Linking Latin

The LiLa Knowledge Base of Interlinked Linguistic Resources for Latin

Marco Passarotti, Flavio Massimiliano Cecchini, Greta Franzini, Eleonora Litta, Francesco Mambrini, Giovanni Moretti, Paolo Ruffolo, Rachele Sprugnoli, Marinella Testori

CLARIN Annual Conference 2020
Bazaar | October 7

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):

- 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

CLARIN
Common Language Resources and Technology Infrastructure

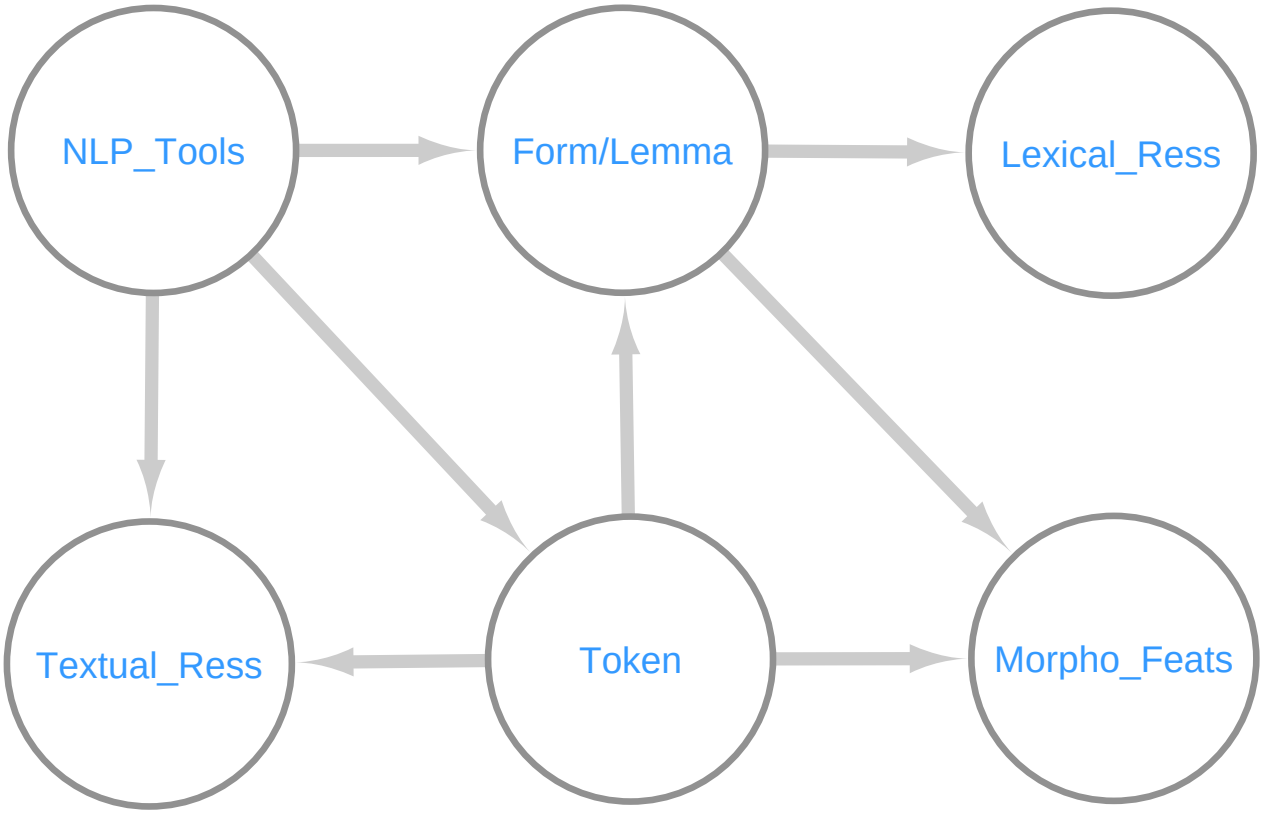
Infrastructure

LiLa
Linking Latin

Interoperability

LiLa Knowledge Base

Lexically-based architecture and (meta)data sources

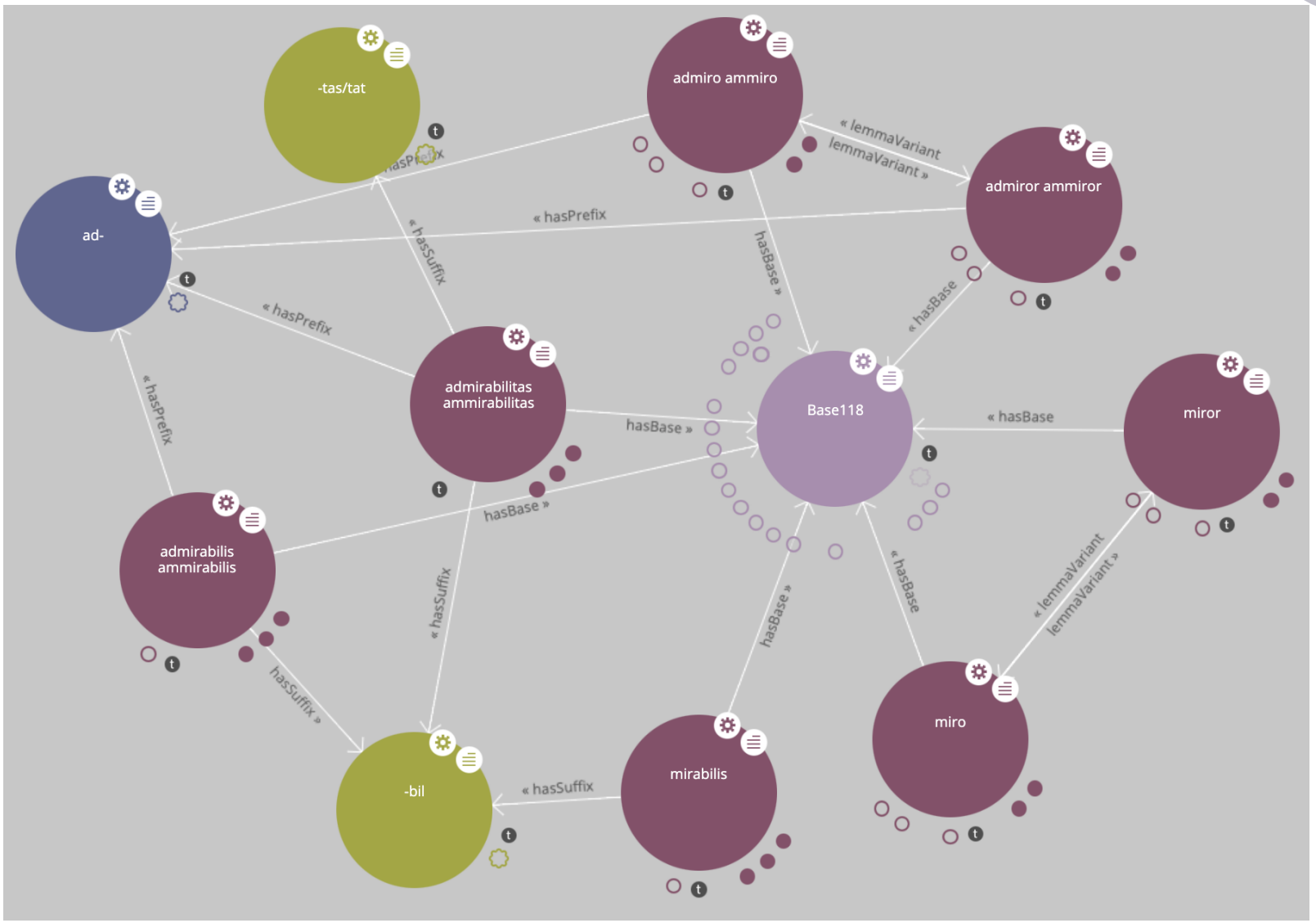


Lemma *admiror* ‘admire, respect’

