

Hulse, Mark C. "Determining Authorship: A New Technique."

Citation details

Early Modern Digital Review, vol. 4, no. 4, 2021, <https://doi.org/10.33137/rr.v44i4.38648>.

Renaissance and Reformation / Renaissance et Réforme, vol. 44, no. 4, 2021, pp. 204-210, <https://doi.org/10.33137/rr.v44i4.38648>.

Peer review

This is a peer-reviewed article in *Early Modern Digital Review*, distributed in print by *Renaissance and Reformation / Renaissance et Réforme*.

Copyright

Early Modern Digital Review materials are published under a Creative Commons 4.0 license (CC BY 4.0) that permits the right to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) the material, provided that the author and source are credited. The full description of CC BY 4.0 can be found on creativecommons.org/licenses/by/4.0/.

Early Modern Digital Review

Early Modern Digital Review is an online, open-access, and refereed journal publishing high-quality reviews of digital projects related to early modern society and culture. It is committed to productive evaluation of both established digital resources and recent tools and projects. Its publications are distributed online by the journal and its partners, and in print by *Renaissance and Reformation / Renaissance et Réforme*.

That match shows how InfoRapid can cope with variant spellings. It is a flexible tool and can cope with vast amounts of text: 999 pages, the website claims. I warmly recommend it.

BRIAN VICKERS

Institute of English Studies, London

<https://doi.org/10.33137/rr.v44i4.38647>

Works Cited

- Foakes, R. A., and R. T. Rickert, eds. 1961. *Henslowe's Diary*. Cambridge: Cambridge University Press.
- Harbage, A., and S. Schoenbaum. 1964. *Annals of English Drama, 975–1700*. Philadelphia: University of Pennsylvania Press.
- Vickers, Brian, ed. Forthcoming. *The Complete Works of Thomas Kyd*. 2 vols. Suffolk: Boydell and Brewer.
- Vickers, Brian. 2007. "Incomplete Shakespeare: Or, Denying Co-authorship in *1 Henry VI*." *Shakespeare Quarterly* 58:311–52.
- Wiggins, Martin, and Catherine Richardson. 2012–. *British Drama 1533–1642*. Oxford: Oxford University Press.

Jackson, MacDonald P., researcher.

"Determining Authorship: A New Technique." Method using LION. *Research Opportunities in Renaissance Drama* 41 (2001):1–14.

Accessed 7 March 2022.

LION: about.proquest.com/en/products-services/literature_online.

Highlighting verbal similarities between texts has always featured prominently in attribution research, and parallel passages sometimes resonate better than statistical readouts of function word usage, graphs of metrical patterns, or other technical features that require mining for data at a different level than one typically experiences in literary works. Still, the more arcane mathematical approaches have multiplied in the last two decades and are unlikely to diminish in importance. In a 2001 article, MacDonald Jackson presented one possible method for bridging the quantitative-qualitative divide via comprehensive and methodical searching for

more subtle lexical combinations within disputed segments of texts.¹ This is made possible by electronic databases that can house nearly all extant printed works from the early modern period. The following reflection on Jackson's experiments, and other critics' variations on the underlying premise, offers guidelines for applying such techniques in future research.

Exhaustive listing of verbal echoes can create overwhelming amounts of data when the threshold is low for identified correspondences, but if we measure relevance as a function of rareness it is possible to pare down results. Jackson's 2001 study, which outlined his technique, divided two excerpts from *Titus Andronicus* into searchable terms: seventeen lines by Peele (1.1.1–17) and twenty by Shakespeare (2.3.10–29). He then fed combinations of “words, phrases, and collocations” into the Literature Online (LION) database while making use of what proximity operators its search engine allowed (such as “near” or “followed by”), counting hits when the canon of Shakespeare or Peele (but not both) contained the same combination.² Peele earned the majority of the unique hits in the opening seventeen lines as predicted (10 to 2), while Shakespeare won the 2.3 passage handily (7 to 2).³

