Lifeless Winter without Break: Ovid's Exile Works and the LiLa Knowledge Base

Aurora Alagni[1,*], Francesco Mambrini[1] and Marco Passarotti[1]

[1]Università Cattolica del Sacro Cuore, Largo Gemelli 1, Milano, 20123, Italy

Abstract

In this paper we describe the process of semi-automatic annotation and linking performed to connect two works by the Latin poet Ovid to the LiLa Knowledge Base. Written after Ovid's exile from Rome, the *Tristia* and the *Epistulae ex Ponto* mark the beginning of the "literature of exile". In spite of their importance, no lemmatized version existed and the two collections were not part of the major annotated corpora linked to LiLa. The paper discusses the workflow used to annotate and publish the works as Linked Open Data connected to the LiLa Knowledge Base. On account of their subject and the emotional tone attached to the theme of exile, the two works are particularly relevant for sentiment analysis. We discuss some results of a lexicon-based analysis that is enabled by the interlinking with LiLa. We use LatinAffectus, a manually-generated polarity lexicon for Latin nouns and adjectives, to perform Sentiment Analysis on the aforementioned works and interpret the (replicable) results by consulting and simultaneously enriching the available literary scholarship with new information.

Keywords

Linked Open Data, Lemmatization, Latin, Sentiment Analysis, Humanities Computing

1. Introduction

The World Wide Web provides Latin scholars with a plethora of free, high-quality resources, issued from a long tradition of linguistic and philological study; many digital libraries, such as the Perseus Digital Library [1] or the Digital Latin Library [2], supply electronic and often machine-actionable versions of some of the most studied texts in world literature. In the last years, the CIRCSE Research Center has developed the LiLa Knowledge Base with the objective of making the distributed knowledge about Latin texts interoperable through the application of the principles of the Linked Data paradigm [3]. LiLa (presented below in sec. 3) now includes a number of lexicons and annotated corpora. In particular, the *Opera Latina* LASLA corpus, a manually lemmatized and morphosyntactically annotated corpus of more than 1.5 million words mainly belonging to Classical Latin literature that was recently added to LiLa [4], has significantly expanded the textual heritage within the LiLa Knowledge Base, which now provides a Linked Open Data (LOD) compliant edition of many widely studied literary works.

Publius Ovidius Naso (anglicized as Ovid, 43 BCE - 17 CE) is arguably one of the most influential writers in the history of Western literature. His mythological poem in 15 books (the *Metamorphoses*, written between 2 and 8 CE) has been a crucial source of inspiration for artists like Dante, Shakespeare, or Titian. His body of elegiac poetry of erotic subject won him immense popularity during his life and afterwards. In spite of his importance, the work of Ovid is not represented in full neither in the LiLa network, nor in any other annotated corpora. The LASLA corpus provides only his earlier works (*Ars Amatoria, Remedia Amoris, Medicamina, Amores, Heroids*) and other poems (*Fasti, Halieutica, Ibis*), while the annotation of the *Metamorphoses* is listed as "in progress".

Among the works that are utterly missing figure two of the last books of Ovid's career, the *Tristia* ("Sorrows" or "Lamentations", written between 9–13 CE) and the *Epistulae ex Ponto* ("Letters from the Black Sea", 12–17 CE, henceforth *Epistulae*) that were partly published after the poet's death. These two poetic collections center around Ovid's forced departure from Rome and exile to the town of Tomis (modern-day Constanța in Romania), at the furthest ends of the Roman empire. Despite his many attempts, Ovid would never come back from this "utmost part of an unknown world" (*extremis ignoti partibus orbis*, *Tr.* 3.3.3[1]) nor was he ever restored to his previous status. The two works are a fundamental source for the biography of the poet. Moreover, they are a foundational archetype of a peculiar sub-genre that is still influential in modern days, the "exile literature" [6].

Ovid's exilic works were banished from libraries, and although they survived, were often judged unfavorably by the critics [7, xxxvi]. The present study aims, in part, at revoking the ban that still seems to weigh on these

CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy

*Corresponding author.

†These authors contributed equally.

aurora.alagni01@icatt.it (A. Alagni);
francesco.mambrini@unicatt.it (F. Mambrini);
marco.passarotti@unicatt.it (M. Passarotti)
0000-0003-0834-7562 (F. Mambrini); 0000-0002-9806-7187 (M. Passarotti)

[1]All English translations are by Wheeler [5].

Ovidian poetic collections, allowing them to enter the LiLa network. In what follows we describe how we prepared a lemmatized and part-of-speech (POS) tagged version of the two poems and how we linked this edition to the network of textual and lexical resources for Latin connected to LiLa. Our work fills the significant gap created by the absence of the exilic works of Ovid from the available annotated corpora. In addition, it also links to LiLa two collections of poems that, on account of their subject, foreground the emotional tone, and were successful in shaping the conventions of exilic literature; these works established the literary codification of the psychological reactions to banishment, within a veritable poetics of exile. Their content and historical relevance make them ideal candidates for a computationally based study on the sentiment analysis of literary texts.

