

DHCS Proposal
Brad Pasanek and D. Sculley

A Study of Parody: Literary Criticism and Machine Learning

Linda Hutcheon defines parody as “a form of repetition with ironic critical distance, marking difference rather than similarity” (*A Theory of Parody* xiii). *Repetition* (countable?) with *distance* (in a metric space?) to mark a *difference* (measurable?) -- these characteristics of parody appear to be computable quantities, and our efforts to bring computation to criticism contribute to the theory of intertextuality. Where a Foucauldian critic such as Margaret Rose emphasizes the importance of incongruity or discontinuity in her theory of parody; a Russian Formalist like Victor Shklovsky, who is interested in conventionality laid bare, may emphasize the opposite. In this work, we explore the application of machine learning methods in the identification of parody, and we compare machine-generated results with similar identifications of difference and similarity by literary critics.

Pastiche and parody are two red threads; our texts a braid or a knot. We look to translations, adaptations, and burlesques of *Don Quixote* in eighteenth-century England. (The history of the novel, it has been suggested, may be no more than one sustained rewriting of Quixote.) We focus next on those texts by Samuel Richardson, Henry Fielding, and Eliza Haywood at the center of what has been called, by critic William Warner, “the *Pamela* media event.” And we treat Sterne’s endless parody in *Tristram Shandy* against this background before taking up three pairings of authors—Defoe and Coetzee, Pope and Nabokov, Austen and authors of Austen fan fiction. Moving back and forth between the eighteenth and twentieth centuries, we hope to gauge how texts are written and rewritten, one against another. Our work treats parody across remote historical periods in order to pose new questions about the Enlightenment and its relation to postmodernity.

We apply machine learning methods to our constellation of texts in the following explorations of imitation, quotation, and parody. We explore the concept of distance and repetition in parody by using a variety of formal distance metrics to find those passages that are most similar in pairs and clusters of texts. The concept of difference in parody is investigated by finding passages that most distinguish the pairs of texts, using an exhaustive search variant of active learning. Finally, we find those passages that are most likely to be confused with passages from a paired text by using a classifier-based method of anomaly detection. Additionally, we collect a parallel set of results from humans (both professional literary critics and college students) in each of these three exercises so that we have two collections of similar passages, most distinguishing passages, and passages most likely to be confused with the other text. We provide both quantitative analysis of the differences in results generated by

machine learning methods and humans, and also perform close readings of the selections found by both the machine learning methods and humans.

We pose the question: is parody a matter of vocabulary or style, idiom or genre? As a pendant and punchline to the foregoing experiments then, we generate short synthetic parody texts, using Markov models, from each of our authors. These machine-authored texts are compared to short parody texts generated by humans.