

# *Auceps syllabarum*: A Digital Analysis of Latin Prose Rhythm\*

TOM KEELINE AND TYLER KIRBY

## ABSTRACT

*In this article we describe a series of computer algorithms that generate prose rhythm data for any digitised corpus of Latin texts. Using these algorithms, we present prose rhythm data for most major extant Latin prose authors from Cato the Elder through the second century A.D. Next we offer a new approach to determining the statistical significance of such data. We show that, while only some Latin authors adhere to the Ciceronian rhythmic canon, every Latin author is ‘rhythmical’ — they just choose different rhythms. Then we give answers to some particular questions based on our data and statistical approach, focusing on Cicero, Sallust, Tacitus and Pliny the Younger. In addition to providing comprehensive new data on Latin prose rhythm, presenting new results based on that data and confirming certain long-standing beliefs, we hope to make a contribution to a discussion of digital and statistical methodology in the study of Latin prose rhythm and in Classics more generally. The Supplementary Material available online (<https://doi.org/10.1017/S0075435819000881>) contains an appendix with tables, data and code. This appendix constitutes a static ‘version of record’ for the data presented in this article, but we expect to continue to update our code and data; updates can be found in the repository of the Classical Language Toolkit (<https://github.com/cltk/cltk>).*

**Keywords:** Latin prose rhythm; clausulae; Latin prose style; digital analysis; statistical analysis; Cicero

To sum up, we may accept that Zielinski’s statistics, while they are far from perfect, do nevertheless give a tolerably accurate picture of Cicero’s clausulae ... It is conceivable that in the future computer technology may allow accurate statistics to be produced for large amounts of material, such as whole authors, at the touch of a button. But until that day arrives, Zielinski’s figures for Cicero’s speeches ... may suffice. They are the best we have and, until computers come to our aid, will not be improved upon.<sup>1</sup>

For over a century, scholars studying Latin prose rhythm have relied on the statistics generated by Theodor Zielinski’s pioneering *Das Clauselgesetz in Ciceros Reden*.<sup>2</sup> They have also complained about his methodology and its inadequacies.<sup>3</sup> Zielinski read his

\* We thank Kathleen Coleman, Myles Lavan, Tim Moore, Christopher Whitton and the *Journal*’s anonymous readers: all have substantially improved this article. We also thank Christopher Kelly for his friendly and efficient editorial work.

<sup>1</sup> Berry 1996a: 50.

<sup>2</sup> Although born Tadeusz Stefan Zieliński, when publishing in German he went by Theodor Zielinski. For a full biography, see Srebrny 2013, with prose rhythm discussed at 149–51. An accessible introduction to Zielinski’s methods and their results is provided by Clark 1905.

<sup>3</sup> For a comprehensive critique of Zielinski, see Oberhelman 2003: 90–106.

own Russian translations of Cicero's speeches out loud in order to develop a feel for where sense breaks (and so clausulae) occurred in the Latin;<sup>4</sup> he arbitrarily decided that the cretic was the basis for Latin prose rhythm;<sup>5</sup> he did not compare his observed frequencies of clausular patterns to any expected values, thus ignoring the naturally occurring rhythms of the Latin language;<sup>6</sup> he came up with dubious rules for word division and resolutions within his clausular categories.<sup>7</sup> But this was path-breaking scholarship, for Zielinski had no real predecessors — and he has had no successors either.<sup>8</sup> In the 115 years since *Das Clauselgesetz*, no scholar has had the *Sitzfleisch* to do what Zielinski did: he counted, by hand, 17,902 clausulae in Cicero's speeches. He analysed his results in detail and produced elegant summary tables, all without the aid of electronic calculators. The result is an imposing and apparently authoritative monument.

The real problem with Zielinski's analysis, however, is not its methodological basis. About his methodology Zielinski is an exemplar of openness and honesty: he lays out his assumptions and reasoning at every step of the process, and in exhaustive detail. While all of these have been questioned, no one would expect the first explorer of uncharted terrain to map it perfectly. A bigger problem is that Zielinski's results are unverifiable and unreproducible. He seems to provide a deluge of data, but readers must trust that he has scanned and counted and tabulated correctly, for he provides comprehensive scansion for only one speech. Zielinski was a great scholar, but in most fields of scientific inquiry we do not simply accept unverifiable pronouncements. And yet it is not just Zielinski: all scholars of Latin prose rhythm who give even partial statistics have presented their varying results and varying methodologies from a black box that could not be inspected or verified.<sup>9</sup> Furthermore, there looms a potentially even bigger problem: if one wanted to modify Zielinski's methodology — disregarding some of his strictures on word division and resolution, say — it would require recounting *everything* from scratch.

Fortunately, computers have come to our aid.<sup>10</sup> In this article, we describe a series of interrelated algorithms and modules that can produce a comprehensive analysis of the prose rhythms of a given corpus of Latin literature with a few keystrokes. This digital approach presents entirely new possibilities for the study of prose rhythm. With complete openness and transparency, we can calculate prose rhythm statistics from across the whole of extant Latin literature. Furthermore, we can be absolutely consistent in our procedures and confident in our statistics, and yet we are not bound to any one

<sup>4</sup> Zielinski 1904: 7. See Laurand 1936–38: 2.199–200; Berry 1996a: 49–50 with details on what Zielinski did — and did not — count using this method.

<sup>5</sup> He also arbitrarily allowed molossi, choriamb and epitrites to be substituted for a cretic. See Shewring 1930: 165; cf. Aili 1979: 67–8. Cicero himself points to the double trochee as the fundamental unit (*Orat.* 2.12–15); see further Winterbottom 2011.

<sup>6</sup> See, for example, De Groot 1921: 18–20; Shewring 1930: 165; Berry 1996a: 48.

<sup>7</sup> See, for example, Shewring 1930: 165; Berry 1996b: 52 n. 253.

<sup>8</sup> So Winterbottom 2011: 265 n. 17 on Zielinski's results: 'still the only source for complete figures on Cicero's speeches'. Zielinski was not in fact the first to study prose rhythm (see Novotný 1929: 2–16), but his results were so novel and comprehensive that, for all intents and purposes, they sprang fully formed from his head and revolutionised the field.

<sup>9</sup> So, for example, Bornecque 1907; De Groot 1921; 1926; Broadhead 1922; Primmer 1968; Aili 1979; Aumont 1996.

<sup>10</sup> An earlier digital tool developed especially for the analysis of *cursus* rhythms is described in Spinacce 2014 (and is online at: <http://cursusinclausula.uniud.it/public/>). Because this tool requires users to intervene manually in cases of potentially ambiguous prosody, and because so many clausulae contain instances of such prosody (for example, *puellā* vs *puellā*), this tool is of only the most limited application. (In the *exordium* of *Pro Milone*, which consists of about 575 words, twelve user interventions were required, and several other errors were generated. Each user intervention must be hand encoded into the text being scanned.) Furthermore, it is hard to get the programme to work, and the results it produces are not presented in a useful format. A truly remarkable early pioneer in applying computers to the study of Greek prose rhythm was McCabe 1981; the understated description of what he managed to accomplish with the technology of forty years ago (at 82–118) is awe-inspiring.

methodology. If it becomes clear, for example, that we should treat elision differently, we can do so and generate new numbers and new statistics — instantly. Zielinski laboriously counted 17,902 clausulae by hand over years: we count hundreds of thousands of clausulae in seconds. Furthermore, all of our results are verifiable from the highest to the lowest level: we can show how any individual phrase has been scanned and categorised, and all of our code and data are open source. We can thus answer fundamental and challenging questions about prose rhythm, and answer them with speed, consistency and transparency.

## I METHODOLOGY

Latin prose rhythm sometimes looks like a species of philological witchcraft, albeit one without the seductive power of most black magic. In part this is because the ancient testimonia on the subject are confused or confusing, and ancient theory does not always seem to match ancient practice.<sup>11</sup> But it is clear that ancient orators and rhetoricians perceived prose rhythm as a real phenomenon, and they cannot be faulted for failing to reduce a complex and intuitively felt system to a set of clear rules. Indeed, it was not just prose rhythm that caused headaches for ancient linguistic theorists: everything from the Latin stress accent to its ablative case proved obstacles for ancient authorities trying to systematise the properties of their language.<sup>12</sup> Self-diagnosis is hard.

With the distance of two millennia and a bevy of statistics, we may actually stand a better chance today of describing the practice of ancient prose rhythm. Modern theories continue to proliferate, and we do not propose to adjudicate among them here.<sup>13</sup> After much trial and error, we have settled on a system that both seems generally reasonable and accounts for the data. Our typology accords fairly well with modern scholarly approaches, but it is fundamentally a pragmatic choice, adopted because it yields useful and interesting results.<sup>14</sup> It is not meant to be the last word. Again, a virtue of the digital approach is that we can adjust — and have adjusted — our methods and classification as our understanding improves. Looking at the data that we provide, new readers may detect other points of interest which have eluded us.

We divide all possible clausular patterns into seven main categories. Of these seven categories, the first four — cretic-trochaic, double cretic/molossus cretic, ditrochaic and hypodochmiac — are traditionally considered ‘rhythmic’. These are the rhythmic preferences that seem to have been developed for Greek prose by the shadowy Hegesias (third century B.C.) and are exemplified in Latin by Cicero.<sup>15</sup> When scholars talk about

<sup>11</sup> Most notably Cicero’s: see, for example, *De or.* 3 and *Orat.* 168–238. Further testimonia are collected in Bornecque 1907: 5–166; Clark 1909; recent discussion in Oberhelman 2003: 27–67. Zielinski 1904: 4, among many others, reasonably concluded that Cicero did not know his own practice. A few scholars have tried to show that Cicero’s theory and practice do align, most notably Laurand 1936–38: 2.159 and *passim*; Schmid 1959 (followed by, for example, Koster 2011), but this requires exceptional creativity. Quintilian likewise seems hopeless; in the *Institutio oratoria*, ‘there is hardly a single type of ending to a Latin sentence that is not recommended’ (Winterbottom 2011: 263). On these issues, see sensibly Hutchinson 2018: 5–10, 16–19.

<sup>12</sup> Latin accent: Allen 1978: 83–4; ablative case: Taylor 1991.

<sup>13</sup> For a summary of research from the Renaissance through the early twentieth century, see Novotný 1929: 2–33. Useful sketches of work from the nineteenth century onwards include Wilkinson 1963: 237–40; Aili 1979: 8–15; Aumont 1996: 11–58. A comprehensive survey and evaluation of all major modern studies of prose rhythm is provided by Oberhelman 2003: 69–184.

<sup>14</sup> The system that we describe here is very similar to (but not exactly identical with) the schemata of Nisbet 1990; Hutchinson 2018: 11–12.

<sup>15</sup> For Hegesias and his system, see concisely Hutchinson 2018: 5–10, 16–19; full testimonia in *FGrH* 142, *RE* 7.2, cols 2607–8. Hutchinson 2013: 233–5 argues that Cicero introduced these rhythms to Latin; this is at least a plausible suggestion, but given the fragmentary evidence for Latin prose before Cicero, certainty is impossible. The seemingly Ciceronian rhythmic propensities of the non-Ciceronian *Rhet. Her.* may raise some doubts (on which see Hutchinson 2013: 235).

'rhythmic' authors, they usually mean those who follow this system.<sup>16</sup> Hegesias' doctrines were very influential and found a number of adherents; as we will see, looking at authors' differing preferences for so-called 'rhythmic' and 'non-rhythmic' clausulae has real explanatory power. But we will also see that all Latin authors have their own rhythmic preferences, even those who do not follow this artificial system. Thus, in a slight but significant terminological shift, we will avoid calling authors 'rhythmic' and 'non-rhythmic', even as we still find it useful to compare the artificially 'artistic' rhythms (the first four categories below) with 'non-artistic' rhythms (the last three).

1. Cretic-Trochaic:  $\text{--}\text{--}\ \text{--}\text{x}$   
Resolved:
  - a.  $\text{v}\text{v}\text{v}\text{--}\ \text{--}\text{x}$
  - b.  $\text{--}\text{v}\text{v}\text{v}\text{--}\text{x}$
  - c.  $\text{--}\text{v}\text{v}\text{v}\text{x}$
2. Double cretic/molossus cretic:  $\text{--}\text{--}\ \text{--}\text{x}$  or  $\text{--}\ \text{--}\ \text{--}\ \text{--}\text{x}$   
Resolved:
  - a.  $\text{v}\text{v}\text{v}\text{--}\ \text{--}\text{x}$
  - b.  $\text{--}\text{v}\text{v}\text{v}\text{--}\text{x}$
  - c.  $\text{--}\text{v}\text{v}\text{v}\text{x}$
  - d.  $\text{v}\text{v}\text{--}\ \text{--}\ \text{--}\text{x}$
  - e.  $\text{--}\text{v}\text{v}\text{--}\ \text{--}\text{x}$
  - f.  $\text{--}\ \text{--}\text{v}\text{--}\text{x}$
  - g.  $\text{--}\ \text{--}\ \text{--}\text{v}\text{v}\text{x}$
  - h.  $\text{--}\text{v}\text{--}\ \text{--}\ \text{--}\text{x}$ <sup>17</sup>
3. Double trochee:  $\text{--}\text{--}\text{x}$   
Resolved:
  - a.  $\text{v}\text{v}\text{--}\text{x}$ <sup>18</sup>
  - b.  $\text{--}\text{v}\text{v}\text{x}$ <sup>19</sup>
4. Hypodochmiac:  $\text{--}\text{v}\text{v}\text{--}\text{x}$ <sup>20</sup>  
Resolved:
  - a.  $\text{v}\text{v}\text{v}\text{--}\text{x}$
  - b.  $\text{--}\text{v}\text{v}\text{v}\text{x}$
5. Spondaic:  $\text{--}\ \text{--}\ \text{--}\text{x}$  (no resolutions)
6. Heroic:  $\text{--}\text{v}\text{v}\text{--}\text{x}$  (no resolutions)
7. Miscellaneous (everything else)<sup>21</sup>

In the first four categories, we allow one resolution of a long into two shorts. Despite some temptation, we have nowhere permitted two or more resolutions in a single clausula. Once you allow more than one resolution, clausulae quickly begin to lose their individual character: should  $\text{v}\text{v}\text{v}\text{v}\text{--}\text{x}$  count as a twice-resolved cretic-trochaic or a once-resolved double trochee? There are ways around this problem, but complications immediately

<sup>16</sup> So explicitly Hutchinson 2015: 789.

<sup>17</sup> Epitrite substitution: Zielinski 1904: 85–92; Berry 1996b: 51.

<sup>18</sup> The pattern  $\text{--}\text{v}\text{v}\text{--}\text{x}$  is counted as a resolved cretic trochee, not a resolved double trochee.

<sup>19</sup> The patterns  $\text{--}\text{v}\text{v}\text{v}\text{--}\text{x}$  and  $\text{--}\ \text{--}\ \text{--}\ \text{v}\text{v}\text{x}$  are counted as resolved double cretics/molossus cretics, not resolved double trochees.

<sup>20</sup> Hypodochmiac clausulae are rare, occurring less frequently than double spondees even in authors with a predilection for 'artistic' rhythms. But Hutchinson 1995: 485–6, looking at the alternation of *atque/ac* before consonants in Cicero, is a simple and persuasive piece of evidence in favour of treating them as artistic. If they were treated as non-artistic, however, very little would change in the following discussion.

<sup>21</sup> We do track certain forms of 'everything else' individually, for example, first paeons ( $\text{--}\text{v}\text{--}\text{x}$ ) that do not constitute parts of a once-resolved cretic trochee ( $\text{--}\text{v}\text{v}\text{--}\text{x}$ ), or choriamb-trochees ( $\text{--}\text{v}\text{--}\ \text{--}\text{x}$ ), but their numbers are generally so small that it makes most sense to lump them all together in a miscellaneous group.

multiply, and we doubt whether something like ~~~~~~x could ever be felt as anything other than a very long series of shorts.<sup>22</sup>

We have used the Packard Humanities Institute (PHI) Latin texts as our corpus of data.<sup>23</sup> These texts are of high quality and freely available, although they require extensive preprocessing for machine analysis. First they must be reformatted to Unicode and extra spaces and line breaks must be removed, along with section numbers and book divisions and so forth. Then their orthography must be made uniform: we have converted consonantal *i* and *u* to *j* and *v* throughout, and systematically incorporated certain unusual features of Latin prosody (for example, *huius* → *hujjus*). Then the texts must be ‘macronised’: vowels that are long by nature must be so marked. This is a non-trivial process for which we have used the excellent tool of Johann Winge, which shows a remarkably high degree of accuracy for classical Latin texts (95–98 per cent).<sup>24</sup> This done, the texts must be syllabified, i.e., separated out into their constituent syllables; here again we have made use of an open-source tool, this time from the Classical Language Toolkit (CLTK).<sup>25</sup> Finally, problematic elements must be removed from our sample and tracked separately: we exclude clausulae that contain abbreviations (most notably proper names), Roman numerals, textual corruptions marked by editors (daggers, brackets and the like) or fewer than four syllables.

