

Sunoikisis Digital Classics

Session 18

Text and digital editing Geography and Lexicography

June 1, 2016

Stylianos Chronopoulos (Freiburg)
Tariq Yousef (Leipzig)
Chiara Palladino (Bari/Leipzig)

Syllabus

- Recap: XML and digital editions; a gentle introduction to TEI as a markup language
- Modules for technical sources
- Dictionaries as a genre: specificities and peculiarities (and problems)
- Markup strategies for lexicography
- Geography as a genre: specificities and peculiarities (and problems)
- Markup strategies for digital editions of geographical texts
- A simple practical application: Transforming xml documents: indexes of place names
- Importing georeferenced PlaceNames through automatic extraction: a deeper investigation in QGIS (visualizing relations to reference points; visualizing different categories of places)
- Dynamic Lexica:
 - General introduction: what are Dynamic Lexica?
 - An example of dynamic lexicon from the ancient world

Recap: XML and Digital editions

Why use digital publications?

Texts in machine-readable format:

- Make semantic distinctions (*emphasis*, *<i>italics</i>*) and degrees of meaning (*<foreign>foreign languages</foreign>*, *<quote>quotations</quote>*) explicit and reusable
- Allow multiple outputs by transforming an XML document with XSLT: you can have HTML, PDF, DOC, EPUB
- Allow indexing: Named Entities, Tables of Contents, Concordances
- Can be used for text-mining
- Can be used for machine-learning
- Allow interchange with other projects (e.g. Linked Open Data)
- Allow re-use that we cannot forecast at the moment

A gentle introduction to markup languages

Some terminology

- Markup language: a set of markers or tags for an electronic text. It makes the text machine-readable, providing distinguished functionality to some of its features.
 - Markup, encoding, tagging: inserting those markers in a text
- Validating Schema: defines a set of rules associated with the markers in a given context
 - The set contains all tags that may appear in a given context
- Markup Vocabulary: defines a set of markers specifically employed with a certain function in a given schema

XML (EXtensible Markup Language)

- XML uses tags:

```
<element attribute="value">content</element>  
<div type="textpart">some random text</div>
```

- XML must be well-formed:
 - All tags must be closed
 - All tags must be nested properly (i.e. contained by all ancestors)
 - The whole document must be contained by one element named ROOT
- XML may validate against a schema:
 - You can validate your document by associating it to the schema
 - The parser checks that your document conforms to the specifications given in the schema

What is TEI (Text Encoding Initiative)

- An inline markup language
- Has:
 - Elements, which have names
 - Attributes, which have values
 - A basic schema, which can be customized by selecting a range of features (there are also various types of subschemes already available)
- The main reference: the TEI Guidelines ([current version](#))
 - The Guidelines have only a few mandatory parts
 - They can be adapted to specific needs (designed to be flexible)

TEI likes grouping information

- Modules
 - Each module declares particular XML elements and attributes that have structural aspects in common
 - Modules have a semantic ground: the grouped element declarations are considered as part of a specific textual structure, for example a Bibliography or the critical apparatus
 - The Core Module is recommended: it contains the basic elements and attributes that should be used in ANY TEI document
- Classes
 - Group the features of elements and attributes: those belonging to a class inherit all the features of that class (also for subclasses)
- Macros & datatypes
 - Define big groups of frequently used models and datatypes (e.g. **data**.
certainty)

Dictionaries and Geography together.

Why?

complex scientific/scholarly genres

- complex in terms of
 - internal structure(s)
 - external reference
 - density of information
- inconsistencies
- the linearity of the textual form impedes the clear and structured presentation of the information
- in print culture: development of complex typographical structures to accommodate for the logical/structural complexity
- TEI XML: provides a sufficient(?) device to annotate these texts both on the level of logical structure and on the level of representation

The Genre(s): “Dictionary”

genres

- two fundamental different approaches:
 - from word → meaning(/information): semasiological dictionaries
 - from “notion” → words: onomasiological dictionaries
- great variety depending on purpose, e.g.:
 - monolingual – bi-/multilingual dictionaries
 - etymological dictionaries
 - dictionaries of technical terms
 - author’s or genre’s dictionaries
- integration of types of information concerning different linguistic levels (phonology, morphology, syntax, semantics, pragmatics)

structures

- macrostructure: the overall principle of organisation of the material
 - alphabetic organisation
 - “onomastic” organisation
- microstructure: the way(s) each “entry” is organised
 - which elements/information are present
 - how elements are ordered

Markup Strategies for Dictionaries using TEI XML

general aim

- encode ancient texts similar in type and function to modern dictionaries
- annotate explicit and implicit information to enable
 - various types of visualization
 - connections between different dictionaries
 - extraction of quantitative data
- apply modern categories and models to ancient texts
- experience the limits of TEI XML with dictionaries that are not in a “catalogue”-similar structure and with complex, not clearly structured material

the entry “adnotatio” in Oxf. Lat. Dict.

adnotatio ~onis, f. [adnoto+-tio] 1 the writing or making of notes. b a note or comment. in hoc genere prorsus recipio hanc breven ~onem libellosque quo vel manu teneantur Qunit.Inst. 10.31. b a te librum meum cum ~onibus tuis exspecto Plin.Ep. 7.20.2. 2 a notice, intimation. b notice, attention. annus exinde computandus est ex quo ea ~o. .publice innotuit Macer dig. 48.17.4. b dignum ~one est quod aquam non nisi ex castello duci permittit Fron. Aq. 106. 3 symbolization manifestum est eam (sc. notam) sedeciens ducendam ex ~one denarii Maecian. iur. 63.

the entry “adnotatio” in Oxf. Lat. Dict.

adnotatio ~onis, f. [adnoto+-tio] 1 the writing or making of notes. b a note or comment. in hoc genere prorsus recipio hanc breven ~onem libellosque quo vel manu teneantur Qunit.Inst. 10.31. b a te librum meum cum ~onibus tuis exspecto Plin.Ep. 7.20.2. 2 a notice, intimation. b notice, attention. annus exinde computandus est ex quo ea ~o. .publice innotuit Macer dig. 48.17.4. b dignum ~one est quod aquam non nisi ex ...

29; παρονομαστα, quae dicitur ~o QUINT.Inst.9.3.66.

adnoscō: see AGN-.

adnotāmentum ~ī, n. [ADNOTO+-MENTVM]
A note or comment.

multam.. locutionum talium copiam offendimus atque his uulgo ~is inspersimus GEL.1.7.18; 17.2.1.

adnotātiō ~ōnis, f. [ADNOTO+-TIO]

1 The writing or making of notes. b a note or comment.

in hoc genere prorsus recipio hanc breuem ~onem libellosque qui uel manu teneantur QUINT.Inst.10.7.31. b a te librum meum cum ~onibus tuis exspecto PLIN.Ep.7.20.2; GEL.pr.3; columniosa est.. illa ~o posse legari seruo et quamdiu seruiat PAUL.dig.31.1.82.2.

2 A notice, intimation. b notice, attention. annus exinde computandus est, ex quo ea ~o.. publice innotuit MACER dig.48.17.4. b dignum ~one est quod aquam non nisi ex castello duci permittit FRON.Aq.106.

