



4th Summer Datathon on Linguistic Linked Open Data
Cercedilla, Spain
30th May – 3 June 2022

LiLa – Linking Latin Introduction

Francesco Mambrini (Università Cattolica del Sacro Cuore, Milan, Italy)





UNIVERSITÀ
CATTOLICA
del Sacro Cuore



LiLa - Linking Latin

An introduction

Francesco Mambrini

4th Summer Datathon on Linguistic Linked Open Data
Cercedilla, Spain
30 May 2022



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.

Overview



Introduction and Fundamentals

LiLa: Mission and Architecture

LiLa now!

Lemma Bank and Lexical Resources

Textual Resources

Services and Tools

To sum up

Table of Contents



Introduction and Fundamentals

LiLa: Mission and Architecture

LiLa now!

Lemma Bank and Lexical Resources

Textual Resources

Services and Tools

To sum up

Studying Latin (and other ancient languages)

how it was done



Studying Latin in the digital era



Searching database for **mulier** ⓘ

Bibliography criteria: **None**

Hits 1 - 25 of 670

[View occurrences with context](#) [View occurrences line by line \(KWC\)](#)

Sort results by: [None](#) ⓘ [None](#) ⓘ [None](#) ⓘ [Sort](#)

1. Ammianus Marcellinus, Res ... s' autem quad vixerint late palantur, ut alibi **mulier** nubat, in loco periat alio, liber.
2. Ammianus Marcellinus, Res ... er his incitatum propositum ad nocendum aliqua **mulier** vilis, quae ad palatum (ut popos).
3. Ammianus Marcellinus, Res ... spiravit oppressa, et in abrupto necessitatibus **mulier** Rufinum totius machinae confitefa.
4. Ammianus Marcellinus, Res ... e, Barbatio epistulam suscepisse confessus, et **mulier** scripsisse documento convicta non.
5. Ammianus Marcellinus, Res ... tuando quereretur, et imperator Prosequere ait **mulier**, siquid te laesam existimis; hic.
6. Ammianus Marcellinus, Res ... ulisse dicebatur Anepsiiae, morte multatus est. **Mulier** vero ut contineendas vitam spem filii.
7. Ammianus Marcellinus, Res ... am. Apud proconsulam Asiae Dolabellam Smyrnaea **mulier** subolem proprie et maritum venit.
8. Ammianus Marcellinus, Res ... t nega filium tuum esse aut vero abscedo hinc. **Mulier** compellitur ab sposo, et coepit.
9. Ammianus Marcellinus, Res ... e iussit in conspectu suo sisti. Cui et dixit: **Mulier**, filius tuus aduersus te rogar;
10. Ammianus Marcellinus, Res ... s' autem quod vixit, postea. Quicquid ex eis. Etiamque et mulierum invenerunt. I' omnes respondebat. Hoc enim dicitur.

[A logon.uchicago.edu: x](https://logon.uchicago.edu/) + <https://logon.uchicago.edu/mulier>

AOTEION

mulier

mulier

Short Definition

mulier, a, woman, female

Frequency

mulier is the 493rd most frequent word

BWL

median mulierum complectitur

he embraced his wife around her middle

FrieseDennisonVergil

mulier, ari, f: a woman, 7,661.

LewisShort

DMLBS

Gaffiot 2016

Lewis

La >

mulier, éris, t [molar, comp. of molis, q. v.], a woman, a female, whether married or not.

I. Lit.

A. In gen. Plaut. Poen. 1, 2, 36: mulieres omnes propter infirmatae consili, magiores in tutorum potestate esse voluerent, Ccc. Mur. 12, 27: of a virgin, etc. Verr. 2, 1, 25, § 64: mulieres omnes discut, quascunque sexus feminis sunt. Dss. 34, 2, 26: nō permittit mulier sibi. Juv. 6, 457.—

B. In part., c. wife, opp. to a maid: equa virgo sit aut mulier digna, etc. Cc. Ver. 2, 1, 24, § 63: Cicero obprobrius, quod sepius exasperans Publilianum virgines duxisset, "rasa mulier" inquit, Quir. 6, 3, 75: si vigo fuit primo, postea mulier, Lact. 1, 17, 8 - alia: pudica mulier, Hor. Epod. 2, 39. Suer. Vit. 3, 75: si vigo fuit

Boethius, *Consolatio Philosophiae*
James J. O'Donnell, Ed.

