Title	English Bards and Unknown Reviewers: a Stylometric Analysis of Thomas Moore and the Christabel Review
Author(s)	Benatti, Francesca; Tonra, Justin
Publication Date	2015
Publication Information	Benatti, Francesca; Tonra, Justin (2015) 'English Bards and Unknown Reviewers: a Stylometric Analysis of Thomas Moore and the Christabel Review'. Breac: A Digital Journal of Irish Studies, 4 .
Publisher	University of Notre Dame
Link to publisher's version	https://breac.nd.edu/
Item record	http://hdl.handle.net/10379/4821

English Bards and Unknown Reviewers: a Stylometric Analysis of Thomas Moore and the Christabel *Review.*

Francesca Benatti, Open University

Justin Tonra, National University of Ireland, Galway

Introduction

Fraught relations between authors and critics are a commonplace of literary history. The particular case that we discuss in this article, a negative review of Samuel Taylor Coleridge's *Christabel* (1816), has an additional point of interest beyond the usual mixture of amusement and resentment that surrounds a critical rebuke: the authorship of the review remains, to this day, uncertain. The purpose of this article is to investigate the possible candidacy of Thomas Moore as the author of the provocative review. It seeks to solve a puzzle of almost two hundred years, and in the process clear a valuable scholarly path in Irish Studies, Romanticism, and in our understanding of Moore's role in a prominent literary controversy of the age.

The article appeared in the September issue of the *Edinburgh Review*, a major quarterly journal and standard-bearer for literary reviewing from its establishment in 1802. The assessment of *Christabel* enraged Coleridge, who believed his work "was assailed with a malignity and a spirit of personal hatred."[1] However, the *Edinburgh*, along with many other nineteenth-century periodicals, maintained a strict a policy of anonymity for its contributors so a focus for Coleridge's fury was wanting. The poet acknowledged a number of suspects, and the identity of periodical reviewers was often an open secret in literary circles,[2] but on occasion articles would remain stubbornly anonymous.

Such is the case with the *Edinburgh*'s review of *Christabel*. Though a scholarly debate about its authorship has simmered since the beginning of the twentieth century, attempts to

---

[1] Samuel Taylor Coleridge, *Biographia Literaria; Or, Biographical Sketches of My Literary Life and Opinions*, vol. 2, 2 vols. (London: Rest Fenner, 1817), 298.

[2] Joanne Shattock. *Politics and Reviewers: The "Edinburgh" and the "Quarterly" in the Early Victorian Age*. Leicester: Leicester University Press, 1989, 15-16.

attribute the article have yet to reach a consensus.[3] Most recently, Duncan Wu has

attributed the article to William Hazlitt and included it in his edition of Hazlitt's unpublished

writings, basing his attribution upon Coleridge's own suspicions and the other negative

reviews of his works that Hazlitt published during 1816-17.[4] However, later on in 1829,

Coleridge came to suspect Irish author Thomas Moore of having "undertaken to strangle the

Christabel,"[5] and he has since been one of the leading candidates identified by subsequent

scholars.[6]

A growing interest in the importance of Moore to the fields of Irish Literature, Irish History,

and Irish Studies is evident in the renewed scholarly focus of recent years.[7] In Romanticism,

Moore's centrality to the contemporary literary culture has been demonstrated in James

Chandler's case study of the year 1819 in literature.[8] The writer's candidacy for the

authorship of the *Christabel* review underscores that pivotal importance, while a successful

attribution of the article is significant for our understanding of several aspects of Moore's

career and has consequences for those disciplinary fields where he has been the focus of

growing attention.

A successful attribution of the *Christabel* review is desirable for a number of reasons. It

would answer several pertinent questions that have arisen during the debate about its

---

[3] John Beer, "Coleridge, Hazlitt, and 'Christabel,'" *Review of English Studies* 37, no. 145 (1986): 40–54; P. L. Carver, "The Authorship of a Review of 'Christabel' Attributed to Hazlitt," *Journal of English and Germanic Philology* 29, no. 4 (1930): 562–78; Kathleen H. Coburn, "Who Killed Christabel?," *TLS* (May 20, 1965): 397; Wilfred S. Dowden, "Thomas Moore and the Review of 'Christabel,'" *Modern Philology* 60, no. 1 (1962): 47–50; Thomas Hutchinson, "Coleridge's Christabel," *Notes and Queries* X, 9th S (1903): 170–2; Hoover H. Jordan, "Thomas Moore and the Review of 'Christabel,'" *Modern Philology* 54, no. 2 (1956): 95–105; Elisabeth Schneider, "The Unknown Reviewer of Christabel: Jeffrey, Hazlitt, Tom Moore," *PMLA* 70, no. 3 (1955): 417–32; Elisabeth Schneider, "Tom Moore and the Edinburgh Review of Christabel," *PMLA* 77, no. 1 (1962): 71–76.

[4] See William Hazlitt. *New Writings of William Hazlitt.* Edited by Duncan Wu. Oxford; New York: Oxford University Press, 2007, 203-206.

[5] Samuel Taylor Coleridge. Notebook 42 ff 10-11 BM Add. Man. 47537, Sept-Oct. 1829. Quoted in Coburn, "Who Killed *Christabel*?," 397.

[6] Schneider is Moore's most determined advocate, but he is also considered a candidate by Stanley Jones. See Stanley Jones. *Hazlitt. A Life. From Winterslow to Frith Street* (Oxford: Clarendon Press, 1989), 223-24.

[7] A host of scholarly articles have accompanied new biographies by Ronan Kelly and Linda Kelly, a collection, *Thomas Moore: Texts, Contexts, Hypertext*, and new editions of Moore's writings by Jane Moore, Emer Nolan and Seamus Deane, and Jeffery Vail.

