

Linguistic Linked Open Data for Humanists

Anas Fahad Khan, Giulia Pedonese, Michele Mallia



Lisbon Summer School in Linguistics, July 1-5, 2024



Sources

- Giulia Pedonese. (2022, giugno 10). The LiLa Knowledge Base of Interoperable Linguistic Resources for Latin. Architecture and Current State. Zenodo. <https://doi.org/10.5281/zenodo.6637965>
- Eleonora Litta, & Giulia Pedonese. (2023, giugno 13). Sense and Sensibility: compiling the revised Latin WordNet and the new Latin Vallex. Zenodo. <https://doi.org/10.5281/zenodo.8033735>
- Giulia Pedonese, & Mirko Tavoni. (2023, giugno 14). Dante in LiLa: UDante and Dante Latin Loanwords. Zenodo. <https://doi.org/10.5281/zenodo.8036751>

This presentation uses pictures from <https://www.flaticon.com/>

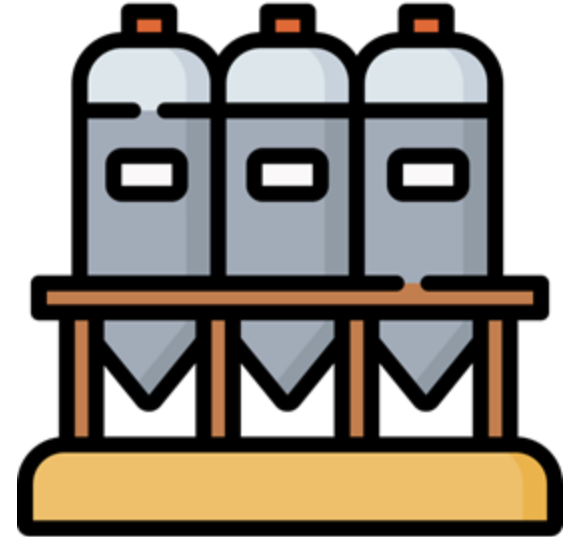
Index

- Current state of language resources for Latin
- What is LiLa
- The technical structure of the Knowledge Base
- Linguistic Resources in LiLa
- Building interoperability: three use cases
 - The annotation and linking of a treebank: UDante
 - Linking a lexical resource to LiLa: Dante Latin Loanwords
 - How it was made: modelling the LatinWordNet and LatinVallex

Data silos

Linguistic (meta)data are often scattered in different collections. This prevents the users from fully benefiting their potentials in interoperable ways.

One of the main issues is that linguistic resources may use different conceptual models and (meta)data can be represented via annotation schemas often incompatible with other systems



The LiLa project

«the *LiLa*: Linking Latin project (2018-2023) was awarded funding from the European Research Council (*ERC*) to build a Knowledge Base of linguistic resources for Latin based on the Linked Data paradigm, i.e. **a collection of multifarious, interlinked data sets described with the same vocabulary of knowledge description** (by using common data categories and ontologies). The project's ultimate goal is to **make full use of the linguistic resources and NLP tools for Latin** developed thus far, in order to bridge the gap between raw language data, NLP and knowledge description.»

Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin / Passarotti, Marco; Mambrini, Francesco; Franzini, Greta; Cecchini Flavio, Massimiliano; Litta, Eleonora; Moretti, Giovanni; Ruffolo, Paolo; Sprugnoli, Rachele. - In: STUDI E SAGGI LINGUISTICI. - ISSN 0085-6827. - 58:1(2020), pp. 177-212.
<https://doi.org/10.4454/ssl.v58i1.277>, p. 180

The LiLa-Linking Latin project



- ERC Grant 2018-2023
- CIRCSE, Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione, Interdisciplinary Research Centre, Università Cattolica del Sacro Cuore, Milan, Italy
- Research Team covering a range of expertise: Computational Linguistics, Latin Linguistics, NLP, Classica Philology and Humanities Computing



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

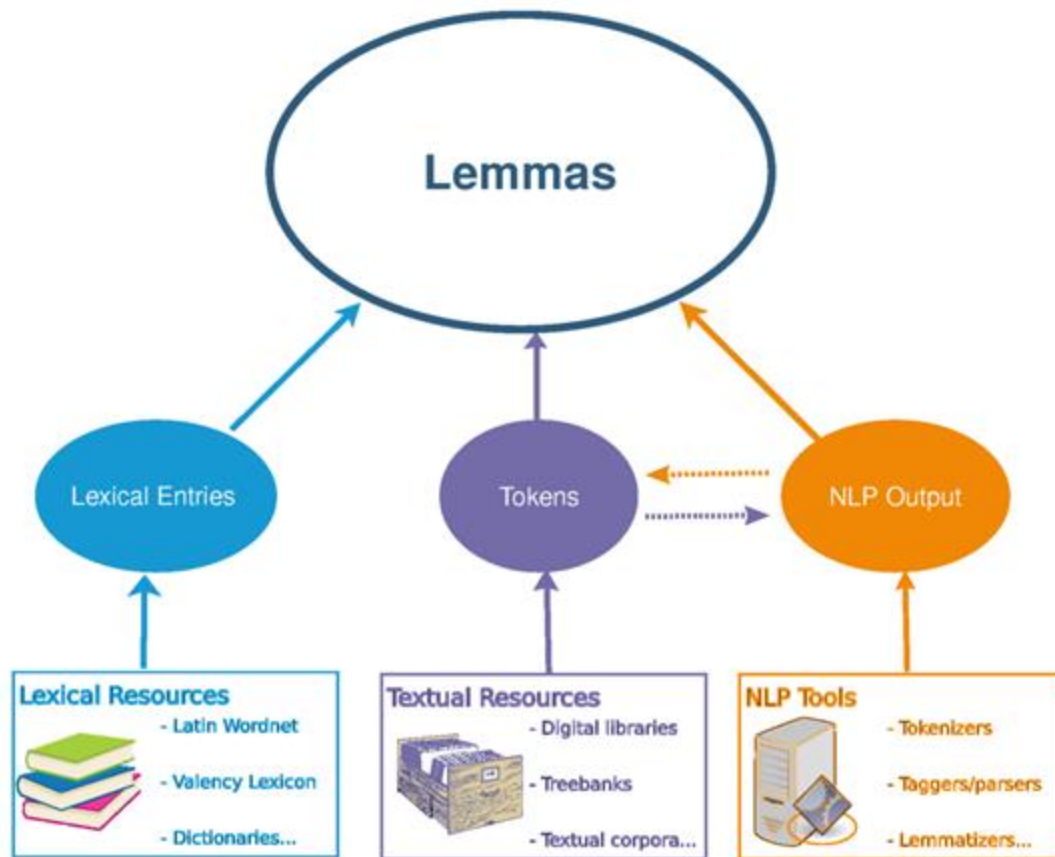


The architecture

The **Lemma Bank** as a centralised hub to connect different resources

Lemmatisation as the common layer of annotation among different Linguistic Resources:

- Lexical Resources index lexical entries using lemmas
- Lemmas facilitate search in corpora (inflectional morphology)
- Lemmas are the result of basic NLP annotation (text pre-processing)



