The aim of this project is to fill a somewhat puzzling lacuna by combining the study of Latin poetry with state-of-the-art techniques in computational stylometry. As with many projects in computational stylometry, the project sits between disciplines—in this case, on the border between Classical Philology and Computational Linguistics.

The importance of Latin poetry to the modern literary tradition (in many languages!) hardly needs to be rehearsed—nor indeed its impact on art, philosophy and popular culture. Indeed, many of the standard stories of 'Greek' mythology are, in fact, known only through Ovid's version in the *Metamorphoses*. It has long been accepted that reasonably large samples are required to perform meaningful stylometry. However, the standard stylometric techniques look mostly at how authors use *words* (which words, how often, how many, etc). But poetry is much more stylistically dense than prose when considering all of the available stylistic markers (metrical features, sound, line patterning, ...). This, logically, should mean that meaningful results are achievable with much smaller samples, provided that we explore areas beyond lexical features (i.e. words). It is these *extra-lexical features* that will be the main subject of research.

With this in mind, this project identifies three focus areas for investigation. Area 1 aims to consider 'sonic style'. By transforming the texts to a phonetic representation, the intention is to explore stylistic markers that may be present in features such as rhyme (or assonance, consonance etc) and alliteration. Area 2 aims to explore the stylistic fingerprints that poets leave when they employ quotation or allusions to other works. Area 3 is focused on what we might broadly call 'grammatical style'. Beginning with the simplest ideas, the project will explore some simple multivariate statistics (which tenses, cases or declensions do the authors prefer? How complex are their clauses?). From there we plan to extend to more complex technical questions like hyperbaton, tmesis, chiasmus and so forth. These three areas will lay out a broad foundation for the study of extra-lexical style, although additional concepts may be investigated as time and opportunity permit.

In the 19th and 20th centuries there were already a significant number of what might be called "quantitative" (as opposed to "statistical") philological studies. These authors were highly innovative in identifying stylometric features; features of versification, rhythm, rhyme, metrical variation, or features of diction and structure. In many cases, the intricate tabulations of thousands of lines of verse were done without computers! Unfortunately, however, most of the analyses have been fairly ad hoc in nature. This project, therefore, aims to significantly advance the field by combining philological intuition with modern statistical approaches and methodological rigor. It is not enough to identify useful-seeming stylometric features—the results should be carefully examined for empirical and theoretical significance, for accuracy, and cross-tested and validated against multiple works. This is all part of the regular practice of computational stylometry, and should prove to be a very welcome complement to the historical work. With modern and open scientific methodology, stylometric results for Latin poetry will become more reliable, independently verifiable and more expedient (via the open publication of tools, techniques and documentation). In addition, some of the approaches developed here should be applicable to languages other than Latin, and will thus form part of the frontier of computational stylometry itself, potentially offering new ideas or approaches to existing practitioners.