

Towards the Modeling of Polarity in a Latin Knowledge Base*

Rachele Sprugnoli^[0000-0001-6861-5595], Francesco
Mambrini^[0000-0003-0834-7562], Giovanni Moretti, and
Marco Passarotti^[0000-0002-9806-7187]

CIRCSE Research Centre, Università Cattolica del Sacro Cuore
Largo Agostino Gemelli 1, 20123 Milano
{rachele.sprugnoli, francesco.mambrini, giovanni.moretti,
marco.passarotti}@unicatt.it

Abstract. In this paper, we describe the process of inclusion of a prior polarity lexicon of Latin lemmas, called *LatinAffectus*, in a knowledge base of interoperable linguistic resources developed within the *LiLa: Linking Latin* project. More specifically, a manually-curated list of lemma-sentiment pairs is linked to a comprehensive collection of Latin lemmas by using Semantic Web and Linked Data standards and practices. *LatinAffectus* is modeled relying on three formal representation frameworks: Lemon and Ontolex to describe the lexicon, and the Marl ontology to describe the sentiment properties of each of its lexical entries. We present the lexicon, the methodology and the results of the linking process, as well as a use case and the planned future work.¹

Keywords: Linguistic Linked Open Data · Sentiment Analysis · Latin.

1 Introduction

Throughout the recent years, several linguistic resources and tools were created for many languages to support sentiment analysis, i.e. the task of automatically classifying a piece of text according to the sentiment conveyed by it. Although the main applications of such resources and tools fall into categories like social media and customer experience monitoring, there is a growing interest in the research community to develop resources and tools to perform sentiment analysis of texts written in ancient languages. Such interest mirrors the substantial growth of the area dedicated to building and using linguistic resources for ancient and historical languages, which has primarily concerned Latin and Ancient Greek as essential media for accessing and understanding the so-called Classical tradition.

* This work is supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme via the “LiLa: Linking Latin” project - Grant Agreement No. 769994.

¹ Copyright ©2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In particular, Latin plays a central role in this context, as texts written in Latin are spread all over Europe, covering a time span of almost two millennia and being testimonials of the common, but still diverse, past that contributed to shape the cultural heritage of Europe. Exploiting the most advanced techniques for preserving, investigating and sharing such heritage assets that have survived from the past times is at the same time a challenge and an obligation for the research area dealing with developing linguistic resources and tools. Given the wide variety of the Latin texts in terms of their era, place and literary genre, the achievements of this field of research promise to impact a large and heterogeneous community made of historians, philologists, archaeologists and literary scholars, in different ways all dealing with textual and lexical data written in Latin.

The recent launch of projects aimed at automatically extracting structured knowledge from ancient sources provided by linguistic resources, like for instance *eAqua*², *Logeion*³ and *Corpus Corporum*⁴, shows how the current availability of linguistic resources for ancient languages, and particularly Latin, is such that there is a large need for making them interact.

To address the issue of interoperability between lexical and textual resources for Latin, the *LiLa: Linking Latin* project (2018-2023)⁵ was launched with the objective of building a Knowledge Base (KB) of linguistic resources for Latin based on the Linked Data paradigm, i.e. a collection of several data sets represented using the same vocabulary of knowledge description and linked together [25]. Within the LiLa project, aside from interlinking the already available resources for Latin, we are also building a number of new ones, among which is *LatinAffectus*, a lexicon that assigns a prior sentiment score to a selection of Latin adjectives and nouns [28].

This paper describes the process of inclusion of *LatinAffectus* into the LiLa KB and presents a simple use-case showing how the interaction of the linguistic resources currently linked through LiLa can be exploited to address a specific research question.

