

Contributing to the annotation of the WoPoss corpus

Francesca Dell'Oro

francesca.delloro@unine.ch

PI of the WoPoss project



Roadmap

1. WoPoss project
2. WoPoss corpus
3. how you could contribute
 - the correction of the linguistic annotation
 - the annotation of modality
 - Exercises

1. The WoPoss project

- WoPoss = *A World of Possibilities. Modal pathways over an extra-long period of time: the diachrony of modality in the Latin language*
- project funded by the Swiss National Science Foundation
- team: 5 people at the moment

1. Goals

- main goal: reconstruct the evolution of modal meanings from the prehistory of the Latin language up to the 7th century CE
- resource to achieve the goal: setting up a diachronic balanced corpus of Latin texts

2. WoPoss corpus

- main features:
 - diachronic
 - balanced
- annotation
 - automatic
 - manual

Query the corpus

- <https://woposs.unine.ch/form.html>

The screenshot shows a web browser window with the title "WoPoss: A World of Possibilities". The URL in the address bar is "woposs.unine.ch/form-marker.html". The page content is titled "Search by modal marker" and includes a heading "Choose a modal marker (required)". Below this, there are several sections of checkboxes for different parts of speech and constructions:

- Adjectives**: aequus, aptus, certus, dubius, illicitus, impossibilis, incertus, ineptus, licitus, necessarius, possibilis
- Adverbs**: certe, dubie, forsani, forsitan, fortasse, forte, indubitate/indubitanter, necessarie, possibiliter, probabiliter
- Constructions with *esse***: est + infinitive, ius est, meum est, necesse/um/us est, opus est, usus est
- Constructions with *habeo***: habeo + infinitive, necesse/um/us habeo
- Nouns**: dubium, facultas, possibilitas, potestas, necessitas, necessitudo, probabilitas, voluntas
- Verbal adjectives**: -bilis, -ndus, -turus
- Verbs**: debedo, decet, licet, nequeo, nolo, malo, oportet, possum, queo, valet, volo
- Optional filters**:
 - Description of the marker
 - Description of the scope
 - Description of the modal relation
 - Description of the work

A blue "Submit" button is located at the bottom of the form.

2. Automatic annotation

We use models trained on treebank data (UD Perseus) in order to carry out an automatic annotation of:

tokenisation

PoS (parts of speech)

lemmas

morpho-syntactic features

[dependencies]

*If you are interested in models from
treebanks...*

SunoikisisDC Fall 2020

Session 7. Using Treebanks

Thursday Nov 19, 2020

- YouTube link: <https://youtu.be/EfHnR0nl9tE>

If you want to have an idea of the level of performance of the model(s), try...

https://github.com/WoPoss/automatic_annotation

3. How can you contribute?

- the performance of the automatic annotation is not bad, but there are some errors
- we appreciate *external help* to correct such errors

3. Examples of errors in the automatic annotation

- tokens: the text elements have not been correctly separated e.g.,
`<w>etego</w>` instead of
`<w>et</w>`
`<w>ego</w>`
- PoS: the token has not been attributed to the correct part of speech
e.g., *ego* associated with pos = “VERB” instead of
`<w pos="VERBPRON">ego</w>`

- lemma: the token has not been associated to the correct lemma e.g., *facio* as lemma for *ego* instead of

```
<w pos="PRON" lemma = "facioego">ego</w>
```

- morpho-syntactic description: the form has not been correctly analysed e.g.,

ego analysed as an accusative instead of

```
<w Case=AccNom|Gender=Masc|Number=Sing pos="PRON" lemma =  
"ego">ego</w>
```

3. In concrete terms, how can I contribute?

1. go to GitHub (<https://github.com/>) and create an account
2. go to the WoPoss repository ‘Contributing’
(<https://github.com/WoPoss-project/Contributing>)
3. fork it
4. open the file and do the changes
5. commit them to your fork
6. pull a request
7. we will be able to accept the changes

Step 2: in the WoPoss-project in GitHub find the repository ‘Contributing’

The screenshot shows the GitHub interface for the 'WoPoss' repository. The top navigation bar includes links for Overview, Repositories (22), Projects, Packages, Teams, People (2), and Settings. A search bar at the top allows users to 'Find a repository...'. Below the search bar are filters for Type, Language, and Sort, along with a green 'New repository' button. The main content area displays a card for the 'Contributing' repository, which is public. The card shows the repository name in blue, its status as Public, and various metrics: 1 fork, 0 stars, 0 issues, and 1 pull request. It also indicates that the repository was updated 29 minutes ago.

