

Euterpe's Hidden Song: Patterns in Elegy

- Scheirer, Walter
Center for Brain Science, Harvard University
wscheirer@fas.harvard.edu
- Forstall, Christopher
Department of Classics, State University of New York at Buffalo
forstall@buffalo.edu

Summary

This poster presentation is the culmination of an extensive statistical study of Latin elegy. The notion that poetic sound represents a distinct layer of information where discrete patterns can be found is an intriguing one. In a theoretical context, sound captures a "metalanguage" that is universal to all discourse, but only exposed in poetry. This poster develops a model for metalanguage and applies it to Latin elegiac poetry, with a study of the primary poets working in this mode. Through the application of statistics, the structure of this metalanguage is uncovered, leading to new critical conclusions with respect to poetic style. The following abstract provides a brief overview of the work.

1. Introduction

It is widely acknowledged that classical Latin poetry was heavily influenced by Greek oral poetic traditions. The perception of sound is critical to the experience of the poetic work, and lyrical virtuosity is its hallmark. As Ezra Pound reminds us, "poetry begins to atrophy when it gets too far from music" [1]. Recent scholarship in Latin poetry emphasizes the primacy of meter as a framework to organize sound [2]. Classical notions of sound differ, however, with Plato and Aristotle suggesting that music itself is far more organic – a reflection of natural sounds and emotion that emanates deep within the self [3] [4]. In our own study of the role of sound in poetry, we have found patterns in the elegiac form at extraordinarily fine levels, only recently detectable with the use of methods from the digital humanities, and never remarked upon by any ancient or contemporary commentator. These patterns are a signature of the process that generates language, forms meter, and enables the creative act.

2. Aspects of Poetic Composition

Consider the phoneme, which is the fundamental building block of language. Philodemus of Gadara posits that the formation of poetry is contingent upon an agreeable arrangement of phonetic units [5]. Views on how such an arrangement could come to be varied in classical times. *Anomalists* like Varro argue for an injection of novelty into language to draw our attention, while *Analogists* including Julius Caesar recommend an adherence to a perceived natural ordering of linguistic elements [6]. In all of these discussions, the relationship between basic elements is key.

Saussure assumes that differential relationships between phonemes are necessarily negative and reduced to opposition [7]. Deleuze provides a critique of this based on the deeply plausible possibility that a positive relationship can exist. He states, "For opposition teaches us nothing about the nature of that which is thought to be opposed. The selection of phonemes possessing pertinent value in this or that language is inseparable from that of morphemes as elements of grammatical constructions" [8]. An interplay between *reciprocal* sounds leads one to consider alternatives to Saussure's theory.

The linguist Gustave Guillaume writes of a hidden nature of language formulation in the mind before speech actualization [9]. Guillaume argues that *potential* phonetic, lexical, and semantic forms interact in myriad combinations, defining a "metalanguage" that supports natural language. Deleuze seizes upon this idea, stating that metalanguage "cannot be spoken in the empirical usage of a given language, but must be spoken and can be spoken only in the poetic usage of speech" [10]. It is this profound and fundamental insight that we scrutinize in this work.

Is there some way to resolve the opposition between the linguistic theories of Saussure and Guillaume? After all, both systems hint at a probabilistic model that follows some set of rules to enforce cognitive economy, which facilitates memory. Successful "linguistic throws of the dice" [11] yield novel turns of phrase, while other

combinations fall flat. The link from theory to experimentation that we pursue can be stated as follows: language is constructed by complex unseen interactions between linguistic elements in the mind (Saussure and Guillaume); poetry is unique, in that it provides a window into the aforementioned process (Deleuze). Based on our analysis of a large corpus using the statistical techniques of distant reading [12] [13], this model appears to have some explanatory power.

3. The Elegiac Form

As a first case study, we examine the poetic form of the elegiac couplet under the lens of the above model. The elegiac meter is used for a variety of themes, most notably erotic love [14]. The elegiac couplet is a pair of two different one-line "verses". The first line is identical to dactylic hexameter; the second, often called the "pentameter" line of the couplet, is shorter by two half-feet. The scansion is shown in Fig. 1. In our analysis, we consider all of the extant elegies from Catullus, Propertius, Tibullus, Ovid, and Martial. A selection of non-elegiac poems is also considered for comparison. All texts come from the Perseus Digital Library [15].