The core component of LiLa is a large collection of Latin lemmas, whose role is to connect the different (and possibly distributed) linguistic resources that interact in LiLa [17]. Particularly, the textual resources are included into LiLa by linking the occurrences of the words in their texts to the lemmas of LiLa, while lexical resources connect to LiLa by linking the contents of their lexical entries to the lemmas of the KB. The result is an interlinked ecosystem where textual and lexical (meta)data provided by several resources become interoperable. Including *LatinAffectus* into LiLa enhances a subset of the lemmas provided by the LiLa collection with a prior positive/negative polarity. Such a black or white approach is at the same time a limitation and an advantage of *LatinAffectus*. As for the former, the lexicon does not account for the different meanings that words may have, some of which can show different polarity values, thus failing to represent

² <http://www.eaqua.net/>

³ <https://logeion.uchicago.edu/lexidium>

⁴ <http://www.mlat.uzh.ch/MLS/>

⁵ <https://lila-erc.eu>

the span of possible sentiments of a word. As for the latter, assigning one prior, prototypical polarity value to the lexical entries helps the application of the (meta)data from *LatinAffectus* to real texts. Indeed, no sufficiently accurate tools for word sense disambiguation are currently available for Latin, which prevents from analyzing texts with the help of sentiment lexicons that provide different polarity values for the same word, as this implies to consider all the possible values while computing the overall polarity of a sentence, or a text.⁶ Instead, by grounding on one, prior polarity value it becomes possible to apply *LatinAffectus* to Latin texts without the need of pre-processing data with a layer of word sense disambiguation. This aspect becomes an added value when *LatinAffectus* interacts with all the other resources included in LiLa, because its (meta)data are not anymore available in isolation, but they are interoperable with those of other resources, thus exploiting to the best the contribution provided by each of them in applications to address research questions.

The paper is organized as follows. Section 2 provides a brief overview of the related work on polarity lexicons and on the strategies to represent linguistic resources and services for sentiment analysis in the Linked Data framework. Section 3 describes *LatinAffectus* and Section 4 details the process of modeling it including it into the LiLa KB. Section 5 presents a simple use case to show how the interoperability between the resources connected in LiLa, and particularly *LatinAffectus*, can support research in the Humanities. Finally, Section 6 concludes the paper with a discussion about the need to make the linguistic resources for Latin interact and sketches our future work.

2 Related Work

Sentiment Analysis and related tasks [16], such as emotion analysis, subjectivity detection and opinion mining, are very popular both in academic research and in business applications where the focus is mostly given to the analysis of contemporary texts like product or service reviews [11] and social media posts [24]. In this tasks, the creation of polarity lexicons, that is lists of words associated to their out-of-context sentiment orientation, is of fundamental importance but also a very time-consuming process. Several approaches have been developed to automatize this process and build multi-lingual lexicons covering also less-resources and ancient languages. For example, Mohammad [22] adopts crowd-sourcing techniques to generate an English valence, arousal, dominance (VAD) lexicon and then automatically translates it into other 103 languages, including Latin. Chan and Skiena [5] use instead a knowledge graph propagation algorithm starting from Wikipedia to build lexicons of positive and negative words in 136 languages inclusive of Latin.

These two resources, although of undoubted value, have two main drawbacks: they are noisy due to the presence of English words, such as *microchip* and *reli-*

⁶ An example of such approach is provided by the API of the *Latin WordNet* project of the University of Exeter, which can perform sentiment analysis of individual strings via HTTP POST requests to <https://latinwordnet.exeter.ac.uk/sentiment/>.

*able*⁷ and their content was not checked by a Latin expert or evaluated against a gold standard. Another limitation is that these lexicons have not been published in the Linked Data framework, thus limiting their usability and semantic interoperability. However, various efforts have been made to develop a formal representation of linguistic resources and services for sentiment analysis. More specifically, the Marl ontology is designed for the publication of data about opinions and the sentiments expressed in them [30] and the EuroSentiment project has proposed a model that integrates it with the lexicon model for ontologies (lemon) [3,19], so to represent lexical resources for sentiment and emotion analysis such as lexicons and annotated corpora [2]. This approach has been applied, for example, to represent polarity information of German compound words, relying, in particular, on Ontolex [20], that is the core module of lemon [7].

In this paper we aim to overcome the aforementioned limitations of the currently available polarity lexicons for Latin by publishing, using Linked Data principles, a list of lemmas with their prior sentiment orientation created by experts of Latin language and culture, and then expanded by exploiting a set of already available manually curated linguistic resources. Each entry in the resources we created is linked to the collection of Latin lemmas provided by the LiLa KB, so to achieve interoperability.