[8] James Chandler, *England in 1819: The Politics of Literary Culture and the Case of Romantic Historicism* (Chicago ; London: University of Chicago Press, 1998), *passim*, particularly 267-99.

authorship, as well as contextualising the many claims, counter-claims, and denials that trailed in its wake. From the perspective of Thomas Moore and Irish studies, a definitive outcome would clarify many issues about an important period in the author's career that are currently subject to speculation and uncertainty. Most notably, it would illuminate Moore's relationship with Coleridge, and complicate his sympathetic statements about his fellow poet during the period of the *Christabel* controversy. It would also provide nuance to our appreciation of how the Irish poet navigated and manipulated the complex dynamics and relationships of the literary and journalistic spheres in London. In a similar fashion, a successful attribution to Moore would help to address unanswered questions about his relationship with Lord Byron which hinge upon references to the "noble bard"[9] in the *Christabel* review. 1816 saw Byron's departure from England, so a clearer picture of the connection between biographer and his future subject is to be welcomed. Additionally, the composition of Moore's canon of writings for the *Edinburgh* review would become more complete with the addition of a freshly attributed piece.

This article describes the process and the results of a new examination of the *Christabel* review based on the use of stylometric analysis and author attribution technologies. We outline the challenges that we have encountered in this work-in-progress, the unique difficulties presented by this case, and the opportunities and implications of this intersection of digital humanities and Irish Studies.

The Christabel *review*

Christabel was published by John Murray on 8 May 1816 in a volume that contained the long titular poem, as well as the shorter 'Kubla Khan' and 'The Pains of Sleep.' Conceived as a five part work, 'Christabel''s first two sections were written in 1797 and 1800, respectively, but the poet's self-confessed "indolence"[10] accounted for the neglect of the remaining

---

[9] Review of *Christabel: Kubla Khan, a Vision; The Pains of Sleep*, by Samuel Taylor Coleridge, *Edinburgh Review* 27 (1816): 58

[10] Samuel Taylor Coleridge, *Christabel: Kubla Khan, a Vision; The Pains of Sleep.* (London: John Murray, 1816), v.

parts.[11] The review of the volume in the *Edinburgh* is a sustained attack on the title poem, with some brief comments on the two shorter works. The reviewer characterises 'Christabel' as "one of the most notable pieces of impertinence of which the press has lately been guilty"; "utterly destitute of value"; and "a mixture of raving and driv'ling."[12] Faced with such effrontery, Coleridge sought to identify its author, but was denied by the journal's scrupulous corporate anonymity. His first suspect was the *Edinburgh*'s editor, Francis Jeffrey, who wrote many of the articles that appeared in the journal's pages, and who exercised full authority to make "retrenchments and verbal alterations"[13] to his contributors' work. When Coleridge expressed his surprise that the attack had come from "a man, who both in my presence and in my absence, has repeatedly pronounced it ['Christabel'] to be the finest poem of its kind in the language,"[14] Jeffrey understood the remark to be aimed at him. In a lengthy footnote to William Hazlitt's review of Coleridge's *Biographia Literaria* for the August 1817 issue of the *Edinburgh*, Jeffrey discussed *Christabel*, and unequivocally stated, "I did not review it."[15] Such was his desire to disavow his authorship of the review that he bypassed the journal's policy of anonymity to sign the footnote with his initials.

Coleridge's next suspect was William Hazlitt, who wrote no fewer than five critical reviews of Coleridge's works between 1816 and 1817.[16] He even reviewed *Christabel* in the *Examiner* newspaper in June 1816,[17] but the essence of that article, though negative, did not match the fervour of the *Edinburgh*. Hazlitt never admitted authorship of the *Edinburgh* review of *Christabel*, though he acknowledged harbouring extensive animosity against Coleridge.[18] Despite the fact that Coleridge did not identify Thomas Moore as a suspect until 1829, there

---

[11] Though Coleridge promises to "embody in verse the three parts yet to come, in the course of the present year" (Coleridge, *Christabel*, v-vi), they do not materialise.

[12] Review of *Christabel*: 66.

[13] William Hazlitt, "Coleridge's Literary Life," Review of *Biographia Literaria*, by Samuel Taylor Coleridge, *Edinburgh Review* 28 (1817): 512.

[14] Coleridge, *Biographia Literaria* vol. 2, 299.

[15] Hazlitt, "Coleridge's Literary Life," 510.

[16] Four reviews were on the *Statesman's Manual* and one on *Biographia Literaria*. See Duncan Wu, *William Hazlitt : the First Modern Man* (Oxford and New York: Oxford University Press, 2008), 190.

[17] William Hazlitt, Review of *Christabel: Kubla Khan, a Vision; The Pains of Sleep*, by Samuel Taylor Coleridge, *Examiner* no. 440 (June 2, 1816): 348-49.

