A Statistical Stylistic Study of Latin Elegiac Couplets

C. W. Forstall[1] and W. J. Scheirer[2]

1. Department of Classics, State University of New York at Buffalo (forstall@buffalo.edu)
2. Department of Computer Science, University of Colorado at Colorado Springs (wjs3@vast.uccs.edu)

Abstract for Paper Presentation
Continuing our study [2, 3] of repetitive sound and its relationship to style in poetry, this talk introduces a variety of statistical features found to be useful descriptors of Latin elegiac couplets (for background, see [4]). Using computational statistical methods, we have undertaken a broad survey of Latin elegiac poets. The elegiac meter is used for a variety of themes, most notably Love [1, p. 322]. The elegiac couplet is a pair of two different one-line "verses":

– ∪ | – ∪ | – ∪ | – ∪ | – ∪∪ | – ⊔
– ∪ | – ∪ | – || – ∪∪ | – ∪∪ | ⊔

In the above, "–" represents a long syllable, "∪" a short syllable, "∪ either one long syllable or two shorts, and ⊔ either one long syllable or a short. The first line is identical to dactylic hexameter; the second, often called the "pentameter" line of the couplet, is shorter by two half-feet. For this work, we are most interested in addressing the nature of the sound that is constrained by this meter—does it reflect the voice of the poet, or the general style of the elegiac form?

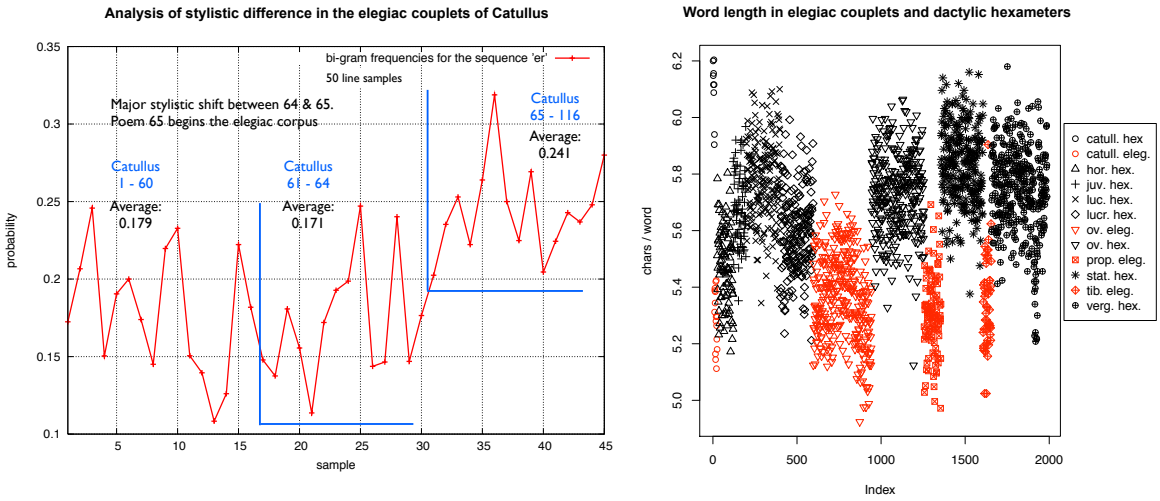
Our *functional n-gram* [2], when applied at the character level, is a feature that describes the most frequent sound oriented information in a text. The values taken on by distinct functional n-grams have been found to vary by poet and meter. They can even reveal much about the style of a single poet. As an illustration, we consider the entire Catullan corpus (left side of figure). An examination of frequently occurring character-level bi-grams shows that the sequence *er* is the top bi-gram that is common to all poems in Catullus (this is also true of all poets considered). Calculating the associated probabilities for this feature over a collection of 50 line samples spanning the entire corpus exposes a distinct break between the polymetric poems (1 – 64) and the elegiacs (65 – 116). Between the short (1 – 60) and long poems (61 – 64), we observe consistency in the average probabilities – 0.179 and 0.171 respectively. For the elegiacs, we observe a sharp jump in frequency to an average probability of 0.241. Thus, we have a distinct style marker in the form of the most common sound pattern.

The significance of *er* to elegiac couplets is not peculiar to Catullus. Samples drawn from the elegies of Ovid, Propertius and Tibullus also exhibit the same probabilistic consistency as Catullus for *er*. However, a question remains over the nature of the author and meter signals and their individual influences on the poetry. A correlation between all elegiac poets considered with respect to the feature *er* suggests that meter influences the sound-level form of the poetry by constraining word choice. But the standard deviation of the bi-gram frequency *er*, calculated over samples drawn from a particular poet, indicates the additional presence of an author signal. For instance, for 50 line samples representing the three individual books of Tibullus, the resulting standard deviations are $\sigma_1 = 0.03226$, $\sigma_2 = 0.04335$, $\sigma_3 = 0.0532$. The highest standard deviation belongs to book 3, which is attributed to a collection of poets, including Tibullus, Sulpicia, and other (often inferior) writers. The presence of multiple authors explains this higher feature variance, which weakens the slight author signal.

A broader survey of multiple authors in two different meters supported the hypothesis that meter

affects bi-gram frequencies. Here, elegiac couplets were compared with stichic (continuous) dactylic hexameter. Samples of 250 randomly-chosen words were used. Several functional bi-gram frequencies showed greater metrical variation than *er* in this case: *nt*, the greatest, but *um* and *am* were also sensitive to a meter signal. Beyond bi-gram frequencies, useful results were obtained from *am*:*um*, the ratio of the number of occurrences of the two bi-grams, and from mean word length, the feature most sensitive to meter (right side of figure). The number of characters per word tended to be higher for dactylic hexameter than for elegiac couplets both within and between authors. Catullus 64 was dramatically higher, separated completely from the rest of the Catulluan corpus, and generally higher than samples from any author in either meter.

A deficiency in this study was the lack of a large data base of authors who wrote in both meters. One way to solve the problem of author- and meter signals' overlapping is to split the elegiac corpus into two halves, a hexameter half and a pentameter half, cutting each couplet in two. This produced equally-sized corpora for each author in each of two metrical forms, and eliminated the effects of whatever genre differences exist between stichic hexameters and elegiac couplets. A preliminary study considered small elegiac corpora by Catullus, Ovid, Propertius, and Tibullus. The hexameter lines were treated separately from the pentameter lines; samples were of 150 randomly-chosen words. As in the hexameter-elegiac study, the bi-gram frequency *nt*, the ratio *um*:*am*, and particularly word length were most sensitive to the difference between hexameter and pentameter. While, as expected, word length was greater for the hexameter half of the elegiac couplet than for the pentameter, it was still not as high as for stichic hexameters. One model to explain this postulates 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 influences word choice in the hexameter line, even though here (as proven by the stichic hexameters) word length is not so constrained by meter.



References

[1] G.B. Conte, *Latin Literature: A History*. Translated by J.B. Solodow. The Johns Hopkins University Press, 1999.

[2] C.W. Forstall and W.J. Scheirer, "Features from Frequency: Authorship and Stylistic Analysis Using Repetitive Sound," in Proc. of the *Chicago Colloquium on DHCS*, Nov. 2009.

[3] C.W. Forstall, S.L. Jacobson and W.J. Scheirer, "Evidence of Intertextuality: Investigating Paul the Deacon's *Angustae Vitae*," presented at *Digital Humanities*, July 2010.

[4] M. Platnauer, *Latin Elegiac Verse: Study of Metrical Uses of Tibullus, Propertius and Ovid*. Cambridge University Press, 1951.