— U | — U | — U | — U | — UU | — U
 — U | — U | — || — UU | — UU | U

Fig. 1: The elegiac couplet. "—" represents a long syllable, "u" a short syllable, "U" either one long syllable or two shorts, and "U" either one long syllable or a short.

4. Statistical Analysis

For stylistic analysis, the choice of style marker is important [16] [17] [18] [19]. In this work, we look at a particular form of character-level bi-grams as a proxy for phonemes. Unlike phonemes, character-level bi-grams do not suffer from potential errors introduced by scholarly judgment of how an ancient word might have sounded. A *functional bi-gram* [20] [21], when applied at the character-level, is an n-gram-based feature [22] that describes the most frequent sound-oriented information in a text. Similar to function words, functional n-grams are those n-grams that are elements of most of the lexicon, necessitating their use. We computed functional bi-grams using a set of custom perl scripts, and calculated statistics via Microsoft Excel. All code and data will be released at DH2014.

In Latin elegies, the most frequent bi-gram is "er". Because of its frequency, this sound alone is quite sensitive to meter, author, and literary era. Other sound choices exert influence upon "er", yielding interesting patterns. It is natural to ask why this is so. In order to see any patterns, we need a large enough feature sampling – otherwise we are simply lost in the noise. This is exactly what a functional bi-gram like "er" is meant to provide.

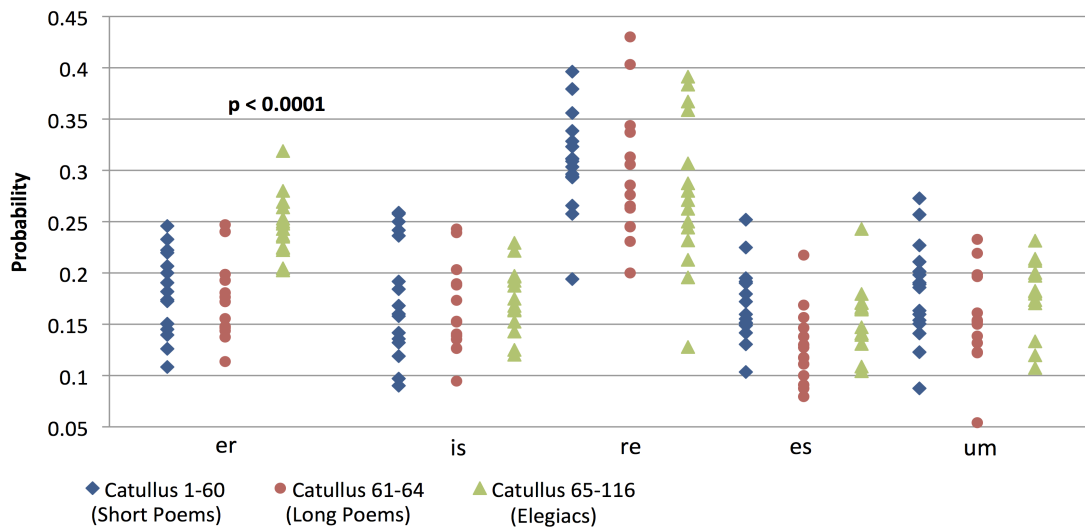


Fig. 2: Functional bi-gram probabilities for Catullus.

Calculating the associated probabilities for "er" over a collection of 50 line samples spanning the entire Catullan corpus exposes a substantial divergence between the polymetric poems, numbered 1-64, and the elegiacs, numbered 65-116, in Fig. 2. The probability of "r" directly following "e" is much higher in the elegiacs. Given just

"er" for a sample, it is not difficult to guess its meter. In a rigorous statistical sense, we can ask if the bi-gram "er" is truly significant compared to the other most frequently occurring functional bi-grams. The null hypothesis states that any bi-gram should occur with roughly the same frequency across the entire corpus. For a multiple comparisons scenario with 100 hypothesis tests, the significance level is 0.0001. "er" is statistically significantly different ($p < 0.0001$) via the two-tailed paired t-test.

