

# The Culture of Litigation and Periodical Invention,

a portion of the book project, *The Culture of Litigation, 1779-1835*,  
Mark Schoenfield, Department of English



## Introduction

My current book project, “The Culture of Litigation 1770-1835,” explores how literature—both periodical and fictional, engaged, abetted, and resisted, the rapid transformation of the British legal system, and in doing so, transformed the public’s relation to both litigation and its notion that adversarial confrontation produced truth, clarity, or resolution. To explore the role of the periodicals in this engagement, I combine close-reading analysis with various text-mining techniques that can illuminate broader trends in the roughly one-million articles represented in the Proquest British Periodicals database (PBP) for the relevant time period.

From 1770 to 1835, roughly 45,000 articles in the database mention both “law” and either “jury” or “judge”: many discuss the question of whether juries have the right to intervene in questions of law, rather than only to determine questions of fact (and the coherence of such a division); others debate the utility of juries in civil matters (an innovation for Scottish courts around 1820); still others deploy the concept of the jury metaphorically or comically, as when the “Ettrick Shepherd” “Timothy Ticker” (caricatures of James Hogg and Robert Sym respectively) debate the propriety and adjudication of talking with one’s mouth full. While these individual instances are both intriguing and significant, this project seeks ways to supplement such readings with broader claims about how key terms circulated in the public press—as a proxy for the public imagination and approximation of public discourse.



## Developing Methods

This is an on-going project, working dialectically between developing methods, having tentative findings, and then refining or altering those methods. I believe my most important recognition in this work is that no finding produced by the machine methods will be certain in itself. Rather, it will develop a more concentrated set of articles (or section of articles) that can be used for close-reading tests of critical hypotheses.

One simple method has been to use SQL for complex searches that can identify articles engaged with a particular topic; because the native search for PBP only identifies the occurrence of a word in an article, not its number of times, SQL allows more precise searches. Such searches can be guided and modified by WordToVec analyses which use Natural Language Processes to assess words which tend to be associated. For example, running a WordToVec analysis on “lawyer,” offered for each decade the 50 most substantial associations. For example, “orator” is consistently associated with “lawyer,” while the association with “Scotchman” only arises in the after 1815. With the caution that these could be artifacts of the dubious quality of the data, some are sufficiently suggestive to warrant further investigation. “Scotsman” might have been associated with “Lawyer” from early on but obscured by OCR errors, or “Scotchman”’s association with lawyers might be an effect of the increase of Scottish periodicals, which may or may not influence the social standing and visibility of lawyers

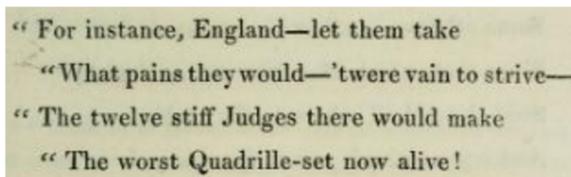
With 50 associations possible for 14 half-decades, there are 700 scores, from Brutus and booby, which occur respectively only once in 1770 and twice in 1820 and 1825, to “Orator” which occurs every year. Poet appears four times, with relatively weak associations in the 1700s, and “quack” appears four times, in the 1800, with a highest ranking of 7. These associations are not equivalent to synonyms (although the clear synonyms of barrister and attorney appear) but rather signal a mathematical association that hints at a range of possible semantic ones.

## OCR Problems: A major challenge

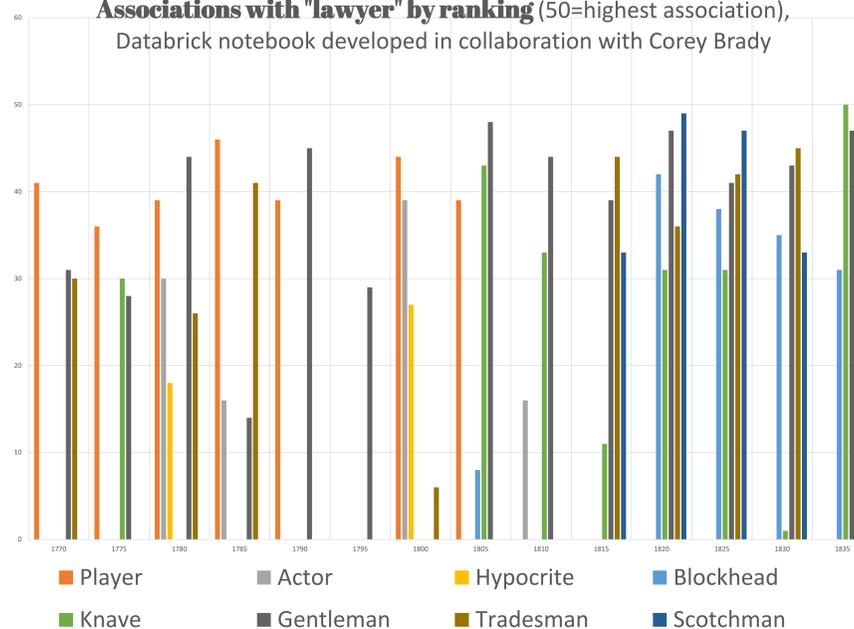
Because British Periodicals relies on its Optical Character Recognition performed two decades ago to produce the texts of the articles, its inaccuracies are a difficult stumbling block to any Natural Language Processing analyses, which depend on accurate correlations between words. Our research had shown that of the roughly 1.5 billion words, some 100million are unique, and therefore almost certainly unique OCR errors; at the same time, we know “the” is misrecognized as “tbe” nearly that same 100million times. The misrecognitions of the letter f for s and of e for c occur frequently enough to produce distortions in topic modelling and Word2Vec methods. We are currently working on several methods to correct the misrecognitions. One involves asking the computer to test single-letter substitutions against a dictionary (so the string “exprefs” becomes “express”; and then proceeding to increasingly complex substitutions. Another method is to use transformers which are trained on a corpus of correctly spelled words in context. This second method has the advantage of being able to use context, and can, for example recognize a misOCR’d single string that should be two distinct words.