After preprocessing, by default we collect up to thirteen syllables worth of clausular data before every mark of ‘heavy’ punctuation, viz. full-stops, semicolons, colons, question marks and exclamation marks (. ; : ? !). This is not a perfect method, since clausulae can and do occur where editors tend to punctuate with commas, as well as in places where there is no punctuation at all.<sup>26</sup> Furthermore, many previous scholars have only looked at clausulae before periods, question marks and exclamation marks.<sup>27</sup> Including semicolons and colons by default seems best to us, but within our framework users can decide for themselves and set which punctuation they would like to consider, and so results with different punctuation patterns can easily be generated.<sup>28</sup>

Then these data must be scanned, sorted and counted. On the one hand, it is easy to write a programme to scan macronised Latin texts. The basic rules are straightforward: if a syllable is closed or ends with a long vowel, it is long. If a syllable is open and ends with a short vowel, it is short. But there are a variety of subtleties that must be accounted for, including elision, instances of mute+liquid and cases of short open syllables before *s impura* (*sc, sm, sp, sq, st, z*; so *ipse sceleratus*); in Cato, at least, one

<sup>22</sup> In cases of ambiguity, Zielinski 1904 took the most rigorous line, attempting to determine the appropriate category for a multiply resolved clausula by considerations of word division, accent and supposed ictus. Even if he managed to be consistent in his choices (unverifiable), the problems with this approach are so considerable as to render it of little practical value.

<sup>23</sup> <http://latin.packhum.org/>.

<sup>24</sup> <https://github.com/Alatius/latin-macronizer>; see Winge 2015. The challenges involved are considerable: *puellā* (nom.) vs *puellā* (abl.), *incīdo* (‘I cut into’) vs *incīdo* (‘I fall upon’), *omnīs* (nom. and gen. sg.) vs *omnīs* (acc. pl.), etc. These problems present a major obstacle for automating scansion, and Winge has done groundbreaking work. His approach uses the RFTagger (<http://www.cis.uni-muenchen.de/~schmid/tools/RFTagger/>) for part-of-speech tagging trained on the Perseus Latin Dependency Treebank ([https://github.com/PerseusDL/treebank\\_data](https://github.com/PerseusDL/treebank_data)) and PROIEL (<https://github.com/proiel>). Larger data-sets of training data and other machine-learning approaches are possible and may increase accuracy still further.

<sup>25</sup> Johnson *et. al.* 2014–.

<sup>26</sup> See, for example, Nisbet 1990 and the earlier investigations of Fraenkel 1968 (building on his own previous work); Habinek 1985. Restricting ourselves to clausulae before punctuation here ensures consistency in our results. Note too that the placement of commas differs widely in different critical texts: compare, for example, the practice of German and English editors.

<sup>27</sup> So, for example, Aili 1979, among others; cf. Berry 1996a: 64, who does include colons and semicolons.

<sup>28</sup> In our testing, as you might expect, considering clausulae only before full-stops, exclamation marks and question marks increases the proportion of artistic rhythms, whereas including commas decreases it.

might even countenance ‘sigmatic ecthlypsis’, or loss of final *s*.<sup>29</sup> By default we do elide, do lengthen a short final open syllable followed by an *s impura*, but do not lengthen a short vowel followed by a mute + liquid. We think that this is the most accurate representation of classical Latin pronunciation.<sup>30</sup> But we also allow users to set these parameters for themselves, and we try to track more fine-grained data as well: so we record whether an elision is of a long vowel/diphthong or of *-m* or of a short vowel and allow users to choose to elide or not elide in any of these categories.<sup>31</sup> We furthermore track word division/word shape and word accent, which may be relevant if we wish to consider iambic shortening or rules for resolutions that depend on word division or hypothetical ‘prose ictus’.<sup>32</sup> In sum, we have built in flexibility to allow users to set their own preferred parameters and slice the data differently.

With these tools, we can generate all manner of reports in seconds.<sup>33</sup> After preprocessing, we can show the complete syllabification, scansion and accentuation of any Latin text; we can show those results divided into clausulae; we can produce data on numbers and percentages of individual clausulae within a text; and of course we can combine all this information to yield comprehensive data on the prose rhythms of any set corpus of Latin literature, as we do below. Such reports allow us to ask and answer with ease questions that would have taken weeks and months and years of tedious (and error-prone) calculation before.

### *Some Limitations*

Our method certainly is not perfect. For example, we currently assume that Latin prosody showed no variation or evolution over time. This is manifestly untrue, most obviously perhaps in the case of final *-o*. We know from verse evidence that in the first century B.C., final *-o* in most words was regularly long (for example, *ergō*, so always in Vergil). But by the time of Lucan, and still more so by that of Martial, final *-o* was usually short. We treat such cases as invariably long. In our current model, we likewise ignore effects like iambic shortening, which presumably was in operation in all ages on at least some words at least some of the time.

<sup>29</sup> On this feature of archaic prosody see Allen 1978: 36–7, and the references collected in Butterfield 2008: 188 n. 4. We have disregarded the possibility of weakening or loss of final *s*.

<sup>30</sup> See Allen 1978: 78–82, 89–90; on *s impura*, Cser 2012 with somewhat different conclusions. Different scholars have treated these cases differently. Aili 1979: 48–9 excludes all such potentially ambiguous prosody from his corpus (although he is content to include instances of aphaeresis like *factumst*), attempting to limit his investigation to cases of certainty. The number of clausulae that he is forced to exclude, however, is enormous, amounting to nearly half of the total in Cicero. Most other scholars have tended to treat these ambiguities on a case-by-case basis, deciding in cases of uncertainty based on an idea of which potential clausula would be ‘better’; the criteria tend to the subjective: so Zielinski 1904: 174–5 had proclaimed that syllables before *s impura* are lengthened ‘without exception’, but Nisbet replied that ‘when I read “ipse sceleratus” before a pause (*Pis.* 28), I hear *esse videatur*’ (Nisbet 1990: 359). Ancient precepts on these questions are often frustratingly vague, for example, Quint., *Inst.* 9.4.36: ‘nonnumquam hiulca etiam decent faciuntque ampliora quaedam.’

<sup>31</sup> Our tests also show that eliding maximises artistically rhythmic patterns: if you exclude all elisions, you vastly increase the number of clausulae in our non-artistic ‘miscellaneous’ category at the expense of artistic clausulae. Similarly, if you allow mute + liquid to lengthen the preceding syllable, you simply increase the number of long syllables, thus favouring more spondaic cadences.

<sup>32</sup> As did Zielinski 1904. Prose ictus, a supposed accent on the first long syllable of each metrical foot in the clausula, continues to play a part in numerous studies of Latin prose rhythm (for example, Aumont 1996: 211–17), despite the lack of any ancient evidence for such a thing. It seems very likely that any apparent tendencies toward coincidence or clash of ‘ictus’ and word accent in prose are epiphenomenal; see especially Oberhelman 2003: 106–10. A similar case has been made (less persuasively) about ictus in Latin poetry: Stroh 1990; Zeleny 2008; Fortson 2011.

<sup>33</sup> It is important to note that each of these tools is modular and can be reused for other purposes; furthermore, it is easy to make changes to one module without affecting the rest of the system and then rerun tests and reports.

With further modifications we could allow users to consider different treatments of prosody, but two risks immediately present themselves: first, if final *-o* is treated differently in, for example, Cicero and Pliny, it may no longer be legitimate to compare results between the two. Second, and perhaps more seriously, it is hard to know in any individual case whether *-o* is pronounced short or long. In Silius Italicus we find only *ergō* — except at 16.217 ‘cui nescire licet? quin ergō tristia tandem’. For Silius, metre guarantees prosody in each instance, including when it differs from our expectations. But what do we do with Pliny the Younger? At *Ep.* 6.19.5 ‘concursant ergo candidati’, *ergō* gives a ‘better’ clausula (molossus ditrochee), and so perhaps *-ō* should be preferred there, but there is no metrical guarantee. For now consistent practice throughout seems methodologically safest.<sup>34</sup>

There is also the fact that our output will necessarily be determined by our input. The PHI texts are meticulous reproductions of standard print editions, but they do not include a critical apparatus, and so we cannot take account of variant readings. More importantly, for the past century editors have made decisions among variant readings and competing emendations at least in part based on their understanding of prose rhythm. Indeed, they have also considered prose rhythm in how they punctuate their texts. Thus, to some degree, prose rhythm has already been ‘baked in’ to these texts, and our results could be circular. While this is admittedly true at a local level — that is, in the case of any given sentence — over a large corpus, the vast majority of clausulae will be free from textual troubles, and most editorial decisions concerning choice of reading and choice of punctuation will not have hinged on prose rhythm. This objection is thus more potent in theory than practice.

Finally, the various component parts of our programme occasionally err. Although the macroniser returns correct results 95–98 per cent of the time, the rest of the time it does not.<sup>35</sup> Even more rarely, sometimes the *u/v* and *ij* converter makes a mistake, as does the syllabifier.<sup>36</sup> While we have tried to make our algorithms as accurate as we can, some error inevitably remains, and we have not adjusted any of our results by hand. We plead the following:

1. The error is very small by comparison with the enormous amounts of data that we can consider. Our sample size is large enough that we can rely on the central limit theorem to justify our statistical analysis. Put plainly: Big Data eliminates small error as a practical issue.
2. The error will be the same in all of our tests. That is to say, we expect that the same types and proportions of error will be present in a text of Cicero or Caesar or Apuleius. Since we use a uniformly consistent methodology, we will always be comparing like with like.
3. It seems very likely that those who count by hand make mistakes too, although because their results are not easily reproducible, it is very hard to determine what kinds of mistakes they have made and how often they have made them.<sup>37</sup>

<sup>34</sup> You can imagine an algorithm that decides prosody (and elision and so forth) in ambiguous cases so as always to produce the ‘more preferred’ clausula. Such a system, however, almost immediately becomes circular and self-reinforcing. This is, writ very large, the methodological difficulty that more subjective scholars face in their categorisation of doubtful clausulae.

<sup>35</sup> In fact the macroniser tends toward the upper-end of the accuracy range on syllables in clausular position because of the types of words that tend to be found there.

<sup>36</sup> Determining algorithmically when *i* and *u* are consonantal is surprisingly hard: consider, for example, *ui*, *iu*, *ii*, *II* (Roman numeral), *ius*, *Seruius*, *uua*, *fluuius*, *mortuus*, *quid* (this last betwixt and between, a digraph). These problems are bound up with syllabification, which also comes with its own challenges: disyllabic *lin-gua* vs trisyllabic *ar-gu-o*, or *sua-de-o* vs *su-a* and *su-ap-te*. And sometimes Latin orthography simply does not represent pronunciation: *abicio* = *abjicio*, *cuius* = *cujus*, etc. This is to say nothing of truly edge cases, where algorithmic perfection is almost impossible: for example, *thy-i-o* (Prop. 3.7.49, if right), *Thyi-as* (Verg., *Aen.* 4.302!)

<sup>37</sup> For mistakes that can be detected, see, for example, Oberhelman 2003: 92 n. 36: ‘Zielinski’s percentages ... are typically at variance with my calculation of the data (from 1 to 2 percent)’; earlier Axer 1980: 21 n. 1. We have observed similar errors in the Plinian rhythms tabulated by Hofacker 1903.

Our method is not perfect, but we believe that the advantages of getting very accurate — but not perfect — results on large swaths of data in an instant are bigger than the advantages of getting ‘perfect’ results on small amounts of data that take a long time to compile, which cannot be verified and from which it is hard to generalise.<sup>38</sup>

## II DATA

Without further ado, we present in tabular form some of the data that our algorithms have generated. We give first a table of the prose rhythms of most major extant Latin prose authors through the age of Trajan, with Suetonius, Gellius and Apuleius appended. There follow tables of Cicero’s speeches, his rhetorical and philosophical works, and his letters. Finally we give detailed results for Tacitus and Pliny, which we will discuss in the next section. The arrangement within each table is broadly chronological, although perfect consistency in arrangement has proved neither possible nor desirable.

Fragmentary and incomplete works have generally been excluded.<sup>39</sup> We have removed passages in verse from Seneca’s *Apocolocyntosis* and Petronius’ *Satyrica*, but otherwise have not systematically taken special account of verses or quotations.<sup>40</sup> In some authors and works particular caution must be exercised. Given the nature of the *Suasoriae* and *Controversiae*, for example, the statistics for Seneca the Elder are probably of little value and are included only for the sake of completeness. Similar warnings apply to certain texts with particularly small sample sizes or those with unusual transmissions. Numbers never absolve readers of the responsibility to think critically, but with the appropriate caveats in mind, we hope that these numbers will be useful.

The columns in the tables are as follows:

- A. Author and title of work.
- B. Total number of clausulae detected in the work.
- C. Total number of clausulae excluded from consideration (those containing abbreviations, editorially marked textual corruptions, fewer than four syllables and so forth).
- D. Total number of clausulae considered (= B - C).
- E. Percentage of cretic trochees (including resolved forms).
- F. Percentage of double cretics and molossus cretics (including resolved forms).
- G. Percentage of double trochees (including resolved forms).
- H. Percentage of hypodochmiacs (including resolved forms).
- I. Percentage of double spondees (no resolutions).
- J. Percentage of heroic clausulae (no resolutions).
- K. Total percentage of ‘artistic’ clausulae (= E + F + G + H).
- L. Total percentage of double spondees and heroic clausulae (= I + J).
- M. Total percentage of miscellaneous (that is, all other) clausulae.

More detailed tables will be found in the Supplementary Material online (<https://doi.org/10.1017/S0075435819000881>).

<sup>38</sup> The limit case of ‘perfect’ data that cannot be generalised is shown in Koster 2011, where *Pro Roscio Amerino* is laid out in scanned cola according to the ideas of Schmid 1959, but without statistics or further commentary; cf. too, this time with abundant statistics, the massive study of Sträterhoff 1995, over 900 pages devoted to just *De imperio Cn. Pompei* and Livy 1.1.1–2.6.8. A somewhat different approach is illustrated by Vretska and Vretska 1979, where colometry and rhythmic analysis of the *Pro Archia* is a fully integrated part — but only one part — of a broader commentary.

<sup>39</sup> We have also excluded Augustus’ *Res Gestae*, which is a forest of brackets and editorial reconstructions. For some comments on its prose rhythm, see Zwierlein 2002: 43–5.

<sup>40</sup> Note that in Petronius we have not separated narration from dialogue; Müller 1983: 449 claims that the former is rhythmic but the latter is not, and at 449–70 analyses Petronian rhythm in detail. Similar questions may be asked of speeches compared to narrative in historiography, on which see our comments on Sallust and Tacitus below.



TABLE I All authors.