Jackson used this promising technique to great effect in corroborating the split authorship of the final scene from the Webster/Rowley/Heywood collaboration *A Cure for a Cuckold*, selecting representative works from all three playwrights and confirming divisions hypothesized by earlier bibliographers.⁴ He reiterated its applicability to Shakespeare in a 2003 book on *Pericles*, finding affinities between the works of George Wilkins and the opening acts of the late romance.⁵ Gary Taylor explored another combination of hands in 2002, endorsing a Middleton/Rowley/Heywood collaboration in *The Old Law*, and has brought the method to bear on Shakespearean studies in his analysis of the suspected Middleton interpolations in *Macbeth*.⁶

In a recent article I challenged an influential argument founded on Jackson's method—William Weber's 2014 claim that *Titus Andronicus* 4.1

1. Jackson, “Determining Authorship.”

2. Jackson, “Determining Authorship,” 6–7.

3. Jackson, “Determining Authorship,” 9–12.

4. Jackson, “Late Webster and His Collaborators.”

5. Jackson, *Defining Shakespeare*.

6. See Taylor, “Middleton and Rowley – And Heywood” and “Empirical Middleton.”

was by Shakespeare—and noted some problems with this manner of parallel-hunting.⁷ However, I would not deny the method’s potential if proper steps were followed, the first essential component being mathematically-balanced canons that give prospective authors their fair chance to demonstrate similarities. If one candidate has a vastly smaller corpus, it is necessary to restrict other canons or adjust the weight of hits accordingly to compensate for the discrepancy. Weber’s study of *Titus* 4.1 chose a handful of Shakespearean works to approximate Peele’s extant writings, but miscalculations resulted in a “restricted canon” that was 85 percent longer than it should have been.⁸ Unfortunately, much blame can be traced back to Jackson, whose 2001 article contained a similar discrepancy in an uncharacteristic oversight.⁹ Doubts may persist about how genre, chronology, and other factors may impact the precise idiolect on display in rival canons, but parity in length is indispensable.

Those attempting to compensate for different canon sizes must also exercise caution, especially when few extant works exist from a candidate. With only brief samples available for comparison, there is the increased danger of a tested passage returning zero unique parallels despite common authorship, or of similarities being exaggerated when multiplied by a factor calculated from the ratios of canon sizes. Limited bodies of (non-disputed) dramatic works has plagued the study of Elizabethan authors like Robert Wilson, Thomas Nashe, Thomas Kyd, and others. But the hazards of compensating for a small canon with questionable multipliers is evident in Gary Taylor’s claim that lines from a scene of *Arden of Faversham* score hits with the poetry of Thomas Watson.¹⁰ Precious little writing survives from Watson, thereby inflating the importance of what might be accidental correspondences. Taylor defends the ascription to Watson on other grounds, but any argument founded on database-searching should be taken with a grain of salt given these sample size limitations.¹¹

Next, there is the persistent dilemma of how to employ disputed texts in ascribing other disputed works, especially when so many plays were printed

7. Weber, “Shakespeare After All?” For my critique, see Hulse.

8. See Hulse, 863. The same error is duplicated in Pruitt, 95n.

9. Hulse, 863–64. Jackson had used a canon of seven Shakespeare works rather than the eight used by Weber and Pruitt, but still gave him a canon that was 60 percent larger than Peele’s.

10. Taylor, “Finding ‘Anonymous’ in the Digital Archives.”

11. For his expanded case in favour of Watson as part author of *Arden*, see Taylor, “Shakespeare, *Arden of Faversham*, and Four Forgotten Playwrights.”

anonymously circa 1590. My critique of the Weber and Pruitt studies took issue with their excluding *The Troublesome Reign of King John*—a play that I think has been convincingly ascribed to Peele but that was not credited as such in the LION database. No online database carves up and flags Shakespeare's sections of his collaborative plays for easy reference, even where consensus exists. Weber admitted into his restricted canon the likely co-authored *The Taming of the Shrew*, and even quarto variants from the 1597 text of *Romeo and Juliet* whose provenance has divided scholars. Jackson searched the “quarrel scene” (3.5) of *Arden of Faversham* and highlighted numerous Shakespearean links compared with other plays acted between 1580 and 1600, but his data suffers from not itemizing hits from the *Henry VI* plays, *Edward III*, *Titus*, or *The Shrew* by suspected author.¹² There is no easy solution to such difficulties other than for critics to be transparent and forthcoming about their decisions.

