Sentiment Analysis of Latin Poetry: First Experiments on the Odes of Horace

Rachele Sprugnoli, Francesco Mambrini, Marco Passarotti, Giovanni Moretti

CIRCSE Research Centre, Università Cattolica del Sacro Cuore

Largo Agostino Gemelli 1, 20123 Milano

{rachele.sprugnoli,francesco.mambrini
marco.passarotti,giovanni.moretti}@unicatt.it

Abstract

In this paper we present a set of annotated data and the results of a number of unsupervised experiments for the analysis of sentiment in Latin poetry. More specifically, we describe a small gold standard made of eight poems by Horace, in which each sentence is labeled manually for the sentiment using a four-value classification (positive, negative, neutral and mixed). Then, we report on how this gold standard has been used to evaluate two automatic approaches for sentiment classification: one is lexicon-based and the other adopts a zero-shot transfer approach.¹

1 Introduction

The task of automatically classifying a (piece of) text according to the sentiment conveyed by it, known as Sentiment Analysis (SA), is usually performed for purposes such as monitoring contents of social media or evaluating customer experience, by analysing texts like tweets, comments, and micro-blogs.

A still under-investigated yet promising research area where developing and applying SA resources and techniques is the study of literary texts written in historical and, particularly, Classical languages (e.g. Ancient Greek and Latin). Actually, investigating the lexical properties of Classical literary texts is a century-long common practice. However, such investigation can nowadays (1) lead to replicable results, (2) benefit from techniques developed for analysing the sentiment conveyed by any type of text and (3) be performed with freely available lexical and textual resources. As for the latter, the research area dedicated to building and using linguistic resources for Classical languages has seen a substantial growth during the last two decades (Sprugnoli and Passarotti, 2020). For what concerns SA, we recently built a polarity lexicon for Latin nouns and adjectives, called LatinAffectus. The current version of the lexicon includes 4,125 Latin lemmas with their corresponding prior polarity value (Sprugnoli et al., 2020b). LatinAffectus was developed in the context of the LiLa: Linking Latin project (2018-2023)² (Passarotti et al., 2020) which aims at building a Knowledge Base of linguistic resources for Latin based on the Linked Data paradigm, i.e. a collection of several data sets described using the same vocabulary of knowledge description and linked together. LatinAffectus is connected to the Knowledge Base, thus making it interoperable with the other linguistic resources linked so far to LiLa (Sprugnoli et al., 2020a).

In this paper we describe the use of *LatinAf-fectus* to perform SA of the *Odes* (*Carmina*) by Horace (65 - 8 BCE). Written between 35 and 13 BCE, the *Odes* are a collection of lyric poems in four books. Following the models of Greek lyrical poets like Alcaeus, Sappho, and Pindar, the *Odes* cover a wide range of topics related to the individual and social life in Rome during the age of Augustus, like love, friendship, religion, morality, patriotism, the uncertainty of life, the cultivation of tranquility and the observance of moderation. In spite of a rather lukewarm initial reception, the *Odes* quickly became a capital source of influence, in particular as a model of authorial voice and

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

¹This 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, 4.2, 5; Marco Passarotti is responsible for Section 1; Francesco Mambrini is responsible for Section 4.1. Giovanni Moretti developed the zero-shot classification script.

²https://lila-erc.eu

identity.³ Considering not only the importance of the *Odes* in the history of Latin and European literature, but also the diversity of the contents and tones of the poems collected therein, we argue that performing SA on such work can lead to interesting results and might represent a use case to open a discussion about the pros and cons of applying SA techniques and resources to literary texts written in ancient languages.

All data presented in this paper are publicly released: https://github.com/CIRCSE/La tin_Sentiment_Analysis.

2 Related Work

The majority of linguistic resources and applications in the field of SA involve non-literary and non-poetic texts, such as news and user-generated content on the web (Medhat et al., 2014). However, affective information plays a crucial role in literature and, in particular, in poetry where authors try to provoke an emotional response in the reader (Johnson-Laird and Oatley, 2016). Annotated corpora of poems and SA systems specifically designed for poetry are not as numerous as those in other areas of research, first of all that of social media, but works have been carried out for several languages,⁴ including Arabic (Alsharif et al., 2013), Spanish (Barros et al., 2013), Odia (Mohanty et al., 2018), German (Haider et al., 2020), Classical Chinese (Hou and Frank, 2015) and, of course, English (Sheng and Uthus, 2020; Sreeja and Mahalakshmi, 2019).

