



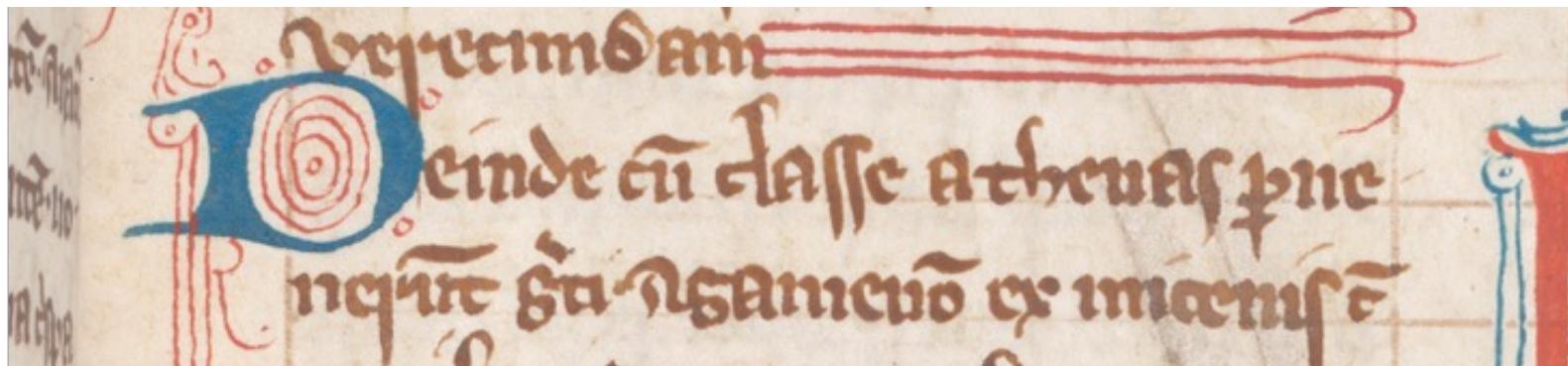
Modelling humanities data with TEI-XML

SCHOLARLY EDITING AND MANUSCRIPT CATALOGUING IN THE DIGITAL AGE

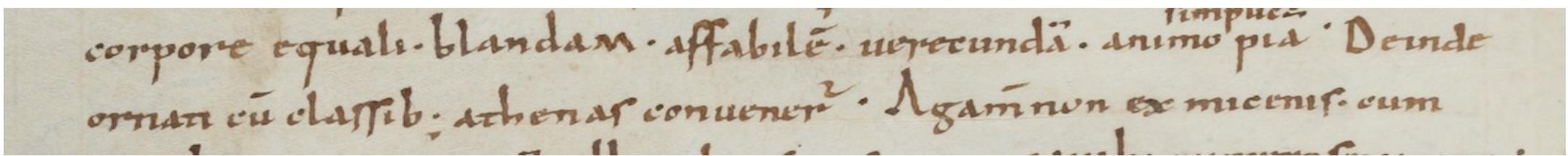
Dr Katarzyna Anna Kapitan
13 November 2024

Transcriptions vs. Editions

Then they arrived at Athens with a fleet



TCD 11500: Deinde cu(m) **classe** athenas **p(er)ueneru(n)t**



Cod. Sang. 197: Deinde **ornati** cu(m) **classib(us)** athenas **conuener(unt)**

Siglorum tabula

- L codex Leidensis s. X
M codex Monacensis s. IX vel X
G codex Sangallensis s. X
F codex Florentinus s. X
B codex Bambergensis s. X
H codex Bernensis s. X
D codex Dresdensis s. s. XII
V codex Vindobonensis s. XII
R codex Vratislaviensis Rhedig. s. XIII
A codex Vratislaviensis s. XIII
O codex Guelferbytanus s. XV

bonum adunum oculis rotundis superciliosum. Palamedem gracilem longum sapientem animo magnum blandum. Podalirium crassum valentem superbū tristem. Machaonem fortem magnum certum prudenter patientem misericordem. Merionem rufum medicorū statura corpore rotundo viriosum pertinacem crudelē inpatientem. Briseidam formosam non alta statura candidam capillo flavo et molli superciliis iunctis oculis venustis corpore aequali blandam affabilem verecundam animo simplici piā.¹⁰

XIV. Deinde ornati cum **classe** Graeci Athenas convenerunt: Agamemnon ex Mycenis cum navibus numero C, Menelaus ex Sparta cum navibus numero LX, Arcesilaus et Prothoenor ex Boeotia cum navibus numero L, Ascalaphus et Ialmenus ex Orchomeno cum navibus numero XXX, Epistrophus et Scheidius ex Phocide cum navibus numero XL, Ajax Telamonius ex Salamina adduxit secum Teucrum fratrem, ex Buprasione Amphimachum Diorem Thalpium Polyxe-

