Shakespeare *at Scale*

CONVENORS

Anupam Basu
Washington University in St Louis

Brett Greatley-Hirsch
University of Leeds

PARTICIPANTS

Paul Brown
De Montfort University

Douglas Bruster
University of Texas at Austin

Jason Cohen
Berea College

Mark Dahlquist
Miami University

Mel Evans
University of Leicester

Jonathan Hope
Arizona State University

Jakob Ladegaard
Aarhus University

Nova Myhill
New College of Florida

Ed Pechter
University of Victoria

Carl G. Stahmer
University of California, Davis

Elizabeth E. Tavares
Pacific University

Christopher N. Warren
Carnegie Mellon University

Michael Witmore
Folger Shakespeare Library

NOTE ON SUGGESTED READINGS

Copies of materials cited in the abstracts are available to download from
http://www.notwithoutmustard.net/saa2019/readings/

Rates of Collaboration in the Early Modern Theatre

Paul Brown (De Montfort University)

ABSTRACT

The most frequently found estimate for the rate of early modern collaborative drama – plays written by more than one playwright – is that it accounts for about 50% of early modern plays. Is that really the case? These estimates were made many decades ago by G. E. Bentley in *The Profession of Dramatist in Shakespeare's Time* (1971). Using data drawn from the published volumes of Martin Wiggins's *Catalogue of Early English Drama*, as well as online databases that contain similar information (on sites like the Database of Early English Playbooks, the Lost Plays Database, and the Non-Shakespearian Drama Database), this paper argues that we have been overestimating the rates of collaborative authorship by 100%.

By empirically examining the records of early modern plays with respect to authorship, this paper argues that the actual rate of collaboration was 25%, half of what we previously thought. This new information allows us to adjust previous assumptions about collaboration. Though it effectively halves the collaborative output of the period, its importance should not diminish accordingly. The number of collaborative plays on the professional stage still runs into the hundreds. But, beyond the field of authorship attribution that tackles who wrote what, we can currently say relatively little about them.

Taking plays and their authorship wholesale allows a macroscopic view of playwriting in the period and allows us to adjust our assumptions about rates of collaborative writing, an adjustment we could not make without such an approach. Knowing how much collaborative writing there was matters to our understanding of the period's theatre.

SUGGESTED READINGS

Hirschfeld, Heather Anne. *Joint Enterprises: Collaborative Drama and the Institutionalization of the English Renaissance Theater.* Amherst: U of Massachusetts P, 2004.

Jackson, MacDonald P. "Collaboration." Arthur F. Kinney (ed.), *The Oxford Handbook of Shakespeare.* Oxford: Oxford UP, 2012. 31–52.

Taylor, Gary. "Collaboration 2016." Dympna Callaghan and Suzanne Gossett (eds), *Shakespeare in Our Time.* London: Bloomsbury Arden Shakespeare, 2016. 141–49.

Shakespeare's Words at Scale

Douglas Bruster (University of Texas at Austin)

ABSTRACT

This essay will revisit the "rare words" approach to Shakespeare's trajectory as a writer. Such scholars as Eliot Slater (1988) and MacDonald P. Jackson, among others, have examined the patterns of words used rarely in Shakespeare's works – words appearing, for example, in two or more works, and from two to eleven times total. The assumption behind this mode of inquiry is that certain words stuck for a time in Shakespeare's mind; they were therefore re-used at a higher rate in chronologically adjacent works before being forgotten (becoming, that is, practically inaccessible as he composed new works using different words). Thus, to find a word appearing in, say, both *The Comedy of Errors* and *Richard III*, but in few other works, would seem to come from (and confirm) what we assume to be those two plays' chronological proximity.

Because they are based on Slater's data, analyzing his figures, Jackson's essays on this topic have pushed Slater's findings about as far as they can go. My essay presents the results of an independent treatment of vocabulary data using (1) versions of the plays that reflect the most current arguments concerning attribution (thus, pared-down versions of *1 Henry VI*, *The Two Noble Kinsmen*, and so forth); and (2) modifications that account for the influence of genre – which scholars commonly acknowledge as an influence on Shakespeare's selection of vocabulary.

SUGGESTED READINGS

Craig, Hugh, and Brett Greatley-Hirsch. *Style, Computers, and Early Modern Drama: Beyond Authorship.* Cambridge: Cambridge UP, 2017.

Jackson, MacDonald P. "Vocabulary Links Between Shakespeare's Plays as a Guide to Chronology: A Reworking of Eliot Slater's Tables." *Shakespeare* 11.4 (2015): 446–58.