Using OpenAI in conjunction with Databricks, in one sample, for example, the sentence OCR’d as “IT is a common observation, that the dwacter of a peopleis in a great meaeae influencedby thei local situation, and the nature of the scenery in which they are placed; and it is impossible to visit the Tyrol without being convinced of the truth of the remark” was corrected to the entirely accurate (except for removing the small caps of the first word and punctuation over-correction): “It is a common observation that the character of a people is in a great measure influenced by their local situation and the nature of the scenery in which they are placed; and it is impossible to visit the Tyrol without being convinced of the truth of the remark.”

In addition, however, to the detriment of implementing this technique wholesale, the OpenAI solution also will over-correct errors which it perceives based on its training not matching Romantic-era periodical styles.



Associations with “lawyer” by ranking (50=highest association),  
Databrick notebook developed in collaboration with Corey Brady



## Next Steps and Objects

In many ways, this fellowship has been a crash course in possibilities, and throughout it, those potentials continue to expand, even as the possibility and even definition of results becomes increasingly complex.

The next steps are to continue working to improve the OCR correction, and to develop Notebooks that are well annotated, and so can be used by both myself and others in exploring British Periodicals.

By using SQL to identify civil cases with small verdicts, I want to explore the arguments and findings presented in the periodicals: how do they position cases fought not for monetary judgment, but rather the claim of being right as a subject before the law.

Ultimately, I hope to find ways of engaging the periodicals in each of the five chapters I envision for the book:

- “Celebrity Lawyers and the Performance of Legal Mastery,” will explore both pictorial and written representations of lawyers as a profession and specific celebrities such as Thomas Erskine, with particular attention to the public circulation of closing arguments and cross-examinations.
- “Duels and the Violent Horizon of Law” will consider the press’s engagement with dueling and the court’s treatment with it.
- “The Jury and the Representational Public” will explore an array of representations of juries and non-jury trials in both fiction and periodicals, to suggest continuities and resistances between the jury and other forms of the represented public, such as periodical readership.
- “Farthing Cases: The Truth’s the Point” considers the rise in civil cases adjudicated for small sums of money, and the motivations highlighted in the coverage of such cases.
- The final chapter, “Property and Propriety,” considers both the law of entail and the laws of marriage, as they circulate in the periodicals and other print forms, as, alternately, mythic and transhistorical and then, topical and of recent invention.

## Influences

Algee-Hewitt, Mark et al. “Canon/Archive: Large-scale Dynamics in the Literary Field” *Literary Lab, Pamphlet II*. (Jan. 2016) 1-13.

Langbein, John. *The Origins of Adversary Criminal Trial*. Oxford University Press, 2005.

Mandell, Laura, *Breaking the Book: Print Humanities in the Digital Age*. Blackwell, 2015.

Piper, Andrew, *Enumerations: Data and Literary Study*. University of Chicago: 2018.

Underwood, Ted, *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago, 2019.

## Acknowledgements

This work was funded by a generous Andrew W. Mellon Fellowship for the Digital Humanities. Additional support was provided by Databricks and the Heard Library. I also thank my colleagues at the Center for Digital Humanities, especially Madeleine Casad, Lynn Ramey, and Debbie Brubaker; my Buchanan compatriots, Cliff Anderson, Corey Brady, and Brian Broll, as well as the undergraduate fellows; my data-scientific or dancing children, Joshua, Michael, and Diana Schoenfield; and Steve Baskauf, who never saw a data problem or a tree he didn’t like.