[18] See Hazlitt, *New Writings*, 204.

were long-standing connections between Moore and the poem. A sequence of surviving

letters outlines them: Lord Byron suggested that Moore should review *Christabel*,[19] but the

latter wrote to Francis Jeffrey in May 1816 to say that he would not do so out of respect for

Coleridge's precarious financial situation at the time.[20] A letter of December 1816 to

publisher John Murray contained a further denial of authorship from Moore.[21]

Theoretical background

While biographers of Hazlitt, Moore, and Coleridge have commented on this controversy, the

most detailed attempts at attributing the controversial review have appeared in scholarly

publications extending from the turn of the twentieth century into the twenty-first. Much of the

scholarship that has attempted to attribute the authorship of the *Christabel* review has

focused on external evidence, but the case is not favoured with that elusive kind of evidence

which "allows us to locate the work's genesis [. . .] at a particular desk in a particular

room."[22] There is a growing recognition of this fact as the debate develops, though fresh

documentary discoveries (none conclusive) are occasionally added to the file of external

evidence. The assessment of the respective merits of external and internal evidence figures

prominently in the rhetoric of the mid-century articles, with Schneider's suggestion that "the

strongest argument is obviously the internal evidence"[23] countered by Jordan's claim that it

is "vexatious to use effectively."[24] Jordan's view is vindicated, to a degree, by the

contradictory conclusions that follow from the use of internal evidence. For example, those

---

[19] George Gordon Byron, *The Works of Lord Byron: With His Letters and Journals, and His Life*, ed. Thomas Moore, 16 vols. (London: John Murray, 1832), 3:183.

[20] "I had some idea of offering myself to you to quiz Christabel [. . .] but I have been lately told that Coleridge is poor—so poor as to be obliged to apply to the Literary Fund—and as this is no laughing matter—why—I shall let him alone—" (Thomas Moore, *The Letters of Thomas Moore*, ed. Wilfred S. Dowden, 2 vols. (Oxford: Clarendon Press, 1964), 1:394-95).

[21] "The article upon Coleridge in the Ed. Rev. was altogether disgraceful both from its dulness and illiberality— You know I had some idea of laughing at Christabel myself—but when you told me that Coleridge was very poor and had been to the Literary Fund, I thought this no laughing matter, and gave up my intention—"(Moore, *The Letters of Thomas Moore*, 1:407).

[22] Harold Love, *Attributing Authorship: An Introduction* (Cambridge: Cambridge University Press, 2002), 51-2.

[23] Schneider, "The Unknown Reviewer of Christabel: Jeffrey, Hazlitt, Tom Moore," 429.

[24] Jordan, "Thomas Moore and the Review of 'Christabel,'" 95.

that argue on the basis of an uncommon word's association with (or dissociation from) a certain author are often refutable by means of simple computer-aided text-searching that was unavailable to these mid-century scholars. Our initial work on testing these kinds of claims rendered some of them doubtful. In contrast to Schneider's assertion that "[b]rilliance was never Hazlitt's criterion of excellence,"[25] our Hazlitt corpus revealed three instances of the word 'brilliant' (including descriptions of 'theories' and 'passages.')[26] Elsewhere, Schneider's claim for Moore's authorship based on his use of the word 'couplet' in a unique sense of two verse lines that do not rhyme is weakened by the discovery of the same sense attributed to the word by Hazlitt.[27] While previous *Christabel* attributionists do not base their claims entirely upon individual pieces of evidence such as these, their simple validation with current digital tools undermines the tendency to accumulate them into an *argumentum verbosium*.

Our ongoing attempts to attribute the *Christabel* review have focused on internal linguistic evidence from the text, and from other texts by the authors that scholarship has identified as the most likely candidates for authorship. Our rationale for this focus is a belief in the inadequacy of the available external evidence coupled with a recognition that previous assessments of internal evidence were limited by their inability to exploit the range of analytic and stylometric tools that have emerged in the digital humanities in recent times. The decision was not based on a philosophical belief in any innate superiority or greater validity of internal evidence, but from a desire to generate *new* evidence that could contribute to a credible attribution. Indeed, while the distinction between internal and external evidence with authorship attribution is fairly solid, the broader division between subjective and objective evidence professed by the two cultures of the humanities and the

---

[25] Schneider, "The Unknown Reviewer of Christabel: Jeffrey, Hazlitt, Tom Moore," 424.

[26] William Hazlitt, "Standard Novels and Romances," Review of *The Wanderer: or, Female Difficulties*, by Madame d'Arblay, *Edinburgh Review* 24 (1815): 333. William Hazlitt, "Sismondi's Literature of the South," Review of *De la Litterature du Midi de l'Europe*, by J. C. L. Simonde de Sismondi, *Edinburgh Review* 25 (1815): 33. Hazlitt, "Coleridge's Lay-Sermon," Review of *The Statesman's Manual*, by Samuel Taylor Coleridge. *Edinburgh Review* 27 (1816): 449.

[27] Schneider, "The Unknown Reviewer of Christabel: Jeffrey, Hazlitt, Tom Moore," 428. William Hazlitt, *A View of the English Stage: Or, a Series of Dramatic Criticisms* (London: John Warren, 1821), 44.

sciences is a false dichotomy. In terms of textual evidence, Jerome McGann's assertion that "there is no such thing as unmarked text"[28] is especially pertinent. For, while his focus is on a residual and indelible materiality of so-called 'plain text'—"all texts implicitly record a cultural history of their artifactuality"[29]—a semantic cultural residue is also present in competing subjective analyses of isolated linguistic phenomena.[30] As the forthcoming description reveals, subjective judgment is inherent in every stage of the attribution process, from the construction of a corpus to the evaluation of the results of algorithmic processes. Thus, while the computing techniques described below are more complete and sophisticated than the manual word-counting employed by Schneider, Jordan, and others, the notion that computing alone will resolve this problem is a fallacy. The rhetoric of authorship attribution promotes the legitimacy of judgments that are made from the empirical facts of a text, but this rhetoric obscures the subjective choices behind the methodologies and workflows, and the interpretation of the facts.[31]

Gathering and processing data

The first stage of stylometric analysis was the selection and preparation of a corpus of articles that would include the potential authors of the *Christabel* review. We decided to follow Jack Grieve's advice that "each author-based corpus should be composed of texts produced in the most similar register, for the most similar audience, and around the same point in time as the anonymous text."[32] We therefore selected a set of texts comprised of articles on literary subjects from the *Edinburgh Review*, written within a two-year interval of the *Christabel* review (1814 to 1818). Attributions were based on the *The Wellesley Index to*

---

[28] Jerome J. McGann, *Radiant Textuality: Literature after the World Wide Web* (New York ; Basingstoke: Palgrave, 2001), 138.

