

Extending and Using a Sentiment Lexicon for Latin in a Linked Data Framework

Rachele Sprugnoli, Marco Passarotti, Marinella Testori, Giovanni Moretti
rachele.sprugnoli@unicatt.it

*Workshop on Sentiment Analysis and Linguistic Linked Data (SALLD-1)
September 1, 2021*



UNIVERSITÀ
CATTOLICA
del Sacro Cuore



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.

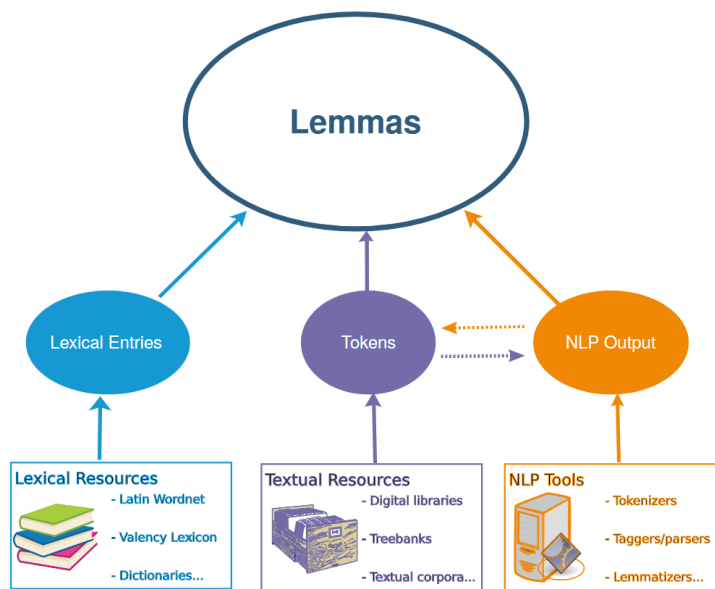
Contributions:

- New **version** of *LatinAffectus*
- **Linking** of *LatinAffectus* to our Knowledge Base
- **Use case** on Dante Alighieri's Latin works
- Modeling of a **sense-based** sentiment lexicon

Resources **available** online:

- Github repository:
https://github.com/CIRCSE/Latin_Sentiment_Lexicons
- SPARQL endpoint: <https://lila-erc.eu/sparql/>

ERC Project: “LiLa: Linking Latin” <https://lila-erc.eu/>



To **provide** the community with a larger lexicon

- Polarity lexicons are fundamental resources for developing Sentiment Analysis systems

To **enrich** our Knowledge Base

- We increase the number of lemmas with polarity information

To **enhance** the interoperability

- The richer and more connected the resources, the more analyses can be done

List of nouns and adjectives associated to their out-of-context sentiment score: from -1 (very negative) to +1 (very positive)

VERSION 1

- Gold Standard: manually created by 2 Latin language and culture experts + 1 supervisor
- Silver Standard: automatically created by deriving new entries from the Gold Standard

[Sprugnoli, R., Passarotti, M., Corbetta, D., & Peverelli, A. (2020). *Odi et Amo. Creating, Evaluating and Extending Sentiment Lexicons for Latin*. In Proceedings of the 12th Language Resources and Evaluation Conference.]

List of nouns and adjectives associated to their out-of-context sentiment score: from -1 (very negative) to +1 (very positive)

EXTENSION

- Gold Standard: manual revision of a list of 816 lemma-sentiment pairs automatically induced from neural word embeddings
- Silver Standard: automatically created by deriving new entries from the Gold Standard

PHASE 1

- assignment of sentiment scores to 100 lemmas by 3 Latin experts (A1, A2, A3) working in parallel
- six-value classification: 1 (fully positive), 0.5 (somewhat positive), 0 (neutral), -0.5 (somewhat negative), -1 (fully negative), **2 (ambiguous)**
- IAA in terms of Cohen's k

	6 CLASSES	4 CLASSES
A1-A2	0.66	0.71
A1-A3	0.59	0.85

PHASE 2

- revision of the scores assigned by A1 done by a team of 4 Latin experts: 115 out of 716 lemmas (16%) received a different score

