

# Tompkinsia: A Data-Driven Commentary Augmentation

Sarah Abowitz  
sarah.abowitz@tufts.edu  
Tufts University  
Medford, Massachusetts, USA

Gregory Crane  
gregory.crane@tufts.edu  
Tufts University  
Medford, Massachusetts, USA

## Abstract

When reading a foreign text from the distant past, commentaries help readers across expertise levels understand obscure linguistic, historical, and/or cultural context. Recent analysis and critique of commentaries have called for clearer, more structurally robust reinforcement of claims made in commentaries. We present Tompkinsia, a hypertext reading environment for an Ancient Greek source text and an associated commentary, where its claims are supported by transparent, accessible, and replicable data-driven analyses leveraging Jupyter notebook integration.

## CCS Concepts

• **Applied computing** → Digital libraries and archives; *Annotation*; • **Information systems** → Digital libraries and archives.

## Keywords

Hypertext reading, Augmenting reading, Digital libraries, Historical texts, Classical commentaries

## 1 Introduction

The commentary tradition has existed for millenia, augmenting reading and translation both on and off the screen. Commentaries are linear lists of annotations on a foreign language source text, and they help readers across expertise levels understand obscure linguistic, historical, and/or cultural context. Commentaries in the Latin and Ancient Greek traditions serve two purposes: first, as a well-informed companion text to help students become acclimated with the language they are learning, and second, as a tool for scholars studying the source text and any topics that text or commentary may resonate with [20]. Recent analysis and critique of commentaries has emphasized three specific problem areas: parallels’ failure to account for “bursty” lexical dispersion, vague and misleading claims, and that commentators often assume a more advanced audience, even when commentators nominally produce a commentary that attempts to serve both novice language learners and veterans equally. Tompkinsia is a hypertext reading environment for an Ancient Greek source text and an associated commentary, where its claims are supported by transparent, accessible, and replicable data-driven analyses leveraging Jupyter notebook integration.

## 2 Background and Motivation

### 2.1 A History of Claims and Commentaries

Since commentaries are linear lists of annotations on a foreign language source text, we argue this construction approaches the definition of a hypertext. Many commentaries try to provide full clarity on the unfamiliar historical, cultural, and linguistic contexts these texts were written in [5]. Scholars wrote commentaries on texts in many languages for millenia [11]. Ancient Greece began

developing its scholarly infrastructure in the 5th century BCE to explain unfamiliar language in the Homeric epics (c. 750 BCE) [7]. Though the Greek tradition is later than others, it continued through Rome and Byzantium, with its presence still felt and its remnants still studied. The first serious editions and commentaries appeared in Alexandria in the 3rd century BCE, and fragments of this work exist on ancient papyri [12]. While a well-developed model for page-based commentaries emerged more than 1,000 years ago, the commentary adopted its modern form by the 18th century. Topics covered by modern print commentaries may have changed, but their format differs little from those produced centuries before [24].

One type of commentary aid, seen in [5], makes a claim about the frequency of a syntactic or semantic tendency’s presence in at least the source text. Often the claim is supported by a set of “parallels,” or references for that phenomenon in the source text or other primary sources. Commentaries and their authors differ in focus and habit, and the quantity and quality of parallel sets have been thoroughly surveyed and questioned by commentary critics [13, 18]. Parallels may be a form of or alongside a similar phenomenon we call “coverage creep.” When writing a commentary on a text with prior commentaries, convention compels the commentator to discuss relevant insights from these commentaries while providing new insights. Opinions change over time, yet what was once scholars in the 1800s referencing one important manuscript has become the duty of modern commentators to cover over 200 years of analysis. Changes in commentaries over the 1900s were also made as part of a larger effort to diversify who reads Greco-Roman texts. As we found in [5], in the 1800s/early 1900s, commentaries were reference dense, but newer commentaries have less references per item. First, one may assume that to not overwhelm newer language learners, student commentaries emerging over the 20th century would lower the number of references per item. Next, these higher reference counts peaked when it was generally assumed that readers were privileged enough to use such a commentary, and could thusly afford all their own copies of the many works referenced. The drop in references could be ascribed to lowering the cost of accessing all works cited to grasp a commentator’s argument. This decrease in cost would be one factor lowering the barrier of entry to engaging more directly with commentaries, with the study of Greco-Roman literature’s audience consciously broadening past the upper class. Other changes emerging during this time include more syntactic aid targeted at newer students, and the diminishing and depersonalization of the commentator’s voice. This is preferable to the more boastful and directly authoritative commentators defending heteropatriarchal opinions over correct syntactic and semantic interpretation before the diversification of the field [38]. Yet there still are known issues in modern commentaries. In [5], we found that while some claims in commentaries cite statistics or secondary sources to support these claims, they most often cite