AUTHOR AND WORK	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
Cato, <i>De agri cultura</i>	2099	244	1855	12.67	20.65	14.66	8.41	16.01	4.69	56.39	20.70	22.91
<i>Rhetorica ad Herennium</i>	2577	265	2312	20.72	13.49	33.61	5.54	14.19	3.59	73.36	17.78	8.87
Varro, <i>De lingua Latina</i>	3186	876	2310	16.36	11.73	16.97	6.58	23.59	5.50	51.65	29.09	19.26
<i>De re rustica</i>	2170	399	1771	17.73	8.07	20.72	3.56	27.39	6.44	50.08	33.82	16.09
Cicero: see separate tables												
Caesar, <i>Bellum Gallicum</i> (1–7)	2484	23	2461	20.97	15.93	22.92	3.66	22.06	3.98	63.47	26.05	10.48
<i>Bellum ciuile</i>	1982	192	1790	22.07	17.43	21.06	4.53	19.61	4.58	65.08	24.19	10.73
<i>Bellum Gallicum</i> 8	301	14	287	26.13	8.36	23.34	1.39	24.04	7.32	59.23	31.36	9.41
<i>Bellum Alexandrinum</i>	562	34	528	23.48	14.96	25.76	2.27	21.21	6.25	66.48	27.46	6.06
<i>Bellum Africanum</i>	589	28	561	21.93	11.05	22.10	4.81	23.53	4.99	59.89	28.52	11.59
<i>Bellum Hispaniense</i>	571	189	382	15.45	13.35	20.94	2.88	27.49	8.38	52.62	35.86	11.52
Sallust, <i>Bellum Iugurthinum</i>	1319	28	1291	10.53	22.08	8.83	4.96	25.10	9.84	46.40	34.93	18.67
<i>Bellum Catilinae</i>	699	20	679	9.57	18.56	14.43	6.04	22.24	11.49	48.60	33.73	17.67
<i>Historiae</i> (speeches and letters)	260	14	246	10.57	20.73	9.35	1.63	26.83	8.94	42.28	35.77	21.95
[Sallust], <i>In Ciceronem</i>	48	2	46	15.22	30.43	13.04	4.35	13.04	10.87	63.04	23.91	13.04
[Sallust], <i>Epistulae ad Caesarem</i>	239	6	233	14.16	20.17	9.44	5.58	21.46	12.02	49.36	33.48	17.17

Continued

Table 1 Continued

AUTHOR AND WORK	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
Cornelius Nepos ( <i>Vitae</i> )	1711	28	1683	15.39	22.22	25.31	7.31	18.78	4.04	70.23	22.82	6.95
Livy, <i>Ab urbe condita</i>	30760	1550	29210	10.71	10.69	13.56	4.76	36.45	7.55	39.71	43.99	16.30
Books 1–10 (and preface)	10224	429	9795	14.02	13.18	16.54	6.41	29.38	6.02	50.15	35.41	14.45
Books 21–30	8610	335	8275	8.34	9.52	10.94	4.04	39.90	8.47	32.83	48.37	18.79
Books 31–40	8349	416	7933	9.33	9.33	11.87	3.68	41.65	8.07	34.21	49.72	16.07
Books 41–45	3577	370	3207	10.10	9.48	15.37	4.21	36.23	8.51	39.16	44.75	16.09
Vitruvius, <i>De architectura</i>	3024	101	2923	16.46	8.96	31.65	3.97	19.26	9.96	61.03	29.22	9.75
Seneca the Elder, <i>Controuersiae</i>	8932	805	8127	18.47	25.68	16.46	8.00	15.02	3.08	68.61	18.10	13.29
<i>Controuersiarum excerpta</i>	2804	103	2701	19.55	24.51	18.62	8.92	12.18	2.52	71.60	14.70	13.70
<i>Suasoriae</i>	1095	131	964	20.02	27.07	15.15	6.74	13.90	4.67	68.98	18.57	12.45
Velleius Paterculus	1225	98	1127	15.08	40.64	11.36	12.60	8.96	2.13	79.68	11.09	9.23
Valerius Maximus	3879	235	3644	20.75	20.28	20.75	6.59	15.72	5.54	68.36	21.27	10.37
Celsus, <i>De medicina</i>	8204	1353	6851	24.51	21.24	16.79	7.01	16.25	3.33	69.54	19.57	10.89
Seneca, <i>Dialogi</i>	5728	192	5536	27.44	29.77	13.42	6.83	11.33	2.24	77.46	13.57	8.98
<i>Apocolocyntosis</i> (no verse)	262	29	233	14.16	20.17	18.45	5.58	19.31	4.72	58.37	24.03	17.60
<i>De beneficiis</i>	3812	133	3679	27.40	27.26	11.77	8.26	11.82	2.72	74.69	14.54	10.76
<i>De clementia</i>	594	20	574	16.38	24.56	14.29	10.28	17.07	3.83	65.51	20.91	13.59
<i>Epistulae morales</i>	11394	570	10824	35.19	28.34	11.45	5.95	8.28	2.36	80.92	10.63	8.44
<i>Naturales quaestiones</i>	3742	232	3510	38.03	28.21	10.83	6.27	7.24	2.45	83.33	9.69	6.98

Continued

Table 1 Continued

AUTHOR AND WORK	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
Q. Curtius Rufus	5244	98	5146	47.84	24.54	9.29	3.61	8.47	1.48	85.29	9.95	4.76
Columella, <i>Res rustica</i> (no bk. 10)	6252	597	5655	16.48	23.15	17.68	6.35	16.60	6.26	63.66	22.86	13.47
<i>De arboribus</i>	534	23	511	11.55	19.37	19.57	8.22	18.00	6.65	58.71	24.66	16.63
Scribonius Largus	2184	137	2047	11.38	18.32	17.00	7.18	20.37	4.15	53.88	24.52	21.59
Pomponius Mela	961	23	938	34.97	32.94	7.14	7.57	5.33	2.24	82.62	7.57	9.81
Asconius	1149	298	851	19.98	16.57	22.33	4.82	21.15	6.11	63.69	27.26	9.05
Petronius, <i>Satyrica</i>	2343	274	2069	24.07	20.15	15.32	7.49	19.77	3.82	67.04	23.59	9.38
Pliny the Elder, <i>Naturalis historia</i>	31008	4664	26344	18.47	20.28	17.45	8.65	16.33	4.27	64.84	20.60	14.56
Quintilian, <i>Institutio oratoria</i>	11433	453	10980	26.26	26.84	18.45	6.68	10.20	2.92	78.23	13.12	8.64
<i>Declamationes minores</i>	8887	522	8365	19.45	20.80	19.19	8.38	15.83	4.05	67.82	19.88	12.30
[Quintilian], <i>Declamationes maiores</i>	5984	293	5691	35.60	22.40	17.89	8.14	6.89	1.90	84.03	8.79	7.19
Pliny the Younger and Tacitus: see separate tables												
Suetonius, <i>Vitae</i>	4155	368	3787	29.23	23.00	22.97	5.36	8.53	2.83	80.57	11.35	8.08
<i>De grammaticis et rhetoribus</i>	199	18	181	20.99	20.99	26.52	4.97	11.05	5.52	73.48	16.57	9.94
Gellius, <i>Noctes Atticae</i>	7536	349	7187	10.67	27.20	19.38	6.21	21.55	3.55	63.46	25.10	11.44
Apuleius, <i>Apologia</i>	1327	146	1181	16.51	18.54	24.89	2.79	21.59	4.06	62.74	25.66	11.60

Continued

Table 1 Continued

AUTHOR AND WORK	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Metamorphoses</i>	2809	83	2726	19.37	26.19	25.46	7.48	8.25	3.41	78.50	11.67	9.83
<i>Florida</i>	390	69	321	17.76	25.55	28.97	3.74	11.21	5.30	76.01	16.51	7.48
<i>De deo Socratis</i>	238	20	218	30.73	24.77	18.81	2.75	11.47	3.21	77.06	14.68	8.26
<i>De deo Socratis</i> (preface)	51	11	40	27.50	30.00	25.00	2.50	7.50	0.00	85.00	7.50	7.50
<i>De Platone et eius dogmate</i>	487	41	446	29.15	22.42	28.03	3.59	9.19	2.91	83.18	12.11	4.71
<i>De mundo</i>	396	45	351	32.48	21.37	22.79	3.13	6.55	4.27	79.77	10.83	9.40

TABLE 2 Cicero's speeches.

SPEECHES												
TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Pro Quintio</i>	670	83	587	21.47	25.89	22.66	6.13	11.93	2.73	76.15	14.65	9.20
<i>Pro Roscio Amerino</i>	901	92	809	21.01	27.32	24.85	5.44	9.02	2.84	78.62	11.87	9.52
<i>Pro Caecina</i>	644	38	606	25.41	24.09	24.59	4.79	10.73	3.14	78.88	13.86	7.26
<i>Pro Tullio</i>	301	90	211	17.06	27.49	22.75	6.16	11.37	0.95	73.46	12.32	14.22
<i>In Caecilium</i>	334	18	316	31.65	26.27	23.10	5.06	6.65	1.90	86.08	8.54	5.38
<i>In Verrem</i> (all)	6728	406	6322	27.00	23.98	24.91	5.96	9.22	2.12	81.86	11.34	6.80
<i>Verr.</i> I	254	11	243	28.81	22.63	20.58	8.64	9.88	3.70	80.66	13.58	5.76
<i>Verr.</i> II.1	1084	69	1015	26.01	23.94	22.56	6.90	11.23	1.67	79.41	12.91	7.68
<i>Verr.</i> II.2	1256	47	1209	25.72	25.06	23.33	6.53	9.10	2.15	80.65	11.25	8.11
<i>Verr.</i> II.3	1778	149	1629	24.19	24.00	28.05	5.34	9.39	2.21	81.58	11.60	6.81
<i>Verr.</i> II.4	1193	82	1111	29.07	25.83	23.94	5.76	7.29	2.34	84.61	9.63	5.76
<i>Verr.</i> II.5	1163	47	1116	30.91	21.24	26.08	5.02	9.05	1.79	83.24	10.84	5.91
<i>Pro Fonteio</i>	273	59	214	36.92	18.69	23.36	5.14	8.88	0.93	84.11	9.81	6.07
<i>Pro Roscio comoedo</i>	490	66	424	19.34	17.45	20.28	4.01	21.70	3.54	61.08	25.24	13.68
<i>Pro lege Manilia</i>	299	15	284	46.13	13.03	31.34	1.41	5.99	1.06	91.90	7.04	1.06
<i>Pro Cluentio</i>	1246	48	1198	30.88	20.95	25.79	5.68	9.85	1.75	83.31	11.60	5.09
<i>De lege agraria</i> (all)	851	65	786	29.77	21.50	25.70	6.87	8.78	1.78	83.84	10.56	5.60
<i>Agr.</i> 1	149	7	142	24.65	21.13	30.99	8.45	9.15	0.00	85.21	9.15	5.63
<i>Agr.</i> 2	606	46	560	31.79	20.54	25.00	6.61	8.39	2.32	83.93	10.71	5.36

Continued

Table 2 Continued

SPEECHES												
TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Agr. 3</i>	96	12	84	25.00	28.57	21.43	5.95	10.71	1.19	80.95	11.90	7.14
<i>Pro Rabirio perduellionis reo</i>	177	36	141	31.21	24.11	24.82	4.26	9.93	2.13	84.40	12.06	3.55
<i>In Catilinam (all)</i>	676	24	652	34.20	25.77	24.39	5.21	4.29	1.38	89.57	5.67	4.75
<i>Cat. 1</i>	200	10	190	34.74	24.21	18.95	8.42	3.68	3.68	86.32	7.37	6.32
<i>Cat. 2</i>	183	4	179	28.49	34.64	24.02	4.47	2.79	0.56	91.62	3.35	5.03
<i>Cat. 3</i>	153	7	146	39.73	19.18	26.03	2.74	7.53	0.68	87.67	8.22	4.11
<i>Cat. 4</i>	140	3	137	35.04	23.36	30.66	4.38	3.65	0.00	93.43	3.65	2.92
<i>Pro Murena</i>	658	54	604	27.98	27.48	25.33	5.13	7.95	0.83	85.93	8.77	5.30
<i>Pro Sulla</i>	524	25	499	31.26	23.25	27.86	3.21	8.02	1.20	85.57	9.22	5.21
<i>Pro Archia</i>	144	12	132	34.85	18.18	36.36	3.79	4.55	0.76	93.18	5.30	1.52
<i>Pro Flacco</i>	841	69	772	26.30	22.93	31.35	4.02	7.51	1.55	84.59	9.07	6.35
<i>Post reditum in senatu</i>	182	12	170	39.41	12.94	32.94	1.18	10.59	1.18	86.47	11.76	1.76
<i>Post reditum ad populum</i>	107	6	101	37.62	10.89	37.62	2.97	4.95	3.96	89.11	8.91	1.98
<i>De domo sua</i>	809	73	736	29.21	19.29	32.61	6.25	6.25	1.63	87.36	7.88	4.76
<i>Pro Sestio</i>	909	39	870	33.22	16.09	28.97	3.79	11.38	2.30	82.07	13.68	4.25
<i>In Vatinius</i>	200	10	190	32.63	24.74	28.95	2.63	8.95	0.00	88.95	8.95	2.11

Continued

Table 2 Continued

SPEECHES												
TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>De haruspicum responsis</i>	423	22	401	35.41	19.45	28.68	2.74	8.98	2.74	86.28	11.72	2.00
<i>De prouinciis consularibus</i>	265	11	254	34.65	23.23	31.50	1.18	5.51	1.57	90.55	7.09	2.36
<i>Pro Caelio</i>	528	14	514	31.71	21.21	32.49	3.50	4.67	1.56	88.91	6.23	4.86
<i>Pro Balbo</i>	348	27	321	34.89	16.51	38.94	1.25	5.30	0.93	91.59	6.23	2.18
<i>In Pisonem</i>	663	46	617	30.63	20.26	29.34	5.02	7.78	1.94	85.25	9.72	5.02
<i>Pro Scauro</i>	220	47	173	33.53	14.45	31.21	1.16	11.56	0.58	80.35	12.14	7.51
<i>Pro Plancio</i>	669	31	638	29.47	21.00	35.11	2.51	5.49	2.19	88.09	7.68	4.23
<i>Pro Rabirio Postumo</i>	303	33	270	21.85	28.52	21.11	6.30	12.59	2.59	77.78	15.19	7.04
<i>Pro Milone</i>	639	40	599	24.21	27.88	27.05	4.84	8.68	2.50	83.97	11.19	4.84
<i>Pro Marcello</i>	150	5	145	23.45	35.86	24.83	10.34	3.45	0.00	94.48	3.45	2.07
<i>Pro Ligario</i>	257	7	250	25.20	26.80	23.20	4.80	8.40	2.00	80.00	10.40	9.60
<i>Pro rege Deiotaro</i>	255	8	247	27.94	33.60	20.65	3.64	8.10	0.81	85.83	8.91	5.26
<i>Philippicae</i>	4187	230	3957	27.70	28.43	23.28	5.00	8.72	1.39	84.41	10.11	5.48
<i>Phil. 1</i>	288	15	273	28.94	27.84	21.98	4.40	11.36	0.73	83.15	12.09	4.76
<i>Phil. 2</i>	980	40	940	29.26	25.74	23.94	4.79	9.47	1.70	83.72	11.17	5.11
<i>Phil. 3</i>	264	15	249	29.72	30.52	26.10	3.61	4.42	0.40	89.96	4.82	5.22
<i>Phil. 4</i>	103	7	96	33.33	30.21	28.13	2.08	2.08	1.04	93.75	3.13	3.13
<i>Phil. 5</i>	397	26	371	28.57	25.61	20.22	6.74	9.16	1.62	81.13	10.78	8.09

Continued

Table 2 Continued

SPEECHES												
TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Phil.</i> 6	162	11	151	30.46	24.50	25.17	3.97	7.28	1.32	84.11	8.61	7.28
<i>Phil.</i> 7	172	14	158	20.89	31.65	23.42	5.70	11.39	1.27	81.65	12.66	5.70
<i>Phil.</i> 8	279	17	262	25.95	28.24	19.47	4.96	12.21	3.05	78.63	15.27	6.11
<i>Phil.</i> 9	87	9	78	33.33	24.36	25.64	7.69	6.41	0.00	91.03	6.41	2.56
<i>Phil.</i> 10	192	15	177	26.55	32.77	19.21	6.21	6.78	2.82	84.75	9.60	5.65
<i>Phil.</i> 11	320	15	305	23.93	30.82	24.92	6.89	8.85	0.66	86.56	9.51	3.93
<i>Phil.</i> 12	291	18	273	24.54	37.73	17.58	5.13	7.69	0.73	84.98	8.42	6.59
<i>Phil.</i> 13	438	22	416	27.40	26.92	24.76	4.81	8.17	1.68	83.89	9.86	6.25
<i>Phil.</i> 14	214	6	208	26.92	28.85	29.81	2.40	8.65	0.48	87.98	9.13	2.88
All speeches	26871	1861	25010	28.42	23.78	26.24	4.98	8.80	1.86	83.42	10.66	5.91
[Cicero], <i>In Sallustium</i>	112	4	108	28.70	14.81	24.07	3.70	18.52	3.70	71.30	22.22	6.48



TABLE 3 Cicero's rhetorica and philosophica.

TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<b>Rhetorica</b>												
<i>De inuentione</i>	2254	87	2167	20.86	17.72	29.95	6.83	13.06	2.81	75.36	15.87	8.77
<i>De oratore</i>	3370	175	3195	30.58	24.95	24.41	6.23	7.39	1.91	86.17	9.30	4.54
<i>De partitione oratoria</i>	600	15	585	28.21	24.96	24.96	6.15	7.35	2.39	84.27	9.74	5.98
<i>De optimo genere oratorum</i>	107	3	104	12.50	38.46	25.96	8.65	7.69	0.00	85.58	7.69	6.73
<i>Brutus</i>	1655	146	1509	21.40	38.50	21.54	7.09	6.30	0.86	88.54	7.16	4.31
<i>Orator</i>	1205	17	1188	22.22	38.64	18.69	9.18	5.47	2.10	88.72	7.58	3.70
<i>Topica</i>	573	15	558	16.85	32.97	20.61	6.81	10.57	2.51	77.24	13.08	9.68
<b>Philosophica</b>												
<i>De re publica</i>	1593	497	1096	17.43	40.88	17.06	8.03	6.20	1.73	83.39	7.94	8.67
<i>De legibus</i>	1235	230	1005	19.10	39.70	21.29	7.86	5.87	1.69	87.96	7.56	4.48
<i>Paradoxa Stoicorum</i>	318	29	289	20.42	31.49	21.11	9.69	7.27	2.08	82.70	9.34	7.96
<i>Lucullus</i>	1163	53	1110	23.42	34.05	18.83	8.20	8.20	1.98	84.50	10.18	5.32
<i>Academica</i>	228	8	220	25.00	25.91	22.27	7.73	9.09	3.18	80.91	12.27	6.82
<i>De finibus</i>	2759	101	2658	21.07	32.73	23.55	7.11	6.77	2.14	84.46	8.92	6.62
<i>Tusculanae disputationes</i>	3386	189	3197	22.46	32.22	21.36	7.13	7.16	2.88	83.17	10.04	6.79

*Continued*

Table 3 Continued

TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>De natura deorum</i>	2148	35	2113	26.03	29.44	23.52	6.67	5.21	4.12	85.66	9.32	5.02
<i>De diuinatione</i>	1987	108	1879	24.06	30.12	20.33	6.07	7.18	4.95	80.57	12.13	7.29
<i>De fato</i>	349	20	329	21.58	30.40	20.67	7.90	8.21	3.34	80.55	11.55	7.90
<i>De senectute</i>	550	27	523	28.49	36.14	17.97	5.74	4.40	3.06	88.34	7.46	4.21
<i>De amicitia</i>	586	36	550	31.09	31.09	20.18	4.36	7.45	1.45	86.73	8.91	4.36
<i>De officiis</i>	1895	84	1811	27.39	33.24	23.14	6.35	3.87	1.88	90.12	5.74	4.14
<i>Timaeus</i>	192	1	191	24.08	38.22	24.61	5.76	2.62	2.62	92.67	5.24	2.09

TABLE 4 Cicero's letters.

