



UNIVERSITÀ
CATTOLICA
del Sacro Cuore



Linking Latin (LiLa)

Building a Knowledge Base of Linguistic Resources for Latin

Marco Passarotti

marco.passarotti@unicatt.it

Word Formation, Grammar and Lexicology in a Multilingual Context
Rome | February 6 2019



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

WHY. Introduction

- Research Question

- Research Need

WHAT & HOW. The LiLa Knowledge Base

- Approach

- The Ontology. Conceptual and Structural Interoperability

- Word Formation in LiLa

Conclusion

We have built and collected (for Latin and more):

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

We have built and collected (for Latin and more):

- ▶ Textual Resources
 - ▶ Digital Libraries (Perseus), Collections of Texts (LASLA, ALIM, Musisque Deoque, Patrologia Latina, PHI, Library of Latin Texts etc.), Treebanks (LDT, IT-TB, PROIEL)
- ▶ Lexical Resources
- ▶ NLP Tools

We have built and collected (for Latin and more):

- ▶ Textual Resources
 - ▶ Digital Libraries (Perseus), Collections of Texts (LASLA, ALIM, Musisque Deoque, Patrologia Latina, PHI, Library of Latin Texts etc.), Treebanks (LDT, IT-TB, PROIEL)
- ▶ Lexical Resources
 - ▶ Valency Lexica (Vallex, IT-VaLex), Latin WordNet, Dictionaries (Du Cange, TLL, OLD, TTLF, DMLBS & DB of Latin Dictionaries c/o Brepols)
- ▶ NLP Tools

We have built and collected (for Latin and more):

- ▶ Textual Resources

- ▶ Digital Libraries (Perseus), Collections of Texts (LASLA, ALIM, Musisque Deoque, Patrologia Latina, PHI, Library of Latin Texts etc.), Treebanks (LDT, IT-TB, PROIEL)

- ▶ Lexical Resources

- ▶ Valency Lexica (Vallex, IT-VaLex), Latin WordNet, Dictionaries (Du Cange, TLL, OLD, TTLF, DMLBS & DB of Latin Dictionaries c/o Brepols)

- ▶ NLP Tools

- ▶ Morphological Analysers (LEMLAT, Words, LatMor), PoS Taggers (TreeTagger, Collatinus, UDPipe), Dependency Parsers (UDPipe), CLTK

We have built and collected (for Latin and more):

- ▶ Textual Resources

- ▶ Digital Libraries (Perseus), Collections of Texts (LASLA, ALIM, Musisque Deoque, Patrologia Latina, PHI, Library of Latin Texts etc.), Treebanks (LDT, IT-TB, PROIEL)

- ▶ Lexical Resources

- ▶ Valency Lexica (Vallex, IT-VaLex), Latin WordNet, Dictionaries (Du Cange, TLL, OLD, TTLF, DMLBS & DB of Latin Dictionaries c/o Brepols)

- ▶ NLP Tools

- ▶ Morphological Analysers (LEMLAT, Words, LatMor), PoS Taggers (TreeTagger, Collatinus, UDPipe), Dependency Parsers (UDPipe), CLTK

Scattered and Unconnected

To make sense of this quantity of empirical data

To make sense of this quantity of empirical data:

- ▶ to extract maximum benefit from our research investments
- ▶ to make computational resources and tools really impact and improve life of Classicists

To make sense of this quantity of empirical data:

- ▶ to extract maximum benefit from our research investments
- ▶ to make computational resources and tools really impact and improve life of Classicists

From Information to Knowledge

Interlinking to Make Interact

LiLa is based on an ontology made of:

LiLa is based on an ontology made of:

- ▶ **Individuals:** instances of objects (one specific token, lemma etc.)

LiLa is based on an ontology made of:

- ▶ **Individuals:** instances of objects (one specific token, lemma etc.)
- ▶ **Classes:** types of objects/concepts (token, lemma, PoS etc.)

LiLa is based on an ontology made of:

- ▶ **Individuals:** instances of objects (one specific token, lemma etc.)
- ▶ **Classes:** types of objects/concepts (token, lemma, PoS etc.)
- ▶ **Attributes:** properties that objects can/must have (morphological features for lemmas/tokens). Each attribute can be a class or an individual

LiLa is based on an ontology made of:

- ▶ **Individuals:** instances of objects (one specific token, lemma etc.)
- ▶ **Classes:** types of objects/concepts (token, lemma, PoS etc.)
- ▶ **Attributes:** properties that objects can/must have (morphological features for lemmas/tokens). Each attribute can be a class or an individual
- ▶ **Relations:** ways in which classes and individuals can be related to one another: RDF triples.

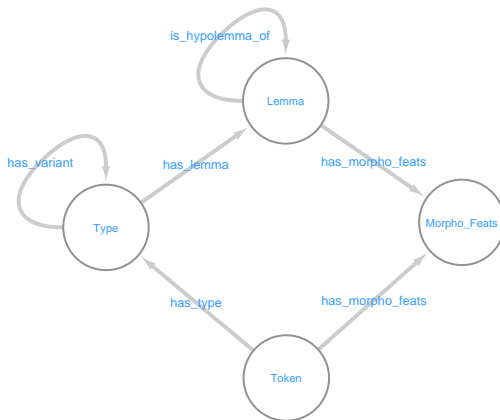
Labels from a restricted dictionary of knowledge description:

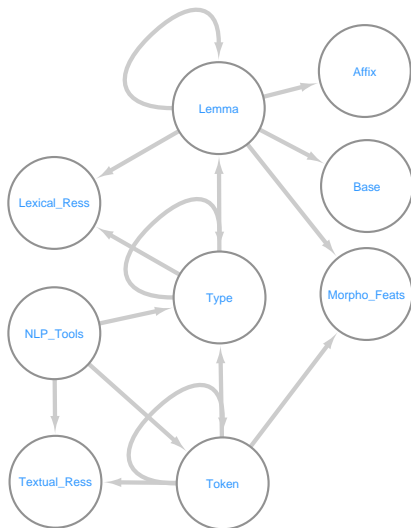
has_lemma, has_PoS

LiLa is based on an ontology made of:

- ▶ **Individuals:** instances of objects (one specific token, lemma etc.)
- ▶ **Classes:** types of objects/concepts (token, lemma, PoS etc.)
- ▶ **Attributes:** properties that objects can/must have (morphological features for lemmas/tokens). Each attribute can be a class or an individual
- ▶ **Relations:** ways in which classes and individuals can be related to one another: RDF triples.
Labels from a restricted dictionary of knowledge description:
has_lemma, *has_PoS*

Each component of the ontology is uniquely identified through a URI.





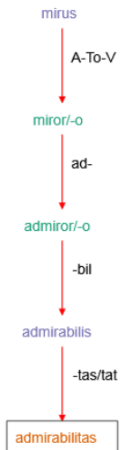
The LiLa Knowledge Base

Word Formation. WFL: Morphotactic and Hierarchical



The LiLa Knowledge Base

Word Formation. WFL: Morphotactic and Hierarchical



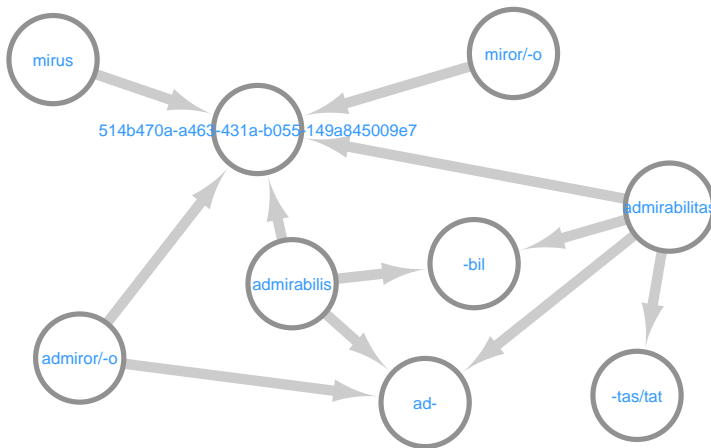
The LiLa Knowledge Base

Word Formation. LiLa: Paradigmatic



Word Formation. LiLa: Paradigmatic





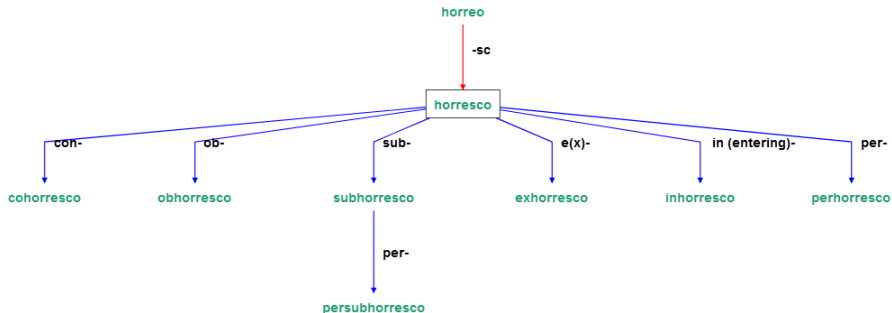
horreo

e(x)-

exhorreo

The LiLa Knowledge Base

Word Formation. WFL: Morphotactic and Hierarchical. Who's First?

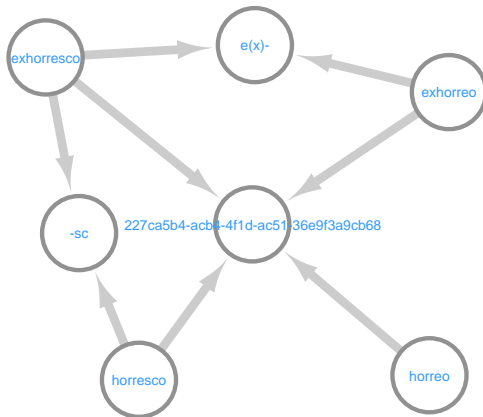


Word Formation. LiLa: Paradigmatic. Nobody's First



The LiLa Knowledge Base

Word Formation. LiLa: Paradigmatic. Nobody's First



What's going on in LiLa (2018-2023):

- ▶ Completing **Word Formation Latin** (E. Litta)
- ▶ Assessing and Extending **Latin WordNet** (G. Franzini)
- ▶ Extending the ***Index Thomisticus* Treebank** (M. Testori)
- ▶ Populating and Testing the **Knowledge Base** (F. Mambrini)
- ▶ Testing and Developing **PoS Taggers** for Latin (F. Cecchini)

Marco Passarotti

Università Cattolica del Sacro Cuore
CIRCSE Research Centre



marco.passarotti@unicatt.it



<https://github.com/CIRCSE>



<https://lila-erc.eu>



<http://wfl.marginalia.it/>



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