The paper is structured as follows. Section 2 reviews related work, with a specific focus on sentiment analysis within the field of Computational Literary Studies. Section 3 introduces the LiLa Knowledge Base and the language resources connected to it. Section 4 describes the workflow followed for the annotation, publication and linking of the works. Section 5 discusses the type of knowledge that can be gained by combining the data from *LatinAffectus*, a prior polarity lexicon of Latin included in LiLa, and the newly prepared edition of the works, for a lexicon-based approach to their sentiment. Section 6 presents the conclusions and discusses plans for future work.

2. Related Work

Sentiment analysis (SA) is the field of study that analyses people's opinions, sentiments, appraisals, attitudes, and emotions toward entities and their attributes expressed in written text [8]. Considering that opinions have now a fundamental role in everyday life, SA is not just an object of research in the field of NLP, but also in business, economic, political, even medical domains. Indeed, sentiment analysis has numerous applications[2], ranging from investigating product reviews to enhance product development [10], analysing news related to the stock market to predict price trends [11], monitoring social media to forecast election outcomes [12], and evaluating public health through tweets about patient experiences [13].

Furthermore, sentiment analysis has recently emerged as one of the most discussed topics within the realm of Computational Literary Studies[3]. This rise in prominence coincides with the so-called "affective turn" in the humanities and social sciences, which has fostered renewed engagement with emotion [15]. However, there remain significant limitations in the application of sentiment analysis within Computational Literary Studies, two of which are addressed in this paper.

First, while the World Wide Web and social media represent an ostensibly infinite repository of emotions, annotated corpora of literary texts are still infrequently available. This is especially true for classical languages. As previously mentioned and will be further illustrated in this paper, this limitation can be mitigated through the development and dissemination of interoperable resources. To our knowledge, there are only a few experiments conducted in classical languages. Sprugnoli et al. [16] evaluated two distinct approaches to automatic polarity classification of eight odes by the Latin author Horace: a lexicon-based approach, grounded in the first version of LatinAffectus, and a zero-shot classification method. Sprugnoli et al. [17] present an example of how to use interoperable resources to analyse the sentiment value of the Latin epistles by Dante Alighieri, employing SPARQL queries that access an extended version of LatinAffectus, the LiLa Knowledge Base, and UDante. Pavlopoulos et al. [18] annotated the sentiment of a modern Greek translation of the first book of the Iliad and demonstrated that a fine-tuned version of GreekBERT can achieve a low error rate. Zhao et al. [19] proposed a model based on transfer learning to classify a dataset of Tang Dynasty Chinese poems and compared the sentiment analysis results with social history analysis. After constructing a sentiment lexicon for Classical Chinese poetry, Hou et al. [20] evaluated it both intrinsically and extrinsically, highlighting that their analysis results align with the main findings established in Classical Chinese literary studies.

Second, although sentiment analysis in the field of Computational Literary Studies is employed to address questions related to literary theory, the results often lack connection to a rigorous analysis, focusing solely on performance metrics. The aforementioned studies exemplify this tendency, particularly since only those conducted on Classical Chinese take literary studies into account. Rarely do they contribute to advancements in literary criticism, an area that could greatly benefit from clear and reproducible results, considering that it typically relies on the intuition of critics. This issue has been highlighted by Rebora [21], who notes that while the strongest connection between literary theory and sentiment analysis occurs in the field of narratology, the actual points of intersection reveal themselves to be problematic and based on questionable assumptions. This paper will also address these concerns, as the results of sentiment analysis conducted on Ovid's exilic works are closely intertwined with the literary scholarship surrounding those texts. Although our findings may not be generalisable due to their

[2] See Wankhade et al. [9] for an in-depth overview of the applications of sentiment analysis, as well as the methods for conducting this task.

[3] For an extensive survey on sentiment and emotion analysis applied to literature, see the paper by Kim and Klinger [14].

basis in a small, yet highly controlled dataset, our method is clearly reproducible and shareable.

3. Latin resources in LiLa

LiLa is a network of interconnected language resources for Latin aimed at insuring interoperability between corpora, lexicons and natural language processing (NLP) tools. To pursue its goal, it adopts the Linked Data paradigm. At the heart of the project, the interlinking between the different components is ensured by the Lemma Bank [22], a collection of canonical forms (lemmas) that can be used to lemmatize texts and index entries in dictionaries. Each lemma of the Lemma Bank is provided with a unique identifier, in the form a URL resolvable on the World Wide Web, and described by a series of properties modeled with the help of OWL ontologies for Linguistic LOD, such as `Ontolex` [23, 45-59].