primary sources. This may seem enough given the tradition of this practice and that commentators often write as humanists first and scientists second. Yet a list of examples does not account for computational linguistic factors like lexical dispersion, and may mislead the reader to believe something occurs more than it does. Lastly, in the late 90s/early 2000s, with the advent of computerized literature and hypertext, critical analysts of commentaries called for varying radical redesigns of the commentary in a new digital medium [15, 29, 38]. Many digital hypertext environments still represent commentaries where their most radical design change is a hyper-linked list of parallels [5, 16]. Though born-digital commentaries present ideas of what commentaries could be, many digitizations of prior commentaries are mostly unaugmented and unrevised. A mainstay in this digital niche could aid further analysis of these claims. Treebanks are XML encodings of sentences where each word has its ID, morphological data, root lemma, and dependence encoded as one edge with a label per word, letting the user construct the sentence as a tree of words [23]. Often these data structures are processed in Python when queried over by digital humanists, and Jupyter notebooks are an effective presentation medium for this work [3, 31]. This paper centers one type of commentary claims. Treebank-verifiable claims are statements made by commentators claiming a linguistic feature occurs in a certain frequency in the text, but by searching across a corpus of ancient Greek works for graph structures found in treebanks that are characterized by any combination of word relation, morphology, and/or lemma, scholars can use these search results to sanity check these claims in a more reliable and transparent way. Treebank-verifiable claims will be analyzed using the GLAUx treebanks [23].

## 2.2 Why Validate Commentary Claims?

As previously mentioned, there are still shortcomings in commentaries that may be clearer to scholars of computational linguistics and accessibility, as well as educators, students, and other unconventional and/or underrepresented types of language learners. First, humanist academics write mostly for their peers. As seen in [5], authors often support their claims with parallels without accounting for lexical dispersion and other factors that may invalidate the claim on further investigation. Next, claims can be vague or misleading. A claim from Charles Smith’s commentary on Thucydides’ *History of the Peloponnesian War* said that the verb  $\lambda\omega\phi\acute{\alpha}\omega$  was “used by Thuc. with reference to sicknesses and grave misfortunes” [37]. Searching for all instances of  $\lambda\omega\phi\acute{\alpha}\omega$  in the *History*, we found 2 instances directly referencing the Athenian plague, and another one that may have been an indirect reference when invoking other misfortunes Athenians overcame. Yet readers may assume that this verb is used more than 3 times in the *History* given the misleading phrasing of this note. In this way, vagueness and misleading claims can pose another barrier to grasping the full insight of a commentary. Lastly, reducing the barrier of entry to commentaries’ insights by showing the mathematical claim analysis is one thing, but these analyses must have accessible explanations. We also found in [5] that when academics try to write a commentary aiming to serve both newer and more advanced language learners, the content is often understood more by the latter. These commentators often

omit the aid newer language learners may need because they assume beginners do not read this. Yet this leads to a commentary excluding more than just the beginners from fully understanding the commentator’s points.

This commentary augmentation will also aid both the hypertext completionists and those uninterested in these notebooks’ detailed analyses. Treebanks have existed for a few decades, yet while many computational linguists have conducted their own analyses over many treebanks with their own bespoke workflows, these analyses are not often front-facing or beginner friendly. This work could lower the barrier of entry into this kind of analysis to those familiar with Python, pandas, and Jupyter notebooks. Plus, augmenting these commentaries with claim analysis may also keep the cursory reader from being misled by claims with notes summarizing their inconsistency or veracity.