https://lila-erc.eu/data/id/lemma/87541

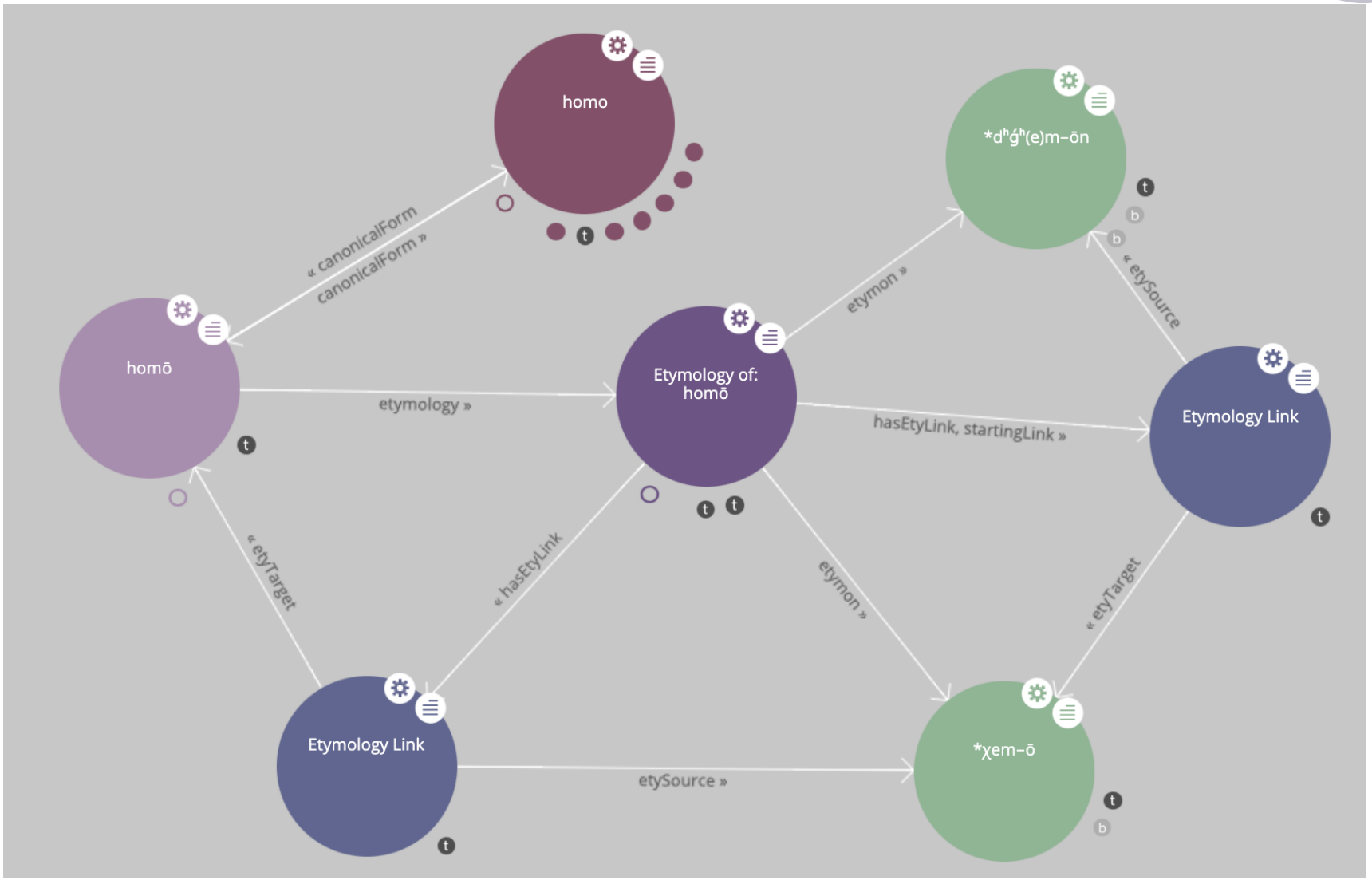
LiLa: Structure. Derivational Morphology