Available annotated corpora of poems differ from each other from at least four points of view: annotation procedure (either involving experts or using crowdsourcing techniques), unit of analysis (verse, stanza, whole poem), granularity of classification (from binary classes, such as positive and negative, to wide sets of emotions), foci of the emotions (annotation of the emotions as depicted in the text by the author or as felt by the reader). With respect to previous work, in this paper we chose to involve experts, to perform annotation at the sentence level (as an intermediate degree of granularity between verse and stanza), to assign four generic classes without defining the specific emotion conveyed by the text, and to focus on the sentiment as depicted by the author.

As for automatic classification systems, the literature reports both lexicon-based (Bonta and Janardhan, 2019) and machine learning approaches, with a constant increasing use of deep learning techniques (Zhang et al., 2018). For example, Mohanty et al. (2018) experiment with Linear-SVM, Naive-Bayes and Logistic Regression classifiers on Odia poems, while Haider et al. (2020) perform multi-label classification on German stanzas with BERT. Given the lack of training data for Latin poetry, in this paper we will instead test unsupervised approaches.

3 Gold Standard Creation

3.1 Annotation

The Gold Standard (GS) consists of eight randomly selected odes,⁵ two from each of the four books that make up the work, for a total of 955 tokens, without punctuation, and 44 sentences (average sentence length: 21, standard deviation: 11). Texts were taken from the corpus prepared by the LASLA laboratory in Liège.⁶ We performed a single-label annotation of the original Latin text by Horace at sentence level. We have chosen the sentence as unit of annotation because it represents an intermediate degree of granularity between that of the verse and that of the stanza. In fact, the limited length of a verse can hinder the full understanding of the sentiment it conveys, while a stanza, being longer, risks to contain very different content and thus, potentially, even opposite sentiments. Furthermore, not all poems can be divided into stanzas, as this depends on the metric scheme of the poem. Instead, sentences can be detected in every poem regardless of its metric scheme, and represent a unit of meaning in their own right.

In the annotation phase, we involved two experts in Latin language and literature (A1 and A2) and another annotator with basic knowledge of Latin but provided with previous experience in sentiment annotation (A3). Annotators were asked to identify the sentiment conveyed by each sentence in the GS, taking into consideration both the vocabulary used by the author and the images that are evoked in the ode. More specifically, annotators were asked to answer the following question: which of the following classes best describes how

³For an orientation on the vast subject of the fortune and reception of the *Odes* see Baldo (2012).

⁴For a recent survey on sentiment and emotion analysis applied to literature, see Kim and Klinger (2018).

⁵Book I: odes 10 and 17; Book II: odes 7 and 13; Book III: odes 13 and 23; Book IV: odes 7 and 11.

⁶http://web.philo.ulg.ac.be/lasla/oper a-latina/.

are the emotions conveyed by the poet in the sentence under analysis?

- positive: the only emotions that are conveyed at lexical level and the only images that are evoked are positive, or positive emotions are clearly prevalent;
- negative: the only emotions that are conveyed at lexical level and the only images that are evoked are negative, or negative emotions are clearly prevalent;
- neutral: there are no emotions conveyed by the text;
- mixed: lexicon and evoked images produce opposite emotions; it is not possible to find a clearly prevailing emotion.

The annotation of the GS was organized in four phases. In the first phase, annotators worked together collaboratively assigning the sentiment class to four of the eight odes (21 sentences): the task was discussed and a common procedure was defined. In the second phase, annotators worked independently on the other four odes (23 sentences): A1 and A2 annotated the original Latin text, while A3 annotated the same odes using an Italian translation (Horace and Nuzzo, 2009) to understand how the use of texts not in the original language can alter the annotation of the sentiment. In the third phase, we calculated the Inter-Annotator Agreement, whereas in the last phase disagreements were discussed and reconciled.