The LiLa ontology

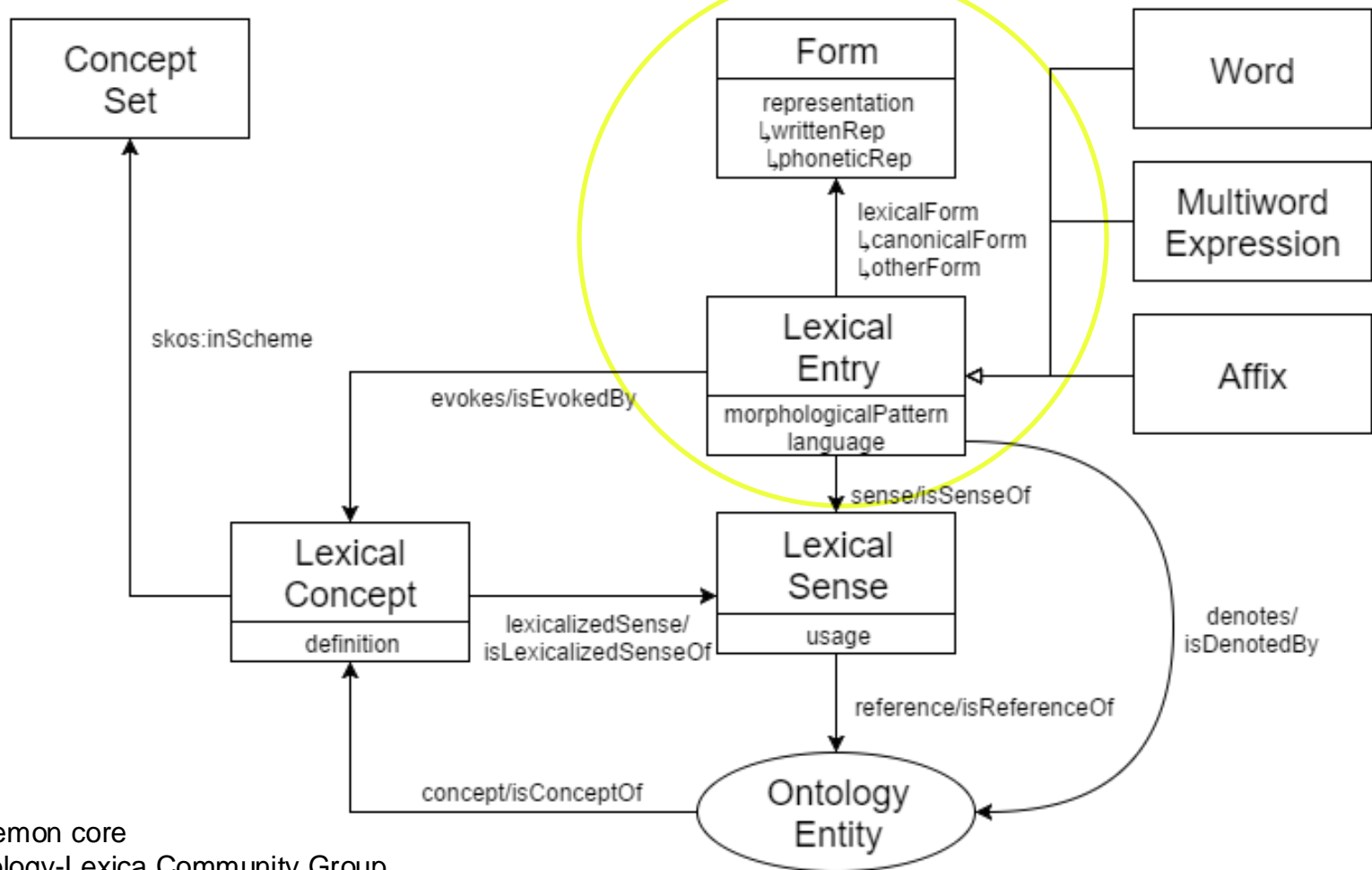
LiLa is based on a **customisation of the Ontolex Lemon core ontology**

A lemma in LiLa is a form chosen as **the citation form of a lexical item**, that is the canonical form representing all the inflected forms.

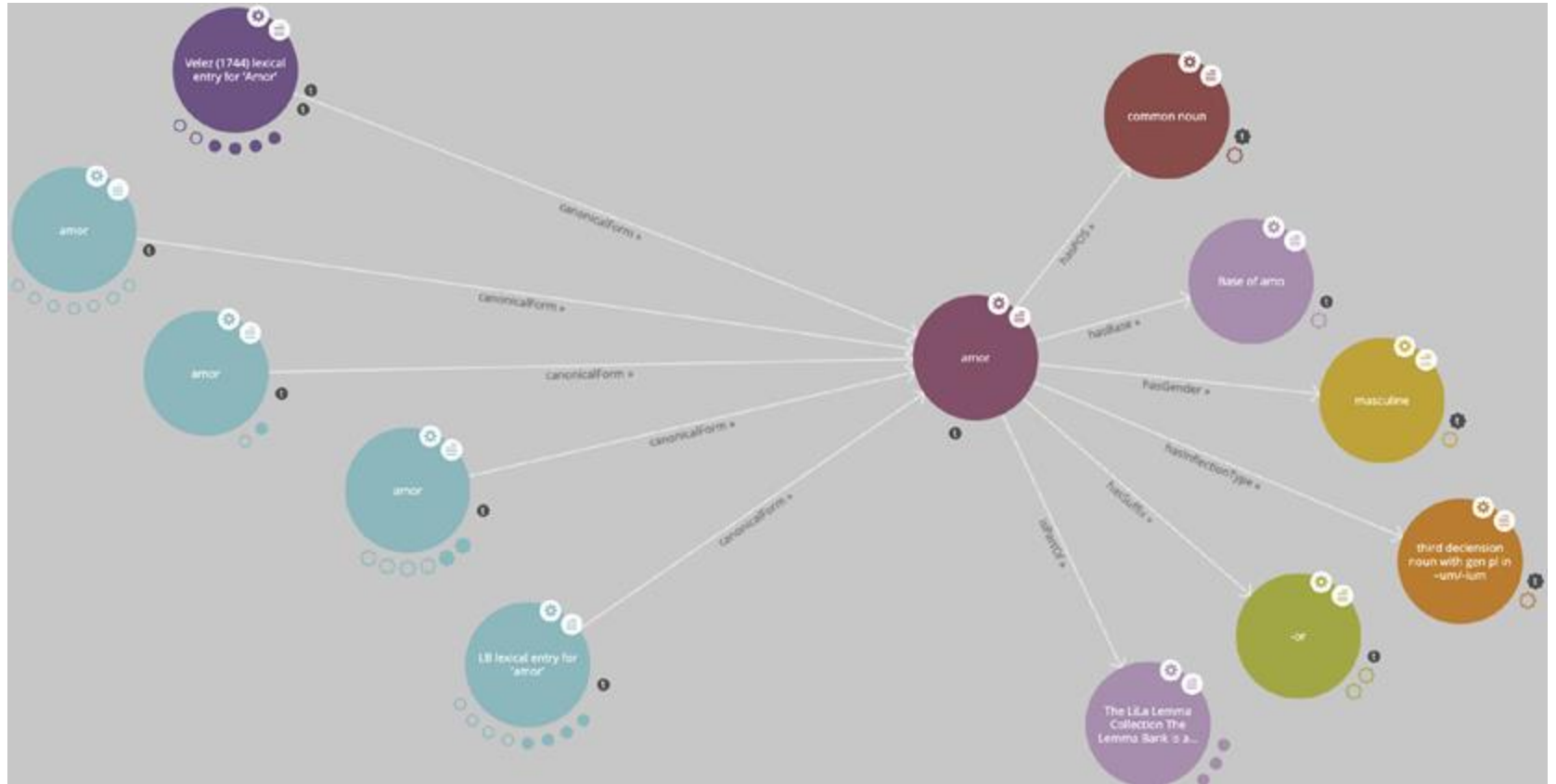
The core of the Knowledge Base is **a large collection of Latin lemmas**.