WoPoss

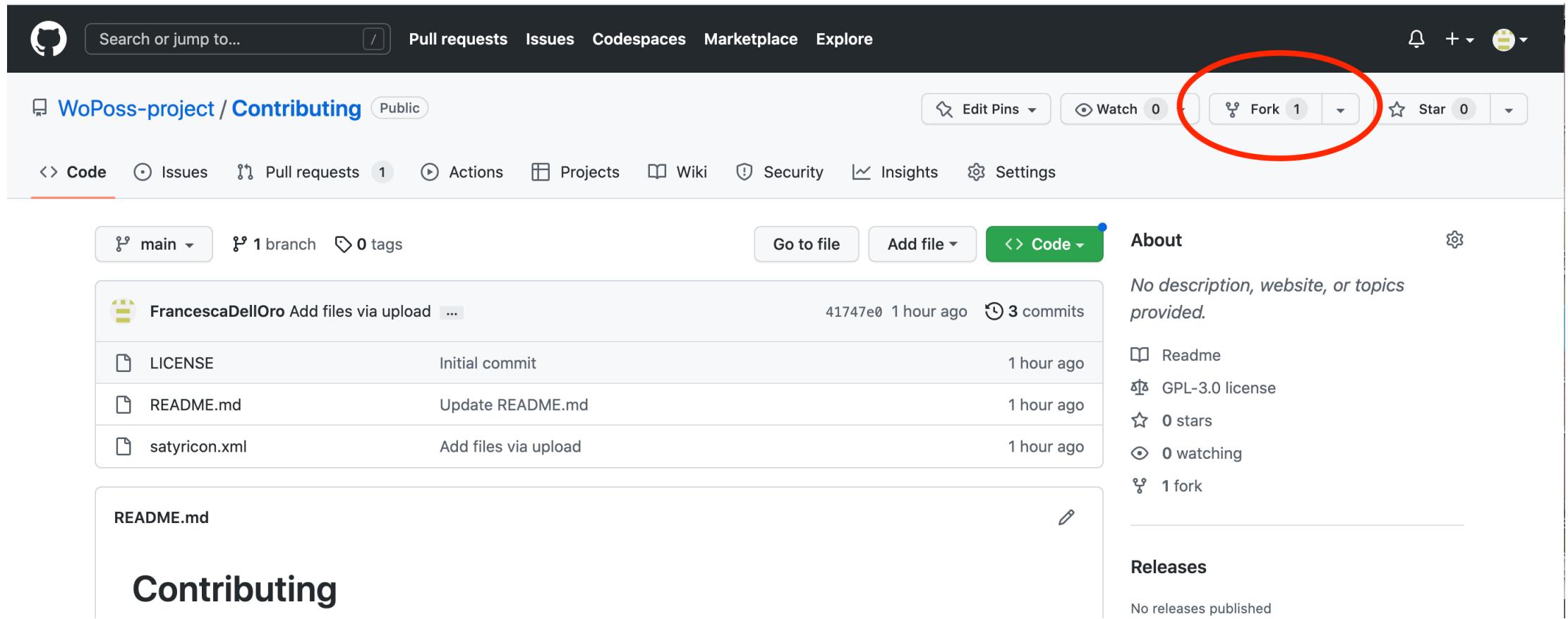
Overview Repositories 22 Projects Packages Teams People 2 Settings

Find a repository... Type Language Sort New repository

Contributing Public

GPL-3.0 1 0 1 Updated 29 minutes ago

Step 3: fork the repository



The screenshot shows a GitHub repository page for "WoPoss-project / Contributing". The top navigation bar includes links for Pull requests, Issues, Codespaces, Marketplace, and Explore. The repository name "WoPoss-project / Contributing" is displayed, along with a "Public" badge. The main navigation bar below shows Code (selected), Issues, Pull requests (1), Actions, Projects, Wiki, Security, Insights, and Settings. On the left, there's a code editor with a commit history for "main" branch, showing four commits from "FrancescaDellOro". The right sidebar contains sections for About (no description, website, or topics provided), Readme, GPL-3.0 license, 0 stars, 0 watching, and 1 fork. A red circle highlights the "Fork" button in the top right corner of the main content area.

[Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)[Edit Pins](#) [Watch](#) [Fork](#) [Star](#)

Your existing forks [X](#)

You don't have any forks of this repository.

+ Create a new fork

Create a new fork

A *fork* is a copy of a repository. Forking a repository allows you to freely experiment with changes without affecting the original project. [View existing forks](#).

Owner *



Repository name *

Contributing



By default, forks are named the same as their upstream repository. You can customize the name to distinguish it further.

Description (optional)

Copy the `main` branch only

Contribute back to WoPoss-project/Contributing by adding your own branch. [Learn more](#).

You are creating a fork in your personal account.

[Create fork](#)

Step 4. open one of the files of the “Satyricon” and do the changes

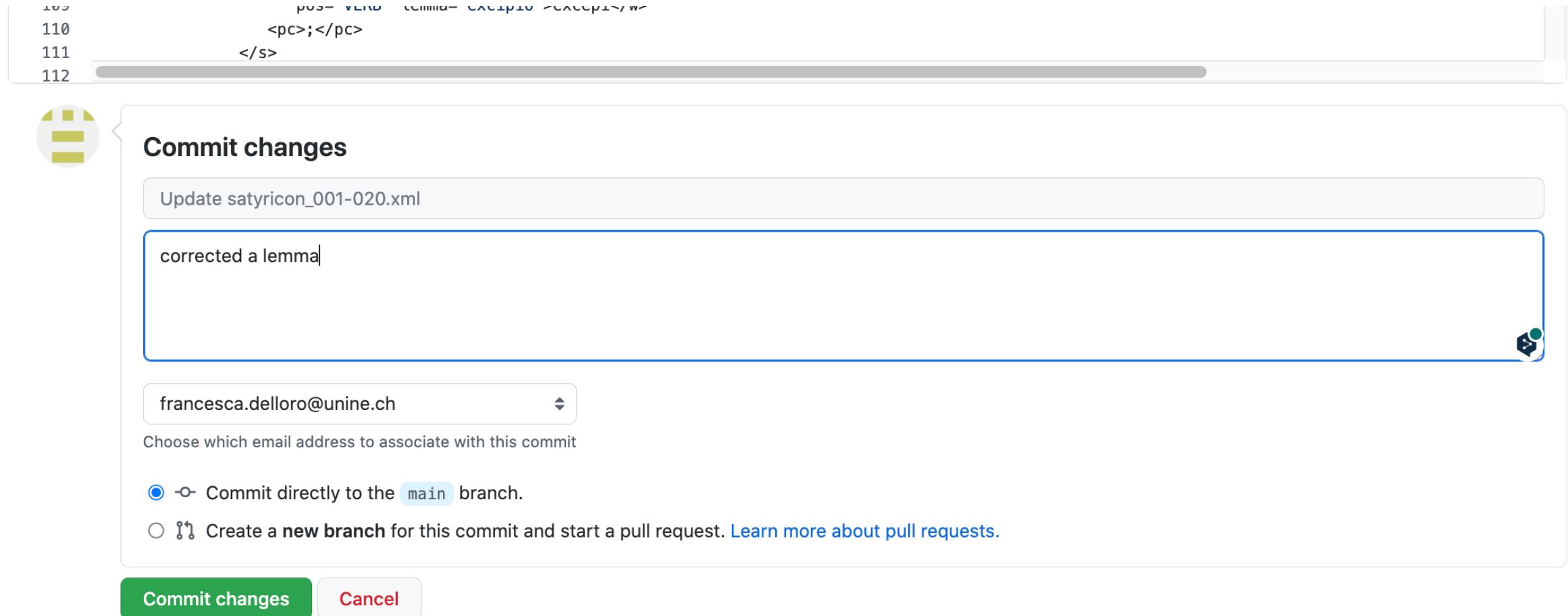
The screenshot shows a GitHub repository page for 'WoPoss-project / Contributing'. The repository is public and has 1 pull request, 2 forks, and 0 stars. The main navigation bar includes Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. Below the navigation, it shows the 'main' branch and the file 'Contributing / satyricon_001-020.xml'. The commit history shows a recent commit by 'HelenaSabel' titled 'satyricon divided in smaller files' made 28 minutes ago. The commit details show 1 contributor. The file content is an XML document with 7375 lines and 378 KB size. A context menu is open over the file content, with the 'Edit this file' option highlighted and circled in green. Other options in the menu include Raw, Blame, and a dropdown menu.