Home Collections/Tests Perseus Catalog Research Grants Open Source About Help Hide browser bar

Your current position in the text is marked in blue. Click anywhere in the line to jump to another position:
book: Boe. Cons. 1.M1
section: Boe. Cons. 1.M1

This text is part of: *Boe. Cons. 1.M1*
Click on a word to bring up persons, dictionary entries, and frequency statistics

Carmina qui epidemus studio florente peregi,
flebilis haec manus cogit exire nos.
ecce milie lacrera dictant scribenda Camenae
et seris elegi flitibus ora rigat,
has saltem nulla potuit peninsula terror,
glori felicitate queat in mea luctu.
solus manet nunc mala fati semis,
unit enim propensa mala inspeca senectus
et dolor acetabul insit inesse sum.
intempestivum funduntur ureticu cani
et tremunt efforce corpora laxa cutis,
mentisq; ardore rorat, et membra dolibus annis
inservit et macta saepe auctoritate uenit
eluen, quam suad miseris auctoritatem aure
et flentes oculos claudere sancta negat!
dine leibus male fida bonis fortuna fauert
et uita uitaq; uitaq; uitaq; uitaq; uitaq; uitaq;
nunq; quis fallax mutatus mobili subtilum
prosternit felicem totius inactias, amici?
qui eccliti stabili non erat illi gradus.

Consolatio Philosophiae: Boethius: James J. O'Donnell.

Notes (Boethius) References (1 total)
+ Commentary references to this page (1): Boethius, Consolatio Philosophiae, Notizien 1

Vocabulary Tool Search

Search

10 Searching in Latin... More search options
Search for:
11 Search for all inflected forms
Search for exact forms only
12 Search for exact forms only
13 Search for exact forms only
Static Identifiers
Citation URI: <http://urn.nbn.se/resolve?urn=urn:nbn:se:umu:diva-50581> (Version 1.1, 1.M1)
Text URI: <http://urn.nbn.se/resolve?urn=urn:nbn:se:umu:diva-50581> (Version 1.1, 1.M1)
Page URI: <http://urn.nbn.se/resolve?urn=urn:nbn:se:umu:diva-50581> (Version 1.1, 1.M1)
Category URI: <http://urn.nbn.se/resolve?urn=urn:nbn:se:umu:diva-50581> (Version 1.1, 1.M1)

Display Preferences
Greek Display: Unicode (precombined)
Arabic Display: Unicode

The diagram illustrates the etymological analysis of the word 'mulier'. It starts with the Latin word 'mulier' at the top, which is analyzed into 'descendens advcl/pred VERB' and 'nescio advmod VERB'. These two components are then further analyzed into 'subito' (ADV), 'admod' (ADV), 'heu' (discourse INTJ), 'mulier' (NOUN), 'fulgor' (NOUN), 'descendens' (advcl/pred VERB), 'nescio' (advmod VERB), and finally 'apparuit' (root VERB). The analysis is presented in a hierarchical tree structure with arrows indicating the flow of analysis from the original word to its constituent parts and then to their individual meanings.

In other words, today...



We have built and collected (for Latin and other languages):

- ▶ Textual Resources
- ▶ Lexical Resources
- ▶ NLP Tools

In other words, today...



We have built and collected (for Latin and other languages):

- ▶ Textual Resources
- ▶ Lexical Resources
- ▶ NLP Tools

Scattered and unconnected

No “One-Ring-to-Rule-Them-All” Approach

Pelagios



What Pelagios isn't | One ring to rule them all



- Data aggregation X
- Standard data representation X
- MEGA search portal X



Connectivity through *common references* rather than a *common schema*



Open Access Published by De Gruyter Saur 2019

Sustaining Linked Ancient World Data

From the book [Digital Classical Philology](#)

Hugh A. Cayless

<https://doi.org/10.1515/9783110599572-004>

Cite this

Citations

1

[Download chapter \(PDF\)](#)