XIV. Dictys Cretensis I 17.

1 bonum] toruum O | fort. aduncum naso 3 Polachrinus G 7 Bereseidam G Breseidam L 8 molli et flauo G 11 classibus G | Graeci add. A ante Athenas 12 micenis LG | cum et numero fere om. G, in eodem numeri litteris scripti sunt 14 archelaus LG archesilaus V Archesilaus Mercerius | protenor LG | boetia G 15 Ascolofus L Ascalaphus G | alimus L alumnus G almenus F corr. in edit. Venet. 1499 | orcoemoenio L oreomeno G corr. Mer-

corporē equali-blandam·affabilē·verecundā·animō pīā·Deinde
ornati cū **classib:** athenas conuener²·Agamīnon ex micenis·cum

Cod. Sang. 197: Deinde ornati cu(m) **classib(us)** athenas conuener(unt)

Encoding Variants

<listWit> (witness list) lists definitions for all the witnesses referred to by a critical apparatus, optionally grouped hierarchically.

<witness> (witness) contains either a description of a single witness referred to within the critical apparatus, or a list of witnesses which is to be referred to by a single sigil.

@xml:id (identifier) provides a unique identifier for the element bearing the attribute.

Encoding Variants

```
<sourceDesc>
  <listWit>
    <witness xml:id="A">Leipzig, Universitätsbibliothek, Cod. Haen. 3518 </witness>
    <witness xml:id="B">Munich, Bayerische Staatsbibliothek, Clm 305v</witness>
    [...]
  </listWit>
</sourceDesc>
```

msDesc in witness

```
<sourceDesc>
  <listWit>
    <witness xml:id="A">
      <msDesc>
        <msIdentifier>
          <country>Germany</country>
          <settlement>Leipzig</settlement>
          <repository>Universitätsbibliothek</repository>
          <idno>Cod. Haen. 3518</idno>
        </msIdentifier>
        <msContents> [...] </msContents>
      </msDesc>
    </witness>
    [...]
  </listWit>
</sourceDesc>
```