Open questions remain about the method's effectiveness at recognizing authorial styles irrespective of imitation, adulteration, or subject matter. Many stylometric traits operate at a level that authors may not realize, minimizing the chance that they could adapt to poetic or rhetorical trends. Verbal echoes, on the other hand, may hinge on conscious attempts to emulate others' habits. Things get further complicated if texts have been revised by a second author or corrupted in transmission. Database searching has yet to untangle a messy work such as *3 Henry VI*, which exists in multiple versions and likely shows some combination of corruption at the printshop and a multi-phase composition involving different playwrights working to complete a story with established characters.¹³ Testing a healthy smattering of passages from across the *Henry VI* plays might advance our understanding of composition history, but the work would be tedious and complicating factors might interfere with the data.

So far, the most evident successes for Jackson's method have involved head-to-head contests between Jacobean playwrights with healthy canons, especially when the relevant parties were known to collaborate, or external evidence implicated them (Wilkins with *Pericles*, or Middleton with *Macbeth*). Less convincing is an experiment that pits Shakespeare against straw-man opponents, such as Taylor does in proposing that unique Shakespearean parallels in the Gertrude-Horatio scene from Q1 *Hamlet* support the “early

12. Jackson, *Determining the Shakespeare Canon*, 17–24. In this experiment Jackson counted combinations that appear five times or fewer in the pool of tested plays rather than unique hits.

13. See, for example, Martin.

draft” theory of the 1603 quarto.¹⁴ (While Taylor’s study chiefly foregrounds n-grams, he includes hits that require substitutions or non-adjacent word strings benefitting Shakespeare.¹⁵) There is no rival playwright whose canon competes for hits in such an experiment, the prevailing alternate hypothesis being that Q1 was cobbled together from remembered lines such that Shakespearean echoes might be expected to predominate. A candidate author must have a testable corpus of works, and differentiating revision from corruption falls outside such a method’s capabilities.

Perhaps the most important difference between persuasive and unpersuasive experiments in this vein is the presence or absence of qualitative follow-up. Weber’s study of *Titus* 4.1 assembled a long list of hits, but the individual correspondences prove quite unimpressive when examined individually, usually requiring common pronouns, prepositions, or function words.¹⁶ Compare this with Jackson’s links between Peele’s works and the opening lines of the play, which show noteworthy usages of rarer terms (e.g., *consecrate*, *diadem*, *ware*, etc.).¹⁷ And while the landslide 7:2 ratio in favour of Shakespearean hits that Jackson found in Tamora’s 2.3 speech is weakened when canons are equalized, there are decidedly stronger connections to Shakespeare’s works than to Peele’s (e.g., “nurse’s song,” the shrill echoing of the hounds, or the image of the “rolled” snake that *2 Henry VI* shares).¹⁸ Interestingly, the discovered echoes largely depend upon hunting scenes in *Venus and Adonis* or *The Shrew*; had the subject matter of the selected passage been different, the correct author (Shakespeare) might not have scored the victory numerically. Even so, the virtue of qualitative meta-analysis is that we can recognize interesting parallels and minimize incidental ones. Weber eschews this important step, as does Taylor in his Q1 *Hamlet* study, and few of their hits reflect an idiolect that is distinctive.

14. Taylor, “Shakespeare’s Early Gothic *Hamlet*.”

15. Taylor, “Shakespeare’s Early Gothic *Hamlet*,” 20–23. Specifically, Taylor freely substitutes the terms *conference* and *conversion* for a unique hit with *Richard II* and allows an inexact sequence of words in *3 Henry VI* to act as another.

16. Hulse, 871. Weber’s list of unique hits for Shakespeare and Peele is not affixed to his article, instead appearing on a personal website (see Works Cited).

17. Jackson, “Determining Authorship,” 8–10.

18. Jackson, “Determining Authorship,” 10–11.

Despite the hazards involved, searchable databases have revolutionized traditional parallel hunting by alleviating the bias that our greater familiarity with Shakespeare's works produces. Control passages should be sampled liberally to establish baseline expectations, and levelled canons are a must. Criteria should also be made explicit regarding what sorts of words qualify as meaningful, keeping experiments replicable by others and, ideally, reducing results to manageable numbers of stronger correspondences. Properly performed, comprehensive searching can complement traditional qualitative analysis, though interpreting data relies upon willingness to assess the garnered hits at a deeper level, and both authors and readers must do their part.