From the book



Digital Classical Philology

Search book



Chapters in this book (24)

Frontmatter

Editor's Preface

Preface

Contents

Introduction

The Free First Thousand Years of Greek

The Digital Latin Library: Cataloging and Publishing Critical Editions of Latin Texts

Hugh A. Cayless

Sustaining Linked Ancient World Data

Abstract: May 31st, 2018 marked the sixth anniversary of the Linked Ancient World Data Institute (LAWDI), a workshop funded by the US National Endowment For the Humanities. This makes it a good time to take stock of the Ancient World Linked Data initiatives that have been around for some time, as well as some that have foundered and some that are new. What makes for sustainable Linked Open Data? Why do some initiatives thrive while others fail? What resources do successful LOD sites need, and how may *they* be obtained? The promise of LOD is that it frees our information from the silos in



ERC Consolidator Grant 2018-2023

A collection of multifarious, interoperable linguistic resources described with the same vocabulary for knowledge description
(by using common data categories and ontologies)

Interlinking as a Form of Interaction

LR: What

“Language resources are collections of data which primarily document communicative acts of humans by some form of recording and/or descriptions, both directly as in corpora, or at higher levels of abstraction in lexicons and ontologies”

(International Standard for Language Engineering, ISLE)

- Pivotal components in NLP
- Machine-readable form
- Data: written, spoken, audio/video
- Tools for acquiring, building, maintaining, sharing, querying LR

LiLa Knowledge Base

Lexically-based architecture and (meta)data sources

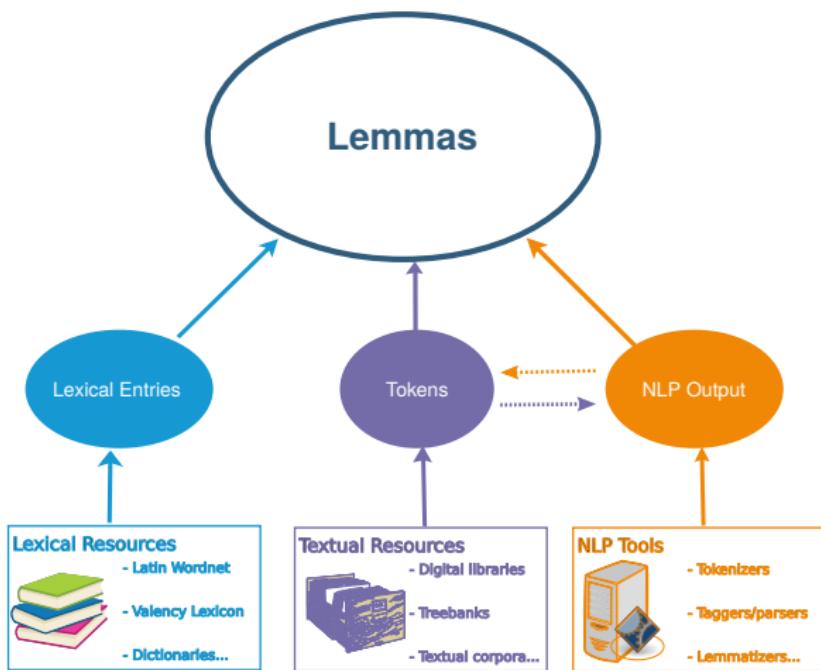


Table of Contents



Introduction and Fundamentals

LiLa: Mission and Architecture

LiLa now!