3 Symbolization.

manifestum est eam (sc. notam) sedeciens ducendam ex ~one denarii MAECIAN.iur.63.

adnotātiuncula ~ae, f. [prec.+-VNCVLA]
A short note.

cum finem proposuerimus ~is istis bellum Poenorum secundum GEL.17.21.50; 19.7.12.

the entry in the book: structure

29; παρονομασία, quae dicitur ~o QUINT. *Inst.* 9.3.66.

adnoscō: see AGN-.

adnotāmentum ~ī, n. [ADNOTO+-MENTVM]

A note or comment.

multam.. locutionum talium copiam offendimus atque
his uulgo ~is inspersimus GEL. 17.18; 17.2.1.

adnotātiō ~ōnis, f. [ADNOTO+-TIO]

1 The writing or making of notes. b a note
or comment.

in hoc genere prorsus recipio hanc breuem ~onem libellos-
que qui uel manu teneantur QUINT. *Inst.* 10.7.31. b a te
librum meum cum ~onibus tuis exspecto PLIN. *Ep.* 7.20.2;
GEL. pr.3; calumniosa est.. illa ~o posse legari seruo et
quamdui seruiat PAUL. *dig.* 31.1.82.2.

2 A notice, intimation. b notice, attention.

annus exinde computandus est, ex quo ea ~o.. publice
innovuit MACER *dig.* 48.17.4. b dignum ~one est quod
quam non nisi ex castello duci permittit FRON. *Aq.* 106.

3 Symbolization.

manifestum est eam (*sc.* notam) sedeciens ducendam ex
~one denarii MAECIAN. *iur.* 63.

adnotātiuncula ~ae, f. [prec.+-VNCVLA]

A short note.

cum finem proposuerimus ~is istis bellum Poenorum
secundum GEL. 17.21.50; 19.7.12.

[entry]

form (headword)

[sense]

“defintion”

cited quotations (examples)

the element “definition” <def>

- different types of dictionaries use different ways to define a sense
 - definitions
 - translations
 - synonyms, antonyms
 - paraphrases
- all of them are “definitions” <def>

the entry in the book: explicit & implicit information

29; παρονομασία, quae dicitur ~o QUINT. Inst. 9.3.66.

adnoscō: see AGN-.

adnotāmentum ~ī, n. [ADNOTO+-MENTVM]

A note or comment.

multam.. locutionum talium copiam offendimus atque
his uulgo ~is inspersimus GEL. I.7.18; I.7.2.1.

adnotātiō ~ōnis, f. [ADNOTO+-TIO]

1 The writing or making of notes. b a note
or comment.

in hoc genere prorsus recipio hanc breuem ~onem libellos-
que qui uel manu teneantur QUINT. Inst. 10.7.31. b a te
librum meum cum ~onibus tuis exspecto PLIN. Ep. 7.20.2;
GEL.pr.3; calumniosa est.. illa ~o posse legari seruo et
quamdui seruiat PAUL. dig. 31.1.82.2.

2 A notice, intimation. b notice, attention.

annus exinde computandus est, ex quo ea ~o.. publice
innovuit MACER dig. 48.17.4. b dignum ~one est quod
quam non nisi ex castello duci permittit FRON. Aq. 106.

3 Symbolization.

manifestum est eam (sc. notam) sedeciens ducendam ex
~one denarii MAECIAN. iur. 63.

adnotatiuncula ~ae, f. [prec.+-VNCVLA]

A short note.

cum finem proposuerimus ~is istis bellum Poenorum
secundum GEL. I.7.21.50; I.9.7.12.

implicit information:
3rd declension class

implicit information:
“noun”

explicit information:
etymology

encoding the structure

29; παρονομαστα, quae dicitur ~o QUINT. Inst. 9.3.66.

adnoscō: see AGN-.

adnotāmentum ~ī, n. [ADNOTO+-MENTVM]

A note or comment.

multam.. locutionum talium copiam offendimus atque
his uulgo ~is inspersimus GEL. 1.7.18; 17.2.1.

adnotatiō ~ōnis, f. [ADNOTO+-TIO]

1 The writing or making of notes. b a note
or comment.

in hoc genere prorsus recipio hanc breuem ~onem libellos-
que qui uel manu teneantur QUINT. Inst. 10.7.31. b a te
librum meum cum ~onibus tuis exspecto PLIN. Ep. 7.20.2;
GEL. pr. 3; calumniosa est.. illa ~o posse legari seruo et
quamdui seruiat PAUL. dig. 31.1.82.2.

2 A notice, intimation. b notice, attention.

annus exinde computandus est, ex quo ea ~o.. publice
innotuit MACER dig. 48.17.4. b dignum ~one est quod
aquam non nisi ex castello duci permittit FRON. Aq. 106.

3 Symbolization.

manifestum est eam (sc. notam) sedeciens ducendam ex
~one denarii MAECIAN. iur. 63.

adnotatiuncula ~ae, f. [prec. +-VNCVLA]

A short note.

cum finem proposuerimus ~is istis bellum Poenorum
secundum GEL. 17.21.50; 19.7.12.

```
<entry n="1">
  <form>adnotatio</form>
  <sense n="1">
    <sense n="a">
      <def>the writing or making of note</def>
    </sense>
    <sense n="b">
      <def>a note or comment</def>
    </sense>
  </sense>
  <sense n="2">
    <sense n="a">
      <def>a notice, intimation</def>
    </sense>
    <sense n="b">
      <def n="2b">notice, attention</def>
    </sense>
  </sense>
</entry>
```

a different possibility for bilingual dictionaries

29; παρονομαστα, quae dicitur ~o QUINT. Inst. 9.3.66.

adnoscō: see AGN-.

adnotāmentum ~ī, n. [ADNOTO+-MENTVM]

A note or comment.

multam.. locutionum talium copiam offendimus atque
his uulgo ~is inspersimus GEL. 1.7.18; 17.2.1.

adnotatiō ~ōnis, f. [ADNOTO+-TIO]

1 The writing or making of notes. b a note
or comment.

in hoc genere prorsus recipio hanc breuem ~onem libellos-
que qui uel manu teneantur QUINT. Inst. 10.7.31. b a te
librum meum cum ~onibus tuis exspecto PLIN. Ep. 7.20.2;
GEL.pr.3; calumniosa est.. illa ~o posse legari seruo et
quamdui seruiat PAUL. dig. 31.1.82.2.

2 A notice, intimation. b notice, attention.

annus exinde computandus est, ex quo ea ~o.. publice
i nnottuit MACER dig. 48.17.4. b dignum ~one est quod
aquam non nisi ex castello duci permittit FRON. Aq. 106.

3 Symbolization.

manifestum est eam (sc. notam) sedeciens ducendam ex
~one denarii MAECIAN. iur. 63.

adnotatiuncula ~ae, f. [prec.+-VNCVLA]

A short note.

cum finem proposuerimus ~is istis bellum Poenorum
secundum GEL. 17.21.50; 19.7.12.

```
<sense n="1">
  <sense n="a">
    <def>the writing or making of note</def>
  </sense>
  <sense n="b">
    <def>a note or comment</def>
  </sense>
</sense>
```

```
<sense n="1">
  <sense n="a">
    <cit type = "translation" xml:lang="en">
      <quote>the writing or making of note</quote>
    </cit>
  </sense>
</sense>
```

encoding implicit and explicit gram. information

29; παρονομαστα, quae dicitur ~o QUINT. Inst. 9.3.66.

adnoscō: see AGN-.

adnotāmentum ~ī, n. [ADNOTO+-MENTVM]

A note or comment.

multam.. locutionum talium copiam offendimus atque
his uulgo ~is inspersimus GEL. 1.7.18; 17.2.1.

adnotatiō ~ōnis, f. [ADNOTO+-TIO]

1 The writing or making of notes. b a note
or comment.

in hoc genere prorsus recipio hanc breuem ~onem libellos-
que qui uel manu teneantur QUINT. Inst. 10.7.31. b a te
librum meum cum ~onibus tuis exspecto PLIN. Ep. 7.20.2;
GEL. pr. 3; calumniosa est.. illa ~o posse legari seruo et
quamdui seruiat PAUL. dig. 31.1.82.2.