Currently, the Lemma Bank includes 226,775 canonical forms, which are used to link 14 lexical resources and 7 corpora. The latter include collections of texts from different times and genres (from the works of Medieval authors like the mathematician Fibonacci [24], Thomas Aquinas [25] or Dante Alighieri [26], to inscriptions from various areas of the Roman Empire [27]). The largest collection of Classical literary texts is provided by the *Opera Latina*, a manually crafted corpus with morphological annotation and lemmatization developed since the 1960s by the LASLA laboratory of the University of Liège. The LASLA corpus (which is still in development) includes 131 Latin works by 19 authors, ranging chronologically from Plautus (c. 254 – 184 BC) to Juvenal (55 – 128 CE). As said, however, even such comprehensive collection does not cover the whole extant production, also for some of the major authors within that time span; Ovid's exilic words are a prominent example of missing texts. To fill the gaps in LASLA, and widen the chronological span of ancient authors to the end of the Roman era in the 6th Century CE, the CIRCSE has launched a new collection (natively linked to LiLa) called the "CIRCSE Latin Library"[4].

Among the lexical resources produced within LiLa[5], LatinAffectus [28] is a manually generated polarity lexicon of Latin adjectives and nouns. The lexicon was designed to support research in Sentiment Analysis (SA) [8], an approach to the linguistic and literary studies of ancient texts that, although still in its infancy, is gaining growing recognition [18][16].

In its latest version, LatinAffectus contains 6,018 lemmas, 2,216 adjectives and 3,802 nouns, to which numerical values expressing their prior polarity, that is their sentiment orientation regardless of the context of use [8], have been associated. The classification adopts five numeric values: -1.0 (fully negative, as e.g. *uulnus*, "wound"), -0.5 (negative, *grauis*, "serious"), 0 (neutral, *ianua*, "door"), +0.5 (positive, *ius*, "justice"), +1.0 (fully positive, *pietas*, "devotion").

In the second part of this paper (Sec. 5) we will make use of data from LatinAffectus to perform lexicon-based Sentiment Analysis of Ovid's exilic works. The results obtained from the SA conducted on the *Tristia* and the *Epistulae*, clear and reproducible, and their interpretation carried on in light of the previous results of literary criticism on the subject allowed us to investigate the evolution of Ovid's poetic journey (Sec. 5.1) and the decline of relationships with those left behind in Rome (Sec. 5.2).

4. Ovid's exile works as LOD

The *Tristia* are a collection of 50 poems in elegiac meter (i.e. couplet of lines with an hexameter followed by a pentameter) divided into 5 books. The *Epistulae* include 46 letters in elegiac couplets divided into 4 books. The poetry in both works mixes the themes of lamentation over the exile and the desperate plead (*peroratio*) directed towards the loved ones and potential allies in Rome.

The starting point of our edition was a plain-text version of the two works, which we obtained from The Latin Library[6]. The two works consists of a total of 43,438 tokens (without punctuation), and 3,061 sentences. Few preprocessing operations were performed over the texts, namely the addition of three missing lines, which were omitted by mistake in the original source (*Tr.* 3.10.44 and 52, *Tr.* 5.12.50), the correction of evident transcription errors (most likely due to OCR issues, e.g. *virunique* for *virumque*, *Tr.* 2.372), the standardization of capitalization usage, and the adoption of the "u" character even for the voiced labiodental fricative [v], following the convention adopted in the LiLa Lemma Bank.

Tokenization, sentence splitting, lemmatization and POS tagging were performed automatically by the LiLa Text Linker, a POS-tagger and lemmatizer for the Latin language developed as one of the user-dedicated services of LiLa that also links the output of the NLP operations to the entries in the Lemma Bank [29]. For POS-tagging and lemmatization the Text Linker uses a custom-trained UDPipe model (as documented in [29]). The output of the tasks performed automatically was systematically reviewed and manually corrected by one annotator adopting a scholarly annotation approach [30]. 42 tokenization errors were identified (on average between 4 and 5 per book), often due to a failure to segment punctuation (e.g. the sequence *legent?* in *Tr.* 5.1.94).

[4] http://lila-erc.eu/data/corpora/CIRCSELatinLibrary/id/corpus.

[5] For a complete list of the resources currently linked to LiLa, see: https://lila-erc.eu/data-page/. Please note that all LiLa's resources are assigned DOIs registered through Zenodo and are also available in CLARIN.

[6] http://www.m.thelatinlibrary.com/ovid.html.

Table 1
Accuracy of POS tagging and lemmatization per book of *Epistulae* and *Tristia* as performed by the LiLa Text Linker

		Accuracy	
Book	Nr. of tokens	POS Tagging	Lemmatization
Ep. 1	5,923	0.95	0.93
Ep. 2	5,770	0.97	0.94
Ep. 3	5,671	0.97	0.95
Ep. 4	7,099	0.97	0.94
Tr. 1	5,805	0.96	0.94
Tr. 2	4,427	0.96	0.93
Tr. 3	6,214	0.96	0.95
Tr. 4	5,311	0.97	0.95
Tr. 5	5,980	0.96	0.94
TOT	52,200	0.94	0.96

