



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

Interlinking Valency Frames and WordNet Synsets

in the LiLa Knowledge Base of Linguistic Resources for Latin

Francesco Mambrini, Marco Passarotti, Eleonora Litta and Giovanni Moretti

SEMANTiCS 2021 | Amsterdam | September 7, 2021



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.

Introduction

- Motivation

- The LiLa project

Valency and Synsets in LiLa

- “LOD-ifying” the Latin Vallex

- PreMOn and the Vallex extension

Conclusions

- ▶ learning Latin vocabulary is hard
 - ▶ most words are polysemic
 - ▶ verb constructions
 - ▶ idiomatic expressions (e.g. *rem publicam gerere* to govern)
- ▶ comprehensive bilingual lexicons (10k+ entries)

- ▶ learning Latin vocabulary is hard
 - ▶ most words are polysemic
 - ▶ verb constructions
 - ▶ idiomatic expressions (e.g. *rem publicam gerere* to govern)
- ▶ comprehensive bilingual lexicons (10k+ entries)
- ▶ ...**but there's more!**

Our approach

Semantic Web technologies for Latin lexicon



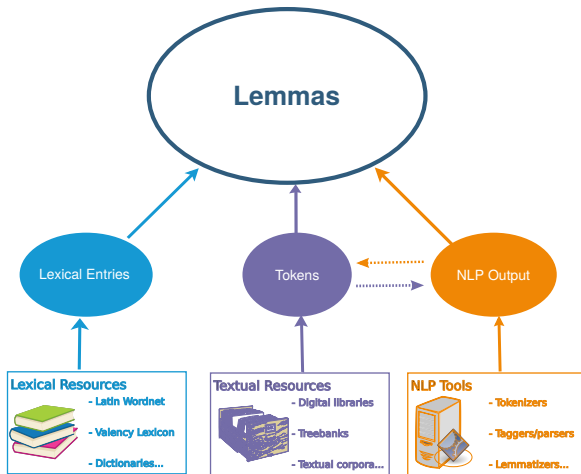
1. we **don't** need yet another dictionary
2. we want to make good use of what is available

Our approach

Semantic Web technologies for Latin lexicon

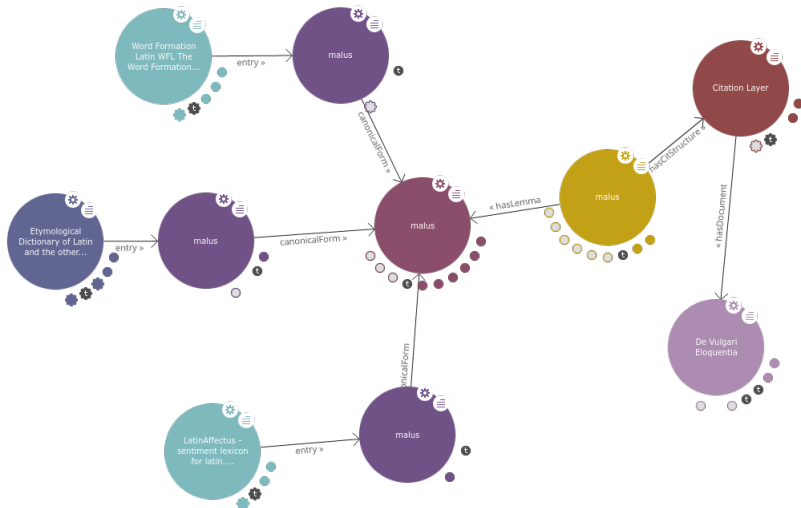


1. we **don't** need yet another dictionary
2. we want to make good use of what is available
3. and this is where **LOD helps!**



Linking through lemmas

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



▶ Corpora

- ✓ Index Thomisticus Treebank: ca. 450,000 tokens
- ✓ Dante Search (UDante treebank): ca. 46,000 token
- ✓ *Querolus sive Aulularia*: ca. 17,000 token
- PROIEL e LLCT treebank; CLaSSES; *Liber Abbaci* (Fibonacci)
- Computational Historical Semantics, LASLA e CroALA

▶ Lexical resources

- ✓ Word Formation Latin: ca. 46,000 entries
- ✓ Etymological dictionary: ca. 1,400 entries
- ✓ LatinAffectus (sentiment lexicon): ca. 2,300 entries
- ✓ IGVLTT (Greek loan words): ca. 1,800 entries
- ✓ Latin WordNet: ca. 1,400 entries (manually checked)
- ✓ Latin Vallex 2.0: ca. 1,000 entries (manually checked)
- Lewis & Short Latin-English dictionary