[29] McGann, *Radiant Textuality*, 138.

[30] The different conclusions about the words 'brilliant' and 'couplet' in the *Christabel* review are examples of how a text is a "galaxy of signifiers, not a structure of signifieds" (Roland Barthes, *S/Z*, trans. Richard Miller (New York: Hill and Wang, 1974), 5).

[31] See also: Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Urbana: University of Illinois Press, 2011), 6.

[32] Jack Grieve. "Quantitative Authorship Attribution: An Evaluation of Techniques." *Literary and Linguistic Computing* 22, no. 3 (9–1, 2007): 255.

Victorian Periodicals, 1824-1900 online edition.[33] We included only those articles where firm attributions had been made by the *Wellesley*, based on documentary evidence.

We found that literary reviews were contributed by a limited group of authors in this period. Editor Francis Jeffrey was by far the most prolific literary reviewer, followed by William Hazlitt and Henry Brougham. Other *Edinburgh Review* regulars like Sir James Mackintosh and John Allen wrote little about literary topics at this time, but contributed articles on cognate subjects such as history or travel, and had both written literary reviews in the past. Other literary reviewers in this period were Thomas Moore, Sir Francis Palgrave and the exiled Italian poet Ugo Foscolo. Additional authors, such as Leigh Hunt, John Cam Hobhouse, Douglas Kinnaird and Sir Walter Scott, contributed one article apiece during 1814-1818. We decided to exclude these last three authors because one single article each would provide an insufficient amount of text for a reliable analysis. Leigh Hunt was likewise excluded because his first (and only) review was published in the issue following the *Christabel* review.[34] We also decided to exclude Foscolo from the corpus based on two considerations: the subject of his literary reviews, which were exclusively on Italian literature, and the fact that all three of his articles were translated from Italian into English by other *Edinburgh* reviewers. The authors we chose to test as potential reviewers of *Christabel* were the following: John Allen, Henry Brougham, William Hazlitt, Francis Jeffrey, Sir James Mackintosh, Thomas Moore, and Sir Francis Palgrave.

The texts for all articles were taken from scans of the *Edinburgh Review* available through Google Books or from archival copies of the original journals. Most of the Google Books scans were of satisfactory legibility. When digital copies proved illegible, scans of physical copies were taken and then processed with ABBYY FineReader 11 Optical Character Recognition (OCR) software. The OCR texts for all articles, whether from Google Books or other sources, were then corrected manually by the investigators against the scanned page

---

[33] *The Wellesley Index to Victorian Periodicals, 1824-1900*. Pro Quest, 2006-2014. http://wellesley.chadwyck.com/. Despite the dates in the title of this resource, it contains a full index of the *Edinburgh Review*.

[34] Francis Jeffrey to Leigh Hunt, 20 October 1816. *Leigh Hunt Online: The Letters*. The University of Iowa Libraries. http://digital.lib.uiowa.edu/cdm/ref/collection/leighhunt/id/7074

images. Original punctuation and spelling were retained, on the assumption that they all conformed to the same "house style" enforced by the *Edinburgh Review* editorial staff. Only obvious spelling mistakes were emended.

Next, the texts were encoded in XML using the Text Encoding Initiative (TEI) P5 guidelines, marking out all quotations from external sources that had not been written by the article's author (including the work or works being reviewed) with <quote> tags. These quotations often comprised a large proportion of the text (about 10 to 30% depending on the article), and their presence could skew the result of our stylometric analysis significantly.[35] An XSLT script was then written to strip out these encoded portions, leaving for each article a plain text file comprised only of the author's words.

The articles available for each author varied both in length and in number. Jeffrey and Brougham often contributed three or four articles per issue, while other authors provided no more than two or three in the four-year period under consideration. In order to preserve a balance in the corpus between frequent and irregular contributors, we elected to include a similar amount of words for each author. We decided upon 10,000 words, divided into at least two articles, as a suitable minimum size for each author-based corpus.[36]

This decision was not without problems, especially for our main focus of interest, Thomas Moore's possible authorship. Previous work on Moore's relationship with the *Edinburgh Review* has shown that Moore wrote only three relatively short articles during 1814-1818, one of them, "Jorgenson's *Travels*" (1817), of tentative attribution.[37] When including only articles of firm attribution, Moore's corpus would fall below 10,000 words. We were therefore forced to include the contribution by Moore nearest in time, an article on "French Novels" from 1820, in order to reach a corpus size equivalent to those of the other candidate authors (see Table 1).

---

[35] Patrick Juola. "Authorship Attribution." *Foundations and Trends in Information Retrieval* 1, no. 3 (2007): 247-8.

[36] John Burrows. "All the Way Through: Testing for Authorship in Different Frequency Strata." *Literary and Linguistic Computing* 22, no. 1 (April 1, 2007): 30.

[37] Moore received payment for an article on "M. de J." in 1818. See Francesca Benatti. "Joining the Press-Gang: Thomas Moore and the *Edinburgh Review*," in *Thomas Moore: Texts, Contexts, Hypertext*, ed. Francesca Benatti, Sean Ryder and Justin Tonra (Oxford: Peter Lang, 2013), 171.

