



UNIVERSITÀ
CATTOLICA
del Sacro Cuore



The LiLa Knowledge Base in a Nutshell

The LiLa Team

LD4LT Annotaton Workshop
Conference *Language, Data and Knowledge* (LDK 2021)
4 September, Zaragoza, Spain



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.

Research question

State of affairs



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

Research question

State of affairs



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

- ▶ Textual Resources

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

- ▶ Textual Resources
- ▶ Lexical Resources

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

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

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

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

Scattered and unconnected

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)

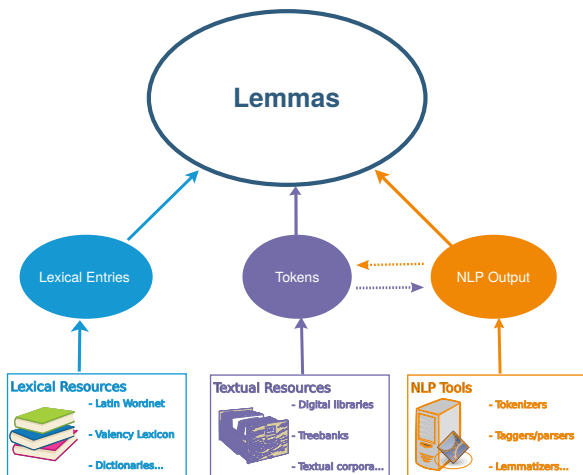
Interlinking as a Form of Interaction



Infrastructure

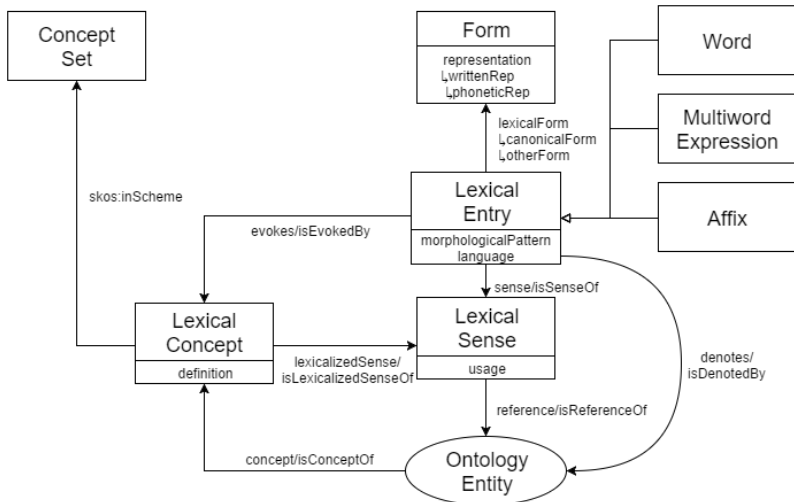


Interoperability



LiLa and Ontolex Lemon

A de facto W3C standard for publishing lexical data as LLOD



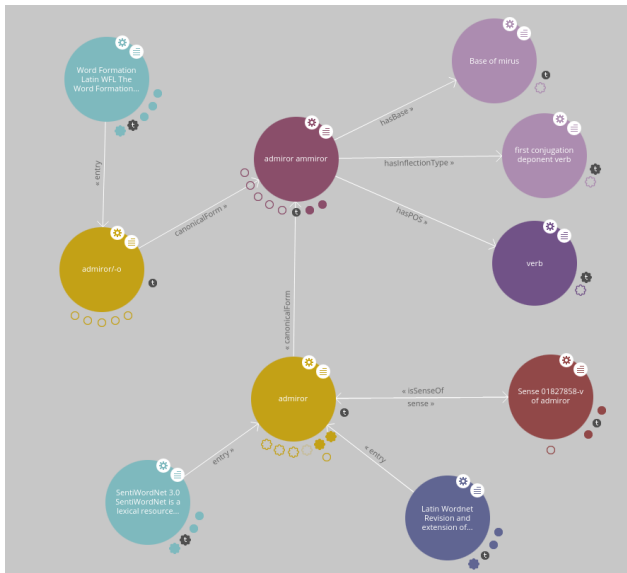
Lemma *admiror* 'to admire, to respect'

<http://lila-erc.eu/data/id/lemma/87541>

- ▶ Lemma Bank
- ▶ A derivational lexicon (Word Formation Latin)
- ▶ A polarity lexicon (LatinAffectus)
- ▶ An etymological dictionary (De Vaan)
- ▶ A Valency Lexicon (Latin Vallex)
- ▶ A manually checked subset of the Latin WordNet

