

Augmented Close Reading for Classical Latin using BERT for Intertextual Exploration

Ashley Gong

Harvard University

ashleygong@college.harvard.edu

Katy Ilonka Gero

Harvard University

katy@g.harvard.edu

Mark Schiefsky

Harvard University

mjschief@fas.harvard.edu

Abstract

Intertextuality, the connection between texts, is a critical literary concept for analyzing classical Latin works. Given the emergence of AI in digital humanities, this paper presents Intertext.AI, a novel interface that leverages Latin BERT (Bamman and Burns, 2020), a BERT model trained on classical Latin texts, and contextually rich visualizations to help classicists find potential intertextual connections. Intertext.AI identified over 80% of attested allusions from excerpts of Lucan’s *Pharsalia*, demonstrating the system’s technical efficacy. Our findings from a user study with 19 participants also suggest that Intertext.AI fosters intertextual discovery and interpretation more easily than other tools. While participants did not identify significantly different types or quantities of connections when using Intertext.AI or other tools, they overall found finding and justifying potential intertextuality easier with Intertext.AI, reported higher confidence in their observations from Intertext.AI, and preferred having access to it during the search process.

1 Introduction

Intertextuality, or the connections and references between texts that impact their meaning and interpretation, is a critical literary concept for the analysis of Latin texts from classical antiquity. Classicists gain new perspectives on ancient works through close reading and searching for allusions: a reference to a previous text as a potential source of inspiration for stylistic choices, semantic concepts, and contextual meaning. These literary connections can be direct, such as verbatim or near-identical quotations, or indirect, through grammatical, metrical, or semantic similarity (Bamman and Crane, 2008; Wills, 1996).

Within the millennia-long tradition of classical Latin scholarship, powerful digital humanities tools for linguistic and literary tasks have emerged in the

last few decades such as morphological and syntactic parsers, digitized manuscript editions, and extensive online dictionaries (Appendix A). Some platforms such as Ingenium (Zhou et al., 2016) are designed to help beginner Latin students grasp foundational grammatical concepts, while others such as the *Thesaurus Linguae Latinae* (1900-) aid experts with comprehensive citations of word usages in Latin.

Still, the higher-order processes of literary analysis and intertextual discovery are primarily analog, aided by commentaries and secondary scholarship, with only a few interfaces such as the Tesseract Project (Coffee et al., 2012) directly proposing digital solutions (Appendix A). Finding allusions between texts and determining whether they are convincing are challenging and subjective inquiries. While some platforms offer advanced searches by similar phrases, poetic meters, and other textual features (Nelis et al., 2017), they do not enable comparisons of the search results in their broader contexts, which can provide macro-level insights that short excerpts do not reveal.

However, developments in transformer models have enabled great strides in automating Latin linguistic tasks. Despite the small extant corpus, models fine-tuned on Latin such as Latin BERT (Bamman and Burns, 2020) and SPhilBERTa (Riemenschneider and Frank, 2023) perform well on tasks like part-of-speech tagging, word sense disambiguation, and semantic similarity retrieval—which can suggest potential intertextuality (Bamman and Crane, 2008). As a result, a range of opportunities, unexplored in existing interfaces, has opened up for analyzing literary connections and augmenting close reading in the Latin language.

Leveraging these advancements in AI for the field of classical studies, we present Intertext.AI, a novel web interface using Latin BERT (Bamman and Burns, 2020) and design choices from popular Latin reading platforms. The interface combines

side-by-side text views with a nearest neighbor query search; users query Latin BERT with an excerpt and target word, for which the model will output the most similar tokens based on the cosine similarity of the model’s word-level contextual embeddings. Users can then view those nearest neighbors in the context of the broader text, starting an inquiry into why the excerpts may be connected. The visual interface is designed to help students quickly develop a sense of which texts may reference each other and contextualize why an AI might register certain sentences as similar.

To evaluate the system’s efficacy in detecting intertextuality, we tested how many allusions attested in classical scholarship Intertext.AI found in select excerpts from Lucan’s *Pharsalia*, a Roman epic poem from the first century CE. We also conducted a user study comparing the efficacy of existing digital tools to Intertext.AI in uncovering potential allusions between texts. We report on the results of both evaluations, which support the ability of Intertext.AI to facilitate intertextual discovery and interpretation by fostering literary comparison.

2 System Design and Feature Usage

Based on formative needfinding conversations (Appendix B), we created Intertext.AI with the following design goal: to enable classicists to **identify potential intertextual correspondences** and parallel constructions across Latin texts to aid comparisons of semantic concepts, themes, and literary features. The interface was implemented using Next.js, Flask, word-level contextual embeddings from Latin BERT (Bamman and Burns, 2020), and a Pinecone vector database, which is queried via a semantic search by cosine similarity. A video demo is available [here](#).¹

The main feature of Intertext.AI is the ability to query for *contextual nearest neighbors*, supported by Latin BERT (Bamman and Burns, 2020, 7-8). Given an excerpt and a target word within it, Latin BERT computes the target word’s contextual embedding and returns the most similar tokens to that word and their context from the corpus, ranked by cosine similarity score (Figure 1). The excerpt is used to identify the sentence in which the target word appears—and thus its use in context. Optionally, the user can use a filter which displays nearest

¹The platform is also available for use at <https://www.ai-latin-close-reading.online/>. The code repository is open source at <https://github.com/ashley-gong/intertext.ai-public>.

neighbors only from the selected texts (it does not recompute embeddings or recalculate scores with a narrower search space). We use an encoder-only model to encourage readers to make their own interpretations of AI-detected similarities. Although more complex LLMs and encoder-decoder models may have the potential to further refine or explain textual correspondences, they introduce the risk of hallucinating false text that can lead readers astray.

After submitting a query, the user can view the sentences that contain the most similar contextualized tokens to the target word and expand the result to read the broader passage in which it is contained (Figure 2). Since the results appear directly next to the original text from which the user inputs a query—and the interface highlights the query in this passage²—a user can compare a result horizontally with the original passage to determine whether the similarity is compelling or vertically with other results to infer a pattern between the AI outputs. The target word is highlighted in yellow to contrast with the blue highlight of the query context, and the result tokens are highlighted either in the same yellow if the result found an instance of the same lemma (root word), or in red if the result token is from a different lemma. Further, within the result’s broader passage, lemmas in common with lemmas from the query excerpt are rendered in orange to emphasize how shared words may communicate contextual similarity (Appendix D). The accentuation of common words between passages is inspired by the similar visualization on the Tesseræ Project interface (Okuda et al., 2022). Intertext.AI uses the Latin lemmatizer from the Stanza library to identify these lemmas (Qi et al., 2020).