LETTERS												
TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Epistulae ad Atticum</i>	11955	1392	10563	16.80	26.06	19.80	7.01	13.31	4.90	69.68	18.21	12.11
<i>Att. 1</i>	686	59	627	19.94	24.24	24.40	4.47	14.19	4.47	73.05	18.66	8.29
<i>Att. 2</i>	805	88	717	19.80	23.99	21.06	7.53	11.85	3.77	72.38	15.62	11.99
<i>Att. 3</i>	492	74	418	18.18	22.25	18.18	5.50	15.55	10.77	64.11	26.32	9.57
<i>Att. 4</i>	731	120	611	19.15	24.55	20.79	6.38	13.42	5.89	70.87	19.31	9.82
<i>Att. 5</i>	753	126	627	13.72	28.23	15.95	7.66	16.91	4.63	65.55	21.53	12.92
<i>Att. 6</i>	660	66	594	15.32	28.45	19.19	7.58	11.62	4.21	70.54	15.82	13.64
<i>Att. 7</i>	899	122	777	15.19	27.28	16.99	7.08	13.77	5.15	66.54	18.92	14.54
<i>Att. 8</i>	595	69	526	15.40	29.28	17.68	6.27	14.26	6.08	68.63	20.34	11.03
<i>Att. 9</i>	851	107	744	14.11	26.08	19.35	8.20	13.44	6.32	67.74	19.76	12.50
<i>Att. 10</i>	710	76	634	15.46	25.08	22.40	7.26	13.25	4.26	70.19	17.51	12.30
<i>Att. 11</i>	662	80	582	16.15	26.98	19.24	6.36	14.60	5.50	68.73	20.10	11.17
<i>Att. 12</i>	918	66	852	17.84	27.23	21.83	7.86	9.62	4.58	74.77	14.20	11.03
<i>Att. 13</i>	1052	108	944	16.53	23.73	20.55	7.63	13.35	3.71	68.43	17.06	14.51
<i>Att. 14</i>	695	64	631	19.02	26.15	19.33	7.61	12.36	3.96	72.11	16.32	11.57
<i>Att. 15</i>	750	96	654	17.89	25.84	19.88	6.88	15.29	3.67	70.49	18.96	10.55
<i>Att. 16</i>	696	71	625	15.52	27.84	18.56	6.24	11.68	4.32	68.16	16.00	15.84
<i>Epistulae ad Quintum fratrem</i>	1426	143	1283	21.82	25.88	21.82	6.39	11.93	3.66	75.92	15.59	8.50

Table 4 Continued

LETTERS												
TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Q. fr.</i> 1.1	198	3	195	29.74	27.18	26.15	9.74	3.59	0.00	92.82	3.59	3.59
<i>Q. fr.</i> , 1.1 excluded	1228	140	1088	20.40	25.64	21.05	5.79	13.42	4.32	72.89	17.74	9.38
<i>Epistulae ad familiares</i>	8031	952	7079	20.38	26.70	22.26	6.06	11.43	2.75	75.41	14.18	10.41
<i>Epistulae ad Brutum</i> (all)	690	49	641	19.66	28.39	17.16	6.86	10.45	6.40	72.07	16.85	11.08
<i>Ad Brut.</i> (1.16 and 17 excluded)	571	43	528	22.16	28.98	19.13	7.20	7.95	5.11	77.46	13.07	9.47
<i>Ad Brut.</i> : Brutus	114	10	104	4.81	27.88	13.46	4.81	16.35	9.62	50.96	25.96	23.08
<i>Ad Brut.</i> : Cicero	457	33	424	26.42	29.01	20.52	7.78	5.66	4.01	83.73	9.67	6.60
<i>Ad Brut.</i> : 1.16, 17	119	6	113	7.96	25.66	7.96	5.31	22.12	12.39	46.90	34.51	18.58
[Cicero], <i>Epistula ad Octavianum</i>	66	3	63	33.33	17.46	33.33	3.17	7.94	0.00	87.30	7.94	4.76

TABLE 5 Tacitus.

TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Dialogus</i>	443	31	412	23.06	18.69	23.30	5.10	17.96	2.67	70.15	20.63	9.22
<i>Agricola</i>	443	14	429	16.32	17.25	19.11	7.46	21.91	4.90	60.14	26.81	13.05
<i>Germania</i>	460	5	455	22.20	18.02	20.00	6.81	17.14	5.49	67.03	22.64	10.33
<i>Historiae</i>	3724	36	3688	17.90	16.46	16.89	5.72	24.08	6.21	56.97	30.29	12.74
<i>Hist.</i> 1	764	9	755	19.47	18.54	17.48	5.96	21.85	5.70	61.46	27.55	10.99
<i>Hist.</i> 2	882	5	877	16.76	16.31	17.33	4.79	24.63	6.73	55.19	31.36	13.45
<i>Hist.</i> 3	837	9	828	16.91	16.79	17.63	5.80	24.28	5.43	57.13	29.71	13.16
<i>Hist.</i> 4	966	9	957	18.18	15.26	15.78	6.27	24.76	6.17	55.49	30.93	13.58
<i>Hist.</i> 5	275	4	271	19.19	14.39	15.50	5.90	25.46	8.49	54.98	33.95	11.07
<i>Annales</i>	5781	120	5661	15.63	17.28	17.12	6.18	23.53	6.43	56.21	29.96	13.83
<i>Ann.</i> 1	769	16	753	14.21	18.73	17.93	6.11	23.37	6.51	56.97	29.88	13.15
<i>Ann.</i> 2	685	8	677	13.29	18.76	16.99	3.40	25.85	5.61	52.44	31.46	16.10
<i>Ann.</i> 3	605	15	590	17.12	15.42	17.80	4.58	21.19	8.31	54.92	29.49	15.59
<i>Ann.</i> 4	646	9	637	13.97	17.27	16.80	7.85	24.96	6.44	55.89	31.40	12.72
<i>Ann.</i> 5	42	9	33	12.12	27.27	15.15	6.06	15.15	6.06	60.61	21.21	18.18
<i>Ann.</i> 6	475	22	453	15.89	18.10	17.66	5.08	22.74	5.52	56.73	28.26	15.01
( <i>Ann.</i> 5 and 6 combined)	517	31	486	15.64	18.72	17.49	5.14	22.22	5.56	57.00	27.78	15.23
<i>Ann.</i> 11	318	8	310	14.84	17.10	19.35	9.03	23.55	5.81	60.32	29.35	10.32
<i>Ann.</i> 12	473	5	468	18.80	14.74	16.45	7.05	23.72	5.34	57.05	29.06	13.89
<i>Ann.</i> 13	443	4	439	14.35	15.49	18.45	6.38	25.74	7.06	54.67	32.80	12.53

Continued

Table 5 Continued

TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<i>Ann.</i> 14	528	11	517	16.83	18.76	14.70	6.77	22.24	6.96	57.06	29.21	13.73
<i>Ann.</i> 15	560	11	549	17.30	16.94	16.03	6.38	23.50	6.56	56.65	30.05	13.30
<i>Ann.</i> 16	237	2	235	18.30	16.17	17.02	8.51	20.43	5.96	60.00	26.38	13.62

TABLE 6 Pliny the Younger.

TITLE	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
Pliny, <i>Epistulae</i> 1-9	4767	391	4376	40.84	21.80	18.26	3.98	6.10	2.29	84.87	8.39	6.74
<i>Ep.</i> 1	533	42	491	38.90	23.01	19.14	4.48	4.89	2.65	85.54	7.54	6.92
<i>Ep.</i> 2	512	32	480	41.25	21.25	20.00	3.54	6.25	2.08	86.04	8.33	5.63
<i>Ep.</i> 3	533	28	505	39.21	18.02	23.37	2.97	8.12	1.39	83.56	9.50	6.93
<i>Ep.</i> 4	531	51	480	40.83	23.75	16.25	4.38	5.83	1.88	85.21	7.71	7.08
<i>Ep.</i> 5	499	33	466	40.56	25.54	16.95	4.08	6.01	1.93	87.12	7.94	4.94
<i>Ep.</i> 6	612	54	558	39.96	22.04	18.46	4.48	5.73	2.33	84.95	8.06	6.99
<i>Ep.</i> 7	537	49	488	42.62	21.93	16.60	2.87	7.38	2.66	84.02	10.04	5.94
<i>Ep.</i> 8	446	33	413	44.55	20.82	15.98	5.08	4.60	2.91	86.44	7.51	6.05
<i>Ep.</i> 9	564	69	495	40.40	20.00	16.97	4.04	5.86	2.83	81.41	8.69	9.90
<i>Epistulae</i> 10	568	20	548	33.39	16.97	19.34	7.30	10.95	3.10	77.01	14.05	8.94
<i>Ep.</i> 10: Pliny	415	18	397	36.27	15.87	22.17	6.30	8.82	2.77	80.60	11.59	7.81
<i>Ep.</i> 10: Trajan	153	2	151	25.83	19.87	11.92	9.93	16.56	3.97	67.55	20.53	11.92
<i>Panegyricus</i>	1288	31	1257	39.30	22.28	20.13	3.66	7.48	1.75	85.36	9.23	5.41

## III ANALYSIS

The foregoing tables provide an order of magnitude more information about Latin clausulae than has been available before, and they provide it all in one place with a consistent methodology. We hope that they will prove useful in a variety of research questions, and we give a sample of such questions below. These only scratch the surface of what we think is possible. We begin with a new approach to determining statistical significance in prose rhythm data, and then proceed to specific questions about the prose rhythm practices of individual authors like Cicero, Sallust, Tacitus and Pliny the Younger.

*How Do You Tell If Any of These Data Are Meaningful? A New Approach*

It is not necessarily obvious that the use of particular sequences of short and long syllables should be regarded as a consciously sought artistic phenomenon in Latin prose. After all, every Latin syllable is long or short, and so every sentence must end with *some* pattern of longs and shorts.<sup>41</sup> Furthermore, the character of the Latin language itself will dictate that some patterns occur more frequently than others: long syllables are more common than short, for example, and so it would surprise no one to hear that -- -x is more common than ~~~x. Likewise, many authors favour verbs at the ends of clauses (i.e., in an important clausal position), and the third person and past tense are disproportionately represented in our surviving texts. These and many other tightly intertwined biases make it extremely hard — we think impossible — to establish any kind of ‘baseline’ expected distribution of rhythms. There is simply no way to say that you would ‘expect’ Latin sentences to end with a cretic-trochee 6 per cent of the time: what do you base your expectations on?

Scholars have generally taken one of three approaches to this question. Some, like Zielinski, ignored it altogether, and simply presented absolute numbers and percentages. But from the beginning it was objected that, for example, reporting that clausulae of the *ēssē uīdēātūr* type occur with 4.7 per cent frequency in Cicero’s speeches whereas the type *ōmnēs ēssēt* occur 6.4 per cent of the time is not in itself useful. What if *ēssē uīdēātūr*-type clausulae naturally occur in Latin 2.4 per cent of the time, while the type *ōmnēs ēssēt* naturally occurs 23.5 per cent of the time? Then the real point of interest would be that Cicero sought out the former and deliberately avoided the latter, but this is hidden behind the absolute frequencies ‘4.7 per cent’ and ‘6.4 per cent’.<sup>42</sup> To determine the significance of any observed frequency, it must somehow be compared against an expected baseline.

A second approach has been to calculate an expected value based on a ‘neutral’ sample of Latin. Albert De Groot at one point tried sampling scholarly translations of Greek texts made in the nineteenth century, but it is almost impossible to say how such scholarly Latin would map onto a native speaker’s intuitions about rhythm.<sup>43</sup> François Novotný looked at the distribution of syllables *not* in clausal position, but this is to compare different things.<sup>44</sup> Others have tried still other approaches: Henri Bornecque, for example, considered the proportions of various patterns in authors whom he deemed

<sup>41</sup> cf. Quint., *Inst.* 9.4.61: ‘neque enim loqui possum nisi e syllabis breuibus ac longis, ex quibus pedes fiunt.’

<sup>42</sup> So De Groot 1926: 20–1, whose examples we have borrowed here. On Cicero’s preference for the *esse uideatur* type, see, for example, Quint., *Inst.* 9.4.73, 10.2.18, Tac., *Dial.* 23.1; for Zielinski’s weakness on this score, see, for example, Bornecque 1907: 212–14; Shewring 1930: 165; Oberhelman 2003: 98–9; cf. Aumont 1996: 13–14.

<sup>43</sup> De Groot 1921; detailed criticism in Aili 1979: 21–5. See also Wilkinson 1963: 140–1.

<sup>44</sup> Summarised in Novotný 1929: 25–7; in more detail Novotný 1926.



unlikely to be rhythmic.<sup>45</sup> But this is arbitrary at best, and circular at worst; deciding that Sallust, say, is unrhythmic, and using his numbers as a baseline, is simply to assume your desired conclusion, and it is not much helped if you add a few other authors into the mix.<sup>46</sup>

In response to the problems of external comparison, Tore Janson and his student Hans Aili pioneered a form of ‘internal comparison’.<sup>47</sup> They looked at a sample of an individual author’s clausulae and determined the frequency of longs and shorts in each position (that is, what percentage of penultimate syllables are long, what percentage of antepenultimate syllables are long and so forth). From this they calculated an expected frequency for each type of clausula in that author, which is simply the product of the observed frequencies for each individual syllable.<sup>48</sup> Then they could compare the observed percentage of a given clausula with its expected value and run statistical tests on their results. This method is ingenious, but it has a fundamental weakness that vitiates any statistics derived from it: these scholars base their ‘expected’ values on the very material that they are trying to observe. If an author systematically seeks certain clausulae and avoids others, those preferences will already be part of the ‘expected’ values and so cannot be called neutral or natural. It is a circular procedure.<sup>49</sup>

We propose a new approach to the question of expected values. We think that the only secure basis for comparison is to look at the tendencies of individual authors and attempt to determine whether there are statistically significant differences in their practices. If so, then we can at least say that the differences among authors are unlikely to be due to random chance. Until now, this task was more or less impossible, because while there exist studies of individual authors’ rhythmic tendencies, the scholars carrying out these studies made different assumptions and employed different methodologies. Our data, by contrast, allow a comparison of like with like across all of Latin prose. Furthermore, in authors with sufficiently large corpora, we can also consider a portion of the corpus and determine whether its rhythmic practices match the rest of the corpus. So with Cicero’s speeches, for example, we can consider each individual speech separately and compare it to the rest of his corpus with that speech removed. Indeed, such a comparison can even be applied to individual letters of Cicero’s to determine whether it is likely that he paid extra attention to rhythm in them, or, with some further work, to compare the rhythmic practices of speeches and narrative in a historian. We will carry out all of these tests in the following sections.

Any such statistical tests must be used with appropriate caution, for their results are wholly determined by the data input. Take Varro and his two substantially extant works, *De lingua Latina* and *De re rustica*. We could consider his distribution of clausulae in five categories (including resolutions in each): cretic-trochaic, double cretic or molossus cretic, double trochee, hypodochmiac, and ‘everything else’. We would then have a table of data like this:

<sup>45</sup> Sall., *Iug.*, Tac., *Ann.* 1, Brutus’ letters to Cicero, Trajan’s letters to Pliny, Fronto’s letters to Marcus Aurelius; Bornecque 1907: 216; critique in, for example, Oberhelman 2003: 115–16. Aumont 1996 still uses Bornecque’s ‘non-metrical’ data (esp. at 67).

<sup>46</sup> For critique of these and other methods, see Aili 1979: 21–32; Orlandi 2005: 396–401.

<sup>47</sup> Janson 1975, esp. 10–34 (applying the method to the medieval *cursus*); Aili 1979, esp. 32–9.

<sup>48</sup> So, labelling the first long of a cretic trochee position 5, the following short position 4, and so on: expected frequency of  $-- -x =$  observed percentage of long in position 5 multiplied by observed percentage of short in position 4 multiplied by observed percentage of long in position 3 multiplied by observed percentage of long in position 2 multiplied by 1 (since the last syllable is indifferent). An example is provided by Aili 1979: 36, another by Oberhelman 2003: 177–9.

<sup>49</sup> As observed by Gotoff 1981: 337; cf. Janson 1975: 26–8. Detailed further criticism in Aumont 1996: 47–57.

