

Profiling of Intertextuality in Latin Literature Using Word Embeddings

Patrick J. Burns,¹ James A. Brofos,² Kyle Li,³ Primit Chaudhuri,¹ Joseph P. Dexter^{4†}

¹ Department of Classics, University of Texas at Austin

² Department of Statistics and Data Science, Yale University

³ Department of Computer Science, Columbia University

⁴ Data Science Initiative and Department of Human Evolutionary Biology, Harvard University

† Corresponding author: jdexter@fas.harvard.edu

Abstract

Identifying intertextual relationships between authors is of central importance to the study of literature. We report an empirical analysis of intertextuality in classical Latin literature using word embedding models. To enable quantitative evaluation of intertextual search methods, we curate a new dataset of 945 known parallels drawn from traditional scholarship on Latin epic poetry. We train an optimized word2vec model on a large corpus of lemmatized Latin, which achieves state-of-the-art performance for synonym detection and outperforms a widely used lexical method for intertextual search. We then demonstrate that training embeddings on very small corpora can capture salient aspects of literary style and apply this approach to replicate a previous intertextual study of the Roman historian Livy, which relied on hand-crafted stylometric features. Our results advance the development of core computational resources for a major pre-modern language and highlight a productive avenue for cross-disciplinary collaboration between the study of literature and NLP.¹

1 Introduction

In “Lonesome Day Blues,” Bob Dylan sings, “I’m gonna spare the defeated...I am goin’ to teach peace to the conquered / I’m gonna tame the proud.” This lyric echoes a passage from Vergil’s ancient Latin epic, the *Aeneid*, as translated by Allen Mandelbaum: “to teach the ways of peace to those you conquer, / to spare defeated peoples, tame the proud” (Thomas, 2012). Such allusions or “intertexts” transmit ideas across space and time, diverse media, and languages. Although researchers focus on those intertextual connections felt to have special literary significance for the works at hand, in principle intertextuality refers to any verbal or

semantic resemblance within the literary system, ranging from direct quotation to topical similarities (Kristeva, 1980; Juvan, 2009). Given the importance of intertextual criticism to literary study, computational identification of text reuse in literature is an active area of research (Bamman and Crane, 2008; Forstall and Scheirer, 2019).

Classical Latin literature is a highly influential tradition characterized by an extraordinary density of allusions and other forms of text reuse (Hinds, 1998). The most widely used tools for the detection of Latin intertextuality, such as Tesseræ and Diogenes, rely on lexical matching of repeated words or phrases (Coffee et al., 2012, 2013; Heslin, 2019). In addition to these core methods, other research has explored the use of sequence alignment (Chaudhuri et al., 2015; Chaudhuri and Dexter, 2017), semantic matching (Scheirer et al., 2016), and hybrid approaches (Moritz et al., 2016; Manjavacas et al., 2019) for Latin intertextual search, complementing related work on English (Smith et al., 2014; Zhang et al., 2014; Barbu and Trausan-Matu, 2017). Much NLP research on historical text reuse, including previous applications of Latin word embeddings, has focused on the Bible and other religious texts (Lee, 2007; Moritz et al., 2016; Bjerva and Praet, 2016; Manjavacas et al., 2019). As such, there is a clear need for enhanced computational methods for classical Latin literature. We describe the optimization of word embedding models for Latin and their application to longstanding questions about literary intertextuality.

2 Evaluation and optimization of word embedding models for Latin

As is typical for many low-resource and pre-modern languages, development of core NLP technologies for Latin remains at an early stage. Following attempts to train word2vec models on unlemmatized corpora of Latin literature shortly after the method’s introduction (Bamman; Bjerva and Praet,

¹Code and data are available at <https://github.com/QuantitativeCriticismLab/NAACL-HLT-2021-Latin-Intertextuality>.