Encoding Variants

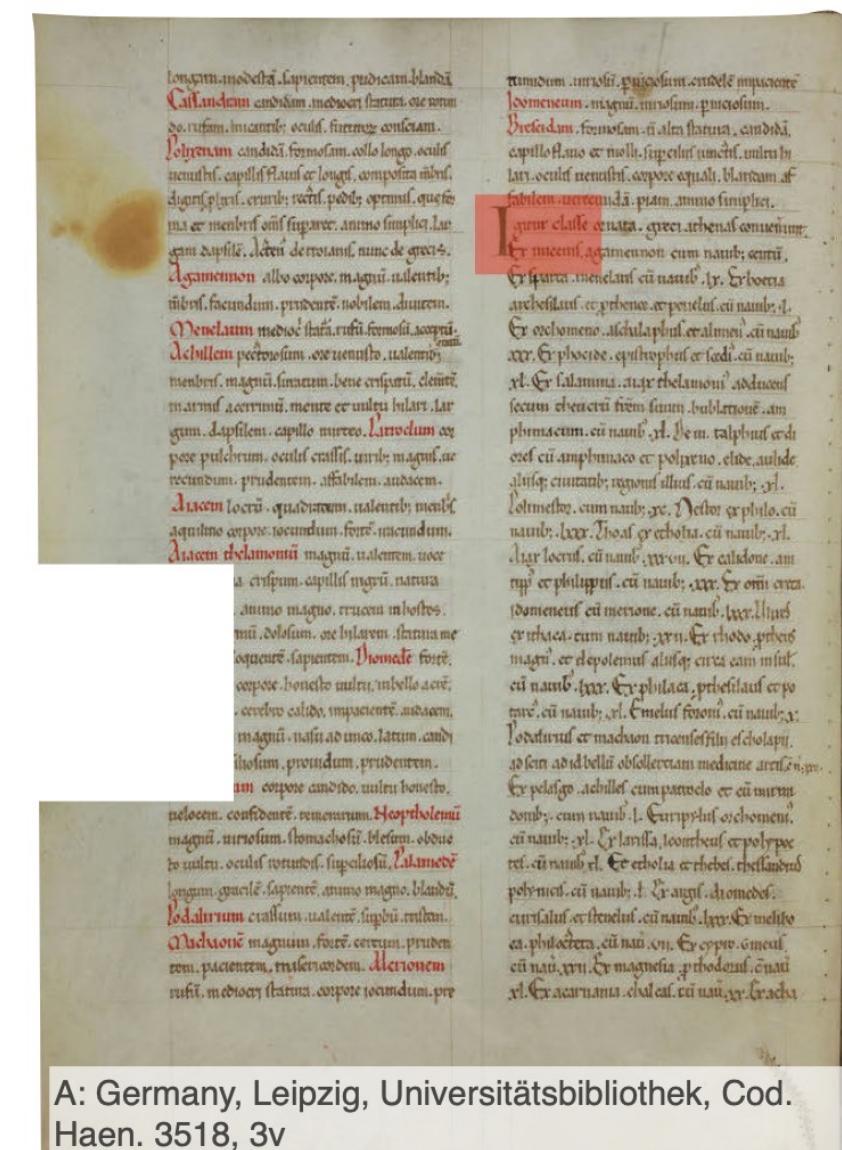
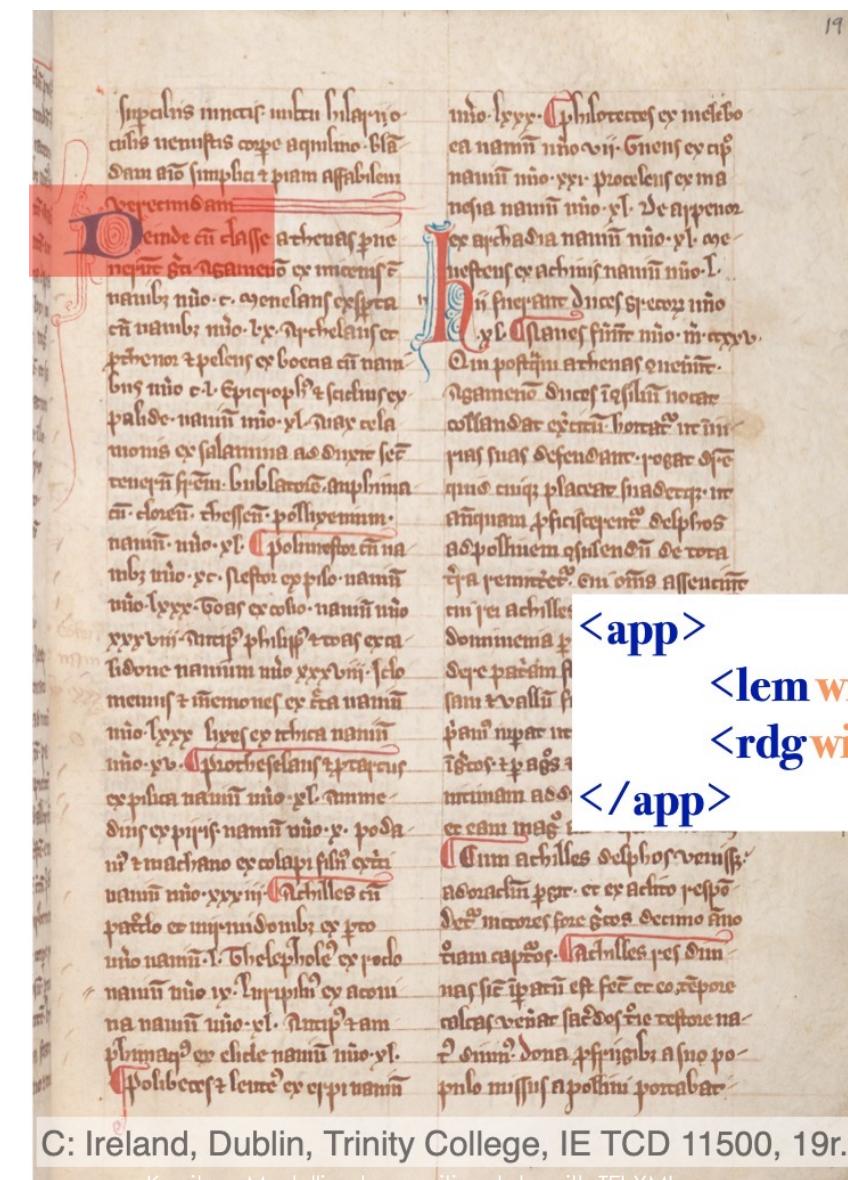
<app> (apparatus entry) contains one entry in a critical apparatus, with an optional lemma and usually one or more readings or notes on the relevant passage.

@type classifies the variation contained in this element according to some convenient typology.

<lem> (lemma) contains the lemma, or base text, of a textual variation

<rdg> (reading) contains a single reading within a textual variation.

@wit (witness or witnesses) contains a space-delimited list of one or more pointers indicating the witnesses which attest to a given reading.