Table 1

Author	Title	Year of publication	Wordcount
John Allen	Lingard's *Antiquities of the Anglo-Saxon Church*	1815	3,075
John Allen	Napoleon Bonaparte	1816	12,811
Henry Brougham	Carnot's *Defence*	1815	4,524
Henry Brougham	Frankiln's *Correspondence*	1817	5,866
Henry Brougham	Junius	1817	5,959
William Hazlitt	Sismondi's *Literature of the South*	1815	8,612
William Hazlitt	Coleridge's *Literary Life*	1817	7,716
Francis Jeffrey	Lord Byron's *Corsair* and *Bride of Abydos*	1814	6,808
Francis Jeffrey	Wordsworth's *Excursion*	1814	6,611
Francis Jeffrey	*Waverley*, a Novel	1814	3,914
Sir James Mackintosh	Godwin's *Lives of Milton's Nephews*	1815	6,705

Sir James Mackintosh	Sir Nathaniel Wraxall	1815	5,924
Thomas Moore	Boyd's *Translations from the Fathers*	1814	4,600
Thomas Moore	French Novels	1820	3,190
Thomas Moore	Lord Thurlow's *Poems*	1814	3,352
Sir Francis Palgrave	Goethe's *Memoirs*	1817	7,281
Sir Francis Palgrave	Ancient German and Northern Poetry	1816	7,219

Stylometric analysis with Stylo

Computational stylistics is the study of how stylistic traits can be measured through statistical methods to trace an author's stylistic "fingerprint," based on the assumption that every author has "a verifiably unique style."[38] Also known as "stylometry," it "relies on advanced statistical procedures to distil significant markers of authorial style from a large pool of stylistic features."[39] It is also referred to, in its applications to investigate the authorship of texts, as "non-traditional authorship attribution."

In recent years, the discipline has been made better known by John Burrows, from his 1987 monograph *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method,* to his Busa Award lectures in 2001, where he presented his "Delta" method,[40]

---

[38] Joseph Rudman. "The State of Non-Traditional Authorship Attribution Studies—2012: Some Problems and Solutions." *English Studies* 93, no. 3 (2012): 265.

[39] Maciej Eder. "Mind Your Corpus: Systematic Errors in Authorship Attribution." *Literary and Linguistic Computing* 28, no. 4 (12–1, 2013): 603.

[40] John Burrows. "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary*

which has been employed to date in more than thirty published studies of authorship attribution, right to the present day.

Methods employed for stylometry can be divided into two broad categories: unsupervised and supervised. Unsupervised methods do not require previous classification of the texts being analysed, and instead "look for superficial patterns in the data," which can nonetheless be extremely relevant for authorship attribution.[41] Unsupervised methods include Principal Component Analysis (PCA) and Cluster Analysis. They require human interpretation of the degree of similarity between the texts being analysed.[42] Supervised methods instead necessitate that "documents be categorized prior to analysis."[43] Examples of supervised machine-learning methods are k-Nearest Neighbours, Nearest Shrunken Centroids and Support Vector Machines. Supervised methods rely upon the division of available texts into a "training" set, containing texts of known authorship that are deemed representative of their author's style, and a "test" set, containing a mixture of texts of known authorship and of anonymous texts, including those whose authorship is being assessed.[44] Supervised methods often include cross-validation, an estimate of the potential margin of error due to inconsistent choice of the training samples, performed through multiple analysis of the dataset, such as ten-fold cross-validation.[45]

We decided to employ Maciej Eder and Jan Rybicki's Stylo suite of scripts for the R statistical software, which is specifically designed for digital stylometric analysis. The Stylo scripts have been constantly developed and enhanced with a graphical user interface since 2010, and are supported by a growing user community. They have been employed in more

---

and Linguistic Computing 17, no. 3 (September 1, 2002): 267–287.

[40] ———. "Questions of Authorship: Attribution and Beyond: A Lecture Delivered on the Occasion of the Roberto Busa Award ACH-ALLC 2001, New York." *Computers and the Humanities* 37, no. 1 (February 2003): 5–32.

[41] Juola, "Authorship Attribution": 273.

[42] Maciej Eder and Jan Rybicki. "Do Birds of a Feather Really Flock Together, or How to Choose Training Samples for Authorship Attribution."*Literary and Linguistic Computing* 28, no. 2 (6–1, 2013): 229.

[43] Juola, "Authorship Attribution", p. 277.

[44] Eder and Rybicki, "Do Birds of a Feather", p. 230

[45] *Ibid.*, 230-1.

than ten scholarly articles and fifteen conference presentations since 2006.

For the purposes of this study, we experimented with two unsupervised methods (Cluster Analysis and Principal Component Analysis) and three supervised ones (Support Vector Machines, Nearest Shrunken Centroids and k-Nearest Neighbours). Due to space restrictions, we will limit our discussion to the results obtained through what proved to be, in our case, the most useful unsupervised and supervised method respectively: Cluster Analysis and Support Vector Machines.

Unsupervised methods

There are many measures possible for stylometric analysis, such as examining the frequency and distribution of characters, words, n-grams, or part-of-speech tags. For the purposes of this study, we decided to concentrate on the analysis of high-frequency words, a method which has shown considerable promise.[46]

One of the most widely tested techniques for the stylometric analysis of most frequent words (MFW) is John Burrows' Delta procedure. In brief, Burrows defines Delta as "the mean of the absolute differences between the z-scores for a set of word-variables in a given text group and the z-scores for the same set of word-variables in a target text."[47] The author of the text group with the smallest value of Delta, i.e. the smallest mean difference from the target text, is "least unlike" it and has the best claim, among the authors tested, to be the author of the target text. In procedural terms, we determine what are the most frequent words in our corpus of literary reviews and how frequently those words occur in each text that we have included in our corpus. We can therefore measure by how much the usage of a given word in the *Christabel* review differs from that of each text in the literary reviews corpus (the z-score). We then repeat the procedure for a given number of most frequent words, add up the

---

[46] High-frequency words have been described as providing "the most consistently reliable results in authorship attribution problems" in Matthew L. Jockers and Daniela M. Witten. "A Comparative Study of Machine Learning Methods for Authorship Attribution." *Literary and Linguistic Computing* 25, no. 2 (6–1, 2010): 215 and as "the most accurate solution" for English texts in Maciej Eder "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." *Literary and Linguistic Computing* (November 14, 2013): 10.