	CRETIC TROCHEE	DOUBLE CRETIC	DOUBLE TROCHEE	HYPODOCHMIAC	OTHER
<i>De lingua Latina</i>	378 (16.36%)	271 (11.73%)	392 (16.97%)	152 (6.58%)	1117 (48.35%)
<i>De re rustica</i>	314 (17.73%)	143 (8.07%)	367 (20.72%)	63 (3.56%)	884 (49.92%)

The most appropriate statistical test to analyse such data, and one with a long history in studies of prose rhythm, is the chi-square test.<sup>50</sup> The details are available in any statistical handbook,<sup>51</sup> but in essence, the chi-square test applied to this data will test the null hypothesis that the two rows of data come from the same distribution and that variation between the two is merely due to chance. (This is not a measure of degree of difference between two samples, but a test of whether these differences are unlikely to arise by chance if both samples were drawn from identical populations.) From our chi-square test statistic is derived a *p*-value; if our *p*-value is below a certain threshold (in this paper, as often, .05), we reject the null hypothesis and conclude that there is a statistically significant difference between the two rows of data.<sup>52</sup> Put plainly, the chi-square test allows us to say whether an apparent difference in authors' use of particular clausulae is in fact statistically significant.<sup>53</sup>

If we run a chi-square test on the above five columns of data, we get  $\chi^2 = 39.796$ ; with four degrees of freedom this results in a *p*-value near zero.<sup>54</sup> Such a value indicates that it is almost impossible for the prose rhythms of these two works to belong to the same distribution. But *a priori* this is very unlikely; Varro wrote both of them, and the rhythms of neither look to be 'artistically' rhythmic in the Ciceronian sense of the term. A test treating these five columns of data appears too sensitive. If, however, we pool the data differently and group our 'artistic' clausulae (cretic trochees, double cretics, ditrochees and hypodochmiacs) together and our 'non-artistic' clausulae (double spondees, heroic clausulae and everything else) together, we can look instead at the two columns of the following table:

<sup>50</sup> The test is used by, for example, Janson 1975; Aili 1979; McCabe 1981; Aumont 1996; Hutchinson 2015; 2018.

<sup>51</sup> See, for example, <https://onlinecourses.science.psu.edu/stat500/node/56/>; a very useful online calculator is Preacher 2001. Hutchinson 2018: 20 tries to explain the test for classicists; similarly Hutchinson 2015: 792, and earlier Aili 1979: 37–9; McCabe 1981: 176–83; Aumont 1996: 69–72.

<sup>52</sup> Note that the chi-square test statistic is also correlated with sample size: the larger the samples, the more statistically significant will be the variation between them. (More random variation is possible in a smaller sample: if paragraph A has two artistic and one non-artistic clausulae, while paragraph B has one artistic and two non-artistic clausulae, the variation may be due to chance. If, on the other hand, text A has 2,000 artistic clausulae and 1,000 non-artistic, whereas text B has 1,000 artistic compared to 2,000 non-artistic, chance is a much less likely explanation for the observed variance.) The chi-square test has certain minimum requirements on sample size, which are met in this paper.

<sup>53</sup> The statistically savvy may wonder about a philosophical question: the chi-square test is usually used to compare two random samples in order to infer whether the populations from which they were drawn are different. Here, however, we might be thought not to have a sample but rather the entire population (all clausulae), thus obviating the need for such a test. In a real sense, however, we do not have the whole population: most of classical literature has perished. Since much of Tacitus' *Annales* and *Historiae* have been lost, for example, what we have is a sample of all of Tacitus. Is our sample random? Admittedly not in the way a statistician would prefer, but it is random in the sense that the works that have been preserved were *not* preserved because of their rhythmic properties (although those properties could sometimes be correlated with other reasons that they were preserved, like 'literary quality'). We thank one of the anonymous *JRS* readers for insightful comments on this issue, which we hope to explore further elsewhere.

<sup>54</sup> In this article, we will give *p*-values to five decimal places, hence here  $p \approx .00000$ . With more decimal places, here  $p = .0000005$ .

	ARTISTIC	NON-ARTISTIC
<i>De lingua Latina</i>	1193 (51.64%)	1117 (48.36%)
<i>De re rustica</i>	887 (50.08%)	884 (49.92%)

A glance at these proportions will show that they are very similar. It is no surprise, then, that a chi-square test on these data yields  $\chi^2 = 0.977$ , producing a  $p$ -value of about 0.32294. This  $p$ -value, by contrast, indicates that it is reasonable to conclude that any deviation in the prose rhythms of these two works is due to random chance. We get the same result if we compare the individual books of *De lingua Latina* and *De re rustica* using a chi-square test of ‘artistic’ vs ‘non-artistic’ clausulae: there are no statistically significant differences in preferences for artistic and non-artistic clausulae among the various books.

These two very different results are a salutary warning that statistical tests must be used cautiously, and always with an eye on the underlying data and reasonable expectations.<sup>55</sup> The choice of collapsing our data into two categories of artistic and non-artistic clausulae is, again, fundamentally a pragmatic one. It produces sensible and interesting results. It has the further virtue of agreeing with many of the theoretical models that have been constructed for Latin prose rhythm. But there may be better — and there are certainly other — ways of dividing the data, and binary tests between ‘artistic’ and ‘non-artistic’ clausulae should simply be seen as one useful tool, not as some kind of definitive measure.

This test also suggests that we should adjust certain assumptions, as another example will make clear. We can compare Varro’s *De re rustica* and Cato’s *De agri cultura* using our ‘artistic’ vs ‘non-artistic’ model as follows:

	ARTISTIC	NON-ARTISTIC
Varro, <i>De re rustica</i>	887 (50.08%)	884 (49.92%)
Cato, <i>De agri cultura</i>	1046 (56.38%)	809 (43.62%)

$\chi^2 = 14.463$ ,  $p$ -value  $\approx 0.00014$ . These two authors, according to our test, almost certainly show different propensities to artistic clausulae. The commonly accepted prior assumption is that neither Varro nor Cato cares about prose rhythm, but we suggest that this assumption is wrong. It is all but certain that any Latin author had intuitive preferences for some rhythms and unconsciously avoided others. Indeed, this is borne out by our data: when we look at our tables for all authors’ prose rhythm preferences, we nowhere see, even in supposedly ‘unrhythmic’ authors, convergence around particular baseline numbers. This should not be surprising: in English no one would expect Jonathan Franzen and David Foster Wallace to share the same rhythmic tendencies, even if they were contemporaries and friends who wrote in the same genres for similar audiences. All Latin authors have their own rhythmic profiles, and thus no universal expected values can be established. But authors can be compared with each other, and furthermore, authors can be compared with the artificial system of ‘artistic’ clausulae adopted by Cicero and many later writers.

So Varro is consistent with Varro, and Caesar is consistent with Caesar:

	ARTISTIC	NON-ARTISTIC
<i>Bellum Gallicum</i> (1–7)	1562 (63.47%)	915 (36.53%)
<i>Bellum ciuile</i>	1165 (65.08%)	625 (34.92%)

<sup>55</sup> See the sensible preliminary cautions of Aumont 1996: 9.

$\chi^2 = 1.173$ ,  $p$ -value  $\approx 0.27879$ : any variation in Caesar's tendency toward artistic clausulae between the *Bellum Gallicum* and the *Bellum civile* is not statistically significant. By contrast, Varro and Caesar clearly differ from each other:

	ARTISTIC	NON-ARTISTIC
Varro	2080 (51.97%)	2001 (49.03%)
Caesar	2727 (64.15%)	1524 (35.85%)

$\chi^2 = 148.224$ ,  $p$ -value  $\approx 0$ : these two authors do not have the same preferences at all. If we say that they are 'not rhythmic', what we really mean is that they do not follow the distribution of clausulae characteristic of Cicero, because they clearly have their own tendencies in how they distribute longs and shorts.<sup>56</sup>

It is pretty clear from our data that no two authors show the same rhythms, although many authors are consistent with themselves in their preferences (so, for example, Sallust). What also seems pretty clear is that some authors deliberately avoid spondaic, heroic and other unusual clausulae in favour of forms of the 'artistic' four (including resolved forms), viz. cretic trochees, double cretics (or molossus cretics), double trochees, and hypodochmiacs. Latin teems with long syllables, and authors who have a markedly lower proportion of – – –x are probably avoiding it deliberately. The effects can be pervasive: in Cicero, for example, *audistis* is found 72 times, *audiuistis* 2, *audisti* 16, *audiuisti* 0.<sup>57</sup> Cicero seems to avoid the sequence of four long syllables. So too does Pliny the Younger show a marked aversion to double spondaic clausulae, which occur in his writings only around 6 per cent of the time. Authors like Tacitus, by contrast, are much less averse to double spondees, which comprise nearly a quarter of his clausulae.

In addition to double spondees, it is especially relevant to consider the frequency of heroic clausulae (that is, hexameter endings). In most authors these are not very frequent, but in certain authors, like Cicero, they are exceptionally rare.<sup>58</sup> The sum of double spondaic and heroic clausulae thus provides an approximate index for how 'artistically' rhythmic an author is; adding in the rare miscellaneous clausulae makes this measure the precise complement of the artistic four.<sup>59</sup> Authors who clearly pay attention to the canons of an artificial doctrine of 'artistic' prose rhythm include (in parentheses is given the author's percentage of artistic clausulae):

1. Cicero (e.g. 83.42 per cent in the speeches taken together)<sup>60</sup>
2. Velleius Paterculus (79.68 per cent)<sup>61</sup>
3. Seneca the Younger (e.g. 80.92 per cent in the *Epistulae morales*)<sup>62</sup>

<sup>56</sup> And this is still to say nothing of tendencies within the works: does Caesar, for example, pay more attention to 'artistic' prose rhythm in speeches? The question has not been sufficiently investigated; see, for example, Gaertner and Hausburg 2013: 71 n. 207; Börner 2016. We will discuss Sallust's and Tacitus' rhythmic tendencies in speeches vs narrative below.

<sup>57</sup> Cicero accounts for the overwhelming majority of *audi(ui)sti(s)* in classical Latin, and so comparisons with other authors are not especially fruitful.

<sup>58</sup> See, for example, Zielinski 1904: 163–6; Shipley 1911; Laurand 1911; 1936–38: 2.179–80; Adams 2013.

<sup>59</sup> Similarly Hutchinson 2018: 19 on using double spondees and their resolved forms (including the heroic clausula) as a gauge for how rhythmic an author is.

<sup>60</sup> The literature on prose rhythm in Cicero is too vast to cite here; see the references collected in Berry 1996b: 49 n. 247, to which can be added Sträterhoff 1995; Hutchinson 1995; 1998: 9–12; Oberhelman 2003; Koster 2011; Winterbottom 2011.

<sup>61</sup> See Bornecque 1907: 571–4; Aili 1979: 126–7; Oakley forthcoming.

<sup>62</sup> See especially Axelson 1933: 7–16; 1939: 23–48; earlier Bourgerly 1910 and, unhelpfully, Zander 1910–14: 2.65–121.

4. Q. Curtius Rufus (85.29 per cent)<sup>63</sup>
5. Pomponius Mela (82.62 per cent)<sup>64</sup>
6. Pliny the Younger (84.87 per cent in *Epist.* 1–9; 85.36 per cent in *Pan.*)<sup>65</sup>
7. Suetonius (80.57 per cent in the *Vitae*)<sup>66</sup>
8. Apuleius (in some works; e.g. 78.50 per cent in *Met.*)<sup>67</sup>
9. [Quintilian], *Declamationes maiores* (84.03 per cent)<sup>68</sup>

In the main, our results confirm earlier scholars' smaller, sample-based studies of individual authors; such replication and verification has long been missing in studies of prose rhythm.<sup>69</sup> So, for example, Velleius Paterculus shows a remarkable affection for double cretic and molossus cretic rhythms, which comprise some 40 per cent of his clausulae. This striking preference is unexpected, unprecedented and not imitated by later authors. Aili looked at a sample of 500 Velleian clausulae, and, although counting only six syllables and presenting his data somewhat differently, found essentially the same tendency.<sup>70</sup>

The great bulk of Latin prose authors, however, seem to have followed their own rhythmical preferences, not a set of Hellenistic precepts. To this generalisation one special case should be noted: both Sallust and especially Livy must have consciously sought out heroic and spondaic rhythms, and to an extraordinary degree (Sall., *Iug.*: 34.93 per cent, *Cat.*: 33.73 per cent; Livy: 43.99 per cent). Livy's preferences moreover intensified over time, being least marked in the first decade (35.41 per cent) but increasingly so in Books 21–30 (48.37 per cent) and 31–40 (49.72 per cent). These authors have deliberately chosen to go in precisely the opposite direction to the Ciceronian system.<sup>71</sup> Whether Livy's and Sallust's predilection for non-artistic clausulae constitutes a 'historical style' is unclear; Tacitus, at any rate, does not follow their example.<sup>72</sup>

In sum, 'expected values' for the distribution of rhythms in unmarked Latin prose simply cannot be established on the basis of surviving evidence, for all authors have their own rhythmic preferences. But there are statistically significant differences in these authorial preferences. Furthermore, an important subset of Latin authors adhered in some fashion to a particular 'artistic' rhythmic canon, and at least a couple deliberately rebelled against it. It is in this sense that we can claim that Latin prose rhythm is not just a chimera that scholarly syllable counters have been chasing after in vain for over a century.

### *Authorial Variation and 'Spurious' Compositions*

Cicero has always provided the notional benchmark against which Latin prose rhythm has been measured, but Cicero's own rhythmical practices vary widely over time and genre and

<sup>63</sup> See Müller 1954: 755–82.

<sup>64</sup> See Havet 1904; Parroni 1984.

<sup>65</sup> See Hofacker 1903; Bornecque 1907: 323–40; Whitton 2013: 28–32 and his index s.v. 'rhythm', and our comments below.

<sup>66</sup> See, for example, Macé 1900: 379–400; Bornecque 1907: 574–8; Fry 2009: 19–20, and further references in Power 2014: 76 n. 47.

<sup>67</sup> 'Apuleius' in the tables above includes works of disputed authorship (the 'preface' to *De deo Socratis*, *De mundo* and *De Platone*). For Apuleian prose rhythm generally, see Bernhard 1927; in *Metamorphoses*, Hijmans 1978; Nisbet 2001. In the philosophical works, where the accentual *cursus* mixes with quantitative rhythm, see Axelson 1987; Redfors 1960: 75–113; Stover 2016: 42–4.

<sup>68</sup> Not only is [Quintilian] not Quintilian, it is not even just one author. For a minutely detailed study of prose rhythm in the *Declamationes maiores*, see Håkanson 2014.

<sup>69</sup> Understandably so: if someone else has already spent a long time counting something, there would seem to be little earthly reward for taking a similarly long time to check the work and pronounce it sound.

<sup>70</sup> Aili 1979: 126–7.

<sup>71</sup> See especially Aili 1979: 69–130.

<sup>72</sup> It also raises the question of whether their prose rhythm is in some sense 'epic': 'historia ... proxima poetis' (Quint., *Inst.* 10.1.31)?

even an individual work. One often reads, for example, that Cicero was less attentive to prose rhythm in his correspondence. While this claim can and should be nuanced, it is clearly right, as can be seen by comparing Cicero's speeches with the *Epistulae ad Atticum*:

	ARTISTIC	NON-ARTISTIC
<i>Speeches</i>	20864 (83.42%)	4146 (16.58%)
<i>Letters to Atticus</i>	7360 (69.68%)	3203 (30.32%)

$\chi^2 = 856.038$ ,  $p$ -value  $\approx 0$ : these distributions are very different. The letters are markedly less concerned with artificially artistic prose rhythm.

Of course, not all letters are created equal.<sup>73</sup> When Cicero is writing for a wider audience, as in his long letter of advice to Quintus during the latter's time as a provincial administrator in Asia, he uses markedly different rhythms than when he writes for his brother's ears alone:

	ARTISTIC	NON-ARTISTIC
<i>Q. fr. 1.1</i>	181 (92.82%)	14 (7.18%)
<i>Q. fr., 1.1 excluded</i>	793 (72.89%)	295 (27.11%)

$\chi^2 = 35.94$ ,  $p$ -value  $\approx 0$ . The polished and public *Q. fr. 1.1* was composed with much more attention to pretty clausulae.

Furthermore, it should be observed that even within Cicero's corpus of speeches we find considerable variation. *Pro Roscio comoedo*, for example, is notably non-artistic in its rhythms, perhaps showing a 'studied negligence' in imitation of comedy.<sup>74</sup> While general trends can be described — the earliest speeches show fewer cretic trochees, say — there exist occasional counter-examples to almost all of them (so the later *Pro Rabirio Postumo* shows a very low percentage of cretic trochees). Given all this variation, can we even talk about Cicero's 'prose rhythm preferences' as some kind of Platonic form? We are sceptical.