Exercise 1: Editing

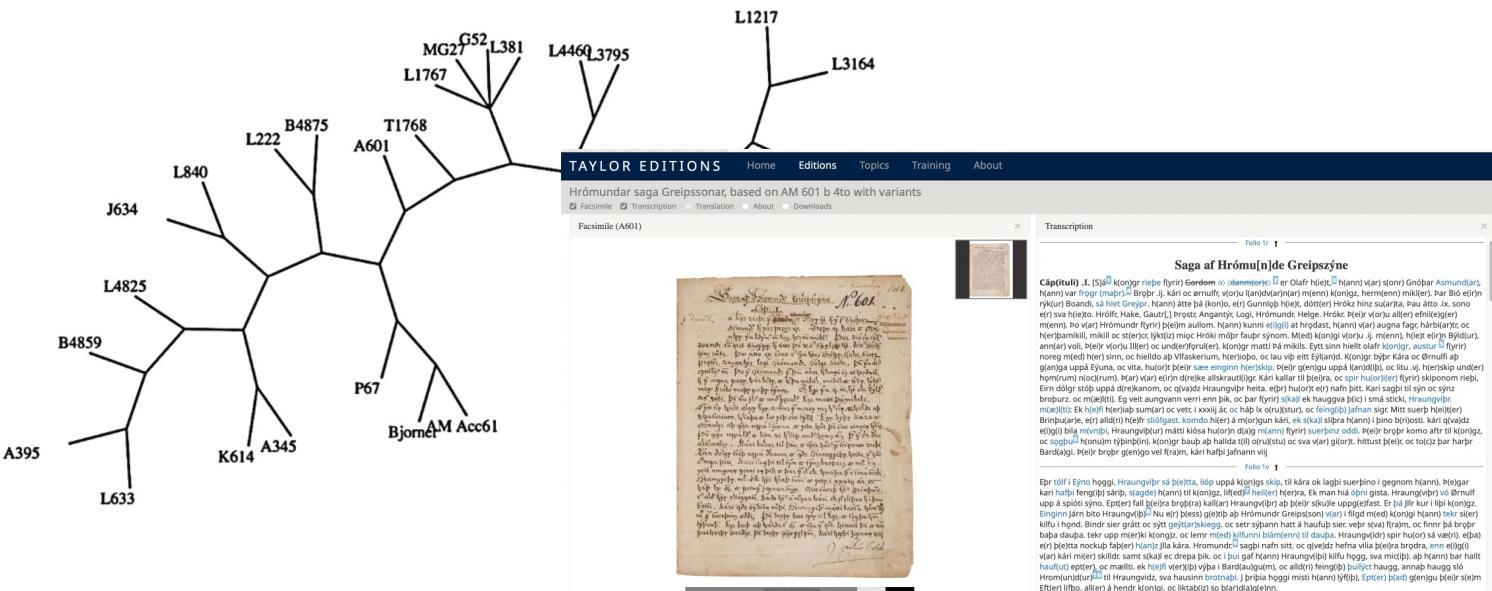
(Group work – 4 people in each group)

- Using Dares_edition_starter.xml prepare an edition of chapter 14 of Dares' *Historia*.
- Encode titleStmt, publicationStmt, sourceDesc with list of witnesses
- Your edition will include variants from four manuscripts:
 - **1. Ireland, Dublin, Trinity College, IE TCD 11500 (Week5/Ex2)**
 - **2. Germany, Munich, Bayerische Staatsbibliothek, Clm 305 (Week5/Ex2)**
 - **3. Switzerland, St Gall, Stiftsbibliothek, Cod. Sang. 197 (Week5/Ex2)**
 - **4. Germany, Leipzig, Universitätsbibliothek, Cod. Haen. 3518 (Week4/Ex2)**
- Entry for each witness should contain msIdentifier and msContents with at least one msItem that gives the locus of Dares' *Historia* (use your Leipzig msDesc for inspiration)

Exercise 1: Transformations

- ▶ Use your file or **Dares_editionSpoiler.xml (Week7/Ex1)**
- ▶ Transform your XML encoding into HTML (with default TEI scenarios)
- ▶ Transform your XML encoding into TEX (with a custom XSLT)
 - ▶ Use the XSLT file **transform_ed_with_app.xsl (Week7/Ex1)** [save as txt]
 - ▶ How to: **Guidelines_Transformation_Scenario.pdf (Week1/Ex3)**

XML to



Editions of Hrómundar saga

Source: https://editions.mml.ox.ac.uk/editions/hromundar_A601/ (left)
<https://ora.ox.ac.uk/objects/uuid:403f3a92-b92c-4d07-8945-d74513c95cac> (right)

Saga af Hrómu(n)de Greipszýne

Cápituli I.

