Renaissance and Reformation
Renaissance et Réforme

Craig, Hugh, director. Craig's Zeta test. Other

Ros Barber

Volume 44, numéro 4, automne 2021

URI : https://id.erudit.org/iderudit/1089359ar
DOI : https://doi.org/10.33137/rr.v44i4.38652

Aller au sommaire du numéro

Éditeur(s)

Iter Press

ISSN

0034-429X (imprimé)
2293-7374 (numérique)

Découvrir la revue

Citer ce compte rendu

Barber, R. (2021). Compte rendu de [Craig, Hugh, director. Craig's Zeta test. Other]. *Renaissance and Reformation / Renaissance et Réforme, 44*(4), 228–233. https://doi.org/10.33137/rr.v44i4.38652

proposed authorship hypothesis. I therefore find stylo() one of the most useful and useable programs.

THOMAS MERRIAM
Basingstoke, England
https://doi.org/10.33137/rr.v44i4.38651

Works Cited

Eder, Maciej, Joanna Byszuk, and Jan Rybicki. "Stylometry with Stylo." Online tutorial. Accessed 17 February 2022. computationalstylistics.github.io/stylo_nutshell/#.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. "Stylometry with R: A Package for Computational Text Analysis." *R Journal* 8 (1):107–21. journal.r-project.org/archive/2016/RJ-2016-007/index.html.

Craig, Hugh, director.
Craig's Zeta test. Other.
Newcastle, UK: Centre for Literary and Linguistic Computing (CLLC), University of Newcastle, 2009. Accessed 15 July 2021.
c21ch.newcastle.edu.au/ia.

The Zeta test has become a popular method, within computational stylistics, to determine the authorship of early modern texts. The Zeta test originated with John Burrows, but it is Hugh Craig's variant that has been most influential in early modern attribution studies.[1] This variant underpins several of the attributions in the New Oxford Shakespeare project.[2]

Zeta is a computational algorithm that is intended to detect a writer's style through determining which words they use more frequently compared with another writer or writers (or indeed themselves in a different period or working in a different genre). For example, an early successful application demonstrated clear shifts in vocabulary in Henry James's early and late styles.[3] The "style"

1. Burrows; Craig and Kinney.

2. Taylor and Egan.

3. Hoover.

of each author (as much as style can be measured by vocabulary choice) is determined by creating a list of "marker" words that are more commonly used in their texts than in the texts in a comparison dataset.

While running Zeta tests requires a degree of technical ability, an appreciation of the variable nature of early modern texts (and an understanding of the limits of our knowledge about their provenance) is essential to the careful interpretation of results. As Joseph Rudman has pointed out, early modern drama is not well-suited to data analysis of this kind.[4] Scientifically speaking, it is not possible to create valid controls for these tests, because we cannot be sure of the extent to which the texts we have are purely authorial. Even with the highly edited texts of Shakespeare's Folio, choosing only those we believe to be solo-authored (though we cannot know this for certain), apparent errors have been introduced that would influence Zeta results. Therefore, significant care must be put into experimental design and thoughtful validation procedures that will check unconscious assumptions.

So far, Zeta's application in authorship attribution, especially when it comes to early modern plays, has been problematic, with flawed test designs resting on unconscious assumptions, insufficient attention paid to the effects of small datasets and the influence of genre, and inappropriate or inadequate validation procedures.

Zeta test results are often misinterpreted. Zeta results are plotted on a graph, showing how many of the "marker words" of two different datasets (usually two different authors, one plotted on the X axis and one on the Y axis) exist in segments of a test text. As Pervez Rizvi points out, the separation of the two "author" datasets on the graph is an outcome of the design of the Zeta test, but some scholars have misunderstood Zeta sufficiently to treat it as though it is a research finding.[5] Much of the interpretation of Zeta tests so far has been through a method that creates a fallacious impression: a bisector line drawn through a point equidistant from the centre of each data cluster, a method "too crude to be reliable."[6] This has often led scholars to wrongly treat segments of the test text as having a verbal affinity to the author whose marker words

4. Rudman, "The State of Authorship Attribution Studies," and Rudman, "Non-Traditional Authorship Attribution Studies."

5. Rizvi, 403.

6. Rizvi, 406.

dominate that side of the graph. However, the full extent of each comparison data cluster needs to be taken into account, and a test segment should be deemed to have affinity with one side or the other by a statistically appropriate method such as standard deviation.[7] Even then, the fact that a text segment (or whole text) shows more affinity for one author than another cannot be taken as proof that it was written by that author without additional evidence and argument. A text could be written by neither of the tested authors and still have more in common with one author's vocabulary than another's for reasons of subject matter, theme, intended audience, literary fashion, a similar education, etc.