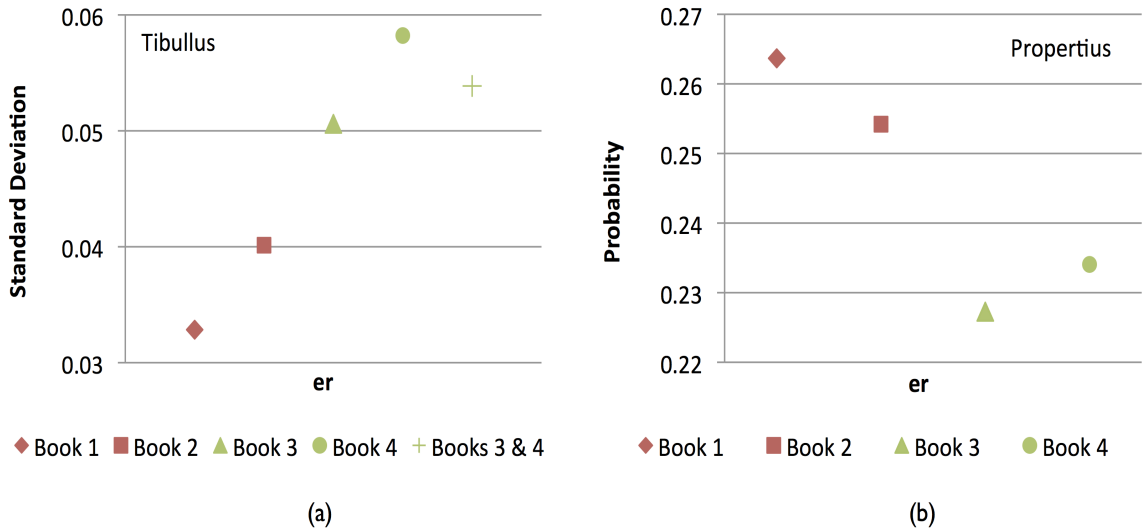


Fig. 3: a. Variation between books 1 & 2 and 3 & 4 of Tibullus; b. "er" bi-gram probabilities for Propertius.

What does "er" tell us about other elegists? In the case of Tibullus, books 3 & 4 are generally acknowledged to be the work of other poets [23], including the perhaps apocryphal contributions of Sulpicia [24]. Considering variation with respect to "er", there is a noticeable increase in standard deviation for books 3 & 4, which we would expect for multiple authors. In another example, poets after Catullus often end the pentameter line with a two-syllable word [25]. Propertius, following Catullus and Tibullus, does not do this in books 1 & 2, but adopts the style in books 3 & 4. From the change in values taken on by the "er" feature, a conclusion can be drawn that this new constraint affects sound choice by increasing the number of sound combinations that are used, thus decreasing the probability of "er".