(Sá¹) kongr rieþe fyrir {Gordom²} ³ danmorc⁴ er Olaf hiet⁵ hann var sonr Gnóþar Astmundar, hann var frög⁶ maf⁷. Broþr .ij. kári oc ɔrnulfr, voru landvarnar menn kongz, hermann mikler.

- ⁵ Þar Bió eirn rýkur Boandi, sá hiet Greýpr. hann átte þá kono, er Gunnloð hiet, dóttir Hrólfur hinz suarta, Þau átto .ix. sono er svá hieto. Hrólfur, Hake, Gautr, Þróstr, Angantýr, Logi, Hrómundr. Helge, Hrókr. Þeir voru aller efnileger menn. Þo var Hrómundr fyrir þeim aullom. hann kunni eigi at hröðast, hann var augna fagr, hárbiart, oc herþamikill, mikill oc sterçr, lyktz miðc Hróki móþr faþur sýnom. Med kongr voru .ii. menn, kleyr eirn Býldur, amþarvar, Þeir voru Iller⁸ oc underfóruðar. Kongr matti þá mikil. Ett sinn hielt olaf kongr, austur⁹ fyrir noreg mikil sinn, en óf að Vlfaskerum. Heriþo, oc lau við Eýland. Kongr býr Kára oc Órnulfi gengið upp Eýuna, ec vita, hundru þremur. Þeir gengið upp Vánheimr, en óf að heimskipu undir homrumn. Þar var sín dreika alskráður. Kári kallar til þeirar, oc spít hútorfinn fyrir skiponum reið. Eirn dolgr stóð upð drékanom, oc qjáðré Hraungviðr heita. eþrið huoñr eñ nafn bitt. Kari sagi til sín oc synz broþur. Þeir miðaði. Eg veit aungvann verri enn pik, oc þær fyrir sér ek skreyði því til káki. Hraungviðr miðaði. Þeir höggva sér með vörum, en óf að hraungviðr. Kári sagi til sín, en óf að heimskipu. Bríþutuðr, eir allidri) Kefr slöfugast, komdo hreñs að mórgum kári, en ekka silra (náði) í því. Kongr sueriz oddi. Þeir broþr komo að til kóngz, og sögðu Konum týlunum, kongr bað ap halli (til) orðustuð os waðr gíðor, hitust þeir; oc toðz par harþardagi. Meir broðr gengið voru fráum, kári hufi jafnum við.

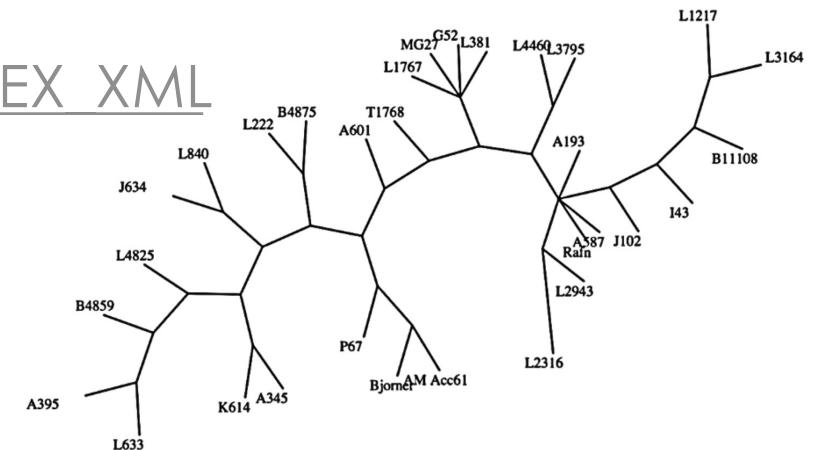
Eirn ríði Eyrn hogg, Hraungviðr sá kámen, legg uppa hraða skop, til kára sé lagð surjensu gegnum hanni. Þeigar kári halfl fengið aðrir, sláðuðu hóum til kóngz, Þeigil hómen Hraungviðr. Ók man hildi línu giza, Hraungviðr við Órnulf upp a spóti syns. Eptir fall þeirra brögðraði kallari Hraungviðr að þeir skuljuðe uppgjæfest. Er þá illur i því Kongz, Einginn járn bito Hraungviðr. Nu erji kress glegð ab Hrómundr Greipsson¹⁰ vian i flig meðd kongri (nánni) teir sér kári. Hraungviðr sér gríðr oc sýrt peysaranking, oc set sjánum hatt að haugfud stendur skul frá með, oc finn jáð brögr báðar. Þeir höggva sér með vörum, en óf að hraungviðr. Kári sagi til sín, en óf að heimskipu. Bríþutuðr (eir) þeitsta mokub fajjey (hanji) til kára. Hrómundr¹¹ sagi nafn sitt, oc qjelvd hefna villa þeirra broþra, enn ekki (var) kári miður skilfð, samt sjáð er drepa bik, oc [þ]uð gaf hænn Hraungviðr kíllo hogg, swa misclj. ab hænni bar hallt haugfud, enn miður skilfð, samt sjáð er drepa bik, oc [þ]uð gaf hænn Hraungviðr kíllo hogg, annan haugg sú Hrómundur¹² til Hraungviðr, swa hæsum broður, j. þeia hogggi mest hænni tylliþ. Eptir] þaði gengið þeir sér