2 A notice, intimation. b notice, attention.

annus exinde computandus est, ex quo ea ~o.. publice
i nnottuit MACER dig. 48.17.4. b dignum ~one est quod
aquam non nisi ex castello duci permittit FRON. Aq. 106.

3 Symbolization.

manifestum est eam (sc. notam) sedeciens ducendam ex
~one denarii MAECIAN. iur. 63.

adnotatiuncula ~ae, f. [prec. +-VNCVLA]

A short note.

cum finem proposuerimus ~is istis bellum Poenorum
secundum GEL. 17.21.50; 19.7.12.

```
<entry n="1">
  <form>
    <orth>adnotatio</orth> [orthograph. form]
    <gramGrp>
      <pos>n</pos> [Part of Speech]
      <gen>f.</gen> [Gender]
      <iType>3</iType> [Declension]
      <case value="genitive">adnotationis</case>
    </gramGrp>
  </form>
  <etym>adnoto+-tio</etym>
  <!-- ... -->
</entry>
```

encoding quoted examples

29; παρονομαστα, quae dicitur ~o QUINT. Inst. 9.3.66.

adnoscō: see AGN-.

adnotāmentum ~ī, n. [ADNOTO+-MENTVM]

A note or comment.

multam.. locutionum talium copiam offendimus atque
his uulgo ~is inspersimus GEL. 1.7.18; 17.2.1.

adnotatiō ~ōnis, f. [ADNOTO+-TIO]

1 The writing or making of notes. b a note
or comment.

in hoc genere prorsus recipio hanc breuem ~onem libellosque
qui uel manu teneantur QUINT. Inst. 10.7.31. b a te
librum meum cum ~onibus tuis exspecto PLIN. Ep. 7.20.2;
GEL. pr. 3; calumniosa est.. illa ~o posse legari seruo et
quamdui seruiat PAUL. dig. 31.1.82.2.

2 A notice, intimation. b notice, attention.

annus exinde computandus est, ex quo ea ~o.. publice
innotuit MACER dig. 48.17.4. b dignum ~one est quod
aquam non nisi ex castello duci permittit FRON. Ag. 106.

3 Symbolization.

manifestum est eam (sc. notam) sedeciens ducendam ex
~one denarii MAECIAN. iur. 63.

adnotatiuncula ~ae, f. [prec. +-VNCVLA]

A short note.

cum finem proposuerimus ~is istis bellum Poenorum
secundum GEL. 17.21.50; 19.7.12.

<entry n="1">

<form>adnotatio</form>

<sense n="1">

<sense n="a">

<def>the writing or making of note

<cit>

<quote>in hoc genere prorsus recipio
hanc breven ~onem libellosque quo vel manu
teneantur</quote>

<bibl> Qunit. Inst. 10.31.</bibl>

</cit>

</sense>

</sense>

</entry>

the lexicon of the “Antiatticist”

- a lexicon of the 2nd century CE
- severely epitomized
- main aim: to argue for the extension the “canonical” vocabulary

models of microstructures in the “Antiatticist”

- clearly structured
- (mainly) 2 different models of microstructure
 - lemma | meaning | sources
 - lemma | grammatical information | sources

“Antiatticist” β 1

Antiatt. β 1:
bios; for livelihood.
Herodotus, Menander
in Demiourgos, Eupolis
in Aiges, the poet
several times.

```
<entry n="β1">
  <form>bios</form>
  <sense n="1">
    <def>for livelihood</def>
    <cit><bibl>Herodotus</bibl></cit>
    <cit><bibl>Menander,
Demiourgos</bibl></cit>
    <cit><bibl>Eupolis
Aiges</bibl></cit>
    <cit><bibl>the poet several
times</bibl></cit>
    </sense>
  </entry>
```

“Antiatticist” α 62

Antiatt. α 62:

argyria; in plural not in singular. Platon in Kleophon.

```
<entry n="α 62">
  <sense>
    <form>argyria</form>
    <usg type="gram">in plural not in
singular.</usg>
    <cit>
      <bibl>
        <persName>Platon</persName>
        <title>in Kleophon.</title>
      </bibl>
    </cit>
  </sense>
</entry>
```

the lexicon of Harpocration

- a lexicon of the 2nd century CE (?): “Lexicon of the ten orators”
- a lexicon explaining the vocabulary (mainly but not exclusively technical terms) used in rhetorical texts of the 5th and 4th cent. BCE
- not always clearly structured
- the microstructure models contain: lemma, meaning, grammatical information, examples/sources, various remarks
- Text (from a print edition) and engl. translation in <http://dcthree.github.io/harpocration>

Harpocration s.v. *abios*

Harp. s .v. ἄβιος

Abios: Antiphon applied “abios” to one who has acquired great wealth, as Homer calls a much wooded wood “axylos”.

```
<entry>
  <form>abios</form>
  <sense ana="wealthy">
    <def>
      <persName>Antiphon</persName>
      applied “abios” to one who has
      acquired great wealth
    </def>
    </sense>
    <note>as Homer calls a much wooded wood
    “axylos”.</note>
  </entry>
```

Geography as a technical genre

Geographical literature for Greeks and Romans

- Geographical sources:
 - *Peripli* (πέριπλοι): lists of places along the coastal line
 - *Periegeseis* (περιηγήσεις) or *Periodoi*: «guided tours around»
 - *Chorographia* (χωρογραφία)
 - *Itineraria*: lists of stations and distances along Roman routes
- Geographical authors:
 - «Major» geographers: Ptolemy, Pausanias, Strabo, Pliny, Pomponius Mela
 - «Minor» geographers: short opuscules by various authors

Beyond the map

- Premodern societies are not map-based
- Cartography was used for different purposes than travelling
- Their notion of space was empirical: functional to travel and spatial practice in general
- The way they described space was the only thing they had for travelling
- **Can we consider premodern geography as a “technical genre”?**

How did they find the way?

- **Distances**
 - The numerical estimate depended on the concrete conditions of travel
 - Several units of measurement (in the widest way possible)
 - Two ways of measuring spatial extents: time or space
- **Systems of orientation**
 - Fixed: winds, stars, cardinal points, etc.
 - Unfixed: unsystematic, environmental, cultural references
- **Spatial knowledge as a «mental model»:**
 - Cultural aspects of spatial practice: associations between places and concepts (e.g. boundaries)

From Leptis to Thermae sail for 60 stades; it is a town, and here in the same way the shoals make the sailing difficult. From Thermae sailing 40 stades you see the promontory against which are two islands staked out with pilings; there is an anchorage. From Adramyte to Aspis 500 stades. It is a high and conspicuous promontory, shaped like a shield. From there sail north so that the promontory appears on the left, for there are many shoals and rocks in this sea. And then Aspis and Neapolis upon it will appear. From the bay of Neapolis to Aspis, sail for 200 stades. It is a high place and the city is on it. It has a harbor looking toward the west, 10 stades beyond the city.

(*Stadiasmus of the Great Sea*, 114-117).