Interoperability is achieved by linking the lexical entries in lexical resources and the tokens in corpora that point to the same lemma.

Conceptual and linguistic challenges of lemmatisation: different corpora, lexica and tools can adopt different strategies



Graphic visualisation of lemmas in the Lemma Bank



Different annotation schemas

Different lemmatisation criteria in pre-existing resources: a word form can be reduced to multiple lemmas

- Participles: VERB or NOUN? e.g. factus < facio
- Deadjectival adverbs: ADJ or ADV? e.g. graviter < gravis
- Polythematic words, e.g. magnus / maior / maximus: comparative form of the base ADJ or different lemma?

Harmonising lemmas

LiLa **does not impose a lemmatisation standard** but rather provides a descriptive set of concepts and properties capable of integrating all solutions adopted by different Latin resources (Passarotti et al. 2020, p. 188).

- Ontolex «written representation» property to accommodate different spelling of canonical forms (e.g. e-oe)
- Two lemmas in the case of Latin words belonging to more than one PoS (e.g. prepositions that are also adverbs, e.g. post, circum, sub)
- Extension of the Ontolex ontology for participles, deponent verbs: hypolemmas, e.g. loquo / loquor, discens < disceo

LEMMA BANK QUERY INTERFACE

QUERY MODULES AVAILABLE













QUERY MODULES SELECTED

Search lemma 

 Regex

Results: 3



Lemma	PoS		
amans	proper noun		
amans	adjective	  	 
amans	common noun	 	

amans

<http://lila-erc.eu/data/id/hypolemma/2061>

rdfs:label	amans
ontolex:writtenRep	amans
rdf:type	lila:Hypolemma ↳ Hypolemma
lila:hasPOS	lila:adjective ↳ adjective
dcterms:isPartOf	< http://lila-erc.eu/data/id/lemma/LemmaBank > ↳ LiLa Lemma Collection
lila:hasDegree	lila:positive ↳ positive
lila:hasInflectionType	lila:n71 ↳ second class adjective (e.g. nom sing in: -s; -x; -r; -l)
lila:isHypolemma	< http://lila-erc.eu/data/id/lemma/88705 > ↳ amo



Linguistic Resources in Lila

<https://lila-erc.eu/data-page/>

LEMMA BANK

- LiLa Lemma Bank: [view](#), [download](#), [cite](#).

LEXICAL RESOURCES

- Word Formation Latin (WFL): [view](#), [download](#), [cite](#).
- Etymological Dictionary of Latin & the Other Italic Languages: [view](#), [download](#), [cite](#).
- Latin Vallex 2.0: [view](#), [download](#), [cite](#).
- Latin WordNet: [view](#), [download](#), [cite](#).
- Index Graecorum Vocabulorum in Linguam Latinam Translatorum: [view](#), [download](#), [cite](#).
- LatinAffectus: [view](#), [download](#), [cite](#).
- Lewis & Short Latin-English dictionary: [view](#), [download](#), [cite](#).
- Glossary of Latin loanwords from the Italian works of Dante Alighieri: [view](#), [download](#).
- Velez's Latin-Portuguese dictionary: [view](#), [download](#).
- Principal Parts Latin: [view](#), [download](#).
- Lexicon der indogermanischen Verben: [view](#), [download](#), [cite](#).
- Dictionary of Medieval Latin in the Czech Lands: [view](#), [download](#), [cite](#).

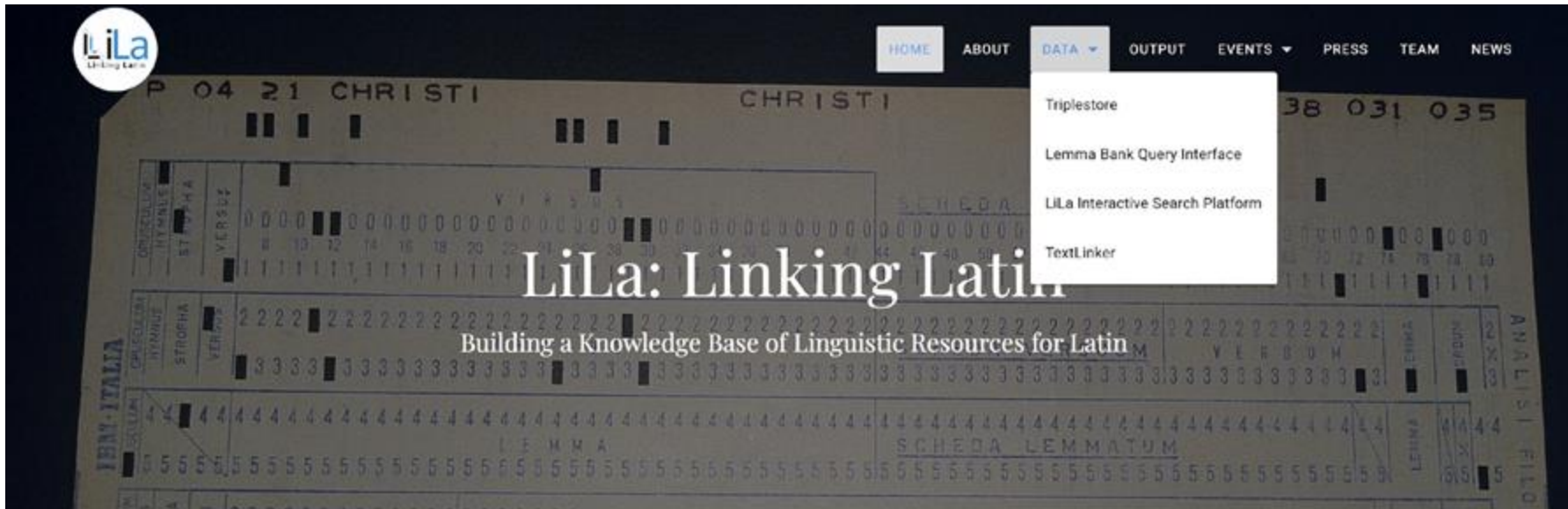
TEXTUAL RESOURCES

- Index Thomisticus Treebank (ITTb): [view](#), [download](#), [cite](#).
- UDante: [view](#), [download](#), [cite](#).
- Querolus sive Aulularia: [view](#), [download](#).
- LASLA: [view](#), [download](#), [cite](#).
- Augustini Confessiones: [view](#), [download](#).
- Computational Historical Semantics Corpus: [view](#), [download](#).
- Fibonacci's Liber Abaci (chapter VIII): [view](#), [download](#), [cite](#).
- Corpus for Latin Sociolinguistic Studies on Epigraphic textS (CLaSSES): [view](#), [download](#), [cite](#).
- Lucani Pharsalia: [view](#), [download](#).
- Augustini De Ciuitate Dei: [view](#), [download](#).
- Augustini De Trinitate: [view](#), [download](#).

