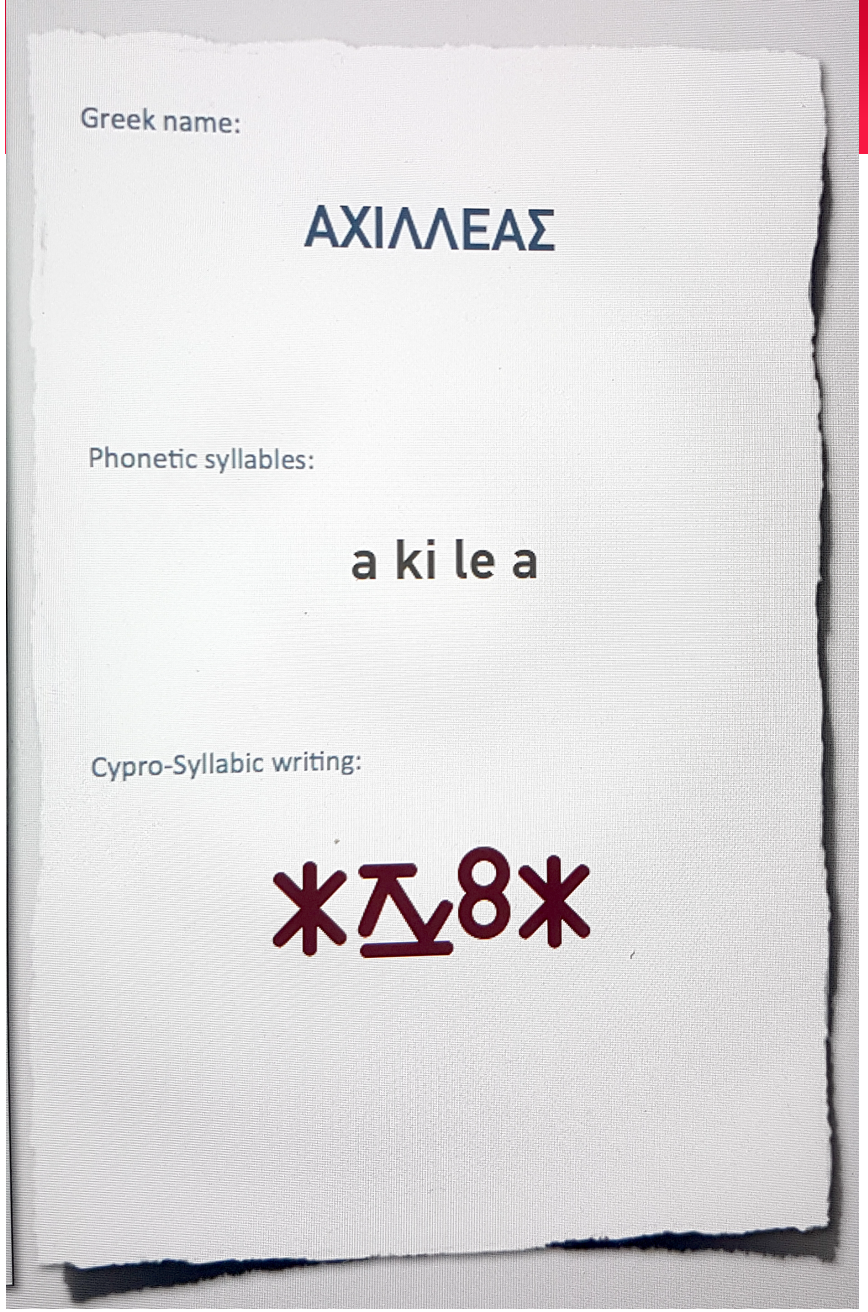
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Transcription

- A transliteration is never ambiguous
- Transcription depends on writing system used
  - The name of Thebes is written Θήβα in Greek alphabet;
  - The sequence has transliteration 'Thēba' in Latin script but has transcription 'Thìva', according with modern Greek pronunciation
  - notice that transcription is based on phonetics, thus pronunciation problems can arise:
    - an English speaker might read 'Thìva' as [ˈθʌɪva] instead of [ˈθiva], and possibly transcribe 'Theeva'