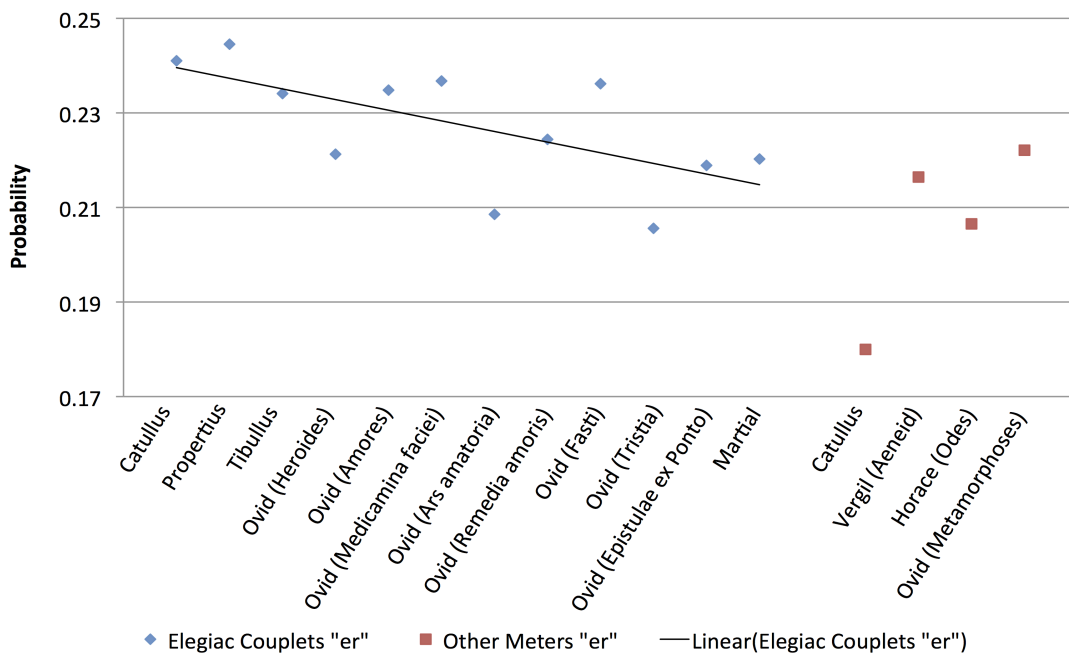


Fig. 4: As the elegiac form evolves, the probability of "er" occurring declines. A linear regression model was fit to the elegists, highlighting the downward trend. The x-axis is arranged chronologically.

A poem's place in time also reflects its style. When viewed chronologically in Fig. 4, the probability of "er" occurring declines. Even within just the Ovidian corpus, this trend is evident. Beyond the change in the pentameter line endings, the pursuit of innovation means more poets look to new sound combinations, which drives down the probability of "er" – and gives some credence to the anomalist argument. Ovid, a master stylist, does not want to sound like his old self as his work progresses. In this vein, a further note can be provided on the magnitude of the meter's role in composition. We find in Fig. 5 a smaller number of longer words in dactylic hexameter compared to elegiac hexameter over the complete works in these meters for thirteen poets. We can explain this as a blending of a genre-dependent signal with the meter signal: the pentameter line tends to have shorter words because of constraints imposed by the meter; this tendency steers word choice towards shorter words in the hexameter line, even though here, as proven by dactylic hexameters, word length is not so constrained by meter.

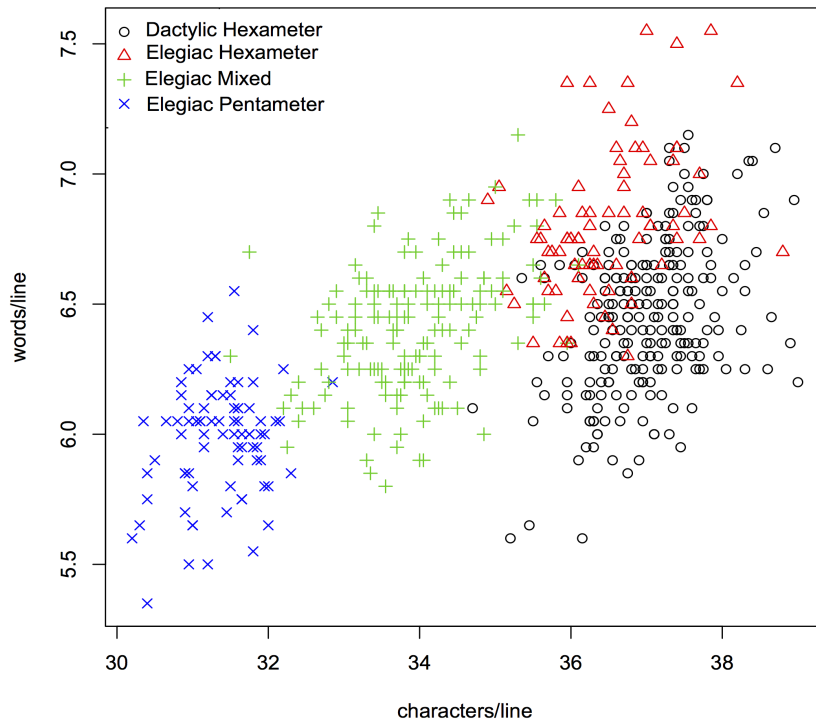


Fig. 5: The interplay between the hexameter and pentameter lines results in shorter hexameter lines with more words, which are atypical outside elegiacs.