Beyond viewing the query and results within their broader contexts, users can also read any passage adjacent to another in a dual text view, enabling them to make aligned comparisons across the original form of the text. Further, an English translation from the Perseus Digital Library (Crane, 2023) accompanies each text in a movable pop-up or a side-by-side scrolling view. Finally, each query returns a histogram that visualizes the distribution of the top 100 cosine similarity scores between the query’s target word and other tokens from the corpus on which Latin BERT is trained. Summary statistics such as the mean, maximum, minimum, and standard deviation, are also included to help

²After a user submits a query, the query’s highlight within the original passage is visible in the single/query view, dual view, and the full text/translation view.

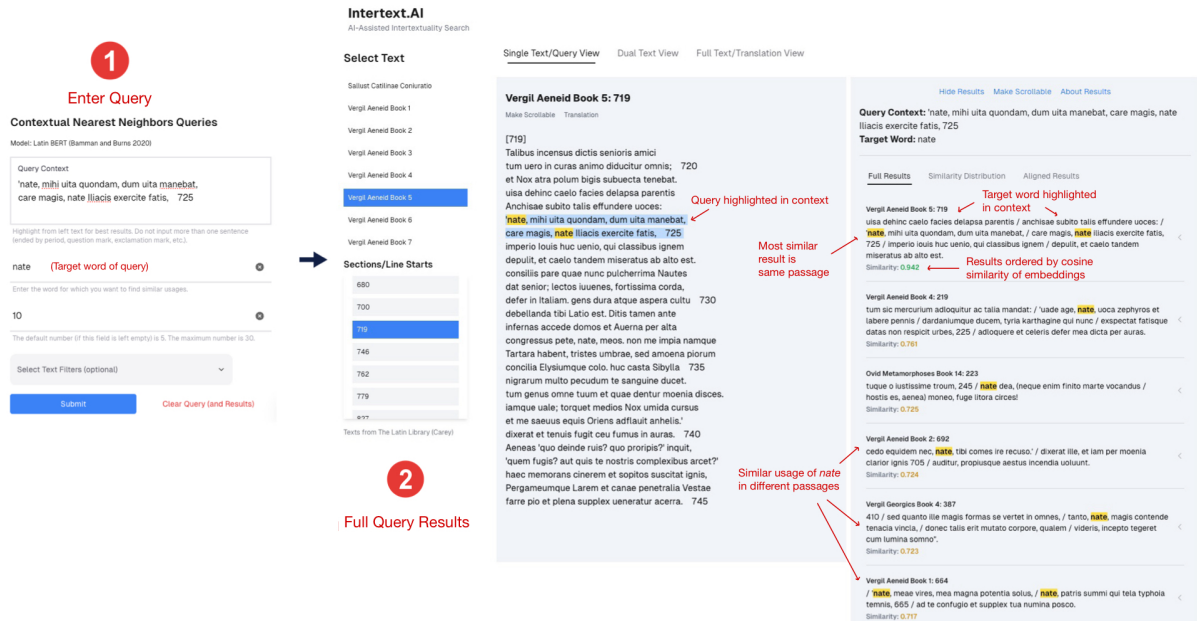


Figure 1: Latin BERT returns excerpts based on the cosine similarity with the query’s target word that may indicate similar word usages, semantic meaning, or stylistic structure. The query at *Aeneid* 5.724-5, *nate, mihi uita quondam, dum uita manebat, / care magis, nate Iliacis exercite fatis* (Son, once dearer to me than life, while life was still remaining for me, son, vexed by Trojan fates), targets the word *nate* ("son"), and the tool returns similar occurrences of *nate* in dialogues and its double usage in adjacent lines (last result).

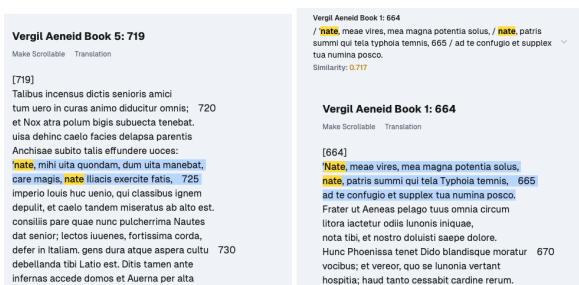


Figure 2: Expanding a result reveals context beyond the immediate phrase, allowing comparison with the highlighted query on the left. The word *nate* (“son”) is similarly used twice in adjacent lines, demonstrating the urgency of the address towards the son in question. The added context around the highlighted lines distinguishes who the son is and shows other nuances in each passage.

contextualize how much more similar the top k results are to the query target than other words in the corpus. See Appendix D for additional figures.

3 Detection of Attested Intertexts

To create a “ground truth” dataset of attested intertextual correspondences found in the first book of Lucan’s *Pharsalia*, we compiled references to other Roman texts from two well-known commentaries on the text: Roche (2009) and Getty (1940). The analysis focuses on references within lines 1.8-32

and 1.67-97, which note 109 distinct connections across both commentaries.

A reference was considered successfully found if it appeared in the top 30 most similar results³ when querying the Lucan excerpt in which it appears in a commentary, using any of the words in the Lucan phrase as the target word, and with or without a text filter used for the non-Lucan referenced text. Using these criteria, Intertext.AI found 89 of the 109 connections from both commentaries, or 81.65% (Appendix D), offering a convincing case for the success with which Latin BERT and the contextual nearest neighbor search engine can find attested allusions when queried.

4 User Study Methodology

We conducted a within-subjects user study with 19 participants with at least an intermediate level of classical Latin study (a mean of over 7.42 years,⁴ $SD=2.32$) to understand how Intertext.AI may help students identify intertextual allusions. All participants were undergraduates except for P2 (Classics PhD candidate), P12 (law student), and P16 (Clas-

³Thirty is the maximum number of results a user can view in detail on Intertext.AI before the latency is noticeably slow.

⁴P16 reported their experience as “over 10 years” as they did not recall an exact number, so we calculated the average using 10 as their number of years of experience.

sics professor), and none had used AI to explore intertextuality before this study.

Participants completed two reading tasks in two conditions, with semi-structured interviews after each stage (more details on study procedure in Appendix C). For 10 minutes, participants were asked to find intertextual connections between two pairs of Latin poems, recording any phrase pairings that could be a convincing allusion due to word-level, grammatical, semantic, or stylistic similarities (based on [Bamman and Crane, 2008](#)).