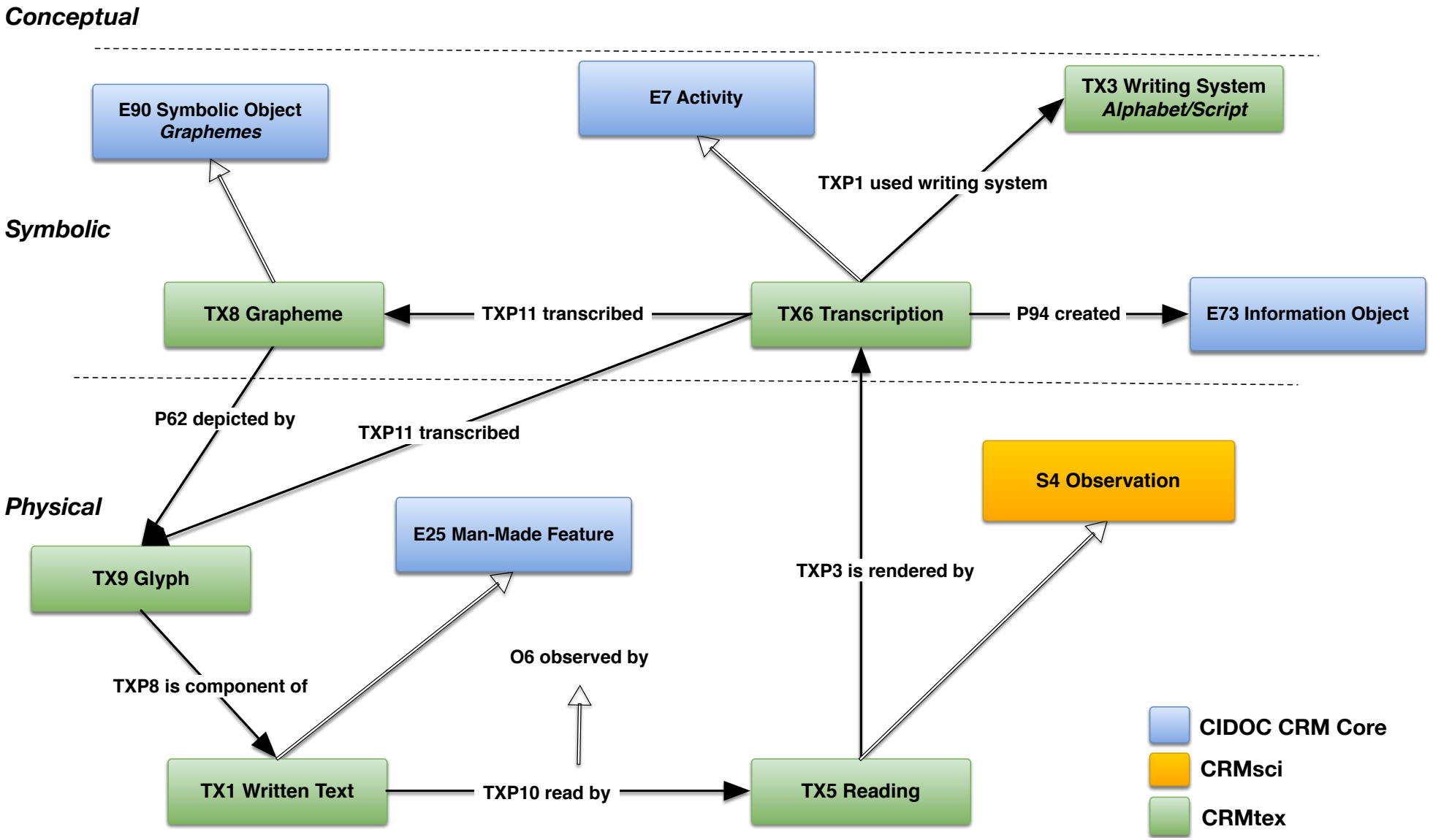
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Transcription



francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Reading and Transcription



francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Text and Text Segments

- Segmentation of the text has various purposes:
  - 1. To describe the layout of the text (columns, pages, etc.)
    - Decided by the writer during a TX2 Writing event for communication
  - 2. To describe the the material condition of a part of the text
    - Operated by scholars during a TX5 Reading event for study
    - Related to the physical feature scholars observe (TX1 Written Text)
  - 3. To analyse the content of the text
    - Operated by scholars during a TX5 Reading event for study
    - Related to the meaning, i.e. the E33 Linguistic Object,
    - Editorial interventions on the text
      - E.g.: **C IULIUS CAEESAR → C(aius) Iulius Cae{e}sar**