3.2 Inter-Annotator Agreement

Cohen's k between A1 and A2 resulted in 0.5, while Fleiss's k among the three annotators (A1-A2-A3) resulted in 0.48 (both these results are considered moderate agreement). In particular, the negative class proved to be the easiest to be annotated (with a Fleiss's k of 0.64), followed by neutral (0.57) and positive (0.45), whereas mixed was the most problematic class (0.23).

We noticed that the Italian translation was sometimes misleading, resulting in cases of disagreement: e.g., the sentence *inmortalia ne speres monet annus et almum quae rapit hora diem*, (ode IV, 7) is translated as 'speranze di eterno ti vietano gli anni e le ore che involano il giorno radioso' (literal translation of the Italian sentence into English: 'hopes of eternity forbid you the years and the hours that steal the radiant day'). A3 marked this sentence as mixed, considering that it is impossible to identify a prevailing emotion between the negativity expressed by the verb 'vietare' ('to forbid') and the positivity of 'giorno radioso' ('radiant day'). However, the translation of the Latin verb *rapio* is not appropriate: the Italian verb 'involare' ('to steal') does not convey the idea of the violent force inherent in *rapio*, which can be more correctly translated with the verb 'to plunder'.⁷

3.3 Reconciliation

Disagreements were discussed and reconciled by the three annotators: Table 1 presents the number of sentences and tokens per sentiment class. Our GS includes a majority of positive sentences (45.4%). Positive (average length: 21, standard deviation: 11), negative (average length: 24, standard deviation: 14), and mixed (average length: 25, standard deviation: 9) sentences are considerably longer than neutral ones (average length: 8, standard deviation: 3). Annotated examples are given in Table 2: English translations by Kaimowitz et al. (2008) are included for clarity.

	Sentences	Tokens
positive	20	411
negative	12	292
neutral	3	23
mixed	9	229
TOTAL	44	955

Table 1: Gold Standard statistics.

4 **Experiments**

4.1 Lexicon-Based Sentiment Analysis

The dataset for this experiment is obtained by means of a simple dictionary lookup of the lemmas in the *LatinAffectus* sentiment lexicon. Entries in the lexicon are assigned a score of: -1.0, -0.5 (negative polarity), 0 (neutral polarity), +0.5, +1.0 (positive polarity). The tokens in the *Odes* that are lemmatized under lemmas that also have an entry in the *LatinAffectus* are assigned the score that is found in the 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. Note

⁷See for instance the English translation by Kaimowitz et al. (2008): "Do not hope for what's immortal, the year warns, and the hour which plunders the day".

Ode	Sent.	Text	Translation	Class		
	hic tibi copia manabit ad plenum	Here for you will flow				
1.17	1.17 103	benigno ruris honorum opulenta cornu	positive			
	venigno runs nonorum opuienia cornu	spills the country's splendors				
		cuncta manus auidas fugient	All that you bestow upon			
4.7 549	heredis amico quae dederis animo	negative				
	nereais amico quae aederis animo	hands of an heir				
	2.13 265		With the Zephyrs cold grows			
		frigora mitescunt Zephyris uer	ora mitescunt Zephyris uer mild, summer tramples			
2 13		proterit aestas interitura simul	springtime, soon to die,	mixed		
2.13		pomifer autumnus fruges effuderit	once productive autumn pours			
	et mox bruma recurrit iners	forth its fruits, and shortly				
		lifeless winter is back				
2.7	235	quem Venus arbitrum dicet bibendi	Who will Venus name as	neutral		
2.1	2.7 235	quem venus arourum accer bibenai	master of the wine?			

Table 2: Annotated examples taken from the Gold Standard.

that 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 dictionary lookup required some manual disambiguation in cases of ambiguity due to homography. For 18 lemmas (corresponding to 49 tokens in the Odes), the sentiment lexicon provides multiple values; in most cases, as with ales 'winged' (adj.), but also 'bird' (n.), the variation is due to a different polarity attributed to the syntactic uses of the word (in the example, to the adjective and the noun). In such cases, the PoS annotation in the LASLA corpus was used to disambiguate and assign the correct score. We also reviewed those words that, although not tagged as nouns or adjectives in LASLA, still yield a match in LatinAffectus. After revision, we decided to keep the scores for a series of lemmas annotated as numerals in the corpus (simplex 'simple, plain', primus and primum 'first', prius 'former, prior') and the indefinite pronoun solus 'alone, only' that in LatinAffectus are marked as adjectives.

