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

Rachele Sprugnoli, Marco Passarotti, Marinella Testori and Giovanni Moretti

CIRCSE Research Centre, Università Cattolica del Sacro Cuore, Milan, Italy

Abstract

In this paper we present the methodology followed to extend a Latin sentiment lexicon (called LatinAffectus), the process of inclusion of the lexicon in a knowledge base of interoperable linguistic resources for Latin and one use case performed on the treebank of Dante Alighieri's Latin works annotated following the Universal Dependencies guidelines. In addition, we report on our first attempt at linking the polarity scores of SentiWordNet 3.0 to a manually revised version of Latin WordNet.

Keywords

Linguistic Linked Data, Sentiment Analysis, Polarity Lexicon, Latin

1. Introduction

The main applications of resources and tools for Sentiment Analysis typically fall within the scope of fields like customer experience and social media monitoring; however, there is an increasing interest in extending their range to applications that can support research in the Humanities. Such applications can potentially impact a large and diverse community made of historians, philologists, archaeologists and literary scholars. Our work focuses on texts written in an Classical language, i.e. Latin: although there is a substantial growth of the area dedicated to building and using datasets and Natural Language Processing (NLP) tools for ancient and historical languages [1], linguistic resources to perform Sentiment Analysis on such languages are still rare. However, this type of resources are essential for accessing and understanding the Classical tradition.

This paper presents the second version of LatinAffectus, a prior polarity lexicon of Latin nouns and adjectives developed within the ERC project LiLa: Linking Latin.[1] In particular, we describe the methodology adopted to obtain a new set of high-quality lemma-sentiment pairs, the linking

This paper is the result of the collaboration between the four authors. For the specific concerns of the Italian academic attribution system, Rachele Sprugnoli is responsible for Sections 2, 3, 4, 6 and Marco Passarotti is responsible for Sections 1 and 7. Section 5 was jointly written by Rachele Sprugnoli and Marco Passarotti. Marinella Testori was the main annotator of the resource presented in the paper, while Giovanni Moretti developed the model and designed the SPARQL queries.

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

rachele.sprugnoli@unicatt.it (R. Sprugnoli); marco.passarotti@unicatt.it (M. Passarotti); marinella.testori@unicatt.it (M. Testori); giovanni.moretti@unicatt.it (G. Moretti)

0000-0001-6861-5595 (R. Sprugnoli); 0000-0002-9806-7187 (M. Passarotti); 0000-0001-5466-3262 (M. Testori); 0000-0001-7188-8172 (G. Moretti)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

[1] https://lila-erc.eu/

of the lexicon to the Linked Data Knowledge Base (KB) of interoperable linguistic resources for Latin developed within LiLa and an example of application using UDante [2], a treebank containing all of Dante Alighieri's Latin works annotated following the Universal Dependencies guidelines [3]. In addition, we report on our first attempt at modeling a sense-based sentiment lexicon built upon a manually revised version of Latin WordNet [4].

2. Related Work

Liu [5] defines Sentiment Analysis as 'the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language'. The interest in this type of research is always growing also because it is considered a valid tool in the business, communication and social fields. In fact, there are many applications that monitor the opinion of customers towards a service or product [6], that study the attitude of users on social media [7] and that try to recognize abusive language and hate speech online [8, 9]. During the last year, due to the outbreak of the COVID-19 pandemic, studies have increased also in the field of health communication and risk prevention [10, 11]. Beyond these works based on current events and focused on the commercial and social domains, there are also works that adopt and adapt methods and techniques of Sentiment Analysis to perform inquiries specific of the Humanities. Linguistic resources and automatic models have been thus developed to deal with the peculiarities of, among others, historical and literary texts [12, 13, 14]. Texts written in Classical languages are especially challenging due, for example, to the lack of native speakers, the limited amount of available data and the presence of unusual textual genres for the Sentiment Analysis task, such as philosophical or documentary texts.

Sentiment Lexicons are essential resources for the development of Sentiment Analysis systems: they are lists of words (forms or lemmas) associated to scores expressing their prior polarity, that is their sentiment orientation regardless of the context of use. Lexicons are often produced (semi-)automatically using, for example, crowdsourcing, corpus-based or corpora-based methods [15]. Such approaches, initially designed for modern languages, have also been tested on ancient languages, as described for Classical Chinese [16] and for Latin [17, 18]. In our previous work [19], we have instead created a prior polarity lexicon following a manual procedure with the aim of obtaining a high quality resource that we called LatinAffectus. We then linked LatinAffectus to the the LiLa KB making it the first sentiment lexicon for Latin to be published in the Linked Data framework [20]. The modeling of LatinAffectus took as references previous efforts in developing formal representation of linguistic resources and services for Sentiment Analysis such as the Marl ontology[2] and the model proposed in the Eurosentiment Project [21]. In this paper we extend our previous work following the Linguistic Linked Data paradigm and underscoring the importance of an interdisciplinary approach combining computational linguistics, semantic web and humanities practices.

