

Evidence of Intertextuality: Investigating Paul the Deacon's *Angustae Vitae*

Forstall, C. W.

forstall@buffalo.edu

Department of Classics, State University of
New York at Buffalo

Jacobson, S.L.

slj25@buffalo.edu

Department of Classics, State University of
New York at Buffalo

Scheirer, W. J

wjs3@vast.uccs.edu

Department of Computer Science, University of
Colorado at Colorado Springs

The study of intertextuality, the shaping of a text's meaning by other texts, remains a laborious process for the literary critic. Kristeva (Kristeva, 1986) suggests that "Any text is constructed as a mosaic of quotations; any text is the absorption and transformation of another. & The nature of these mosaics is widely varied, from direct quotations representing a simple and overt intertextuality, to more complex transformations that are intentionally or subconsciously absorbed into a text. The burden placed upon the literary critic to verify suspected instances of intertextuality is great. The critic must reference a large corpus of possible contributing works, and thus must often be familiar with more texts than was the author whom they are studying. Since, in many cases, the problem is one of pattern recognition, it is a good candidate for automated assistance by computers.

In this work, currently in progress, we propose the use of machine learning and related statistical methods to improve the process by which intertextuality is studied. Specifically, we are working with instances where an author has knowledge of particular texts, and reflects this in discrete passages within their own work. These intertexts may comprise fragmentary quotations, paraphrases, or even stylistic similarity. A passage may be reminiscent of a particular author, or of

a particular literary group. We have defined three different classes of style markers to verify intertexts: phonetic, metric, and dictional. To evaluate the proposed style markers and classification methods, we have chosen an intriguing case study: Paul the Deacon's 8th century poem *Angustae Vitae* (Paul the Deacon, 1881), which we suggest has a strong connection to first-century Neoteric poetry.

Our earlier work in authorship and stylistic analysis (Forstall and Scheirer, 2009) has considered the importance of phonetic style markers, with the observation that sound plays a fundamental role in an author's style, particularly for poets. To capture sound information, we have developed a feature that we call a *functional n-gram*, whereby the power of the Zipfian distribution is realized by selecting the n-grams that occur most frequently as features, while preserving their relative probabilities as the actual feature element. By using more primitive, sound-oriented features, namely, character- and phoneme-level n-grams, we are able to build accurate classifiers with the functional n-gram approach. We have used two different classification algorithms with functional n-grams, yielding very promising results. The first method, a traditional SVM learning approach based on the work of Diederich et al. (Diederich et al., 2003), distinguished authors of Latin poetry with 98.75% accuracy. The second method, a PCA clustering approach (Holmes et al., 2001), showed distinct stylistic separation between the Homeric poems. In light of our previous work, we know that phonetics is an important tool for verifying intertexts, for the same reason it was important for poetics — repetitive sound distinguishes style.

Following this idea of repetitive sound further, we use meter as an additional style marker. For strict meters, it is straightforward to identify their type by analyzing the weights of the syllables in a line. In practice, the nuance of particular poets, or groups of poets, creates unique variations in meter, giving us a discriminating feature. By including meter as another dimension in the feature vector of the SVM learning for the functional n-grams described above, we enhance the discriminatory power of the resulting classifiers. It remains an open question in our work whether meter

alone is powerful enough to achieve the same classification results for individual authors as the functional n-gram. Its utility for group classification is more apparent.

Pulling back from sound, we have also developed a style analysis for diction that is somewhat the opposite of the functional n-gram approach. Considering the Zipfian distribution once again, we turn to elements that occur with lower probabilities. The power of functional n-grams relies on the amount of information carried by the elements at the left side of the Zipfian distribution (assuming the x axis is organized from most frequent to least frequent). For this new style marker, we desire something that is the opposite of functional—features that occur infrequently, but not necessarily *hapax legomena*.

Thus, we fix a desired probability range for words that occur infrequently, and look for n-gram sequences composed of only those words in a particular passage, ignoring all others:

$$(P_{low} < \Pr(\text{word}_1) < P_{high}) \dots (P_{low} < \Pr(\text{word}_2) < P_{high}) \dots (P_{low} < \Pr(\text{word}_n) < P_{high}) \quad (1)$$

where $n \geq 3$. The probability of the resulting n-gram is compared to pre-computed probabilities of the same n-gram (should it exist) for specific authors, or literary groups. This type of style marker is very well suited to our case study, where certain word sequences are common to a particular group (the Latin Neoterics), but uncommon or non-existent in the work of other groups.