PHASE 3

- reconciliation
- removal of lemmas marked as ambiguous

	ADJ	NOUN	+1	+0.5	0	-0.5	-1	ALL
Gold Standard	307	428	112	71	194	112	246	735

PHASE 4

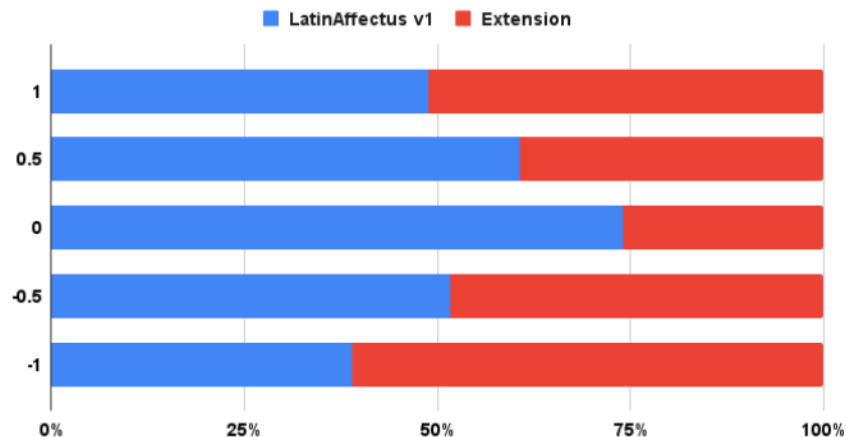
- new lemmas added exploiting derivational, synonym and antonym relations + graphical variants of lemmas present in the Gold Standard
- original polarity scores are propagated or reversed onto the newly derived lemmas

	ADJ	NOUN	+1	+0.5	0	-0.5	-1	ALL
Silver Standard	348	604	121	110	262	150	309	952

LatinAffectus v2



	ADJ	NOUN	TOTAL
LatinAffectus v1	966	1,471	2,437
Extension	655	1,032	1,687
LatinAffectus v2	1,621	2,503	4,124



sanabilis

http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalEntry/lemma_123307

AN ENTITY OF TYPE: **LexicalEntry**

rdfs:label	sanabilis
rdf:type	ontolex:LexicalEntry
ontolex:canonicalForm	< http://lila-erc.eu/data/id/lemma/123307 > ↳ sanabilis
ontolex:sense	< http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalSense/lemma_123307 > ↳ Prior sense of sanabilis

Prior sense of sanabilis

http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalSense/lemma_123307

AN ENTITY OF TYPE: **LexicalSense**

rdfs:label	Prior sense of sanabilis
< http://www.gsi.dit.upm.es/ontologies/marl/ns#polarityValue >	1.0
rdf:type	ontolex:LexicalSense
< http://www.gsi.dit.upm.es/ontologies/marl/ns#hasPolarity >	< http://www.gsi.dit.upm.es/ontologies/marl/ns#Positive >

[Sprugnoli, R., Mambrini, F., Moretti, G., & Passarotti, M. (2020). *Towards the Modeling of Polarity in a Latin Knowledge Base*. In Proceedings of the Third Workshop on Humanities in the Semantic Web (WHiSe 2020)@ ESWC.]

sanabilis

http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalEntry/lemma_123307

AN ENTITY OF TYPE: **LexicalEntry**

rdfs:label sanabilis

rdf:type ontolex:LexicalEntry

ontolex:canonicalForm <<http://lila-erc.eu/data/id/lemma/123307>>
↳ sanabilis

ontolex:sense <http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalSense/lemma_123307>
↳ Prior sense of sanabilis

Prior sense of sanabilis

http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalSense/lemma_123307

AN ENTITY OF TYPE: **LexicalSense**

rdfs:label Prior sense of sanabilis

<<http://www.gsi.dit.upm.es/ontologies/marl/ns#polarityValue>> 1.0

rdf:type ontolex:LexicalSense

<<http://www.gsi.dit.upm.es/ontologies/marl/ns#hasPolarity>> <<http://www.gsi.dit.upm.es/ontologies/marl/ns#Positive>>