[2] http://www.gsi.dit.upm.es/ontologies/marl/1.1/

3. Extending LatinAffectus

LatinAffectus is a polarity lexicon in which a set of Latin nouns and adjectives are associated to their sentiment orientation regardless of the context of use. Its first version [20] is extended with a new Gold Standard (GS), made of 735 lemmas manually annotated and revised by experts in Latin language and culture, and with a new Silver Standard (SS), including 952 lemmas automatically derived from the GS [19].

The GS is built by revising a list of lemma-sentiment pairs automatically induced from neural word embeddings [22]. One annotator (A1) assigned a sentiment score to each lemma using a six-value classification: 1 (fully positive), 0.5 (somewhat positive), 0 (neutral), -0.5 (somewhat negative), -1 (fully negative), 2 (ambiguous). Two further annotators (A2 and A3) assigned a score to 50 randomly chosen lemmas, thus allowing to calculate the inter-annotator agreement (IAA) on a subset of the GS.

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

Table 1
Results of the IAA in terms of Cohen's *k*.

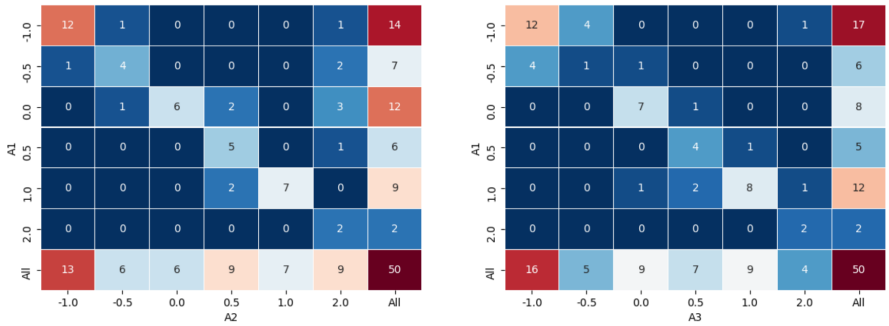

Figure 1: Confusion matrix among annotators (A1 *versus* A2 on the left and A1 *versus* A3 on the right) on score assignment using the six-value classification.

Table 1 reports the results of the IAA considering both the six-value classification and a four-value classification, in which 1 and 0.5 are merged into a unique positive class and -0.5 and -1 are converged under the same negative class. We register a Cohen's *k* higher than that recorded in the first version of LatinAffectus, in which the value of *k* had never exceeded 0.6, not even reducing the classes from 6 to 4. Confusion matrices displayed in Figure 1 show that the greatest differences between A1 and A2 are recorded on the annotation of the ambiguous lemmas, while A1 and A3 differ above all on the intermediate scores; in fact, considering only 4 classes, the IAA between them increases a lot (+26 percentage points). An example of disagreement between A1 and A2 is *stigma*: A1 assigned a somewhat negative score to it (-0.5) by referring to the meanings 'mark on the skin usually impressed upon slave' and 'mark of disgrace', whereas A2 noted a diachronic ambiguity given that, in the Christian era, *stigma*

refers to the mark borne by Christ as result of the Passion and also to the symbol in shape of the Cross, therefore with no negative meaning at all. Disagreements between A1 and A3 were instead recorded for lemmas such as *sanabilis* 'curable/curative' and *sterilitas* 'sterility' to which one annotator assigned intermediate values (0.5 to the former and -0.5 to the latter) and the other assigned totally positive or totally negative values (1 and -1 respectively).

The rest of the lemma-sentiment pairs assigned by A1 were checked by a team of 4 Latin experts who either confirmed the scores or reported any disagreement with respect to the choices made by A1. At the end of this process, all the discrepancies in the annotation were discussed and reconciled. More specifically, 115 lemmas were concerned and their polarity reviewed. After the reconciliation, we composed the final GS by removing the lemmas marked as ambiguous because, due to their semantic and/or diachronic ambiguity, they cannot have a unique a priori polarity score (see the example of *stigma* discussed in the previous paragraph).

	ADJ	NOUN	+1	+0.5	0	-0.5	-1	ALL
Gold Standard	307	428	112	71	194	112	246	735
Silver Standard	348	604	121	110	262	150	309	952
ALL	655	1032	233	181	456	262	555	1687

Table 2
Final composition of the extension of LatinAffectus.

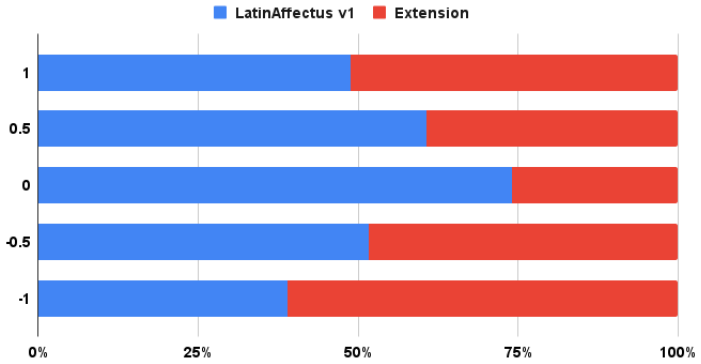

Figure 2: Composition of LatinAffectus v2: the first version of LatinAffectus (blue) merged with the extension presented in this paper (red).