Table 2
Evaluation of POS tagging for the 11 tags with support > 1,000 tokens

POS-Tag	Precision	Recall	F1-score	Support
VERB	0.98	0.97	0.97	10,960
NOUN	0.96	0.97	0.96	10,626
PUNCT	1.00	1.00	1.00	8667
ADJ	0.95	0.90	0.92	4,702
ADV	0.96	0.95	0.95	3,955
DET	0.95	0.99	0.97	3,836
PRON	0.99	0.93	0.96	3,276
CCONJ	0.99	0.99	0.99	1,698
ADP	0.96	0.99	0.98	1,625
PROPN	0.79	0.90	0.84	1,353
SCONJ	0.88	0.94	0.91	1,304

The accuracy score reached by the model of the LiLa Text Linker are reported in table 1[7]. As it can be seen, the tool performed quite satisfactorily in both tasks, reaching an average accuracy across the different books of the two works of 96% and 94% respectively. Accurate lemmatization also lead to good scores for the linking process, with approximately 87% of the word forms uniquely associated with one lemma. Of the remaining lemmas, 10% were ambiguous, as they were associated with two or more potential candidates in LiLa, mainly due to homography (e.g. the lemma string *volo* can be linked to both the first-conjugation verb *volare*, "to fly" and the irregular verb *volere*, "will"), and required manual disambiguation.

Of the 3% of no-matches, most were proper names. Ovid mentions barbarian tribes and figures belonging to Roman cultural circles rarely or never cited elsewhere. In the fourth book of the *Epistulae*, out of a total of 42 tokens not linked to any lemma, 32 are proper names (e.g. the Thracian tribe of the "Corallis", *Ep.* 5.2.37, or the unknown poet "Marius", mentioned in *Ep.* 4.16.24). Table 2 shows the performances of the POS-tagger for the 12 out of 17 tags that were used more than 1,000 times[8]. With an F1-score sensibly under 90%, proper nouns (PROPN) is the most challenging class for the model to predict.

All tasks (tokenization, POS-tagging, lemmatization and linking) are closely interconnected: an error in tokenization inevitably leads to an error in lemmatization and POS tagging, which then causes a wrong or missing linking. For example, 18 forms of the verb *addo*, "to add", in the second person singular imperative, *adde*, "add", were mislabeled as proper nouns (PROPN), and thus assigned to a nonexistent lemma "Ads". Once disambiguations and corrections were performed, the digital editions of the *Tristia* and the *Epistulae* were prepared and published as Linked Data, as part of the "CIRCSE Latin Library"[9].

5. Sentiment analysis and Ovid's exile works

Thanks to the work performed in the linking process, each token of the two exilic poems is now connected to the respective lemma within the Lemma Bank via a dedicated property (`hasLemma`)[10] defined in the OWL ontology of the LiLa project [3]. As the lemma's URI is the same that is used as canonical form for the entries of LatinAffectus, this step effectively enables users to cross-check the textual information within the two works and the scores recorded in the prior polarity lexicon.

Following the same methodology discussed in Sprugnoli et al. for Horace [16, 61-2], we proceeded to match each token of *Tristia* and *Epistulae* to the polarity score recorded in LatinAffectus for their respective lemma. The sentiment scores are obtained by automatically assigning the score found in LatinAffectus to the tokens that are lemmatized under lemmas that also have an entry in the polarity lexicon. For instance, the adjective *malus* "bad" is found with a polarity value of -1.0 in LatinAffectus. All tokens lemmatized as *malus* (adj.) are thus given a score of -1.0. A score of 0.0 is assigned to both words expressly annotated as neutral in LatinAffectus and to those that do not have an entry in the lexicon. The coverage of polarity-laden tokens (both adjectives and nouns) is reported in table 3.

[7] Note that, in the evaluation, we omitted the 3 missing lines that were added in the revision stage. For this reason table 1 has slightly fewer tokens than table 3.

[8] The model uses the Universal POS tagset of Universal Dependencies; see: https://universaldependencies.org/u/pos/index.html.

[9] http://lila-erc.eu/data/corpora/CIRCSELatinLibrary/id/corpus/P.%20Ovidii%20Tristia and http://lila-erc.eu/data/corpora/CIRCSELatinLibrary/id/corpus/P.%20Ovidii%20Epistulae%20ex%20Ponto.

[10] http://lila-erc.eu/ontologies/lila/hasLemma.

Table 3
Token coverage of polarity-laden nouns and adjectives in the books of *Epistulae* and *Tristia*. Per each book, the total nr. of adj. and nouns are reported, as well as the nr. of adj./nouns with polarity score $\neq 0$ (pos/neg)

Book	Nouns		Adjectives		Tot Tokens
	tot	pos/neg	tot	pos/neg	
Epistulae.b1	1,195	1,061	545	360	5,923
Epistulae.b2	1,214	1,088	561	425	5,770
Epistulae.b3	1,135	1,013	452	335	5,671
Epistulae.b4	1,447	1,282	676	464	7,099
Tristia.b1	1,153	995	513	358	5,805
Tristia.b2	922	817	386	268	4,427
Tristia.b3	1,272	1,131	555	386	6,227
Tristia.b4	1,140	1,020	493	353	5,311
Tristia.b5	1,152	1,037	523	386	5,989
TOT	10,630	9,444	4,704	3,335	52,222

In what follows, due to space constraints, only some of the results obtained from the sentiment analysis conducted on Ovid's exilic works will be discussed. In analysing these results, we will focus on the distribution of sentiment-laden words and what this reveals about Ovid's emotional state during his exile.