[Sprugnoli, R., Mambrini, F., Moretti, G., & Passarotti, M. (2020). *Towards the Modeling of Polarity in a Latin Knowledge Base*. In Proceedings of the Third Workshop on Humanities in the Semantic Web (WHiSe 2020)@ ESWC.]

sanabilis

http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalEntry/lemma_123307

AN ENTITY OF TYPE: **LexicalEntry**

rdfs:label	sanabilis
rdf:type	ontolex:LexicalEntry
ontolex:canonicalForm	< http://lila-erc.eu/data/id/lemma/123307 > ↳ sanabilis
ontolex:sense	< http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalSense/lemma_123307 > ↳ Prior sense of sanabilis

Prior sense of sanabilis

http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalSense/lemma_123307

AN ENTITY OF TYPE: **LexicalSense**

rdfs:label	Prior sense of sanabilis
< http://www.gsi.dit.upm.es/ontologies/marl/ns#polarityValue >	1.0
rdf:type	ontolex:LexicalSense
< http://www.gsi.dit.upm.es/ontologies/marl/ns#hasPolarity >	< http://www.gsi.dit.upm.es/ontologies/marl/ns#Positive >

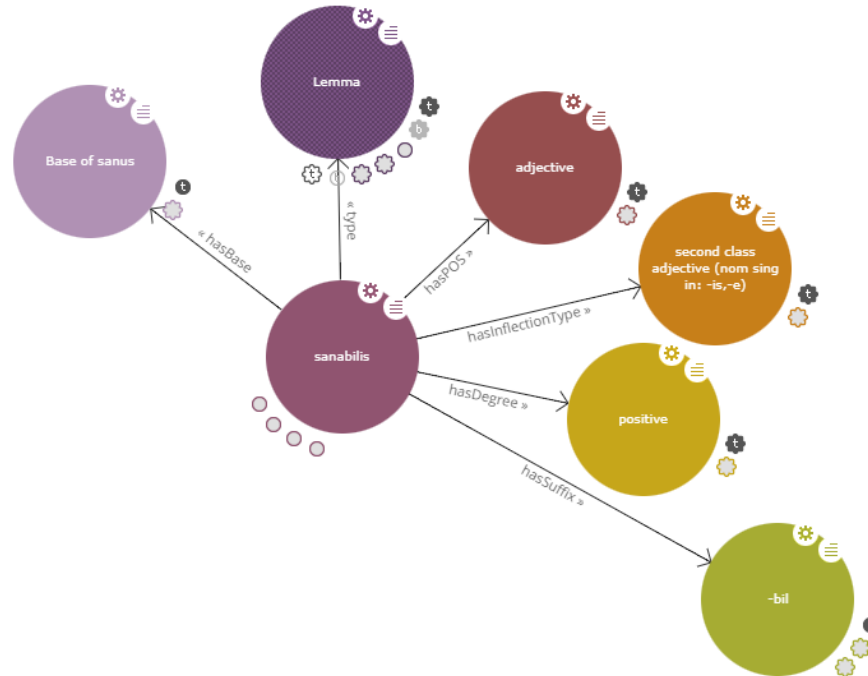
[Sprugnoli, R., Mambrini, F., Moretti, G., & Passarotti, M. (2020). *Towards the Modeling of Polarity in a Latin Knowledge Base*. In Proceedings of the Third Workshop on Humanities in the Semantic Web (WHiSe 2020)@ ESWC.]