Starting from this new GS, we generated a SS by deriving new entries through synonym, antonym and derivational relations with the lemmas in the GS, as well as by adding different written representations of lemmas present in the GS. Polarity scores in the GS were reversed for antonyms and for lemmas derived through the negative prefix *in-*, whereas they were preserved in all the other cases. For example, starting from the lemma *disertus* 'eloquent' with a fully positive polarity (+1), we derived *facundus* as a synonym, then *facundia*, *facunditas*, *infacundus* using morphological derivational relations: the last lemma received a reversed score of -1. At the end of this procedure, we obtained a SS of 952 lemmas that, added to the 735 lemmas of the GS, form an extension of LatinAffectus made of 1,687 lemmas. Table 2 summarizes the final composition of such extension publicly released both as a stand-alone resource and in a

single file along with the first version of LatinAffectus.[3] The new resource created by merging LatinAffectus v1 with the extension presented in this paper is called LatinAffectus v2: it is composed of 4,124 entries, 2,437 coming from the first version of the lexicon and 1,687 from the extension. The stacked bar chart in Figure 2 compares the percentage of lemmas per score in the two sub-parts of LatinAffectus v2 with respect to the total: LatinAffectus v1 (in blue) and the extension (in red).

4. Linking to the LiLa Knowledge Base

We modeled the new entries of LatinAffectus as presented in our previous work [20]: the lexical resource is described using Lemon[4] and Ontolex[5] [23] while the sentiment properties of each entry are defined with the Marl ontology [21]. More specifically, each lexical entry of the resource has only one sense (called 'prior sense') modeled as an instance of an object of the class `ontolex:LexicalSense`. Each prior sense has a relation `marl:hasPolarity` connecting it to the class `marl:Polarity` which indicates whether the sentiment is positive, negative or neutral, and a property `marl:polarityValue` that specifies the decimal score of the sentiment. Figure 3 shows an example of how an entry of LatinAffectus is modeled.

Figure 3: Modeling of the Lexical Entry *sanabilis* (on the left) and of its sentiment proprieties as associated to its prior sense (on the right).

We linked the new entries of LatinAffectus to the LiLa KB, which uses lemmas to interlink distributed resources for Latin following the Linked Data paradigm. Out of 1,687 entries, 61 pointed to more than one possible lemma in the LiLa KB and were manually disambiguated, while 40 were not present in the KB and thus were added. An example of an entry pointing to two possible lemmas in the KB is *laevamentum* to which a polarity of +0.5 is assigned: one lemma, identified by the URI http://lila-erc.eu/data/id/lemma/111147, belongs to the same lexical base of *lĕvis* ('not heavy') and means 'consolation', whereas the other, having the URI http://lila-erc.eu/data/id/lemma/110297, belongs to the same lexical base of *lēvis* ('smooth') and refers to a means of obtaining a smooth surface. Since the lexical entry for *laevamentum* in LatinAffectus is the one for the lemma meaning 'consolation' (thus being assigned a positive polarity), we manually linked it to the former entry in the KB.

[3] https://github.com/CIRCSE/Latin_Sentiment_Lexicons
[4] https://lemon-model.net/
[5] https://www.w3.org/2016/05/ontolex/