In the control condition, participants were allowed to use any online tool designed for Latin. Links to suggested resources were provided, such as the study's designated texts (and translations) in the Loeb Classical Library ([Henderson and Loeb, 2024](#)) or the Latin Library ([Carey, 2021](#)) and Logeion ([Goldenberg and Shanahan, 2024](#)), a digitized Latin dictionary. For the treatment condition, participants completed the same reading task by querying Intertext.AI. Other tools were still available for participants throughout the task, as classicists would likely use many resources during their real research processes. Condition order was counterbalanced such that 10 participants began with the baseline condition while the rest started with the treatment condition; the pairs of texts used for each task were also randomly assigned.

After each task, participants completed a short survey with Likert scale questions (on a 5-point scale) about their task experience with and without Intertext.AI, their confidence in the textual connections they noted, and their evaluation of the usefulness of each Intertext.AI feature. We also asked a few open-ended questions during which participants could verbally provide feedback and notes on their subjective user experience (Appendix C).

5 Results

Besides the intertextual connections participants recorded, we collected both quantitative metrics and qualitative observations about participants' performance, cognitive load, and confidence during their intertextual inquiries. Connections that participants proposed with texts outside of the designated pairs were excluded. We used the Wilcoxon Signed-Rank test (with $\alpha=0.05$) to evaluate statistical significance, as the data was within-subjects and could not be assumed to be normally distributed.

Since Intertext.AI aims to aid the discovery of intertexts, it remains the reader's task to discern

a meaningful correspondence from a more mundane or accidental similarity using the visual aids and context provided on the interface. For this reason, we do not evaluate the supposed "quality" of each intertextual connection along a benchmark of cogency, as the goal is that each participant finds Intertext.AI useful for finding textual parallels they themselves regard as notable.

5.1 Participants' Task Output

All participants found at least one potential textual connection in both tasks. Participants recorded an average of 2.42 connections in the control task ($SD=1.57$) and 2.16 in the treatment task ($SD=1.17$). The median for both conditions is 2. The difference in the number of connections found between the control and treatment conditions is not statistically significant.

We also classified the observed potential connections in terms of thematic, lexical, syntactical, or stylistic similarities.⁵ Table 1 displays the mean number of parallels found of each type. Further, many participants stated that they approached searching for intertextual correspondences differently in the two conditions. P1, P14, and P19 mentioned that in the control condition, they primarily searched for correspondences using English translations from the Loeb Classical Library, which led to the discovery of more thematic similarities, rather than linguistic connections grounded in the Latin. When using Intertext.AI, however, most participants prioritized comparing the Latin texts themselves. Four participants found Intertext.AI more useful for finding words that are close in meaning but not identical. P6 found "particularly similar forms" of words with Intertext.AI, and P15 was "impressed" that a query picked up results with the target word itself and with different words but in "similar contexts," demonstrating both lexical and thematic correspondences.

5.2 Confidence, Ease of Use, and Preference

Participants reported higher ease of finding intertextual connections, higher ease of justifying connections, and higher confidence in connections in the treatment condition (Table 2). The difference in scores was statistically significant for ease of detection ($W=15, p=0.002$) and ease of justification ($W=11, p=0.045$) but not for connection confidence ($W=16, p=0.058$).

⁵See Appendix D for example connections and the distribution found for each type.

Connection Type	Control Mean	Treatment Mean
Lexical	1.37 (SD=1.74)	1.58 (SD=1.66)
Syntactic	0.11 (SD=0.11)	0.05 (SD=0.06)
Thematic	0.89 (SD=0.89)	0.47 (SD=0.45)
Stylistic	0.05 (SD=0.06)	0.05 (SD=0.06)
Total	2.42 (SD=1.57)	2.16 (SD=1.17)

Table 1: The mean number of lexical, syntactic, thematic, and stylistic connections participants found in each condition.

	Control	Treatment
Ease of Finding Connections*	2.58 (SD=1.22)	4.16 (SD=0.83)
Ease of Justifying Connections*	3.58 (SD=1.12)	4.11 (SD=0.99)
Confidence in Connections	3.37 (SD=1.26)	3.84 (SD=0.96)

Table 2: The means of participants’ self-reported Likert scores about the ease of experience (cognitive load) and confidence in task output. Asterisks indicates questions with a statistically significant difference in scores.

Most participants (n=11) expressed that their familiarity with the texts used in the study—or lack thereof—impacted their ability to observe and explain many potential intertextual correspondences. P17 noted that the task of finding connections “was tougher because of less familiarity with the text,” while P2 mentioned that they “benefited from knowing one of these texts incredibly well.”

Ultimately, nearly all participants strongly agreed (n=15) or agreed (n=2) that they would use Intertext.AI again in their future endeavors in classical research (M=4.53, SD=0.90). Despite the lower average connections found in the treatment condition, five participants stated that searching for intertextuality with Intertext.AI felt more efficient. For example, Intertext.AI enabled P4 to find textual connections when they “hadn’t even read one of [the texts],” thus making the exploration more efficient by circumventing the need “to have read everything, ever, to be able to find intertexts.” Most participants (n=17) also stated that they would prefer to have access to Intertext.AI than not when conducting intertextual searches (M=4.36, SD=1.16).

5.3 Engagement with Interface Features

Many participants commended the usability of the interface and feedback on features they wished Intertext.AI offered. Four participants (P8, P10, P12, P15) found the ability to read texts directly adjacent to each other in the dual text display very useful, and four participants (P4, P8, P14, P15) praised the additional context provided in the full search results of the nearest neighbor queries. Table 3 lists

the number of participants who used each feature along with the mean Likert score given for each feature. When asked about interface improvements, seven participants suggested including line numbers in the English translations on Intertext.AI for easier coordination with the Latin texts, and three participants (P11, P12, P19) proposed a feature to gloss individual words in the text.

Feature	% Who Used	Mean Score for Efficacy	
		Close Reading	Intertextual Discovery
Dual Text View	84.21%	4.25 (SD=0.94)	3.94 (SD=1.30)
Full Translation View	26.32%	4.40 (SD=1.07)	3.89 (SD=1.27)
Pop-up Translation	63.16%	3.92 (SD=1.16)	3.82 (SD=1.17)
Full Query Results	100%	–	4.37 (SD=0.76)
Aligned Query Results	57.89%	–	4.18 (SD=1.40)
Similarity Score Distribution	31.58%	–	3.83 (SD=1.60)

Table 3: The means of Likert scores about the efficacy of various Intertext.AI features for close reading and intertextual exploration, along with the percentage of participants who used each feature (out of 19).