Lastly, the scope of this work matters regarding the claims validated by computation. Like the analytical work of Dan Tompkins [40], this design’s namesake, this framework centers data-driven computational linguistic analysis of unambiguous claims (i.e., “It is rare to see a sentence end with a genitive absolute”). A claim like “Odysseus uses the gnomic aorist when speaking with anger” is more ambiguous and thus harder to find with treebanks. It would be easy to find aorists in Odysseus’ dialogue, yet checking each result may be labor intensive, especially since there are no encodings for when an aorist is gnomic or what emotions are there at the atomic level. One could also make a keylist of anger words and use that to further filter the results. Making an accurate keylist with rarer words could take time, but on the other hand, there is always the chance that a valid instance will be excluded. A third method may be training a BERT model to identify passages when Odysseus is angry, and evaluating the results of that. Tompkinsia is a presentation medium for checking claims that are made on a reliable, searchable, and mathematical foundation. Though commentators love agreeing to disagree, this basic design focuses on presenting what can be found and analyzed in an unambiguous way.

## 3 Related Work

Perseus Digital Library [9] digitized about 70 print commentaries, hyperlinking print citations where possible. These links made these commentaries scholarly hypertexts where readers could traverse commentary and source text, clearly seeing connections between both texts [8]. Another digital commentaries platform is the Dickinson College Commentaries [16], known for extending the traditional commentary model, both by including media files, and by linking primary sources and reference works in its content. These commentaries are developed for and with students. Recently New Alexandria [14] built a commentary system where each annotation was peer-reviewed as a separate publication with its own author(s) and metadata. New Alexandria enabled the move beyond the traditional model of single-authored commentaries perpetuating coverage creep [10]. The Ajax Multi-Commentary Project [32] also notably addressed coverage creep [33]. In addition, the Ajax Multi-Commentary Project pioneered work on scalable integration of digitized commentaries and building sustainably minimal computing complex commentary networks [34]. Commentary augmentation is not new, but we have not seen a hypertext reading

environment integrating dedicated Jupyter notebooks for analyzing these commentaries' claims.

Other advancements in this niche of computational linguistics include Tompkins' early computational linguistic analysis of commentary claims [40]. Tompkins' work was near the first to call attention to claims in commentaries worth verifying with computational linguistic means, and he did such verification at a time predating a computer in every home that can run Python and pandas [40]. Like this work, Tompkinsia's analyses aim to transparently walk the reader through claim verification, and comment on the results of such computation. [6] is also a recent user study identifying different patterns of reading behavior while featuring another type of augmented commentary. Students read the Parodos in Sophocles' *Antigone* while accompanied by an augmented version of Richard Jebb's commentary on the text where references were replaced with the relevant excerpt of text referenced. The design of this commentary augmentation is influenced by findings regarding beginner-friendly ways to integrate augmentations without overloading the reader. The original GLAUx treebanks came into existence around 2021 as a result of training multiple classifiers to categorize tokens by Greek morphology, lemmata, part-of-speech, syntactical relation and semantics [23]. It is one of the largest corpora of treebanks spanning many Greek works, so that makes it a suitable candidate for searching over when attempting to verify whether a certain linguistic pattern occurs only in Attic authors. This set of treebanks is also open source, as opposed to fully comprehensive search in the more private TLG [23]. GLAUx also has its own web-based performant treebank query engine, though scholarship on this is under development at current [1, 2]. Web-hosted treebank query engines are nothing new [22, 28], but they are separate websites and this work seeks to integrate analysis more directly into the experience of reading a foreign language source text with a commentary in a hypertext reading environment. Analysis could direct the reader to these other sites, but switching between windows has been cited as one factor of hypertext reading that increases extraneous load [26]. Extraneous load is a form of cognitive load originating from difficulty with navigating a hypertext while reading it [39].

A few works on digital reading influence this work's perception of potential readers. First, [35]'s work on content organizers surfaces the finding that summaries of partitions of hypertext can result in more readers being dissuaded from perusing the same partition of hypertext. One can view a commentary claim similarly: readers may view it as an authoritative statement that is factually accurate, and not in a way they are expected to be able to verify. Providing analysis alongside these claims can challenge the assumption that this work is arcane and outside one's abilities. [21] and [30] identify hypertext reader groups separated by coverage, or what percent of the total hypertext the user traversed. [21] mentions a "disengaged" group and a more completionist group. This is also resonant with types of commentary readers, especially since references and other language may dissuade language learners from engaging deeper with the material itself [20].