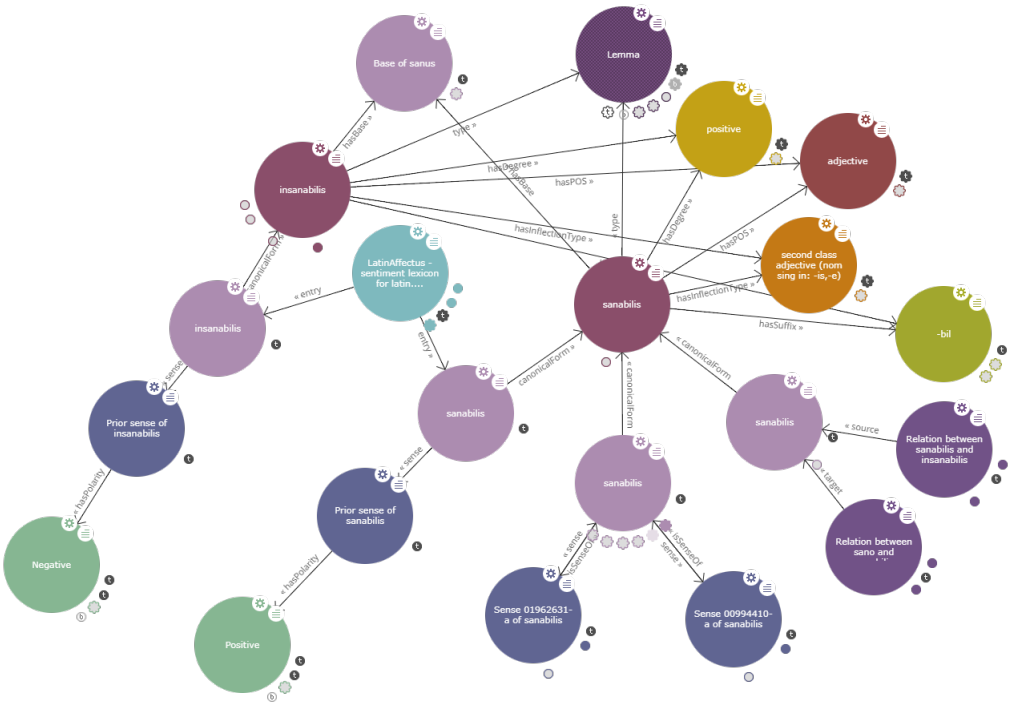

Figure 4: The lemma *sanabilis* in the LiLa Knowledge Base.

Thanks to the linking, each entry of LatinAffectus becomes part of an interoperable ecosystem made of resources of different kinds. For example, as shown in Figure 6, in the LiLa KB the lemma *sanabilis* is associated to a set of grammatical and morphological information (i.e., the part-of-speech, the degree, the inflectional category and the presence of an affix) and also to its lexical base, that is a node of the KB that connects together all the lemmas belonging to the same derivational family. *Sanabilis* shares the same derivational family of *insanabilis*; both lemmas belong to the base of *sanus*. The lemma node of *sanabilis* is also linked to the three canonical forms that represent the lexical entries of as many lexical resources linked to the LiLa KB: a derivational lexicon of Latin called Word Formation Latin [24], a new version of Latin WordNet and LatinAffectus. The left hand side of the figure shows that *sanabilis* has a positive polarity whereas *insanabilis* has a negative polarity.

5. Querying Interlinked Resources

We propose an example of use of the lexicon linked to the LiLa KB to show the benefits provided by interoperability. We query 3 different interlinked resources: i.e., LatinAffectus v.2, the collection of lemmas of the LiLa KB and UDante, the dependency treebank of Dante's Latin works. More specifically, we investigate the sentiment expressed in the *Epistole* 'Epistles', a work that includes 13 letters of various nature and subject, such as poetry, politics and moral

issues. We perform a set of federated queries across the endpoints of LiLa so to understand how the attitude of the author changes depending on the topic of the letter by calculating the sentiment at both letter and paragraph level (queries are reported in the Appendix A). In other words, we perform a lexicon-based Sentiment Analysis[6] not using a script but taking advantage of the power of SPARQL queries. Table 3 reports the overall sentiment for each letter: the absolute value is given in the second row whereas a normalized value is in the third row. The normalized value, that can range from -1 to +1, takes into consideration the number of positive and negative words in the text so to reduce the skewness caused by the different length of the letters [26]: indeed, letters XI and X contain just over 200 tokens whereas letter XIII is over 4,600 tokens long.

	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

Table 3
Sentiment of Dantes's Epistles.

From the results it is worth noticing that the letters with a more positive polarity are the ones written in the name of and on behalf of nobles or institutions for whom Dante was acting, offering his collaboration as a *dictator ab epistulis* (a writer who wrote epistolary documents). Letter I (sentiment score = 0.25) is written in the name of the Captain, the Council and the White party, while Letters VIII (0.25), IX (0.49) and X (0.43) are written for the countess of Battifolle. These last three letters are configured as texts of homage and good wishes, full of courtesy formulas. An example is given by the first paragraph of Letter X (0.83): *Illustrissime atque piissime domine domine Margarite divina providentia Romanorum regine et semper Auguste, fidelissima sua G. de Batifolle Dei et imperialis indulgentie gratia comitissa in Tuscia palatina, cum promptissima recommendatione se ipsam et voluntarium ad obsequia famulatum*, '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'.

On the contrary, the so-called 'Henry letters', written to further the campaign of the emperor Henry VII in Italy, are less positive: these are political letters that witness the bitterness of Dante's exile. In particular, Letter VI has an overall negative sentiment (-0.04): Dante writes to the Florentines, traitors of true freedom and justice warning them not to resist Henry VII, foreseeing the siege and destruction of the city of Florence by the Emperor [27]. Starting from the results provided by the SPARQL query that calculates the sentiment score at paragraph level, it is possible to trace the sentiment orientation throughout the 27 paragraphs of Letter VI (see Figure 5). The first paragraph is antithetical to that of Letter X having a fully negative polarity (-1): *Dantes Alagherii florentinus et exul inmeritus scelestissimis Florentinis intrinsecis*, 'Dante Alighieri, a Florentine undeservedly in exile, to the most iniquitous Florentines within the city'. The negative epithets are repeated in the text (e.g., Paragraph 6 (-0.33): *An ignoratis, amentes*