3 Latin Prior Polarity Lexicons

A Gold Standard (GS) lexicon was manually developed by two Latin language and culture experts who assigned a sentiment score to out-of-context lemmas using a five-value classification: 1 (fully positive), 0.5 (somewhat positive), 0 (neutral), -0.5 (somewhat negative), -1 (fully negative). We chose to take into consideration nouns and adjectives only because their polarity is more easy to define at a lexical level, i.e. out of context, than that of verbs whose semantics is more strictly connected to that of their arguments [12]. Lemmas were taken from the William Whitaker’s *Words* morphological analyzer and digital dictionary⁸, the Cassell’s Latin dictionary⁹ [27] and the lemmatized version of *Opera Latina* [9], a corpus of Classical authors manually annotated with lemmas and Part-of-Speech (PoS) tags. In addition, a Silver Standard (SS) lexicon was built by deriving new entries in two ways: i) by exploiting derivational, synonym and antonym relations with the lemmas in the GS; ii) by adding graphical variants of lemmas present in the GS. Original polarity scores were propagated or reversed onto the newly derived lemmas: for example, scores were preserved in case of synonyms and graphical variants, whereas they were reversed for antonyms (see Table 1). Details on the composition of the GS and the SS are reported in Table 2: the resources are freely available online at <https://github.com/CIRCSE/>

⁷ By manually revising the two lexicons, we calculated the percentage of English words: 14% in the VAD lexicon and 9% in the other.

⁸ <https://mk270.github.io/whitakers-words/>

⁹ <https://github.com/nikita-moor/latin-dictionary>

Latin_Sentiment_Lexicons and their detailed description is given in [28]. The GS and the SS have been merged in a unique resource called *LatinAffectus*.

Table 1. Examples of extensions starting from lemmas in the GS.

Lemma-GS	Score-GS	Extension Type	Lemma-SS	Score-SS
<i>purus</i> ‘pure’	+1	derivational	<i>impurus</i> ‘impure’	-1
<i>innoxius</i> ‘harmless’	+0.5	synonym	<i>innocens</i> ‘innocent’	+0.5
<i>aqua</i> ‘water’	0	derivational	<i>aquarius</i> ‘of/for water’	0
<i>apsentia</i> ‘absence’	-0.5	variant	<i>absentia</i> ‘absence’	-0.5
<i>scelus</i> ‘crime’	-1	antonym	<i>beneficium</i> ‘benefit’	+1

Table 2. Composition of the GS and the SS prior polarity lexicons together with the total number of adjectives and nouns included in the merged resource *LatinAffectus*.

LEXICON	PoS		TOT
	ADJ	NOUN	
Gold Standard	454 (39.7%)	690 (60.3%)	1,144
Silver Standard	512 (39.6%)	781 (60.4%)	1,293
LatinAffectus	966 (39.6%)	1,471 (60.4%)	2,437

4 Modeling and Linking Polarity

For modeling *LatinAffectus*, we rely on three formal representation frameworks: Lemon¹⁰ and Ontolex¹¹ to describe the lexical resource and the Marl ontology¹² to describe the sentiment properties of each entry.

In our approach the polarity lexicon is defined as an instance of the class E31 Document¹³ of the CIDOC Conceptual Reference Model (CRM), an ontology formally describing concepts and relations in the cultural heritage domain [10]. Moreover, *LatinAffectus* is also defined as an object of type `lexicon` following the LInguistic MEtadata (*lime*) module of Ontolex. We link the lexicon to its entries, which are defined as instances of the Ontolex class `LexicalEntry`, through the property called `entry` belonging to the *lime* module. Each lexical entry has a label, an `ontolex:canonicalForm` property connecting it to the corresponding lemma in the LiLa KB, and an `ontolex:sense` property corresponding to the lexical meaning of a lexical entry. Given that *LatinAffectus* deals with prior polarities, each lexical entry has only one sense, modeled as an instance of an object of the class `ontolex:LexicalSense`. Each sense is characterized by a