[47] John Burrows. "Questions of Authorship": 14.

z-scores, and average them. The text with the smallest average difference between its z-scores and *Christabel*'s z-scores is the least unlikely text, or "nearest neighbour."

In order to improve the reliability of Cluster Analysis visualizations based on Delta, we employed Eder's "bootstrap consensus tree" technique, which "starts with pairing the nearest neighbouring samples into two-element groups, and then recursively joins these groups into larger clusters."[48] These connections are represented graphically as a dendrogram tree graph. Texts that the Stylo package classifies as nearest neighbours appear as offshoots of the same branch of the tree, which is represented as a line departing from the centre of the graph and splitting into other lines. If a text is instead linked directly to the centre of the tree by a single line that does not split, Stylo cannot detect a nearest neighbour for it.

We applied a series of tests starting from 50 most frequent words, gradually expanding the range to the 500 most frequent words in increments of 10 words. We thus produced a number of virtual dendrograms which the Stylo script combined into one final dendrogram graph. The consensus strength was initially set at 0.5: texts had to appear as nearest neighbours in at least 50% of the analyses in order to be represented as offshoots of the same branch of the final dendrogram graph.[49] Such a consensus strength is relatively weak and could possibly detect similarities that are only superficial between texts. Any nearest neighbour relationship emerging from a 0.5 consensus strength would therefore need to be validated through more stringent tests with higher consensus settings. Conversely, the absence of linkage between two or more texts at such a tolerant consensus strength is "convincing evidence of their actual significant differentiation."[50]

After the first round of tests were conducted, and the MFW lists examined, it became evident that a further parameter would have to be set. A certain amount of variation in topics among

---

[48] Described in Maciej Eder. "Computational Stylistics and Biblical Translation: How Reliable Can a Dendrogram Be?" in *The Translator and the Computer*, ed.Tadeusz Piotrowski and Lukasz Grabowski (Wroclaw: WSF Press, 2013), 157-8.

[49] *Ibid.*, 164-5 for a description of consensus strength.

[50] *Ibid.,* 166.

the texts being examined has the potential to influence the attribution results.[51] We deliberately excluded texts on political, legal or scientific subjects, but the chosen articles vary from reviews of poetry to reviews of more historical works or travel narratives. Given the years in which they were published were dominated by Waterloo and the Congress of Vienna, there is a significant minority of texts that focuses on the fall of Napoleon or the present state of France, authored for example by Brougham, Allen and Palgrave, but not by Moore or Hazlitt. Some of the content words for these reviews ("Napoleon" and "France", for example) crept into the list of Most Frequent Words and required "culling."[52] So only those words that appear in a given percentage of texts were selected for analysis. We set this percentage on a sliding scale from 50% of texts to 100%, in 10% increments. There were 674 words present in at least 50% of the texts (50% culling), 450 words present in at least 60% of the texts (60% culling), 308 in 70% of the texts, 212 in 80%, 141 in 90%, until at 100% culling, only 98 words were identified as being present in all the texts. At each culling interval, Stylo performed multiple Cluster Analysis tests based on Delta, starting from the 50 most frequent words and ending with the highest available multiple of 10 in increments of 10 words (500 at 50% culling, for a total of 46 tests; 450 at 60% culling, for a total of 41 tests, and so on). In total, Stylo performed 145 Cluster Analysis tests, which it combined into the bootstrap consensus tree graph in Fig. 1 below.

Fig. 1

---

[51] Kim Luyckx and Walter Daelemans. "The Effect of Author Set Size and Data Size in Authorship Attribution." *Literary and Linguistic Computing* 26, no. 1 (4–1, 2011): 51.

[52] Culling was first discussed by David L. Hoover. See "Frequent Word Sequences and Statistical Stylistics." *Literary and Linguistic Computing* 17, no. 2 (6–1, 2002): 170 and "Testing Burrows's Delta."*Literary and Linguistic Computing* 19, no. 4 (11–1, 2004): 456.