Slater, Eliot. *The Problem of* The Reign of King Edward III: *A Statistical Approach.* Cambridge: Cambridge UP, 1988.

Spevack, Marvin. "Shakespeare Microscopic and Panoramic." *Mosaic* 10.3 (1977): 117–27.

Zero Degrees of Francis Bacon:
Citation as Value Generator in Bacon's Works

Jason Cohen (Berea College)

ABSTRACT

This essay examines the citation network of Francis Bacon's work in order to propose a "zero degree" of Francis Bacon. Bacon cites important classical and contemporary authors, but even more crucially, he cites his own earlier works and revises them across the corpus. By suggesting the value and significance of internal citation networks across a diverse body of works, this essay argues for the generative and speculative status of the citation within and among Bacon's works. I consider, additionally, the status of the latent archive: not digital in its inception, this internal corpus of citations remains similarly abstract, distributed, and non-linear as the forms we see in our digital archival moment.

What then can be drawn from this body of citations as they populate the works and cross them over discipline, genre, and time? By reframing the inquiry into Bacon as a question of *citation* and *text-use*, the work here attempts to reframe, at least in a small way, one of the central debates in the tension between Bacon's natural and political systems; perhaps, this essay concludes, the politics of learning remain even more centrally at stake than the politics of nature.

SUGGESTED READINGS

Azoulay, Ariella. "Archive." *Political Concepts* 1.1 (2016).

Burdick, Anne. "Meta! Meta! Meta! A Speculative Design Brief for the Digital Humanities." *Visible Language* 49.3 (2015): 13–33.

De Kosnik, Abigail. *Rogue Archives: Digital Cultural Memory and Rogue Fandom*. Cambridge: MIT P, 2016. (Esp. Introduction and Chapter 1.)

'Substance of a Doubt':
Hamlet, Vector Modelling, and the Scandal of the Digital Humanities

Mark Dahlquist (University of Miami)

ABSTRACT

During the long English Reformation, few English words underwent greater change, or were the subjects of more controversy, than the term "scandal," which for about a century following the publication of the 1583 Douay-Rheims New Testament acquired a distinctive theological and political significance. I have argued elsewhere that this word's largely forgotten doctrinal sense can resolve the notorious textual crux appearing in *Hamlet* (Q2) at 1.4.36-38, which concludes with the word "scandal," uttered by Prince Hamlet, at the moment he first encounters his father's ghost.

This essay considers the capacity of vector modeling and machine learning visualization tools (as exemplified by the Word2vec and t-SNE algorithms) to address traditional hermeneutical and textual challenges, such as those posed by this puzzling crux. While n-gram frequency plots of the word "scandal" demonstrate the term's sharply growing popularity around 1604, as well as a remarkable spike in frequency at the outset of the British Civil Wars, this essay demonstrates that machine-learning algorithms can visualize the changing meaning of terms such as "scandal" in a granular and relative way that can, in conjunction with more familiar lexical and philological tools (such as the OED), support the close-reading of particular passages of text.

While sixteenth century religious controversialists affirmed their own doctrinal orthodoxy, and accused their opponents of innovation, machine learning tools can nevertheless help map lexical and conceptual change in the period. In the case of scandal, the use of these tools illuminates a concept central to Reformation-era discussions of religious idolatry and, more specifically, to the role of material things in matters of human ethics and perception. Thus, this paper illustrates the application of machine learning tools to a traditional interpretive question and does so with an interest in exploring how Reformation-era notions such as scandal provide a groundwork for understanding conceptions of materiality often associated with DH, such as distributed cognition, object-oriented ontology, and other putatively new materialisms.

SUGGESTED READINGS

Dahlquist, Mark. "*Hamlet* and the Snare of Scandal." *Shakespeare Quarterly* 69.3 (2018): 167–87.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." ArXiv:1605.09096 [cs], May 29, 2016.

Mauro, Aaron. "'To Think a World without Thought': Negotiating Speculative Realism in a Digital Humanities Practice." *Digital Studies/Le Champ Numérique* (2014).

Mohr, John W., Robin Wagner-Pacifici, and Ronald L. Breiger. "Toward a Computational Hermeneutics." *Big Data & Society* 2.2 (2015): 1–8.

Spierling, Karen. "'Il Faut Éviter Le Scandale': Debating Community Standards in Reformation Geneva." *Reformation & Renaissance Review* 20.1 (2018): 51–69.