- ⁶ riebel red T1768. ⁷ Þið danmorc¹³ + T1768, A345, P67, L222, B4859.
- ⁸ Astmundar] + d. A587, A193. ⁹ frög maf¹⁴ mafr frægr A587, A193. ¹⁰ sá hiet Greypur hier A345. ¹¹ eigi ei A193, P67, L222. ¹² kongr + A345. ¹³ austur] norþr A345, vestur P67. ¹⁴ szei] seo, A345, B4859. ¹⁵ einginn] eingi A587, engi A193, T1768, A345, L222, B4859. ¹⁶ herkip] skip 5, L222, B4859. ¹⁷ spjrl] + eftir P67. ¹⁸ huoner] hver A587, A193, hvor A345, P67, L222, hueri B4859. ¹⁹ skal] vil P67. ²⁰ Hraungviðr, maf[21] hefer A345, L222, B4859.

Alternatives

- ▶ **XML to Tree with PARS (with Python):**
https://github.com/KAKDH/stemma_XML_Pars

- ▶ **TEX to XML (with Python):**
https://github.com/KAKDH/edition_TEX_XML



Annotations

Kapitan, Modelling humanities data with TEI-XML

For Linguistic Analysis: Part of Speech, Morpho-Syntactic Annotations...

- ▶ **Chapter 18:** Simple Analytic Mechanisms
 - ▶ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/A1.html>

- ▶ **<s>** (s-unit) contains a sentence-like division of a text.
- ▶ **<cl>** (clause) represents a grammatical clause.
- ▶ **<phr>** (phrase) represents a grammatical phrase.
- ▶ **<w>** (word) represents a grammatical word.



@lemma

provides a lemma (base form) for the word, typically uninflected and serving both as an identifier (e.g. in dictionary contexts, as a headword), and as a basis for potential inflections.

@lemmaRef

provides a pointer to a definition of the lemma for the word, for example in an online lexicon.

```
<s xml:lang="la">
  <w lemma="timeo">timeo</w>
  <w lemma="danaii">Danaos</w>
  <w lemma="et">et</w>
  <w lemma="donum">dona</w>
  <w lemma="fero">ferentes</w>
</s>
```

```
<s xml:lang="la">
  <w lemmaRef="http://lexicon.org/latin.xml#timeo">timeo</w>
  <w lemmaRef="http://lexicon.org/latin.xml#danaii">Danaos</w>
  <!-- ... -->
</s>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-s.html>

@pos (part of speech) indicates the part of speech assigned to a token (i.e. information on whether it is a noun, adjective, or verb), usually according to some official reference vocabulary.

Examples of vocabularies: for German STTA (Stuttgart-Tübingen Tagset); for English: CLAWS (The Constituent Likelihood Automatic Word-tagging System); For French: French TreeTagger Part-of-Speech Tags.

The victim's friends told police that
Kruger drove into the quarry
and never surfaced

