



THE LiLa KNOWLEDGE BASE AS A REFERENCE MODEL FOR INTEROPERABILITY BETWEEN LINGUISTIC RESOURCES



Relevant UniDive working groups: WG1, WG2

Flavio Massimiliano Cecchini and Marco Passarotti
Università Cattolica del Sacro Cuore, Milan, Italy

Introduction

Over the past two decades the research area dedicated to **linguistic resources** has seen substantial growth and today covers a **wide span** of languages and language varieties. But despite the increase in the quantity and coverage of linguistic resources, most of these are still locked in their own **“data silos”**, making use of different, non-interacting standards. To overcome this issue for **Latin**, the **LiLa: Linking Latin** project has built a knowledge base of linguistic resources based on

- the **Universal Dependencies (UD)** standard as its basic, shared annotation formalism, and
- the principles of **Linguistic Linked Open Data (LLOD)** to interlink linguistic resources.

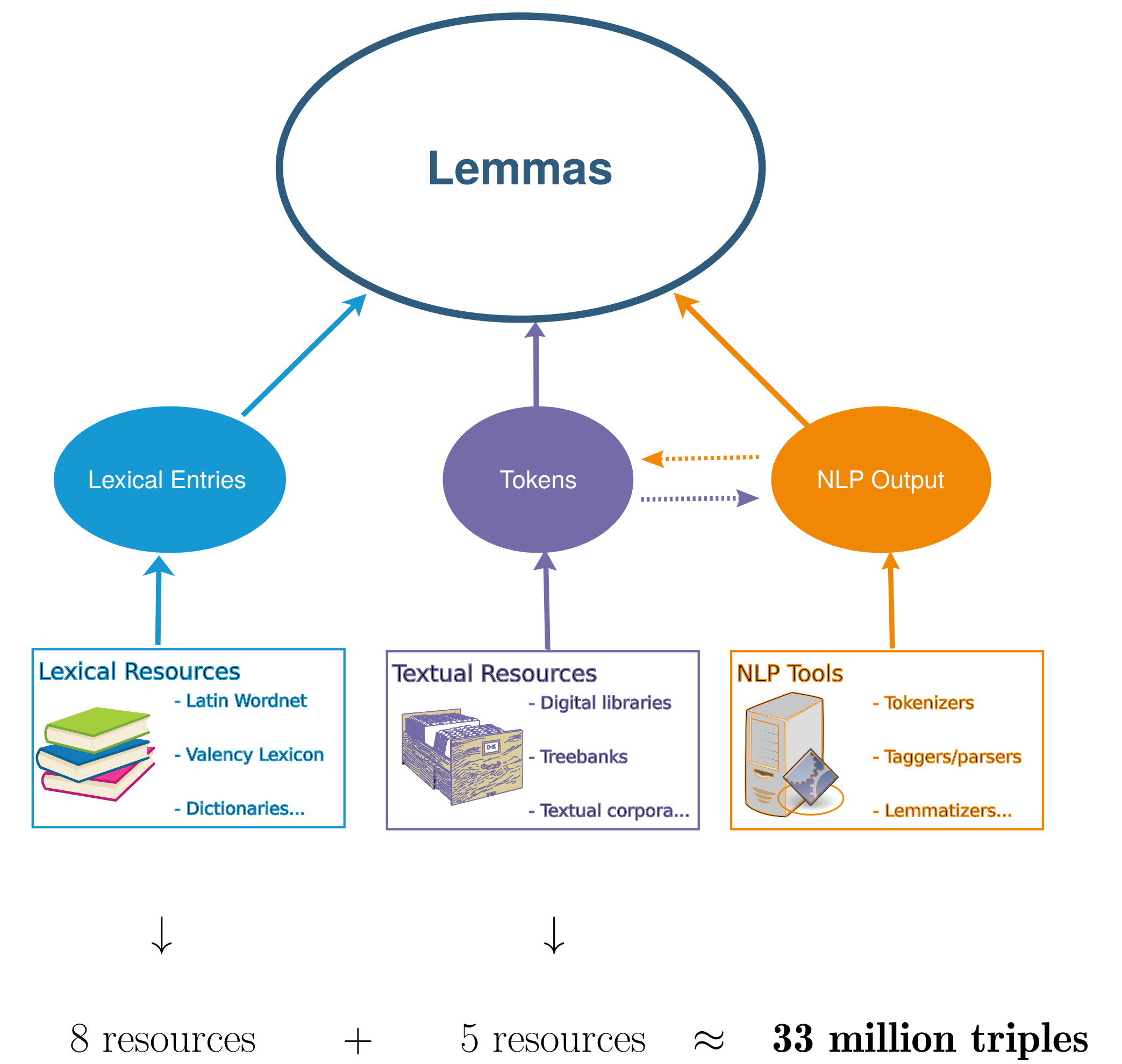
The architecture of the LiLa Knowledge Base is highly **portable** across languages and wants to become a **reference model** to address the issue of communication gaps between linguistic resources, exploiting its effort in harmonizing intralinguistically different standards.