Lemma Bank and Lexical Resources

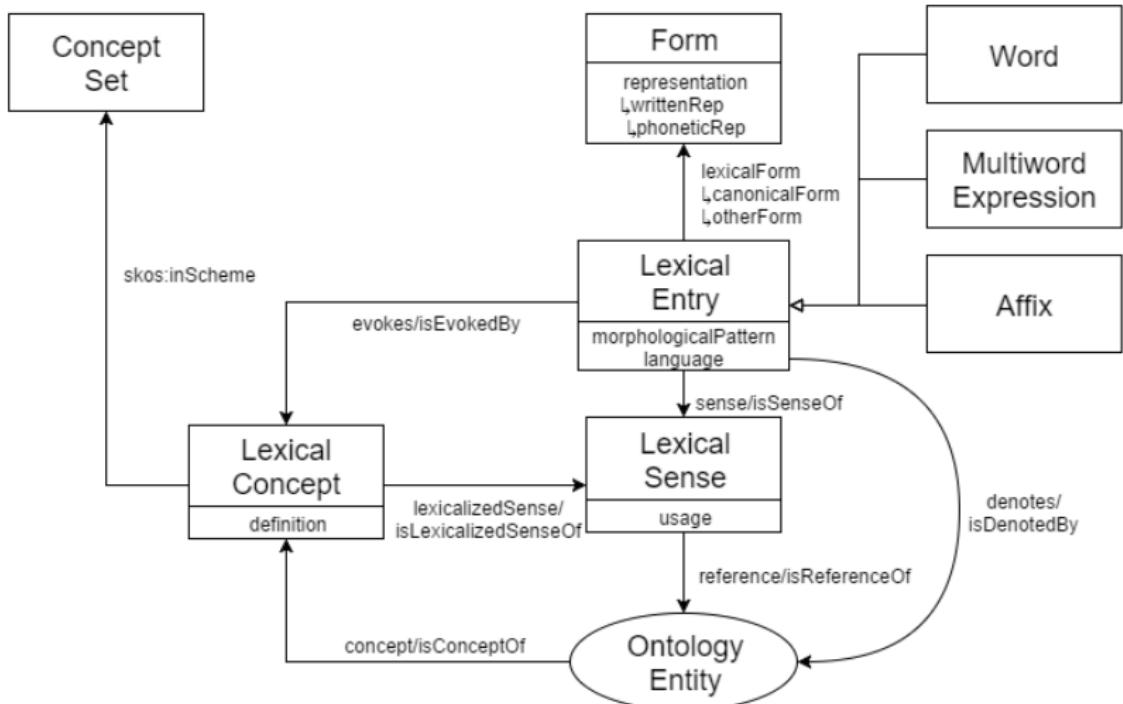
Textual Resources

Services and Tools

To sum up

LiLa and Ontolex Lemon

A *de facto* W3C standard for publishing lexical data as LLOD



Lemma Bank and Lexical Resources

The core of LiLa



Lemma *abdico* ‘to deny, disown’

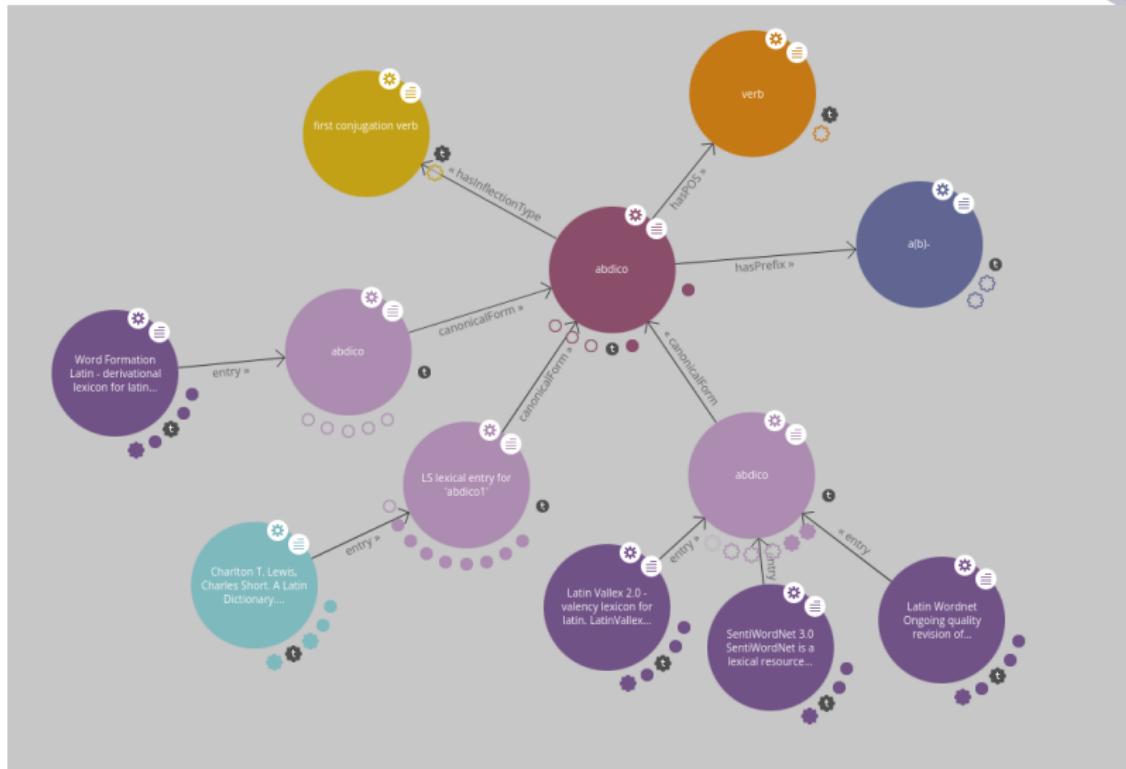
<http://lila-erc.eu/data/id/lemma/86860>