This section ends with discussion of work on the history of commentaries and analysis of their design and content. In [5], we devised and applied a non-exclusive taxonomy for types of aid provided by commentaries across many commentary types to reveal both differences in aid distribution and references provided, as well

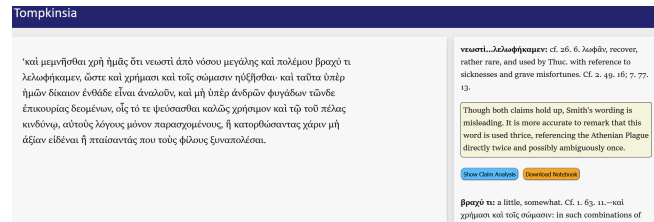


Figure 1: Viewing the source text in Tompkinsia



Figure 2: Notebook view in Tompkinsia

as the cost of and other barriers to accessing the full insight of commentaries. This work demonstrates one way to include code and results supporting claims in commentaries. Multiple critiques over the last 20 years have targeted various aspects of commentary writing, reading, and design [19, 20, 38], yet while these various pieces call for born-digital and digitized commentaries to reckon with these issues, few born-digital commentaries answer this call. Hopefully this work will not be the only response to these pieces.

## 4 Design

In Fig. 1 the Tompkinsia reading environment contains the *History* and Smith's commentary. The source text is in the left pane, and Smith's commentary appears on the right. To view a static Jupyter notebook associated with a claim, clicking a blue button in the claim on the right replaces the contents of the left pane with the static notebook view, as Fig. 2 shows. Here, the notebook represents Markdown, code, and output sections as expected in the Jupyter paradigm, but the code sections are static formatted text and not cells of code that can run. That said, pressing the same blue button reverts the left pane to show the original source text, and the orange button next to it lets the user download the notebook for them to tweak and run on their own. Between the original comment and the buttons is a summary of the analysis, bordered to imply it is an addendum to the original comment. Readers can see the Tompkinsia environment in action at GitHub [4].

## 5 Limitations and Considerations

This design is preliminary, and many objections to this approach are anticipated in this section. First, while hosting Jupyter notebooks dynamically would enable readers to stay in this environment as they explore analyses, the text, and the commentary, hosting notebooks at scale both regarding the quantity of readers and notebooks needed for just one commentary may challenge non-corporate web-hosts. The static design is a compromise ensuring readers can still

enjoy the benefits of these notebooks without sacrificing website availability. Another objection may be that, as one study found [35], summaries of hypertext reduce coverage, or how much of the hypertext readers read, and that a summary of the notebook’s insight above the “Expand Claim Analysis” button may decrease the number of readers that click that button. A recent user study [6] is not the first to find groups of readers differentiated by hypertext coverage [21, 30]. This design provides clear and replicable analysis for anyone curious about the veracity of a claim, but similarly we want to provide those less likely to click the button enough information for them so that they already see where the commentary is misleading.

We also want to recognize that while Tompkinsia was designed for a specific digital reading use case with a primary text and a secondary text, Tompkinsia can be adapted for other purposes. For instance, there are many applications for this paradigm in digital humanist pedagogy, and it is possible that Tompkinsia may be used to introduce undergraduate and pre-college students to quantitative textual analysis outside of Latin and Ancient Greek. The design paradigm of Tompkinsia can also be utilized in lesson plans for how to develop close reading and critical thinking skills while digitally reading. Educators can use Tompkinsia to teach students to apply critical thinking to any auxiliary text assisting with understanding of a main text, digital or otherwise. Inside the commentary tradition, an over-reliance on claims that are under-inspected contributes to the same kind of misunderstanding mentioned in Section 2.2 [20]. Moreover, with the rise of generative AI contributing to a drop in critical thinking skills, there is a greater need for technologies that strengthen our tools for assessing the reliability of information presented to us, rather than merely weakening those same skills [17]. Combining Tompkinsia’s notebook integration with a multi-commentator system such as New Alexandria would also be a game changer, and could be one possible response to a call for a commentary system that allows for transparent dialogue and analysis between those that work on fragmentary texts [38]. While this is keeping with the commentary tradition, we could also see Tompkinsia working with a more layered commentary paradigm, such as the Arabic commentary tradition [25].