[6]The lexicon-based approach is based on the intuition that the polarity of a text can be obtained from the polarity of the words that compose it [25].

et discoli, publica iura cum sola temporis terminatione finiri, et nullius prescriptionis calculo fore obnoxia?, 'Have you to learn, senseless and perverse as you are, that public right can be subject to no reckoning by prescription, but must endure so long as time itself endures?'), as well as the disastrous prophecies about the future of Florence (e.g., Paragraph 15 (-1): *Videbitis edificia vestra non necessitati prudenter instructa sed delitiis inconsulte mutata, que Pergama rediviva non cingunt, tam ariete ruere, tristes, quam igne cremari*, 'The buildings which you have raised, not in prudence to serve your needs, but have recklessly altered to gratify your wantonness, these, ecircled by no walls of a renovated Troy, to your grief you shall see crumble beneath the battering-ram, and devoured by the flames.'). The positive peak at the end is due to the mention of the *faustissimi cursus Henrici Cesaris ad Ytaliam*, that is the 'the very auspicious coming of Henry Caesar in Italy'.

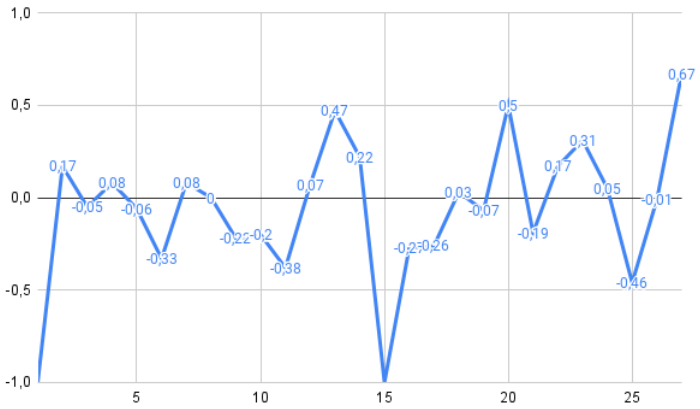

Figure 5: Sentiment orientation throughout Letter VI (scores reported in the figure are normalized).

6. Towards a Sense-Based Lexicon

We have released a set of 6,682 lemma-synset pairs taken from Minozzi's Latin WordNet (LWN) [28], manually revised [4] and mapped onto version 3.0 of Princeton WordNet [29].[7] This new resource comprises 1,073 lemmas distributed across 4,140 synsets: 1,505 synsets are nominal (denoted by the initial n# in the synset ID), 2,091 verbal (v#), 423 adjectival (a#) and 121 adverbial (r#). In order to shift from a prior polarity lexicon, such as LatinAffectus, towards a sense-based lexicon, we added sentiment information at the sense level taking SentiWordNet 3.0 as a reference. SentiWordNet 3.0 [30] contains all the WordNet 3.0 synsets associated to their degrees of positivity, negativity, and neutrality as automatically calculated using a semi-supervised learning step and a random-walk step. We mapped the clean lemma-synset pairs of LWN to those in SentiWordNet 3.0 thus inheriting the sentiment scores from the latter.

Both LWN and the information coming from SentiWordNet are modeled and linked to the LiLa KB [31]. LWN is modeled as the Princenton WordNet RDF [32]. More specifically, the

[7] https://github.com/CIRCSE/latinWordnet-revision

entries of LWN are defined as instances of the Ontolex class `LexicalEntry`. Each lexical entry is linked to the corresponding synsets of Princeton WordNet with the Ontolex property `evokes`; it also has a label, an `ontolex:canonicalForm` property connecting it to the corresponding lemma in the LiLa KB, and an `ontolex:sense` property corresponding to its lexical meaning. Each lexical entry can have several senses, each one with a label and modeled as an instance of an object of the class `ontolex:LexicalSense`. As for sentiment, each LexicalSense has a relation `marl:hasPolarity` connecting it to the class `marl:Polarity` that indicates the polarity (i.e., positive, negative or neutral) and a property `marl:polarityValue` that specifies the score of the polarity (e.g. -1). An example is given by the Lexical Entry *sanabilis* that, as displayed in Figure 6, has two senses: one corresponds to the WordNet synset a#01962631 ('capable of being remedied or redressed') and the other to the synset a#00994410 ('curing or healing is possible'). On the basis of the scores recorded in SentiWordNet 3.0, the former has a neutral polarity whereas the latter has a slightly positive polarity (0.125). The same polarity is shared by the adjectives *curabilis* and *medicabilis*, that are two Lexical Entries of LWN belonging to the same synset a#00994410 of *sanabilis*.