¹⁰ <https://lemon-model.net/>

¹¹ <https://www.w3.org/2016/05/ontolex/>

¹² <http://www.gsi.dit.upm.es/ontologies/marl/1.1/>

¹³ The class E31 “comprises identifiable immaterial items that make propositions about reality”, <http://www.cidoc-crm.org/Entity/e31-document/version-6.2>

label, the relation `marl:hasPolarity` and the property `marl:polarityValue`. More specifically, the relation `marl:hasPolarity` connects the sense to the class `marl:Polarity`, indicating if the sentiment is positive, negative or neutral. On the other side, `marl:polarityValue` specifies the numeric decimal value of the sentiment that, in our case, can be 1.0, 0.5, 0.0, -0.5, or -1.0.

The lexical entries of *LatinAffectus* were linked to their corresponding lemmas in the LiLa KB in a semi-automatic way. First, we performed an automatic matching between the two resources: this revealed the presence of 246 ambiguous lemmas, that is lemmas having the same written representation and the same PoS tag: for example, the entry *fides*, having polarity value 1, can be linked either to the lemma of the fifth declension meaning ‘trust’ or to the lemma of the third declension meaning ‘lyre’. These lemmas were manually disambiguated; thus *fides* was linked to the first of the aforementioned lemmas in the LiLa KB. Further 107 entries, such as Medieval or New Latin words like *praesuppositio* ‘assumption’ and *radioactiuus* ‘radioactive’, were not present in the KB and were therefore added.

Figure 1 illustrates the modeling of the lexical entry *malus* ‘evil’ and of the negative prior polarity of its lexical sense as recorded in *LatinAffectus*.

```
@prefix lemma: <http://lila-erc.eu/data/id/lemma/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
@prefix marl: <http://www.gsi.dit.upm.es/ontologies/marl/ns#> .

<http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalEntry/lemma_111418>
  a ontolex:LexicalEntry ;
  rdfs:label "malus" ;
  ontolex:canonicalForm lemma:111418 ;
  ontolex:sense <http://lila-erc.eu/data/lexicalResources/LatinAffectus/
    id/LexicalSense/lemma_111418> .

<http://lila-erc.eu/data/lexicalResources/LatinAffectus/id/LexicalSense/lemma_111418>
  a ontolex:LexicalSense ;
  rdfs:label "Prior sense of malus" ;
  marl:hasPolarity marl:Negative ;
  marl:polarityValue -1.0 .
```

Fig. 1. Triples in Notation3 format related to the lexical entry *malus* ‘evil’.

Thanks to the linking between entries coming from different resources, among which are *LatinAffectus* and the collection of lemmas of the LiLa KB, it is possible to get a rich set of lexical information. An example is given in Figure 2¹⁴: a number of morphological features, including PoS, degree and inflectional category, are assigned to the node for the lemma *malus* in the KB. This node is also connected to a Base node that plays the role of connecting together all the lemmas belonging to the same derivational family: in this Figure the Base2302 node interlinks the lemmas *malus* and *maleficus* ‘wicked’. The Lemma node *malus* is also connected to its etymology taken from the *Etymological Dictionary of Latin*

¹⁴ The LiLa KB can be explored using a query interface: <https://lila-erc.eu/query/>.

and the other Italic Languages [29] and modeled by relying on the Ontolex-lemmon ontology and the lemonEty extension [14]. In particular, the Lemma node is linked to the canonical form present in the etymological dictionary, which in turn is linked to the Proto-Italic and/or Proto-Indo-European reconstructed forms of the word (e.g. **malo-*) [18]. The left hand side of the image shows that both *malus* and *maleficus* are included in *LatinAffectus* and they both have a negative polarity.