With our style markers in hand, we turn to our case study. In *Angustae Vitae*, Paul the Deacon opposes poetic inspiration and production in the classical world with the writing of poetry in the Christian monastic context. Although he posits the classical and monastic worlds as opposites, the use of Catullan diction and models of poetic exchange recalls the paradigm of the Neoteric, proto-elegiac lover, his beloved, and his poetological concerns. This model is recontextualized to reflect monastic love and poetic exchange. The source of inspiration remains the same — love — but there is a new beloved and a new Muse. While he avoids saying he directly imitates his classical predecessors, Paul the Deacon's poetry is peppered with classical intertexts. Not all of these intertexts are

purposeful allusions. However, it is clear that he was at least well-versed in the poetry of Horace, Virgil, and Ovid. Thus, it becomes our task to verify the portions of the poem believed to be inspired by Catullus.

Further study of the Catullan manuscript tradition and Paul the Deacon's life would be necessary to prove his knowledge of Catullus conclusively. This work will proceed from the *a priori* conclusion that Paul the Deacon had read Catullus. This conclusion is based on the abundance of intertexts and the crucial role they play in coloring Paul the Deacon's poetry. We can gain a sense of these intertexts we are examining by looking at a particular instance.

As the poem opens, the Muses, who have found the cloistered life not of their liking, have abandoned the narrator. More precisely, the Muses are fleeing the fellowship of the cloistered life, *angustae vitae fugiunt consortia Musae*. Thus, there is an opposition: the *consortia angustae vitae versus the Musae* — the fellowship of monastic life versus the classical relationship of the poet and Muses. This opposition continues in lines 2 — 3. The Muses do not wish to live in the fenced-in gardens of monasteries, but rather they desire to play in rosy meadows, *claustrorum septis nec habitare volunt, / per rosulenta magis cupiunt sed ludere prata*. Here, *septa*, the cultivated, enclosed garden of a monastery, is contrasted with the *rosulenta prata*, a wild, open meadow.

These opening lines may reference Eclogue 1 of Vergil, but reading them with Catullus provides a richer understanding of the themes of the poem. Indeed, what the Muses desire in *Angustae Vitae* are cornerstones of Catullan diction. Lines 2 — 3 recall the opening lines of Catullus 2. The Muses desire to play in fields (*cupiunt sed ludere prata*) and to tend to (*colunt*) their delights (*delicias*), just as Catullus' girlfriend is accustomed to play (*ludere solet*) with her own pet/delight (*deliciae*). The classical Muses are compared with the poet's beloved. These are the Muses of elegiac love. *Ludere*, furthermore, is a by-word in Catullus for the production of poetry.

By taking into consideration all such intertext candidates in *Angustae Vitae*, we will show that machine classification is able to produce statistically strong validation results. We

present our study of the three style markers in this context, highlighting strengths and weaknesses. We hope that this case study will serve as a first step towards a more sophisticated and efficient analysis of intertextuality in general. Moreover, this work raises important linguistic questions on the nature of conscious and subconscious influence in style, which is an area we intend to explore in further work.

References

Diederich, J., Kindermann, J., Leopold, E. and Paass G. (2003). 'Authorship Attribution with Support Vector Machines'. *Applied Intelligence*. **19(1-2)**: 109-123.

Forstall, C.W. and Scheirer, W.J. (2009). 'Features from Frequency: Authorship and Stylistic Analysis Using Repetitive Sound'. *Chicago Colloquium on Digital Humanities and Computer Science*. Chicago.

Holmes, D., Robertson, M., and Paez, R. (2001). 'Stephen Crane and the New York Tribune: A Case Study in Traditional and Non-traditional Authorship Attribution'. *Computers and the Humanities*. **35(3)**: 315–331.

Kristeva, J. (1986). 'Word, Dialogue and Novel'. *The Kristeva Reader*. Moi, T. (ed.). New York: Columbia University Press, pp. 34–61.

Paul the Deacon (1881). 'Carmina'. *Monumenta Germaniae Historica, Poeta Latini Aevi Carolini Vol 1*. Diemmler, E. (ed.). Berlin.