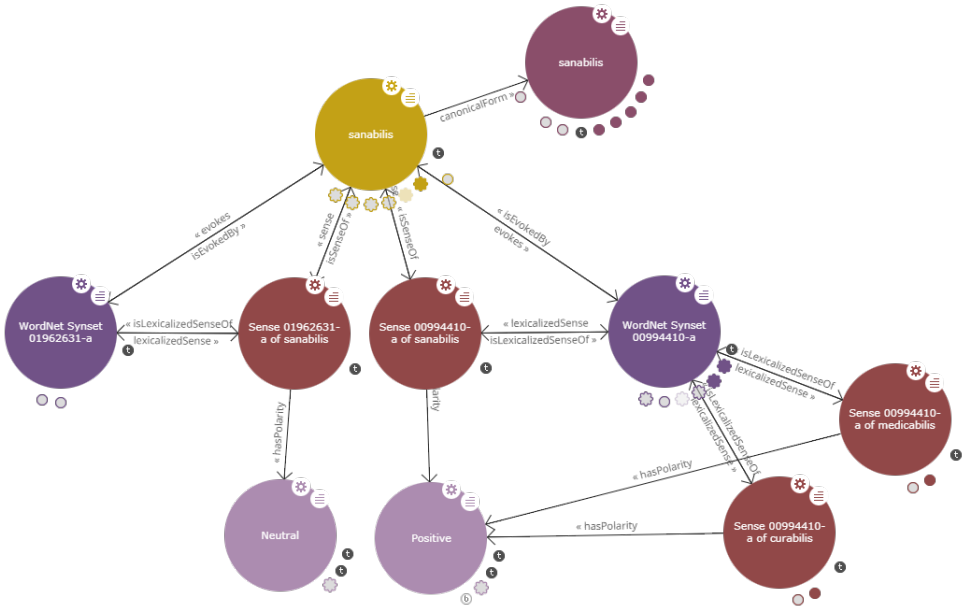

Figure 6: Modeling of polarity, taken from SentiWordNet 3.0, at sense level.

The case of *sanabilis* is an example of how the scores derived from SentiWordNet are not perfectly in line with those of LatinAffectus. Indeed, the prior polarity of *sanabilis* is fully positive and this score seems more appropriate for the two synsets mentioned above. Another example of discrepancy is given by *peccator* 'sinner' that has two senses (n#10285762: 'someone who transgresses; someone who violates a law or command' and n#10601078: 'a person who sins (without repenting)') both with a fully negative polarity in LatinAffectus but with a neutral score in SentiWordNet. More specifically, there are 18 lemmas with a positive polarity in LatinAffectus but with a negative or neutral polarity in SentiWordNet 3.0; 45 lemmas with a negative polarity

in LatinAffectus but with a positive or neutral polarity in SentiWordNet 3.0; 28 lemmas with a neutral polarity in LatinAffectus but with a positive or neutral polarity in SentiWordNet 3.0. It will be important to manually check the scores assigned automatically by SentiWordNet 3.0 in order to create and release a sense-based sentiment lexicon with high accuracy.

7. Conclusion

The practice of performing NLP tasks on literary (and, more generally, cultural) texts, as well as of building linguistic resources to support their automatic analysis is today spreading more and more. Along such line, this paper describes the process we performed to extend a sentiment lexicon for Latin and to interlink it with other available linguistic resources for Latin, presenting an example of how to use the interoperable resources to analyze the sentiment value of the Latin epistles by Dante Alighieri. While in this paper we just briefly comment the results provided by some SPARQL queries both at single epistle and paragraph level, their interpretation would deserve a deeper expertise than ours, thus requiring an interdisciplinary collaboration between scholars coming from computational linguistics and from literary studies, respectively. However, while leaving aside the age-old separation between the two areas, currently such a collaboration has to face a number of practical limitations on both sides.

On the one side, computational tools and digital resources for literary studies still provide too low accuracy rates and too small sets of textual data enhanced with linguistic annotation. As for the specific case reported in this paper, the most evident limitation is the current lack of any tool performing high/mid-quality automatic Word Sense Disambiguation for Latin, which would make it possible to switch from labeling lemmas with prior polarity values (lexical level) to assigning context-based values to their single textual occurrences (textual level).

On the other side, the scholarly area concerned with literary studies still lacks to fully realize the real impact of using NLP tools and computational lexicons in support of its research questions. Such a realization must be strictly bound to a methodological, empirical turn in the Humanities, which would result in supporting scholarly interpretations with evidence and replicable results. Hopefully, such a data-based collaboration between computational linguistics and literary studies promises to start a kind of virtuous circle, where better results from computational linguistics provide solid empirical evidence supporting better insights from literary studies. Our work within the LiLa project wants to join these two areas, aiming to strengthen the use and exploitation of computational methods in the Humanities, thus supporting the turn towards Computational Humanities.

Acknowledgments

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.

References

[1] R. Sprugnoli, M. Passarotti, Proceedings of lt4hala 2020-1st workshop on language technologies for historical and ancient languages, in: Proceedings of LT4HALA 2020-1st Workshop on Language Technologies for Historical and Ancient Languages, 2020.