5.1. Ovid's "last metamorphosis"

To investigate how Ovid's attitude evolves throughout his exilic works, we calculated the overall sentiment for each book (fig. 1). Specifically, we summed the polarity scores and divided the total by the number of sentences to mitigate skewness resulting from the varying lengths of the books [17][11]. This book-level score reveals a negative emotional state persisting until the first book of the *Epistulae*. From the second book onward, however, the sentiment undergoes a polarity shift, becoming positive and remaining so until the last book. The reasons behind such a radical change in the poet's emotional state are worth investigating. Ovid's polarity lexicon, that is, the most frequently used sentiment-laden words in the exilic works, does not show any particular change in the 9 books considered here. An interesting change that we do observe in the last books concerns the distribution of the personal pronouns. In *Epistulae* 1, the relative frequency of the 1st p. singular pronoun, *ego*, is 0.018 (93 over 4,983 lemmas), while for the second person singular pronoun, *tu*, it is 0.010 (52 occurrences). In *Epistulae* 2, the former has an identical relative frequency (0.018, or 89 occurrences over 4,920), while the latter increases significantly, reaching 0.020 (99 occurrences). The focus of Ovidian epistles seems to split, with the once uncontested domain of the "I" beginning to be accompanied by the equally large realm of the "you". The solipsism of the sender starts to giving way to the celebration of the recipient, transmuting the once famous and now banished elegiac poet into a potential celebratory poet, who could exceptionally glorify his future patron if only he is given the chance to (and after, of course, being recalled back home).

Commentators have never doubted that Ovid, after some attempts in the third book (e.g. *Ep.* 3.4-5), dedicates himself to panegyric poetry in the fourth book, no doubt in order to win powerful allies who could intercede for his return [31, 120-121] [32]. However, this intention was never noted or at least imagined for the *Epistulae*'s second book.

It is undeniable that we witness the last metamorphosis in the poetic trajectory of Ovidian elegy. Our results suggest that this metamorphosis, still so premature that it has not been detected by critics, is clearly recorded by sentiment analysis already in the second book of the *Epistulae*. Indeed, when the sentiment analysis is conducted at a finer grain, and thus at the level of individual compositions , it reveals an increase in positivity precisely in the verse-epistles sent to new and powerful recipients. This reflects a new poetic purpose for Ovid's poetry.

5.2. Facing the abandonment

Another advantage of lexicon-based SA is the possibility to directly engage with a list of sentiment words mostly used by an author in their entire production or in specific works of interest. A close observation of this specialized lexicon can lead to interesting outcomes too.

The sentiment words used in the exilic works are relatively stable in quality and quantity. Five distinct semantic spheres [33, 203] can be identified: friendship, politics, justice, intellect, and sadness (fig. 2). Among

[11] Ovid's sentences tend to correspond with the elegiac couplet. The two works have 3,4044 sentences with an average length of 17.16 tokens (stdev = 11.43). The books tend to have a rather similar number of sentences, ranging from 261 (*Tr.* 2) to 388 (*Ep.* 4), with a mean length of 338.22 (stdev = 37.74). Note, however, that we relied on the sentence splitter of TextLinker and the results were not corrected manually.

these, the semantic sphere of friendship and love contains abstract qualities and feelings (*amor* "love", *fides* "trust, faith", *honor* "honor", *nobilitas* "nobility", *pietas* "devotion", *virtus* "virtue"), as well as nouns and qualifying adjectives typical of friendly and romantic relationships (*bonus* "good", *carus* "dear", *dignus* "worthy", *pius* "dutiful, affectionate"). Although this sphere is frequently recurring throughout the exilic production, the words composing it do not appear with the same consistency. Between the third and fourth book of the *Tristia*, new lemmas become part of this semantic sphere, indicating a change in Ovid's relationship with the affections left behind in Rome.

In *Tristia* 3, the only epithets fitting for his friends (lemma *amicus*, 10) were "dear" (*carus*, 10) and "good" (*bonus*, 6). These friends, along with the wife, represented Ovid's only hope of salvation. In *Tristia* 4, Ovid reaches the fourth year of exile and sees the possibility of relying on them slipping further out of his grasp. The poet begins to perceive that the friendship and love shown to him in Rome and at the height of his success might have been more superficial than he believed. His friends fail to write (*Tr.* 4.7.3-5) and Ovid catches himself wondering if his wife still thinks of him (*Tr.* 4.3.10). However, the bonds of friendship and marriage could still be exploited.