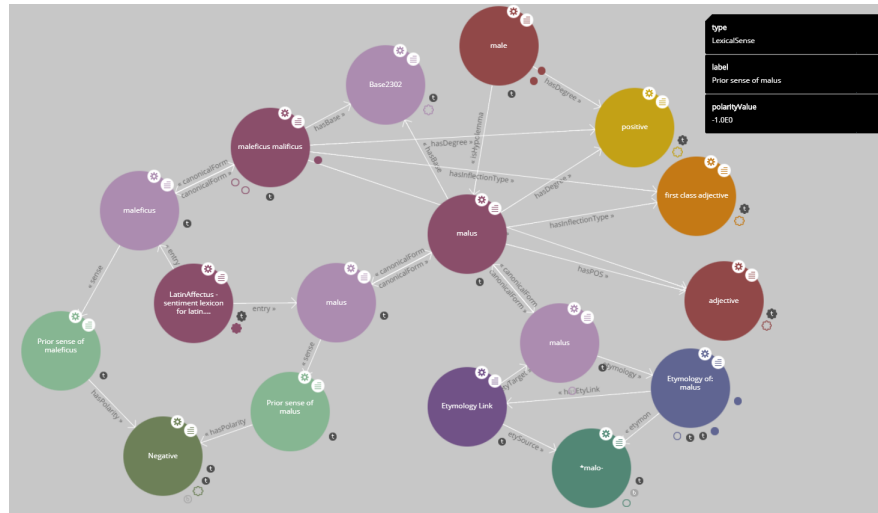


Fig. 2. The lemma *malus* in the LiLa KB including its etymology and its prior polarity.

5 Use Case

The only corpus that is presently connected to the LiLa KB is the *Index Thomisticus* Treebank (ITTb) [26], which provides a complete morpho-syntactic annotation, based on a form of dependency grammar, to the Latin works of the philosopher Thomas Aquinas (13th Century). At the moment, the LiLa KB stores the connections between the 277,547 tokens, taken from the first four books of the treatise *Summa contra Gentiles* (SCG), and the corresponding lemma under which each token is lemmatized. For every token we report also the basic morpho-syntactic information derived from the treebank, such as the link between head and dependent in the dependency tree and the label of their syntactic relation (e.g. “Subject” or “Predicate”).

Although the original annotation stored in the ITTB already allows users to perform complex queries on the language of the SCG, the inclusion of the corpus into the LiLa KB expands the range of possible research to integrate new crucial dimensions for researcher; polarity provides an outstanding example.

One of the philosophical problems that Thomas Aquinas engaged in his teaching and writing is the nature of evil.¹⁵ From the linguistic point of view, one of the prototypical constructions for providing definitions to concepts is the nominal predicate, where a subject is associated with a predicate via the copula verb, like “to be” in English.¹⁶ One example of this form of definition is the sentence: “evil is a deficiency of good”. While the ITTB already allows the users to retrieve the occurrences of copular constructions, it is only by crossing the information with *LatinAffectus* that we can specify the constraint on the polarity of either of the terms (the subject or the predicate nominal). In this way, we can explore the definitions of the negative pole across the work of Thomas Aquinas, and consider its relation to the problem of the nature of evil.

Table 3. The 5 most frequent negative subjects of copular constructions in the ITTB.

Lemma	English translation	Occurrences
<i>malum</i>	‘evil’	37
<i>corruptio</i>	‘destruction’, ‘decay’, ‘corruption’	7
<i>fornicatio</i>	‘fornication’	4
<i>labor</i>	‘labour’, ‘toil’, ‘exertion’	3
<i>occisio</i>	‘killing’, ‘murder’	3

With a series of federated queries across the three endpoints of LiLa¹⁷ (the collection of lemmas, called Lemma Bank, the corpora, and the lexical resources), it is possible to obtain such results. On the negative pole, the ITTB includes 67 tokens labeled with a negative polarity that are the subjects of copular constructions; the 5 most frequently attested lemmas are reported in Table 3. Not surprisingly, by far the most numerous attestations are those of the technical word for the concept of evil itself, *malum*. As Davis observes [6, 14], the term *malum* is, in its philosophical sense, broader than English ‘evil’. The latter carries very strong connotations and is usually not applied to what is perceived as generically unpleasant or troublesome; on the other hand, in the philosophical literature, *malum* covers all the entities that can be said to fall short of the opposite pole of *bonum* (‘good’).