Semi-automatic linking of the the lexical entries coming from the Extension to their corresponding lemmas in the LiLa KB:

- 1,586 automatically matched (94%)
- 61 pointing to more than one lemma
- 40 to be added

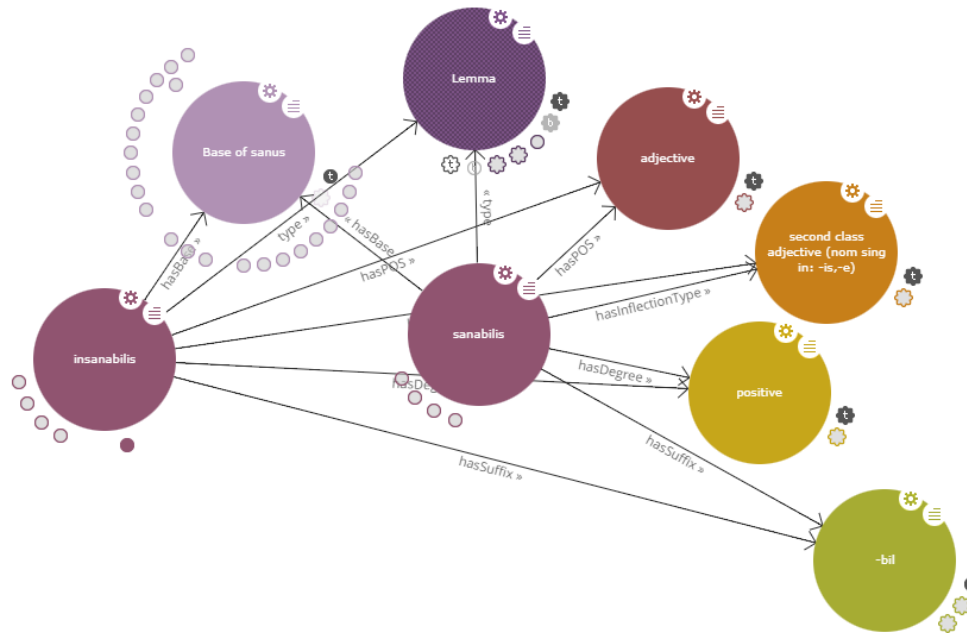
Example

https://lila-erc.eu/lodlive/app_en.html?http://lila-erc.eu/data/id/lemma/123307



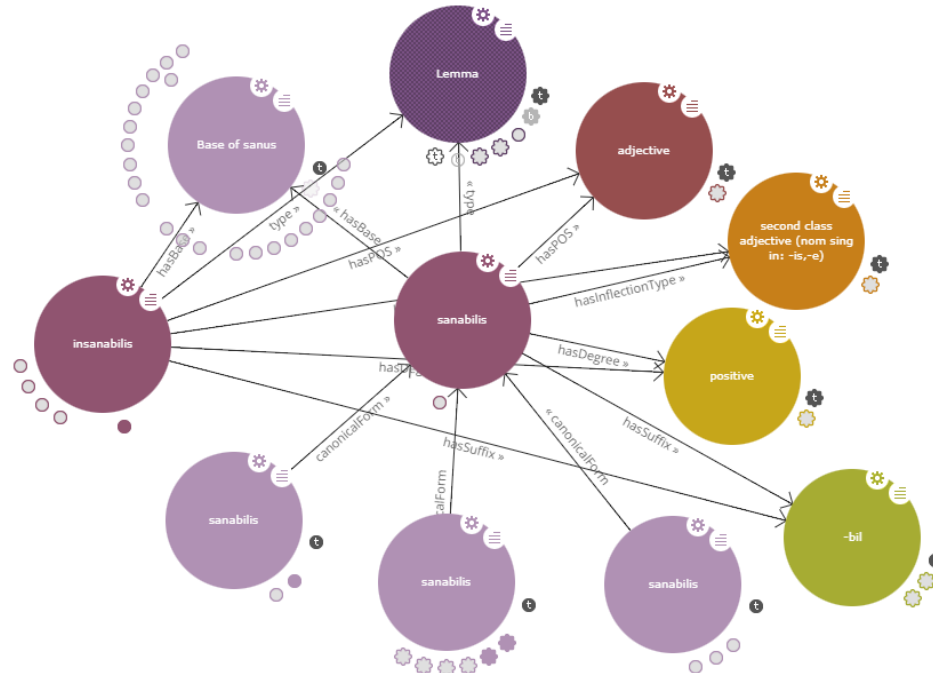
Example

https://lila-erc.eu/lodlive/app_en.html?http://lila-erc.eu/data/id/lemma/123307



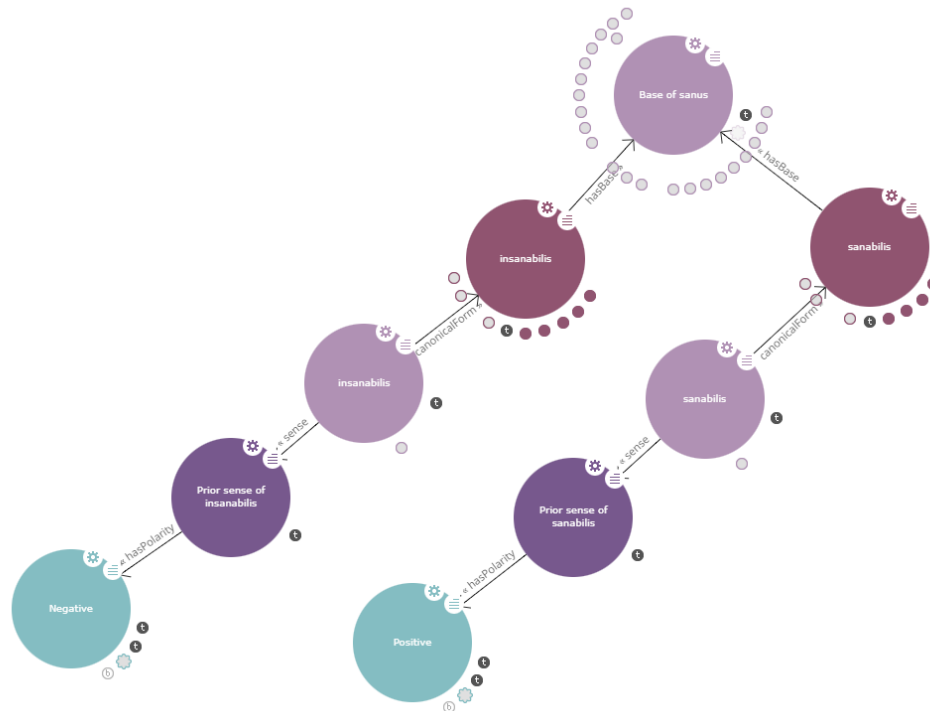
Example

https://lila-erc.eu/lodlive/app_en.html?http://lila-erc.eu/data/id/lemma/123307



Example

https://lila-erc.eu/lodlive/app_en.html?http://lila-erc.eu/data/id/lemma/123307



How the **interoperability** between the resources connected in LiLa, and particularly LatinAffectus, can support research in the Humanities?

UDante:

- dependency treebank of Dante's Latin works
- five works, 55,697 tokens
- literary texts: letters, treatises, poetry
- morphological and syntactic annotation: only UD style
- <https://lila-erc.eu/data/corpora/DanteSearch/id/corpus>

How the attitude of the author changes depending on the topic of the letter in the *Epistole*?



SPARQL QUERY to perform a lexicon-based Sentiment Analysis at letter and paragraph level - <https://github.com/CIRCSE/SPARQL-queries>

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
Abs.	33.1	8.6	28.6	3.9	34.1	-13.6	12.7	13	19.6	23	14.1	8	197.1
Norm.	0.25	0.12	0.24	0.07	0.16	-0.04	0.05	0.25	0.49	0.43	0.04	0.09	0.24

Letters with a **more positive** polarity are the ones written in the name of or on behalf of others:

- Letter I (0.25), Letter VIII (0.25), Letter IX (0.49), Letter X (0.43)