2015) and inclusion of Latin in large-scale multilingual releases of FastText and BERT (Grave et al., 2018; Devlin et al., 2019), in the past year there has been increased interest in systematic optimization and evaluation of Latin embeddings. Spurred by the recent EvaLatin challenge (Sprugnoli et al., 2020), a number of Latin models have been trained for use in lemmatization and part-of-speech tagging (Bacon, 2020; Celano, 2020; Straka and Straková, 2020; Stoeckel et al., 2020), complementing new literary applications to Biblical text reuse and neo-Latin philosophy (Manjavacas et al., 2019; Bloem et al., 2020). In addition, Sprugnoli et al. (2019) recently introduced a synonym selection dataset, based on the TOEFL benchmark for English, which they used to evaluate word2vec and FastText models trained on the LASLA corpus of Latin literature.

To the best of our knowledge, there have been no attempts to compare the performance of these models on standard evaluation tasks. To establish a baseline for further language-specific optimization and to inform our research on intertextuality, we evaluate five Latin models for which pretrained embeddings are publicly available. These models encompass a variety of training corpora and methods, including word2vec, FastText, and nonce2vec (Appendix). We consider two tasks involving synonym matching. The first is the selection task introduced by Sprugnoli et al. (2019); the task is to distinguish the true synonym of a Latin word from three distractors ($N = 2,759$). The second task, which is modeled on one of the English evaluation datasets from Mikolov et al. (2013), involves unrestricted search for the synonyms of 1,910 words found in an online dictionary of Latin near-synonyms (Appendix). In addition, we train word2vec embeddings on a large corpus of Latin compiled from the Internet Archive (Bamman and Crane, 2011; Bamman and Smith, 2012), which we first lemmatize using either the Classical Language Toolkit (Johnson, 2021) or TreeTagger (Schmid, 1994).

The results of the comparative evaluation are summarized in Table 1. For the synonym search task, we consider the number of correct matches found in the top 1, 10, and 25 results by cosine similarity, as well as the mean reciprocal rank (MRR). We find that our models achieve state-of-the-art performance on both tasks compared to the five published models. The improvement in performance may be due to the combination of training on lem-

matized text, which Sprugnoli et al. (2019) identified as an important optimization for Latin, and use of a lower-quality but much larger training corpus (1.38 billion tokens, compared to 1.7 million tokens in the curated LASLA corpus).

3 Construction of benchmark intertextuality dataset

Despite the enormous number of Latin intertextual parallels recorded in the scholarship, computational research on literary text reuse is hampered by a lack of benchmark datasets. Existing benchmarks tend to focus either on binary comparisons, such as between Vergil and Lucan (Coffee et al., 2012), or on specialized forms of religious intertextuality (Moritz et al., 2016; Manjavacas et al., 2019). To enable validation testing of general NLP methods for intertextual search, we assemble a new benchmark dataset based on Valerius Flaccus’ *Argonautica*, an epic poem dating from the 1st century C.E. which recounts the myth of Jason and the Argonauts. For Book 1 of the *Argonautica* we record 945 verbal intertexts with four major epics (Vergil’s *Aeneid*, Ovid’s *Metamorphoses*, Lucan’s *Pharsalia*, and Statius’ *Thebaid*) that are noted in the commentaries of Spaltenstein (2002), Kleywegt (2005), and Zissos (2008). Our dataset thus contains a substantial number of intertexts of established literary interest with coverage across Book 1.

4 Analysis of intertextuality in Latin literature

4.1 Enhanced intertextual search

Several widely used computational search methods for Latin intertextuality rely on lexical matching of related words. We present an alternative approach in which potential intertextual phrases are ranked using word embeddings.

According to this method, we compare a bigram of interest to all bigrams in another text subject to the constraint that the distance between the words does not exceed a fixed interval. The interval parameter is determined by the number of words occurring between the words comprising the bigram of interest and is usually, but not exclusively, between 0 and 2. The choice of bigrams as the basic unit conforms to ancient poetic practice, in which allusive phrases frequently consist of two words (although they can also be single words or longer phrases), and hence also conforms to modern intertextual search methods such as Tesseræ (Coffee