It is also informative to extract all the couplets subject-predicate nominal where the subject is negative. The word that is constantly associated with *fornicatio* ‘fornication’ as predicate nominal in the ITTB is *peccatum* ‘sin’, as all the four occurrences come from a section of the SCG discussing ‘why simple fornication is a sin according to divine law’ (*qua ratione fornicatio simplex secundum legem divinam sit peccatum*, SCG 3.120).

¹⁵ Thomas Aquinas authored a full treatise *On Evil* in the form of a *disputatio*, a formalized treatment of a subject articulated into questions and arguments and counterarguments, issuing from public debates or school seminars [6, 3-53].

¹⁶ On the copular sentences and on the history of the notion of copula, which is also closely tied to the history of Western philosophy, see Moro [23, 248-261].

¹⁷ <https://lila-erc.eu/sparql/index.html>

The words *labor* ‘labour, toil, exertion’ and especially *corruptio* ‘destruction, corruption’ are interesting as well. The former is a complex term that embraces the notion of physical effort, then that of economic production, but also of pain and fatigue. While its main association with extortion justifies the negative polarity, it is a word that could also carry positive association, in the moral and economic sphere, both in the Pagan and Christian cultures [15]. Indeed, Thomas Aquinas also uses it with the adjective *bonus*, to refer (in the plural) to the ‘good words [...] by which we satisfy God for our sins’ [8, 622]. In the ITTB, all the three occurrences of *labor* are coupled with the adjective *necessarius* ‘necessary’, to point to the economic prerequisite of manual work for survival (SCG 3.140) and to rebut the claim that it is equally necessary on moral grounds (SCG 3.140.15 and 16).

The definitions of *corruptio* point to a more complex picture. With the basic meaning of ‘destruction’, ‘decay’ the word has a clear negative orientation. At the same time, the nuances in its use reflect the complexities in the question of ‘evil’. Destruction can in itself be a good thing, if it implies the destruction of evil, and indeed this is exactly one of the first points raised by the philosopher in a sentence that provides a striking example of association of a negative subject with a positive predicate nominal: *cum corruptio mali sit bona* ‘for the destruction of evil is good’ (SCG 3.11.4). Again, there are particular goods for which the corruption (*corruptio*) of the ones means the generation (*generatio*) of the other (SCG 1.11.12). In fact, the opposition between the two terms (*corruptio* and *generatio*) [8, 252] is well reflected also in another passage (SCG 3.140.4) where the concept that an evil is subordinated to the creation of a good is exemplified by the fact that the corruption (*corruptio*) of air is the generation of fire (*generatio*).¹⁸

This case study has presented only a very preliminary and partial analysis, as we have focused our attention only on one of the many syntactical constructions that may be worth investigating.¹⁹ In addition, it is important to note that one crucial issue that could limit the value of the results presented above is the coverage of our polarity lexicon. Given the limited number of lemmas included in the lexicon, it is possible that other negative terms, which do not have a polarity annotation, are used in copular constructions and are not detected by our query. In order to verify this situation, we extracted the list of the 150 most frequent lemmas that are used as subject of copular constructions. In the list we identified 4 lemmas (out of a total of 5, including *malum*, which ranked 37th) that should have had a negative polarity but are not included in *LatinAffectus*: *privatio* ‘deprivation’ (25 occurrences), *paupertas* ‘poverty’ (9), *peccatum* ‘sin’ (7), *defectus* ‘difect’ (6). As these lemmas should have ranked between the sec-

¹⁸ It is worth noticing that, although in Thomas Aquinas the two terms are opposite and *corruptio* is assigned a negative polarity, *generatio* does not have a polarity label in *LatinAffectus*.

¹⁹ Another interesting example could be the instances of coordination, to extract all terms that are associated to positive and/or negative concepts in ‘and’ or ‘or’ construction.

ond and the third position of Table 3, it is clear that this form of evaluation is required before using the corpus data. Nevertheless, it is already possible to see how fruitful is the crossing between lexical data on polarity and linguistic annotation for a broad range of corpus-based researches.

6 Conclusion and Future Work

In this paper, we have described the process of inclusion of a sentiment lexicon for Latin (*LatinAffectus*) into the LiLa KB, an infrastructure of interoperable linguistic resources based on the Linked Data paradigm. Instead of developing from scratch a new model for representing the lexicon in Linked Data, we selected three already available and widely used models, following the general recommendation of the Linked Data world to re-use existing ontologies and vocabularies as much as possible, in order to enhance interoperability with other resources.