6 Discussion and Future Work

Our findings suggest that Intertext.AI successfully helped participants find intertextual connections they were confident about, supporting the initial design goal. Although the average number of connections was lower for the treatment condition than the control, participants’ greater confidence in the parallels they found and higher ease of finding them with Intertext.AI suggested that the interface can lead readers to more easily make more fruitful literary comparisons. Since many felt that familiarity with the texts influenced the ease of the search, Intertext.AI would likely be most useful to those who already have some previous experience conducting intertextual inquiries and know what words could appear across different texts. Limitations in the model interpretability of Latin BERT, the scope of the ground truth evaluation, and the user study sample population’s size and variability necessitate more extensive evaluations of the system. Future work could enhance Intertext.AI and research in this AI-Classics intersection by incorporating multilingual—particularly ancient Greek—classical capabilities, fine-tuning the model with attested allusions, improving phrase-level search, and investigating whether generative LLMs, after mitigating the possible hallucinations of false Latin text or justifications, could help automate more convincing explanations and proposals of potential allusions.

References

- David Bamman and Patrick J. Burns. 2020. [Latin BERT: A Contextual Language Model for Classical Philology](#). *arXiv preprint*. ArXiv:2009.10053 [cs].
- David Bamman and Gregory Crane. 2008. The Logic and Discovery of Textual Allusion.
- Alessandro Barchiesi. 1993. [Future Reflexive: Two Modes of Allusion and Ovid's *Heroides*](#). *Harvard Studies in Classical Philology*, 95:333–365.
- Alessandro Barchiesi. 2001. *Speaking Volumes: Narrative and Intertext in Ovid and Other Latin Poets*. Duckworth, London.
- Patrick J. Burns, James A. Brofos, Kyle Li, Prमित Chaudhuri, and Joseph P. Dexter. 2021. [Profiling of Intertextuality in Latin Literature Using Word Embeddings](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4900–4907, Online. Association for Computational Linguistics.
- William Carey. 2021. [The Latin Library](#).
- 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(2):383–422.
- Emanuela Colombi, Luca Mondin, Luigi Tassarolo, and Bacianini. 2011. [Pedecerto](#).
- Gregory Crane. 2023. [The Perseus Digital Library and the Future of Libraries](#). *International Journal on Digital Libraries*, 24(2):117–128.
- Katy Ilonka Gero, Chelse Swoopes, Ziwei Gu, Jonathan K. Kummerfeld, and Elena L. Glassman. 2024. [Supporting Sensemaking of Large Language Model Outputs at Scale](#). In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–21, Honolulu HI USA. ACM.
- Robert J. Getty, editor. 1940. *M. Annaei Lucani. De Bello Civili: Liber I*. Cambridge University Press, Cambridge.
- Josh Goldenberg and Matt Shanahan. 2024. [Logeion](#).
- Jeffrey Henderson and James Loeb. 2024. [Loeb Classical Library](#).
- Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. 2015. [On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges](#). In *Eurographics Conference on Visualization (EuroVis) - STARS*. The Eurographics Association.
- Massimo Manca, Linda Spinazzè, Paolo Mastandrea, and Luigi Tassarolo. 2011. [Musique Deoque: Text Retrieval on Critical Editions](#). *Journal for Language Technology and Computational Linguistics*, 26(2):129–40.
- Jeremy March. 2024. [philolog.us](#).
- Aditi Muralidharan, Marti A. Hearst, and Christopher Fan. 2013. [WordSeer: A Knowledge Synthesis Environment for Textual Data](#). In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 2533–2536, San Francisco California USA. ACM.
- Damien Nelis, Christopher W Forstall, and Lavinia Galli Milić. 2017. [Intertextuality and Narrative Context: Digital Narratology?](#) *Journal of Data Mining and Digital Humanities*.
- Nozomu Okuda, Jeffery Kinnison, Patrick Burns, Neil Coffee, and Walter Scheirer. 2022. [Tesserae Intertext Service](#). *DHQ: Digital Humanities Quarterly*, 16(1).
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. [Stanza: A python natural language processing toolkit for many human languages](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.
- Frederick Riemenschneider and Anette Frank. 2023. [Graecia capta ferum victorem cepit: Detecting Latin Allusions to Ancient Greek Literature](#). *arXiv preprint*. ArXiv:2308.12008 [cs].
- Paul A. Roche. 2009. *Lucan, De Bello Ciuili, Book I*. Oxford University Press, Oxford.
- Scaife Viewer|Home. [Scaife Viewer|Home](#). Accessed February 7, 2025.
- Thesaurus Linguae Latinae. 1900-. *Thesaurus Linguae Latinae*. Walter de Gruyter GmbH, Berlin.
- Jeffrey Wills. 1996. *Repetition in Latin Poetry: Figures of Allusion*. Clarendon Press, Oxford.
- Tariq Yousef and Stefan Janicke. 2021. [A Survey of Text Alignment Visualization](#). *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1149–1159.
- Sharon Zhou, Ivy J. Livingston, Mark Schiefsky, Stuart M. Shieber, and Krzysztof Z. Gajos. 2016. [Ingenium: Engaging Novice Students with Latin Grammar](#). In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 944–956, San Jose California USA. ACM.

A Related Work

A.1 Augmented Reading Interfaces

Platforms designed to encourage sensemaking—the process of gleaning meaning (Muralidharan et al., 2013)—often use spatial text alignment principles (Yousef and Janicke, 2021) to encourage visual comparisons. Gero et al. (2024) developed an interface to facilitate readers’ sensemaking of

English text generated by large language models (LLMs) and a novel algorithm that determines similarities in tokens' contextual positions and semantic content. When LLMs stochastically produce varying outputs to the same prompts, aligned visualizations help users understand and evaluate the quality of model responses. *WordSeer* (Muralidharan et al., 2013) also takes a visual approach to sensemaking through the lens of distant reading (Jänicke et al., 2015) by allowing users to view texts in slices—in a tree form or filtered by syntactic quality—and statistical metadata about a corpus such as word frequencies and grammatical constructions. The interface introduced in this paper, rather than facilitating multiple stages of the sense-making process, focuses on the initial exploration of intertextuality a Latin student may undertake by keeping the text as unaltered as possible during the comparison of phrases in context.

A.2 Digital Tools for Reading Classical Latin

There are abundant tools online for Classics students and scholars to facilitate Latin reading, many of which have digitized reliable editions of canonical Latin works. The Perseus Digital Library (Crane, 2023) offers one of the most comprehensive online collections of classical texts and allows users to click on any Latin word in its reading interface to display its morphological parsing and definition via the Word Study Tool. More recently, Perseus has released the Scaife Viewer (Scaife Viewer|Home), a reading environment with the same word-level parsing, and now a side-by-side display of works and their translations. Other databases include the Latin Library (Carey, 2021) and the Loeb Classical Library (Henderson and Loeb, 2024), a collection of digitized critical editions with adjacent translations. Intertext.AI builds on top of these foundational digital features, particularly the side-by-side display, by integrating AI to suggest potential instances of intertextuality.