A sentence score (S) was computed by summing the values of all words. Thus, we attributed the label positive to all the sentences with score S > 0 and negative where S < 0. For S = 0, we attributed neutral to sentences where all words had a score of 0 and mixed where positive and negative words were equivalent. The overall accuracy of this method is 48% (macro-average F1 37, weighted macro-average F1 44) with unbalanced scores among the four classes: 70% for positive, 42% for

negative, 67% neutral, while no correct predictions were given for mixed.

4.2 Zero-Shot Classification

We trained a language model for SA on English and tested it on our GS by relying on two stateof-the-art multilingual models. More specifically, we fine-tuned Multilingual BERT (mBERT) (Pires et al., 2019) and XLM-RoBERTa (Conneau et al., 2020) with the GoEmotions corpus (Demszky et al., 2020) using the Hugging Face's PyTorch implementation.⁸ GoEmotions is a dataset of comments posted on Reddit manually annotated for 27 emotion categories or Neutral. In order to adapt this dataset to our needs, we mapped the emotions into sentiment categories as suggested by the authors themselves. For example, joy and love were converged into a unique positive class, whereas fear and grief were merged under the same negative class. The neutral category remained intact and comments annotated with emotions belonging to opposite sentiments were marked as mixed. Comments labeled with ambiguous emotions (i.e. realization, surprise, curiosity, confusion) were instead left out.⁹ With this procedure, we built a training set made of 18,617 positive, 10,133 negative, 1,965 neutral and 1,581 mixed comments. For fine-tuning, we chose the

⁸https://huggingface.co/transformers/ index.html

⁹For the full mapping, please see: https://github .com/google-research/google-research/blo b/master/goemotions/data/sentiment_mappi ng.json.

Language	Test Set	Genre	mBERT	XLM-RoBERTa
English	GoEmotions	social media	86%	73%
	AIT-2018	social media	64%	59%
	Poem Sentiment	literary - poetry	50%	70%
Italian	MultiEmotions-It	social media	70%	75%
	AriEmozione	literary - opera	50%	52%
Latin	Horace GS	literary - poetry	32%	30%

Table 3: Accuracy of the mono-lingual and cross-lingual (zero-shot) classification method.

	Lexicon-Based SA			Zero-Shot mBERT		Zero-Shot XML-RoBERTa			
	Р	R	F1	Р	R	F1	Р	R	F1
positive	0.56	0.70	0.62	0.83	0.25	0.38	1.00	0.10	0.18
negative	0.62	0.42	0.50	0.75	0.50	0.60	0.53	0.67	0.59
neutral	0.25	0.67	0.36	0.10	1.00	0.18	0.11	1.00	0.20
mixed	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: Precision (P), recall (R) and F1-score (F1) for the lexicon-based method and for the zero-shot classification experiment.

following hyperparameters: 32 for batch size, 2e-5 for learning rate, 6 epoches, AdamW optimizer.¹⁰

We evaluated the trained model on different datasets, including our GS. For each of the following test sets, we randomly selected 44 texts so to have the same number of input data as in our GS:

- GoEmotions: test set taken from the same corpus used for training the English model.
- Poem Sentiment: collection of English verses annotated with the same sentiment classes as in our GS (Sheng and Uthus, 2020).
- AIT-2018: English data of the emotion classification task of SemEval-2018 Task 1: Affect in Tweets (Mohammad et al., 2018). Each tweet is annotated as neutral or as one, or more, of eleven emotions. The original annotation was mapped onto our four sentiment classes, leaving out ambiguous emotions.
- AriEmozione: verses taken from 18th century Italian opera texts annotated with one or two emotions and the level confidence of the annotators (Fernicola et al., 2020). We randomly selected our test set from verses with high confidence scores, mapping emotions onto our four sentiment classes. Since the dataset does not contain verses annotated with opposite emotions, the class mixed is not present in the test set we built.