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="https://raw.githubusercontent.com/WoPoss-project/annotation-schemes/master/schemas/out/tei_woposs.rng" type="application/xml"?
3 <?xml-model href="https://raw.githubusercontent.com/WoPoss-project/annotation-schemes/master/schemas/out/tei_woposs.rng" type="application/xml"?
4 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="satyricon">
5   <teiHeader>
6     <fileDesc>
7       <titleStmt>
8         <title>Satyricon</title>
9         <title type="short">Satyricon</title>
10        <author>G. Petronius Arbiter</author>
11        <principal>Francesca Dell'Oro</principal>
12      </titleStmt>
13      <editionStmt>
```

Some errors in the first passage of the Satyricon

```
</teiHeader>
<text>
  <body>
    <div type="section" n="1">
      <ab>
        <s n="2">
          <anchor type="solid" subtype="start" />
          <w pos="ADV" lemma="nyum">Num</w>
          <w msd="Case=Abl|Gender=Neut|Number=Sing" pos="ADJ" lemma="alius">alio</w>
          <w msd="Case=Abl|Gender=Neut|Number=Sing" pos="NOUN" lemma="genus">genere</w>
          <w msd="Case=Gen|Gender=Fem|Number=Plur" pos="NOUN" lemma="furia">furiarum</w>
          <w msd="Case=Nom|Gender=Masc|Number=Plur" pos="NOUN" lemma="declamo"
              >declamatores</w>
          <w msd="Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin|Voice=Pass"
              pos="VERB" lemma="inquieto">inquietantur</w>
          <pc>, </pc>
          <w msd="Case=Nom|Gender=Masc|Number=Plur" pos="PRON" lemma="qui">qui</w>
          <w msd="Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin|Voice=Act"
              pos="VERB" lemma="clamo">clament</w>
          <pc>:</pc>
        </s>
        <s n="3">
          <pc>'</pc>
          <w msd="Case=Acc|Gender=Neut|Number=Plur" pos="NOUN" lemma="fiha">haec</w>
          <w msd="Case=Acc|Gender=Neut|Number=Plur" pos="NOUN" lemma="vulnus">vulnera</w>
```

Step 5. Commit and propose the changes



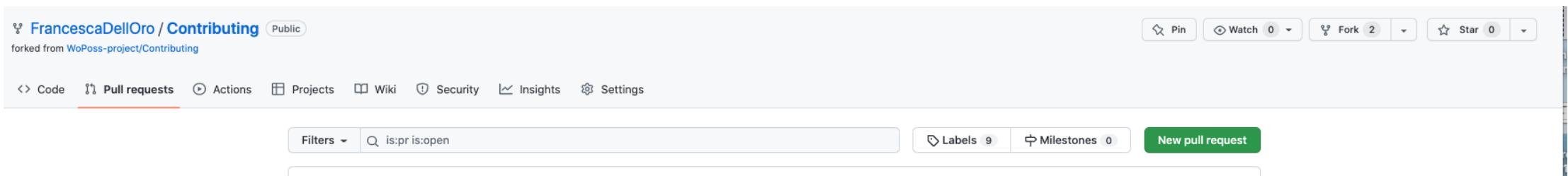
The screenshot shows a GitHub commit dialog over a code editor background. The code editor displays XML code with lines 110 and 111 highlighted:

```
109
110      <pc>;</pc>
111      </s>
112
```

The commit dialog has the following fields:

- Commit changes**: A button with a green icon.
- Update satyricon_001-020.xml**: The file being updated.
- corrected a lemma**: The commit message.
- francesca.delloro@unine.ch**: The email address associated with the commit, with a dropdown arrow.
- Choose which email address to associate with this commit**: A placeholder text.
- Commit directly to the main branch.**: A selected option (radio button).
- Create a new branch for this commit and start a pull request.**: An alternative option.
- Commit changes** and **Cancel**: Buttons at the bottom.

And pull a request



Your changes are tracked

Conversation 0 Commits 1 Checks 0 Files changed 1 +1 -1

Changes from all commits ▾ File filter ▾ Conversations ▾ Jump to ▾ ⚙ ▾ 0 / 1 files viewed Review changes ▾

satyricon.xml

Line	Change	Text
85	-	<w msd="Case=Abl Gender=Neut Number=Sing" pos="ADJ" lemma="alius">alio</w>
86	-	<w msd="Case=Abl Gender=Neut Number=Sing" pos="NOUN" lemma="genus">genere</w>
87	-	<w msd="Case=Gen Gender=Fem Number=Plur" pos="NOUN" lemma="furia">furiarum</w>
88	-	<w msd="Case=Nom Gender=Masc Number=Plur" pos="NOUN" lemma="declamo"
88	+	<w msd="Case=Nom Gender=Masc Number=Plur" pos="NOUN" lemma="declamator"
89	-	>declamatores</w>
90	-	<w msd="Mood=Ind Number=Plur Person=3 Tense=Pres VerbForm=Fin Voice=Pass"
91	-	pos="VERB" lemma="inquieto">inquietantur</w>

Now you know how to
contribute to the correction of
the automatic annotation

3. Contributing to the manual annotation

- semantics
 - annotation of modality: in the project we rely on the *WoPoss guidelines for annotation*
 - <https://zenodo.org/badge/DOI/10.5281/zenodo.3560951.svg>
 - Inception platform

If you want to know more...

SunoikisisDC Summer 2021

Session 4. Linguistic Annotations of Greek and Latin Inscriptions

Thursday May 6, 2021

YouTube link: https://youtu.be/7jvWi_levC0

3. What is modality?

Expression of the notional domains of necessity, possibility and volition with regard to an imagined situation (state of affairs).

- lexical markers: *possum, debeo, volo, forsitan...*
- morphological markers: *-bilis, -ndus, -turus*
- see I. Nuyts & J. van der Auwera 2016

3. Types of modality

- **epistemic** modality:
 - *That's probably the postman bringing today's newspaper.*
- **deontic** modality: typically obligation and permission
 - *You can go now.*
- **dynamic** modality: typically abilities and needs and general possibilities and necessities
 - *He can stand on his head without using his hands.*

3. How can I contribute to the semantic annotation?

- in order to contribute to the semantic annotation, you need an Inception account (write to the WoPoss team: francesca.delloro@unine.ch)
 1. choose a document among Cicero's letters to Brutus
 2. read the text and look for modal markers (there is a list in the WoPoss guidelines, pp. 8-9: <https://zenodo.org/record/6417878#.Y3ynH-zMJTY>)
 3. when you find a marker, you annotate it as a marker
 4. then you look for the scope of marker (modalised state of affairs)
 5. finally, if you think you have recognised a type of modality, you annotate it

Step 1. Enter in the Inception platform and choose a document

The screenshot shows a web browser window for the INCEpTION platform at the URL <https://woposs.unine.ch/inception/p/woposs---crowdsourced/annotate?5>. The browser interface includes standard navigation buttons (back, forward, search, etc.) and a user profile bar with 'Help', 'fdelloro', 'Log out', and a timer '29 min'.

The main content area displays a modal dialog titled 'Choose Document'. The dialog has a dropdown menu set to 'User fdelloro (me)'. A search bar contains the placeholder 'Filter by name... (open with enter when one is left)'. Below the search bar is a table with the following columns:

State Name
<input checked="" type="radio"/> letters_cicero_brutus_1.01.txt
<input type="radio"/> letters_cicero_brutus_1.02.txt
<input type="radio"/> letters_cicero_brutus_1.02a.txt
<input type="radio"/> letters_cicero_brutus_1.03.txt
<input type="radio"/> letters_cicero_brutus_1.03a.txt
<input type="radio"/> letters_cicero_brutus_1.04.txt
<input type="radio"/> letters_cicero_brutus_1.04a.txt
<input type="radio"/> letters_cicero_brutus_1.05.txt