Model	Selection (%)	Ranking (%)			
		top 1	top 10	top 25	MRR
Bamman	66.6	0.4	2.1	3.3	17.5
Grave et al. (2018)	74.0	0.2	1.2	1.7	11.8
Sprugnoli et al. (2019) (word2vec CBOW)	81.1	2.4	11.3	15.9	19.8
Sprugnoli et al. (2019) (FastText SG)	86.9	1.7	9.3	14.3	18.2
Bloem et al. (2020)	84.8	0.3	3.9	7.0	10.1
word2vec (CLTK)	84.9	3.2	14.5	20.4	22.7
word2vec (TT)	87.7	3.5	15.0	21.0	20.6

Table 1: Evaluation of five published and two new Latin word embedding models on two synonym detection tasks.

et al., 2012, 2013). A key difference in our approach, however, is that bigram pairs may share only one or even zero words in common. The bigrams are drawn from the dataset of commentators’ annotations; in cases where commentators note only a single-word intertext or a phrase longer than a bigram, we supplement or select words on a case-by-case basis, giving preference to those words that bear a semantic or syntactic similarity to one or more words in the intertext.

The similarity score for a bigram pair is calculated by taking the cosine similarities of the embeddings of the four possible pairs of words across both bigrams, and averaging the highest cosine similarity and the score for the remaining pair of words. The bigram pair *flammifero Olympo* (“fiery Olympus”) and *flammi fera nocte* (“fiery night”), for example, generates the four lemmatized pairs *flammifer ~ flammifer*, *flammifer ~ nox*, *Olympus ~ flammifer*, and *Olympus ~ nox*. Hence, the similarity score for the bigram pair is the average of 1.0 for the exact match, *flammifer ~ flammifer*, and 0.35 for the other remaining word pair, *Olympus ~ nox* (i.e., 0.67). In this way, the similarity score for an intertext noted by the commentators is ranked against all other bigrams in the relevant text, the size of which we set at a single book of poetry (i.e., equivalent to the text on which the dataset is based). Although the choice to use one unit of text rather than another is somewhat arbitrary—one could consider complete works rather than constituent books, for example—the use of single books has several advantages, notably provision of a large but not overwhelming number of comparison phrases while maintaining ancient textual units with distinct episodes and themes.

Following this approach, we compute a ranking for each of the 945 parallels in the Valerius Flaccus benchmark. For embeddings we use our word2vec model trained on CLTK-lemmatized text, which by

MRR performs best in the synonym ranking task (Table 1). The precision@ k and recall@ k for $k = 1, 3, 5, 10, 25, 50, 75, 100$, and 250 are summarized in Fig. 1.

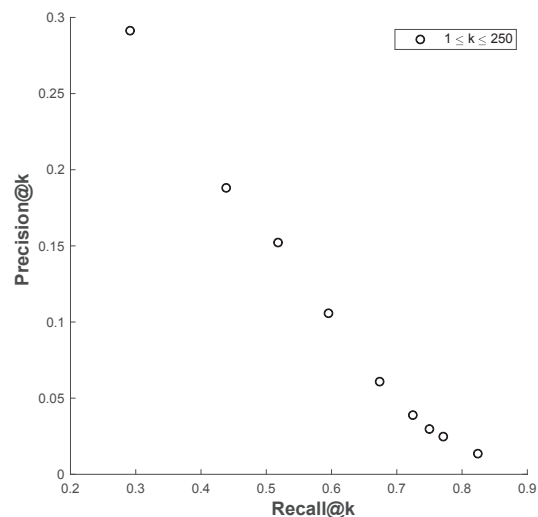


Figure 1: Precision and recall for Latin intertextual search using an optimized word2vec model.

We next compare our method and the Tesseract search tool, which is regarded as state-of-the-art for Latin intertextual search (Bernstein et al., 2015; Forstall and Scheirer, 2019). Using their public web-based interface, we run Tesseract searches comparing Book 1 of the *Argonautica* with each of the four texts in the benchmark dataset. Tesseract produces lists of repeated bigrams ranked according to a hand-crafted scoring formula that considers the rareness and proximity of the words in each bigram. For the complete set of Tesseract results, the recall is 33.9%, and the precision is 0.97%; with $k = 250$, our method achieves a comparable precision (1.4%) but higher recall (82.4%). An important advantage of the Tesseract tool, however, is that it searches for similar phrases in parallel and does not require a list of specific queries as input. As such, the results aggregated for this comparison come from a much

smaller number of Tesseract searches than the 945 embedding-based searches we run. For this reason, Tesseract is likely to be more suitable than our method for applications in which the user does not have predetermined phrases of interest.