Online dictionaries and Latin learning assistants beyond digital editions also help classicists read and learn Latin. The *Thesaurus Linguae Latinae* (1900-) is an international project to define and cite all usages of every lemma in the Latin corpus, and other platforms such as Logeion (Goldenberg and Shanahan, 2024) and philolog.us (March, 2024) have digitized reputable Latin dictionaries and support queries by lemma. Finally, *Ingenium* (Zhou et al., 2016) helps beginner Latin students develop grammatical understanding through a block-based

interface; blocks contain inflected Latin words that snap into place when a clause is grammatically sound. The AI-enhanced system offered by Intertext.AI targets students more advanced than those who are still learning Latin grammar to encourage the first steps of higher-order literary analysis through textual allusions.

A.3 Digital Intertextual Discovery in Classical Languages

Nelis et al. (2017) provide a survey of existing digital tools classicists use to assist their search for intertextuality. The *Musique Deoque* (Manca et al., 2011) provides a more advanced search for word sequences and their variants in Latin poetry to stimulate comparisons of diction and meter. *Pedecerto* (Colombi et al., 2011) similarly enables poetry searches by metrical pattern, word forms, and other more advanced features. The tool that resembles the interface in this paper the most is the Tesseræ Project (Coffee et al., 2012), which can return a ranked list of potential parallels between two texts through word-level n-gram matching. Intertext.AI, in contrast, searches for intertextuality using a transformer model, Latin BERT (Bamman and Burns, 2020), and enables users to view potential references in the context of the entire text, rather than in isolated snippets.

Beyond Latin BERT, Riemenschneider and Frank (2023) trained SPhilBERTa, a multilingual sentence BERT model that can detect cross-lingual similarities—many of which are known to scholars as allusions—across English, Latin, and Ancient Greek. Furthermore, Burns et al. (2021) trained static embeddings on Latin, rather than contextual BERT embeddings, to detect synonyms and stylistic similarities within a small corpus with around 87 percent accuracy, leveraging a dataset of scholar-supported intertextual parallels from Latin epic poetry. This paper introduces the novel integration of a Latin transformer model with a user interface designed for intermediate to advanced students as well as domain experts.

B Needfinding

Conversations with four students with experience reading Latin and one faculty instructor of beginner to intermediate Latin informed the need thesis for this system by reinforcing current gaps in digital tools for Latin. Students noted that they most often explore intertextuality between two passages

known to contain references; they reason about the exact connections after this starting point. A flexible side-by-side display of any two texts that can highlight potentially similar matches could thus effectively support these comparisons. With an interface that identifies similar sentences from various texts and visualizes them in parallel with the original, students can more quickly develop a sense of what phrases may be alluding to each other.

Two students also commented on the usefulness and comprehensive scope of the *Thesaurae Linguae Latinae* (*Thesaurus Linguae Latinae*, 1900-) for finding word usages in various works but noted usability challenges due to the definitions being in Latin and an overwhelming amount of data to sift through. The organization of online dictionaries (Goldenberg and Shanahan, 2024; March, 2024) by lemma also makes it difficult to search for a contextualized word in a certain passage or for inflected variants of words in context. Many intertextual references go beyond a single word to incorporate a phrase-level match or similar semantic concepts, so enabling a multi-word query within which a user can pinpoint a word captures this variation in allusion and contextualizes word searches more effectively.

Ultimately, these discussions identified challenges in conducting contextualized inquiries about intertextual allusions without having to filter an excessive list of results. Intertext.AI aims to address this gap by facilitating comparisons with interactive side-by-side displays and AI-augmented, context-driven searches for similar usages of Latin vocabulary.

C Additional Procedure Details and Interview Questions

This study received IRB approval under Harvard University's IRB25-0130. Participants were recruited primarily within the Harvard Classics department through emails to the department's mailing list, undergraduate student mailing lists, individual students who expressed interest, and faculty who agreed to share the opportunity with their students. The eligibility requirements consisted of being fluent in English, over 18 years of age, and having a reading level sufficiently proficient for upper-level Latin language courses in the Harvard Classics curriculum, or approximately 4 years of study (though having over 4 years of formal Latin study was not required). Before the study, partici-

pants indicated their informed consent via written or electronic signature and completed a survey to explain their background and experience reading classical Latin, as well as their familiarity with the texts used in the study. If a participant indicated a familiarity level of less than 2 out of 5 on a Likert scale (1=unknown text, 5=very familiar), they were offered the option to read short synopses for each text prior to starting the tasks to learn an overview of the narrative. Further, participants received a short tutorial on how to use each feature of Intertext.AI and interpret the results of the contextual nearest neighbor query engine at a basic level before the start of the treatment condition.

The pairs of texts chosen for these tasks were Ovid's *Heroides* 7 with Vergil's *Aeneid* Book 4 and Ovid's *Heroides* 10 with Catullus 64. The *Heroides*, a collection of fictional letters written from the perspectives of scorned mythological heroines to their lovers, were selected for their comparable reading difficulty, genre, content, and documented intertextual influences from preceding Roman texts (Barchiesi, 1993, 2001). According to self-reported Likert scores on a scale from 1 to 5 about their familiarity with the tasks' texts, participants, though varying widely, were on the whole more familiar with the *Aeneid* and the Catullus, which are often taught in school as central texts in the classical Latin canon, than the *Heroides*, which are read relatively less frequently.

C.1 Pre-study Background Survey Questions

To help me understand how Intertext.AI could be used by people at different stages of their Latin learning, please describe your background/comfort level in reading classical Latin (more specifically, works/language generally between 100 BCE and 200 CE).

1. How many years have you studied classical Latin?
2. What Latin courses (if any) have you taken at Harvard?⁶
3. Rank your familiarity with the following text on a scale of 1-5 (1=unknown, 5=very familiar): Ovid *Heroides* 7.
4. Rank your familiarity with the following text on a scale of 1-5 (1=unknown, 5=very familiar): Virgil *Aeneid* (Book 4).

⁶This question was asked because the study population was sampled from Harvard.

5. Rank your familiarity with the following text on a scale of 1-5 (1=unknown, 5=very familiar): Ovid *Heroides* 10.
6. Rank your familiarity with the following text on a scale of 1-5 (1=unknown, 5=very familiar): Catullus 64.
7. Any additional context that would be useful to know about your background in studying Latin?

C.2 Post-task Interview Questions: Control

The following Likert-scale questions were asked in a short survey following the reading task (described in Chapter 4.1) in the control condition, during which they were provided links to the Loeb Classical Library (Henderson and Loeb, 2024) and Logeion (Goldenberg and Shanahan, 2024)—though other resources were permitted for use as well.

