

Plebeius Sermo: A Corpus Study of Cicero's Letters to Atticus

Cic. *Fam.* 9.21.1: *verum tamen quid tibi ergo videor in epistulis? Nonne plebeio sermone agere tecum?*

“But what do I seem like in my letters? Surely I use everyday language with you?”

From Tyrell & Purser to Halla-aho, scholars have taken Cicero's letters to his closest friends to be genuine examples of informal Latin, unintended for publication and unmarred by an editor's pen (Halla-Aho, 2011; Shackleton Bailey, 1999). The letters to Atticus, given Cicero's close relationship to him and the features of their style, are considered especially good evidence of informal Latin (Hofmann, 1951). However, the possibility the letters were edited and the influence of Cicero's rhetorical training adds uncertainty. The validity of the claim that the letters as a collection are genuine informal communication is strong but does not come with concrete proof. I propose a re-analysis of the letters using text datasets that have recently become available. I argue that the Letters' genuine informal nature can not only be verified but also further explored using corpus methods. It's enticing to believe we have a rare window into the everyday spoken language of the Romans. The prospect is even more enticing given the current pedagogical climate: living Latin has recently become much more popular in schools. I have seen 6th grade students speaking Latin, responding to commands, and writing too. An understanding of beginner Latin outside Caesar and Livy is needed for the modern classroom.

A text as well pored over as the Letters to Atticus can provide more insights through computational analysis. Corpus linguistics has become a valuable tool in the past several decades for teasing out the properties of different registers such as speech and writing, letters and broadcast, which Douglas Biber's research advanced. The premise of corpus linguistics is that a

computer-aided study of tens of thousands, hundreds of thousands, or millions of words results in rigorous and significant results. Large language models like ChatGPT are evidence this is an effective approach, as taking in massive amounts of data reduces the effect of outliers and edge cases. Imagine if ChatGPT was only trained on *shakespeare.mit.edu*: it would hardly be a successful chatbot using *doths* and *thous* and *yonders*. As it is, the chatbot is trained on internet English at large, so the effect of outliers like 16th c. English is minimized in the face of a huge representative sample. The result is an effective and uncanny understanding of English usage.

A multivariate analysis of the letters to Atticus lets us put a comparison with other formal and informal texts on firmer ground. The results of a multivariate analysis are twofold. First, it shows whether the similarities between the Letters to Atticus and other colloquial texts are significant enough to be unlikely to be due to chance. Second, the analysis reveals the most relevant features separating informal from formal prose, building our knowledge of informal Latin grammar. My work for my master's thesis has already suggested the former is true, reconfirming the anecdotal evidence; the goal is to present this prior work and expand on the latter question. Recent datasets make a quantitative study of the syntax and grammar of different registers possible. The PROIEL treebanks contain annotated texts of the letters and have been a valuable resource for performing author attribution (Vainio et al., 2019), and studying word order (Gulordava and Merlo, 2015).

The new statistical evidence adds rigor to the assertion that the Letters belong to the informal category. A gold standard informal text with rigorous evidence backing it up paves the way for future study of informal Latin. There is uncertainty about what aspects of syntax in comedy, for example, are due to their informal nature, the meter, or archaism. A statistically backed baseline for informal language provides a tentative path forward for isolating the

colloquial features of those texts. A quantitative study of informal Latin based on these well-laid foundations has further implications for authorship attribution, commenting on texts, and creating resources for students learning spoken Latin in the classroom.

Bibliography

- Gulordava, Kristina, and Paola Merlo. “Diachronic Trends in Word Order Freedom and Dependency Length in Dependency-Annotated Corpora of Latin and Ancient Greek.” *Proceedings of the Third International Conference on Dependency Linguistics*, 2015, pp. 121–30.
- Halla-Aho, Hilla. “Epistolary Latin.” *A Companion to the Latin Language*, edited by James Clackson, Wiley-Blackwell, 2011, pp. 427–44. Blackwell Companions to the Ancient World.
- Hofmann, Johann Baptist. *Lateinische Umgangssprache*. 3rd ed., C. Winter, 1951.
- Shackleton Bailey, D. R., translator. *Letters to Atticus, Volume I*. Vol. 1, Harvard University Press, 1999.
- Vainio, Raija, et al. “Reconsidering Authorship in the Ciceronian Corpus through Computational Authorship Attribution.” *Ciceroniana on Line*, vol. 3, no. 1, 2019, pp. 15–48.