The Zeta algorithm is strongly influenced by genre, and the design of any Zeta test needs to acknowledge this. Because of its focus on words with high information content (so-called "lexical words" as opposed to "function words" such as prepositions, pronouns, etc.), the subject matter of texts can unduly influence results. For example, most of the top "Marlowe marker words" found in the Joan of Arc sections of *Henry VI Part 1* that led to Hugh Craig attributing them to Marlowe were words associated with warfare and were Marlowe marker words chiefly because his small canon is dominated by three battle-heavy plays, *Tamburlaine Parts 1* and *2 and Edward II*.[8] No control-style Zeta test has been done to discover whether battle scenes in other early modern plays have an affinity with these Marlowe marker words. In a demonstration of the influence of genre on the algorithm, Zeta could not detect *The Taming of the Shrew* as Shakespeare's when his markers words were derived from a dataset of his tragedies (with Marlowe's full canon as the comparison) but gave 92 percent of it to Shakespeare when the comparison dataset comprised Shakespeare comedies.[9] Any Zeta test design that isn't taking account of genre will generate invalid results.

The comparison datasets must be fair, balanced, and logical. A Zeta test can be designed with a wide number of variables which will lead to different style markers. Different marker words will arise from determining an author's style against all other writers of the period, or against a smaller subset of authors writing in the same genre, or against a single other author. Using a comparison set of 130 plays in all genres across six decades to define the marker words

7. Barber, "Big Data," 2.

8. Barber, "Big Data," 14.

9. Barber, "Big Data," 12.

of a single author with a small canon, as Craig did in his earliest work with Zeta, is not advisable.[10] Burrows and Craig found that Zeta gave high levels of accuracy for telling apart Shakespeare and Marlowe plays when Marlowe's canon was compared with a particular set of eight early Shakespeare tragedies and histories (excepting a highly anomalous result for Marlowe's *The Jew of Malta* which has to be explained away) but any variation in the make-up of that set reduces accuracy considerably. When used with a statistically valid interpretation method, the validation accuracy of these datasets is reduced, with plays such as *Anthony and Cleopatra*, *The Tempest, and Henry V* scoring as less than 75 percent Shakespeare's.[11] In addition, Zeta fails to recognize the majority of the text in the three supposedly co-authored *Henry VI* plays as either Shakespeare's or Marlowe's.[12] Therefore, even Zeta tests designed with seemingly effective datasets should undergo more extensive validation testing than is currently fashionable in order to understand their robustness.

The size of the comparison datasets (and their size relative to each other) needs careful management. Many early modern dramatists have canons of one hundred thousand words or fewer, and Zeta can be inaccurate with datasets of this size. In experiments with Shakespeare's canon, using a limited set of one hundred thousand words of his tragedies to generate marker words, the Zeta algorithm did not recognize *Hamlet* as Shakespeare's until the comparison dataset was matched for size, genre, and period (six tragedies by others from 1600 to 1605). Even then, only just over 60 percent of Hamlet fell within three standard deviations of the vocabulary range of Shakespearean tragedy. Other Shakespeare tragedies were also poorly recognized.[13] When additional validation tests were run with identically sized datasets of one hundred thousand words for each author, the Zeta algorithm was inaccurate more often than it was accurate.[14] It is therefore recommended that comparison datasets are larger than one hundred thousand words, and one a low multiple of the other, with significant validation testing to ensure that Zeta is accurate under these specific conditions.