<input type="radio"/> letters_cicero_brutus_1.06.txt
<input type="radio"/> letters_cicero_brutus_1.07.txt
<input type="radio"/> letters_cicero_brutus_1.08.txt
<input type="radio"/> letters_cicero_brutus_1.09.txt
<input type="radio"/> letters_cicero_brutus_1.10.txt
<input type="radio"/> letters_cicero_brutus_1.11.txt
<input type="radio"/> letters_cicero_brutus_1.12.txt
<input type="radio"/> letters_cicero_brutus_1.13.txt
<input type="radio"/> letters_cicero_brutus_1.14.txt
<input type="radio"/> letters_cicero_brutus_1.15.txt
<input type="radio"/> letters_cicero_brutus_1.16.txt
<input type="radio"/> letters_cicero_brutus_1.17.txt
<input type="radio"/> letters_cicero_brutus_1.18.txt
<input type="radio"/> letters_cicero_brutus_2.01.txt
<input type="radio"/> letters_cicero_brutus_2.02.txt

At the bottom of the dialog is a large red 'X' button. The footer of the page includes the text 'Technische Universität Darmstadt --- Computer Science Department --- INCEpTION --- 25.4 (2022-11-08 12:45:47, build 1aeb414a)'.

Step 2. read the text and look for a modal marker

fdelloro WoPos - crowdsourced letters_cicero_brutus_1.01.txt 11-20 / 23 sentences [doc 1 / 26]

11 12

11	ADJ (MorFea...) multus	PRON (MorFea...) is	NOUN (MorFea...) indicia	CCONJ sed	ADP ad	VERB (MorFea...) scribo	ADV non	ADJ (MorFea...) necessarius	PUNCT .
12	VERB (MorFea...) volo	SCONJ enim	NOUN (MorFea...) testimonium	PRON (MorFea...) hic	PRON (MorFea...) tibi	VERB (MorFea...) video	ADV (MorFea...) potius	NOUN (MorFea...) epistula	PUNCT .

Step 3. annotate the marker selecting the layer “modal unit”

The screenshot shows the INCEption interface for morphological annotation. The main window displays a grid of Latin words from sentences 11 to 15, each with its part of speech and morphological features. Sentence 11 contains the word 'necessaria' highlighted with a yellow box and labeled '(potential) marker'. A context menu is open at this word, with 'Modal unit' selected. The annotation details panel shows the following information:

- Layer:** Modal unit
- Text:** necessaria
- a) Type of modal unit:** (potential) marker
- b) Pertinence:** fragmentary text

No links or relations connect to this annotation.

Step 4. Annotate the scope of markers (usually the verb indicating the state of affairs)

INCEpTION Projects Dashboard

Help fdelloro Log out 29 n

11-20 / 23 sentences [doc 1 / 26]

Layer Modal unit

Create a **Modal relation** relation by drawing an arc between annotations of this layer.

Annotation Delete Clear

Layer Modal unit

No links or relations connect to this annotation.

a) Type of modal unit

(potential) marker
scope unit

11 multa eius indicia sed ad scribendum non necessaria .

12 volo enim testimonium hoc tibi videri potius quam epistulam .

13 augo Antoni beneficio est .

14 is ipse beneficu magnu pars ab tu sum .

15 itaque eum salvis nobis .

Step 4. Propose a type of modality

fdelloro WoPoss - crowdsourced letters_cicero_brutus_1.01.txt 11-20 / 23 sentences [doc 1 / 26]

Layer Modal unit

Create a **Modal relation** relation by drawing an arc between annotations of this layer.

Annotation

Delete Reverse Clear

From scribendum

To necessaria

a) Type of modality

- dynamic
- deontic
- dynamic
- epistemic
- possibility

b) Dynamic subtype (participant control)

Annotations for sentence 11:

- 11 multa eius indicia sed ad scribendum non necessaria .
- 12 volo enim testimonium hic tu video potius quam epistula .
- 13 augeo Antoni beneficio est .
- 14 is ipse benefici magna pars ab tu sum .
- 15 itaque eum salvis nos .
- 16 vellet salvum .

Annotations for sentence 12:

- 12 volo enim testimonium hic tu video potius quam epistula .

Annotations for sentence 13:

- 13 augeo Antoni beneficio est .

Annotations for sentence 14:

- 14 is ipse benefici magna pars ab tu sum .

Annotations for sentence 15:

- 15 itaque eum salvis nos .

Annotations for sentence 16:

- 16 vellet salvum .

Now you know how to
contribute to the annotation of
modality

Why contributing to WoPoss?

- read Latin texts
- improve your knowledge of Latin language and literature (or inscriptions)
- improve your Latin grammar
- improve your ability to interpret Latin texts (modality annotation)
- your work is citable

3. How is my work recognised?

- responsibility statement in the files in GitHub (open access)