Outside of this paper, the authors are also developing a query engine for ease of conducting and providing augmented claim analysis as shown here in this work. One observation from this upcoming research is that the GLAUx treebanks, which are the result of training a classifier on Greek grammar and then using it to generate treebanks for a large corpus of Greek text, occasionally misclassifies (i.e., mislabeling *λώφρησιν* as a form of *λωφάα* or mislabeling *ἀναλοῦν* as a form of *ἀναλέω* when two commentaries claim it is a rare form of *ἀναλίσκω* [27, 36]). The GLAUx team recognizes this, and it is easy to understand how what cannot be recognized by Morpheus would be classified by Lemming, which is reported to have a lower accuracy rate than Morpheus of .1 [23]. That said, a relevant takeaway from the world of commentaries is that when errors like this reveal themselves, it is more beneficial to engage in academic discussion on how best to correct these errors together, rather than leaving those errors unassessed and still affecting those reading the relevant text. One last anticipated remark is that online query engines using data sets such as GLAUx already contain enough data to verify these claims reliably, or that the analyses

presented could have centered GLAUx searches [2]. A query engine like GLAUx may still prove useful in computational linguistics research and pedagogy, but Tompkinsia as a hypertext environment is more beginner-friendly and provides a reading experience where the design mitigates the extraneous cognitive load of switching between windows and running one’s own verification. This is not a choice of one or the other; query engines focused on datasets such as GLAUx and reading environments such as Tompkinsia may coexist as solutions to different problems, but GLAUx could identify instances for assessment of a phenomenon and Tompkinsia could demonstrate different types of analyses one could do with those results. So why is Tompkinsia not using GLAUx initially? GLAUx has no API for requesting results of a query from a notebook cell. A version of our query engine can be imported into a notebook and invoked with a query that returns results in an inline human-readable format that is conducive both to the static notebook view without constituting more extraneous load and to computation that can run offline in the actual notebook file itself, given minimal setup. In the commentary analysis community, there have been demands made of born-digital commentaries to explain claims and other observations like these, to include missing but needed syntactic aid that was once excluded under the incorrect assumption that readers had learned enough of Greek or Latin that printing such basic syntactic aid would prove redundant [38]. Whether or not these authors realize it, commentary critique is a unique subdomain of accessibility research, and including what was once excluded in a transparent and easy-to-follow way lowers the barrier to entry to discussions like these, minimizing the amount of statements readers accept on the grounds that the editor simply knows something that they do not know enough to verify for themselves.

## 6 Conclusion and Future Work

The commentary tradition has existed for millennia, and helps readers across expertise levels understand obscure linguistic, historical, and/or cultural context. Our analysis and critique of commentaries has drawn attention to a few problem areas: parallels’ failure to account for “bursty” lexical dispersion, vague and misleading claims, and that commentaries are often written for a more advanced audience, even when commentators nominally produce a commentary that attempts to serve both novice language learners and veterans alike. Tompkinsia, presented here, is a design for a hypertext reading environment for an Ancient Greek source text and a corresponding augmented commentary, where data-driven analyses in a computational notebook makes commentary claims transparent, accessible, and replicable. Tompkinsia’s next chapter will involve a user study with a more developed form of the software to observe how different hypertext readers do or do not engage with our design. This study will also include surveys focusing on reader trust regarding these claims. Another area of interest for this work would be to augment an existing open-source digitized commentary, though we found in [5] that born-digital commentaries frequently have an audience of newer language learners. That said, we would be open to working with commentators to develop a born-digital commentary that include more advanced language learners while continuing this conversation with reliable and transparent analysis.