Beyond Latin

LiLa also includes resources which branch into **related languages**, both **diachronically and synchronically**:

- **Proto-Italic** and **Proto-Indo-European** roots for all Latin lemmata
- a collection of **Ancient Greek** loanwords into the Latin lexicon
- a selection of Latin loanwords used in **Medieval Italian** by Dante Alighieri (XIII c.)

The LiLa Knowledge Base



Annotational Endeavor

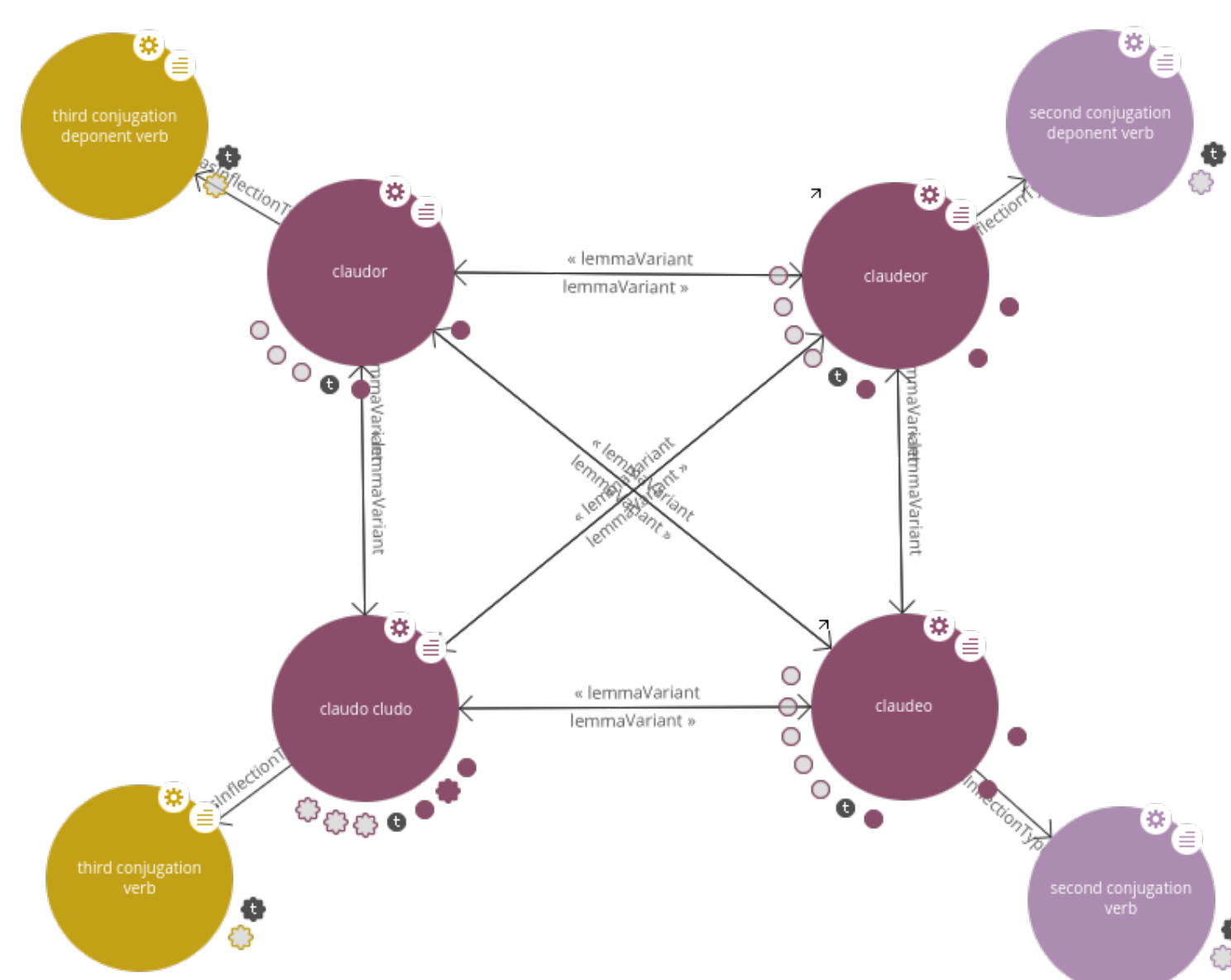
Within the LiLa project, several textual resources have been **annotated** and **interlinked** through the Lemma Bank, in particular:

- the *Index Thomisticus* treebank (Latin works by Thomas Aquinas, XIII c.), originally in the Prague Dependency Treebank format, has been **converted** into a UD version
- the *UDante* treebank (consisting of all Latin works by Dante Alighieri, XIII c.) has been annotated **from scratch** following the UD formalism

These efforts, in **interaction** with the LiLa Knowledge Base, have brought forth an ongoing effort of **harmonization** between Latin treebanks and the definition of more **comprehensive** and **typologically grounded** guidelines for Latin, besides enriching LiLa's **Lemma Bank**.

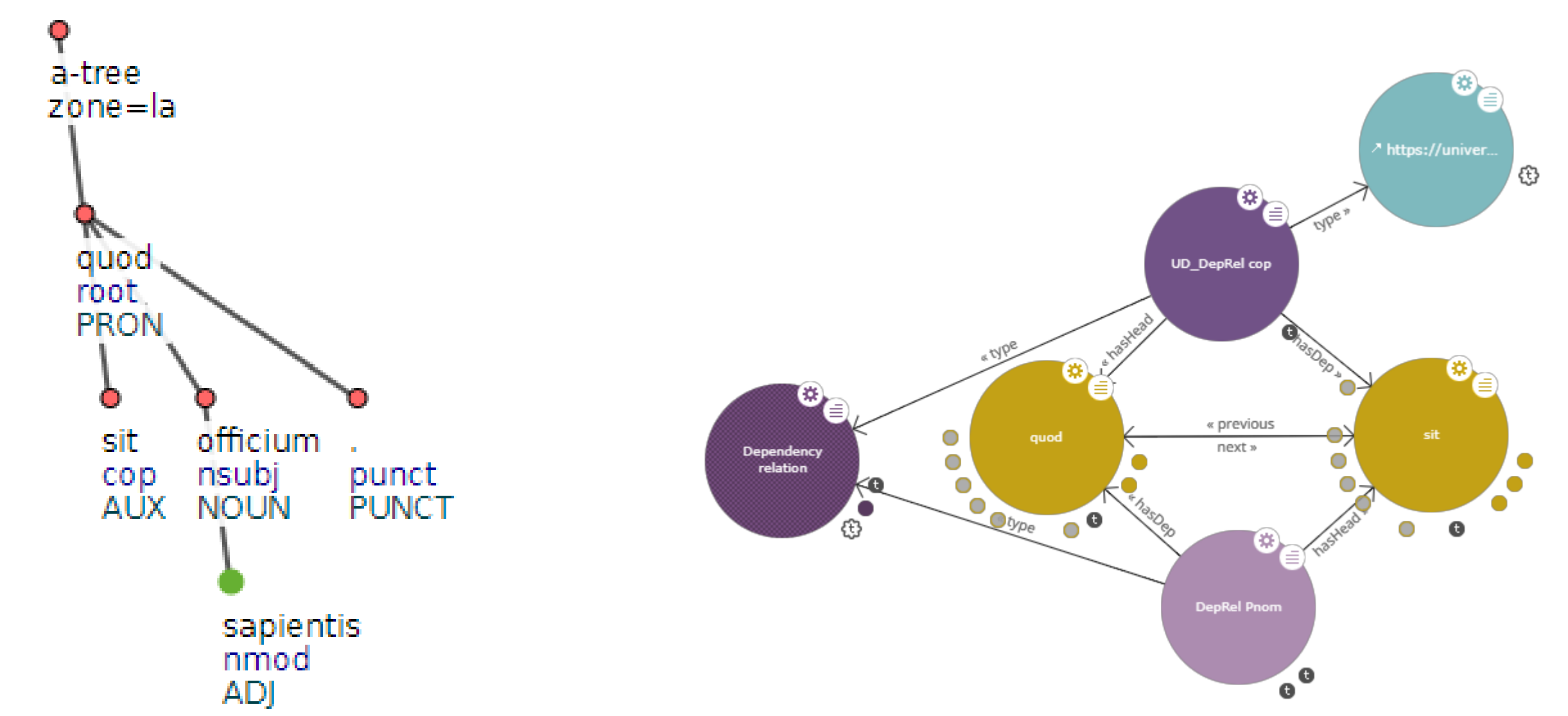
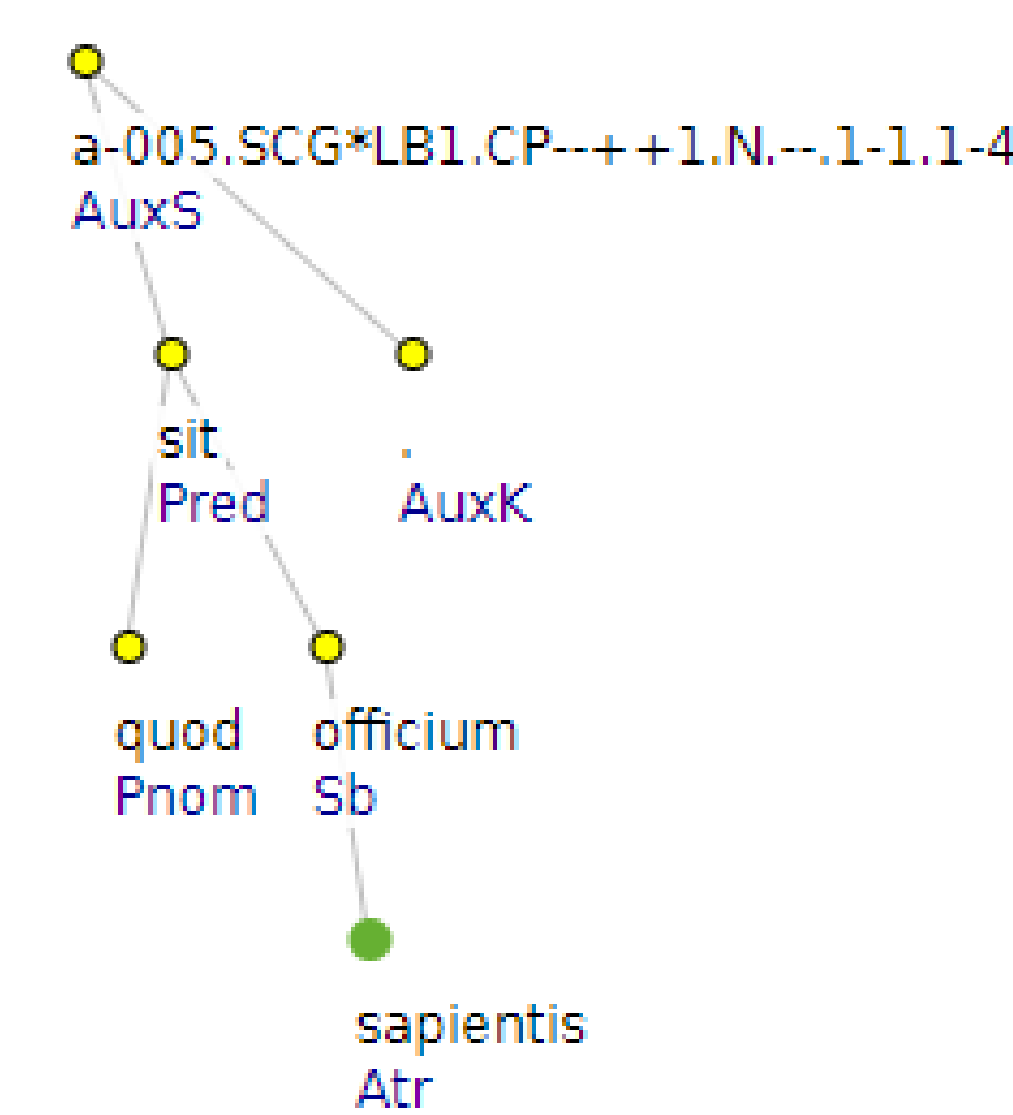
Lexical Variability

claudeo 'to limp' and its variants with different features



Syntactic Annotation: Bringing Different Standards Together

quod sit officium sapientis. (Index Thomisticus treebank, *Summa contra gentiles* lib. 1, cap. 1, n. 1)
'[What be] the office of the wise man.'



Interoperability between Resources

Looking for: all **tokens** in any **document** of the LASLA **corpus** representing a word with the same **base** of *lux* 'light' being a suffix-**derived** denominal adjective having positive **polarity**...
→ 5 occurrences of LUCULENTUS 'bright, splendid', e.g. *luculentum* in the Epidicus by Plautus (III-II c. BCE) | Try it yourself at <http://lila-erc.eu:8080/lila-lisp/>!



Token	Document
luculentam	Catilina
luculentam	Philippica Oratio
luculentam	Plinius Epistulae
luculenti	Philippica Oratio
luculentum	Epidicus