[2] F. M. Cecchini, R. Sprugnoli, G. Moretti, M. Passarotti, Udante: First steps towards the universal dependencies treebank of dante's latin works, in: Seventh Italian Conference on Computational Linguistics, CEUR-WS. org, 2020, pp. 1–7.

[3] J. Nivre, M.-C. De Marneffe, F. Ginter, Y. Goldberg, J. Hajic, C. D. Manning, R. McDonald, S. Petrov, S. Pyysalo, N. Silveira, et al., Universal dependencies v1: A multilingual treebank collection, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 1659–1666.

[4] G. Franzini, A. Peverelli, P. Ruffolo, M. Passarotti, H. Sanna, E. Signoroni, V. Ventura, F. Zampedri, Nunc est aestimandum: Towards an evaluation of the latin wordnet, in: Sixth Italian Conference on Computational Linguistics (CLiC-it 2019), Accademia University Press, 2019, pp. 1–8.

[5] B. Liu, Sentiment analysis and opinion mining, Synthesis lectures on human language technologies 5 (2012) 1–167.

[6] K. Zvarevashe, O. O. Olugbara, A framework for sentiment analysis with opinion mining of hotel reviews, in: 2018 Conference on information communications technology and society (ICTAS), IEEE, 2018, pp. 1–4.

[7] L. Yue, W. Chen, X. Li, W. Zuo, M. Yin, A survey of sentiment analysis in social media, Knowledge and Information Systems 60 (2019) 617–663.

[8] A. Rodríguez, C. Argueta, Y.-L. Chen, Automatic detection of hate speech on facebook using sentiment and emotion analysis, in: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), IEEE, 2019, pp. 169–174.

[9] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, Y. Chang, Abusive language detection in online user content, in: Proceedings of the 25th international conference on world wide web, 2016, pp. 145–153.

[10] A. Alamoodi, B. Zaidan, A. Zaidan, O. Albahri, K. Mohammed, R. Malik, E. Almahdi, M. Chyad, Z. Tareq, A. Albahri, et al., Sentiment analysis and its applications in fighting covid-19 and infectious diseases: A systematic review, Expert systems with applications (2020) 114155.

[11] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, J. Kim, Covidsenti: A large-scale benchmark twitter data set for covid-19 sentiment analysis, IEEE Transactions on Computational Social Systems (2021).

[12] T. Haider, S. Eger, E. Kim, R. Klinger, W. Menninghaus, Po-emo: Conceptualization, annotation, and modeling of aesthetic emotions in german and english poetry, in: International Conference on Language Resources and Evaluation (LREC), 2020.

[13] P. Koncar, A. Fuchs, E. Hobisch, B. C. Geiger, M. Scholger, D. Helic, Text sentiment in the age of enlightenment: an analysis of spectator periodicals, Applied Network Science 5 (2020) 1–32.

[14] T. Schmidt, J. Dangel, C. Wolff, Senttext: A tool for lexicon-based sentiment analysis in digital humanities, in: Information between Data and Knowledge, volume 74 of *Schriften*

zur Informationswissenschaft, Werner Hülsbusch, Glückstadt, 2021, pp. 156–172. Session 3: Digital Humanities.
[15] M. Kaity, V. Balakrishnan, Sentiment lexicons and non-english languages: a survey, Knowledge and Information Systems (2020) 1–36.
[16] Y. Hou, A. Frank, Analyzing sentiment in classical chinese poetry, in: Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), 2015, pp. 15–24.
[17] Y. Chen, S. Skiena, Building sentiment lexicons for all major languages, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2014, pp. 383–389.
[18] S. Mohammad, Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 174–184.
[19] R. Sprugnoli, M. Passarotti, D. Corbetta, A. Peverelli, Odi et amo. creating, evaluating and extending sentiment lexicons for latin., in: Proceedings of The 12th Language Resources and Evaluation Conference, 2020, pp. 3078–3086.
[20] R. Sprugnoli, F. Mambrini, G. Moretti, M. Passarotti, Towards the modeling of polarity in a latin knowledge base, in: Third Workshop on Humanities in the Semantic Web (WHiSe 2020), CEUR-WS. org, 2020, pp. 59–70.
[21] P. Buitelaar, M. Arcan, C. A. Iglesias, J. F. Sánchez-Rada, C. Strapparava, Linguistic Linked Data for Sentiment Analysis, in: Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and linking lexicons, terminologies and other language data, 2013, pp. 1–8.
[22] U. Sidarenka, Sentiment analysis of German Twitter, doctoralthesis, Universität Potsdam, 2019. doi:10.25932/publishup-43742.
[23] J. P. McCrae, J. Bosque-Gil, J. Gracia, P. Buitelaar, P. Cimiano, The Ontolex-Lemon model: development and applications, in: Proceedings of eLex 2017 conference, 2017, pp. 19–21.
[24] E. Litta, M. Passarotti, (When) inflection needs derivation: a word formation lexicon for Latin, in: N. Holmes, M. Ottink, J. Schrickx, M. Selig (Eds.), Lemmata Linguistica Latina. Volume 1. Words and Sounds, De Gruyter, Berlin, Boston, 2019, pp. 224–239.
[25] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Computational linguistics 37 (2011) 267–307.
[26] A. Jurek, M. D. Mulvenna, Y. Bi, Improved lexicon-based sentiment analysis for social media analytics, Security Informatics 4 (2015) 1–13.
[27] L. Marcozzi, L'epistola di dante ai fiorentini: memoria scritturale, profetismo e tracce umanistiche dell'invettiva dantesca, in: Le lettere di Dante, De Gruyter, 2020, pp. 329–352.
[28] S. Minozzi, Latin wordnet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'information retrieval, Strumenti digitali e collaborativi per le Scienze dell'Antichità (2017) 123–134.
[29] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. J. Miller, Introduction to wordnet: An on-line lexical database, International journal of lexicography 3 (1990) 235–244.
[30] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining., in: Lrec, volume 10, 2010, pp. 2200–2204.
[31] M. Passarotti, F. Mambrini, E. Litta, G. Moretti, Interlinking Valency Frames and WordNet

