Computation and Rhetorical Invention: Finding Things to Say With word2vec

Mark Wolff

In his recent book *Friending the Past*, Alan Liu laments the waning of a rhetorical regime that until recently had held sway in literary studies as a means of making sense of the past. Instead of using what Liu names as *rhetoric-representation-interpretation* to convey an understanding of history, we are now stuck in an ambiguous regime of *communication-information-media* where it is not clear how we reach an understanding of anything (2-3). The shift from *rhetoric-representation-interpretation* to *communication-information-media* is not unique in the history of literary studies, however. It follows another shift that occurred over 100 years ago in how rhetoric was deployed. Gérard Genette observed in 1966 that literary studies had not always emphasized representations. Before the end of the nineteenth century, literary studies revolved around the art of writing. Texts were not objects to interpret but models to imitate: students demonstrated their understanding of literature by mastering elocution and reproducing figures of style in the works they read. With the institution of literary history as a nationalist project at the end of the nineteenth century, academic reading approached texts as objects to be explained according to prescribed methods for documenting how literature represented a national identity. This new way of studying literature stressed disposition, or the arrangement of ideas in the service of ideology.

The methods of literary history would eventually be used by literary scholars in the twentieth century to turn narratives about literature away from nationalism toward other priorities, most notably poststructuralism and the critique of cultural hegemonies.
Recent developments in information technology have further challenged paradigms for reading literature. Digital tools for text analysis allow for the study of large corpora using quantitative methods. As Ted Underwood, Andrew Piper and others have shown, large-scale computational text analysis has called into question fundamental concepts in literary history such as periodization, nationality, and genre. Using computational methods can enable us to develop models for literary studies, but these models are not limited to interpretation. Computational techniques such as topic modeling and word vector spaces can facilitate investigations into the possibilities for literary creation. Technology has the potential for exploring invention, or the finding of ideas to express through language given a context that can be parameterized.

If, in literary studies, an emphasis on elocution or style served the perpetuation of social hierarchies, and if an emphasis on disposition or argumentation challenged these hierarchies by promoting forms of knowledge and ideologies, a new rhetorical emphasis is needed to respond to the ontological condition of the communication-information-media era. We are surrounded by data with no clear way to make sense of it, and we need to explore inventional methods of finding things to say within this state of being. Digital environments today constitute in part the material context for suasive activity, and as Thomas Rickert argues, contemporary rhetoric must attend to how humans and the world are in this context (xv). The affordances of networked access to texts and computational processing contribute to a rhetorical ambience that grants a degree of agency to the environment in what is said about the world, which includes literature.
I will consider two examples of how tools for computational literary studies lend themselves to inventional practices. The first is *ReRites*, a year-long project by David (Jhave) Johnston who used a neural net trained on various corpora to produce poetry (“Why A.I.?” 172). The raw text of the poetry was generated by computation but Johnston edited the output. In terms that emphasize the materiality of computation, Johnston describes his role as “carving the text.” Neither the computer nor Johnston writes these poems in the sense we usually give to writing: they emerge from the world in which a machine and a human find themselves. After performing complex analyses on very large corpora, the machine produces something the human takes to find something to say with language.

The second is *SonGenApp*, a web application I developed that enables a user to select verses from a large corpus of sonnets to assemble a new poem. From all the verses in the corpus a word embedding is modeled with *word2vec*, and from the model the user selects verses semantically with an analogy based on a pair of words. The user can modify a selected verse as long as it follows the rules of scansion and rhyme for sonnets. With the application attending to formal constraints, the task of the user is to find verses that are meaningful in some way at the moment of using the application. The user can always read the source texts for selected verses and base the construction of the generated poem on a knowledge of literary themes and history. But this prior knowledge is not necessary. The user can encounter verses in the corpus by changing the analogy as if it were a knob on a black box.

The quantity of digital texts at our disposal opens possibilities for discovery in rhetorical invention. Stephen Ramsay has described a “hermeneutics of screwing
“around” where browsing resources leads serendipitously to the pleasure of finding things one had not anticipated. Computation has the potential to afford the same discovery in finding things to express through writing.

Works Cited