How to query the LiLa Knowledge Base

We have already seen a screenshot of the Lemma Bank Query Interface, but we have two other options:

- Triplestore: SPARQL endpoint
- Lemma Bank Query Interface: a user-friendly application
- LiLa Interactive Search Platform: a user-friendly interface to query the Knowledge Base



<https://lila-erc.eu/#page-top>

Practice

Exercise 1

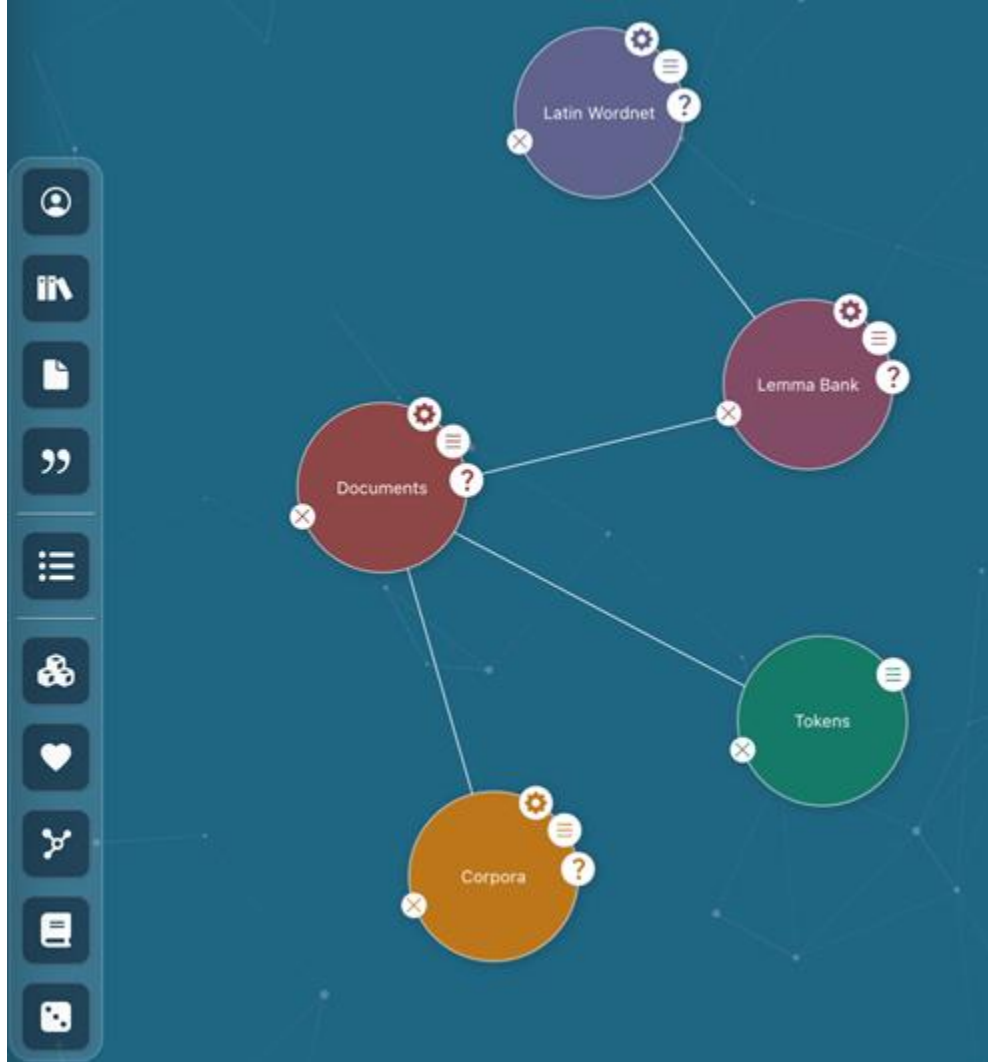
- Go to the Lemma Bank Query Interface
- Search for the lemma «populus»
- How many results do you have? What is the difference between them?
- How many resources are connected to the lemma «populus» with ID <https://lila-erc.eu/data/id/lemma/118463>?
- What is its meaning?



LiLa Interactive Search Platform

Selecting on the left:

- Lemma Bank: «amo»
- Corpus: Index
Thomisticus Treebank
- Displaying the results in terms of synsets and tokens in each one of Thomas Aquinas' works



Results:

Results

Token	Document
amamus	Summa contra Gentiles
amamus	Summa contra Gentiles
amamus	Summa contra Gentiles
amamus	Summa contra Gentiles
amamus	Summa contra Gentiles
amamus	Summa contra Gentiles

1–100 of 154

amamus



http://lla-erc.eu/data/corpora/ITTB/id/token/005.SCG*LB1.CP-7++5.N.-3.3-3.6-2W5

ENTITÀ DI TIPO: Terminal

Token context:
quae enim propter se **amamus**, volumus esse optima, et semper meliorari et multiplicari ...

References:
Summa contra Gentiles, Liber 1, Caput 75, Numerus, Paragraphus 3

rdfs:label	amamus
powla:hasStringValue	amamus
rdf:type	powla:Terminal
powla:hasLayer	<http://lla-erc.eu/data/corpora/ITTB/id/citationUnit/DocumentLayer> ↳ Document Layer
powla:next	<http://lla-erc.eu/data/corpora/ITTB/id/token/005.SCG*LB1.CP-7++5.N.-3.3-3.6-2W6> ↳ ,
powla:previous	<http://lla-erc.eu/data/corpora/ITTB/id/token/005.SCG*LB1.CP-7++5.N.-3.3-3.6-2W4> ↳ &&
lla:hasLemma	<http://lla-erc.eu/data/id/lemma/88705> ↳ amo

PREFIXES: [RDF](#) [RDFS](#) [DC](#) [DCTERMS](#) [LILA](#) [LILACORPORA](#) [ONTOLEX](#) [LIME](#) [LEMONETY](#) [LEXINFO](#) [POWLA](#) [MARL](#)
Query  [EXAMPLES](#)

```

1 PREFIX lime: <http://www.w3.org/ns/lemon/lime#>
2 PREFIX lila: <http://lila-erc.eu/ontologies/lila/>
3 PREFIX lemonEty: <http://lari-datasets.ilc.cnr.it/lemonEty#>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX ontollex: <http://www.w3.org/ns/lemon/ontollex#>
6 PREFIX powla: <http://purl.org/powla/powla.owl#>
7 PREFIX lexinfo3: <https://lexinfo.net/ontology/3.0/lexinfo.owl#>
8
9 #Counts in the Latin works of Dante Alighieri the number of occurrences of those Latin words that are etyma of Latin loanwords in the Italian works of Dante Alighieri (provided by the "Dante Latin Loanwords" lexical resource), including the part of speech (POS) of each etymon.
10
11 SELECT ?etylabel ?pos (count(?etylabel) as ?count)
12 WHERE {
13   ?etymon a lemonEty:Etymon ;
14           ontollex:canonicalForm ?lemma ;
15           rdfs:label ?etylabel ;
16   ?pos a lexinfo3:partOfSpeech ?etylabel
17   ?etymology lemonEty:etymon ?etymon .
18   ?lexicalEntry lemonEty:etymology ?etymology .
19   <http://lila-erc.eu/data/lexicalResources/DanteLatinLoanwords/Lexicon> lime:entry ?lexicalEntry .
20   ?token lila:hasLemma ?lemma ;
21           powla:hasLayer ?tokenlayer .
22   ?tokenlayer powla:hasDocument ?tokendocument .
23   <http://lila-erc.eu/data/corpora/UDante/id/corpus> powla:hasSubDocument ?tokendocument .
24 }
25 GROUP BY ?etylabel ?pos
26 ORDER BY DESC(?count)
27