Lemmas and Lexical Entries

abdico, 'deny, disown'



12



Lexical resources

And the adopted ontologies



- ▶ Lemma Bank: Ontolex, custom
- ▶ A Lat.-Eng. dictionary (Lewis & Short): Ontolex + lexicog
- ▶ A derivational lexicon (WFL): Morph
- ▶ A polarity lexicon (LatinAffectus): Ontolex + Marl
- ▶ An etymological dictionary (De Vaan): LemonEty
- ▶ A dictionary of Greek loanwords: LemonEty
- ▶ A Valency Lexicon (Latin Vallex): PreMOn
- ▶ A manually checked subset of the Latin WordNet: Ontolex

Table of Contents



Introduction and Fundamentals

LiLa: Mission and Architecture

LiLa now!

Lemma Bank and Lexical Resources

Textual Resources

Services and Tools

To sum up

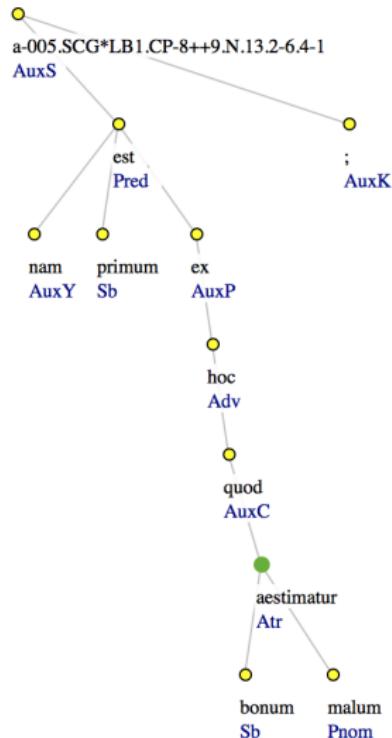
(Annotated) Corpora in LiLa

Source: The *Index Thomisticus* Treebank (original scheme)



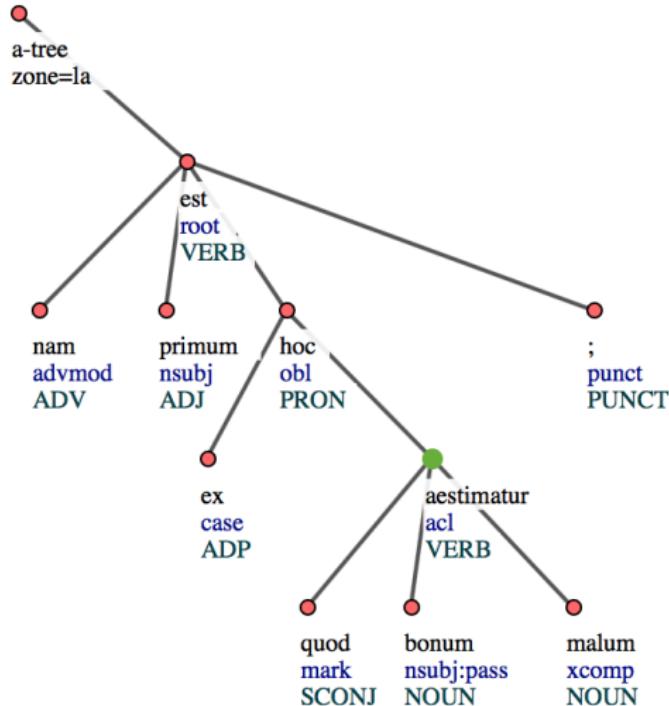
*nam primum est ex hoc
quod bonum **aestimatur**
malum; (IT-TB: SCG, lib. 1,
cap. 89, n. 13)*