In *Tristia*'s book 4, as the occurrences of the adjectives "friend" (*amicus*, 3) and "dear" (*carus*, 3) decrease, the use of words such as "devoted", "virtuous", "worthy", and "husband" increases. This lexicon here suggests a form of conditional praise: only by proving themselves worthy of the friend and spouse in need can those left in Rome earn their title. Thus, if his friends are truly "virtuous" (*bonus*, 8) and "devoted" (*pius*, 6) and wish their "fame" (*fama*, 7) to be such among contemporaries and posterity, they must show themselves worthy of such a connotation. His wife must, similarly, prove herself worthy of being his husband's (*vir*, 13) wife, even though he is exiled. Consequently, Ovid would sooner credit a ten-verses long series of *adynata* rather than believe that his friend decided to abandon him (*Tr.* 4.7.10-20). At the same time, his wife, dutiful as she is (*Tr.* 4.3.71), surely must be existing solely to work for and diligently lament her absent husband (*Tr.* 4.3.17-38). Moreover, his misfortune gives her a unique chance for fame, for her loyalty to be forever remembered (*Tr.* 4.3.81-84). This logic of coercion begins to be employed in book 4 of the *Tristia*, and finds full employment in the *Epistulae*. It consists of imposing fundamental moral models and values of the Roman citizen on his recipients through targeted praises, so that the recipients feel obliged to comply with the requests. Here too, sentiment analysis reveals in its embryonic state what the critical eye has only caught later in full development.

6. Conclusion and future work

The work that we presented in the paper had two outcomes. Firstly, our LOD edition of Ovid demonstrates the benefits of interoperability among resources for Latin. Interoperability greatly facilitates the work of scholars, allowing them to benefit from lexicon, corpora, and NLP tools useful for every stage of their research through a single point of access. The LiLa project already provides a paradigm of this model, but to continue doing so, it requires constant integration. This is true not only for corpora, whose enrichment this paper testifies to. Despite the important results that SA conducted with LatinAffectus already provides for Ovid, there remains several ways for enhancing its performance. The coverage of LatinAffectus is extensive with regard to nouns and adjectives, as clearly demonstrated by its performance on the dataset discussed in this paper (see table 3). However, it is evident that a current limitation is its failure to account for the sentiment of verbs. This is why LatinAffectus, like the other linguistic resources available in LiLa, should not be regarded as a static resource, but rather as one that is continually evolving and being updated. Additionally, improvements could be made by accounting for syntactic phenomena such as polarity shifters [16] and by taking into consideration the poetic nature of the text (e.g. by providing access to metrical information[12]). In a broader sense, there is a lack of sufficient consideration for the context in which sentiment words are collocated. However, context-sensitive sentiment analysis is still in its early stages within NLP[13], and clearly, much work remains to be done to effectively incorporate context into sentiment analysis.

The second outcome is in suggesting the undeniable potential of a hybrid approach, such as the one employed in this study, crossing literary criticism with the use of quantitative methods and computational resources. The theories developed within literary criticism and the investigative tools provided by computational linguistics can and should effectively collaborate, mutually enriching each other. In this specific context, the reflections developed within literary criticism regarding Ovid's exile works were crucial for interpreting the data derived from sentiment analysis. In turn, sentiment analysis was fundamental for confirming and deepening these observation, providing interpretable and reproducible data.

If a classic is a book which has never exhausted all it has to say to its readers (as Calvino wrote [35, 5]), it is also because scholars are capable of interrogating it with new methods to address longstanding and unresolved questions.

[12] For instance, this can be achieved by linking existing resources, such as Musisque Deoque, to LiLa.

[13] See Teng et al. [34] paper for an overview of state-of-the-art studies on context-sensitive sentiment analysis.

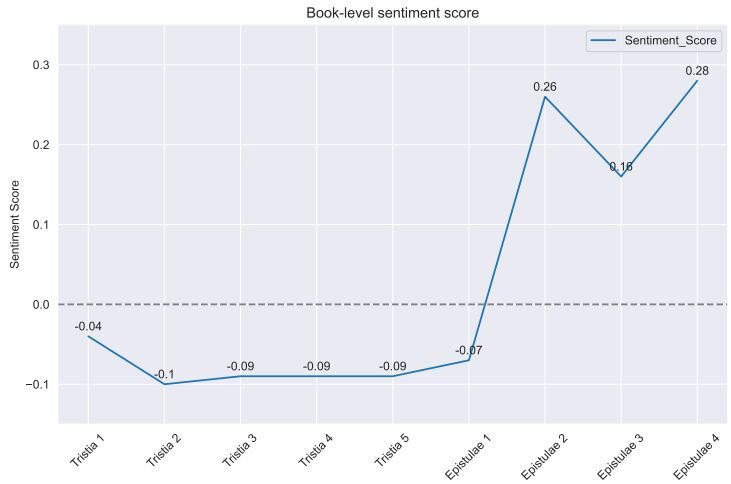

Figure 1: Ovid's overall sentiment (i.e. sum of all polarity words in each book divided by the number of sentences in each book) across the 5 books of the *Tristia* and the 4 books of the *Epistulae*.