```

 Table
  Response

How this query works

- Based on two resources in LiLa: Udante and Dante Latin Loanwords (DLL)
- **It retrieves all the Latin lemmas attested in Dante's Latin works** (Udante) that are also **etyma of Latin borrowings** in his Italian works (DLL)

Dante Alighieri, Italian writer and poet (1265-1321)



How this query works

- Prefixes define the namespaces used to shorten the names of classes and properties
- SELECT retrieves the label of the Latin etymon and counts its occurrences
- GROUP BY gathers the results by etylabel
- ORDER BY gives the list in decrescent order based on the number of occurrences

```
1 PREFIX lime: <http://www.w3.org/ns/lemon/lime#>
2 PREFIX lila: <http://lila-erc.eu/ontologies/lila/>
3 PREFIX lemonEty: <http://lari-datasets.ilc.cnr.it/lemonEty#>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX ontollex: <http://www.w3.org/ns/lemon/ontollex#>
6 PREFIX powla: <http://purl.org/powla/powla.owl#>
7 PREFIX lexinfo3: <https://lexinfo.net/ontology/3.0/lexinfo.owl#>
8
```

How this query works

WHERE specifies the conditions for selecting data: etyma that have a canonical form (lemma) and a label (etymlabel)

- The relation between the etymology and the etymon and the lexical entry and the etymology is modelled through the LemonEty Ontology
- The token is defined as having a lemma in the LiLa Lemma Bank and having a specific annotation layer (tokenlayer)
- The annotation layer is linked to a specific document to consider only the tokens in the specified corpus (Udante)

Case studies



Examples of LOD resources:

1. Creating and linking a treebank: UDante
2. Modelling etymologies: Dante Latin Loanwords
3. Refining and extending the Latin WordNet and its connection to theLatin Vallex

1. Linking a Lexical Resource to LiLa: UDante

- The source: DanteSearch
- Creating the corpus
- Linguistic annotation of the resource
- Linking to the LiLa Knowledge Base

Reference: *Le opere latine di Dante tra annotazione linguistica e web semantico* / Passarotti, Marco; Sprugnoli, Rachele; Pedonese, Giulia in "Linguistica e letteratura : XLVI, 1/2, 2021, Pisa : Fabrizio Serra, 2021 , 1724-0522 - Casalini id: 5232312" - 45-71 p. - Permalink: <http://digital.casalini.it/10.19272/202101602003> - Casalini id: 5236140

Udante: what is it?

A syntactic annotated corpus of Dante Alighieri's Latin Works based on the Universal Dependencies (UD) formalism

- *Monarchia, De Vulgari Eloquentia, Eclogues, Letters, Questio de aqua et terra*
- Texts from DanteSearch
- Tokenisation, sentence splitting, lemmatisation, PoS tagging, morphological analysis, syntactic annotation
- 55,666 tokens

Already encoded texts

Ordina per: [= [Testo](#) | [Numeri occorrenze](#) | [Ordine cronologico](#)]

Trovate 782 sezioni in 18 opere.

☰ [Commedia - Inferno: codifica sintattica](#), Dante Alighieri

1. [Canto I](#)
2. [Canto II](#)
3. [Canto III](#)
4. [Canto IV](#)
5. [Canto V](#)
6. [Canto VI](#)
7. [Canto VII](#)
8. [Canto VIII](#)
9. [Canto IX](#)
10. [Canto X](#)
11. [Canto XI](#)
12. [Canto XII](#)
13. [Canto XIII](#)
14. [Canto XIV](#)
15. [Canto XV](#)
16. [Canto XVI](#)
17. [Canto XVII](#)
18. [Canto XVIII](#)
19. [Canto XIX](#)
20. [Canto XX](#)

Commedia - Inferno: codifica sintattica

Dante Alighieri

Canto I

1. (Nel mezzo del cammin di nostra vita
2. mi ritrovai per una selva oscura,
3. ch  la diritta via era smarrita.)
4. (Ahi quanto a dir qual era   cosa dura
5. esta selva selvaggia e aspra e forte
6. che nel pensier rinova la paura!)
7. (Tant'  amara che poco   pi  morte;
8. ma per trattar del ben ch'  vi trovai,
9. dir  de l'altre cose ch'  v'ho scorte.)
10. (Io non so ben ridir com'  v'intrai,
11. tant'era pien di sonno a quel punto
12. che la verace via abbandonai.)
13. (Ma poi ch'  fui al pi  d'un colle giunto,
14. l  dove terminava quella valle
15. che m'avea di paura il cor compunto,
16. guardai in alto e vidi le sue spalle

Commedia - Inferno: edizione elettronica lemmatizzata

Dante Alighieri

Canto I

1. Nel mezzo del cammin di nostra vita
2. mi ritrovai per una selva oscura
3. ch  la diritta via era smarrita.
4. Ahi quanto a dir qual era   cosa dura
5. esta selva selvaggia e aspra e forte
6. che nel  pensier rinova la paura!
7. Tant'   amara che poco   pi  morte;
8. ma per trattar del   ben ch' i' vi trovai,
9. dir  de l' altre cose ch' i' v' ho scorte.
10. Io non so ben ridir com' i' v' intrai,
11. tant' era pien di sonno a quel punto
12. che la verace via abbandonai.
13. Ma poi ch' i' fui al pi  d' un colle giunto,
14. l  dove terminava quella valle
15. che m' avea di paura il cor compunto,
16. guardai alto, e vidi le sue spalle

Universal Dependencies: a De Facto Standard

Universal Dependencies is a project developing cross-linguistically consistent treebank annotation across different human languages. **Its goal is to facilitate multilingual parser development, cross-lingual learning, and parsing research.** The annotation scheme is based on an evolution of Stanford Dependencies (de Marneffe et al., 2006, 2008, 2014), Google universal part-of-speech tags (Petrov et al., 2012), and the Intersect interlingua for morphosyntactic tagsets (Zeman, 2008).

UD is an **open community effort with over 600 contributors producing over 200 treebanks in over 150 languages.** It aims to provide a universal inventory of categories and guidelines to facilitate consistent annotation of similar constructions across languages while allowing language-specific extensions when necessary.