A minority of intertexts in the dataset contain no shared lemma and hence present a challenge for existing detection methods based on lexical matching but are recoverable using our search method. The phrases *e clausis [antris]* (“from the enclosed [cave],” *Arg.* 1.417) and *circum claustra [fremunt]* (“[they roar] around the gate,” *Aen.* 1.56), for example, contain no words in common but have similar syntax (the prepositions *e* and *circum*) and semantics (words indicating enclosure). Similarly, the phrases *Phlegethontis operti* (“hidden Phlegethon,” *Arg.* 1.735) and *Acherontis aperti* (“open Acheron,” *Theb.* 11.150) both refer to rivers in the underworld and contain near-antonymic adjectives. Word embeddings can thus be used to identify intertexts of literary interest in a way that complements existing methods.

4.2 Anomaly detection

Computational analysis of literary intertextuality is typically treated as an information retrieval problem, as in the previous section. Here we consider an alternative framework of studying intertextuality through anomaly detection (Forstall et al., 2011). For this approach, we train word embeddings on highly restricted corpora, so that the resulting models capture aspects of authorial style. We use those restricted embeddings as features to predict instances of similarity between authors, which can indicate intertextuality. To illustrate this approach we describe a case study involving Latin historiography and the development of prose style.

In particular, we examine patterns of stylistic influence between the Roman historian Livy, his source material, and other Latin prose literature. As assessment of similarities in literary style is inherently subjective, we consider the task of replicating two experiments from a previous computational study of Livy, which employed a hand-crafted set of Latin stylometric features such as syntactic markers and function words, using word embeddings. Our approach to evaluation of a subjective task is thus similar to that of Bamman et al. (2014), who tested a set of preregistered hypotheses about literary characters.

Like most historical writing, Livy’s monumental

history of Rome drew on a wide range of source material, such as earlier historiography and political speeches, most of which is no longer extant. The extent to which Livy cited these earlier sources, and their influence on Livy’s compositional practice, remain important open questions for ancient historians. Dexter et al. (2017) demonstrated previously that anomaly detection could be used to distinguish a database of 439 putative citational passages from the remainder of Livy. To replicate this analysis, we train a word2vec model on all of Livy’s surviving history and use the embeddings as input for a one-class support vector machine (SVM). Following Dexter et al. (2017), we set the detection rate of the one-class SVM to 20% and train on a random selection of 30,000 5-sentence passages of Livy. We find that the one-class SVM labels $38.2 \pm 0.8\%$ of passages from the citation database as anomalous, compared to $18.4 \pm 2.0\%$ of a validation set with 439 passages of general Livy (mean and standard deviation from $N = 3$ runs). These results provide further evidence that citational passages of Livy exhibit an anomalous writing style, whether due to source use or stylistic modulation, corroborating the earlier analysis.

Finally, we consider the stylistic similarity of Livy to 17 other works of Latin literature analyzed by Dexter et al. (2017). Again using a one-class SVM trained on Livy, we predict the “Livianess” of each work (Fig. 2). Our results confirm the major trends identified by the prior stylometric analysis, including the expected dissimilarity to Livy of the verse texts and the consistent similarity of contemporary and early imperial historiography. The primary difference between the two sets of results is that the stylometric features indicate greater similarity between Livy and non-historiographical prose, such as Augustine’s *Confessions* and Vitruvius’ *De architectura*, than do word embeddings, which may reflect a relative lack of shared diction.