Word Formation Latin



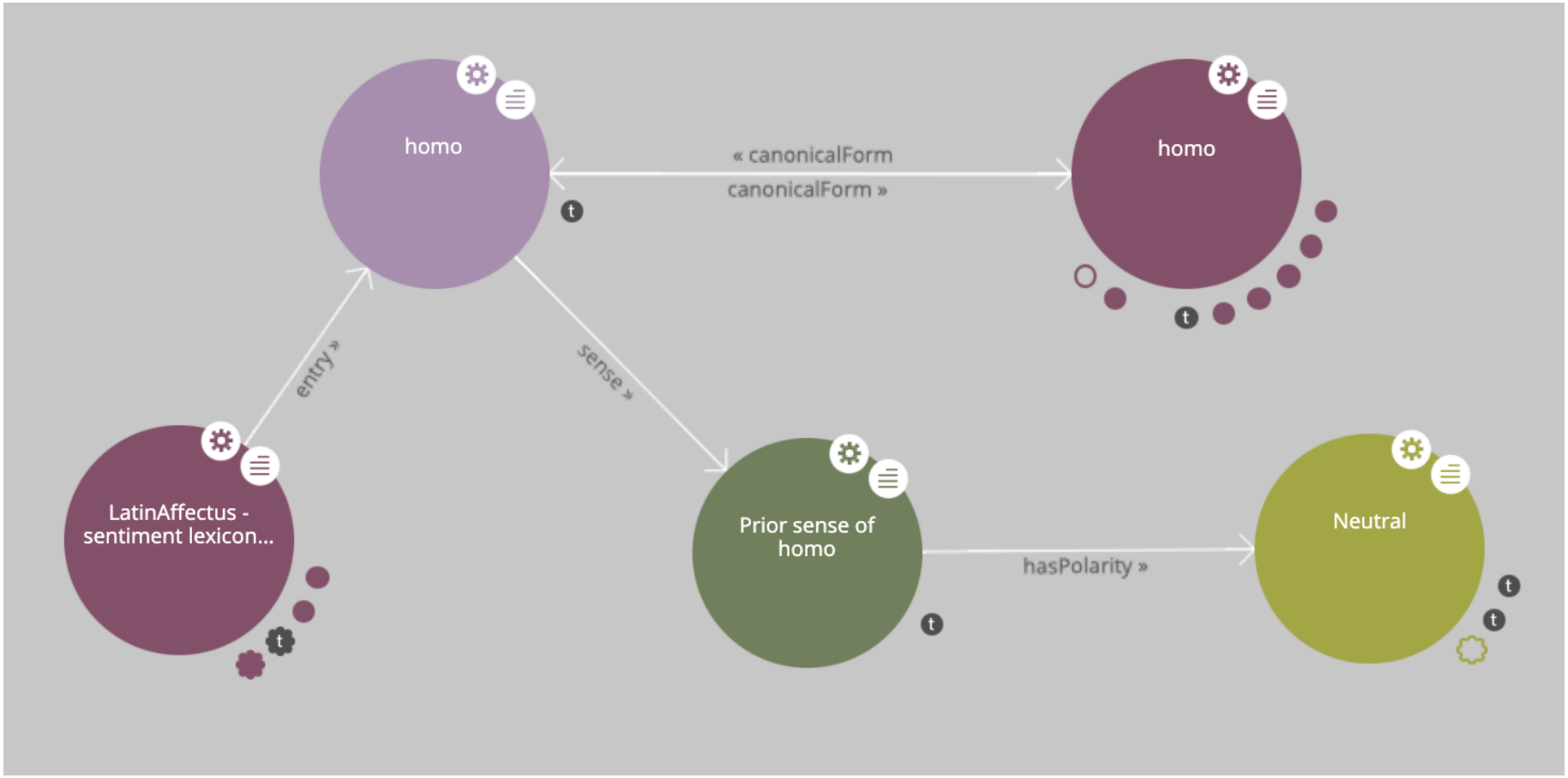
LiLa: Structure. Etymology

Etymological dictionary of Latin and the other Italic Languages