Authors, Genre and Time in Early Modern Drama:
A Case Study of Aphra Behn

Mel Evans (University of Leicester)

ABSTRACT

The precise nature of the relationship between authorial style, genre and chronology is an essential question in the investigation of early modern dramatic language. Computational stylistics offers new ways to document the tri-directional relationship between these facets of dramatic style, due to its ability to identify large-scale linguistic patterns across diverse datasets (e.g. Craig and Greatley-Hirsch 2017; Rybicki 2016). My paper explores this hypothesis using findings from the AHRC-funded project 'Editing Aphra Behn in the Digital Age'. The project gathers new evidence of Behn's authorial style, *vis a vis* the style of her contemporaries and early modern predecessors, using computational stylistic techniques; and uses these findings to evaluate texts of dubious authorship. I outline some of the challenges, and opportunities, that arise in a study of a generically diverse writer with a long career, focussing on two case studies: the dating of *The Young King* (see Evans 2018) and the collaborative authorship of Behn's posthumous play *The Younger Brother.* The findings suggest that computational stylistic techniques provide a valuable means of documenting how authorial style, genre and chronology intersect at the macro-level. However, I propose that the descriptive prowess of the computational stylistic approach should be complemented with micro-level perspectives, drawing on related fields such as historical pragmatics, to ensure interpretative rigour: to lead us closer to a 'why', as well as a 'what'. Combining the quantitative and the qualitative – reading the small words as words, as well as numbers – is a vital part of tracking the evolution of early modern dramatic style.

SUGGESTED READINGS

Craig, Hugh, and Brett Greatley-Hirsch. *Style, Computers, and Early Modern Drama: Beyond Authorship.* Cambridge: Cambridge UP.

Evans, Mel. "Style and Chronology: A Stylometric Investigation of Aphra Behn's Dramatic Style and the Dating of *The Young King.*" *Language and Literature* 27.2 (2018): 103–32.

Rybicki, Jan. "Vive La Différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies." *Digital Scholarship in the Humanities* 31.4 (2016): 746–61.

Inheritance and Social Networks in Three Seventeenth-Century Comedies

Jakob Ladegaard (Aarhus University)

ABSTRACT

This paper studies social networks in three early modern comedies with shared plot features. The plays all revolve around an inheritance conflict between an uncle and his prodigal nephew and heir. The plays in question are Thomas Middleton's *A Trick to Catch the Old One* (c. 1605), Philip Massinger's *A New Way to Pay Old Debts* (c. 1625) and Richard Brome's *A Mad Couple Well Matched* (c. 1639). The study compares the ways in which the heirs in competition with their uncles negotiate two interwoven social networks to regain their inheritance; an economic network of credit relations and a network of amorous alliances.

To do that, we use social network analysis. We first compare the overall network metrics of the plays to understand the basic structure of the economic and amorous networks in question. We then focus on the relationships between heirs and uncles, using two types of centrality measures (metric measures and count measures) to analyze their relative centrality in the networks. Finally, we expand the comparison to encompass key actors in the economic and amorous networks to see what this reveal about the relative importance of the two networks. We find that the plays differ widely in size and density and discuss the meaning of this for their treatment of inheritance. We also find that betweenness centrality as well as relative scores on metric measures and count measures help explain relations between uncles and heirs. And finally, it seems that the metric centrality of female characters is connected to the relative importance of the amorous network in the plays. The work presented in this paper is part of the larger research project 'Unearned Wealth: A literary history of inheritance, 1600–2015', where we use digital methods to study representations of inherited wealth in English literature from the early modern period until now.

SUGGESTED READINGS

Algee-Hewitt, Mark. "Distributed Character: Quantitative Models of the English Stage, 1550–1900." *New Literary History* 48.4 (2017): 751–82.

Fischer, Frank, *et al.* "To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930)." *DH 2018* (2018): 193–200.

Masías, Victor Hugo, *et al.* "Exploring the prominence of *Romeo and Juliet*'s characters using weighted centrality measures." *Digital Scholarship in the Humanities* 32.4 (2017): 837–58.

Visualizing the Repertory

Nova Myhill (New College of Florida)

ABSTRACT

The most complete information we have on patterns in the early modern theatrical repertory are the sections of Henslowe's Diary that record performances and income for various companies including Strange's, Queen's, Sussex's, the Lord Chamberlain's and the Lord Admiral's in the 1590s. While this information has been available since its discovery by Malone in the 18th century, it is difficult to work with in aggregate, and has more frequently been considered in terms of information about individual plays or performance. For this paper, I plan to code that information to create data visualizations that I think can help us to understand some things about how the repertory functioned as a whole, particularly in terms of shifts in repertory, how old plays fell into or out of frequent production, and how new plays entered the repertory. While in the grand scheme of things, this is small-scale data, it offers a window on how quantitative analysis makes some forms of evidence newly legible and how varying methods of data aggregation allow one to tell significantly different stories from the same data. Finally, attempts to examine this particular data set invites consideration of how the availability of certain data sets both drives and limits the kinds of questions we can ask about theatre history.

