Macro-Etymological Literary Analysis

Book Proposal

Jonathan Reeve

2019-11-30

Consider the words *kingly*, *royal*, and *regal*. Although lexically synonymous, they have subtle but crucial differences in tone. *Kingly* feels the least formal, owing to its origin in the Anglo-Saxon language, while *royal* and *regal*, as descendants of French and Latin, evoke increasingly serious connotations. Triplets like these, of which English has many, not only speak to the level of discourse of a text, but hint at its genre, characterization, and affective infrastructures. Macro-etymological text analysis is a new computational technique for measuring these resonances, by quantifying the histories of the words in a text. In other words, it is a method of digitally reading a work, by leveraging the language histories of its words. This book, tentatively titled *Macro-Etymological Text Analysis: Computational Applications of Language History to Liteary Criticism*, introduces the method, and provides four demonstrations, covering much of the history of British and Irish literature.

Each of the four chapters in this short book corresponds to a case study of an individual writer: Geoffrey Chaucer, John Milton, Mary Shelley, and James Joyce. These writers are not only chosen for their canonicity in literary studies, or for their representativeness of certain centuries (14th, 17th, 19th, and 20th, respectively), but because macro-etymological analysis is uniquely positioned to reveal a new dimension to their works, and to contribute to the rich scholarly traditions that surround them. The works discussed in each chapter, *The Canterbury Tales*, *Paradise Lost*, *Frankenstein*, and *A Portrait of the Artist as a Young Man*, all possess latent structures which this new analysis foregrounds.

This book's introduction begins by presenting a brief account of the colorful history of the English language, showing the circumstances under which the language has borrowed from others. The use of Latin among the clergy in the middle ages, for instance, accounts for Latinate liturgical terminology, and the translation history of ancient Greek mathematical works accounts for Hellenic words in mathematics. The English aristocrats' use of French following the Norman conquest of 1066, and the Anglo-Saxon in use among the lay classes, may explain why the French-descended words for meats (*pork*, *beef*) are so different from their corresponding animals (*pig*, *cow*). From this foundation, I explain the algorithm used to perform the analyses, a computer program called `macro-etym`. One of the very few tools created specifically for use in computational literary analysis, it navigates the generational hierarchies of the words in a text, and—using methods initially suggested by 16th century scholars of Chaucerian etymology—it selects branches of their language history according to the resonances they retain. It is also among the first to use the

Etymological Wordnet, a highly multilingual database of etymological relations between words, originally developed by the computer scientist Gerard de Melo.

The first chapter, on Chaucer's *Canterbury Tales*, highlights the permeability of Chaucer's Middle English with its linguistic neighbors French and Latin, and the ways in which Chaucer leveraged this to build nuanced voices for his narrators. I reproduce an early, analogue etymological analysis of Chaucer's work, and extend it by showing differences in etymological trends among the narrators of the tales. The oft-cited *quiteing* dynamic between Chaucer's narrators, for instance—the ways in which the narrators answer and echo each others' tales—appears in the ways they use words of varying language families. Furthermore, Chaucer's orthographic habits show divergent etymologies even among variations of the same word.

Milton's epic poem *Paradise Lost*, the subject of the second chapter, is another prime example of linguistic-historic tenor used to effect poetic structure. Only here, interlinguistic tensions are used more for evoking locale: I show how proportions of Latinate words in Milton correlate with his cosmological setting, whether heaven, hell, or earth. I also build on Miltonic scholarship to quantify the use of affixes in the poem, in English, usually markers of Greek or Latin roots (*pan-*, *poly-*, *pre-*), but also combined with words of other linguistic ancestry. I show how Milton uses *re-* as if it were a Latin, rather than English, prefix.

A study of Mary Shelley's *Frankenstein* provides the third chapter. An epistolary novel, its form carries both the formality of scientific record and the intimacy of personal correspondence. Since Shelley renders Frankenstein's moods of surprise and terror in Germanic terms, and his more documentarian moods in Latinate, this allows for a further analysis of the linguistic fingerprints of those moods. In this chapter, I show the interplay between the macro-scale structual patterns of etymologies and the micro-scale patterns of words and word segments, engaging with the scholarship that discusses the novel's epistolary affordances, as well as that which pays particular attention to word choice.

The fourth and final chapter is devoted to James Joyce's *A Portrait of the Artist as a Young Man*. This chapter will be adapted from my earlier chapter, "A Macro-Etymological Analysis of *A Portrait of the Artist as a Young Man*," first published in *Reading Modernism with Machines* (Palgrave Macmillan, 2016). Here, I show that the novel's modernist subjectivities, evoked in part through its epiphanic/antiepiphanic structure, is apparent in the etymologies of its words. Not only does the narrator, Stephen Dedalus, exhibit a more Latinate vocabulary as his Jesuit education progresses, but his quasi-religious epiphanies show themselves to be significantly more Latin and French than his Germanic antiepiphanies. In this novel, Joyce pays special attention to words, their sounds, and their origins, and I trace many of these to his use of Walter Skeat's etymological dictionary.

This book will be the first monograph (or "minigraph") to introduce macro-etymological text analysis as a method, and it will situate it among methods in both traditional and quantitative literary studies. The technique, and the textual insights it produces, will serve as a significant contribution to digital literary studies, and to literary criticism.