▶ NLP software

- ✓ LEMLAT (lemmatizer and morph analyzer): ca. 150,000 lemmas

▶ TOTAL: ca. 13 million triples

▶ Corpora

- ✓ Index Thomisticus Treebank: ca. 450,000 tokens
- ✓ Dante Search (UDante treebank): ca. 46,000 token
- ✓ *Querolus sive Aulularia*: ca. 17,000 token
- PROIEL e LLCT treebank; CLaSSES; *Liber Abbaci* (Fibonacci)
- Computational Historical Semantics, LASLA e CroALa

▶ Lexical resources

- ✓ Word Formation Latin: ca. 46,000 entries
- ✓ Etymological dictionary: ca. 1,400 entries
- ✓ LatinAffectus (sentiment lexicon): ca. 2,300 entries
- ✓ IGVLTT (Greek loan words): ca. 1,800 entries
- ✓ **Latin WordNet**: ca. 1,400 entries (manually checked)
- ✓ **Latin Vallex 2.0**: ca. 1,000 entries (manually checked)
- Lewis & Short Latin-English dictionary

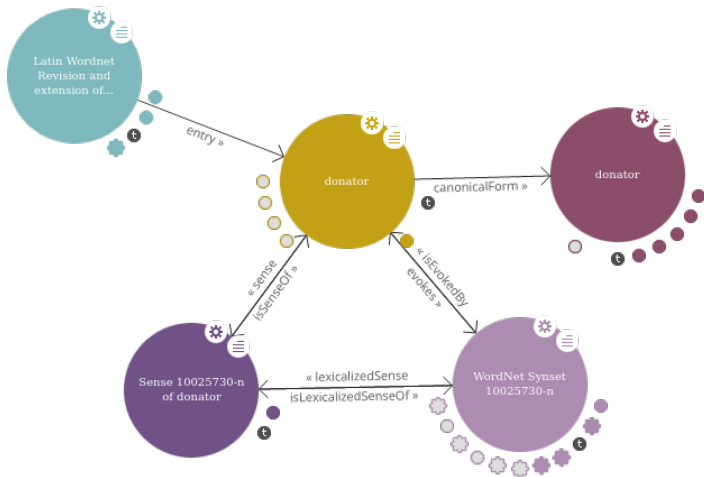
▶ NLP software

- ✓ LEMLAT (lemmatizer and morph analyzer): ca. 150,000 lemmas

▶ TOTAL: ca. 13 million triples

Latin WordNet Lemmas

donator: person who makes a gift of property (syn 10025730-n)



vallex 4.0

all

show filters ▾

search (2772 lexemes)



absolvovat
absorbovat
adresovat
akceptovat
aktivovat
aktivovat se
aktualizovat
analyzovat
angažovat
angažovat se
apelovat
aplikovat
argumentovat
asistovat
balit, balívat
bát se, bávat se
bavit, bavívat
bavit se, bavívat se
bdít, bdívat
běhat, běhávat
běžet
bit
bit se



apelovat^{impf}

① vyzývat; tlačit

frame ACT^{obl}₁ ADDR^{obl}_{k+3,na+4} PAT^{obl}_{k+3,proti+3,inf,aby,af,že}

example apeloval na své kolegy, aby práci dokončili včas; apelovat proti trestu smrti; apelovat k přijetí opatření; Apeloval na účastníky vyslechnout každého v duchu tolerance. (SYN) Obě strany nyní apelují k Evropské unii, aby pomohla. (SYN)

less of ① ^

class communication

control ADDR

recipr ACT-ADDR vzájemně na sebe apelovali, aby ...

reflex ACT-ADDR Proti Slavii jsme byli skoro bez šance, apeluji na sebe i další ofenzivní hráče, abychom něco předvedli.