SUGGESTED READINGS

Algee-Hewitt, Mark. "Distributed Character: Quantitative Models of the English Stage, 1550–1900." *New Literary History* 48.4 (2017): 751–82.

Knutson, Roslyn L. "Repertory System." Arthur F. Kinney (ed.)., *The Oxford Handbook of Shakespeare*. Oxford: Oxford UP, 2012. 404–19.

Weingart, Scott B. "Demystifying Networks, Parts I & II." *Journal of Digital Humanities* 1.1 (2011).

Can Quantitative Analysis Go Beyond Authorship?

Ed Pechter (University of Victoria)

ABSTRACT

Brett and Anupam cut us lots of slack. Engage any topic in the quantitative study of Shakespeare, adopting any approach, with only one exception: no authorship attribution. This is like *Areopagitica* – it's all good, as long as it's not Papistry – but for different reasons. Milton excluded Catholics from the free market of ideas because he thought Catholics excluded everyone else. For Anupam and Brett, attribution is problematic not because it is hostile to the values of their enterprise, but because it has instantiated them with such extraordinary success. The stunning achievements in attribution have established a substantial position for quantitative analysis (QA) in Shakespeare studies, but more attribution might now reinforce a tendency to sequester the method within a single topic – a kind of one-trick pony. Better to go *beyond authorship*, extending the reach of QA into territories so-far unexplored.

I infer all this from Hugh Craig and Brett Greatley-Hirsch's (CG's) *Style, Computers, and Early Modern Drama: Beyond Authorship*, whose subtitle is incorporated into my title and reiterated at the end of the preceding paragraph, and whose Introduction develops along the lines sketched out above. That "Authorship has been the main focus of computational stylistics in early modern drama to date" is the situation they set out to correct: "our aim is to build on the striking advances … in authorship attribution and apply similar methods to other aspects of literary history." Even as CG boldly go where few digital humanists have gone before, they affirm continuity with previous work, procedurally if not thematically, using "similar methods" to those that have served attribution scholars in the past. This assumption – that the quantitative methods underwriting attribution are transferable to "wider stylistic contexts" beyond authorship – is what I'm looking at here.

SUGGESTED READINGS

Craig, Hugh, and Arthur F. Kinney (eds). *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge UP, 2009.

———, and Brett Greatley-Hirsch. *Style, Computers, and Early Modern Drama: Beyond Authorship*. Cambridge: Cambridge UP, 2017.

Dreyfus, Hubert L. *What Computers Can't Do: The Limits of Artificial Intelligence*. rev. ed. New York: Harper, 1979.

Liu, Alan. "Where is Cultural Criticism in the Digital Humanities?" Matthew K. Gold (ed.)., *Debates in the Digital Humanities*. Minneapolis: U of Minnesota P, 2012. 490–509.

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

The Range of Discourse:
Mapping the Semantic Landscape of Early Modern Popular Print

Carl G. Stahmer (University of California, Davis)

ABSTRACT

Tessa Watt notes that broadside ballads inhabited a "shared culture" of print that reach across social and geographic boundaries in early modern Europe, a trait that the 17th century bibliophile Samuel Pepys described as "promiscuous." As such, as a body of literature they provide a window into the range of early modern discourse: which discursive topics circulated through which paths and in what combinations both with each other and with other symbolic objects (images and tunes). To date, however, we lack a comprehensive, macro understanding of this complexity. For example, treatments of the interplay of discourses of politics and gender are frequent in the scholarly record, but to what extent are these topics truly correlated when analyzed across the body of extant works from the period?

This paper will report on a study that employed computational methods to identify and define the range of semantic topics that appear in the English Broadside Ballad Archive (EBBA) and to derive a macro-scale map of topic, image, and tune reuse and combination. The EBBA collection consists of nearly 9,000 broadside ballads printed in English in the 16th and 17th century, comprising nearly 90% of all extant broadside ballads from the period, each fully transcribed and thickly catalogued. As such, it presents a model corpus for this type of analysis. Methods that were employed in the analysis and that will be discussed include Topic Modelling with Latent Dirichlet Allocation, the application of content-based image recognition software to track the reuse of images across ballads, the use of acoustic analysis and comparison of Music Encoding Initiative (MEI) representations of tunes to track tune re-use, and statistical analysis to determine the frequency of co-occurrence both within and across the classification categories.