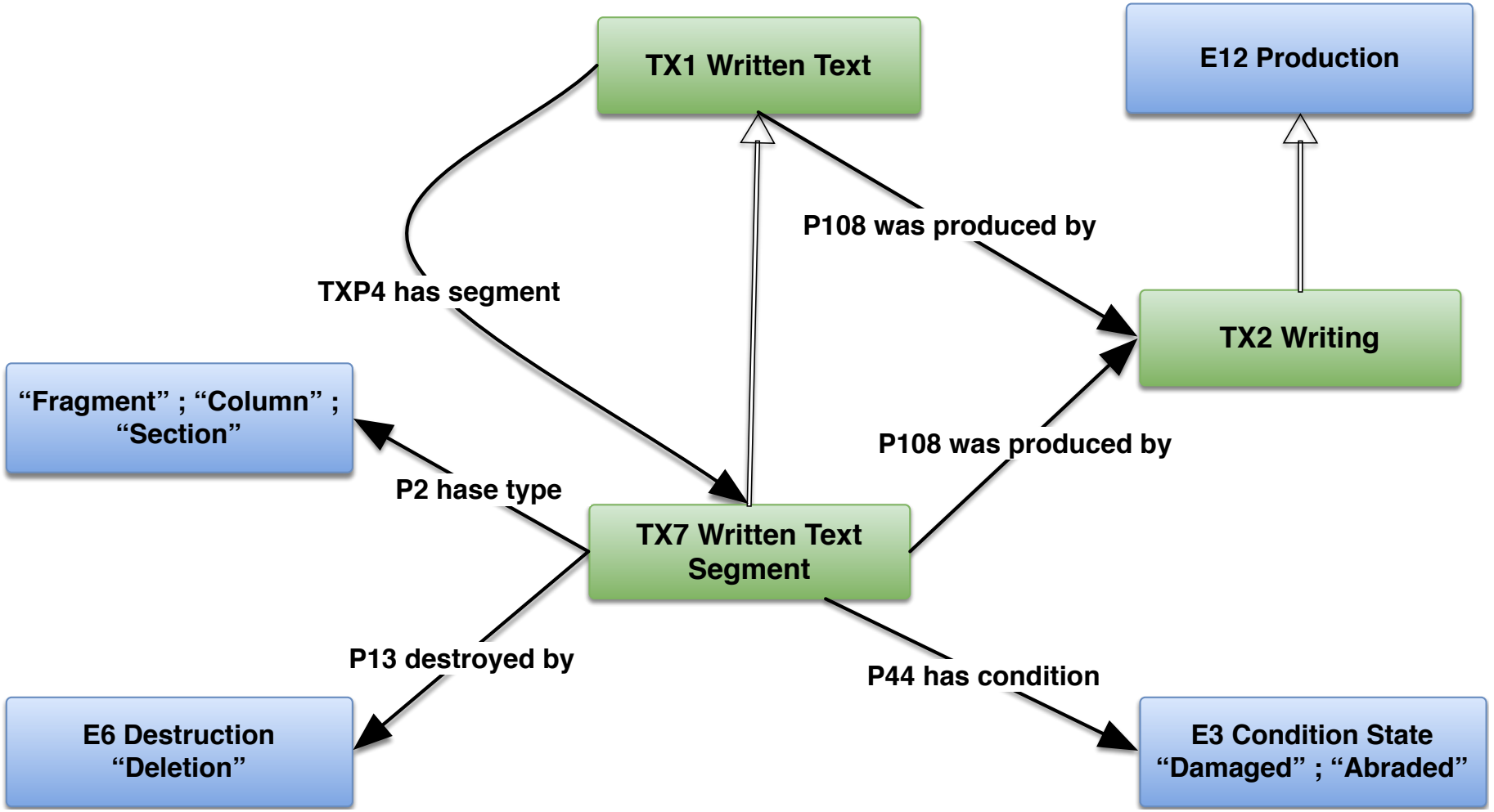
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Text and Text Segments

- **TX7 Written Text Segment.** Subclass of TX1 Written Text
- To identify portions of text considered to be of particular significance by scholars, as witnesses of a certain meaning or bearers of a particular phenomenon relevant to the investigation, study and understanding of the ancient text.
  - columns, fragments, sections, paragraphs, single words or letters, or other components of the written text.
- A single production event (TX2) or destruction event (E6) could be associated to each fragment
  - letters or words damaged or worn by atmospheric agents or human interventions
- Specific conditions (E3) for documenting its status during the observation process (S4).

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

Text segments and physical conditions

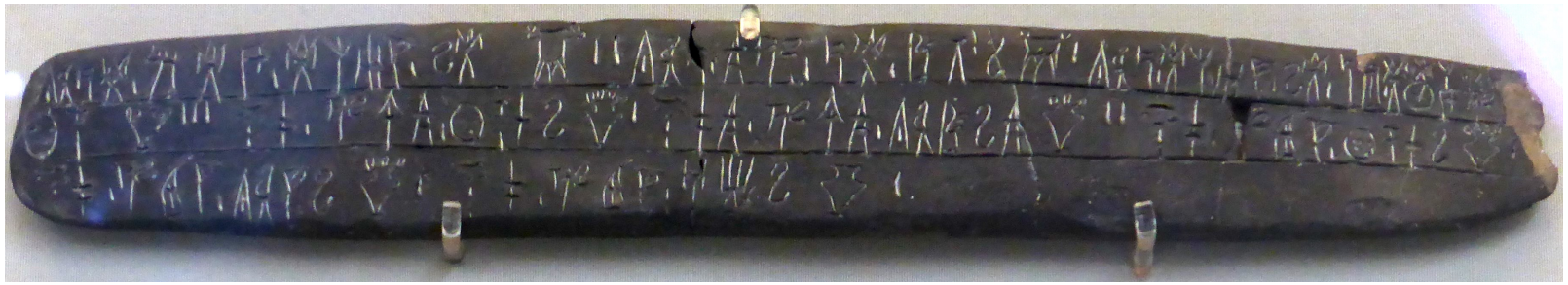


francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

An example: the Tripod Tablet from Pylos

Linear B inscription on a clay tablet (PY TA 641)