5 Conclusions

We present an empirical analysis of Latin intertextuality using word embedding models. In addition to its specific contributions to literary criticism and the digital humanities, our work makes several methodological advances of interest to the broader NLP community. We conduct a comparative evaluation of Latin word embedding models for two synonym matching tasks and report an optimized model that achieves state-of-the-art performance,

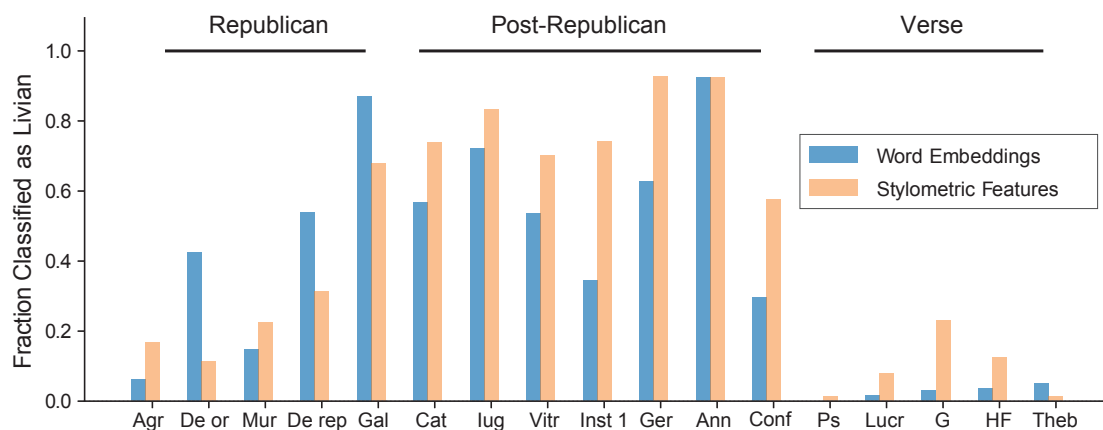


Figure 2: Comparison of Livy to 17 other Latin works. The stylometric data is reprinted from Dexter et al. (2017), which used passages of 35 sentences each.

which we apply to intertextual search of Latin poetry. By capturing similarities other than exact repetition of words and phrases, our method complements existing search tools, such as Diogenes and Tesserae. Given the diversity and complexity of references employed by Latin authors, taking a multifaceted approach is essential to the computational study of Latin intertextuality. Although our initial work focuses on static embeddings, one potential avenue for improving our search method would be to leverage context-aware embeddings such as multilingual or Latin BERT (Devlin et al., 2019; Bamman and Burns, 2020). In addition, we illustrate how intertextuality can be studied using anomaly detection, and we replicate previous stylometric research about the Roman historian Livy, which was informed by domain knowledge, using an unsupervised approach. We hope that this work will strengthen cross-disciplinary collaboration between classics, the digital humanities, and NLP.

Acknowledgements

This work was conducted under the auspices of the Quantitative Criticism Lab (www.qcrit.org), an interdisciplinary group co-directed by P.C. and J.P.D. and supported by an American Council of Learned Societies Digital Extension Grant and a National Endowment for the Humanities Digital Humanities Advancement Grant (Grant No. HAA-271822-20). P.C. was supported by a Mellon New Directions Fellowship, and J.P.D. by a Neukom Fellowship and a Harvard Data Science Fellowship. The material contributed by J.A.B. is based upon work supported by the National Science Founda-

tion Graduate Research Fellowship under Grant No. 1752134. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Adriana Cásarez, James Patterson, and Ariane Schwartz for assistance with compiling the Valerius Flaccus intertextuality dataset, and Tommaso Spinelli for sharing the dictionary of Latin near-synonyms.

References

- Geoff Bacon. 2020. [Data-driven choices in neural part-of-speech tagging for Latin](#). In *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 111–113, Marseille. European Language Resources Association (ELRA).
- David Bamman. [11K Latin texts](#).
- David Bamman and Patrick J. Burns. 2020. [Latin BERT: A contextual language model for classical philology](#).
- David Bamman and Gregory Crane. 2008. [The logic and discovery of textual allusion](#). In *Proceedings of the 2008 LREC Workshop on Language Technology for Cultural Heritage Data*.
- David Bamman and Gregory Crane. 2011. [Measuring historical word sense variation](#). In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, page 1–10, New York, New York. Association for Computing Machinery.
- David Bamman and David Smith. 2012. [Extracting two thousand years of Latin from a million book library](#). *Journal on Computing and Cultural Heritage*, 5.