5. Discussion

What can be said of these findings? The most frequent sound in a poem provides important clues to the overall construction of language in its aesthetic and historical contexts. Poetry is wonderful for many reasons – that it is a window into the mind is perhaps the most striking.

Acknowledgements

This work was supported by NEH Start-Up Grant Award #HD-51570-12.

References

1. Pound, E. 1960. *ABC of Reading*, New Directions, p. 14.
2. Morgan, L. 2010. *Musa Pedestris: Metre and Meaning in Roman Verse*. Oxford University Press.
3. Grube, G and C. Reeve. 1992. *Plato: Republic*, Second Edition. Hackett Publishing Company, pp. 52-80.
4. Lord, C. 2013. *Aristotle's Politics: Second Edition*. University of Chicago Press, pp. 223-239.
5. D. Armstrong, 1995. "The Impossibility of Metathesis: Philodemus and Lucretius on Form and Content in Poetry." In Obbink, D. ed. *Philodemus and Poetry: Poetic Theory and Practice in Lucretius, Philodemus, and*

- Horace. Oxford University Press, pp. 210-233.
6. Colson, F. 1919. "The Analogist and Anomalist Controversy," *The Classical Quarterly*, 13.1.24-36.
 7. Saussure, F. 1966. *Course in General Linguistics*. McGraw-Hill.
 8. Deleuze, G. 1968. *Difference and Repetition*. Columbia, p. 205.
 9. Guillaume, G. 1984. *Foundations for a Science of Language*. John Benjamins.
 10. Deleuze, G. 1968. *Difference and Repetition*. Columbia, p. 193.
 11. *Ibid.*, p. 205.
 12. Moretti, F. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.
 13. Jockers, M. 2013. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.
 14. Morgan, L. 2010. *Musa Pedestris: Metre and Meaning in Roman Verse*. Oxford University Press.
 15. Perseus Digital Library. Ed. Gregory R. Crane. Tufts University. <http://www.perseus.tufts.edu>. Accessed October 17, 2013.
 16. Burrows, J.F. 1989. "An Ocean Where each Kind...: Statistical Analysis and Some Major Determinants of Literary Style," *Computers & the Humanities*, 23.309-321.
 17. Eder, M. 2008. "How Rhythmical is Hexameter: a Statistical Approach to Ancient Poetry," *Digital Humanities*, 2007.
 18. Juola, P. 2008. "Authorship Attribution," *Foundations and Trends in Information Retrieval* 1.3.233-334.
 19. Hoover, D. 2013. "The Full-Spectrum Text-Analysis Spreadsheet," *Digital Humanities*, 2013.
 20. Forstall, C.W. and W.J. Scheirer. 2009. "Features from Frequency: Authorship and Stylistic Analysis Using Repetitive Sound," *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1.2.1-23.
 21. Forstall, C.W., Jacobson, S.L. and W.J. Scheirer. 2011. "Evidence of Intertextuality: Investigating Paul the Deacon's *Angustae Vitae*," *Literary and Linguistic Computing* 26.3.285-296.
 22. Jurafsky, D. and J. Martin. 2009. *Speech and Language Processing: Second Edition*. Pearson Prentice Hall.
 23. Conte, G.B. 1999. *Latin Literature: A History*. Johns Hopkins, pp. 330-331.
 24. Hubbard, T. 2005. "The Invention of Sulpicia," *The Classical Journal*, 100.2.177-194.
 25. Platnauer, M. 1951. *Latin Elegiac Verse: Study of Metrical Uses of Tibullus, Propertius and Ovid*. Cambridge, p. 17.