From Leptis to Thermae sail for 60 stades; it is a town, and here in the same way the shoals make the sailing difficult. From Thermae sailing 40 stades you see the promontory against which are two islands staked out with pilings; there is an anchorage. From Adramyte to Aspis 500 stades. It is a high and conspicuous promontory, shaped like a shield. From there sail north so that the promontory appears on the left, for there are many shoals and rocks in this sea. And then Aspis and Neapolis upon it will appear. From the bay of Neapolis to Aspis, sail for 200 stades. It is a high place and the city is on it. It has a harbor looking toward the west, 10 stades beyond the city.

(*Stadiasmus of the Great Sea*, 114-117).

From Leptis to Thermae sail for 60 stades; it is a town, and here in the same way the shoals make the sailing difficult. From Thermae sailing 40 stades you see the promontory against which are two islands staked out with pilings; there is an anchorage. From Adramyte to Aspis 500 stades. It is a high and conspicuous promontory, shaped like a shield. From there sail north so that the promontory appears on the left, for there are many shoals and rocks in this sea. And then Aspis and Neapolis upon it will appear. From the bay of Neapolis to Aspis, sail for 200 stades. It is a high place and the city is on it. It has a harbor looking toward the west, 10 stades beyond the city.

(*Stadiasmus of the Great Sea*, 114-117).

From Leptis to Thermae sail for 60 stades; it is a town, and here in the same way the shoals make the sailing difficult. From Thermae sailing 40 stades you see the promontory against which are two islands staked out with pilings; there is an anchorage. From Adramyte to Aspis 500 stades. It is a high and conspicuous promontory, shaped like a shield. From there sail north so that the promontory appears on the left, for there are many shoals and rocks in this sea. And then Aspis and Neapolis upon it will appear. From the bay of Neapolis to Aspis, sail for 200 stades. It is a high place and the city is on it. It has a harbor looking toward the west, 10 stades beyond the city.

(*Stadiasmus of the Great Sea*, 114-117).

From Leptis to Thermae sail for 60 stades; it is a town, and here in the same way the shoals make the sailing difficult. From Thermae sailing 40 stades you see the promontory against which are two islands staked out with pilings; there is an anchorage. From Adramyte to Aspis 500 stades. It is a high and conspicuous promontory, shaped like a shield. From there sail north so that the promontory appears on the left, for there are many shoals and rocks in this sea. And then Aspis and Neapolis upon it will appear. From the bay of Neapolis to Aspis, sail for 200 stades. It is a high place and the city is on it. It has a harbor looking toward the west, 10 stades beyond the city.

(*Stadiasmus of the Great Sea*, 114-117).

A spatial system defined by relations

- **Place A**
 - Distance estimate
 - Direction (conventional system of orientation)
 - Underspecified indication (unconventional system of orientation): various degrees of relations between places
- **Place B**

Premodern geography as a technical genre

- Geographical language is technical:
 - A specific system made up of concepts and paradigms
 - A specific way of “encoding” spatial information in the language
- Geographical language is *not* technical (yet):
 - High degree of implicit information (in written sources and in maps)
 - Strong variability in the linguistic representation of information:
 - According to the context
 - According to the period
- How can we “encode” geographical information in digital editions? (and can we at all?)

Markup strategies for Digital Editions of Geographical texts

Preliminary observations

- You should choose your strategy according to the results you want
- TEI is an option:
 - It provides Guidelines
 - has a large community willing to help
 - It does not demand a lot of time to be learned
-certainly not the only option:
 - You can create your own XML Schema if you think you need a very specific one
 - There are other methods (e.g. standoff annotation)
- ...maybe not the best option:
 - If your documents have a high semantic demand, TEI may not be the solution (but it can provide a customization model)

Let's start “simple”: place names

13. Names, Dates, People and Places

<placeName>

contains an absolute or relative place name

<geogName>

contains a name associated with some geographical feature, belonging to physical geography (“Red Sea”, “Mount Rushmore”)

<geogFeat>

nested into <geogName>, indicates geographical feature names (“Sea”, “Mount”)

Example:

- You can tag types of spatial entities with the attribute @type
- A proper nesting of `<geogName>` with `<geogFeat>` associated also uses `<name>`, as usual for named entities consisting of multiple parts
- You should be consistent in the choice of the tag

```
<div type="textpart" subtype="section" n="9">
<p> After the <geogName type="water">
    <name>Aegean</name>
    <geogFeat>Sea</geogFeat>
</geogName> comes the <placeName type="region">Hellespont</placeName>
ending at <placeName type="settlement">Abydus</placeName> and <placeName
    type="settlement">Sestus</placeName>, then the <geogName
    type="water">Propontis</geogName> ending at <placeName
    type="settlement">Chalcedon</placeName> and <placeName
    type="settlement">Byzantium</placeName>, where the narrows are from
which the <geogName type="water">Pontus</geogName> begins, then
    <geogName type="water">
        <geogFeat>Lake</geogFeat>
        <name>Maeotis</name>
    </geogName>. Again, from the beginning of <placeName type="entity">
        >Europe</placeName> and <placeName type="entity">Libya</placeName>
the <geogName type="water">
    <name>Iberian</name>
    <geogFeat>Sea</geogFeat>
</geogName> from the <placeName>Pillars</placeName> to <geogName
    type="mountain">
        <geogFeat>Mount</geogFeat>
        <name>Pyrene</name>
    </geogName>, the <geogName type="water">Ligurian</geogName> as far as
the borders of <placeName type="country">Etruria</placeName>, the
    <geogName type="water">Sardinian</geogName> beyond <placeName
    type="island">Sardinia </placeName> bending down towards <placeName
    type="entity">Libya</placeName>, the <geogName type="water">
        >Tuscan</geogName> ending as far as <placeName type="island">
            >Sicily</placeName> and beginning from the <geogName
            type="promontory">capess</geogName> of <placeName type="region">
```

Authority lists with placeNames

- For the creation of indexes and controlled lists of named entities
 - Encoded in TEI or EpiDoc, or in an external database format (e.g. gazetteers)
 - A project that tags named entities is strongly advised to link them to an authority list, either internal or external
1. Internal authority lists are indicated with <`listPlace`>: contains a list of places, optionally followed by a list of relationships defined amongst them (see below)
 2. External authority lists are referred to by the `@ref` attribute: points at an external database which stores controlled information about the toponym (e.g. [Pleiades](#))

Example:

- Suggestion: use **@nymRef** for sources pointing to the canonical form (also called “nym”) of a named entity: the value of this attribute indicates the standard form of the name, not the entity itself
- **@ref** contains an URL or URI pointing to an external reference
- Using nyms can be useful for indexing types of names or places that need to be referenced to a standard list in order to allow advanced applications in some standardized form (like creating a map)

```
<p n="1">Αναξίμανδρος ό
<placeName type="ethnic" nymRef="Μιλήτος" ref="http://pleiades.stoa.org/places/606335">
Μιλήσιος</placeName> ἀκουστῆς Θάλεω πρώτος ἀπετόλμησε τὴν οἰκουμένην
ἐν πίνακι γράψαι, μεθ' ὅν Έκαταῖος ό <placeName type="ethnic" nymRef="Μιλήτος"
ref="http://pleiades.stoa.org/places/606335">Μιλήσιος</placeName>
ἀνήρ πολυπλανῆς διηκρίβωσεν ὡστε θαυμασθῆναι τὸ πρᾶγμα·
Ἐλλάνικος γάρ ό <placeName type="ethnic" nymRef="Λέσβος"
ref="http://pleiades.stoa.org/places/550696">Λέσβιος</placeName>
ἀνήρ πολυίστωρ ἀπλάστως παρέδωκε τὴν ἴστορίαν. εἴτα Δαμάστης ό
<placeName type="ethnic" nymRef="Κίτιους" ref="http://pleiades.stoa.org/places/707556">
Κίτιεύς</placeName> τὰ πλείστα ἐκ τῶν Έκαταίου μεταγράψας περίπλουν ἔγραψεν.
ἔξης Δημόκριτος καὶ Εὔδοξος καὶ ἄλλοι τινὲς γῆς περιόδους καὶ περίπλους ἐπραγματεύσαντο.
</p>
```