10. Craig, 62.

11. Barber, "Big Data," 6, Table 3.

12. Barber, "Big Data," 18, Table 8.

13. Barber, "Big Data," 10.

14. Barber, "Big Data," 11.

The functioning and set up for individual Zeta tests are explained in the published research articles and chapters that use this test, and more comprehensively in the introduction to Hugh Craig and Arthur F. Kinney's *Shakespeare, Computers, and the Mystery of Authorship.* However, documentation for the Intelligent Archive Javascript software (which includes the ability to run Zeta tests) is negligible, with technical release notes but no instructions. The current interface of the Intelligent Archive is rudimentary and not geared towards usability, with no guidance or help feature. Despite its name, the Archive does not come pre-loaded with texts; these must be individually uploaded from suitable repositories in XML or TEI format.[15] Researchers wishing to run their own tests may prefer the fully guided method of running and analyzing Zeta tests (including a database of early modern plays from 1552 to the Restoration) available in the Goldsmiths data repository.[16]

ROS BARBER
Goldsmiths, University of London
https://doi.org/10.33137/rr.v44i4.38652

Works Cited

Barber, Ros. 2020. "BDNE Zeta Dataset." Goldsmiths Research Online. London: Goldsmiths, University of London. doi.org/10.25602/GOLD.00028390.

Barber, Ros. 2020. "Big Data or Not Enough? Zeta Test Reliability and the Attribution of Henry VI." *Digital Scholarship in the Humanities* 36 (3):542–64. doi.org/10.1093/llc/fqaa041.

Burrows, John. 2007. "All the Way Through: Testing for Authorship in Different Frequency Strata." *Literary and Linguistic Computing* 22 (1):27–47. dx.doi.org/10.1093/llc/fqi067.

Craig, Hugh. 2009. "The Three Parts of Henry VI." In Craig and Kinney, 44–77. Cambridge: Cambridge University Press.

Craig, Hugh, and Arthur F. Kinney. 2009. *Shakespeare, Computers, and the Mystery of Authorship.* Cambridge: Cambridge University Press.

15. For Shakespeare's texts, the best source is Folger Digital at shakespeare.folger.edu/download/. For other early modern plays, the best source is EarlyPrint, texts.earlyprint.org/works/.

16. Barber, "BDNE Zeta Dataset."

Hoover, David L. 2007. "Corpus Stylistics, Stylometry, and the Styles of Henry James." *Style* 41 (2):174–203. jstor.org/stable/10.5325/style.41.2.174.

Rizvi, Pervez. 2019. "The Interpretation of Zeta Test Results." *Digital Scholarship in the Humanities* 34 (2):401–18. doi.org/10.1093/llc/fqy038.

Rudman, Joseph. 2016. "Non-Traditional Authorship Attribution Studies of William Shakespeare's Canon: Some Caveats." *Journal of Early Modern Studies* 5:307–28. doi.org/10.13128/JEMS-2279-7149-18094.

Rudman, Joseph. 1997. "The State of Authorship Attribution Studies: Some Problems and Solutions." *Language Resources and Evaluation* 31 (4): 351–65.

Taylor, Gary, and Gabriel Egan, eds. 2017. *The New Oxford Shakespeare Authorship Companion.* Oxford: Oxford University Press. doi.org/10.1093/actrade/9780199591169.001.0001.

Kaufer, David, and Suguru Ishizaki, creators.
DocuScope.
Pittsburgh, PA: Carnegie Mellon University. Accessed 25 November 2021. cmu.edu/dietrich/english/research/docuscope.html.

DocuScope is described by its creators as "a computer-aided text analysis tool that allows researchers to conduct both quantitative and qualitative analyses of how the designed reader experience is created by writers through the selection of micro linguistic composing patterns, i.e., words and phrases."[1] The program is able to generate a huge amount of potentially useful data—breaking down texts individually into tagged components, which can be compared quantitatively. One of the best aspects of the way in which the data is outputted is that a corpus can be viewed both at the level of a corpus and at the level of an individual text from that corpus. This allows the data to be interrogated at a much more detailed level than is perhaps usual: both in how the analysis for each word of a text can be viewed, and in comparison with output from a wider corpus, or group of texts.

From the point of view of early modern textual linguistics, however, the program uses a modern dictionary. My own use of DocuScope was limited by

1. *DocuScope*, 4.