- Discovered in the Archive room of the Mycenaean palace at Englianos (Pylos, Messenia, Greece)
- National Archaeological Museum Athens
- Date: Late Bronze Age, 13th c. B.C.
- Three lines of text, separated by two horizontal lines
- Ductus: from left to right
- Deciphered in 1952 by Michael Ventris
- The identification of the ideograms to the corresponding ancient Greek words, proved that Linear B was a form of Greek



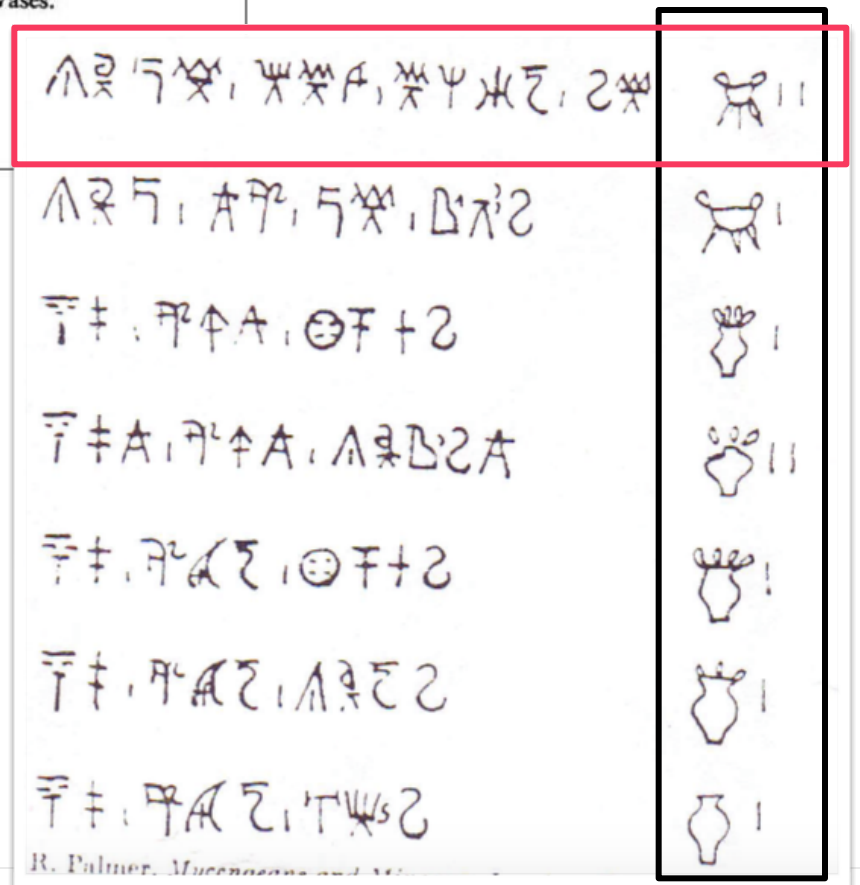
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

VAST - LAB

Pylos Tablet PY 641-1952 (Ventris)

'Two tripods - Aigeus the Cretan (?) brings them. 2 Tripods.' 'Two big three-handled vases. 2 Vases.'

with one possible translation of Line 1, with which I disagree in part. Richard



- List of object brought or present in the Palace
- It uses syllabic characters together with ideograms of vases
- List of cooking vessels, such as the grill (eschara, e-ka-ra), bowl (phiale, pi-je-ra), brazier (pyraustro, pu-ra-u-to-ro) and shovel (cho[s]teria, ko-te-ri-ja),
- The mentioned tripod (tripous, ti-ri-po), gives this tablet its name

ti

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

VAST - LAB

Pylos Tablet PY 641-1952 (Ventris)

'Two tripods - Aigeus the Cretan (?) brings them. 2 Tripods.' 'Two big three-handled vases. 2 Vases.'

with one possible translation of Line 1, with which I disagree in part. Richard

Transliteration of the first segment (in Latin script)