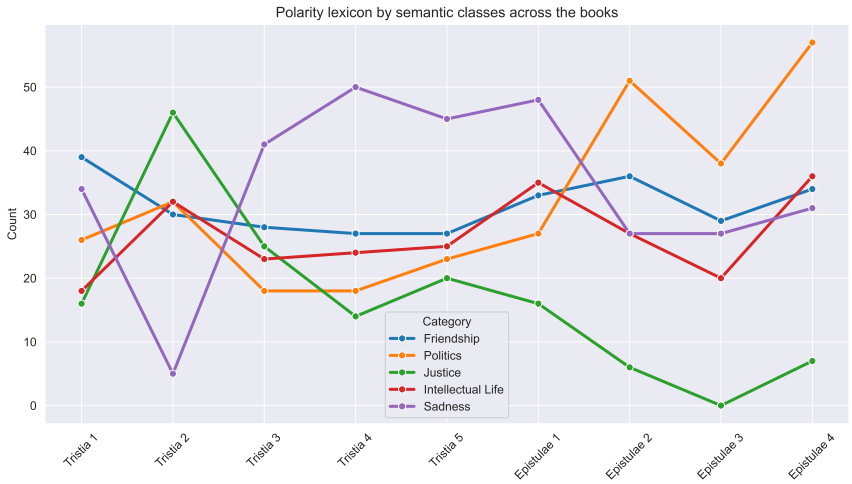

Figure 2: Distribution of polarized words according to semantic class across the 5 books of the *Tristia* and the 4 books of the *Epistulae*.

Appendix

The appendix contains the figures cited in section 5.

References

[1] G. Crane, The perseus digital library and the future of libraries, International Journal of Digital Libraries 24 (2024) 117–128. URL: https://doi.org/10.1007/s00799-022-00333-2.

[2] S. J. Huskey, The digital latin library: Cataloging and publishing critical editions of latin texts, in: M. Berti (Ed.), Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution, De

Gruyter, Berlin, Boston, 2019, pp. 19–34. doi:doi:10.1515/9783110599572-003.

[3] M. Passarotti, F. Mambrini, G. Franzini, F. M. Cecchini, E. Litta, G. Moretti, P. Ruffolo, R. Sprugnoli, Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin, Studi e Saggi Linguistici 58 (2020) 177–212. doi:10.4454/ssl.v58i1.277, number: 1.

[4] M. Fantoli, M. Passarotti, F. Mambrini, G. Moretti, P. Ruffolo, Linking the LASLA Corpus in the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin, in: Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 26–34.

[5] A. L. Wheeler, Publius Ovidius Naso. Tristia. Ex Ponto, Harvard University Press, Cambridge, MA, 1959.

[6] J.-M. Claassen, Ovid revisited: The poet in exile, Bloomsbury Academic, London, 2008.

[7] P. Green, Ovid. The Poems of Exile. Tristia and the Black Sea Letters, University of California Press, Berkeley, 2005.

[8] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, 2nd edition ed., Cambridge University Press, Cambridge ; New York, 2020.

[9] M. Wankhade, A. C. S. Rao, C. Kulkarni, A survey on sentiment analysis methods, applications, and challenges, Artif. Intell. Rev. 55 (2022) 5731–5780. URL: https://doi.org/10.1007/s10462-022-10144-1. doi:10.1007/s10462-022-10144-1.

[10] R. Bose, R. Dey, S. Roy, D. Sarddar, Sentiment Analysis on Online Product Reviews, 2018.

[11] F. Xing, E. Cambria, R. Welsch, Natural language based financial forecasting: a survey, Artificial Intelligence Review 50 (2018). doi:10.1007/s10462-017-9588-9.

[12] B. O'Connor, R. Balasubramanyan, B. Routledge, N. Smith, From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series, Proceedings of the International AAAI Conference on Web and Social Media 4 (2010) 122–129. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14031. doi:10.1609/icwsm.v4i1.14031, number: 1.

[13] E. M. Clark, T. A. James, C. A. Jones, A. Alapati, P. Ukandu, C. M. Danforth, P. S. Dodds, A sentiment analysis of breast cancer treatment experiences and healthcare perceptions across twitter, ArXiv abs/1805.09959 (2018). URL: https://api.semanticscholar.org/CorpusID:44063573.

[14] E. Kim, R. Klinger, A Survey on Sentiment and Emotion Analysis for Computational Literary Studies, Zeitschrift für digitale Geisteswissenschaften (2019). URL: http://arxiv.org/abs/1808.03137. doi:10.17175/2019_008, arXiv:1808.03137 [cs].

[15] P. C. Hogan, B. J. Irish, L. P. Hogan (Eds.), The Routledge Companion to Literature and Emotion, Routledge, London, 2022. doi:10.4324/9780367809843.

[16] R. Sprugnoli, F. Mambrini, M. Passarotti, G. Moretti, The Sentiment of Latin Poetry. Annotation and Automatic Analysis of the Odes of Horace, IJCoL. Italian Journal of Computational Linguistics 9 (2023). doi:10.4000/ijcol.1125.

[17] R. Sprugnoli, M. C. Passarotti, M. Testori, G. Moretti, Extending and using a sentiment lexicon for latin in a linked data framework, 2021. URL: https://api.semanticscholar.org/CorpusID:248149526.