• MultiEmotions-It: a multi-labeled emotion dataset made of Italian comments posted on YouTube and Facebook (Sprugnoli, 2020). The original emotion labels were converted into our four classes.

Table 3 reports the results of mono-lingual and cross-lingual classification for the different datasets briefly described above and for the two pre-trained multilingual models. There is no clear prevalence of one model over the other: results vary greatly from one dataset to another. On the same language (thus without zero-shot transfer), we notice a drop in the performance for both mBERT and XML-RoBERTa when moving from Reddit comments, that is the same type of text as the training data, to tweets, but even more so when they are evaluated on poems. As for the zero-shot classification, results on Italian YouTube and Facebook comments are better than the ones registered on English tweets, but accuracy drops when applied to opera verses. However, the worst results are recorded for Latin with an accuracy equal to, or slightly above 30% (for mBERT: macro-average F1 29, weighted macro-average F1 35; for XML-RoBERTa: macro-average F1 24, weighted macro-average F1 26). For both mBERT and XML-RoBERTa, we register the same trend at class level: perfect accuracy for neutral, good accuracy for negative (50% with mBERT and 67% with XML-RoBERTa), low accuracy for positive (25% with mBERT and 10% with

¹⁰We adapted the following implementation: https:// gist.github.com/sayakmisra/b0cd67f406b4e 4d5972f339eb20e64a5.

XML-RoBERTa) and no correct predictions for mixed.

5 Conclusions and Future Work

In this paper we have presented a new GS, made of odes written by Horace, for the annotation of sentiment in Latin poetry. The extension of the manually annotated dataset is one of our future work: the goal is to have a sufficient amount of data to test supervised systems. We have also experimented two different SA approaches that do not require training data: both of them are not able to correctly identify sentences with mixed sentiments, which, in any case, are the most problematic also for human annotators. Table 4 reports a comparison in terms of precision, recall and F1score among the lexicon-based approach and the zero-shot classification experiments with both the mBERT and the XML-RoBERTa models. The former performs better on the positive class whereas the zero-shot method achieves a higher F1-score on the negative one even if this class is not the most frequent in the training data. Both mBERT and XML-RoBERTa obtain a very high precision on the sentences marked as positive (0.83 and 1.00 respectively) but the recall is extremely low (0.25 and 0.10 respectively). On the contrary, for the neutral class, the recall is perfect (1.00 for both models) but the precision is very low (0.10 and 0.11 respectively).

A manual inspection of the output of the lexicon-based method revealed two main problems of that approach: i) the limited coverage of LatinAffectus and ii) sentiment shifters are not properly taken into consideration. As for the first point, LatinAffectus covers the 43% of nominal and adjectival lemmas in the GS, leaving out lemmas with a clear sentiment orientation. To overcome this issue, we are currently working on the extension of the lexicon with additional 10,000 lemmas. Regarding the sentiment shifters, their impact is exemplified by the following sentence: cum semel occideris et de te splendida Minos fecerit arbitria non Torquate genus non te facundia non te restituet pietas ('When you at last have died and Minos renders brillant judgement on your life, no Torquatus, not birth, not eloquence, not your devotion will bring you back.' - ode IV, 7). Here, the sentiment score calculated by the script is very positive (3) because it does not handle the frequent negations: however, the particle non should reverses the positive polarity of *facundia* 'eloquence' and *pietas* 'devotion'. This problem could be mitigated by modifying the script with rules that take into account negations and their focus.

Regarding the zero-shot classification approach, the very low performances on Latin deserve further investigation. It is possible that the problem lies in the data used to build the pre-trained models: i.e., Wikipedia for mBERT and Commoncrawl for XML-RoBERTa. Both resources were developed by relying on automatic language detection engines and are highly noisy due to the presence of languages other than Latin and of terms related to modern times. An additional improvement may also come from using for finetuning an annotated in-domain corpus in a wellresource language, that is a corpus of annotated poems: unfortunately, the currently available corpora are not big enough for such purpose.