<https://universaldependencies.org/>

Format conversion

- TEI-XML files > CoNLL-U files
- Conversion of the annotation codes from DanteSearch to UD. This process was not one-to-one:

e.g. "a" for "adjective» > ADJ, NUM, DET

- Mapping of morphological features: UD provides a more granular set of labels e.g. "pp" for past participle in DanteSearch corresponds to the following UD features:

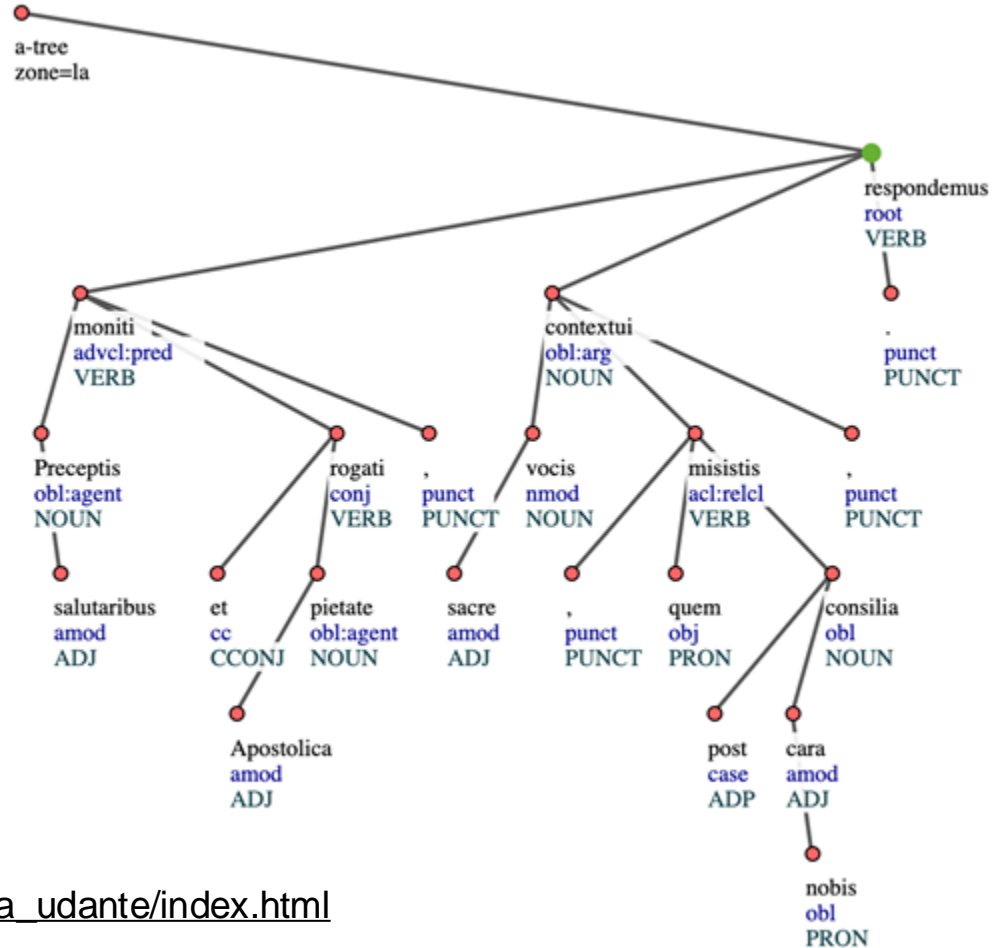
Aspect=Imp|Degree=Pos|Tense=Pres|VerbForm=Part|Voice=Act

UD Dependency trees

According to the UD framework, the syntactic structure of sentences in UDante has a dependency annotation style

- each sentence is represented by **acyclic graphs called 'trees'** whose nodes are oriented from a root
- The **predicate is at the centre and its complements depend on it** without distinction between arguments and adjuncts
- The nodes are made of **words or punctuation marks**
- difference between **semantically full words and functional words** that tend to occupy 'leaf' nodes i.e., that are not heads of any syntactic relation
- To each dependency relation is assigned a label that describes its syntactic function (dependency relation: deprel)

Example of a syntactic tree: graphic representation



https://universaldependencies.org/treebanks/la_udante/index.html

Example of a syntactic tree: the CoNLL-U format

```
1 # sent_id = 4
2 # text = Hanc quidem secundariam Greci habent et alii , sed non omnes : ad habitum vero huius pauci perveniunt , quia non nisi per spatium temporis et studii assiduitatem regu
3 1 Hanc hic DET ddisfa Case=Acc|Gender=Fem|NounClass=LatPron|Number=Sing|PronType=Dem 3 det _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
4 2 quidem quidem ADV r 5 discourse _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
5 3 secundariam secundarius ADJ afs1a Case=Acc|Degree=Pos|Gender=Fem|NounClass=IndEurA|Number=Sing 5 obj _ CitationHierarchy=Liber_Primum,
6 4 Greci graecus ADJ Snp2n Case=Nom|Gender=Masc|NounClass=IndEur0|Number=Plur 5 nsubj _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
7 5 habent habeo VERB va2ipp3 Aspect=Imp|Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbClass=LatE|VerbForm=Fin|Voice=Act 0 root _ CitationHierarc
8 6 et et CCONJ co 7 cc _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
9 7 alii alius DET dmpn Case=Nom|Gender=Masc|NounClass=LatPron|Number=Plur|PronType=Ind 4 conj _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
10 8 , , PUNCT Pu 11 punct _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
11 9 sed sed CCONJ co 11 cc _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
12 10 non non PART r Polarity=Neg 11 advmod:neg _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
13 11 omnes omnis DET amp2n Case=Nom|Gender=Masc|NounClass=IndEurI|Number=Plur|PronType=Tot 5 conj _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
14 12 : : PUNCT Pu 18 punct _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
15 13 ad ad ADP e AdpType=Prep 14 case _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
16 14 habitum habitus NOUN sms4a Case=Acc|Gender=Masc|NounClass=IndEurU|Number=Sing 18 obl:arg _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
17 15 vero vero CCONJ r 18 cc _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
18 16 huius hic DET ddisfg Case=Gen|Gender=Fem|NounClass=LatPron|Number=Sing|PronType=Dem 14 nmod _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
19 17 pauci paucus DET amp1n Case=Nom|Gender=Masc|NounClass=IndEur0|Number=Plur|PronType=Ind 18 nsubj _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
20 18 perveniunt pervenio VERB va4ipp3 Aspect=Imp|Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbClass=LatI|VerbForm=Fin|Voice=Act 5 conj _
21 19 , , PUNCT Pu 29 punct _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
22 20 quia quia SCONJ cs 29 mark _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
23 21 non non PART r Polarity=Neg 29 advmod:neg _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
24 22 nisi nisi SCONJ cs 24 mark _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
25 23 per per ADP e AdpType=Prep 24 case _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
26 24 spatium spatium NOUN sns2a Case=Acc|Gender=Neut|NounClass=IndEur0|Number=Sing 29 advcl _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
27 25 temporis tempus NOUN sns3g Case=Gen|Gender=Neut|NounClass=IndEurX|Number=Sing 24 nmod _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
28 26 et et CCONJ co 28 cc _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
29 27 studii studium NOUN sns2g Case=Gen|Gender=Neut|NounClass=IndEur0|Number=Sing 28 nmod _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
30 28 assiduitatem assiduitas NOUN sfs3a Case=Acc|Gender=Fem|NounClass=IndEurX|Number=Sing 24 conj _ CitationHierarchy=Liber_Primum,1,Paragr
31 29 regulamur regulo VERB vpl1pp1 Aspect=Imp|Mood=Ind|Number=Plur|Person=1|Tense=Pres|VerbClass=LatA|VerbForm=Fin|Voice=Pass 18 advcl _ Citatio
32 30 et et CCONJ co 31 cc _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
33 31 doctrinamur doctrino VERB vpl1pp1 Aspect=Imp|Mood=Ind|Number=Plur|Person=1|Tense=Pres|VerbClass=LatA|VerbForm=Fin|Voice=Pass 29 conj _
34 32 in in ADP e AdpType=Prep 33 case _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
35 33 illa ille DET ddisfb Case=Abl|Gender=Fem|NounClass=LatPron|Number=Sing|PronType=Dem 29 obl _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
36 34 , , PUNCT Pu 5 punct _ CitationHierarchy=Liber_Primum,1,Paragraphus_3
```