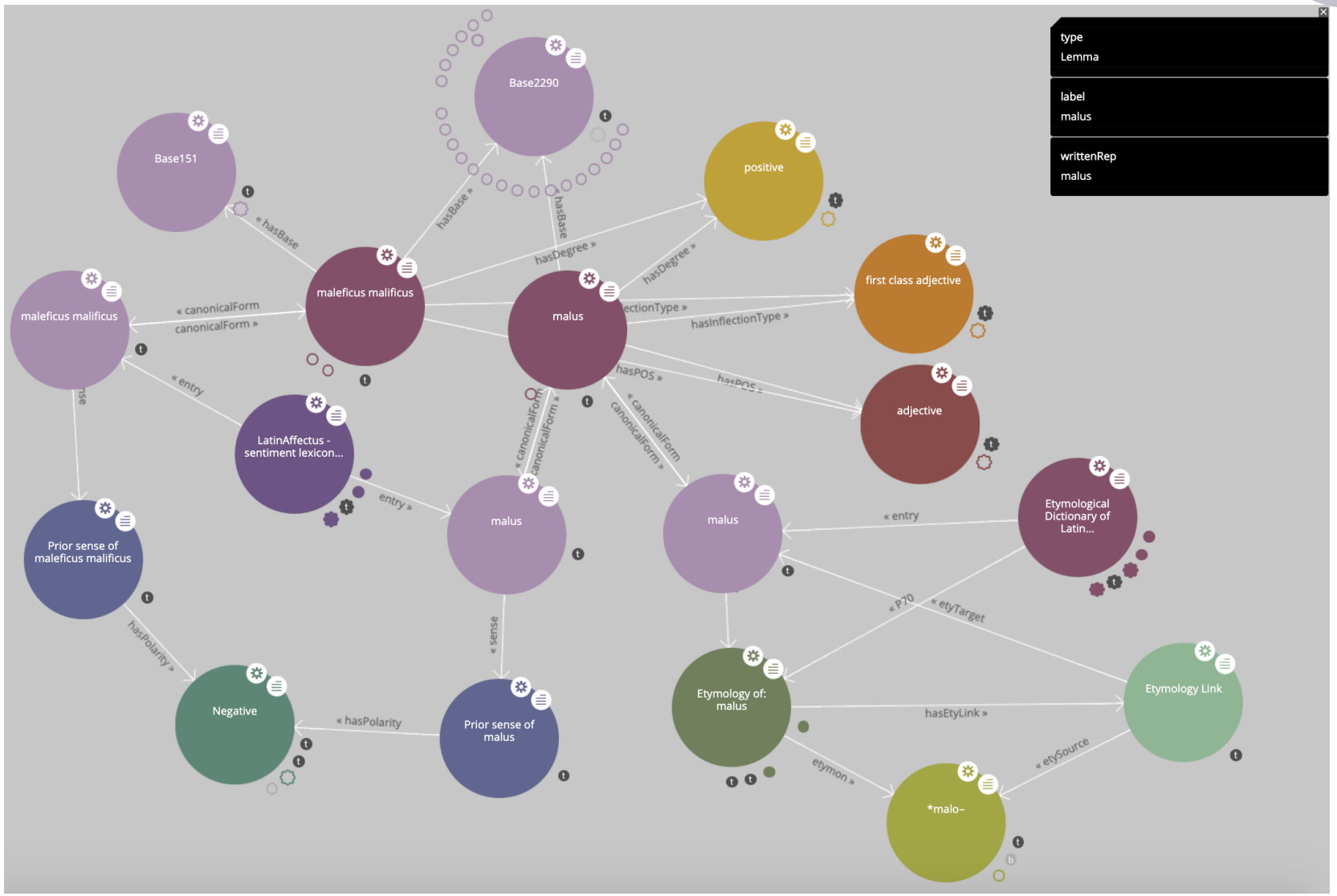
LiLa: Structure. Polarity

LatinAffectus



LiLa: Structure. All together now!