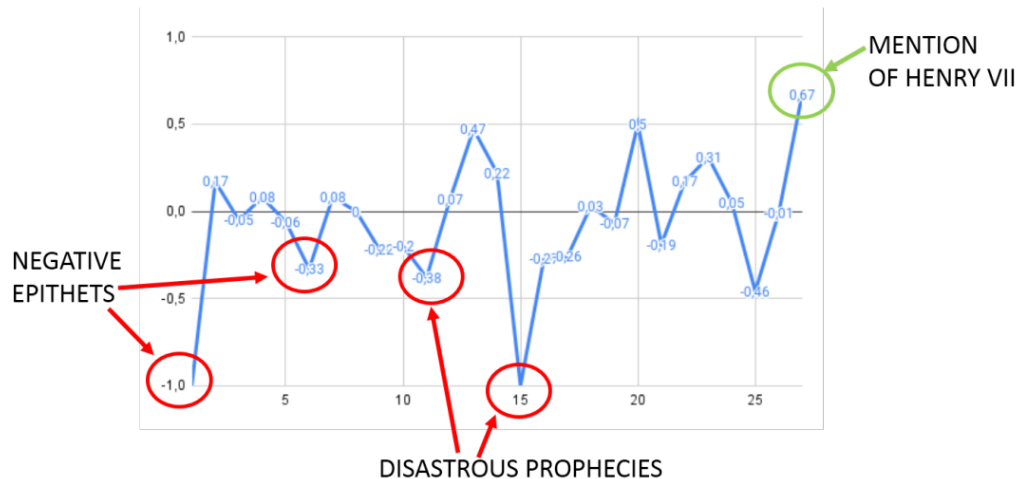
Example: Letter X, paragraph 1, score 0.83

EN: *“The illustrious renown of your Magnificence, which wakeful Fame spreads abroad as she flies, affects divers persons in divers ways, so that some it uplifts with the hope of good fortune, while others it casts down with the dread of destruction. The report whereof, overtopping all deeds of recent times, I erstwhile did deem extravagant, as going beyond the appearance of truth.”*

“Henry letters”, written to further the campaign of the emperor Henry VII in Italy, are **less positive**:

- Letter V (0.16), **Letter VI (-0.04)**, Letter VII (0.05)

Paragraph 1, score -1
EN: “Dante Alighieri, a Florentine undeservedly in exile, to the most iniquitous Florentines within the city.”



Towards a sense-based lexicon



We added sentiment information to a subset of Minozzi's Latin WordNet (LWN) manually revised and mapped onto version 3.0 of Princeton WordNet:

- 6,682 lemma-synset pairs
- LWN modeled as the Princeton WordNet RDF
- sentiment scores taken from **SentiWordNet 3.0**

Towards a sense-based lexicon



sanabilis

http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalEntry/L_123307

rdfs:label	sanabilis
rdf:type	ontolex:LexicalEntry
ontolex:canonicalForm	< http://lila-erc.eu/data/id/lemma/123307 > ↳ sanabilis
ontolex:sense	< http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalSense/L_123307_00994410-a > ↳ Sense 00994410-a of sanabilis ----- < http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalSense/L_123307_01962631-a > ↳ Sense 01962631-a of sanabilis
ontolex:evokes	< http://wordnet-rdf.princeton.edu/wn30/00994410-a > ----- < http://wordnet-rdf.princeton.edu/wn30/01962631-a >

Towards a sense-based lexicon



Sense 01962631-a of sanabilis

http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalSense/L_123307_01962631-a

rdfs:label	Sense 01962631-a of sanabilis
<http://www.gsi.dit.upm.es/ontologies/marl/ns#polarityValue>	0.0
<http://www.w3.org/2003/06/sw-vocab-status/ns#term_status>	reviewed
rdfs:type	ontolex:LexicalSense
<http://www.gsi.dit.upm.es/ontologies/marl/ns#hasPolarity>	<http://www.gsi.dit.upm.es/ontologies/marl/ns#Neutral>
ontolex:isLexicalizedSenseOf	<http://wordnet-rdf.princeton.edu/wn30/01962631-a>
ontolex:isSenseOf	<http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalEntry/L_123307> ↳ sanabilis