1. On a scale of 1 to 5 (0=not used, 1=strongly disagree, 3=neutral, 5=strongly agree), I found the Loeb Classical Library texts easy to use and helpful for close reading the texts.
2. On a scale of 1 to 5 (0=not used), the Logeion online dictionary assisted me with close reading the texts.
3. I found the Loeb Classical Library texts usable and helpful for discovering potential instances of intertextuality.
4. The Logeion online dictionary assisted me with discovering potential instances of intertextuality.
5. It was easy to find potential instances of intertextuality.
6. It felt natural to explain or justify the connections I listed.
7. I feel confident about the strength of the connections I listed.

Participants also answered the following open-ended questions verbally:

1. Please explain your experience using the tools you did.
2. If there was anything you wish the tools you used could do to facilitate this process, what would that be?
3. Any other thoughts, feedback, opinions, etc.?

C.3 Post-task Interview Questions: Treatment

The following Likert-scale questions were asked in a short survey following the reading task (described in Chapter 4.1) in the treatment condition, during which they were instructed to use Intertext.AI.

1. On a scale of 1 to 5 (1=strongly disagree, 3=neutral, 5=strongly agree), I found Intertext.AI easy to use.
2. On a scale of 1 to 5, the dual text view of Intertext.AI was useful for close reading the texts. (0 = not used)
3. On a scale of 1 to 5, the dual text view of Intertext.AI was useful for finding potential instances of intertextuality. (0 = not used)
4. The inclusion of side-by-side translations on Intertext.AI was useful for close reading the texts. (0 = not used)
5. The inclusion of side-by-side translations on Intertext.AI was useful for finding potential instances of intertextuality. (0 = not used)
6. The inclusion of movable and collapsible translations on Intertext.AI was useful for close reading the texts. (0 = not used)
7. The inclusion of movable and collapsible translations on Intertext.AI was useful for finding potential instances of intertextuality. (0 = not used)
8. The full query results were informative and helpful for determining potential intertextual allusions.
9. The aligned query results were informative and helpful for determining potential intertextual allusions. (0 = not used)
10. The distribution of similarity scores were informative and helpful for understanding the degree of similarity of each query results. (0 = not used)
11. It was easy to find potential instances of intertextuality using Intertext.AI.
12. It felt natural to explain or justify the connections I listed.
13. I feel confident about the strength of the connections I listed.

14. I would use this interface again in my Classics/Latin research endeavors.
15. I prefer having access to and using this interface to only using a dictionary or commentary to find intertextual connections.

Similar to the control condition, participants received the following open-ended questions verbally:

1. Please explain your experience using Intertext.AI.
2. If there was anything you wish Intertext.AI could have done to facilitate the task, what would that be?
3. Any other thoughts, feedback, opinions, etc.?

D Additional Tables and Figures

The screenshot shows the Intertext.AI interface. On the left, there is a snippet of text from Vergil's Aeneid Book 5, line 719. The text is: "Talibus incensus dictis senioris amici tum uero in curas animo diducitur omnis; 720 et Nox atra polum bigis subuecta tenebat. uisa dehinc caelo facies delapsa parentis Anchisae subito talis effundere uoces: **nate, mihi uita quondam, dum uita manebat, care magis, nate** Iliacis exercite fatis, 725 imperio louis huc uenio, qui classibus ignem depulit, et caelo tandem miseratus ab alto est. consiliis pare quae nunc pulcherrima Nautes dat senior; lectos iuuenes, fortissima corda, defer in Italiam. gens dura atque aspera cultu 730 debellanda tibi Latio est. Ditis tamen ante infernas accede domos et Auerna per alta congressus pete, nate, meos. non me impia namque Tartara habent, tristes umbrae, sed amoena piorum concilia Elysiumque colo. huc casta Sibylla 735 nigrarum multo pecudum te sanguine ducet. tum genus omne tuum et quae dentur moenia disces. iamque uale; torquet medios Nox umida cursus et me saeuus equis Oriens adflauit anhelis.' dixerat et tenuis fugit ceu fumus in auras. 740 Aeneas 'quo deinde ruis? quo proripis?' inquit, 'quem fugis? aut quis te nostris complexibus arceat?' haec memorans cinerem et sopitos suscitauit ignis, Pergameumque Larem et canae penetralia Vestae farre pio et plena supplex ueneratur acerra. 745".

On the right, the search results are displayed. The query context is: "nate, mihi uita quondam, dum uita manebat, care magis, nate Iliacis exercite fatis, 725". The target word is "nate". The results show a similarity score of 0.602. The results are vertically aligned with the original text, showing a high similarity score of 0.602 for the target word 'nate'. The results are: "Full Results", "Similarity Distribution", "Aligned Results". The results are: "Catullus: 64", "o decus eximium magnis uirtutibus augens, / emathiae tutamen, opis carissime **nato**, / accipe, quod laeta tibi pandunt luce sorores, / ueridicum oraclum: sed uos, quae fata sequuntur, / currite ducentes subtegmina, currite, fusi.", "Similarity: 0.602".

Below the search results, there is a section for "Catullus: 64" with a similarity score of 0.563. The text is: "namque ferunt olim, classi cum moenia diuae / linquentem gnatum uentis concrederet aegaeus, / talia complexum luueni mandata dedisse: / **gnate** mihi longa iucundior unice uita, / **gnate**, ego quem in dubios cogor dimittere casus, / reddite in extrema nuper mihi fine senectae, / quandoquidem fortuna mea ac tua feruida uirtus / eripit inuito mihi te, cui languida nondum / lumina sunt gnati cara saturata figura, / non ego te gaudens laetanti pectore mittam, / nec te ferre sinam fortunae signa secundae, / sed primum multas expromam mente querelas, / canitiem terra atque infuso puluere foedans, / inde infecta uago suspendam lintea malo, / nostros ut luctus nostraeque incendia mentis / carbasus obscurata dicet ferrugine hibera.", "Similarity: 0.563".

Figure 3: A user can compare particular results to each other vertically and to the original passage vertically, which follows the screen cursor as the user scrolls. The result is from Catullus 64.215-6, *gnate mihi longa iucundior unice uita, / gnate, ego quem in dubios cogor dimittere casus* ("My only son, sweeter to me than my long life, son, I am forced to send you into doubtful circumstances"). The user can take note of the similar placement of the target words at the front of the lines and the common lemmas *mihi* and *uita*, bolded in orange, indicating similar semantic and lexical material. The red highlight on *gnate*, despite being the same root word as *nate*, highlights the usage of a different (and archaic) form of *nate*.