The lemma admiror and its connections

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



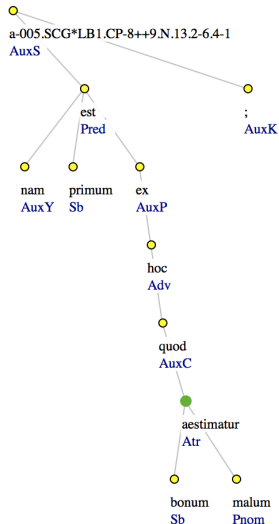
(Annotated) Corpora in LiLa

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



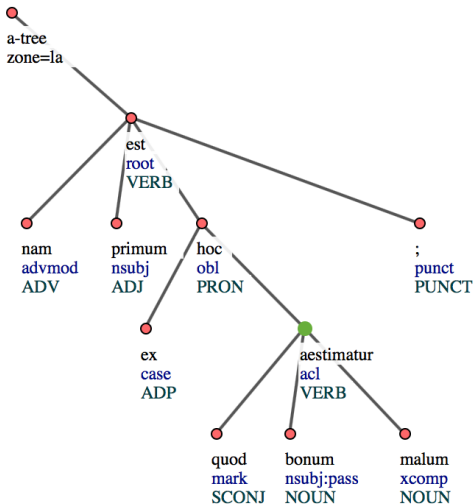
*nam primum est ex hoc
quod bonum **aestimatur**
malum; (IT-TB: SCG, lib. 1,
cap. 89, n. 13)*

*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)

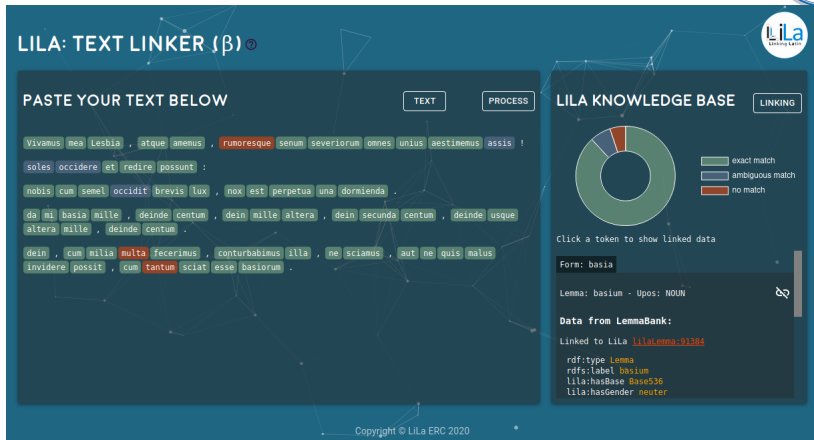


Token *aestimatur*

`http://lila-erc.eu/lodview/data/corpora/
ITTB/id/token/005.SCG*LB1.CP-8++9.N.13.
2-6.4-1W8`



Figure: LiLa's Text Linker



LILA: TEXT LINKER (β)

PASTE YOUR TEXT BELOW

TEXT PROCESS

Vivamus mea Lesbia, atque amemus, rumoresque senum severiorum omnes unius aestimemus assis!

soles occidere et redire possunt:

nobis cum semel occidit brevis lux, nox est perpetua una dormienda.

da mi basia mille, deinde centum, dein mille altera, dein secunda centum, deinde usque altera mille, deinde centum.

dein, cum milia multa fecerimus, conturbabimus illa, ne sciamus, aut ne quis malus invidere possit, cum tantum sciat esse basiorum.

LILA KNOWLEDGE BASE LINKING

Click a token to show linked data

Form: basia

Lemma: basium - Upos: NOUN

Data from LemmaBank:

Linked to LiLa [LiLaLemma:91394](#)

rdf:type Lemma
rdfs:label basium
lila:hasBase Base536
lila:hasGender neuter

Legend: exact match (green), ambiguous match (blue), no match (red)

Copyright © LiLa ERC 2020

Figure: Text processed against the LiLa Knowledge Base

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

► Corpora

- ✓ Index Thomisticus Treebank (*Summa contra Gentiles*): ca. 450,000 nodes
- ✓ Dante Search (700th death anniversary): ca. 46,000 tokens
- ✓ Querolus sive Aulularia: ca. 17,000 tokens
- PROIEL and LLCT treebanks
- Computational Historical Semantics, LASLA and CroALa Corpora

► Lexica

- ✓ Word Formation Latin: ca. 46,000 lemmas (Classical Latin)
- ✓ Etymological dictionary of Latin & the other Italic Langs.: ca. 1,400 entries
- ✓ LatinAffectus: ca. 2,300 entries
- ✓ Index Graecorum Vocabulorum in Linguam Latinam: ca. 1,800 entries
- ✓ Latin WordNet: ca. 1,000 manually checked entries
- ✓ Latin Vallex 2.0: Valency Lexicon
- Lewis & Short Dictionary

► NLP tools

- ✓ LEMLAT (lemma bank): ca. 150,000 lemmas

► TOTAL: approximately 13 million triples

Query Interface, Triplestore and Linker

- ▶ Query interface; Triplestore
- ▶ Linker

Linguistic Resources. Corpora

- ▶ Index Thomisticus Treebank
- ▶ Dante Search
- ▶ *Querolus sive Aulularia*

Linguistic Resources. Lexica

- ▶ Word Formation Latin
- ▶ Etymological Dictionary of Latin & the Other Italic Languages
- ▶ LatinAffectus
- ▶ Index Graecorum Vocabulorum in Linguam Latinam
- ▶ Latin WordNet
- ▶ Latin Vallex 2.0