“capable of being remedied or redressed”

Sense 00994410-a of sanabilis

http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalSense/L_123307_00994410-a

rdfs:label	Sense 00994410-a of sanabilis
<http://www.gsi.dit.upm.es/ontologies/marl/ns#polarityValue>	0.125
<http://www.w3.org/2003/06/sw-vocab-status/ns#term_status>	reviewed
rdfs:type	ontolex:LexicalSense
<http://www.gsi.dit.upm.es/ontologies/marl/ns#hasPolarity>	<http://www.gsi.dit.upm.es/ontologies/marl/ns#Positive>
ontolex:isLexicalizedSenseOf	<http://wordnet-rdf.princeton.edu/wn30/00994410-a>
ontolex:isSenseOf	<http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalEntry/L_123307> ↳ sanabilis

“curing or healing is possible”

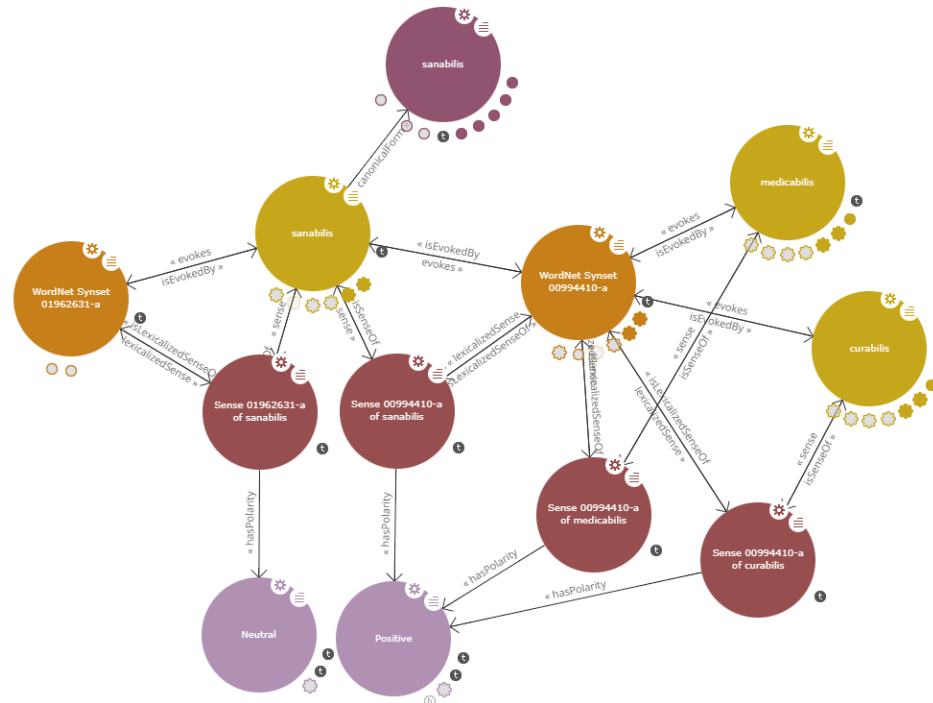
Example

https://lila-erc.eu/lodlive/app_en.html?http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalEntry/_123307



Example

https://lila-erc.eu/lodlive/app_en.html?http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalEntry/_123307



- To extend the coverage of LatinAffectus
WORK IN PROGRESS: manual score assignment to the most frequent lemmas (nouns and adjectives) of “Opera Latina”
- To refine the quality of LatinAffectus
WORK IN PROGRESS: manual revision of lemma-sentiment pairs in the Silver Standard
- To release a sense-based sentiment lexicon with high accuracy
FUTURE WORK: manual revision of the scores automatically assigned by SentiWordNet 3.0

Thanks! Get in touch...




Università Cattolica del Sacro Cuore
CIRCSE Research Centre

 <https://lila-erc.eu>

 info@lila-erc.eu / rachele.sprugnoli@unicatt.it

 @ERC_LiLa / @RSprugnoli

 Largo Gemelli 1, 20123 Milan, Italy