Edinburgh_1820
Bootstrap Consensus Tree
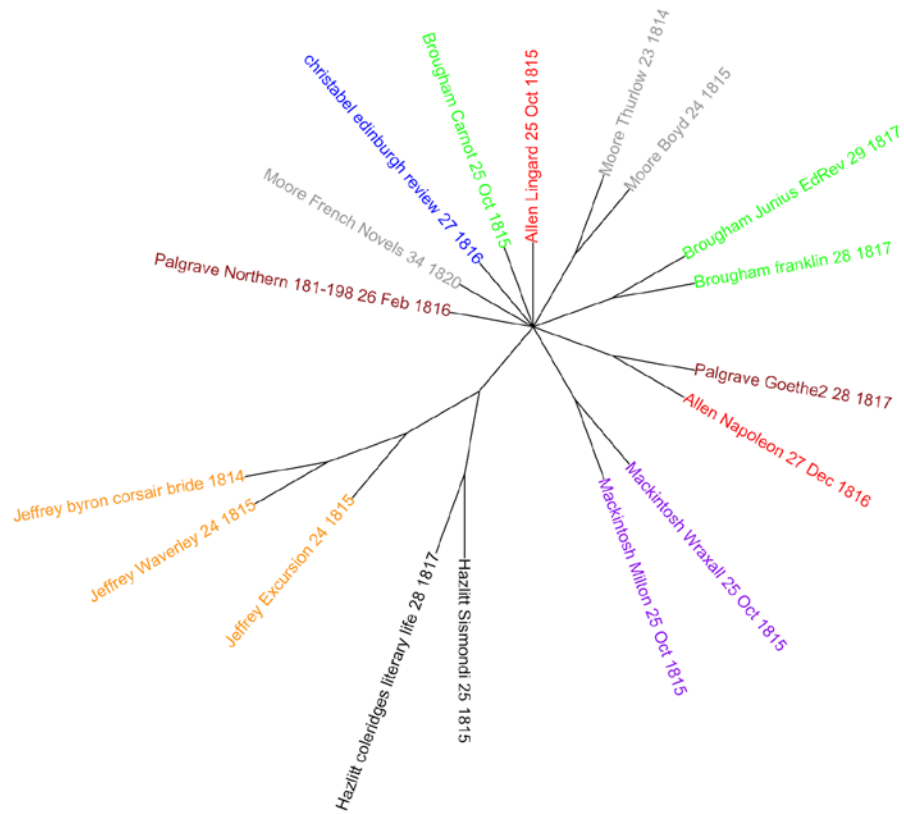
50-500 MFW Culled @ 50-100%
Classic Delta distance Consensus 0.5

From the branching of the dendrogram tree, it is evident that Jeffrey's articles form a clear cluster among themselves but are also closely linked to Hazlitt's, which in turn form another cluster. Mackintosh's articles are grouped together, while Moore's 1814-1815 articles form a distinct cluster, with the 1820 article not being detected as its nearest neighbour. Two of Brougham's articles are detected as by the same author, but interestingly, another is not. The analysis fails to discriminate between Palgrave and Allen, grouping together two of their articles. Most significantly for the purposes of our study, the *Christabel* review is not detected as the nearest neighbour of any other text, as it is connected directly with the

centre of the tree graph. It sits with four other texts for which, like *Christabel*, no nearest neighbour emerged in at least 50% of the tests Stylo performed.

There are several possible explanations for this failure, which are discussed in the Evaluation section below.

Supervised methods

In addition to Cluster Analysis with Delta, we also decided to experiment with supervised attribution methods, which have been employed successfully in authorship attribution contexts by Jockers, Juola and others.[53]

The corpus was divided between a training set comprising the same texts employed for Cluster Analysis, including between 11,000 and 17,000 words per author in at least two texts, and a test set. The composition of the test set was driven by the need to test the accuracy of the chosen methods on articles of different length. There is still no consensus over the minimum amount of words needed for accurate attribution, with Eder indicating 5,000 as the minimum reliable sample for fiction in English or other languages,[54] Burrows employing 2,000 as the minimum threshold for his Delta, Zeta and Iota tests, with declining results for shorter texts,[55] and Holmes, Gordon and Wilson stating 1,000 words as the minimum sample size for "stylometric reliability."[56] The *Christabel* review, when stripped of all quotations, measures 2,686 words. Besides searching for the author of the *Christabel* review, one significant goal of our tests was to measure how successful machine learning methods would be at attributing articles shorter than 3,000 words. The composition of the test set is given in Table 2 below. Articles whose attribution is not supported by certain evidence are indicated with an asterisk.

---

[53] See for example Juola. "Authorship Attribution": 277-8 and 285-6 and Matthew L. Jockers. "Testing Authorship in the Personal Writings of Joseph Smith Using NSC Classification."*Literary and Linguistic Computing* 28, no. 3 (9–1, 2013): 371–381.

[54] Maciej Eder. "Does Size Matter?": 14.

[55] Burrows, "All the Way Through": 28.

[56] David I. Holmes, Lesley J. Gordon, and Christine Wilson. "A Widow and Her Soldier: Stylometry and the American Civil War." *Literary and Linguistic Computing* 16, no. 4 (11–1, 2001): 406.

Table 2

Author	Title	Year of Publication	Wordcount
Henry Brougham*	Forsyth's *Remarks on Italy*	1814	2086
Henry Brougham	Dr King's *Memoirs*	1819	2461
Henry Brougham	Lord Nelson's *Letters to Lady Hamilton*	1814	4469
Henry Brougham*	Restoration of Poland	1814	2960
Henry Brougham*	*Rome, Naples and Florence* [by Stendhal]	1817	2585
William Hazlitt	*Christabel [Examiner]*	1816	794
William Hazlitt	Coleridge's *Lay Sermon*	1816	4026
William Hazlitt	Coleridge's *Statesman's Manual [Examiner]*	1816	1696
William Hazlitt and Francis Jeffrey*	Hunt's *Story of Rimini*	1816	1790
Francis Jeffrey	Miss Edgeworth's	1817	5807

	Tales		
Francis Jeffrey	Rogers' *Human Life*	1819	2075
Francis Jeffrey	Wordsworth's *White Doe*	1815	2145
Thomas Moore	*French Official Life*	1826	3168
Thomas Moore*	Jorgenson's *Travels*	1817	5230
Thomas Moore	Irish Novels	1826	4170
Sir Francis Palgrave	Herbert's *Helga*	1815	4526
unknown *	Coleridge's *Christabel*	1816	2686

The test set comprised, besides the *Christabel* review, one article per author over 3,000 words in length (with the exception of Mackintosh and Allen, for whom no articles were included in the test set); several articles between 2,000 and 3,000 words in length for those authors that had contributed such pieces (Brougham, Hazlitt, Jeffrey); some very short articles under 2,000 words in length for Brougham, Hazlitt and Jeffrey. Included among the test set were truly anonymous articles (the *Christabel* review), others whose attribution in the *Wellesley Index* is probable, and others where attribution is based on solid documentary evidence, such as a mention in an author's correspondence or a reprint in their personal essay collections.