Spatial relatedness 1: locations

<location>

defines the location of a place as a set of geographical coordinates, or in terms of other named geo-political entities

<offset>

marks the part of a relative temporal or spatial dimension indicating the direction of the offset between two places or times (note it can also be used to encode vague temporal information)

<geo>

normally used to mark coordinates in a standard system, such as WGS84, KML or GML

Example:

- Sets of instructions or relations between places can be treated with `<location>` and `<placeName>` all nested into one `<place>` or `<placeName>` element (this strategy is also used for personal names with additional information)
- `<location>` simply defines a position by means of nested elements
- NOTE: EpiDoc only allows `<place>` in lists (`<listPlace>`). This nesting has to be done into one `<placeName>` element

```
<placeName>
  <location>
    <placeName type="island">Crete</placeName>
      <offset>near</offset>
      <geogName>Peloponnesus</geogName>
      <offset>off</offset>
      <geogName type="cape">Malea</geogName>
    </location>
  </placeName>
```

Spatial relatedness 2: relations

<relation>

describes any kind of relationship or linkage amongst a specified group of places, events, persons, objects or other items

a standoff-like markup with a good level of precision for hierarchical relations

<listRelation>

provides information about relationships defined amongst people, places or organizations, either informally or as formally expressed relation links

can be located at the Header in order to provide basic semantics to the relations of the document

Example:

- A preliminary <listPlace> provides IDs to all the places involved in the relation
- @name gives a standard reference to the relation and provides some sort of semantic specification to it (e.g. <relation name="partOf">)

for those going over from Europe both by land and by the islands lying in a row:

```
<listPlace>

    <place xmlid="Eub">
        <placeName>Euboea</placeName>
    </place>
    <place xmlid="Andr">
        <placeName>Andros</placeName>
    </place>
    <place xmlid="Ten">
        <placeName>Tenos</placeName>
    </place>
    <place xmlid="Myc">
        <placeName>Myconos</placeName>
    </place>
    <place xmlid="Ic">
        <placeName>Icaria</placeName>
    </place>
    <place xmlid="Samos">
        <placeName>Samos</placeName>
    </place>
    <place xmlid="Myca">
        <placeName>Mycale</placeName>
    </place>
    <relation name="route"
        mutual="#Eub #Andr #Ten #Myc #Myca #Ic #Samos"/>

</listPlace>
```

Disadvantages

<location>, <offset>

- Poor semantic specification of the nature of the relation: everything relies on the nesting
- Easy use for standard information (“north of...”) and simple locations, but infinite nesting for complicated locations and limited possibilities for long instructions

```
<placeName>
  <location>
    <offset>After the</offset>
    <geogName type="water">
      <name>Aegean</name>
      <geogFeat>Sea</geogFeat>
    </geogName>
    <offset>comes the</offset>
    <placeName type="region">
      >Hellespont</placeName>
      <offset>ending at</offset>
      <placeName type="settlement">Abydus</placeName>
      <placeName type="settlement">Sestos</placeName>
    </location>
  </placeName>
```

<relation>

- Only for hierarchical or mutual relations (@mutual, @active, @passive)
- As standoff-like markup it is not incorporated in the text as it is and requires the isolation/manipulation of junks of complex texts (pure place lists are rare in historical texts)
- Very restrictive usage in TEI and EpiDoc: requires customization for “proper” geographical texts

Distances

<measure>

contains a word or string referring to some quantity, usually comprising a number, a unit and a commodity name.

Example:

- <measure> can indicate any type of measurement: @type needs to specify whether it is a linear distance, an extent, a perimeter, a length, a width...
- units in premodern texts are expressed either in time or space: @unit can be adopted to specify that, instead of the specific unit
- there is much variety in the expression of units of measurement

```
<placeName>
  <offset>from</offset>
  <placeName type="island">Paphos</placeName>
  <offset>to</offset>
  <placeName type="settlement">Alexandria</placeName>
  <geogFeat>the passage is</geogFeat>
  <offset>from the north</offset>
  <measure type="distance" unit="space">
    3,800 stades
  </measure>
</placeName>.
<placeName>
  the perimeter of
  <placeName type="island">Rhodes</placeName> is
  <measure type="perimeter">1300 stades</measure>
</placeName>
```

Disadvantages

- Poor specification of types
- No standard reference for modern equivalents of ancient units of measurement
- A distance is a relation in itself: how do we record that?
- Encoding implicit information:

“The perimeter of Rhodes is 1330 stades, of Cos 540 stades, of Samos 630 stades ... The perimeter of Chios is 660 stades, of Lesbos 1100 stades.”

Desiderata

- An EASY system for tagging spatial relations
- A linguistic reference structure
- A dataset of semantic relations between spatial entities
- An integration of relational markup (RDF, for example) in digital editions, so that every element is connected in the context in the appropriate way

Some simple applications of encoded
geographical texts

Creating indexes with XSLT transformations

τὸν λεικόνοτον ἥτοι λιβόντον, μέσου δὲ 27 ἀπαρκτίου καὶ ἀργέστου θρασκίαν ἥτοι κίρκιον ὑπὸ τῶν

περιοικῶν <ονομαζόμενον>, ἔθη δὲ οἰκεῖν τὰ πέρατα κατ’ ἀπλώσην <placeName type="ethnic" nymRef="Βάκτρια" ref="http://pleiades.stoa.org/places/971712">Βάκτριανούς</placeName>, κατ’ εὖρον <placeName type="ethnic" nymRef="Τινδική" ref="http://pleiades.stoa.org/places/50004">Τινδούς</placeName>, κατὰ φοίνικα 28 <placeName ref="http://pleiades.stoa.org/places/39290">Ερυθράν θάλασσαν</placeName> καὶ <placeName ref="http://pleiades.stoa.org/places/39274">Αἰθιοπίαν</placeName>, κατὰ νότον <placeName nymRef="Αἰθιοπία" ref="http://pleiades.stoa.org/places/39274">τὴν ύπερ Αἴγυπτον Αἰθιοπίαν</placeName>, κατὰ λεικόνοτον <placeName type="ethnic" nymRef="Ταράμαντες" ref="http://pleiades.stoa.org/places/354116">τοὺς ύπερ Σύρτεις Γαράμαντας</placeName>, κατὰ λίβα <placeName type="ethnic" nymRef="Αἰθιοπία" ref="http://pleiades.stoa.org/places/334481">Αἰθίο [καὶ] δυσμικούς ύπερ Μαύρους</placeName>, 29 κατὰ ζέφυρον <placeName ref="http://data.pastplace.org/search?q=152854">Στηλας</placeName> καὶ ἀρχάς <placeName ref="http://data.pastplace.org/search?q=6542791">Λιβύνης</placeName> καὶ <placeName ref="http://pleiades.stoa.org/places/893990">Εὐρώπης</placeName>, κατὰ ἀργέστην <placeName ref="http://pleiades.stoa.org/places/1027">Ιβριανὰ τὴν νῦν Ισπανίαν</placeName>, κατὰ [βέ] θρασκίαν <placeName type="ethnic" nymRef="Κελτική" ref="http://pleiades.stoa.org/places/993">Κελτούς</placeName> καὶ τὰ ὅμορα, κατὰ ἀπαρκτίουν> <placeName type="ethnic" nymRef="Σκωθία" ref="http://pleiades.stoa.org/places/1273">τοὺς ύπερ Θράκην Σκύθας</placeName>, κατὰ [βέ] βορρᾶν <placeName nymRef="Πόντος" ref="http://pleiades.stoa.org/places/1224">Πόντον</placeName>