```
19      <name>Helena Bermúdez Sabel</name>
20  </respStmt>
21  <respStmt>
22      <resp>Semantic annotation and correction of the linguistic annotation of modal
23          passages</resp>
24      <name>Séverine Reymond</name>
25  </respStmt>
26  <respStmt>
27      <resp>Review and validation of the annotation</resp>
```

References

Bermúdez Sabel, Helena; Dell’Oro, Francesca (2020). “[Automatic annotation of Latin and Greek texts](#)”. GitHub.

Dell’Oro, Francesca. 2022. WoPoss guidelines for annotation. Revised version. Swiss National Science Foundation. DOI: <https://doi.org/10.5281/zenodo.6417878>
INCEpTION: <https://inception-project.github.io/>

Klie, J.C., Bugert, M., Boullosa, B., Eckart de Castilho, R. and Gurevych, I. The INCEpTION platform: machine-assisted and knowledge-oriented interactive annotation. Proceedings of the 27th international conference on computational linguistics: system demonstrations , Association for Computational Linguistics, 2018:5–9. <https://www.aclweb.org/anthology/C18-2002/>

Universal Dependencies: <https://universaldependencies.org/>

UD Latin Perseus: https://universaldependencies.org/treebanks/la_perseus/index.html

Acknowledgements

- Swiss National Science Foundation: project n° 176778
(<https://data.snf.ch/grants/grant/176778>)
- Center for Hellenic Studies (<https://chs.harvard.edu/>)

additional material

UD tagset for the PoS

- ADJ: adjective
- ADP: adposition (utilisé pour les prépositions)
- ADV: adverb
- CCONJ: coordinating conjunction
- INTJ: interjection (par exemple, « heu »)
- NOUN: noun
- NUM: numeral
- PRON: pronoun
- PUNCT: punctuation
- SCONJ: subordinating conjunction
- VERB: verb

UD tagset for the morpho-syntactic description

- Case : NOM = nominative ; GEN = genitive ; DAT = dative ; ACC = accusative ; ABL = ablative ; VOC = vocative ; LOC = locative
- Gender : FEM = feminine ; MASC = masculine ; NEUT = neuter
- Number : Sing = singular ; Plur = plural
- Degree : Cmp = comparative ; Sup = superlative
- Aspect : Perf = perfective ; Imp = imperfective
- Mood : Imp = imperative ; Ind = indicative ; Sub = subjunctive
- Person : 1 = first person ; 2 = second person ; 3 = third person
- VerbForm : FIN = finite forms; PART = participle ; GDV = gerundive ; GER = gerund ; INF = infinitive
- Voice : ACT = active ; PASS = passive

Thank you!
Merci beaucoup!