*for the first arises because
the good **is judged** to be
evil; (Trans. Anton C. Pegis)*



(Annotated) Corpora in LiLa

Source: The *Index Thomisticus* Treebank (UD scheme)



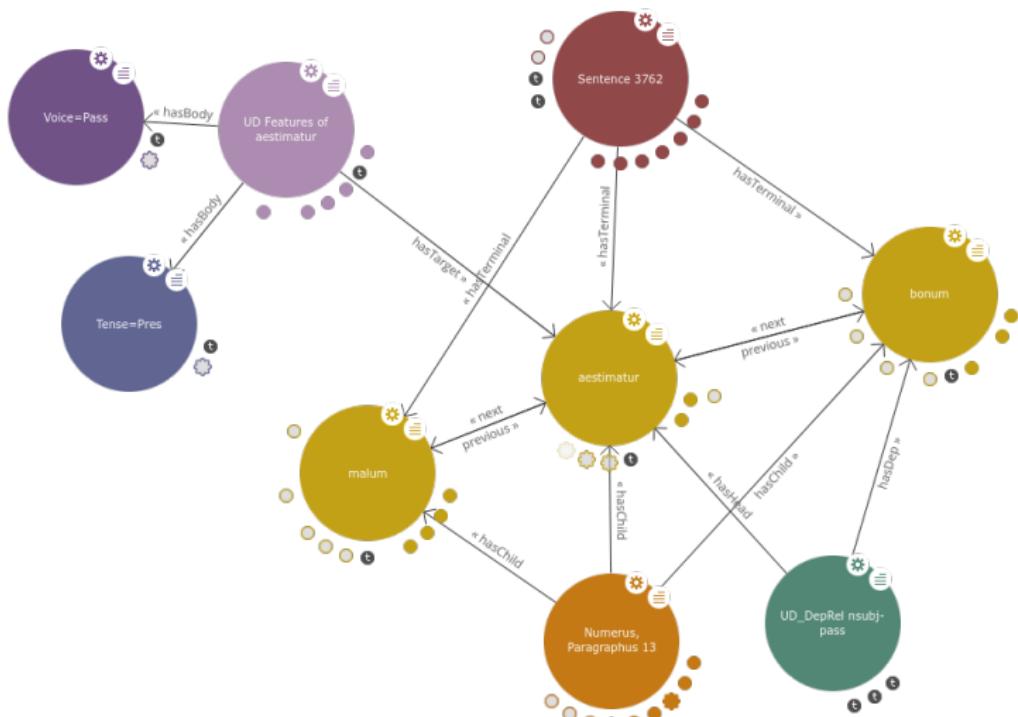
Token *aestimatur*

[http://lila-erc.eu/lodview/data/corpora/
ITTB/id/token/005.SCG*LB1.CP-8++9.N.13.](http://lila-erc.eu/lodview/data/corpora/ITTB/id/token/005.SCG*LB1.CP-8++9.N.13.)