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <?xmlstylesheet type="text/xsl" href="toponymsforPelagiosdata.xsl"?>
4  ▷ <TEI xmlns="http://www.tei-c.org/ns/1.0">
5  ▷ <teiHeader>
6  ▷ <fileDesc>
7  ▷ <titleStmt>
8  <title>Agathemerus, Sketch of Geography</title>
9  </titleStmt>
```



```
          </a>
        </h4>
      <ul data-role="listview">
        <xsl:for-each select="current-group()">
          <li>
            <a target="_blank"
               href="{concat('edition.html#chapter',t:div[@type='textpart']/@n)}">
              <xsl:value-of select="text()"/>
              <xsl:text>
                <xsl:value-of
                  select="concat(ancestor::t:div/@n,'.',ancestor::t:p/@n)"/>
                <xsl:text>
              </xsl:for-each>
            </a>
          </li>
        </xsl:for-each>
      </ul>
    </div>
```

A raw index of Named Entities for Agathemerus' *Sketch of Geography*

Places

@type

Αιγαίον πέλαγος

- [Αιγαίον \(3.9\)](#)
- [Αιγαίον \(3.9\)](#)

Αιγύπτιον

- [Αιγύπτιον \(3.9\)](#)

Αιθιοπία

- [Αιθιοπίαν \(2.7\)](#)
- [τὴν ὑπὲρ Αἴγυπτον Αιθιοπίαν \(2.7\)](#)

Αἰλανίτου μυχός

- [Αἰλανίτου μυχοῦ \(3.14\)](#)

Βυζάντιον

- [Βυζάντιον \(3.9\)](#)

@nymRef

3. ἔστι δὲ ἡ ...

ἔστι δὲ ἡ μεγάλη Σύρτις σταδίων ,ε, ἡ δὲ μικρὰ σταδίων ,αχ'. τὸ δὲ ἄκρας Ἰαπυγίας ἐπὶ Κεραύνια ὅρη τῆς Ήπείρου .

διδέχεται τὸ Αιγαῖον <πέλαγος> Ἐλλήσποντος λίγων εἰς Ἀβυδον τὸν δὲ Πόντος ἀρχεται· είτα ἡ Μαιῶτις λίμνη. πάλιν δὲ ἀπ' ἀρχῆς Εύρη τῆς Τυρρηνίας περάτων, Σαρδῶν δὲ τὸ ύπερ Σαρδῶν νεύον πρὸς Λιάκρων, είτα Λιβυκόν, είτα Κρητικὸν καὶ Σικελικὸν καὶ Ιόνιον καὶ Άς 35 ἡτοι Αλκυνίδα θάλασσαν . τὸ δὲ Σουνιώ καὶ Σκυλλαίω περιεχόμενον Παμφύλιον καὶ Αιγύπτιον . ύπερ δὲ τὸ Ικαρίον ἔξις ἀναχεῖται τὸ Αιγύπτιον .

You are here: Home → Ancient Places → Aethiopia

Aethiopia

Creators: D.T. Potts

Contributors: R. Talbert, T. Elliott, spilles

Copyright © The Contributors. Sharing and remixing permitted under terms of the Creative Commons Attribution 3.0 License (cc-by).

Last modified Feb 14, 2012 06:37 PM — History

An ancient place, cited: BAtlas 4 A3 Aethiopia

Canonical URI for this page:

<http://pleiades.stoa.org/places/39274>



Locations:

• Barrington Atlas location (Attested dates needed)

Names:

• *Aethiopia* (Attested dates needed)

Place type:

unknown



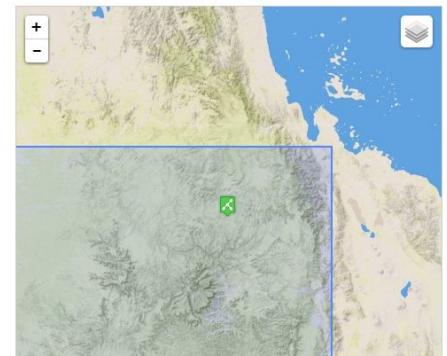
Makes a connection with:



Has a connection with:

• *A(u)ksumē* — by D.T. Potts — last modified Mar 08, 2014 10:20 PM
Present-day Aksum in Etiopia. The center of the Aksumite Kingdom, a maritime trading power that flourished from the first to the tenth centuries AD. The city was listed as UNESCO World Heritage Site in 1980.

References:



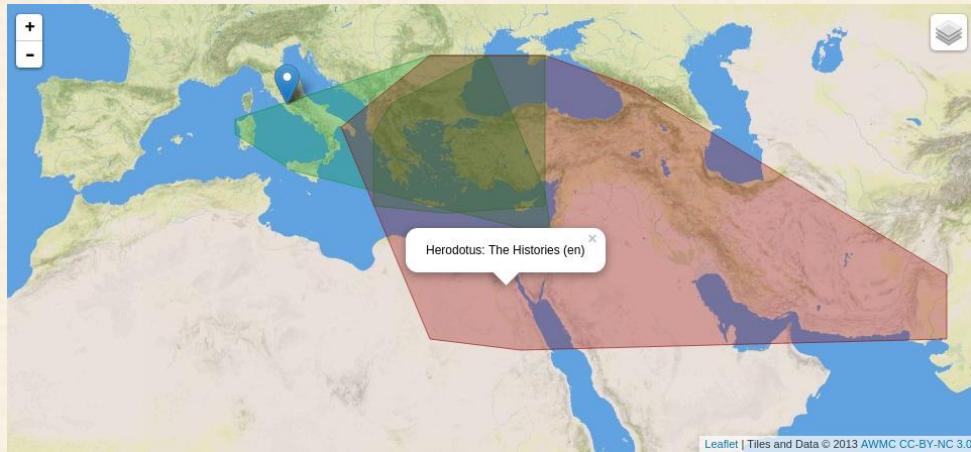
Visualizing toponyms and spatial footprints



Places categorized as types of natural feature



Selection of @type="ethnic" only



Spatial footprint using
Pelagios APIs and Leaflet

An advanced application: dynamic lexica

dynamic lexica

- digital products through combination of:
 - words and meanings in different languages
 - definitions and semantic relationships
 - dictionary definitions and texts (corpora)
- are expandable
- provide a much more accurate image of the lexicon of a language

Greek-Latin Dynamic Lexicon

Data source(Perseus Digital Library):

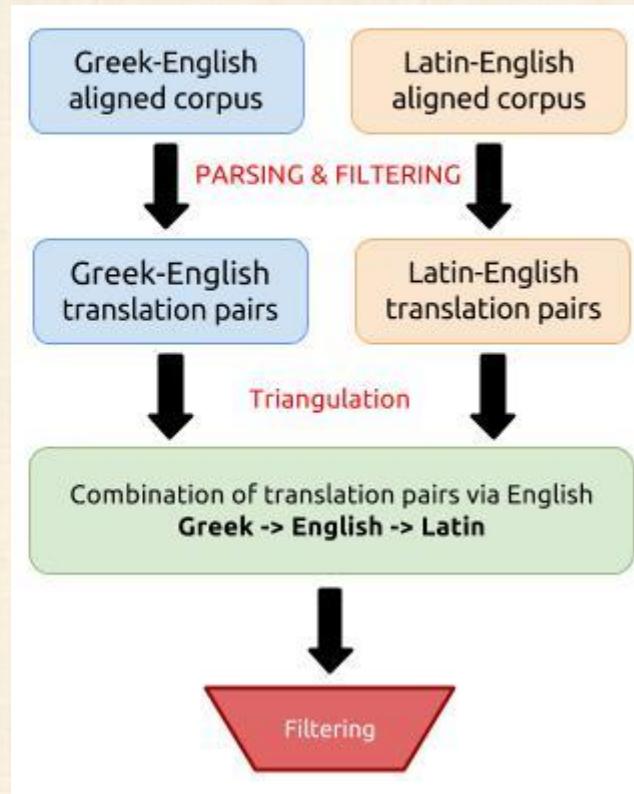
- Greek-English aligned texts (104 files)
- Latin-English aligned texts (59 files)



The parallel texts are aligned on a sentence level using **Moore's Bilingual Sentence Aligner**. Then the **Giza++** toolkit is used to align the sentence pairs at the level of individual words.

	Files	Sentences	Words	Distinct pairs
Ancient Greek	104	210 K	4,32 M	43 K
Latin	59	132 K	2,33 M	39 K

The proposed method



One-to-One Alignment