```
<s><w pos="AT0">The</w> <w pos="NN1">victim</w><w pos="POS">'s</w> <w pos="NN2">friends</w>  
<w pos="VVD">told</w> <w pos="NN2">police</w> <w pos="CJT">that</w> <w pos="NP0">Kruger</w>  
<w pos="VVD">drove</w> <w pos="PRP">into</w> <w pos="AT0">the</w> <w pos="NN1">quarry</w>  
<w pos="CJC">and</w> <w pos="AV0">never</w> <w pos="SENT">surfaced</w><pc pos="PUN">. </pc></s>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-s.html>

```
<s type="sentence">
<phr ana="#n">
<phr ana="#gn">
<w ana="#ATO">The</w>
<w ana="#NN1">victim</w>
<m ana="#POS">'s</m>
</phr>
<w ana="#NN2">friends</w>
</phr>
<phr ana="#v">
<w ana="#VVD">told</w>
<phr ana="#n">
<w ana="#NN2">police</w>
</phr>
<cl ana="#fn">
<w ana="#CJT">that</w>
<phr ana="#n">
<w ana="#NP0">Krueger</w>
</phr>
<phr ana="#v">
<phr ana="#v1">
<w ana="#VVD">drove</w>
<phr ana="#pr">
<w ana="#PRP">into</w>
<phr ana="#n">
<w ana="#ATO">the</w>
<w ana="#NN1">quarry</w>
</phr>
</phr>
</phr>
<w ana="#CJC">and</w>
<phr ana="#v2">
<w ana="#AV0">never</w>
<w ana="#VVD">surfaced</w>
</phr>
</phr>
</cl>
</phr>
<c ana="#pun">.</c>
</s>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-s.html>

The victim's friends
told police
that Kruger drove into the quarry
and
never surfaced

More linguistic attributes in the TEI Guidelines:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.linguistic.html>

For Personography:

► Chapter 14: Names, Dates, People, and Places

► <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>

<persName> (personal name) contains a proper noun or proper-noun phrase referring to a person, possibly including one or more of the person's forenames, surnames, honorifics, added names, etc.

<surname> (surname) contains a family (inherited) name, as opposed to a given, baptismal, or nick name.

<forename> (forename) contains a forename, given or baptismal name.

```
<persName>
  <forename type="first">Franklin</forename>
  <forename type="middle">Delano</forename>
  <surname>Roosevelt</surname>
</persName>
<persName>
  <forename type="given">Margaret</forename>
  <forename type="unused">Hilda</forename>
  <surname type="birth">Roberts</surname>
  <surname type="married">Thatcher</surname>
</persName>
```



<listPerson> (list of persons) contains a list of descriptions, each of which provides information about an identifiable person or a group of people

<person> (person) provides information about an identifiable individual, for example a participant in a language interaction, or a person referred to in a historical source.

```
<person>
  <persName>
    <forename>Miss Major</forename>
    <surname>Griffin-Gracy</surname>
  </persName>
  <birth when="1940-10-25"/>
  <gender value="transWoman"
    evidence="selfIdentification">trans woman</gender>
  <persPronouns value="she"
    evidence="selfIdentification"/>
  <note>
    <p>Veteran of the Stonewall Riots. Founder of the
      Griffin-Gracy Educational Retreat and Historical
      Center (the House of GG). Activist and advocate for
      transgender and gender-nonconforming people of
      color.</p>
  </note>
</person>
```

Kapitan, Modelling humanities data with TEI-XML.



@xml:id (identifier) provides a unique identifier for the element bearing the attribute.

@ref (reference) provides an explicit means of locating a full definition or identity for the entity being named by means of one or more URIs.

```
<person xml:id="ArnMag">
  <persName xml:lang="is">Árni Magnússon</persName>
  <persName xml:lang="da">Arne Magnusson</persName>
  <persName xml:lang="la">Arnas Magnæus</persName>
</person>
```

In 1702 <persName ref="#ArnMag">Árni</name> wrote
in one of his letters to his friend...

Dares, ch. 14:

Deinde ornati cum classe Graeci Athenas convenerunt:

Agamemnon ex Mycenis...

```
  ...
  ...
  ...
21 <listPerson>
22   <person xml:id="Aga_001">
23     <persName>Agamemnon</persName>
24     <idno type="wikidata">Q128176</idno>
25     <note>King of Mycenae; son of Atreus in Greek mythology.</note>
26   </person>
27 </listPerson>
28 </sourceDesc>
29 </fileDesc>
30 </teiHeader>
31 <text>
32   <body>
33     <div type="chapter" n="14">
34       <p>
35         <s>Deinde ornati cum classe Graeci Athenas convenerunt:</s>
36         <s><persName ref="#Aga_001">Agamemnon</persName> ex Mycenis cum navibus numero
            type="ships" value="100"> C. </s> <s> Menelaus ex Sparta cum navibus numero
            type="ships" value="100"> C. </s>
```



Item Discussion

Read

View history

Search Wikidata

Agamemnon (Q128176)

son of Atreus in Greek mythology

▼ In more languages

Configure

Language	Label	Description	Also known as
English	Agamemnon	son of Atreus in Greek mythology	
French	Agamemnon	personnage de la mythologie grecque	
Spanish	Agamenón	héroe de la Grecia antigua	Agamemnon Agamenon
German	Agamemnon	König der griechischen Mythologie	

All entered languages

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data

Create a new Lexeme
Recent changes
Random Lexeme

Tools

What links here
Related changes
Special pages
Modelling human
Permanent link

Statements

instance of



mythological Greek character

edit

0 references

Search

Select Field:

Personal Names

Select Index:

All VIAF

Search Terms:

Agamemnon

Search

Agamemnon, mythologie grecque 

VIAF ID: 173096818 (Personal)

Permalink: <http://viaf.org/viaf/173096818>

Preferred Forms

 100 0 _ ta Agamemnon, tc mythologie grecque



Selected Titles

Kelton, Modeling humanities data with TEI-XML

For Spatial Analysis:

- ▶ **Chapter 14:** Names, Dates, People, and Places
 - ▶ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>
- <placeName>** contains an absolute or relative place name.