ti-ri-po-de a3-ke-u ke-re-si-jo we-ke *201VAS 2

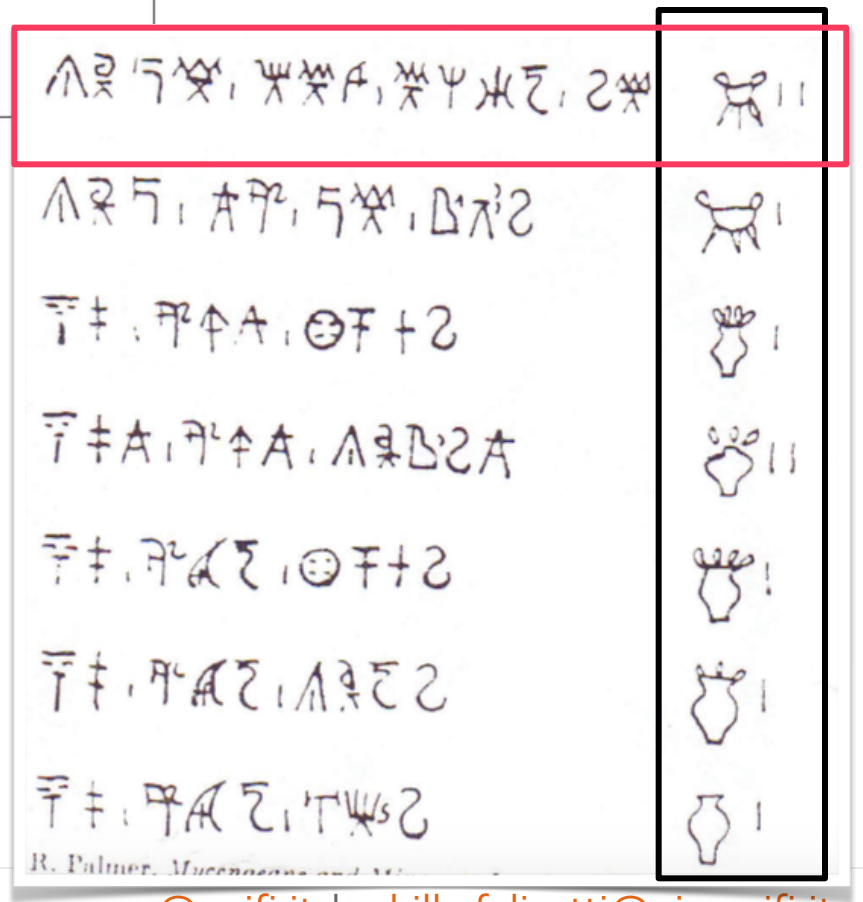
Transcription (in Latin script)

tripode aigeus krēsjowergēs *201VAS 2

Translations

"Two tripod-cauldrons of the Aikeu type" (Ventris)

"Two tripods goat-shaped of Cretan workmanship" (current translation)



francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

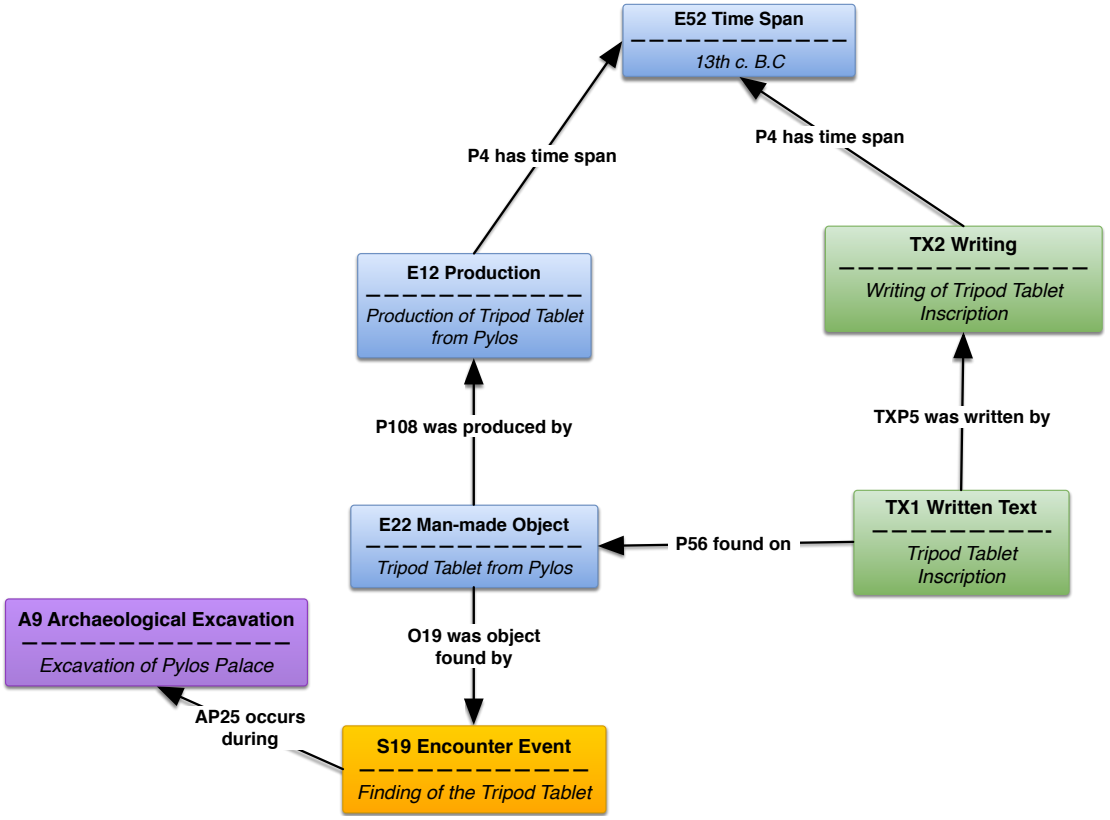
Symbolic signs in text

- Apart the TX8 Grapheme class, other classes (subclasses of E90 Symbolic object) are maybe needed to describe symbolic signs in texts
- On the basis of the semiotic nature of its instances and according to FdS and Peirce regarding way of denoting objects, we can identify:
- **Icons**: conventional signs that resembles what they `stand for' in the reality (referent)
  - e.g. the icon of the vase in the Tripod Tablet inscription