Interoperability is the key word here. To show the benefit of working with linguistic resources that interact with each other, we presented a simple use case, where (meta)data from different resources for Latin (a dependency tree-bank, the *LatinAffectus* lexicon and the collection of lemmas of the LiLa KB) interact to investigate a basic topic of Thomistic philosophy. Although the work of philosophical investigation on the texts of Thomas Aquinas is centuries-old, it was never possible until now to join automatically (and, thus, to exploit to the best) the information provided by separate resources, like lexicons, dictionaries and corpora, to find the answers to fundamental research questions. As for the specific case of Thomas Aquinas, this goes back to the history itself of linguistic computing, since his texts were among the first to be automatically processed with computers, when in the 1950s the Jesuit Roberto Busa started to use IBM machines to build the large corpus of the *Index Thomisticus* [4].

Today we can make the data of the *Index Thomisticus* (now partly tree-banked) speak the same language of several other linguistic resources for Latin that were created across the last decades. It is from the synergy of the (meta)data provided by such resources that we can draw the overall picture as the necessary condition to grasp the textual data, which in turn makes it possible to better understand their content and, ultimately, to yield new knowledge.

We are convinced that now is the time for the research area dealing with the development and distribution of linguistic resources for Latin to find a way to harmonize such differences in data and metadata, as a requirement raising both from data providers and from data users. Indeed, the lack of interoperability between resources prevents them from benefiting the large research community working in the broad area of the Humanities, which often deals with Latin texts, but is not provided with sufficient expertise to make distributed resources using different annotation schemes, tag sets and data formats interact. The result of such situation is that a large set of valuable linguistic resources for Latin, built with remarkable effort by data providers in long lasting projects, still remains unused (and sometimes unknown) by their reference community.

LiLa was launched just to overcome such state of affair. The first result of the project was to provide the LiLa KB with its very core component, i.e. the collection of Latin lemmas, which is used as the connecting point between the resources that LiLa wants to make interact. Once the lemma collection was ready, we started to include into the KB the first linguistic resources for Latin, among which is *LatinAffectus*. In the near future, we plan to include the *Latin Wordnet* ([13], [21]) so that, besides the prior polarity sentiment score provided by *LatinAffectus*, we will be able to assign a specific score to the single meanings of the words (assigned to different synsets of *WordNet*), by relying on previous work done for the *SentiWordNet* resource [1].

References

1. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: *Lrec.* vol. 10, pp. 2200–2204 (2010)
2. Buitelaar, P., Arcan, M., Iglesias, C.A., Sánchez-Rada, J.F., Strapparava, C.: 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.* pp. 1–8 (2013)
3. Buitelaar, P., Cimiano, P., Haase, P., Sintek, M.: Towards linguistically grounded ontologies. In: *European Semantic Web Conference.* pp. 111–125. Springer (2009)
4. Busa, R.: *Index Thomisticus: Sancti Thomae Aquinatis operum omnium indices et concordantiae in quibus verborum omnium et singulorum formae et lemmata cum suis frequentis et contextibus variis modis referuntur.* Index thomisticus: Sancti Thomae Aquinatis operum omnium indices et concordantiae, Frommann-Holzboog (1974)
5. Chen, Y., Skiena, S.: Building sentiment lexicons for all major languages. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).* pp. 383–389 (2014)
6. Davis, B. (ed.): *Thomas Aquinas. On Evil.* Oxford University Press, Oxford (2003)
7. Declerck, T.: Representation of Polarity Information of Elements of German Compound Words. In: *LDL 2016 5th Workshop on Linked Data in Linguistics: Managing, Building and Using Linked Language Resources.* p. 46 (2016)
8. Deferrary, R., Barry, I.: *A lexicon of St. Thomas Aquinas based on the Summa theologica and selected passages of his other works.* Catholic University of America Press, Washington (1948)
9. Denoos, J.: *Opera Latina: une base de données sur internet.* *Euphrosyne* **32**, 79–88 (2004)
10. Doerr, M.: The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine* **24**(3), 75–75 (2003)
11. Fang, X., Zhan, J.: Sentiment analysis using product review data. *Journal of Big Data* **2**(1), 5 (2015)
12. Fillmore, C.J.e.a.: *Linguistics in the morning calm.* Linguistics Society of Korea. Frame Semantics. Seoul: Hanshin (1982)
13. Franzini, G., Peverelli, A., Ruffolo, P., Passarotti, M., Sanna, H., Signoroni, E., Ventura, V., Zampedri, F.: *Nunc Est Aestimandum. Towards an Evaluation of the Latin WordNet.* In: *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019).* CEUR-WS. org (2019)