```
<sentence id="6">
  <wds lnum="L1">
    <w n="6-1" nrefs="6-3">illuc </w>
    <w n="6-2" nrefs="6-1">regredere</w>
    <w n="6-3" nrefs="6-4">ab</w>
    <w n="6-4" nrefs="6-6">ostio </w>
    <w n="6-5" nrefs="6-7">. </w>
  </wds>
  <wds lnum="L2">
    <w n="6-1" nrefs="6-2">get</w>
    <w n="6-2" nrefs="">away</w>
    <w n="6-3" nrefs="6-1">there </w>
    <w n="6-4" nrefs="6-3">from</w>
    <w n="6-5" nrefs="">the</w>
    <w n="6-6" nrefs="6-4">door</w>
    <w n="6-7" nrefs="6-5">!</w>
  </wds>
</sentence>
```

illuc regredere ab ostio .
get away there from the door !

One-to-many Alignment

```
<sentence id="8" >
  <wds lnum="L1">
    <w n="61-1" nrefs="8-1">καὶ</w>
    <w n="61-2" nrefs="8-2 8-3">πῶς</w>
    <w n="61-3" nrefs="8-5">λέγει</w>
    <w n="61-4" nrefs="8-6">;</w>
  </wds>
  <wds lnum="L2">
    <w n="61-1" nrefs="8-1">and</w>
    <w n="61-2" nrefs="8-2">what</w>
    <w n="61-3" nrefs="8-2">does</w>
    <w n="61-4" nrefs="">he</w>
    <w n="61-5" nrefs="8-3">say</w>
    <w n="61-6" nrefs="8-4">?</w>
  </wds>
</sentence>
```

καὶ πῶς λέγει ;

and what does he say ?

Alignment errors

The Latin word (**et**) occurred 55034 times in the text and is aligned to:

Word	Frequency	Percentage
and	48185	88%
;	1422	2,5%
; and	1145	2%
,	818	1,5%
, and	545	1%
other translations	1919	4%

Source of Errors:

The alignment is noisy and done automatically by the Giza++ software

Alignment errors filtering

Word	Frequency	Percentage
and	48185	88%
;	1422	2,5%
; and	1145	2%
,	818	1,5%
, and	545	1%
other translations	1919	4%

Lemmatization

Latin and Greek are highly inflected languages, whereas English is considered a weakly inflected language.

eius cupio **filiam** virginem mihi desponderi.
i' d like his maiden **daughter** to be promised me in marriage.

piliae et **filiae** salutem.
love to pilia and your **daughter**.

mea haec erilis gestitavit **Filia**.
this **daughter** of my mistress had them.

Lemmatization

Latin and Greek are highly inflected languages, whereas English is considered a weakly inflected language.

eius cupio **filiam** virginem mihi desponderi.

i' d like his maiden **daughter** to be promised me in marriage.

piliae et **filiae** salutem.

love to **pilia** and your **daughter**.

mea haec erilis gestitavit **filia**.

this **daughter** of my mistress had them.

Lemmatization solves this problem

filiam, filiae, filia → **daughter**

filia → **daughter**

Lemmatization

Latin and Greek are highly inflected languages, whereas English is considered a weakly inflected language.

καὶ ἐπειδὴ ἔτεκεν **uiόv**, ἔξαρνος ἦν μὴ εἶναι ἐξ αὐτοῦ τὸ παιδίον
but when she gave birth to a **son**, (callias) denied that the child was his

uiòv δὲ Τεισίου τοῦ Ῥαμνουσίου
the **son** of teisias of rhamnos

ούμὸς **uióç**
my **son**

Lemmatization solves this problem

uiòv, uióç, uióv → **son**

uióç → **son**

MORPHEUS

Lemmatization of Latin and Greek words is done using **Morpheus**.

Morpheus is a morphological parsing and lemmatizing tool for Ancient Greek and Latin produced by the Perseus Project and available as a **web service**.

It takes as input a token (Greek or Latin) and returns the lemma from which this token derives, and the full morphological information (part of speech, case, number, gender, person, mood, etc.).

Lemmatization

Lemmatization of English translations will produce better results.

E.g., translation candidates of the Greek word **λέγειν**

Translation	Freq	Precentage
say	551	36%
speak	492	32%
tell	149	9.7%
speaking	110	7%
said	89	6%
saying	54	3.5%
mention	45	2.9%
says	25	1.5%
spoke	19	1.2

Translation probabilities

Translation	Freq	Precentage
say	551	36%
speak	492	32%
tell	149	9.7%
 speak	110	7%