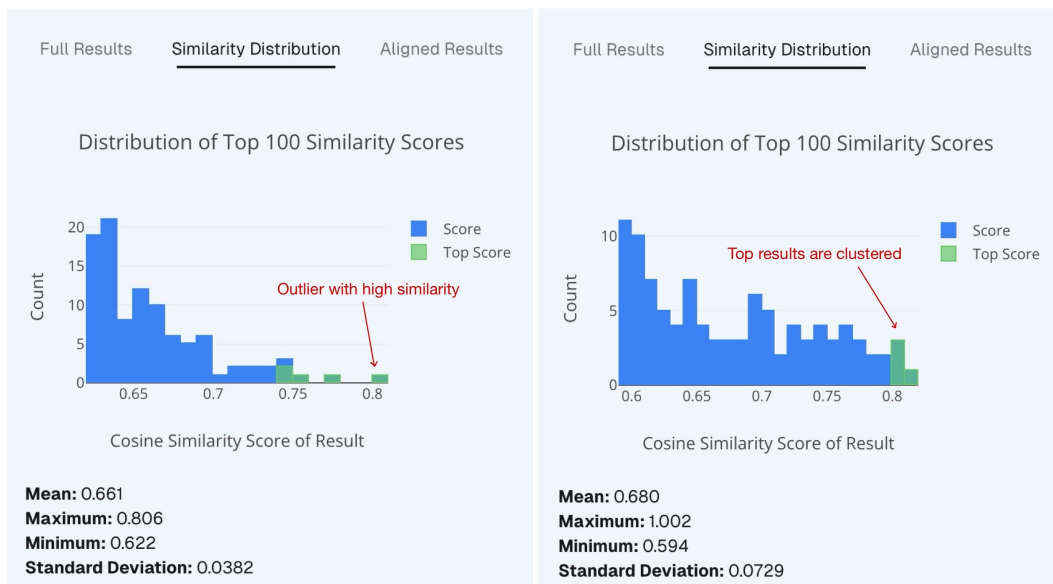


Figure 4: A comparison between two distributions of the top 100 results' similarity scores for two nearest neighbor queries. On the left, the result with the highest score stands out in the histogram, indicating a potential outlier to explore further, while on the right, the top results are all clustered around the same values.

[Hide Results](#) [Make Scrollable](#) [About Results](#)

Query Context: ad effeminandos animos
Target Word: effeminandos

Full Results	Similarity Distribution	Aligned Results
Caesar Gallic Wars Book 1: 1	...mercatores saepe commeant atque ea quae ad	effeminandos animos pertinent important, prox...
Cicero In Catilinam 3: 20	...f decem dies facti sunt, neque res ulla, quae ad	placandos deos pertineret, praetermissa est....
Caesar Civil Wars Book 3: 79	...eleritati studebat, et suis ut esset auxilio, et ad	opprimendos adversarios ne occasione tempori...
Caesar Gallic Wars Book 2: 21	...caesar, necessariis rebus imperatis, ad	cohortandos milites, quam [in] partem fors obt...
Caesar Gallic Wars Book 6: 43	...caesar rursus ad	vexandos hostes profectus magno coacto num...

Figure 5: A user can view results aligned by the most similar tokens returned by the model to facilitate visual comparison. All result tokens have an "-nd-" infix, indicating a gerundive form, and *ad* ("for the sake of") precedes them, demonstrating the identical grammatical construction.

Catullus: 1
Make Scrollable Translation

[1] I. ad Cornelium

Cui dono lepidum novum libellum
arida modo pumice exolitum?
Corneli, tibi: namque tu solebas
meas esse aliquid putare nugas.
Iam tum, cum ausus es unus Italorum
omne aevum tribus explicare cartis...
Doctis, Iuppiter, et laboriosis!
Quare habe tibi quidquid hoc libelli—
qualecumque, quod, o patrona virgo,
plus uno maneat perenne saeclo!

Catullus ×

To whom inscribe my charming new book—just out and with ashen pumice polished? Cornelius, to you! for you used to deem my triflings of account, and at a time when you alone of Italians dared unfold the ages' abstract in three chronicles—learned, by Jupiter!—and most laboriously written. Therefore take this booklet, such as it is, and, O Virgin Patroness, may it outlive generations more than one.

Translations from Perseus Scaife Viewer (2024) Close

Figure 6: A user can open a small, movable window containing a translation of the current passage.

Translations from Perseus Scaife Viewer (2024)

Caesar Gallic Wars Book 1
Make Scrollable

[1] Gallia est omnis divisa in partes tres, quarum unam incolunt Belgae, aliam Aquitani, tertiam qui ipsorum lingua Celtae, nostra Galli appellantur. Hi omnes lingua, institutis, legibus inter se differunt. Gallos ab Aquitanis Garumna flumen, a Belgis Matrona et Sequana dividit. Horum omnium fortissimi sunt Belgae, propterea quod a cultu atque humanitate provinciae longissime absunt, minimeque ad eos mercatores saepe commeant atque ea quae ad effeminandos animos pertinent important, proximique

Caesar Gallic Wars Book 1
Make Scrollable

All Gaul is divided into three parts, one of which the Belgae inhabit, the Aquitani another, those who in their own language are called Celts, in our Gauls, the third. All these differ from each other in language, customs and laws. The river Garonne separates the Gauls from the Aquitani; the Marne and the Seine separate them from the Belgae. Of all these, the Belgae are the bravest, because they are furthest from the civilization and refinement of [our] Province, and merchants least frequently resort to them, and

Figure 7: A user can read a Latin text in full with an adjacent translation.

Type	Connected Phrases (Found in Control)	Explanation of Connection
Lexical	<p><i>tibi litore mitto / unde tuam sine me vela tulere ratem</i> (Ov. Her. 10.5-6) "I send this [letter] to you from the shore where your sails carried off your ship without me"</p> <p><i>tempore Theseus / egressus curvis e litoribus Piraei</i> (Cat. 64.73-4) "At that time, Theseus, having left the winding shores of Piraeus"</p>	"Litore, litoribus, lots of ocean and wave language, setting out from the shores" (P8) ("shore")
Syntactic	<p><i>Thesea prensuras semisupina manus: nullus erat. referoque manus iterumque retempto perque torum moveo brachia: nullus erat.</i> (Ov. Her. 10.12-4) "Half-laying down, I [moved] my hands to lay hold of Theseus: there was no one. And I move my arms all over the bed: there was no one."</p> <p><i>o nimis optato saeculorum tempore nati heroes, salvete deum genus! o bona matrum</i> (Cat. 64.22-3) "O heroes, born at the most desired time of the ages, hail, offspring of the gods! O good [children] of mothers"</p>	"Parallelism in both" (P17), <i>nullus erat</i> and <i>o...</i> repetition ("there was no one", "O...")
Thematic	<p><i>uror, ut inducto ceratae sulphure taedae</i> (Ov. Her. 7.23) "I am inflamed [with love], as torches covered with wax that are dipped in sulfur"</p> <p><i>at regina gravi iamdudum saucia cura vulnus alit venis et caeco carpitur igni.</i> (Virg. Aen. 4.2) "But the queen, long since wounded by a heavy love feeds her wound with her veins and is seized by a hidden fire."</p>	"Theme of fire, burning with love" (P13)
Stylistic	<p><i>scelerate revertere Theseu!</i> (Ov. Her. 35) "Return, wicked Theseus!"</p> <p><i>divos scelerare Penates</i> (Cat. 64.19-21) "to defile the divine household"</p>	"[scelerate/scelero] rarely used in Catullus, epithet in Ovid" (P5) ("wicked"/"defile")

Table 4: An example of each connection type found during participants' control conditions. The query is listed in the row above the result, and the words of interest are bolded in the Latin excerpts.