Lemma Bank, Derivational Morphology, Etymology and Polarity

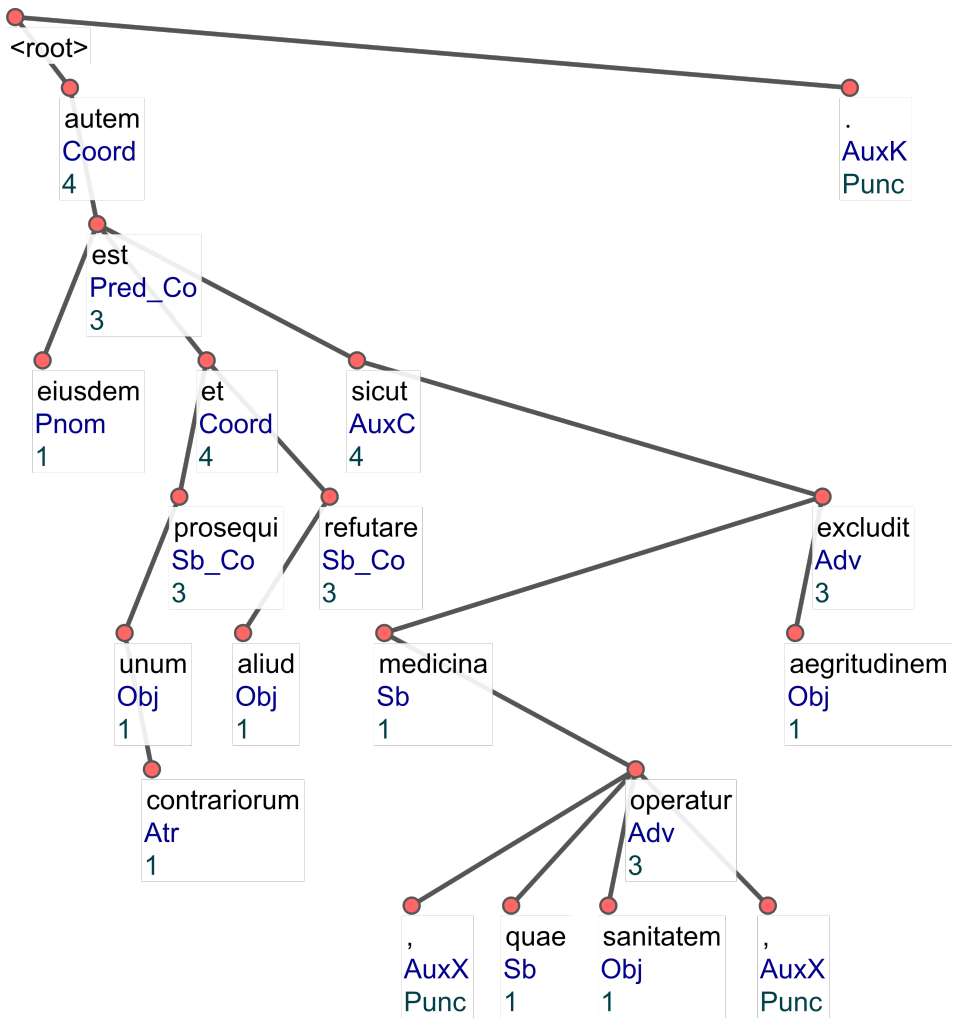


LiLa: Structure. Syntax

The *Index Thomisticus* Treebank: Dependency trees

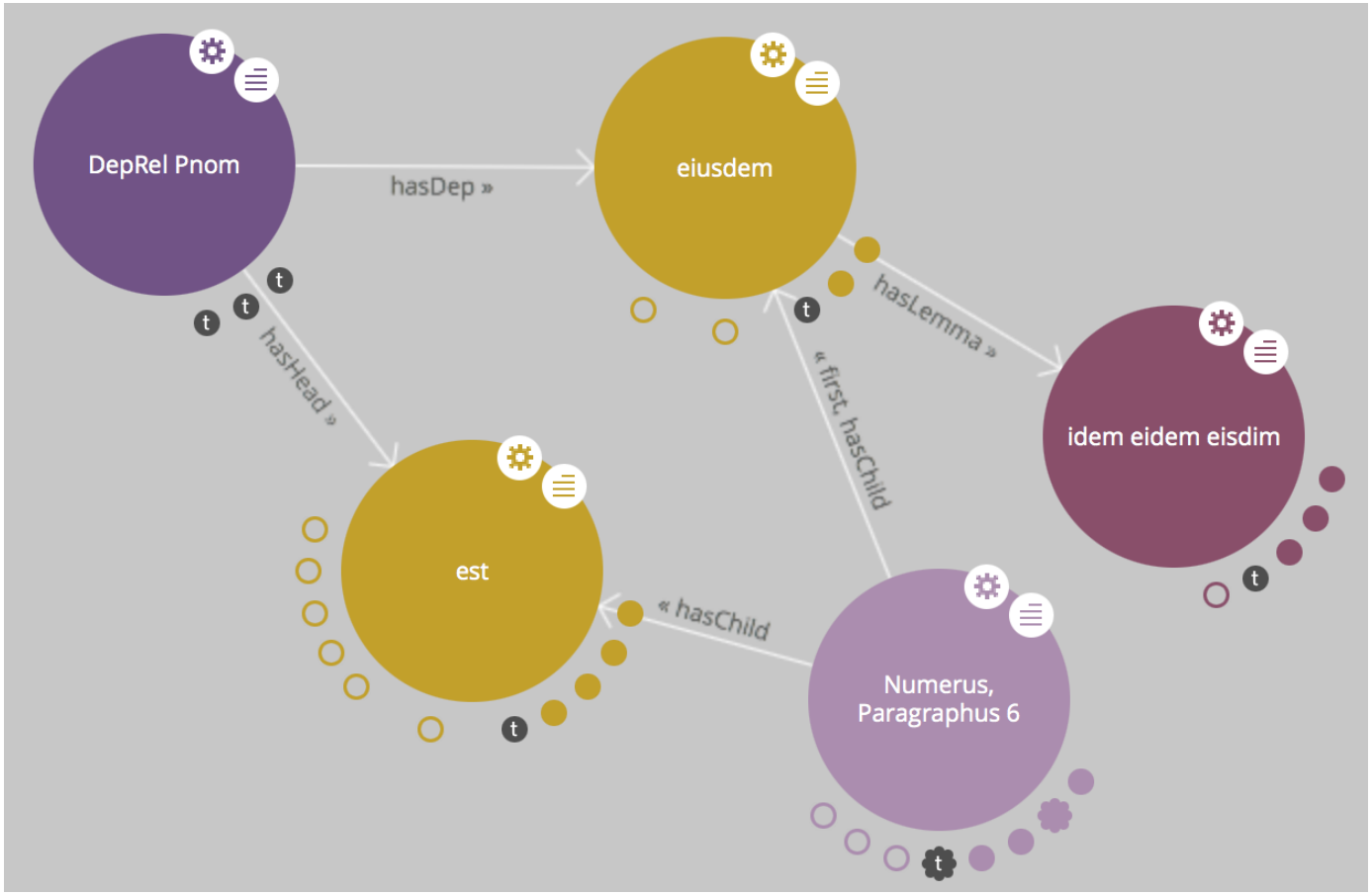
eiusdem *autem* ***est*** *unum contrariorum prosequi et aliud refutare sicut medicina , quae sanitatem operatur , aegritudinem excludit .* (IT-TB: SCG, lib. 1, cap. 1, n. 6)

Now ***it belongs to the same thing*** *to pursue one contrary and to remove the other: thus medicine, which effects health, removes sickness.* (Trans. Laurence Shapcote)



LiLa: Structure. Syntax

Phenomena and noumena



LiLa: Overview

Resources connected and upcoming connections

- **Corpora**
  - ☑ Index Thomisticus Treebank (*Summa contra Gentiles*): ca. 400,000 nodes
  - ☑ Dante Search (700th death anniversary coming up!): ca. 46,000 tokens
  - ☑ *Querolus sive Aulularia*: ca. 17,000 tokens
  - ☐ PROIEL and LLCT treebanks
  - ☐ Computational Historical Semantics Corpus
- **Lexica**