SUGGESTED READINGS

Fumerton, Patricia. "Digitizing Ephemera and Its Discontents: EBBA's Quest to Capture the Protean Broadside Ballad." Kevin D. Murphy and Sally O'Driscoll (eds), *Studies in Ephemera: Text and Image in Eighteenth-Century Print.* Lewisburg: Bucknell UP, 2013. 55–97.

Fumerton, Patricia, Carl Stahmer, Kris McAbee, and Megan Palmer Browne. "Vexed Impressions: Towards a Digital Archive of Broadside Ballads Illustrations." Brent Nelson and Melissa Terras (eds), *Digitizing Medieval and Early Modern Material Culture.* Tempe: ACMRS, 2012. 257–85.

Koehl, Arthur, and Carl Stahmer (devs). *Archv.* https://github.com/cstahmer/archv

Embedded Lives:
Shakespeare, Literature, and the *Oxford Dictionary of National Biography*

Christopher N. Warren (Carnegie Mellon University)

ABSTRACT

The *Oxford Dictionary of National Biography* (ODNB), first published in 2004, is of course a monumental resource for historians, critics, and many other scholars whose work intersects with British history. While Shakespeareans frequently consult the ODNB, few are aware of Shakespeare's symbolic significance to the entire enterprise. Of Oxford University Press's twelve internal categories for commissioning and editing biographies, only one category is named after a single individual. Alongside blocks for Medieval, Tudor, Stuart, Hanoverian and Victorian biographies, there exists an entire category of biographies commissioned and organized under the heading "Shakespeare." "Shakespeare" isn't one person: like Henry IV, he has "many marching in his coats" – 1,781 people to be exact. For this paper, I have trained a word embedding model on all of the Shakespeare biographies put together while also modeling each of the other 11 main ODNB categories. I intend to discuss what distinguishes the ODNB's corporate "Shakespeare" from those subsumed under the Dictionary's competing organizing categories.

SUGGESTED READINGS

Almeida, Felipe, and Geraldo Xexéo. "Word Embeddings: A Survey." ArXiv:1901.09069 [CS, Stat], January 25, 2019.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings." ArXiv:1607.06520 [CS, Stat], July 21, 2016.

Gavin, Michael. "Vector Semantics, William Empson, and the Study of Ambiguity." *Critical Inquiry* 44.4 (2018): 641–73.

Heuser, Ryan James. "Word Vectors in the Eighteenth Century." *DH* 2017.

Schmidt, Ben. "Vector Space Models for the Digital Humanities." *Bookworm* (blog), October 25, 2015. http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html.

Warren, Christopher N. "Historiography's Two Voices: Data Infrastructure and History at Scale in the *Oxford Dictionary of National Biography* (ODNB)." *Journal of Cultural Analytics*, November 22, 2018.

A Map of Early English Texts

Michael Witmore (Folger Shakespeare Library)
Jonathan Hope (Arizona State University)

ABSTRACT

In this paper we try to map the whole of TCP – our metric is Docuscope LATs, and the statistical analysis is PCA. There are two main questions.

The first is a theoretical one: does it make any kind of sense to map the variation across 60,000 texts? That is, does linguistic variation viewed at this scale tell us anything about the cultural forms that produced the texts in TCP? Our initial findings are that the results are, to us at least, interpretable, in ways that are not banal – which seems to justify the application of analytic techniques at this scale and may suggest new ways of approaching the history of writing in English.

The second question is a methodological one: without consistent genre metadata for the texts in TCP, how can we make sense of where our 60,000 dots end up in PCA-space? We try to address this through a two-stage process. First via an unsupervised analysis which seeks to characterise the PCs and spaces independently of any human-ascribed genre labels; then we read onto these spaces the results of a reduced study of 1,080 TCP texts which do have human-generated genre metadata. The results of this study suggest that the unsupervised procedure identifies human-understandable patterns which map coherently and consistently onto genres and types of writing.

SUGGESTED READINGS

Hope, Jonathan, (ed.). "The Great Work Begins: EEBO-TCP in the Wild." *Wine Dark Sea*, March 25, 2016. http://winedarksea.org/?p=2627

Witmore, Michael. "Finding 'Distances' between Shakespeare's Plays 1." *Wine Dark Sea*, June 23, 2015. http://winedarksea.org/?p=2225

Witmore, Michael. "Finding 'Distances' between Shakespeare's Plays 2: Projecting Distances onto New Bases with PCA." *Wine Dark Sea*, July 6, 2015. http://winedarksea.org/?p=2271