## References

- [1] 2024. *GLAUx: How?* Retrieved February 10, 2026 from <https://glaux.be/how.php>
- [2] 2024. *W3.CSS Template*. Retrieved February 10, 2026 from <https://glaux.be/search.php>
- [3] Sarah Abowitz. 2025. *lepidopterane-at-smith/treebank\_morphin*. Retrieved February 10, 2026 from [https://github.com/lepidopterane-at-smith/treebank\\_morphin/blob/main/morph\\_search.ipynb](https://github.com/lepidopterane-at-smith/treebank_morphin/blob/main/morph_search.ipynb)
- [4] Sarah Abowitz. 2026. *lepidopterane-at-smith/tompkinsia: A data-driven commentary augmentation design*. Retrieved February 10, 2026 from <https://github.com/lepidopterane-at-smith/tompkinsia/tree/main>
- [5] Sarah Abowitz, Alison Babeu, and Gregory Crane. 2024. Bridging the Understanding Gap: Helping Readers Engage Directly with Foreign-Language Sources More Easily. In *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3677389.3702539>
- [6] Sarah Abowitz and Gregory Crane. 2025. Student Use of Commentaries with Inline Reference Resolution. In *Proceedings of the 36th ACM Conference on Hypertext and Social Media (HT '25)*. Association for Computing Machinery, New York, NY, USA, 146–155. doi:10.1145/3720553.3746681
- [7] Lorenza Bannardo and Kenneth W. Yu. 2023. Graeco-Roman Commentary beyond Alexandria: Problems and Prospects. In *Practices of Commentary*, Amanda Goodman, Suzanne Conklin Akbari, and Carol Symes (Eds.). Amsterdam University Press, 9–28. doi:10.1515/9781802701593-003
- [8] Gregory Crane. 1996. Building a digital library: The Perseus Project as a case study in the humanities. In *Proceedings of the first ACM international conference on Digital libraries*. 3–10. <https://dl.acm.org/doi/pdf/10.1145/226931.226932>
- [9] Gregory Crane. 2004-. *Perseus Digital Library Project*. Retrieved February 10, 2026 from <http://www.perseus.tufts.edu/hopper/>
- [10] Gregory Crane, Alison Babeu, Lisa M Cerrato, Amelia Parrish, Carolina Penagos, Farooosh Shamsian, James Tauber, and Jake Wegner. 2023. Beyond translation: engaging with foreign languages in a digital library. *International Journal on Digital Libraries* 24, 3 (2023), 163–176.
- [11] R. Criboire. 1996. *Writing, Teachers, and Students in Graeco-Roman Egypt*. Scholars Press. <https://books.google.com/books?id=NQwcAQAALAAJ>
- [12] Eleanor Dickey. 2007. *Ancient Greek scholarship: a guide to finding, reading, and understanding scholia, commentaries, lexica, and grammatical treatises, from their beginnings to the Byzantine period*. Oxford University Press, Oxford.
- [13] Elaine Fantham. 2002. Commenting on Commentaries: A Pragmatic Postscript. In *The Classical Commentary: Histories, Practices, Theory*, Roy Gibson and Christina Kraus (Eds.). Brill, 403–421. doi:10.1163/9789047400943\_017
- [14] Center for Hellenic Studies. 2023. *New Alexandria Open Commentary Platform*. Retrieved August 1, 2024 from <https://oc.newalexandria.info/>
- [15] Don Fowler. 1999. Criticism as commentary and commentary as criticism in the age of electronic media. In *Commentaries/Kommentare*, Glenn Most (Ed.). Vandenhoeck Ruprecht, 426–442.
- [16] Christopher Francese. 2011-. *Dickinson College Commentaries*. Retrieved August 2, 2024 from <https://dcc.dickinson.edu/about-dcc>
- [17] Michael Gerlich. 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* 15, 1 (2025). doi:10.3390/soc15010006
- [18] R.K. Gibson. 2002. Cf. e.g.: a Typology of Parallels and the Role of Commentaries on Latin Poetry. In *The Classical Commentary: Histories, Practices, Theory*, Roy Gibson and Christina Kraus (Eds.). Brill, 331–358.
- [19] Amanda Goodman and Suzanne Conklin Akbari (Eds.). 2023. *Practices of Commentary: Medieval Traditions and Transmissions*. Arc Humanities Press. doi:10.2307/jj.6253300
- [20] Barbara Graziosi. 2009. Commentaries. In *The Oxford Handbook of Hellenic Studies*. Oxford University Press. arXiv:<https://academic.oup.com/book/0/chapter/334629323/chapter-ag-pdf/44445706/book,8587,ection,34629323.ag.pdf> doi:10.1093/oxfordhb/9780199286140.013.0068
- [21] Carolin Hahnel, Dara Ramalingam, Ulf Kroehne, and Frank Goldhammer. 2023. Patterns of reading behaviour in digital hypertext environments. *Journal of Computer Assisted Learning* 39, 3 (2023), 737–750. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12709> doi:10.1111/jcal.12709