  - ☑ Word Formation Latin: ca. 46,000 lemmas (Classical Latin)
  - ☑ BRILL Etymological dictionary of Latin and the other Italic Languages: ca. 1,400 entries
  - ☑ LatinAffectus (sentiment lexicon for Latin): ca. 2,300 entries
  - ☑ Index Graecorum Vocabulorum in Linguam Latinam: ca. 1,800 entries
  - ☐ Latin WordNet & Latin Vallex 2.0
- **NLP tools**
  - ☑ LEMLAT (lemma bank): ca. 150,000 lemmas
- **TOTAL: approximately 5 million triples**

Query interface and Triplestore

- https://lila-erc.eu/query/
- https://lila-erc.eu/sparql/

Corpora

- https://lila-erc.eu/data/corpora/ITTB/id/corpus
- https://lila-erc.eu/data/corpora/DanteSearch/id/corpus
- https://lila-erc.eu/data/corpora/Querolus/id/citationUnit/QuerolussiveAulularia

Lexica

- https://lila-erc.eu/data/lexicalResources/BrillEDL/Lexicon
- https://lila-erc.eu/data/lexicalResources/LatinAffectus/Lexicon
- https://lila-erc.eu/data/lexicalResources/IGVLL/Lexicon

Thanks!

Get in touch

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.