Studies of prose rhythm often hold out the promise of uncovering an author's unique rhythmic fingerprint, a sort of unchanging stylistic essence. Such a fingerprint could be of enormous use in questions of authenticity. Some authors, as we have seen, do present a very consistent fingerprint: Caesar is consistent with Caesar; Varro is consistent with Varro. Other authors, however, are chameleons, adapting their rhythms to circumstances. Cicero is a chameleon. Such authorial variation and adaptability means that we cannot naively rely on prose rhythm to distinguish between genuine and spurious compositions.

This claim is most easily demonstrated by using our artistic vs non-artistic test for each of Cicero's speeches set against the corpus of the rest of his speeches. In effect, we are conducting a thought experiment in which we ask, 'If this work were not known to be Cicero's, would it fit rhythmically with the rest of his corpus?' Table 7 shows Cicero's surviving speeches, sorted from most to least artistically rhythmic.

The test that we have just described would identify fully twenty-two of these speeches as suspect:

- Non-artistic to a statistically significant degree (9): *Quinct.*, *Rosc. Am.*, *Caecin.*, *Tul.*, *Verr. 2.1* and *2.2*, *Q. Rosc.*, *Rab. Post.*, *Phil. 8*.
- Artistic to a statistically significant degree (13): *Leg. Man.*, *Catil. 2* and *4*, *Arch.*, *Dom.*, *Vat.*, *Prov. cons.*, *Cael.*, *Balb.*, *Planc.*, *Marcell.*, *Phil. 3* and *4*.

<sup>73</sup> See especially Hutchinson 1998: 9–12; earlier, for example, Bornecque 1907: 565–70.

<sup>74</sup> Von Albrecht 2003: 23 n. 72. For full details of Cicero's prose rhythm practices in this speech, see Axer 1980: 21–4.

TABLE 7 Cicero's speeches (ranked).

TITLE	ARTISTIC CLAUSULAE	NON-ARTISTIC CLAUSULAE	PERCENTAGE ARTISTIC	PERCENTAGE NON-ARTISTIC
<i>Pro Marcello</i>	137	8	94.48	5.52
<i>Phil. 4</i>	90	6	93.75	6.25
<i>Cat. 4</i>	128	9	93.43	6.57
<i>Pro Archia</i>	123	9	93.18	6.82
<i>Pro lege Manilia</i>	261	23	91.90	8.10
<i>Cat. 2</i>	164	15	91.62	8.38
<i>Pro Balbo</i>	294	27	91.59	8.41
<i>Phil. 9</i>	71	7	91.03	8.97
<i>De prouinciis consularibus</i>	230	24	90.55	9.45
<i>Phil. 3</i>	224	25	89.96	10.04
<i>Post reditum ad populum</i>	90	11	89.11	10.89
<i>In Vatinius</i>	169	21	88.95	11.05
<i>Pro Caelio</i>	457	57	88.91	11.09
<i>Pro Plancio</i>	562	76	88.09	11.91
<i>Phil. 14</i>	183	25	87.98	12.02
<i>Cat. 3</i>	128	18	87.67	12.33
<i>De domo sua</i>	643	93	87.36	12.64
<i>Phil. 11</i>	264	41	86.56	13.44
<i>Post reditum in senatu</i>	147	23	86.47	13.53
<i>Cat. 1</i>	164	26	86.32	13.68
<i>De haruspicum responsis</i>	346	55	86.28	13.72
<i>In Caecilium</i>	272	44	86.08	13.92
<i>Pro Murena</i>	519	85	85.93	14.07
<i>Pro rege Deiotaro</i>	212	35	85.83	14.17
<i>Pro Sulla</i>	427	72	85.57	14.43
<i>In Pisonem</i>	526	91	85.25	14.75
<i>Leg. agr. 1</i>	121	21	85.21	14.79
<i>Phil. 12</i>	232	41	84.98	15.02
<i>Phil. 10</i>	150	27	84.75	15.25
<i>Verr. II.4</i>	940	171	84.61	15.39
<i>Pro Flacco</i>	653	119	84.59	15.41
<i>Pro Rabirio perduellionis reo</i>	119	22	84.40	15.60
<i>Pro Fonteio</i>	180	34	84.11	15.89

*Continued*

Table 7 Continued

TITLE	ARTISTIC CLAUSULAE	NON-ARTISTIC CLAUSULAE	PERCENTAGE ARTISTIC	PERCENTAGE NON-ARTISTIC
<i>Phil. 6</i>	127	24	84.11	15.89
<i>Pro Milone</i>	503	96	83.97	16.03
<i>Leg. agr. 2</i>	470	90	83.93	16.07
<i>Phil. 13</i>	349	67	83.89	16.11
<i>Phil. 2</i>	787	153	83.72	16.28
<i>Pro Cluentio</i>	998	200	83.31	16.69
<i>Verr. II.5</i>	929	187	83.24	16.76
<i>Phil. 1</i>	227	46	83.15	16.85
<i>Pro Sestio</i>	714	156	82.07	17.93
<i>Phil. 7</i>	129	29	81.65	18.35
<i>Verr. II.3</i>	1329	300	81.58	18.42
<i>Phil. 5</i>	301	70	81.13	18.87
<i>Leg. agr. 3</i>	68	16	80.95	19.05
<i>Verr. I</i>	196	47	80.66	19.34
<i>Verr. II.2</i>	975	234	80.65	19.35
<i>Pro Scauro</i>	139	34	80.35	19.65
<i>Pro Ligario</i>	200	50	80.00	20.00
<i>Verr. II.1</i>	806	209	79.41	20.59
<i>Pro Caecina</i>	478	128	78.88	21.12
<i>Phil. 8</i>	206	56	78.63	21.37
<i>Pro Roscio Amerino</i>	636	173	78.62	21.38
<i>Pro Rabirio Postumo</i>	210	60	77.78	22.22
<i>Pro Quinctio</i>	447	140	76.15	23.85
<i>Pro Tullio</i>	155	56	73.46	26.54
<i>Pro Roscio comoedo</i>	259	165	61.08	38.92
[Cicero], <i>In Sallustium</i>	77	31	71.30	28.70

Now these data are not without use. We have already commented on the exceptional *Pro Roscio comoedo*, which is, rhythmically speaking, far and away Cicero's 'least Ciceronian' speech. It is probably not coincidence that most of the other less 'artistic' speeches cluster at the beginning of Cicero's career; it would not be surprising to find that his rhythmic preferences evolved and were refined over time, and any such change has been flattened out in this test. And yet *Philippic 8* is rather unexpected; Cicero's tendency towards more artistic clausulae is hardly a fixed law. On the other hand, sometimes Cicero seems to have gone out of his way to be especially 'artistic' in his rhythms. Such speeches include



some of Cicero's most important, like the *Catilinarians* (a sign of careful revision?), as well as particularly literary efforts like *Pro Archia* and *Pro Caelio*.<sup>75</sup>

But while the data are not useless, a test showing that fully 38 per cent of Cicero's speeches appear 'non-Ciceronian' is clearly not the appropriate instrument to determine authorship of a potentially Ciceronian speech.<sup>76</sup> For Cicero, prose rhythm is not just a signature of authorship; it is in fact a form of content. A too simple application of statistical tests to prose rhythm to resolve questions of authenticity risks conflating variation in content with variation in authorship.

We still think that such tests can sometimes be applied with profit, but they must be applied very carefully. They work best with authors who do not appear to vary their rhythmic practices depending on content, like Sallust. As our tables show, Sallust exhibits the same rhythmic profile in all of his historical works, and we shall soon see that he does not evince any differences between his narrative and set-piece speeches within those works either. The author of the pseudo-Sallustian *Inuectiua in Ciceronem*, on the other hand, has a markedly different set of preferences for artistic and non-artistic clausulae:

	ARTISTIC	NON-ARTISTIC
Sall., <i>Cat.</i> and <i>Iug.</i>	929 (47.16%)	1041 (52.84%)
[Sall.], <i>In Ciceronem</i>	29 (63.04%)	17 (36.94%)

$\chi^2 = 4.549$ ,  $p$ -value  $\approx 0.03293$ .<sup>77</sup> One might still try to argue that this is simply an instance of generic differences dictating different rhythms, but in any case we can say that overall propensity to artistic clausulae does not encourage belief in Sallustian authorship.<sup>78</sup> By contrast, preferences for artistic clausulae at least do not militate against the claim that Sallust wrote the *Epistulae ad Caesarem*:

	ARTISTIC	NON-ARTISTIC
Sall., <i>Cat.</i> and <i>Iug.</i>	929 (47.16%)	1041 (52.84%)
<i>Epistulae ad Caesarem</i>	115 (49.36%)	118 (50.64%)

$\chi^2 = 0.404$ ,  $p$ -value  $\approx 0.52503$ . The rhythms of the *Epistulae ad Caesarem* are indistinguishable from Sallust in his historical works; if they are not genuine, the imitator showed a remarkably accurate knowledge of Sallust's unusual rhythmic tendencies.<sup>79</sup>

<sup>75</sup> For prose rhythm in the *Pro Archia*, see Vretska and Vretska 1979.

<sup>76</sup> The only speech in the above results whose authorship has been seriously questioned is *De domo sua*. While we find that the speech is 'artistic' to a statistically significant degree (*contra* Zielinski 1904: 218–19; Nisbet 1939: xxxii–xxxiii), these would be very weak grounds to reject Ciceronian authorship in any event, and would not go along with supposed stylistic defects in other aspects of the speech. For a recent explanation of some of the apparent oddity of this speech, see Kenty 2018.

<sup>77</sup> Note that the chi-squared test statistic, while yielding a statistically significant  $p$ -value, is still relatively small here because of the small sample size of the *Inuective*; see n. 52 above.

<sup>78</sup> Few scholars believe that Sallust wrote the *Inuective*; see Novokhatko 2009: 111–29; Santangelo 2012: 29–32.

<sup>79</sup> Similarly few scholars believe that Sallust wrote the *Epistulae*, but see Posadas 2016, who does, with further bibliography on the question in his n. 2. For the other side, see Mastrorosa 2017, with comprehensive bibliography on both sides of the debate in her nn. 2–3. The extent to which later imitators perceived and replicated the prose rhythm of their models, and whether (or how) such sensitivity changed over the centuries, merits further investigation. As we will see below, Tacitus, for one, is not concerned to be especially Ciceronian in his 'Ciceronian' *Dialogus*. The pseudo-Ciceronian *In Sallustium* (of uncertain date) would be Cicero's least artistically rhythmic speech, the *Pro Roscio comoedo* excepted; the *Epistula ad Octavianum* (also of uncertain date), on the other hand, one of his most artistically rhythmic letters.

But such applications are perhaps more limited than we might want. *Rhetorica ad Herennium*, for example, is rhythmically indistinguishable from *De inuentione*, but this is not a function of Ciceronian authorship: you might guess that similarity in content is the reason that their rhythms converge. Tests using this method can measure real differences between texts, and this is of value, but such variation may be tied to any number of factors, most notably variation in content. While in certain circumstances, particularly when an author shows very stable rhythmic practices, these tests can be a piece of evidence in the discussion of authenticity, prose rhythm is very far from a panacea for resolving the attribution of a disputed work.

### *Variation Within a Text: Speeches vs Narrative in Sallust and Tacitus*

We have just seen that some authors vary their prose rhythm practices in different genres (private letters vs public speeches, say), and that indeed some authors show remarkable variation even within a single broad genre (Cicero's orations). This naturally leads to the question of whether authors show different rhythmic practices *within* an individual work. In Latin historiography, for example, is there a difference in prose rhythm between narrative and inset speeches?<sup>80</sup> We have looked at the cases of Sallust and Tacitus. For Sallust, the answer is a clear no. For Tacitus the situation is more complex: Tacitus does seem to have different rhythmic profiles, and they do sometimes correlate with the distinction between narrative and speeches — but not always.

To arrive at these answers we must first separate the historians' corpora into speeches and narrative. While it is perhaps not impossible to do this programmatically, it is a challenge,<sup>81</sup> and we have simply segregated by hand. We have included only longer instances of direct speech, excluding both short utterances and all indirect discourse.<sup>82</sup> Our corpora of speeches are as follows:<sup>83</sup>

Sall., *Cat.* 20, 33, 51, 52, 58; *Iug.* 10, 14, 31, 85, 102, 110; *Hist. Or.* Lepidus, Philippus, Cotta, Macer.

Tac., *Agr.* 30–2, 33–4; *Hist.* 1.15–16, 29–30, 37–8, 83–4; 2.47, 76–7; 3.2, 20; 4.32, 42, 58, 64–5, 73–4, 77; 5.26; *Ann.* 1.22, 28, 42–3, 58; 2.37–8, 71, 77; 3.12, 16, 46, 50; 4.8, 34–5, 37, 40; 6.6, 8; 11.24; 12.37; 13.21; 14.43–4, 53–4, 55–6; 15.2, 20; 15.22, 31.

<sup>80</sup> This question has been explored to some degree for Sallust, Livy and Tacitus (see Ullmann 1925; Aumont 1996: 383–7, including also Caesar, for whom see further n. 56 above), but most extensively for Tacitus. A summary of scholarship on Tacitean prose rhythm is provided by Hellegouarc'h 1991: 2437–45. Discussions of narrative vs speech in Tacitus include Ullmann 1925; 1931; Salvatore 1950: 143–68; Andreoni 1968; Dangel 1991: 2496–504. None of these treatments has been able to perform a consistent comparison on all the clausulae in question, leading to unreliable conclusions.

<sup>81</sup> Editors typically denote the beginning of direct speech with ' and its end with ', but they differ in how they treat a single speech that continues over multiple paragraphs (for example, repeat the ' at the beginning of each paragraph or not?), and ' is also sometimes used for other purposes (for example, M.' = Manius). This is one of many cases where a corpus marked up with metadata would prove useful; see our remarks in conclusion.

<sup>82</sup> Our reasons for not considering indirect discourse separately from narrative and direct speech are strictly pragmatic: it is much harder to find and segregate such instances of reported speech. Their rhythms thus remain an open question, but note that if they agree with the rhythms of direct speech, then all the rhythmic differences between speech and narrative found here will be magnified.

<sup>83</sup> We have listed the section or section range where the speech is found, but we have only included in our corpus the portion of that section which contains direct speech. Our corpora are similar but not identical to those of Ullmann 1925: 67, 72; 1931: 72; Andreoni 1968: 304–5.

TABLE 8 Sallust and Tacitus, speeches vs narrative.

AUTHOR AND WORK	TOTAL CLAUSULAE	TOTAL CLAUSULAE EXCLUDED	TOTAL CLAUSULAE CONSIDERED	PERCENTAGE CRETIC-TROCHEE	PERCENTAGE DOUBLE CRETIC (OR MOLOSSUS CRETIC)	PERCENTAGE DOUBLE TROCHEE	PERCENTAGE HYPODOCHMIAC	PERCENTAGE SPONDAIC	PERCENTAGE HEROIC	PERCENTAGE 'ARTISTIC'	PERCENTAGE SPONDAIC + HEROIC	PERCENTAGE OTHER
<b>Sallust</b>												
<i>Bellum Iugurthinum</i> (narrative)	1077	27	1050	10.57	22.38	9.14	4.95	25.43	9.62	47.05	35.05	17.90
<i>Bellum Iugurthinum</i> (speeches)	242	3	239	10.46	20.92	7.53	4.60	23.85	10.46	43.51	34.31	22.18
<i>Bellum Catilinae</i> : narrative	490	19	471	9.98	17.83	15.71	4.88	21.02	12.31	48.41	33.33	18.26
<i>Bellum Catilinae</i> : speeches	209	2	207	8.70	20.29	11.59	8.70	25.12	9.18	49.28	34.30	16.43
<i>Historiae</i> : speeches	186	9	177	10.17	23.73	9.60	1.69	25.99	7.34	45.20	33.33	21.47
<b>Tacitus</b>												
<i>Dialogus</i>	443	31	412	23.06	18.69	23.30	5.10	17.96	2.67	70.15	20.63	9.22
<i>Agricola</i> : narrative	380	13	367	14.99	17.98	18.26	7.63	23.16	5.45	58.86	28.61	12.53
<i>Agricola</i> : speeches	63	3	60	23.33	13.33	25.00	5.00	15.00	1.67	66.67	16.67	16.67
<i>Germania</i>	460	5	455	22.20	18.02	20.00	6.81	17.14	5.49	67.03	22.64	10.33
<i>Historiae</i> : narrative	3465	35	3430	17.70	16.09	16.59	5.71	24.81	6.27	56.09	31.08	12.83
<i>Historiae</i> : speeches	259	2	257	20.62	21.40	21.01	5.45	14.40	5.45	68.48	19.84	11.67
<i>Annales</i> : narrative	5467	115	5352	15.73	17.10	17.02	6.17	23.75	6.43	56.02	30.18	13.81
<i>Annales</i> : speeches	314	7	307	13.68	20.52	18.89	6.51	19.87	6.19	59.61	26.06	14.33

For Sallust the results are plain.<sup>84</sup> For example, in the *Bellum Iugurthinum*:

	ARTISTIC	NON-ARTISTIC
<i>Iug.</i> : narrative	494 (47.05%)	556 (52.95%)
<i>Iug.</i> : speeches	104 (43.51%)	135 (56.49%)

$\chi^2 = 0.977$ ,  $p$ -value  $\approx 0.32294$ . The *Bellum Catilinae* shows an even greater similarity:

	ARTISTIC	NON-ARTISTIC
<i>Cat.</i> : narrative	228 (48.40%)	243 (51.60%)
<i>Cat.</i> : speeches	102 (49.28%)	105 (51.72%)

$\chi^2 = 0.043$ ,  $p$ -value  $\approx 0.83573$ . Even the longer speeches of Sallust's *Historiae* seem to fit this pattern. We here compare them with the *Bellum Iugurthinum* and *Bellum Catilinae*, because the fragmentary state of the remainder of the *Historiae* makes any inferences drawn against them unreliable at best:

	ARTISTIC	NON-ARTISTIC
<i>Iug.</i> and <i>Cat.</i> : narrative	722 (47.47%)	799 (52.53%)
<i>Hist.</i> : speeches	80 (45.20%)	97 (54.80%)

$\chi^2 = 0.328$ ,  $p$ -value  $\approx 0.56684$ . Sallust shows an apparently unshakable consistency in his preferences for artistic and non-artistic clausulae, both across his various works and within them, making no distinctions between speeches and narrative.