2-6.4-1W8

The token *aestimatur*

Morpho-syntactic information



The token *aestimatur*

Other metadata

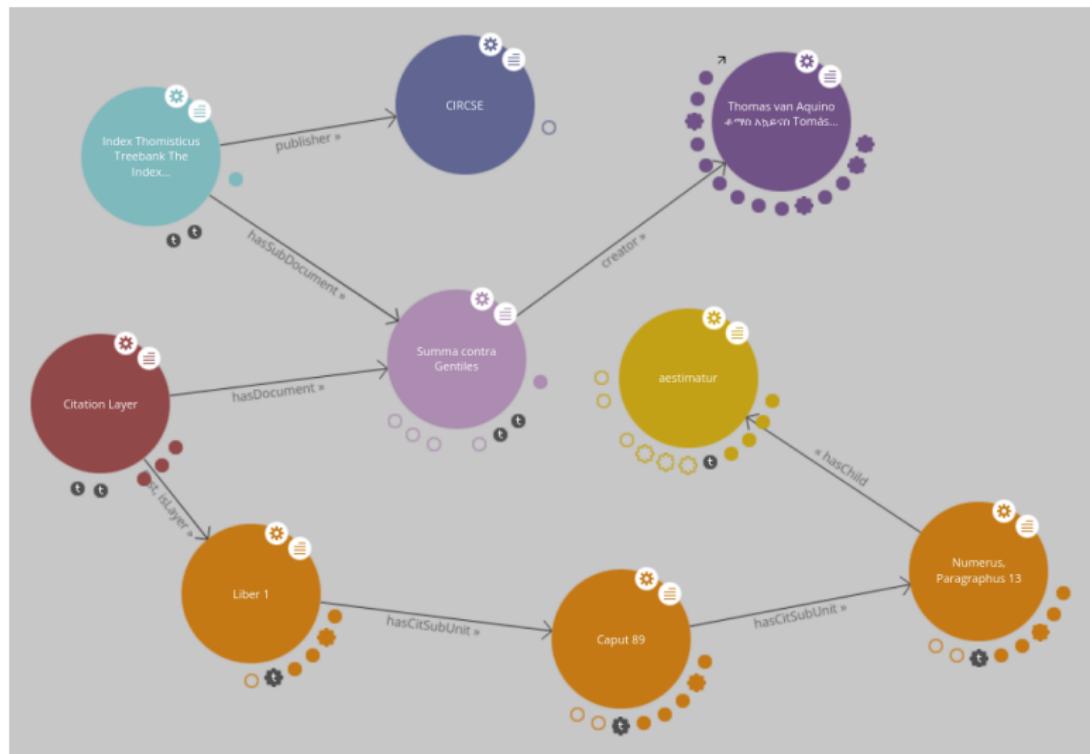


Table of Contents



Introduction and Fundamentals

LiLa: Mission and Architecture

LiLa now!

Lemma Bank and Lexical Resources

Textual Resources

Services and Tools

To sum up

Lemma Bank Query Interface

<https://lila-erc.eu/query/>

SPARQL Access Point

<https://lila-erc.eu/sparql/>

TextLinker

<http://lila-erc.eu:8080/LiLaTextLinker/>

Table of Contents



Introduction and Fundamentals

LiLa: Mission and Architecture

LiLa now!

Lemma Bank and Lexical Resources

Textual Resources

Services and Tools

To sum up

► Textual Resources

- Index Thomisticus Treebank (*Summa contra Gentiles*): ca. 400,000 nodes
- UDante Treebank: ca. 46,000 tokens
- Querolus sive Aulularia*: ca. 17,000 tokens
- Liber Abbaci* (ch. VIII) by Leonardo Fibonacci: ca. 30,000 tokens
- LASLA Corpus: ca. 1.7 million tokens
- PROIEL and LLCT treebanks, CompHistSem, CroALa, Musisque DeoQue

► Lexical Resources

- Lemma Bank: ca. 200,000 canonical forms
- Word Formation Latin: ca. 36,000 lemmas (Classical Latin)
- Etymological Dict. of Latin & the Other Italic Langs.: ca. 1,500 entries
- LatinAffectus: ca. 3,300 entries
- Index Graecorum Vocabulorum in L. Latinam Transl.: ca. 1,800 entries
- Latin WordNet: ca. 2,500 manually checked entries
- Latin Vallex 2.0: ca. 2,000 entries
- Lewis & Short Dictionary: ca. 50,000 entries
- Lexikon der Indogermanischen Verben (LIV). Wiktionary, BabelNet

TOTAL: approximately 33 million triples

Keep in touch!!

Thank you



26

LiLa: Linking Latin

Università Cattolica del Sacro Cuore
CIRCSE Research Centre

✉ 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.