<https://github.com/CIRCSE/UDante?tab=readme-ov-file>

Results after applying LOD standards

- A treebank aligned with standard annotation criteria for syntactic corpora
- Same format as hundreds of other Universal Dependencies treebanks
- Interoperable with other Linguistic Resources such as Lexica, Corpora and Dictionaries through the LiLa Knowledge Base
- Part of a Semantic Web ecosystem where the resource is best valued



2. An Italian Lexical resource

The glossary of Latin Loanwords from the Italian works of Dante Alighieri

- It includes 765 Latin loanwords attested in the four Italian works by Dante Alighieri (Rime, Vita Nova, Convivio and Commedia)
- From the same reference editions of DanteSearch
- It adds information on the historical linguistic framing of the loan, e.g. the loan type and its relative position of first attestation or hapax in XIIIth Century Italian

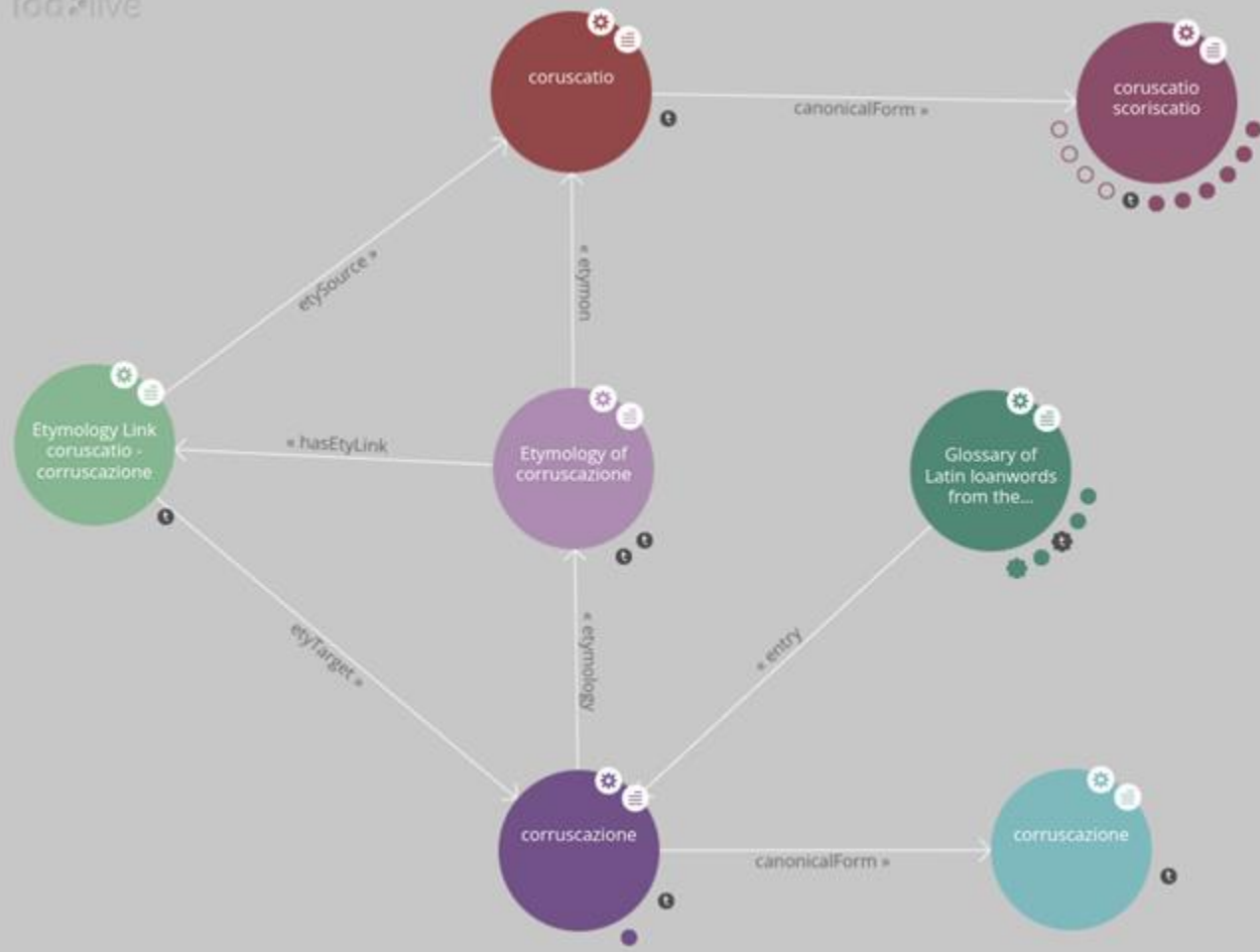
The linking process

Each one of the 765 Italian lemmas of the resource was linked to the corresponding Latin lemma in

the Lemma Bank:

- 740 lemmas were automatically matched with a 1:1 ratio
- 20 lemmas had a 1:0 match in the Lemma Bank: 15 new lemmas (e.g. *contrapassum*, *inconsummabilis*), 2 new written representations (e.g. *armonizo*, *harmonizo*) and 3 new lemma variants were added (e.g. *divitia divitiae*)
- 5 lemmas had a one-to-many match: 4 were cases of lemma variants and 1 lemma was left ambiguous as the etymology is still discussed (*adultus*, *adolesco* = to grow, *adoleo* and *adolesco* = to burn)

Repository: <https://github.com/CIRCSE/DanteLatinLoanWords>



Lemon ETY

Sources:

- Francesco Mambrini and Marco Passarotti. 2020. [Representing Etymology in the LiLa Knowledge Base of Linguistic Resources for Latin](#). In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, pages 20–28, Marseille, France. European Language Resources Association.
- Khan, Anas & Frontini, Francesca. (2024). Toward a Representation of Semantic Change in Linked Data. *Languages*. 9. 215. [10.3390/languages9060215](https://doi.org/10.3390/languages9060215).