- David Bamman, Ted Underwood, and Noah A. Smith. 2014. [A Bayesian mixed effects model of literary character](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 370–379, Baltimore, Maryland. Association for Computational Linguistics.
- Miruna-Stefania Barbu and Stefan Trausan-Matu. 2017. [Intertextuality detection in literary texts using word2vec models](#). In *2017 21st International Conference on System Theory, Control and Computing (ICSTCC)*, pages 262–265. Institute of Electrical and Electronics Engineers.
- Neil Bernstein, Kyle Gervais, and Wei Lin. 2015. [Comparative rates of text reuse in classical Latin hexameter poetry](#). *Digital Humanities Quarterly*, 9.
- Johannes Bjerva and Raf Praet. 2015. [Word embeddings pointing the way for late antiquity](#). In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, pages 53–57, Beijing. Association for Computational Linguistics.
- Johannes Bjerva and Raf Praet. 2016. [Rethinking intertextuality through a word-space and social network approach – the case of Cassiodorus](#).
- Jelke Bloem, Maria Chiara Parisi, Martin Reynaert, Yvette Oortwijn, and Arianna Betti. 2020. [Distributional semantics for neo-Latin](#). In *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 84–93, Marseille. European Language Resources Association (ELRA).
- Giuseppe Celano. 2020. [A gradient boosting-Seq2Seq system for Latin POS tagging and lemmatization](#). In *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 119–123, Marseille. European Language Resources Association (ELRA).
- Pramit Chaudhuri and Joseph P. Dexter. 2017. [Bioinformatics and classical literary study](#). *Journal of Data Mining & Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages.
- Pramit Chaudhuri, Joseph P. Dexter, and Jorge A. Bonilla Lopez. 2015. [Strings, triangles, and go-betweens: Intertextual approaches to Silius' Carthaginian debates](#). *Dictynna*, 12.
- Neil Coffee, Jean-Pierre Koenig, Shakthi Poornima, Roelant Ossewaarde, Christopher Forstall, and Sarah Jacobson. 2012. [Intertextuality in the digital age](#). *Transactions of the American Philological Association*, 142:383–422.
- Neil Coffee, Jean-Pierre Koenig, Shakthi Poornima, Roelant Ossewaarde, Christopher Forstall, and Sarah Jacobson. 2013. [The Tesseræ Project: Intertextual analysis of Latin poetry](#). *Literary and Linguistic Computing*, 28:221–228.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Joseph P. Dexter, Theodore Katz, Nilesh Tripuraneni, Tathagata Dasgupta, Ajay Kannan, James A. Brofos, Jorge A. Bonilla Lopez, Lea A. Schroeder, Adriana Casarez, Maxim Rabinovich, Ayelet Haimson Lushkov, and Pramit Chaudhuri. 2017. [Quantitative criticism of literary relationships](#). *Proceedings of the National Academy of Sciences*, 114:E3195–E3204.
- Christopher W. Forstall, Sarah L. Jacobson, and Walter J. Scheirer. 2011. [Evidence of intertextuality: investigating Paul the Deacon's *Angustae Vitae*](#). *Literary and Linguistic Computing*, 26:285–296.
- Christopher W. Forstall and Walter J. Scheirer. 2019. [Quantitative Intertextuality: Analyzing the Markers of Information Reuse](#). Springer International Publishing, Cham.
- Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. [Learning word vectors for 157 languages](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 3483–3487, Miyazaki. European Language Resources Association (ELRA).
- Peter Heslin. 2019. [Diogenes](#).
- Stephen E. Hinds. 1998. *Allusion and Intertext: Dynamics of Appropriation in Roman Poetry*. Cambridge University Press, Cambridge.
- Kyle P. Johnson. 2021. [CLTK: The Classical Language Toolkit](#).
- Marko Juvan. 2009. *History and Poetics of Intertextuality*. Purdue University Press, West Lafayette, Indiana.
- A. J. Kleywegt. 2005. *Valerius Flaccus, Argonautica, Book I*. Mnemosyne, Bibliotheca Classica Batava. Supplementum. Brill, Leiden.
- Julia Kristeva. 1980. Word, dialogue, and novel. In Leon S. Roudiez, editor, *Desire in Language: A Semiotic Approach to Literature and Art*, pages 64–91. Columbia University Press, New York, New York.
- John Lee. 2007. [A computational model of text reuse in ancient literary texts](#). In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 472–479, Prague. Association for Computational Linguistics.

