



UNIVERSITÀ  
CATTOLICA  
del Sacro Cuore



# UDante: First Steps Towards the Universal Dependencies Treebank of Dante's Latin Works

Flavio M. Cecchini, Rachele Sprugnoli, Giovanni Moretti, Marco Passarotti

{flavio.cecchini,rachele.sprugnoli,giovanni.moretti,marco.passarotti}@unicatt.it

*CLiC-it 2020*  
March 1, 2021



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant Agreement No. 769994.

## UDante Treebank:

- ▶ All five works in Latin by Dante:
  - ▶ *Monarchia*
  - ▶ *De vulgari eloquentia*
  - ▶ Eclogues (+ those by Giovanni del Virgilio)
  - ▶ Letters (I–XIII)
  - ▶ *Questio de aqua et terra*
- ▶ 55,666 tokens

To be published with the next **UD release** (v2.8)!

Gold Standard:

**Phase 1–3:** three sets of 15, 8, 10 sentences each as common training

**Phase 4:** first 10 sentences from each work (except *Questio*) as control

<https://github.com/CIRCSE/UDante>

Conversion of the *tei-xml* files of *DanteSearch* into CoNLL-U format:

1. script
2. ad-hoc rules
3. manual revision

```
<LM lemma="resono" catg="va1cis3">resonaret</LM>
```

v → VERB

a → Voice=Act

1 → InflClass=LatA

ci → Aspect=Imp|Mood=Sub|Tense=Past|VerbForm=Fin

s → Number=Sing

3 → Person=3

Two parts and four phases performed using *ConlluEditor*<sup>1</sup>

- ▶ Part 1: **training**, parallel annotation to calculate IAA, increasing complexity
  - Phase 1: basics of the UD formalism → 15 sentences
  - Phase 2: intermediate level → 5 sentences
  - Phase 3: complex syntactic structures → 10 sentences

	Phase 1	Phase 2	Phase 3
EDGES	80%	83%	79%
DEPRELs	84%	92%	91%

- ▶ Part 2: **control**
  - Phase 4: individual annotation of a specific work → 10 sentences

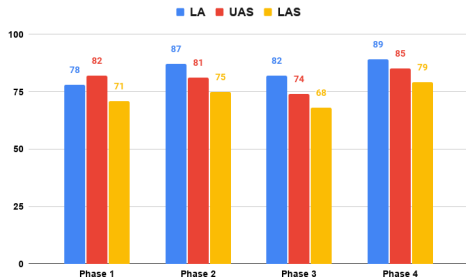
---

<sup>1</sup><https://github.com/Orange-OpenSource/conlleditor>

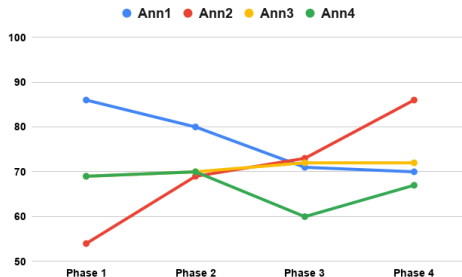
# Results

Comparison with the Gold Standard

(a) Attachment scores



(b) F-Measure



## LiLa: Linking Latin

Università Cattolica del Sacro Cuore  
CIRCSE Research Centre



[info@lila-erc.eu](mailto:info@lila-erc.eu)



<https://github.com/CIRCSE>



<https://lila-erc.eu>



@ERC\_LiLa



Largo Gemelli 1, 20123 Milan, Italy



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme - Grant Agreement No. 769994.