3. Refining and Expanding the revised Latin WordNet

- **Latin WordNet:** Latin Lexical database grouping words by sets of cognitive synonyms (Synsets) which are linked by conceptual-semantic and lexical relations: hyponymy, hyperonymy, meronymy ecc.
- **Latin Vallex:** a valency lexicon of Latin assigning a valency frame to each sense of a word (tagset. Czech VALLEX). In the LiLa KB it shares the same lexical entry as the Latin WordNet

Different versions

- Minozzi's Latin WordNet (2004): 9,3778 lemmas and 8,973 synsets <https://edizionicafoscari.unive.it/it/edizioni4/libri/978-88-6969-183-6/latin-wordnet-una-rete-di-conoscenza-semantica-per/>
- Michael Short: (Exeter): Automatically extended version: more than 70,000 words <https://latinwordnet.exeter.ac.uk/>
- LiLa version: manually annotated set of around 9.000 words

LiLa's Latin WorNet: the refining process

Manual removal of noise, namely:

- **Senses not applicable to Latin**, e.g. *pictura*, n: a representation of a person or scene in the form of a print or transparent slide; recorded by a camera on light-sensitive material n#03925226
- **Concrete vs. abstract meaning**, e.g. *ager*, n a branch of knowledge n#05996646 a piece of land cleared of trees and usually enclosed n#08569998

Example of a lexical entry: responsum

responsum, n.

- **Lewis & Short s.v. respondeo:** 1 an answer, reply, response. 2 an answer, reply of a lawyer, priest, oracle, etc.; an opinion, response, oracle
- **Oxford Latin Dictionary s.v. responsum:** 1. An answer, reply (to a question, request, etc.); b. given by a ruler to envoys, petitioners, etc.; c. a reply to an argument; 2. (spec.) An answer given by an oracle, soothsayer or sim. b. an answer given by a jureconsult to a legal query

How to assign synsets

responsum, n. (existing, **eliminated**, **added** synsets)

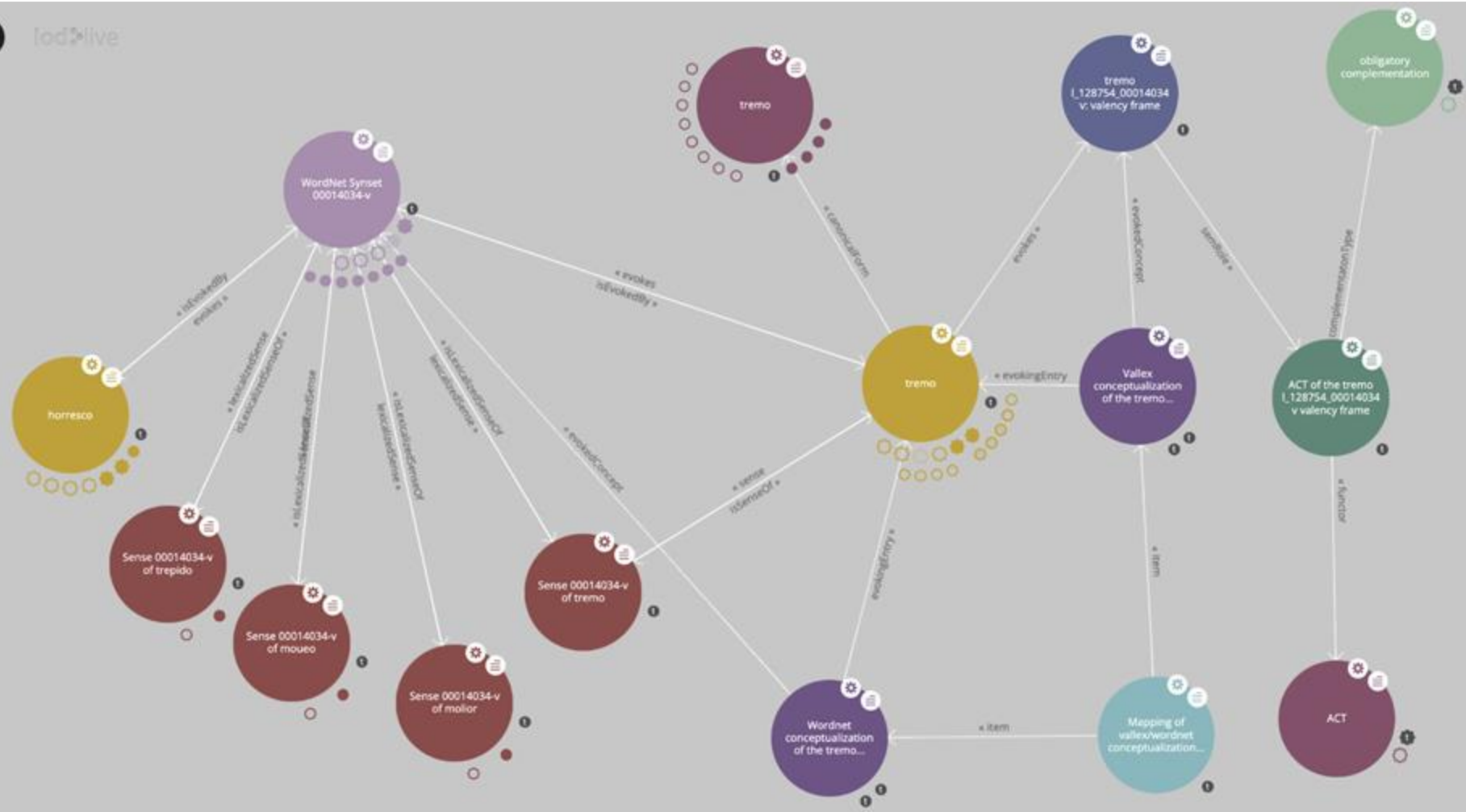
- n#06746005 a statement (either spoken or written) that is made to reply to a question or request or criticism or accusation
- n#07200527 the speech act of replying to a question
- **n#00859001 a bodily process occurring due to the effect of some antecedent stimulus or agent**
- **n#07199565 the speech act of continuing a conversational exchange**
- n#06750339 a prophecy (usually obscure or allegorical) revealed by a priest or priestess; believed to be infallible
- **n06782019 a message expressing a belief about something; the expression of a belief that is held with confidence but not substantiated by positive knowledge or proof**

Link to the Latin Vallex

Senses of words are represented as PWN synsets evoked by LWN lexical entries, each valency bearing synset is linked to a valency frame

e.g. moveo:

- 01762528-v cause to be agitated, excited, or roused – ACT, PAT
- 01419473-v change the arrangement or position of – ACT, PAT
- 01850315-v cause to move or shift into a new position or place, both in a concrete and in an abstract sense – ACT, PAT, DIR3
- 00095377-v put forth and grow sprouts or shoots – ACT



LiLa Linguistic Resources in CLARIN

The [Catholic University of the Sacred Heart](#) (Milano) joined CLARIN-IT by signing a Scientific Collaboration Agreement with ILC-CNR (Pisa) within CLARIN on 8th March 2021. The research unit involved in CLARIN-IT is the [Interdisciplinary Research Center "CIRCSE"](#)
<https://www.clarin-it.it/en/content/consortium>

LILA collection in the VLO: Search > Collections "CIRCSE"
<https://vlo.clarin.eu/?1&fq=collection:CIRCSE&fqType=collection:or>