Type	Connected Phrases (Found in Treatment)	Explanation of Connection
Lexical	<p><i>ut rate felici pacata per aequora labar;</i> <i>temperet ut ventos Aeolus; exul ero.</i> (Ov. <i>Her.</i> 10.67-8) “Even if I glide through calm seas on a lucky boat and Aeolus tempers the winds; I will still be an exile.”</p> <p><i>quae simul ac rostro uentosum / proscidit aequor</i> (Cat. 64.12-3) “As soon as she split the windy sea with the ship’s beak”</p>	Collocation of the sea “with <i>rate/rostro</i> , <i>ventosum/ventos</i> ” (P16) (“boat”/“ship’s beak”, “windy”/“winds”)
Syntactic	<p><i>nec nova Karthago, nec te crescentia tangunt</i> <i>moenia nec scepro tradita summa tuo?</i> (Ov. <i>Her.</i> 7.11-2) “Do new Carthage, her rising walls, and supreme power handed to your scepter not touch you?”</p> <p><i>quid bella Tyro surgentia dicam / germanique minas?</i> (Virg. <i>Aen.</i> 4.43-4) “Why should I speak of the wars rising from Tyre and the threats of your brother?”</p>	“Similar form between <i>crescentia/surgentia</i> , similar meaning as well though in different contexts (rising walls vs. rising wars)” (P6) (“rising/rising”)
Thematic	<p><i>mitius inveni quam te genus omne ferarum;</i> <i>credita non ulli quam tibi peius eram.</i> (Ov. <i>Her.</i> 10.3-4) “I have found every species of beast milder than you; I had been entrusted to no one more wicked than you.”</p> <p><i>quaenam te genuit sola sub rupe leaena,</i> <i>quod mare conceptum spumantibus expuit undis</i> (Cat. 64.154-5) “What lioness gave birth to you under a lone rock, what sea spit you out, conceived by the foaming waves?”</p>	“Stock form of curse” (P15), insulting the target as a beast or wild animal
Stylistic	<p><i>tum denique fleui; / torpuerant molles ante dolore genae.</i> (Ov. <i>Her.</i> 45-6) “Then at last I wept; my soft cheeks had grown stiff from my grief before this.”</p> <p><i>tum Thetidis Peleus incensus fertur amore,</i> <i>tum Thetis humanos non despexit hymenaeos,</i> <i>tum Thetidi pater ipse iugandum Pelea sensit.</i> (Cat. 64.19-21) “Then it is said that Peleus was inflamed with love for Thetis, then Thetis did not look down upon human nuptials, then the Father himself felt that Peleus must be joined to Thetis.”</p>	“Epic use of starting a line [or new sentence] with <i>tum</i> ” (P7)

Table 5: An example of each connection type found during participants’ treatment tasks. The query is listed in the row above the result, and the target words are bolded in the Latin excerpts.

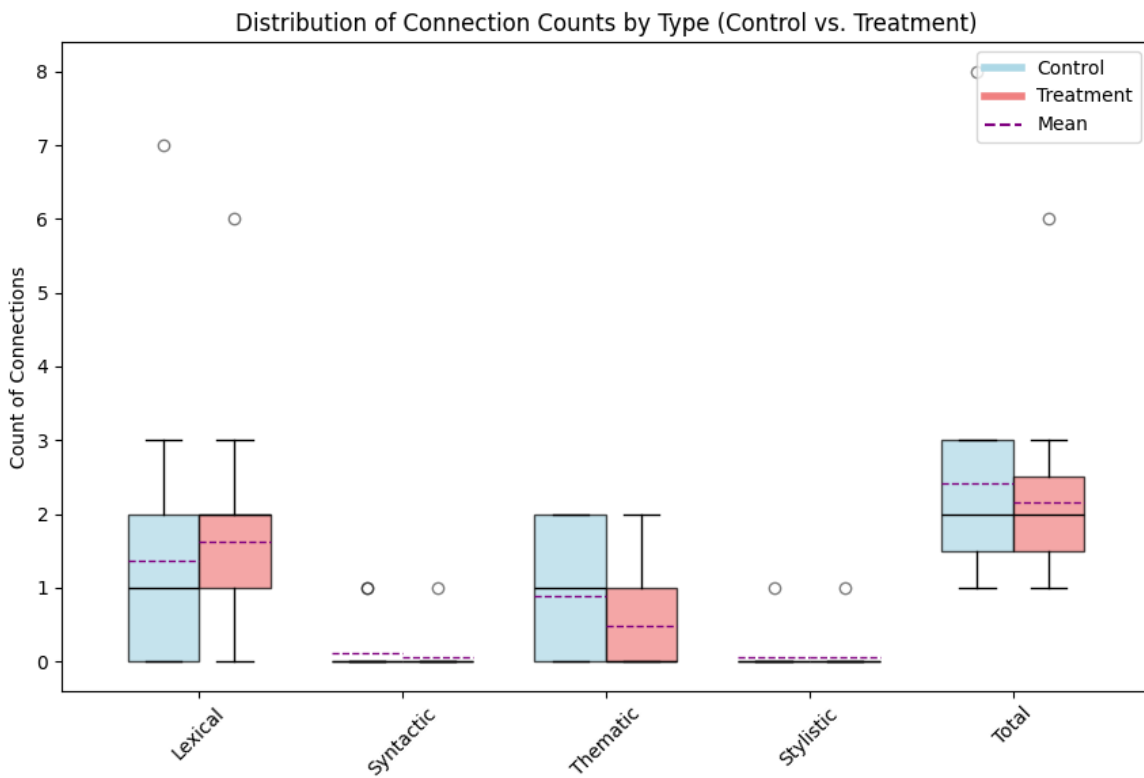


Figure 8: The distribution of lexical, syntactic, thematic, and stylistic connections participants found in each condition. The box plot indicates the median, interquartile ranges, and outliers, while the dashed purple line is the mean count of each type.