diat deagent často se na něj apelovalo, aby byl pilnější

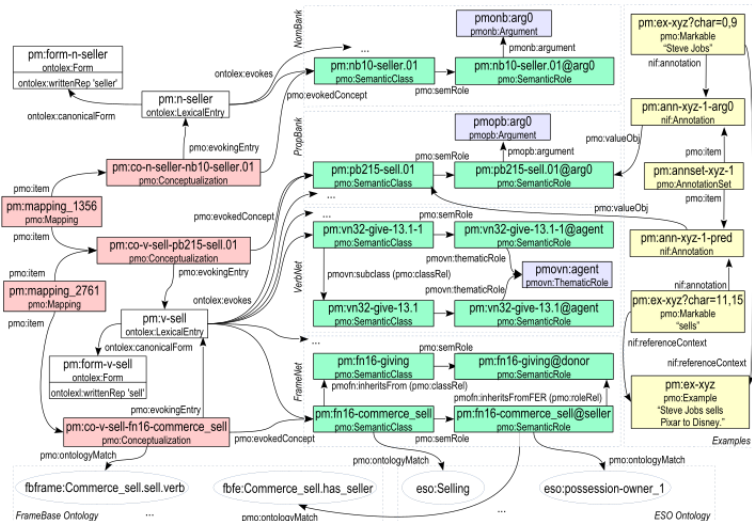
② klást důraz; dovolávat se

frame ACT^{obl}₁ PAT^{obl}_{na+4}

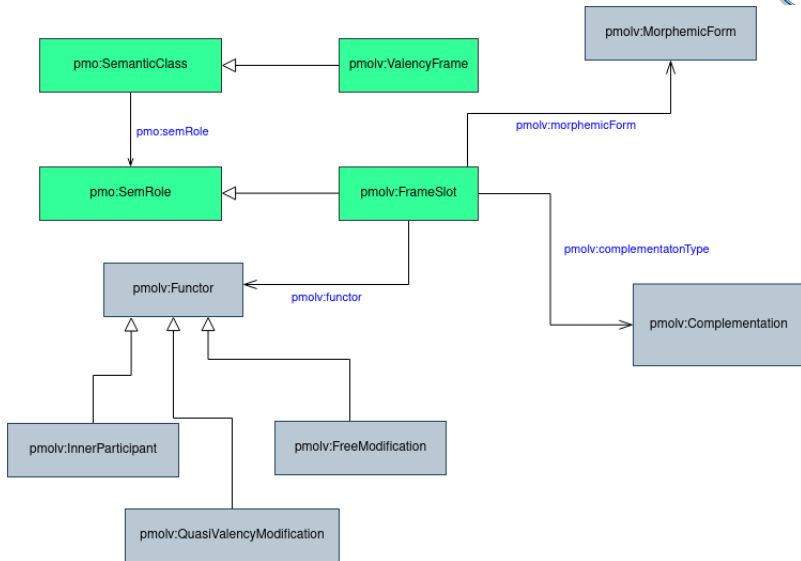
example apeloval na morálku

less of ② ^

diat deagent v jeho rodině se stále apeluje na morálku

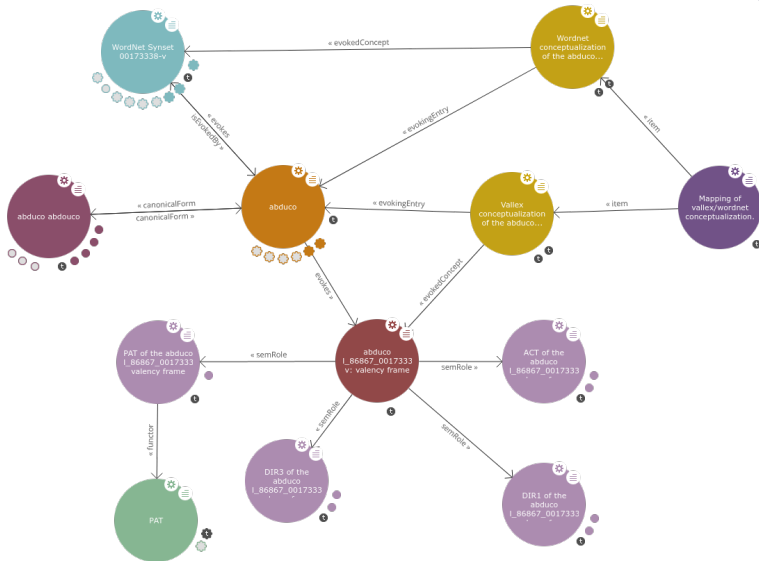


The Vallex Extension



Bringing all together

Vallex, WordNet, etc...



