Computational Profiling of Genre and Speech Styles in Latin Literature

This paper presents a quantitative approach to analyzing genre and speech in Latin literature. We build stylometric profiles of the canonical genres of Latin prose and verse based on the frequencies of function words, syntactic features, and other non-lexical markers, and use machine learning to situate the style of each genre within the corpus of classical Latin literature. We then extend these methods to analyze progressively finer distinctions: authors writing within the same genre, differences between subgenres, and finally the styles of speakers in individual works. The resulting profiles offer a multidimensional portrait of the stylistic tendencies typical of the major Latin genres (drama, elegy, and epic for verse, epistolography, historiography, oratory, philosophy, and technical treatise for prose), as well as detailed information about intra-generic and intra-authorial heterogeneity in style.

Recent work in computational literary studies has shown that taking the coherence of genre as a working assumption can enable productive lines of interpretation and investigation of cross-temporal trends (Moretti 2013, Jockers 2013, Wilkens 2016, Underwood 2019). Yet while genre is a central concern of literary criticism, especially within Classics, there is little consensus regarding specific generic definitions, or even how much interpretative weight should be given to the concept due to the fluidity of generic boundaries (Harrison 2007, Derrida 1980). To the extent that concepts of genre cohere, they do so through relationships of meter, diction, theme, and reception. Small-scale features that occur frequently provide potentially new evidence about generic style but are prohibitively difficult to count without computation. In our feature set, we incorporate stylometric features long considered important to Latin literary style, such as *atque*

followed by a consonant (Adams 1972), as well as markers drawn from computational studies of English genres (Jockers 2013).

In the first part of the study, we use supervised machine learning to predict genre based on style. In our experiment, an algorithm is given a subset of texts labeled by genre and learns to identify associations between our quantitative stylistic data and the genre labels. Afterwards, the algorithm predicts the genre of the remaining texts based solely on twenty-six stylometric features. With this approach, we achieve greater than 85% accuracy in genre identification across the whole corpus. To aid in the interpretation of these results, we use statistical feature ranking to identify the characteristics that best distinguish each genre from the rest of the corpus. For example, we find that historiography is distinguished by a low frequency of conditional sentences except in passages of direct speech. To explore the implications of this corpus-wide finding, we analyze the accounts of the rape and suicide of Lucretia in Livy (1.57-60) and Augustine (*De civitate dei* 1.19), which, although covering the same subject matter, differ markedly in narrative structure and use of conditionals (four in 1,069 words for Livy, thirteen in 672 words for Augustine; cf. Trout 1994).

In the second part, we employ dimensionality reduction techniques with the same feature set to examine variation within genres and individual works. We perform a detailed stylometric comparison of narrative and didactic epic, finding clear enrichment of hypotactic markers, such as longer sentences and more frequent use of relative clauses, in didactic. The paper concludes with an extended case study of mortal and divine speech in Vergil's *Aeneid* and Statius' *Thebaid*. We find that, for both epics, the speech of gods and human characters can be distinguished using principal component analysis of the stylometric data, with one notable exception – Juturna's speech in *Aeneid* 12 appears human in style. Against that background, we argue that Juturna's

final lines, in which she imagines the impossibility of being sent down to the shades, casts an especially ironic light on her use of the word *deam* of herself and on the narrator's use of *dea* two lines later (*Aeneid* 12.884-886). She is not just a goddess who regrets her immortality and exits the poem in a form of burial; in a fundamental sense she simply does not sound like a god.

Bibliography

Adams, J.N. (1972). "The Language of the Later Books of Tacitus' *Annals*" in *CQ* 22.2: 350-373.

Derrida, J. (1980). "The Law of Genre" in *Glyph* 7: 202-232.

Harrison, S.J. (2007). *Generic Enrichment in Vergil and Horace*. Oxford.

Jockers, M. (2013). *Macroanalysis*: *Digital Methods and Literary History*. Champaign, Il.

Moretti, F. (2013). *Distant Reading*. London.

Trout, D. (1994). "Re-Textualizing Lucretia: Cultural Subversion in the *City of God*." in *JECS* 2: 53-70.

Wilkens, M. (2016). "Genre, Computation, and the Varieties of the Twentieth-Century U.S. Fiction" in *Journal of Cultural Analytics*.

Underwood, T. (2019). *Distant Horizons: Digital Evidence and Literary Change*. Chicago, Il.