- [22] Nick Kallen. 2013. *nkallen/pseudw: Language learning software*. Retrieved February 10, 2026 from <https://github.com/nkallen/pseudw>
- [23] Alek Keersmaekers. 2021. The GLAUx corpus: methodological issues in designing a long-term, diverse, multi-layered corpus of Ancient Greek. In *Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021*, Nina Tahmasebi, Adam Jatowt, Yang Xu, Simon Hengchen, Syrielle Montariol, and Haim Dubossarsky (Eds.). Association for Computational Linguistics, Online, 39–50. doi:10.18653/v1/2021.lchange-1.6
- [24] Christina S Kraus and Christopher A. Stray. 2015. Form and Content. In *Classical Commentaries: Explorations in a Scholarly Genre*, Christina S. Kraus and Christopher Stray (Eds.). Oxford University Press, 1–18. doi:10.1093/acprof:oso/9780199688982.003.0001
- [25] Cornelis van Lit and Dirk Roorda. 2024. Neither Corpus Nor Edition: Building a Pipeline to Make Data Analysis Possible on Medieval Arabic Commentary Traditions. *Journal of Cultural Analytics* 9, 3 (June 2024). doi:10.22148/001c.116372
- [26] Anne Mangen, Bente R. Walgermo, and Kolbjørn Brønnekk. 2013. Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research* 58 (2013), 61–68. doi:10.1016/j.ijer.2012.12.002
- [27] Edgar Cardew Marchant (Ed.). 1897. *Thucydides, Book 6*. (1st ed.). London Macmillan, London, United Kingdom.
- [28] Scott Martens. 2012. TüNDRÄ: A Web Application for Treebank Search and Visualization. In *Proceedings of The Twelfth Workshop on Treebanks and Linguistic Theories (TLT12) (TLT '12)*. The Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria, 133–144. <https://www.academia.edu/download/87484106/TLT12Proceedings.pdf#page=139>
- [29] Willard McCarty. 2002. A Network with a Thousand Entrances: Commentary in an Electronic Age. In *The Classical Commentary: Histories, Practices, Theory*. Brill, 359–402. doi:10.1163/9789047400943\_016
- [30] Tiziano Picardi, Miriam Redi, Giovanni Colavizza, and Robert West. 2020. Quantifying Engagement with Citations on Wikipedia. In *Proceedings of The Web Conference 2020 (Taipei, Taiwan) (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 2365–2376. doi:10.1145/3366423.3380300
- [31] Charles Pletcher. 2025. *pletcher/aristotle\_corpus\_experiments*. Retrieved February 10, 2026 from [https://github.com/pletcher/aristotle\\_corpus\\_experiments/blob/0a9956350654d4e795c264a6580887fa478c5c86/lab.ipynb#L4](https://github.com/pletcher/aristotle_corpus_experiments/blob/0a9956350654d4e795c264a6580887fa478c5c86/lab.ipynb#L4)
- [32] Matteo Romanello. 2020-. *Ajax Multi-Commentary Project*. <https://mromanello.github.io/ajax-multi-commentary/>.lastaccessed=
- [33] Matteo Romanello and Sven Najem-Meyer. 2024. A Named Entity Annotated Corpus of 19th Century Classical Commentaries. *Journal of Open Humanities Data* 10, 1 (Jan. 2024). doi:10.5334/johd.150
- [34] Matteo Romanello, Sven Najem-Meyer, and Bruce Robertson. 2021. Optical Character Recognition of 19th Century Classical Commentaries: the Current State of Affairs. In *The 6th International Workshop on Historical Document Imaging and Processing (HIP '21)*. Association for Computing Machinery, Lausanne, Switzerland, 1–6. doi:10.1145/3476887.3476911
- [35] M. Sanchiz, F. Amadieu, J. Lemarié, and A. Tricot. 2022. Do graphic and textual interactive content organizers have the same impact on hypertext processing and learning outcome? *Journal of Computing in Higher Education* 35 (June 2022). doi:10.1007/s12528-022-09328-z
- [36] Cynthia W. Shelmerdine (Ed.). 1989. *Thucydides Book VI: commentary* (1st ed.). Bryn Mawr Commentaries, Bryn Mawr, PA.
- [37] Charles Forster Smith (Ed.). 1913. *Thucydides. Book VI* (1st ed.). Ginn and Company Boston, Boston, MA.
- [38] Susan Stephens. 2002. Commenting on Fragments. In *The Classical Commentary: Histories, Practices, Theory*, Roy K. Gibson and Christina Shuttleworth Kraus (Eds.). Brill, 67–88. doi:10.1163/9789047400943\_005
- [39] Omar Taky-eddine and Redouane Madaoui. 2024. Cognitive Overload in the Hypertext Reading Environment. *International Journal of English Language Studies* 6, 2 (May 2024), 94–100. doi:10.32996/ijels.2024.6.2.13
- [40] Daniel P. Tompkins and Adam Parry. 1972. *Stylistic characterization in Thucydides: Nicias and Alcibiades*. Cambridge University Press, 181–214.

Received 12 February 2026