 say	89	6%
 say	54	3.5%
mention	45	2.9%
 say	25	1.5%
 speak	19	1.2

Lemmatization of English words

Translation	Freq	Precentage
say	710	46.8%
speak	621	40.6%
tell	149	9.7%
mention	45	2.9%

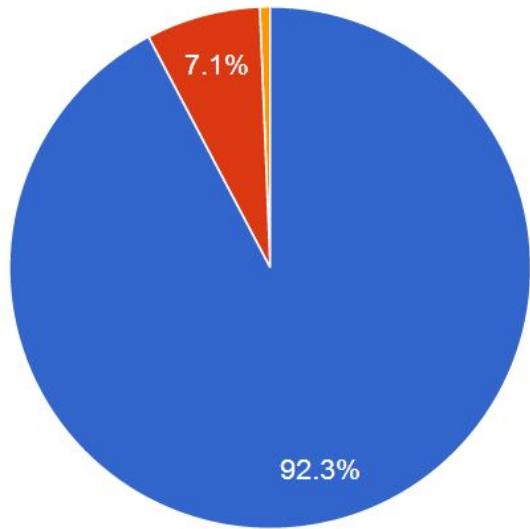
Group the results and recalculate
the probabilities

Triangulation

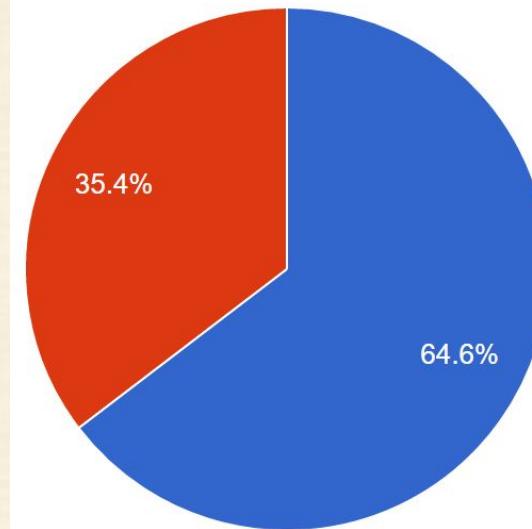
Triangulation is based on the assumption that two expressions are likely to be translations if they are translations of the same word in a third language.

DEMO

Greek Translation of (city)



Latin Translation of (city)



- πόλις (πόλιν, πόλεως, πόλει, πόλιος)
 - ἄστυ (ἄστυ, ἄστει, ἄστεως, ἄστεος)
 - πολέω (πόλεις)
- civitas (civitate, civitatem, civitatis, civitas, civitati)
 - urbs (urbem, urbe, urbis, urbs, urbi)

Triangulation

The English word (**city**) is translated to

(92.3% **πόλις**), (7.1% **ἄστοι**), (0.6% **πολέων**)

(**city**) is translated to Latin

(64.6% **civitas**), (35.4% **urbs**)

The extracted pairs via triangulation:

(**πόλις -civitas**),

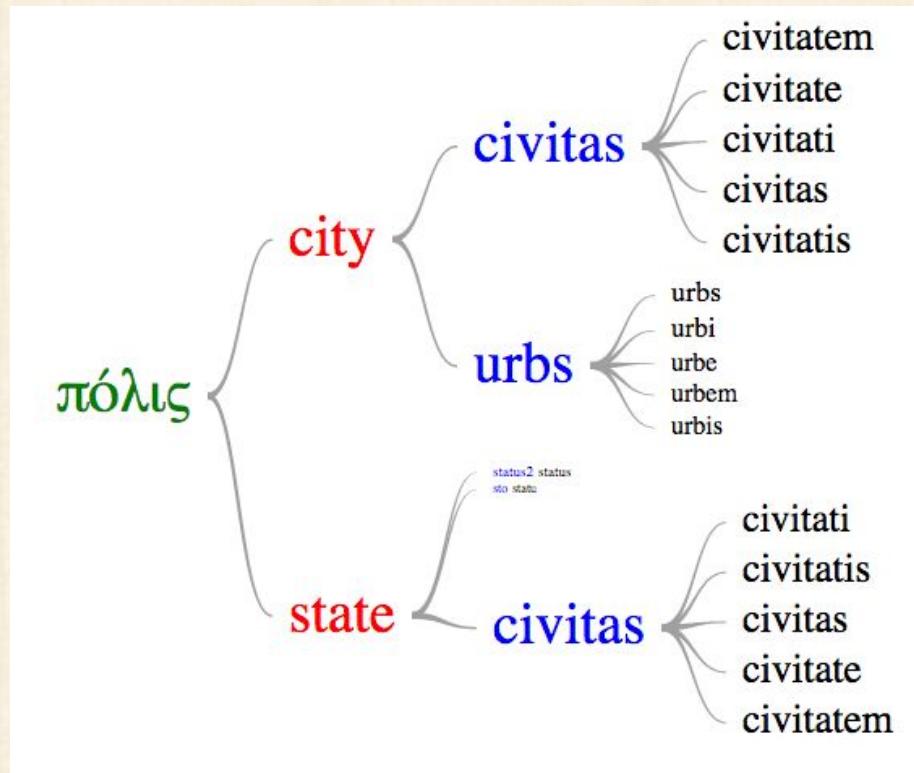
(**πόλις -urbs**),

(**ἄστοι-civitas**),

(**ἄστοι-urbs**),

(**πολέων-civitas**),

(**πολέων-urbs**).



Translation-Pairs filtering

We used **Jaccard coefficient** as a similarity metric to measure the similarity or the relatedness between every Greek-Latin pairs.

$$J(A, B) = \frac{A \cap B}{A \cup B}$$

The relatedness between the Greek word (**πόλις**) and the Latin word (**civitas**) can be calculated as follows:

Translation-Pairs filtering

πόλις		
city	7432	74 %
state	1911	19 %
town	366	3.5 %
athens	357	3.5 %

πολέω		
city	6673	64.3 %
of	2352	22.6 %
state	823	7.9 %
town	530	5.2 %

άστυ		
city	6747	91 %
citizen	399	4 %
town	236	3 %
others	96	1%

urbs		
city	44395	94.2 %
rome	2190	4,6 %
town	531	1.2 %

civitas		
city	8748	72.9 %
state	2340	19.5 %
citizenship	420	3.5 %
citizen	372	3.1 %

Translation-Pairs filtering

πόλις - cīvitas	city, state	$(74 + 19 + 72 + 19.5)/200 = 92.25 \%$	92.25 %
πόλις - urbs	city, town	$(74 + 3.5 + 94.2 + 1.2)/200= 86.45 \%$	86.45 %
πολέω - civitas	city, state	$(64.3 + 7.9 + 72 + 19.5)/200 = 81.85 \%$	81.85 %
πολέω - urbs	city, town	$(64.3 + 5.2 + 94.2 + 1.2)/200= 82.45 \%$	82.45 %
ἄστυ - cīvitas	city, citizen	$(91+ 4 + 72 + 3.1)/200 = 85.05 \%$	85.05 %
ἄστυ - urbs	city, town	$(91 + 3 + 94.2 + 1.2)/200= 94.7 \%$	94.7 %

Translation-Pairs filtering

Ρώμη		
rome	8044	100 %

urbs		
city	44395	94.2 %
rome	2190	4.6 %
town	531	1.2 %

$$J(\text{Ρώμη}, \text{urbs}) = (100+4.6)/200 = 51.3 \%$$

Evaluation

Mean reciprocal rank (MRR)

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} RR_i$$

We selected randomly 200 translation pairs obtained via the proposed method with different frequencies (high and low) and different JACCARD Co values.

Each pair should be assigned into four categories: **Correct, small difference, big difference and incorrect.**

Evaluation

Incorrect ✗	Big difference ⓘ	Small difference ⓘ	Correct ✅	πόλει - civitatis (city) [0.910379]
Incorrect ✗	Big difference ⓘ	Small difference ⓘ	Correct ✅	uiέ - gnato (son) [0.947044]
Incorrect ✗	Big difference ⓘ	Small difference ⓘ	Correct ✅	λέγοντες - diceret (say) [0.906456]
Incorrect ✗	Big difference ⓘ	Small difference ⓘ	Correct ✅	θεῶν - di (god) [0.935497]

Category	Reciprocal Rank	Jaccard co	60% <	70% <	80% <	90% <
Correct	1	MMR	61.25 %	74 %	87.5 %	94.5 %
Small difference	0.75					
Big difference	0.25					
Uncorrect	0					

The quality of the method depends on two factors:

- **The size of aligned-parallel corpora**

Bigger corpora produce better translation probability distribution and more translation candidates.

- **The quality of the aligner**

Manually aligned corpora yield more accurate results.