- **Symbols**: signs with a higher degree of conventionality and arbitrariness
  - The dove symbol to signify the peace in Latin inscriptions
  - Paragraphemes, i.e. diacritic signs (commas, dots, prosodic pitches, …)

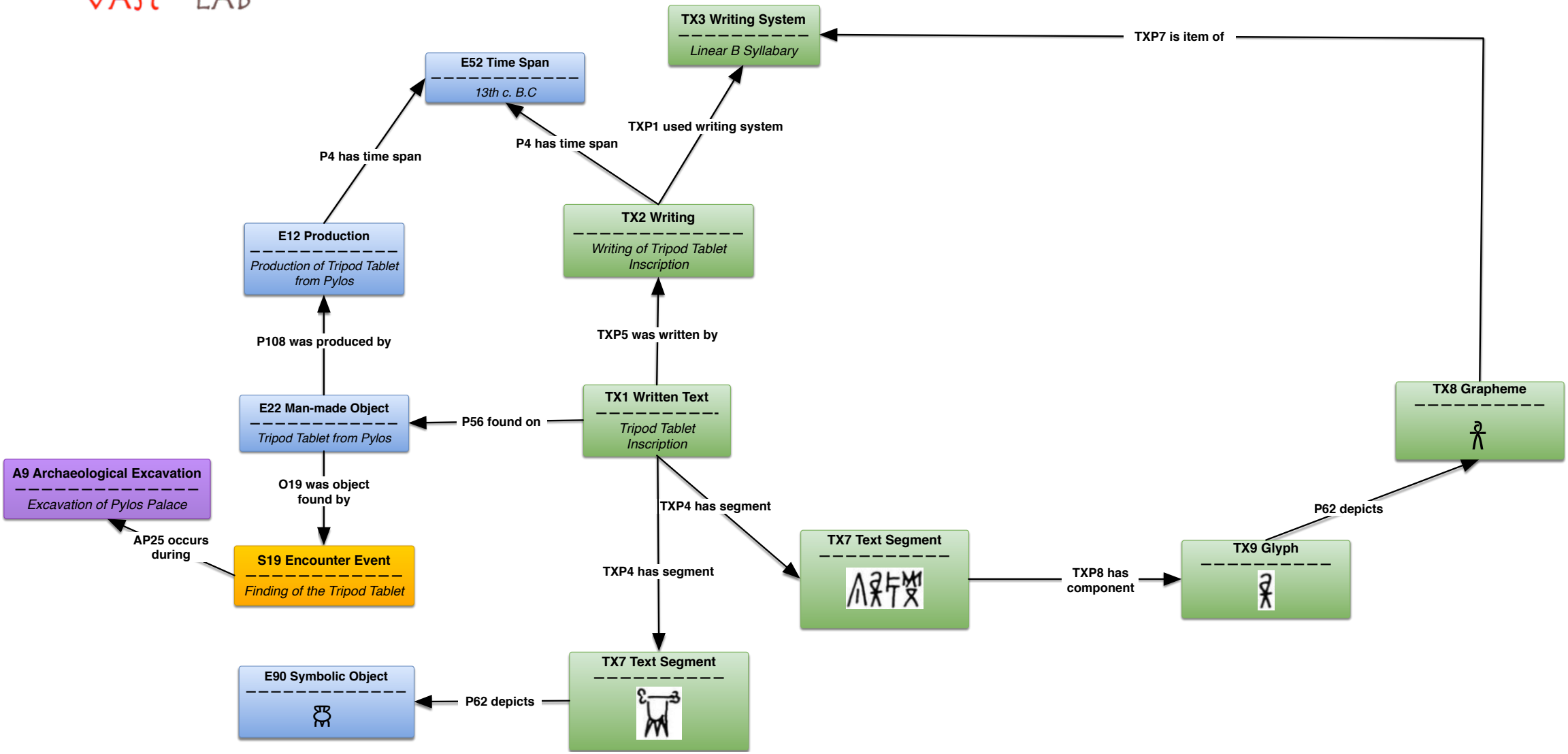
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