14. Khan, A.F.: Towards the Representation of Etymological Data on the Semantic Web. *Information* **9**(12), 304 (Dec 2018)
15. Lau, D.: *Der lateinische Begriff Labor*. Fink, Munich (1975)
16. Liu, B.: *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press (2015)
17. Mambrini, F., Passarotti, M.: Harmonizing Different Lemmatization Strategies for Building a Knowledge Base of Linguistic Resources for Latin. In: *Proceedings of the 13th Linguistic Annotation Workshop*. pp. 71–80. Association for Computational Linguistics, Florence, Italy (Aug 2019)
18. Mambrini, F., Passarotti, M.: Representing Etymology in the LiLa Knowledge Base of Linguistic Resources for Latin. In: Kernerman, I., Krek, S. (eds.) *Proceedings of Globalex Workshop on Linked Lexicography (GLOBALEX 2020)*. European Language Resources Association (ELRA), Paris, France (May 2020)
19. McCrae, J., Spohr, D., Cimiano, P.: Linking lexical resources and ontologies on the semantic web with lemon. In: *Extended Semantic Web Conference*. pp. 245–259. Springer (2011)
20. McCrae, J.P., Bosque-Gil, J., Gracia, J., Buitelaar, P., Cimiano, P.: The Ontolex-Lemon model: development and applications. In: *Proceedings of eLex 2017 conference*. pp. 19–21 (2017)
21. Minozzi, S.: Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'Information Retrieval. In: Mastandrea, P. (ed.) *Strumenti digitali e collaborativi per le Scienze dell'Antichità*, pp. 123–134. No. 14 in *Antichistica* (2017), <http://doi.org/10.14277/6969-182-9/ANT-14-10>
22. Mohammad, S.: 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)*. pp. 174–184 (2018)
23. Moro, A.: *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge University Press, Cambridge (1997)
24. Nakov, P., Rosenthal, S., Kiritchenko, S., Mohammad, S.M., Kozareva, Z., Ritter, A., Stoyanov, V., Zhu, X.: Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts. *Language Resources and Evaluation* **50**(1), 35–65 (2016)
25. Passarotti, M.C., Cecchini, F.M., Franzini, G., Litta, E., Mambrini, F., Ruffolo, P.: The LiLa Knowledge Base of Linguistic Resources and NLP Tools for Latin. In: *2nd Conference on Language, Data and Knowledge (LDK 2019)*. pp. 6–11. CEUR-WS.org (2019)
26. Passarotti, M.C.: The Project of the Index Thomisticus Treebank. In: Berti, M. (ed.) *Classical Philology. Ancient Greek and Latin in the Digital Revolution*, pp. 299–320. Berlin, Boston: De Gruyter (2019)
27. Simpson, D.P.: *Cassell's Latin dictionary*. Simon & Schuster Macmillan Company (1959)
28. Sprugnoli, R., Passarotti, M., Corbetta, D., Peverelli, A.: Odi et Amo. Creating, Evaluating and Extending Sentiment Lexicons for Latin. In: *Proceedings of LREC 2020* (2020)
29. de Vaan, M.: *Etymological Dictionary of Latin: and the other Italic Languages*. Brill, Amsterdam (2008), <https://brill.com/view/title/12612>
30. Westerski, A., Sánchez-Rada, J.F.: *Marl Ontology Specification, V1*. 0 May 2013 (2013)