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

- Ouais Alsharif, Deema Alshamaa, and Nada Ghneim. 2013. Emotion classification in arabic poetry using machine learning. *International Journal of Computer Applications*, 65(16).
- Gianluigi Baldo. 2012. Horace (Quintus Horatius Flaccus), Carmina. In Christine Walde and Brigitte Egger, editors, *Brill's New Pauly Supplements I Volume 5 : The Reception of Classical Literature*. Brill, Amsterdam, October. Publisher: Brill.
- Linda Barros, Pilar Rodriguez, and Alvaro Ortigosa. 2013. Automatic Classification of Literature Pieces by Emotion Detection: A Study on Quevedo's Poetry. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pages 141–146. IEEE.
- Venkateswarlu Bonta and Nandhini Kumaresh2and N Janardhan. 2019. A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, 8(S2):1–6.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised

cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online, July. Association for Computational Linguistics.

- Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A Dataset of Fine-Grained Emotions. In 58th Annual Meeting of the Association for Computational Linguistics (ACL).
- Francesco Fernicola, Shibingfeng Zhang, Federico Garcea, Paolo Bonora, and Alberto Barrón-Cedeño. 2020. AriEmozione: Identifying Emotions in Opera Verses. In *Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it* 2020). Accademia University Press.
- Thomas Haider, Steffen Eger, Evgeny Kim, Roman Klinger, and Winfried Menninghaus. 2020. PO-EMO: Conceptualization, annotation, and modeling of aesthetic emotions in German and English poetry. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1652– 1663, Marseille, France, May. European Language Resources Association.
- Horace and Gianfranco Nuzzo. 2009. *I quattro libri delle Odi e l'Inno secolare di Quinto Orazio Flacco*. Flaccovio.
- Yufang Hou and Anette Frank. 2015. Analyzing sentiment in classical chinese poetry. In *Proceedings* of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), pages 15–24.
- Philip N. Johnson-Laird and Keith Oatley. 2016. Emotions in music, literature, and film. In Lisa Feldman Barrett, Michael Lewis, and Jeannette M. Haviland-Jones, editors, *Handbook of emotions*, chapter 3, pages 82–97. The Guildford Press.
- Jeffrey H Kaimowitz, Ronnie Ancona, et al. 2008. *The* odes of Horace. Johns Hopkins University Press.
- Evgeny Kim and Roman Klinger. 2018. A survey on sentiment and emotion analysis for computational literary studies. *arXiv preprint arXiv:1808.03137*.
- Walaa Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113.
- Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. In Proceedings of International Workshop on Semantic Evaluation (SemEval-2018), New Orleans, LA, USA.
- Gaurav Mohanty, Pruthwik Mishra, and Radhika Mamidi. 2018. Kabithaa: An annotated corpus of odia poems with sentiment polarity information. In

Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Paris, France, may. European Language Resources Association (ELRA).

- Marco Passarotti, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. 2020. Interlinking through lemmas. the lexical collection of the LiLa knowledge base of linguistic resources for Latin. *Studi e Saggi Linguistici*, 58(1):177–212.
- Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How Multilingual is Multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001.
- Emily Sheng and David C Uthus. 2020. Investigating societal biases in a poetry composition system. In Proceedings of the Second Workshop on Gender Bias in Natural Language Processing, pages 93– 106.
- Rachele Sprugnoli and Marco Passarotti, editors. 2020. Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages, Marseille, France, May. European Language Resources Association (ELRA).
- Rachele Sprugnoli, Francesco Mambrini, Giovanni Moretti, and Marco Passarotti. 2020a. Towards the Modeling of Polarity in a Latin Knowledge Base. In Proceedings of the Third Workshop on Humanities in the Semantic Web (WHiSe 2020), pages 59–70.
- Rachele Sprugnoli, Marco Passarotti, Daniela Corbetta, and Andrea Peverelli. 2020b. Odi et Amo. Creating, Evaluating and Extending Sentiment Lexicons for Latin. In Proceedings of the 12th Language Resources and Evaluation Conference, pages 3078– 3086.
- Rachele Sprugnoli. 2020. MultiEmotions-it: A new dataset for opinion polarity and emotion analysis for Italian. In Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020), pages 402–408. Accademia University Press.
- PS Sreeja and GS Mahalakshmi. 2019. Perc-an emotion recognition corpus for cognitive poems. In 2019 International Conference on Communication and Signal Processing (ICCSP), pages 0200–0207. IEEE.
- Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253.