- Enrique Manjavacas, Brian Long, and Mike Kestemont. 2019. [On the feasibility of automated detection of allusive text reuse](#). In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 104–114, Minneapolis, Minnesota. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. [Efficient estimation of word representations in vector space](#).
- Maria Moritz, Andreas Wiederhold, Barbara Pavlek, Yuri Bizzoni, and Marco Büchler. 2016. [Non-literal text reuse in historical texts: An approach to identify reuse transformations and its application to Bible reuse](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1849–1859, Austin, Texas. Association for Computational Linguistics.
- Walter Scheirer, Christopher Forstall, and Neil Coffee. 2016. [The sense of a connection: Automatic tracing of intertextuality by meaning](#). *Digital Scholarship in the Humanities*, 31:204–217.
- Helmut Schmid. 1994. [Probabilistic part-of-speech tagging using decision trees](#). In *Proceedings of the International Conference on New Methods in Language Processing*, Manchester.
- David A. Smith, Ryan Cordell, Elizabeth Maddock Dillon, Nick Stramp, and John Wilkerson. 2014. [Detecting and modeling local text reuse](#). In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, page 183–192. Institute of Electrical and Electronics Engineers.
- François Spaltenstein. 2002. *Commentaire des “Argonautica” de Valérius Flaccus (livres 1 et 2)*. Éd. Latomus, Bruxelles.
- Rachele Sprugnoli, Marco Passarotti, Flavio Massimiliano Cecchini, and Matteo Pellegrini. 2020. [Overview of the EvaLatin 2020 evaluation campaign](#). In *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 105–110, Marseille. European Language Resources Association (ELRA).
- Rachele Sprugnoli, Marco Passarotti, and Giovanni Moretti. 2019. [Vir is to moderatus as mulier is to intemperans: Lemma embeddings for Latin](#). In *Proceedings of the Sixth Italian Conference on Computational Linguistics*, Bari.
- Manuel Stoeckel, Alexander Henlein, Wahed Hemati, and Alexander Mehler. 2020. [Voting for POS tagging of Latin texts: Using the flair of FLAIR to better ensemble classifiers by example of Latin](#). In *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 130–135, Marseille. European Language Resources Association (ELRA).
- Milan Straka and Jana Straková. 2020. [UDPipe at EvaLatin 2020: Contextualized embeddings and treebank embeddings](#). In *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 124–129, Marseille. European Language Resources Association (ELRA).
- Richard F. Thomas. 2012. The streets of Rome: The classical Dylan. In William Brockliss, Primit Chaudhuri, Ayelet Haimson Lushkov, and Katherine Wasdin, editors, *Reception and the Classics: An Interdisciplinary Approach to the Classical Tradition*, pages 134–159. Cambridge University Press, Cambridge.
- Qi Zhang, Jihua Kang, Jin Qian, and Xuanjing Huang. 2014. [Continuous word embeddings for detecting local text reuses at the semantic level](#). In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, page 797–806, New York, New York. Association for Computing Machinery.
- Andrew Zissos. 2008. *Valerius Flaccus’ Argonautica Book I, Edited with Introduction, Translation, and Commentary*. Oxford University Press, Oxford.