MARK C. HULSE

Jackson College

<https://doi.org/10.33137/rr.v44i4.38648>

Works Cited

- Hulse, Mark. 2021. "The Case for Peele's Authorship of *Titus Andronicus* 4.1: Cross-Examining Attribution Methods in a Disputed Scene." *The Review of English Studies* 72 (2021):860–81. doi.org/10.1093/res/hgab032.
- Jackson, MacDonald P. 2003. *Defining Shakespeare: "Pericles" as Test Case*. Oxford: Oxford University Press. doi.org/10.1093/acprof:oso/9780199260508.001.0001.
- Jackson, MacDonald P. 2001. "Determining Authorship: A New Technique." *Research Opportunities in Renaissance Drama* 41:1–14.
- Jackson, MacDonald P. 2014. *Determining the Shakespeare Canon: "Arden of Faversham" and "A Lover's Complaint"*. Oxford: Oxford University Press. doi.org/10.1093/acprof:oso/9780198704416.001.0001.
- Jackson, MacDonald P. 2001. "Late Webster and His Collaborators: How Many Playwrights Wrote *A Cure for a Cuckold*." *The Papers of the Bibliographical Society of America* 95 (3):295–313. doi.org/10.1086/pbsa.95.3.24304488.
- Martin, Randall. 2002. "*The True Tragedy of Richard Duke of York and 3 Henry VI*: Report and Revision." *The Review of English Studies* 53 (209):8–30. doi.org/10.1093/res/53.209.8.
- Pruitt, Anna. 2017. "Refining the LION Collocation Tests: A Comparative Study of Authorship Test Results for *Titus Andronicus* Scene 6 (= 4.1)."

- In *The New Oxford Shakespeare: Authorship Companion*, edited by Gary Taylor and Gabriel Egan, 92–106. Oxford: Oxford University Press.
- Taylor, Gary. 2014. “Empirical Middleton: Macbeth, Adaptation, and Micro-authorship.” *Shakespeare Quarterly* 65 (3): 239–72. doi.org/10.1353/shq.2014.0030.
- Taylor, Gary. 2019. “Finding ‘Anonymous’ in the Digital Archives: The Problem of *Arden of Faversham*.” *Digital Scholarship in the Humanities* 34:855–73. doi.org/10.1093/llc/fqy075.
- Taylor, Gary. 2002. “Middleton and Rowley – And Heywood: *The Old Law* and New Attribution Technologies.” *The Papers of the Bibliographical Society of America* 96 (2):165–217. doi.org/10.1086/pbsa.96.2.24295710.
- Taylor, Gary. 2020. “Shakespeare, *Arden of Faversham*, and Four Forgotten Playwrights.” *The Review of English Studies* 71:867–95. doi.org/10.1093/res/hgaa005.
- Taylor, Gary. 2019. “Shakespeare’s Early Gothic *Hamlet*.” *Critical Survey* 31 (1/2):4–25. doi.org/10.3167/cs.2019.31010202.
- Weber, William. Personal Website. Accessed 2 Dec. 2021. williamweatherford-weber.wordpress.com.
- Weber, William. 2014. “Shakespeare After All? The Authorship of *Titus Andronicus* 4.1 Reconsidered.” *Shakespeare Survey* 67:69–84.

Rizvi, Pervez, creator.

Collocations and N-grams. Database.

London, 2017. Accessed 29 July 2021.
shakespearestext.com/can/index.htm.

Pervez Rizvi’s electronic corpus of 527 plays dated between 1552 and 1657, titled Collocations and N-grams, is an invaluable aid for researchers aiming to ascertain the authorship and chronology of early modern texts. Results of automated searches enable scholars to check for phrasal repetitions between plays of the period. Rizvi’s project, which is unfunded and not affiliated with any institution, is a gift to the scholarly community. Launched in 2017, it has already led to many fascinating discoveries concerning the dating of *Alphonsus, Emperor of Germany*;¹⁹ the possibility of Cyril Tourneur’s hand in *The Honest*

19. Jackson, “The Date of *Alphonsus, Emperor of Germany*.”