Formation, senses and syntax

From *pes* (foot.n) to verbs via prefix



```
PREFIX ontalex: <http://www.w3.org/ns/lemon/ontalex#>
PREFIX vartrans: <http://www.w3.org/ns/lemon/vartrans#>
PREFIX wfl: <http://lila-erc.eu/ontologies/lila/wfl/>
```

```
select ?deriv ?rellab ?rel where {
  ?lex ontalex:canonicalForm
    <http://lila-erc.eu/data/id/lemma/117499>.
  ?rel vartrans:source ?lex ;
    vartrans:target ?deriv ;
    rdfs:label ?rellab ;
    wfl:hasWordFormationRule ?rule .
  ?rule a
    <http://lila-erc.eu/ontologies/lila/wfl/Prefixation/NounToVerb>
}
```

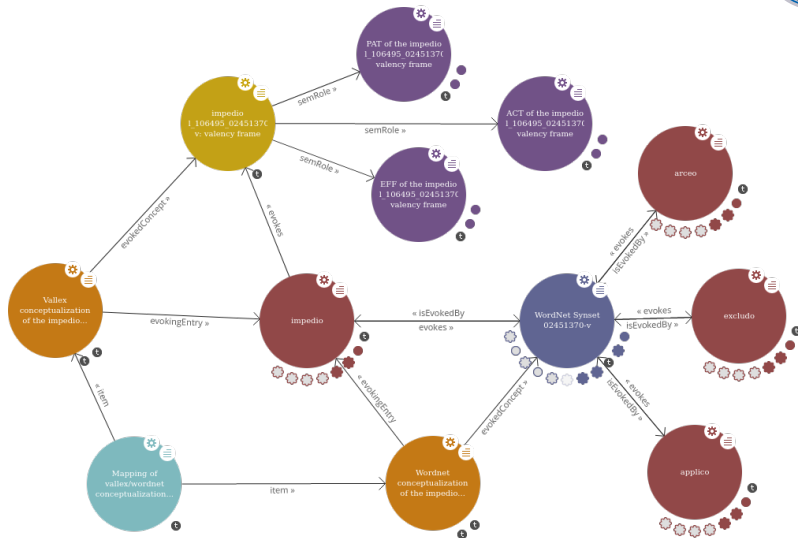
Try it at: <https://lila-erc.eu/sparql/>

- ▶ *expedio* (23 senses, mostly 2 arg slots):
 - ▶ make ready or suitable or equip in advance for a particular purpose or for some use, event, etc
 - ▶ synonyms: *apto, facio, constituo, comparo...*
 - ▶ free or remove obstruction from
 - ▶ remove the outer cover or wrapping of
 - ▶ release from something that holds fast, connects, or entangles
 - ▶ ...
- ▶ *impedio* (3 senses, 2 or 3 arg slots):
 - ▶ prevent the progress or free movement of
 - ▶ be a hindrance or obstacle to
 - ▶ hinder or prevent the progress or accomplishment of

- ▶ *expedio* (23 senses, mostly 2 arg slots):
 - ▶ make ready or suitable or equip in advance for a particular purpose or for some use, event, etc
 - ▶ synonyms: *apto, facio, constituo, comparo...*
 - ▶ free or remove obstruction from
 - ▶ remove the outer cover or wrapping of
 - ▶ release from something that holds fast, connects, or entangles
 - ▶ ...
- ▶ *impedio* (3 senses, 2 or 3 arg slots):
 - ▶ prevent the progress or free movement of
 - ▶ be a hindrance or obstacle to
 - ▶ hinder or prevent the progress or accomplishment of

Impedio, 3-slot sense

Mapped to synset: "impede, hinder – be a hindrance or obstacle to"



Summing up

- ▶ new resources for syntax/semantics of Latin
- ▶ an ontology to map predicate structures in Latin to the LWN
- ▶ a LOD environment: all resources are **interoperable**
- ▶ LiLa is a powerful tool to study Latin vocabulary!

Future work:

- ▶ more Vallex entries!
- ▶ more LWN entries!
- ▶ word-sense disambiguation on Latin texts?

Thanks!

Get in touch



Francesco Mambrini¹, Eleonora Litta¹, Marco Passarotti¹ and Giovanni Moretti¹

¹ Università Cattolica del Sacro Cuore, Milan



info@lila-erc.eu



[@ERC_LiLa](https://twitter.com/ERC_LiLa)



<https://github.com/CIRCSE>



<https://lila-erc.eu>



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.