PY TA 641: Text and support



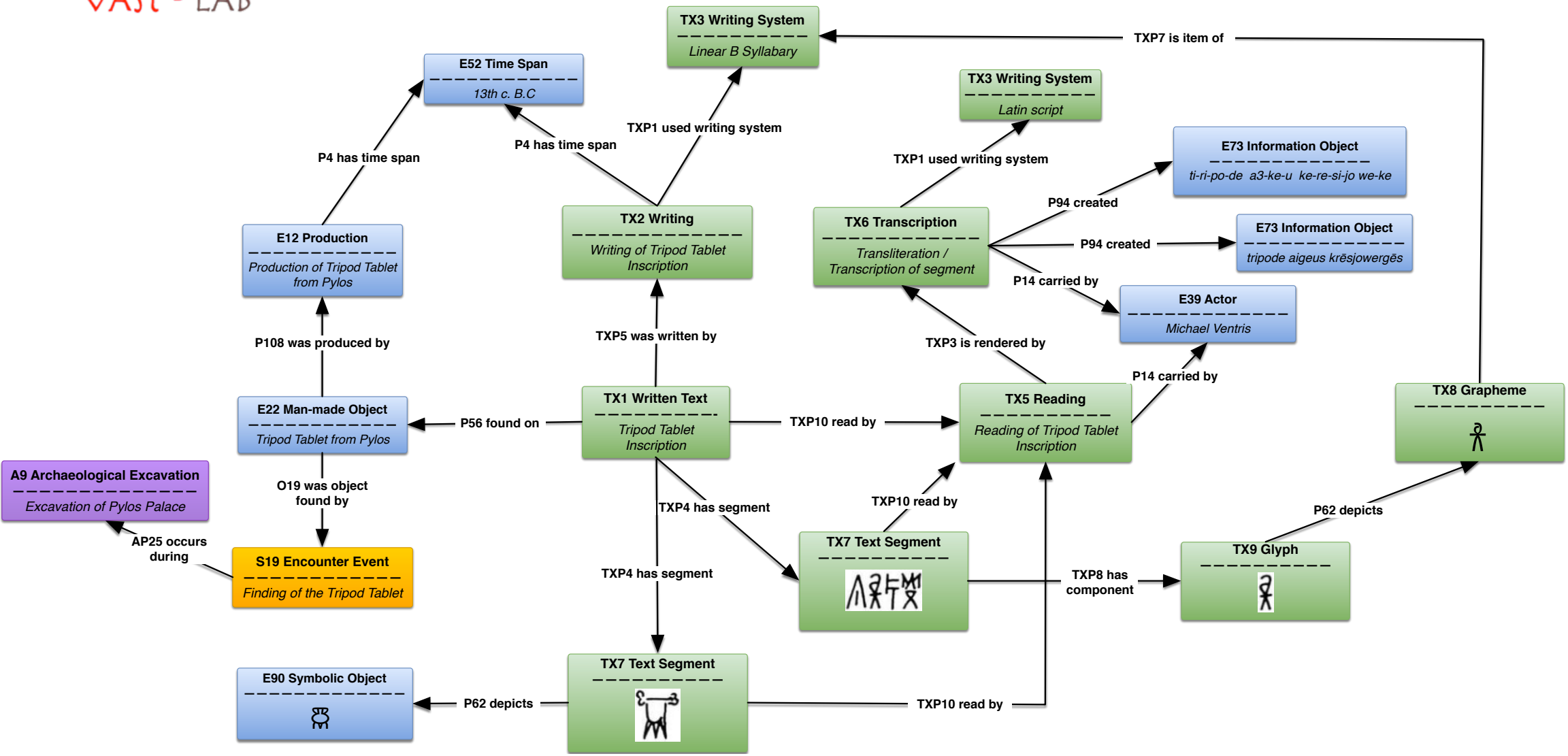
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