Synsets in the LiLa Knowledge Base of Linguistic Resources for Latin, in: 17th International Conference on Semantic Systems (SEMANTiCS), 2021.
[32] P. Cimiano, C. Chiarcos, J. P. McCrae, J. Gracia, Applying linked data principles to linking multilingual wordnets, in: Linguistic Linked Data, Springer, 2020, pp. 215–228.

A. SPARQL queries

Queries to replicate the results reported in Section 5 using the LiLa endpoint available at https://lila-erc.eu/sparql/dataset.html?tab=query&ds=/corpora.

- Use the following query to calculate the sentiment value for each Letter by Dante Alighieri. The query returns a table with 4 columns: the URI of the text (*ep*), the absolute value of the sentiment per text (*sentVal*), the number of positive and negative lemmas with the text (*count*) and the sentiment value normalized with respect to the number of polarized lemmas (*normVal*).

```
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#>
prefix lila: <http://lila-erc.eu/ontologies/lila/>
prefix lime: <http://www.w3.org/ns/lemon/lime#>
prefix marl: <http://www.gsi.dit.upm.es/ontologies/marl/ns#>
prefix powla: <http://purl.org/powla/powla.owl#>
prefix lila_corpora: <http://lila-erc.eu/ontologies/lila_corpora/>

SELECT  ?ep (SUM(?v) as ?sentVal) (COUNT(?v) as ?count)
((SUM(?v) / COUNT(?v)) as ?normVal)
WHERE {
    service <https://lila-erc.eu/sparql/lexicalResources/query> {
    VALUES ?polarity { marl:Positive marl:Negative }
        ?le ontolex:canonicalForm ?lemma ;
        ontolex:sense ?sense .
    ?sense marl:hasPolarity ?polarity ;
          marl:polarityValue  ?v ;
    }

    ?token lila:hasLemma ?lemma ;
      ^powla:hasChild/^lila_corpora:hasCitSubUnit ?ep.
   ?CitationLayer powla:hasDocument
     <http://lila-erc.eu/data/corpora/DanteSearch/id/corpus/Epistole>;
                  lila_corpora:isLayer ?ep .
} group by  (?ep)
order by ?ep
```

- Use the following query to calculate the sentiment value for each paragraph of each Letter by Dante Alighieri. The query returns a table with 5 columns: the URI of the text (*ep*), the URI of the paragraph (*par*), the absolute value of the sentiment per text (*sentVal*), the number of positive and negative lemmas with the text (*count*) and the sentiment value normalized with respect to the number of polarized lemmas (*normVal*).

```
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#>
prefix lila: <http://lila-erc.eu/ontologies/lila/>
prefix lime: <http://www.w3.org/ns/lemon/lime#>
prefix marl: <http://www.gsi.dit.upm.es/ontologies/marl/ns#>
prefix powla: <http://purl.org/powla/powla.owl#>
prefix lila_corpora: <http://lila-erc.eu/ontologies/lila_corpora/>

SELECT  ?ep ?par (SUM(?v) as ?sentVal) (COUNT(?v) as ?count)
((SUM(?v) / COUNT(?v)) as ?normVal)
WHERE {
   ?CitationLayer powla:hasDocument
   <http://lila-erc.eu/data/corpora/DanteSearch/id/corpus/Epistole>;
                  lila_corpora:isLayer ?ep .
  ?ep lila_corpora:hasCitSubUnit ?par .
  ?par powla:hasChild ?token.
  ?token lila:hasLemma ?lemma .
    service <https://lila-erc.eu/sparql/lexicalResources/query> {
    VALUES ?polarity { marl:Positive marl:Negative }
        ?le ontolex:canonicalForm ?lemma ;
        ontolex:sense ?sense .
    ?sense marl:hasPolarity ?polarity ;
       marl:polarityValue  ?v ;
    }

} group by  ?ep ?par
order by ?ep ?par
```