A Appendices

A.1 Links to published Latin word embedding models

- Bamman: <https://www.cs.cmu.edu/~dbamman/latin.html>
- Grave et al. (2018): <https://fasttext.cc/docs/en/crawl-vectors.html>
- Sprugnoli et al. (2019) (word2vec CBOW): (“allLASLAlemmi-vector-100-nocase-w5-CBOW.vec”), <https://embeddings.lila-erc.eu/samples/download/word2vec>
- Sprugnoli et al. (2019) (FastText SG): (“allLASLA-lemmi-fast-100-SKIP-win5-min5.vec”), <https://embeddings.lila-erc.eu/samples/download/fasttext>
- Bloem et al. (2020): (“bamman-c50-d100”), <https://github.com/bloemj/nonce2vec/tree/nonce2vec-latin>

A.2 Training details for new Latin word2vec models

- word2vec (CLTK) is trained using Gensim word2vec with the following parameters: size = 300, min_count = 100, iter = 1. The model is trained on plaintext files from Bamman (<https://www.cs.cmu.edu/~dbamman/latin.html>) lemmatized with CLTK BackoffLatinLemmatizer (Johnson, 2021).
- word2vec (TT) is trained using Gensim word2vec with the following parameters: size = 50, min_count = 100, iter = 1. The model is trained on plaintext files from Bamman (<https://www.cs.cmu.edu/~dbamman/latin.html>) lemmatized with Tree-Tagger (Schmid, 1994).

A.3 Links to publicly available datasets

- Synonym selection dataset from Sprugnoli et al. (2019): <https://embeddings.lila-erc.eu/samples/syn/syn-selection-benchmark-Latin.tsv>
- Dictionary of Latin near-synonyms from Spinelli (2018): <https://github.com/tommasospinelli/Online-Dictio>

[nary-of-Latin-Near-Synonyms/blob/master/Latin%20Near-Synonyms%20dataset.txt](https://github.com/qcrit/PNAS_2017_QuantitativeCriticism/blob/master/Latin%20Near-Synonyms%20dataset.txt)

- Database of possible citational passages in Livy from Dexter et al. (2017): https://github.com/qcrit/PNAS_2017_QuantitativeCriticism/blob/master/Code/LivyPassages.csv

A.4 Parameters for Tesseract searches

We run four Tesseract searches comparing Book 1 of Valerius Flaccus’ *Argonautica* to Vergil’s *Aeneid*, Ovid’s *Metamorphoses*, Lucan’s *Pharsalia*, and Statius’ *Thebaid*, respectively. For these searches, we use version 3 of the online Tesseract tool (<http://tesseract.caset.buffalo.edu/>) with all parameters set to their default values: unit = line, feature = lemma, number of stop words = 10, stoplist basis = corpus, score basis = word, frequency basis = corpus, maximum distance = 10 words, distance metric = frequency, drop scores below = 6.

A.5 Latin texts

The texts of the Latin epic poems included in our intertextual search analysis are from the Tesseract corpus (<https://github.com/tesseract/tesseract/tree/master/texts>), which is derived from the Perseus Digital Library (<http://www.perseus.tufts.edu/hopper/>).

Following Dexter et al. (2017), we consider 17 texts for the stylistic analysis of Livy, abbreviated in Fig. 2 according to the *Oxford Classical Dictionary*: Agr, Cato’s *De agri cultura*; Ann, Tacitus’ *Annals*; Conf, Augustine’s *Confessions*; De or, Cicero’s *De oratore*; De rep, Cicero’s *De republica*; Cat, Sallust’s *De coniuratione Catilinae*; G, Vergil’s *Georgics*; Gal, Caesar’s *Bellum Gallicum*; Ger, Tacitus’ *Germania*; HF, Seneca’s *Hercules Furens*; Inst 1, Quintilian’s *Institutio Oratoria* 1; Iug, Sallust’s *Bellum Iugurthinum*; Lucr, Lucretius’ *De rerum natura*; Mur, Cicero’s *Pro Murena*; Ps, Plautus’ *Pseudolus*; Theb, Statius’ *Thebaid*; Vitruvius’ *De architectura*. Further information about this corpus is available in Dexter et al. (2017).