For Tacitus the story is more nuanced. In the *Annales*, he shows a slight tendency toward more artistic clausulae in speeches, but it is slight and not statistically significant:

	ARTISTIC	NON-ARTISTIC
<i>Ann.</i> : narrative	2998 (56.02%)	2354 (43.98%)
<i>Ann.</i> : speeches	183 (59.61%)	124 (40.39%)

$\chi^2 = 1.523$ ,  $p$ -value  $\approx 0.21717$ . In his last work, it appears that Tacitus did not differentiate speeches from narrative rhythmically, or at any rate that any differentiation is so small that it may well have arisen by chance.

But in his earlier works the tendency toward artistic clausulae in speeches is more pronounced. So in the *Agricola*:

	ARTISTIC	NON-ARTISTIC
<i>Agr.</i> : narrative	216 (58.86%)	151 (41.14%)
<i>Agr.</i> : speeches	40 (66.67%)	20 (33.33%)

$\chi^2 = 1.31$ ,  $p$ -value  $\approx 0.25239$ . The chi-square test statistic here is small both because the difference in the proportion of artistic clausulae is not large and, importantly, because the sample size of speeches in the *Agricola* is so small. But these proportions are very nearly

<sup>84</sup> Full data for both Sallust's and Tacitus' speech and narrative prose rhythms are available in the Supplementary Material online.

what we see in the *Historiae*, where the larger sample size allows for more statistical confidence:

	ARTISTIC	NON-ARTISTIC
<i>Hist.</i> : narrative	1924 (56.09%)	1506 (43.91%)
<i>Hist.</i> : speeches	176 (68.48%)	81 (31.52%)

$\chi^2 = 14.969$ ,  $p$ -value  $\approx 0.00011$ . The difference between speech and narrative here is large and statistically significant. The narrative portion of the *Historiae* shows almost the exact same propensity to artistic clausulae as the narrative of the *Annales* (and the *Agricola*). The speeches of the *Historiae*, however, resemble nothing so much as the *Dialogus* and *Germania*, from which they are indistinguishable in their preferences for artistic clausulae.

	ARTISTIC	NON-ARTISTIC
<i>Dial.</i>	289 (70.15%)	123 (29.85%)
<i>Hist.</i> : speeches	176 (68.48%)	81 (31.52%)

$\chi^2 = 0.207$ ,  $p$ -value  $\approx 0.64913$ .

	ARTISTIC	NON-ARTISTIC
<i>Germ.</i>	305 (67.03%)	150 (32.97%)
<i>Hist.</i> : speeches	176 (68.48%)	81 (31.52%)

$\chi^2 = 0.157$ ,  $p$ -value  $\approx 0.69193$ .

What do all these numbers mean? They seem to indicate that while Sallust has a uniformly consistent set of (dis)preferences for artistic clausulae, Tacitus has at least two separate rhythmic profiles that he can use. These two separate profiles sometimes correlate with the distinction between speech and narrative (so in the *Dialogus*, *Agricola* and *Historiae*), but not always: in the *Annales*, Tacitus shows roughly the same proportion of artistic clausulae in both speech and narrative, and in the *Germania*, which is exclusively narrative, Tacitus exhibits the rhythmic preferences that he shows elsewhere for speeches. More investigation is needed here, but it is plain that prose rhythm is part of Tacitus' literary artistry, and that he sometimes varies his practice for some kind of effect. It would certainly be a mistake to claim, as many scholars have, that Tacitus is indifferent to prose rhythm.<sup>85</sup>

### Tacitus, *Dialogus de oratoribus*

We have just seen that Tacitus makes use of a particular rhythmic profile in the *Dialogus de oratoribus*. Now in that work he imitates Cicero in numerous and varied points of diction. He postpones *igitur* to second position; he uses the word *autem* some twenty times (compared to six instances in all of the *Historiae* and *Annales*); he indulges in a number of synonymous doublets.<sup>86</sup> One might wonder whether his rhythmic preferences in the *Dialogus* are a sought-out imitation of Cicero too, as Gregory Hutchinson claims.<sup>87</sup>

<sup>85</sup> Starting with Norden 1918 [1898]: 2.942: 'Dagegen [sc. in contrast to Pliny the Younger] ignoriert Tacitus ... den Rhythmus der Klausel durchaus.' Further references in Aili 1979: 128–9; Hellegouarc'h 1991: 2445; Dangel 1991: 2496.

<sup>86</sup> For Cicero and Ciceronianisms in the *Dialogus*, see van den Berg 2014: esp. 208–40; Keeline 2018: 223–76, neither considering prose rhythm.

<sup>87</sup> Hutchinson 2018: 9.

It is in some sense true that the *Dialogus* is Tacitus' 'least Tacitean' work in its propensity to artistic clausulae. A test of its numbers of artistic and non-artistic clausulae against those of the rest of Tacitus' corpus marks it as a clear outlier:

	ARTISTIC	NON-ARTISTIC
<i>Dialogus</i>	289 (70.15%)	123 (28.85%)
Tacitus ( <i>Dial.</i> excluded)	5846 (57.13%)	4387 (42.87%)

$\chi^2 = 27.483$ ,  $p$ -value  $\approx 0$ . But as we have already seen, that is only part of the story. The *Germania* too, for example, shows the same rhythmic profile, as do the speeches in the *Agricola* and the *Historiae*.

Moreover, this propensity to artistic clausulae is not necessarily 'Ciceronian'. The best point of comparison between the *Dialogus* and 'Cicero' is not completely clear. Does the *Dialogus* map onto the prose rhythm of Cicero's speeches?

	ARTISTIC	NON-ARTISTIC
Cicero's <i>Speeches</i>	20864 (83.42%)	4146 (16.58%)
Tac., <i>Dialogus</i>	289 (70.15%)	123 (29.85%)

$\chi^2 = 51.135$ ,  $p$ -value  $\approx 0$ . No, it is not even close. What about Cicero's own dialogues? Here it is hard to know what corpus to pick, but the *Dialogus* is less artistically rhythmic than any of Cicero's surviving dialogues. If we compare it, for example, with all of Cicero's extant rhetorical and philosophical works pooled together, we get:

	ARTISTIC	NON-ARTISTIC
Cicero's <i>Rhetorica</i> and <i>Philosophica</i>	22183 (84.42%)	4094 (15.58%)
Tac., <i>Dialogus</i>	289 (70.15%)	123 (29.85%)

$\chi^2 = 62.125$ ,  $p$ -value  $\approx 0$ . Again, not even close; even further away, in fact.

The rhythms of the *Dialogus* are clearly different from the narrative portions of Tacitus' historical works, but they resemble the *Germania* and the speeches of the *Agricola* and the *Historiae*. What Tacitus is doing with this varying propensity toward artistic clausulae calls for further study, but we can say with confidence that neither in the *Dialogus* nor anywhere else does he even approach a true rhythmic imitation of Cicero.

### *Pliny the Younger*

Pliny the Younger offers an interesting test case for a variety of questions, not least because he, like Sallust, presents such a consistent set of rhythmic preferences. We can thus use our statistical tests to answer questions such as: do Pliny's private letters (*Ep.* 1–9) differ from his correspondence with Trajan (*Ep.* 10)? Is there any variation within the books of private correspondence? Does Trajan's prose rhythm in *Ep.* 10 differ from Pliny's? And what of the rhythms of the *Panegyricus*, an epideictic speech perhaps liable to entirely different generic conventions from a book of stylish letters?

In the first instance we can observe that Pliny is an author with a marked preference for artistic rhythms. He shuns spondaic and heroic clausulae (even more than Cicero did in his speeches, to say nothing of his letters), and he favours cretic-trochaic rhythms to an almost unprecedented degree and with remarkable consistency across the private correspondence: they comprise some 40 per cent of the clausulae in *Ep.* 1–9.<sup>88</sup> These preferences combine to

<sup>88</sup> The figures in Whitton 2013: 29, reporting 29 per cent cretic trochees, do not seem to include resolutions. Pliny's only predecessor to show such a love for cretic-trochaic rhythms is Quintus Curtius.

yield an extraordinarily stable rhythmic profile across the private letters. Indeed, those similarities extend even to the *Panegyricus*. Consider a detailed chi-square test of the sort that showed different distributions for Varro's two works:

	CRETIC TROCHEES	DOUBLE CRETICS	DOUBLE TROCHEES	HYPODOCHMIACS	OTHER
<i>Ep.</i> 1–9	1787 (40.84%)	954 (21.80%)	799 (18.26%)	174 (3.98%)	662 (15.13%)
<i>Panegyricus</i>	494 (39.30%)	280 (22.28%)	253 (20.13%)	46 (3.66%)	184 (14.64%)

$\chi^2 = 5.902$ ,  $p$ -value  $\approx 0.20658$ . The *Panegyricus*, even on a very fine-grained test, cannot be distinguished from the letters, and the individual books of letters are themselves all but indistinguishable from each other.<sup>89</sup>

The exception, of course, is Book 10. Trajan's replies show a clearly different rhythmic fingerprint. If we compare the pooled artistic and non-artistic patterns in *Ep.* 1–9 with Trajan's replies to Pliny in Book 10, the latter are conspicuously less artistic:

	ARTISTIC	NON-ARTISTIC
<i>Ep.</i> 1–9	3714 (84.87%)	662 (15.13%)
<i>Ep.</i> 10: Trajan	102 (67.55%)	49 (32.45%)

$\chi^2 = 33.083$ ,  $p$ -value  $\approx 0$ . Trajan's rhythms in Book 10 are completely different from Pliny's in Books 1–9. Indeed, Trajan's rhythms in Book 10 are completely different from Pliny's in Book 10:

	ARTISTIC	NON-ARTISTIC
<i>Ep.</i> 10: Pliny	320 (80.60%)	77 (19.40%)
<i>Ep.</i> 10: Trajan	102 (67.55%)	49 (32.45%)

$\chi^2 \approx 10.53$ ,  $p$ -value  $\approx 0.00117$ . Trajan (or his chancery secretary) speaks in his own voice and with his own cadences.

The prose rhythm of Pliny's own letters in Book 10 is only slightly less 'artistic' than that of Books 1–9, although the difference does rise to statistical significance:

	ARTISTIC	NON-ARTISTIC
<i>Ep.</i> 1–9	3714 (84.87%)	662 (15.13%)
<i>Ep.</i> 10: Pliny	320 (80.60%)	77 (19.40%)

$\chi^2 = 5.066$ ,  $p$ -value  $\approx 0.02439$ . Nevertheless, prose rhythm appears to have been a natural part of Pliny's composition process in a way that it was not for Cicero in his letters, although it must still be a learned part, because his preferences are so distinctive — or, just maybe, he revised Book 10 for publication himself and took some care for its rhythmic properties.<sup>90</sup>

<sup>89</sup> In a comparison of all five rhythmic categories, only Book 3 stands out slightly, where Pliny has a particular preference for double trochees and lower than usual affection for cretic trochees and double cretics. This difference disappears, however, in a pooled comparison of artistic vs non-artistic categories. Interestingly, in the latter comparison it is Book 9 that looks slightly unusual, because it is overall a bit less artistically rhythmic, and yet when comparing all five categories it looks normal.

<sup>90</sup> To us this hypothesis seems unlikely (see especially Coleman 2012), but it is currently in vogue: see, for example, Gibson and Morello 2012: 259–64; Woolf 2015, with further references.

Finally, as we have already seen, although the *Panegyricus* is a speech, in it Pliny uses almost exactly the same rhythmical patterns as he does in the *Epistulae*. But to think of the rhythmic preferences of the *Panegyricus* as the same as those of the *Epistulae* is probably to put the cart before the horse. In his own lifetime, Pliny was above all an orator, and it is a simple twist of fate that we happen to have ten books of Pliny's letters and only one preserved speech. It seems very likely that the prose rhythms we find in his letters have their origin in the preferences that he developed for his speeches. This is probably a deliberate (and artistic) affectation, since one might have expected his correspondence, like Cicero's, to be looser about such details, and it is another reason we should consider Pliny's letters highly polished literary compositions.

#### IV CONCLUSIONS

Our algorithms and the data that they generate provide a powerful tool to answer questions like the ones posed above, a list which can be extended indefinitely. Because we are using computers and code, we can change assumptions or look at different texts or divide our existing texts up differently — and immediately generate refreshed data for the entirety of the corpus that we are considering. Furthermore, although it is in most cases impossible to replicate previous scholars' methodologies with absolute precision, in broad outline we can nevertheless check their results almost instantaneously. This process of replication and verification has long been absent from studies of Latin prose rhythm. Since all our code and data are open source and publicly available, our own results can also be easily checked (and perhaps improved).

Improvements and extensions of these data may take a variety of forms. A different approach to locating clausulae, one that does not rely on punctuation, might help advance exploration of 'internal' clausulae, a topic which has thus far resisted rigorous analysis. More extensively marked up texts would facilitate other kinds of investigations: for example, does Cicero use different rhythms in his *exordia*, or *narrationes* or *perorationes*? Annotating his speeches with consistent metadata would allow for more detailed study. More sophisticated data manipulation techniques, like Principal Component Analysis, might give us other profitable ways to categorise our data beyond just 'artistic' and 'non-artistic'.<sup>91</sup> And this is to say nothing of further work that can be done with the data that we have already collected, like that on word division and word accent in clausulae, which would necessarily be crucial in studying the rhythms of late antique texts as the *cursus* begins to develop.

Of course, none of the broad brush pictures painted by statistical analysis can give insights at the level of an individual clausula in an individual sentence in an individual author's text. Such an analysis of the details of prose rhythm in the context of a speech or a letter is eminently worthwhile and can have great explanatory power.<sup>92</sup> So when Cicero describes the same event twice in almost the same words in *Pro Milone*, he once writes 'respondit triduo illum aut summum quadriduo esse periturum' (*Mil.* 26), but later 'audistis ... periturum Milonem triduo' (*Mil.* 44). It seems likely that he wrote *esse periturum* in the first case because it was in clausular position (= *esse uideatur*), whereas in the second the infinitive came in the middle of the phrase and so he preferred simply *periturum*. Prose rhythm is one of the keys to unlocking the secrets of Latin word order and word choice, revealing points of emphasis and rhetorical artifice, and understanding

<sup>91</sup> In this paper we consciously chose to group all possible clausulae into seven patterns; we then sub-divided those seven into 'artistic' and 'non-artistic'. Principal Component Analysis, by contrast, is a data reduction technique that would ignore ancient and modern prose rhythm classifications and instead seek algorithmically to group together clausular patterns into the 'principal components' (whatever those may be) that best account for the observed variance between samples: see Jolliffe 2002.

<sup>92</sup> See, for example, Vretska and Vretska 1979; Hutchinson 1995; 1998: 9–12; 2015; 2018; Riggsby 2010.



it at the local level is essential for appreciating an author's verbal artistry. Much of this artistry must have been put into practice subconsciously or unconsciously (see, for example, Quint., *Inst.* 9.4.119–20), and we remain sceptical of accounts that attempt to quantify the force of any individual clausula, but it is clear that ancient authors and ancient audiences could perceive and appreciate rhythmic prose.<sup>93</sup> Today, without native speaker *Sprachgefühl*, we can only recover these effects by philological analysis.