PY TA 641: composition and segments



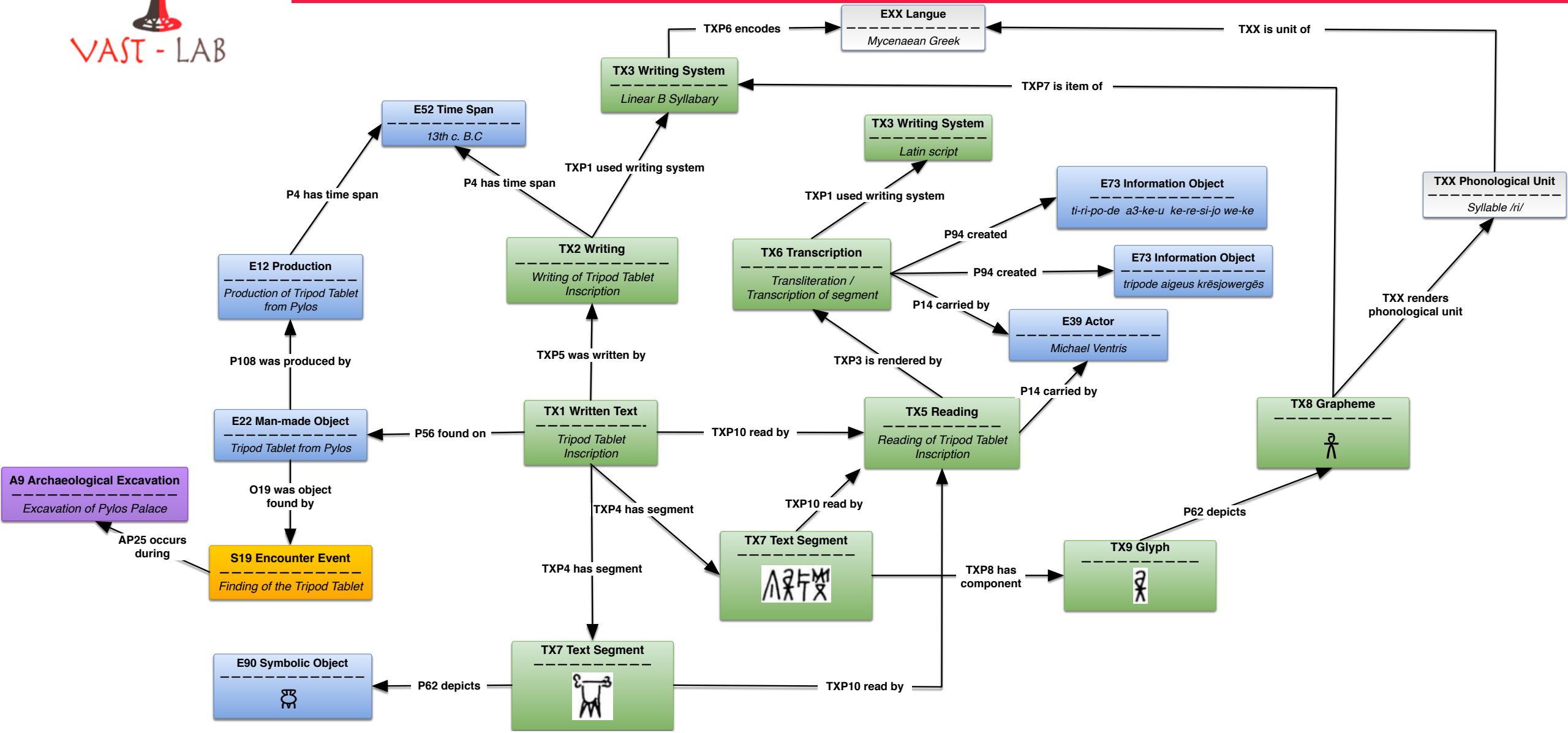
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

PY TA 641: Text reading and transcriptions



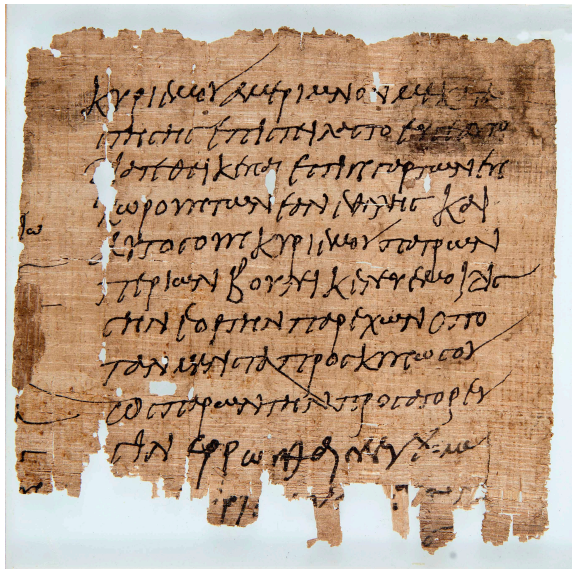
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

PY TA 641: Phonology and langue



francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

CRMtex documentation and initiatives



- Publications:
  - A. Felicetti, F. Murano, P. Ronzino, F. Niccolucci: *CIDOC CRM and Epigraphy: a Hermeneutic Challenge*, EMF-CRM@ TPDL, 2015, http://ceur-ws.org/Vol-1656/paper5.pdf
  - A. Felicetti, F. Murano, *Scripta manent: a CIDOC CRM semiotic reading of ancient texts*, Int. Journal of Digital Libraries, 2017 https://doi.org/10.1007/s00799-016-0189-z
  - A. Felicetti, F. Murano, *Ce qui est écrit et ce qui est parlé. CRMtex for modelling textual entities on the Semantic Web*, paper submitted to Semantic Web Journal, November 2019
- Documentation: http://www.cidoc-crm.org/crmtex/ (v1.0 - available for review soon after SIG)
- Initiatives:
  - CRMtex as component of Epigraphy.info ontology (under development)
  - CRMtex candidate to be used in ARIADNEplus as Application Profile for Inscriptions (task 4.4.13)

francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

PIN
POLO UNIVERSITARIO CITTÀ DI PRATO

Thank you …

Francesca Murano, Achille Felicetti
DILEF, University of Florence, Italy
VAST-LAB, PIN
francesca.murano@unifi.it | achille.felicetti@pin.unifi.it

VAST - LAB