[18] J. Pavlopoulos, A. Xenos, D. Picca, Sentiment Analysis of Homeric Text: The 1st Book of Iliad, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 7071–7077. URL: https://aclanthology.org/2022.lrec-1.765.

[19] H. Zhao, B. Wu, H. Wang, C. Shi, Sentiment analysis based on transfer learning for Chinese ancient literature, in: 2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2014), 2014, pp. 1–7. URL: https://ieeexplore.ieee.org/document/7059510. doi:10.1109/BESC.2014.7059510.

[20] Y. Hou, A. Frank, Analyzing Sentiment in Classical Chinese Poetry, in: K. Zervanou, M. van Erp, B. Alex (Eds.), Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), Association for Computational Linguistics, Beijing, China, 2015, pp. 15–24. URL: https://aclanthology.org/W15-3703. doi:10.18653/v1/W15-3703.

[21] S. Rebora, Sentiment Analysis in Literary Studies. A Critical Survey, Digital Humanities Quarterly 17 (2023). URL: https://www.proquest.com/scholarly-journals/sentiment-analysis-literary-studies-critical/docview/2842908301/se-2?accountid=9941, place: Providence.

[22] F. Mambrini, M. C. Passarotti, The lila lemma bank: A knowledge base of latin canonical forms, Journal of Open Humanities Data (2023). doi:10.5334/johd.145.

[23] P. Cimiano, C. Chiarcos, J. P. McCrae, J. Gracia, Linguistic Linked Data: Representation, Generation and Applications, Springer International Publishing, Cham, 2020. doi:10.1007/

978-3-030-30225-2.

[24] F. Grotto, R. Sprugnoli, M. Fantoli, M. Simi, F. M. Cecchini, M. C. Passarotti, The annotation of Liber Abbaci, a domain-specific latin resource, in: Proceedings of the Eighth Italian Conference on Computational Linguistics (CLiC-it 2021), aAccademia University Press, Milan, 2021, pp. 176–183. URL: https://doi.org/10.4000/books.aaccademia.10659.

[25] F. Mambrini, M. Passarotti, G. Moretti, M. Pellegrini, The Index Thomisticus Treebank as Linked Data in the LiLa Knowledge Base, in: C. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, S. Odijk, Janand Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022), European Language Resources Association (ELRA), Marseille, France, 2022, pp. 4022–4029. URL: https://aclanthology.org/2022.lrec-1.428.

[26] F. M. Cecchini, R. Sprugnoli, G. Moretti, M. Passarotti, UDante: First Steps Towards the Universal Dependencies Treebank of Dante's Latin Works, in: J. Monti, F. Dell'Orletta, F. Tamburini (Eds.), Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020, Bologna, Italy, March 1–3 2021), Associazione italiana di linguistica computazionale (AILC), Accademia University Press, Turin, Italy, 2020, pp. 99–105. URL: http://ceur-ws.org/Vol-2769/paper_14.pdf.

[27] I. De Felice, L. Tamponi, F. Iurescia, M. Passarotti, Linking the corpus classes to the lila knowledge base of interoperable linguistic resources for latin, in: Proceedings of CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 — Dec 02, 2023, CEUR Workshop Proceedings, Venice, 2023, pp. 1–7. URL: https://ceur-ws.org/Vol-3596/paper20.pdf.

[28] R. Sprugnoli, M. Passarotti, D. Corbetta, A. Peverelli, Odi et Amo. Creating, Evaluating and Extending Sentiment Lexicons for Latin., in: Proceedings of the Twelfth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 3078–3086.

[29] M. Passarotti, F. Mambrini, G. Moretti, The services of the LiLa knowledge base of interoperable linguistic resources for Latin, in: Proceedings of the 9th Workshop on Linked Data in Linguistics @ LREC-COLING 2024, ELRA and ICCL, Torino, Italia, 2024, pp. 75–83.

[30] D. Bamman, F. Mambrini, G. Crane, An Ownership Model of Annotation: The Ancient Greek Dependency Treebank, in: Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories, EDUCatt, Milan, Italy, 2009, pp. pp. 5–15.

[31] J.-M. Claassen, Displaced Persons: The Literature of Exile from Cicero to Boethius, Duckworth, 1999. Google-Books-ID: 1FkXAQAAIAAJ.

[32] M. Labate, Elegia triste ed elegia lieta. Un caso di riconversione letteraria, Materiali e discussioni per l'analisi dei testi classici (1987) 91–129. doi:10.2307/40235896.

[33] M. C. Gaetano Berruto, La linguistica. Un corso introduttivo, 3. edizione ed., UTET Università, [Grugliasco], 2022.

[34] Z. Teng, D. T. Vo, Y. Zhang, Context-Sensitive Lexicon Features for Neural Sentiment Analysis, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 1629–1638. URL: http://aclweb.org/anthology/D16-1169. doi:10.18653/v1/D16-1169.

[35] I. Calvino, Why read the classics? Perchè leggere i classici?, Penguin., Londra, 2009.