While interpreting prose rhythm at the level of the sentence and clause requires close reading and analysis, at the global level, questions of prose rhythm cry out for an open-source, Big Data approach. We have offered one such approach, producing algorithms to detect and categorise the rhythms of any Latin prose text, providing comprehensive data generated by these algorithms for most of extant classical Latin prose, presenting a new statistical approach to analysing the significance of those data, and giving several examples of how to use our data and procedures to answer particular questions about authors' propensity toward artistic rhythms. For example, we can confirm that Cicero's letters are significantly less concerned with 'artistic' prose rhythm than are his speeches, but we can also show how certain letters, like the lengthy and polished *Q. fr.* 1.1, take particular care to be artistically rhythmical. We can with a few clicks compare the prose rhythms of the perhaps spurious *Inuictiua in Ciceronem* or *Epistulae ad Caesarem senem* with those of the undisputedly genuine Sallust: the former does not look at all Sallustian, but the latter actually does. We can compare the rhythms of speeches and narrative in authors like Sallust and Tacitus: Sallust's rhythms never change, but Tacitus has at least two distinct rhythmic profiles (neither of which, even in the *Dialogus*, counts as 'Ciceronian'). We can see almost at a glance that Trajan's replies to Pliny's letters in Book 10 have an entirely different rhythmic fingerprint from Pliny's, while in the *Panegyricus* Pliny mirrors the rhythmic preferences that he shows in the *Epistulae*. It may be an exaggeration to claim that technology will revolutionise the study of Latin prose rhythm — the fundamental insights as worked out over a century ago seem to stand correct and confirmed — but it will certainly replough the entire field, offering fresh data and the possibility of countless new results. Nothing will ever make the study of Latin prose rhythm easy, but computers will certainly make it a lot easier.

Department of Classics, Washington University in St Louis  
[keeline@wustl.edu](mailto:keeline@wustl.edu)

Independent scholar  
[tyler.kirby9398@gmail.com](mailto:tyler.kirby9398@gmail.com)

#### SUPPLEMENTARY MATERIAL

For Supplementary Material for this article please visit <https://doi.org/10.1017/S0075435819000881>.

#### BIBLIOGRAPHY

- Adams, E. D. 2013: *Esse videtur: Occurrences of the Heroic Clausulae in Cicero's Orations*, MA thesis, University of Kansas.  
 Aili, H. 1979: *The Prose Rhythm of Sallust and Livy*, Stockholm.

<sup>93</sup> For an attempt at measuring the *Schlußwert* of individual clausulae, see Primmer 1968 with the critique of Aili 1979: 25–32. On ancient audiences' perception of prose rhythm, from the Greek side, see Vatri *forthcoming*.

- Allen, W. S. 1978: *Vox Latina: A Guide to the Pronunciation of Classical Latin* (2nd edn), Cambridge.
- Andreoni, E. 1968: 'Le clauseole nei discorsi dell'*Agricola*, delle *Historiae* e degli *Annales*', *Rivista di cultura classica e medioevale* 10, 299–320.
- Aumont, J. 1996: *Métrique et stylistique des clausules dans la prose latine. De Cicéron à Pline le Jeune et de César à Florus*, Paris.
- Axelson, B. 1933: *Senecastudien. Kritische Bemerkungen zu Senecas Naturales quaestiones*, Lund.
- Axelson, B. 1939: *Neue Senecastudien. Textkritische Beiträge zu Senecas Epistulae morales*, Lund.
- Axelson, B. 1987: 'Akzentuierender Klauselrhythmus bei Apuleius. Bemerkungen zu den Schriften *De Platone* und *De mundo*', in B. Axelson (A. Önnersfors and C. Schaar (eds)), *Kleine Schriften zur lateinischen Philologie*, Stockholm, 233–45. (Reprint of Vetenskapssocieteten i Lund, *Årsbok* 1952, 3–20.)
- Axer, J. 1980: *The Style and Composition of Cicero's Speech Pro Q. Roscio comoedo: Origin and Function*, Warsaw.
- Bernhard, M. 1927: *Der Stil des Apuleius von Madaura. Ein Beitrag zur Stilistik des Spätlateins*, Stuttgart.
- Berry, D. H. 1996a: 'The value of prose rhythm in questions of authenticity: the case of *De optimo genere oratorum* attributed to Cicero', *Papers of the Leeds International Latin Seminar* 9, 47–74.
- Berry, D. H. 1996b: *Cicero: Pro P. Sulla oratio*, Cambridge.
- Bornecque, H. 1907: *Les clausules métriques latines*, Lille.
- Börner, K. 2016: 'Klauselrhythmus in den direkten Reden des *Corpus Caesarianum*', *Acta Antiqua Academiae Scientiarum Hungaricae* 56, 81–92.
- Bourgery, A. 1910: 'Sur la prose métrique de Sénèque le philosophe', *Revue de philologie* 34, 167–72.
- Broadhead, H. D. 1922: *Latin Prose Rhythm: A New Method of Investigation*, Cambridge.
- Butterfield, D. J. 2008: 'Sigmatic ecthlypsis in Lucretius', *Hermes* 136, 188–205.
- Clark, A. C. 1905: 'Zielinski's *Clauselgesetz*', *Classical Review* 19, 164–72 (review of Zielinski 1904).
- Clark, A. C. 1909: *Fontes prosae numerosae*, Oxford.
- Coleman, K. M. 2012: 'Bureaucratic language in the correspondence between Pliny and Trajan', *Transactions of the American Philological Association* 142, 189–238.
- Cser, A. 2012: 'Resyllabification and metre: the issue of *s impurum* revisited', *Acta Antiqua Academiae Scientiarum Hungaricae* 52, 363–73.
- Dangel, J. 1991: 'Les structures de la phrase oratoire chez Tacite. Étude syntaxique, rythmique et métrique', *Aufstieg und Niedergang der römischen Welt* 2.33.4, 2454–538.
- De Groot, A. W. 1921: *Der antike Prosarhythmus. Zugleich Fortsetzung des Handbook of Antique Prose-Rhythm*, Groningen.
- De Groot, A. W. 1926: *La prose métrique des anciens*, Paris.
- Fortson, B. W. 2011: 'Latin prosody and metrics', in J. Clackson (ed.), *A Companion to the Latin Language*, Malden, MA, 92–104.
- Fraenkel, E. 1968: *Leseproben aus Reden Ciceros und Catos*, Rome.
- Fry, C. 2009: 'De *Tranquilli elocutione*. Suétone en utilisateur de sa langue', in R. Poignault (ed.), *Présence de Suétone. Actes du colloque tenu à Clermont-Ferrand, 25–27 novembre 2004*, Tours, 15–29.
- Gaertner, J. F. and Hausburg, B. 2013: *Caesar and the Bellum Alexandrinum: An Analysis of Style, Narrative Technique, and the Reception of Greek Historiography*, Göttingen.
- Gibson, R. K. and Morello, R. 2012: *Reading the Letters of Pliny the Younger: An Introduction*, Cambridge and New York.
- Gotoff, H. C. 1981: 'The prose rhythm of Sallust and Livy', *Classical Philology* 76, 335–40 (review of Aili 1979).
- Habinek, T. N. 1985: *The Colometry of Latin Prose*, Berkeley.
- Håkanson, L. 2014: 'Der Satzrhythmus der 19 Größeren Deklamationen und des Calpurnius Flaccus', in L. Håkanson (B. Santorelli (ed.)), *Unveröffentlichte Schriften, Vol. 1. Studien zu den pseudoquintilianischen Declamationes maiores*, Berlin, 47–130.
- Havet, L. 1904: 'La prose de Pomponius Méla', *Revue de philologie* 28, 57–9.
- Hellegouarc'h, J. 1991: 'Le style de Tacite. Bilan et perspectives', *Aufstieg und Niedergang der römischen Welt* 2.33.4, 2385–453.
- Hijmans, B. L. 1978: 'Asinus numerosus', in B. L. Hijmans and R. T. van der Paardt (eds), *Aspects of Apuleius' Golden Ass: A Collection of Original Papers*, Groningen, 189–209.

- Hofacker, C. 1903: *De clausulis C. Caecili Plini Secundi*, Bonn.
- Hutchinson, G. O. 1995: 'Rhythm, style, and meaning in Cicero's prose', *Classical Quarterly* 45, 485–99.
- Hutchinson, G. O. 1998: *Cicero's Correspondence: A Literary Study*, Oxford.
- Hutchinson, G. O. 2013: *Greek to Latin: Frameworks and Contexts for Intertextuality*, Oxford.
- Hutchinson, G. O. 2015: 'Appian the artist: rhythmic prose and its literary implications', *Classical Quarterly* 65, 788–806.
- Hutchinson, G. O. 2018: *Plutarch's Rhythmic Prose*, Oxford.
- Janson, T. 1975: *Prose Rhythm in Medieval Latin from the 9th to the 13th Century*, Stockholm.
- Johnson, K. et al. 2014–: *CLTK: The Classical Language Toolkit*, <https://github.com/cltk/cltk>, DOI 10.5281/zenodo.593336.
- Jolliffe, I. T. 2002: *Principal Component Analysis* (2nd edn), New York.
- Keeline, T. J. 2018: *The Reception of Cicero in the Early Roman Empire: The Rhetorical Schoolroom and the Creation of a Cultural Legend*, Cambridge.
- Kenty, J. 2018: 'The political context of Cicero's *De domo sua*', *Ciceroniana online* 2, 245–64.
- Koster, S. 2011: *Ciceros Rosciana Amerina. Im Prosarhythmus rekonstruiert*, Stuttgart.
- Laurand, L. 1911: 'Les fins d'hexamètre dans les discours de Cicéron', *Revue de philologie* 35, 75–88.
- Laurand, L. 1936–38: *Études sur le style des discours de Cicéron. Avec une esquisse de l'histoire du 'cursus'* (4th edn, 3 vols), Paris.
- Macé, A. 1900: *Essai sur Suétone*, Paris.
- Mastrorosa, I. G. 2017: 'Les épîtres à César du Pseudo-Salluste. Des conseils pour gouverner dans l'antiquité tardive?' in É. Gavoille and F. Guillaumont (eds), *Conseiller, diriger par lettre*, Tours, 155–72.
- McCabe, D. F. 1981: *The Prose-Rhythm of Demosthenes*, New York.
- Müller, K. 1954: *Geschichte Alexanders des Grossen. Lateinisch und deutsch*, Munich.
- Müller, K. 1983: *Petronius Satyrice: Schelmenszenen. Lateinisch – deutsch*, Munich.
- Nisbet, R. G. 1939: *M. Tulli Ciceronis De domo sua ad pontifices oratio*, Oxford.
- Nisbet, R. G. M. 1990: 'Cola and clausulae in Cicero's speeches', in E. M. Craik (ed.), *Owls to Athens: Essays on Classical Subjects Presented to Sir Kenneth Dover*, Oxford, 349–59. (Reprinted in R. G. M. Nisbet (S. J. Harrison (ed.)), *Collected Papers on Latin Literature*, Oxford, 1995, 312–24.)
- Nisbet, R. G. M. 2001: 'Cola and clausulae in Apuleius' *Metamorphoses* 1.1', in A. Kahane and A. Laird (eds), *A Companion to the Prologue of Apuleius' Metamorphoses*, Oxford, 16–26.
- Norden, E. 1918: *Die antike Kunstprosa vom VI. Jahrhundert v. Chr. bis in die Zeit der Renaissance* (3rd edn, 2 vols), Leipzig.
- Novokhatko, A. A. 2009: *The Invectives of Sallust and Cicero*, Berlin.
- Novotný, F. 1926: 'Le problème des clausules dans la prose latine', *Revue des études latines* 4, 221–9.
- Novotný, F. 1929: *État actuel des études sur le rythme de la prose latine*, Lwów.
- Oakley, S. P. forthcoming: 'Point and periodicity: the style of Velleius Paterculus and other Latin historians writing in the early Principate'.
- Oberhelman, S. M. 2003: *Prose Rhythm in Latin Literature of the Roman Empire: First Century B.C. to Fourth Century A.D.*, Lewiston, NY.
- Orlandi, G. 2005: 'Metrical and rhythmical clausulae in medieval Latin prose: some aspects and problems', in T. Reinhardt, M. Lapidge and J. N. Adams (eds), *Aspects of the Language of Latin Prose*, Oxford, 395–412.
- Parroni, P. 1984: *Pomponii Melae De chorographia libri tres*, Rome.
- Posadas, J. L. 2016: 'Los consejos de Salustio a César antes de la guerra civil', *Florentia Iliberritana* 27, 195–205.
- Power, T. 2014. 'The endings of Suetonius' *Caesars*', in T. Power and R. K. Gibson (eds), *Suetonius the Biographer: Studies in Roman Lives*, Oxford, 58–77.
- Preacher, K. J. 2001: 'Calculation for the chi-square test: an interactive calculation tool for chi-square tests of goodness of fit and independence', <http://www.quantpsy.org/chisq/chisq.htm>.
- Primmer, A. 1968: *Cicero numerosus. Studien zum antiken Prosarhythmus*, Vienna.
- Redfors, J. 1960: *Echtheitskritische Untersuchung der apuleischen Schriften De Platone und De mundo*, Lund.
- Riggsby, A. 2010: 'Form as global strategy in Cicero's *Second Catilinarian*', in D. H. Berry and A. Erskine (eds), *Form and Function in Roman Oratory*, Cambridge, 92–104.

- Salvatore, A. 1950: *Stile e ritmo in Tacito*, Naples.
- Santangelo, F. 2012: 'Authoritative forgeries: late republican history re-told in Pseudo-Sallust', *Histos* 6, 27–51.
- Schmid, W. 1959: *Über die klassische Theorie und Praxis des antiken Prosarhythmus*, Wiesbaden.
- Shewring, W. H. 1930: 'Prose-rhythm and the comparative method', *Classical Quarterly* 24, 164–73.
- Shipley, F. W. 1911: 'The heroic clausula in Cicero and Quintilian', *Classical Philology* 6, 410–18.
- Spinacce, L. 2014: "'Cursus in clausula'", an online analysis tool of Latin prose', in F. Tomasi, R. Rosselli del Turco and A. M. Tammaro (eds), *Proceedings of the Third AIUCD Annual Conference on Humanities and their Methods in the Digital Ecosystem*, <https://dl.acm.org/citation.cfm?id=2802635>, New York.
- Srebrny, S. 2013: 'Tadeusz Zieliński (1859–1944)', *Eos* C (fasciculus extra ordinem editus electronicus), 118–63 ([http://eos.uni.wroc.pl/wp-content/uploads/2016/07/118\\_EOS-2013-CD.pdf](http://eos.uni.wroc.pl/wp-content/uploads/2016/07/118_EOS-2013-CD.pdf)).
- Stover, J. A. 2016: *A New Work by Apuleius: The Lost Third Book of the De Platone*, Oxford.
- Sträterhoff, B. 1995: *Kolometrie und Prosarhythmus bei Cicero und Livius*. De imperio Cn. Pompei und Livius I, 1–26, 8 kolometrisch ediert, kommentiert und statistisch analysiert (2 vols), Oelde.
- Stroh, W. 1990: 'Arsis und Thesis, oder: Wie hat man lateinische Verse gesprochen?' in M. von Albrecht and W. Schubert (eds), *Musik und Dichtung. Neue Forschungsberichte, Viktor Pöschl zum 80. Geburtstag gewidmet*, Frankfurt, 87–116. (Reprinted in W. Stroh (J. Leonhardt and G. Ott (eds)), *Apocrypha. Entlegene Schriften*, Stuttgart, 2000, 193–216.)
- Taylor, D. J. 1991: 'Latin declensions and conjugations: from Varro to Priscian', *Histoire Épistémologie Langage* 13, 85–109.
- Ullmann, R. 1925: 'Les clausules dans les discours de Salluste, Tite Live et Tacite', *Symbolae Osloenses* 3, 65–75.
- Ullmann, R. 1931: 'Les clausules dans les discours de Tacite', in H. Holst and H. Mørland (eds), *Serta Rudbergiana*, Oslo, 72–9.
- van den Berg, C. S. 2014: *The World of Tacitus' Dialogus de Oratoribus: Aesthetics and Empire in Ancient Rome*, Cambridge.
- Vatri, A. forthcoming: 'The nature and perception of Attic prose rhythm', *Classical Philology*.
- von Albrecht, M. 2003: *Cicero's Style: A Synopsis*, Leiden.
- Vretska, H. and Vretska, K. 1979: *Marcus Tullius Cicero Pro Archia Poeta. Ein Zeugnis für den Kampf des Geistes um seine Anerkennung*, Darmstadt.
- Whitton, C. 2013: *Pliny the Younger: Epistles Book II*, Cambridge.
- Wilkinson, L. P. 1963: *Golden Latin Artistry*, Cambridge.
- Winge, J. 2015: *Automatic Annotation of Latin Vowel Length*, BA thesis, Uppsala University.
- Winterbottom, M. 2011: 'On ancient prose rhythm: the story of the Dichoreus', in D. Obbink and R. Rutherford (eds), *Culture in Pieces: Essays on Ancient Texts in Honour of Peter Parsons*, Oxford, 262–76.
- Woolf, G. 2015: 'Pliny/Trajan and the poetics of empire', *Classical Philology* 110, 132–51.
- Zander, C. 1910–14: *Eurhythmia vel compositio rhythmica prosae antiquae* (3 vols), Leipzig.
- Zeleny, K. 2008: *Itali modi. Akzentrhythmen in der lateinischen Dichtung der augusteischen Zeit*, Vienna.
- Zielinski, T. 1904: *Das Clauseigesetz in Ciceros Reden. Grundzüge einer oratorischen Rhythmik*, Leipzig.
- Zwierlein, O. 2002: 'Augustins quantitierender Klauselrhythmus', *Zeitschrift für Papyrologie und Epigraphik* 138, 43–70.