```
<placeName>
  <settlement>Rochester</settlement>
  <region>New York</region>
</placeName>
```



<listPlace> (list of places) contains a list of places, optionally followed by a list of relationships (other than containment) defined amongst them.

<place> (place) contains data about a geographic location

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

<geo> (geographical coordinates) contains any expression of a set of geographic coordinates, representing a point, line, or area on the surface of the earth in some notation.

Dares, ch. 14:

Deinde ornati cum classe Graeci **Athenas** convenerunt:
Agamemnon...

```
27 </listPerson>
28 <listPlace>
29   <place xml:id="Athe_001">
30     <placeName>Athens</placeName>
31     <location>
32       <geo>37.9839 23.7278</geo>
33       <!-- lat long -->
34     </location>
35     <idno type="wikidata">Q1524</idno>
36   </place>
37 </listPlace>
38 </sourceDesc>
39 </fileDesc>
40 </teiHeader>
41 <text>
42   <body>
43     <div type="chapter" n="14">
44       <p>
45         <s>Deinde ornati cum classe Graeci <placeName ref="#Athe_001">Athens</placeName> convenerunt:</s>
```



The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.

Athens [advanced search] all countries

enter a location name, ex: "Paris", "Mount Everest", "New York"

Browse the names

- [Countries](#)
- [Postal codes](#)
- [Country statistics](#)
- [Recent modifications](#)

Information

- [About GeoNames](#)
- [Data Sources](#)
- [User manual](#)
- [Ambassadors and Team](#)
- [Forum](#)
- [Blog](#)
- [Mailing list](#)
- [Commercial Support and Consulting](#)

Download

- [Info](#)
- [Free Gazetteer Data](#)
- [Free Postal Code Data](#)
- [Premium Data](#)

Web Services

- [Overview](#)
- [Documentation](#)
- [Client Libraries](#)
- [Premium Web Services](#)

Sponsoring

Kapitan, Modelling humanities data with TEI-XML

Search Tools & Databases

- Library Catalog
- ▶ Getty Research Portal
- Collection Inventories & Finding Aids
- ▶ Photo Archive
- ▶ Research Guides & Bibliographies
- Digital Collections
- ▶ Article & Research Databases
- ▶ Collecting & Provenance Research
- BHA & RILA

▼ Getty Vocabularies

- ▶ Art & Architecture Thesaurus (AAT) ®
- ▶ Cultural Objects Name Authority (CONA) ®
- ▼ Getty Thesaurus of Geographic Names (TGN) ®
- About TGN

Getty Thesaurus of Geographic Names® Online

Search the TGN

[? Help](#)[Search](#)

Place Type:

[Lookup](#)[Clear](#)

Nation:

[Lookup](#)[Pop-up Search](#)[Browse the TGN hierarchies](#)[Copyright information](#)

Search Tips

For the Find Name field, you may use **AND** and **OR** [e.g., 1) san carlos, 2) carlos OR charles, 3) carl* OR charl*, 4) san AND carlos, 5) carlos AND (san OR saint), 6) (carlos OR charles) AND (san OR saint)] Boolean operators must be in all caps (AND and OR). Wildcard is the asterisk (*); right truncation only. To find an exact match rather than a key word, use quotes

Have a Question?

- ✉ Contact the Vocabulary Program

More Applications



Drama:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DR.html>



Letters:

<https://encoding-correspondence.bbaw.de/v1/index.html>



Language corpora:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CC.html>



Dictionaries:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>



Music:

<https://music-encoding.org>

Exercise 2

- ▶ Download the XML file **Dares_Historia_Chapter_14_annotations_starter.xml**
- ▶ Divide the document into sentences **<s>**
- ▶ Start annotating **names** and **numbers** following the examples provided in the document, using **<persName>**, **<placeName>** and **<num>** for numbers of ships.
- ▶ **Remember to add new people and places to your Source Description and give them unique identifiers (xml:id)!**
- ▶ **Send your annotated file to Katarzyna by Tuesday 19 November, 23:59.**