Of the supervised methods employed, Support Vector Machines (SVM) yielded the highest overall percentage of correct attributions of the test set at 74% correct attributions over 1,600 tests on individual articles. We employed a linear kernel and set the cost of constraints violation at 1 (the default value for Stylo). Stylo analysed the training set and formulated a set of rules, or classifier, based on the stylistic patterns detected in the training set. It

subsequently applied the classifier to the test set. The inclusion of articles of known

authorship allowed us to assess the accuracy of the classifier's deductions. After preliminary

tests, culling was found not to improve significantly the performance of the classifier, and

was therefore not employed in the final set of analysis. Tests were performed on the 16

articles in the test set, at 50-500 MFW in increments of 50 MFW, with 10-fold cross-

validation, randomly swapping and substituting texts between the training and test sets a

total of 10 times. Each article in the test set was thus tested 100 times.[57]

Evaluation

Neither unsupervised nor supervised methods provided a probable attribution for the

Christabel review.

The Delta-based unsupervised Cluster Analysis was able to correctly isolate certain patterns

in the data, and to unveil the existence of others, such as the stylistic similarities between

Jeffrey's articles and Hazlitt's. However, it indicated that the *Christabel* review does not bear

immediate resemblance to the contemporary literary articles published in the *Edinburgh*

Review. Even with a tolerant consensus strength of 0.5, no article emerged as its nearest

neighbour, thus adding a stylometric corroboration of the difficulties faced by traditional

attributionists.

The supervised analysis with SVM provided further confirmation of these uncertainties. The

classifier identified Jeffrey as the most likely author in 63% of the tests, with Moore second

in 28%, followed by Brougham in 8% and Mackintosh in 1%. Out of 100 tests, the classifier

did not identify William Hazlitt as the likely author of the *Christabel* review on even one

occasion.[58]

Puzzled by the total absence of what is the leading candidate in traditional attribution, we

analysed in greater detail the performance of the classifier on those texts that shared

---

[57] Eder and Rybicki, "Birds of a Feather": 230-31 suggest that even 10-fold validation may be insufficient for literary attribution tests. It is nonetheless a widely employed validation method.
[58] This total absence of attribution to Hazlitt was replicated in the preliminary tests as well.

significant traits with *Christabel*, to judge its performance in these specific areas:

1. attribution to William Hazlitt
2. article length between 2,000-3,000 words
3. editorial intervention by Francis Jeffrey

The four articles by William Hazlitt included in the test set were all attributed successfully to their rightful author, as shown in Table 3 below.

Table 3

Title	Wordcount	First candidate	Second Candidate	Third Candidate
Christabel [*Examiner*]	794	Hazlitt (73%)	Moore (25%)	Mackintosh (2%)
Coleridge's *Lay Sermon*	4,026	Hazlitt (100%)	-	-
Coleridge's Statesman's Manual [*Examiner*]	1,696	Hazlitt (75%)	Mackintosh (13%)	Brougham (8%)
Hunt's Story of Rimini	1,790	Hazlitt (61%)	Jeffrey (21%)	Moore (18%)

The longest text by Hazlitt in the test set, the review of Coleridge's *Lay Sermon* for the *Edinburgh Review*, was attributed most convincingly, in 100% of the analyses. This result was consistent with those for the other texts above 3,000 words in length, which were attributed correctly in between 93% and 100% of the tests (with one exception, which will be

discussed below). The two short reviews of Coleridge that Hazlitt published for the *Examiner* newspaper were also attributed to him, though with a lower margin of success, in 73% of the analyses for his review of *Christabel* and 75% for the review of the *Lay Sermon*. This behaviour is very likely to be influenced by the short length of both articles, with the *Christabel* review for the *Examiner* being the shortest text in the test set at 794 words. Hazlitt nonetheless emerges as the most likely author of both, thus suggesting that his total absence from the list of candidates for the *Edinburgh* review of *Christabel* is not due to the model's inability to identify Hazlitt's authorial fingerprint. Even in the case of an article where editorial intervention by Jeffrey is likely to have been extensive, such as the review of Hunt's *Story of Rimini*,[59] Hazlitt's voice does not seem to be obliterated but merely confused. It is also worth noting that, outside of the four articles in Table 3, Hazlitt is never indicated as the author of any other article in any of the tests we performed with SVM.

At 2,686 words, the review of *Christabel* may be considered by some scholars to be below the minimum length for reliable attribution (see page 17). To assess whether its short length could be responsible for the lack of a clear attribution, Table 4 contains the results of the tests performed on all remaining articles below 3,000 words in length.

Table 4

Article	Traditional attribution	First candidate	Second candidate	Third candidate
Forsyth's *Remarks on Italy*	Brougham*	Jeffrey (100%)		
Dr King's *Memoirs*	Brougham	Brougham (85%)	Jeffrey (13%)	Mackintosh (2%)

---

[59] Wu in Hazlitt, *New Writings*, 177-8, believes the article is Hazlitt's except for three paragraphs by Jeffrey that are more critical of Hunt's poetry.

Restoration of Poland	Brougham*	Brougham (36%)	Jeffrey (36%)	Mackintosh (26%)
Rome, Naples and Florence [by Stendhal]	Brougham*	Brougham (81%)	Palgrave (11%)	Jeffrey (7%)
Rogers' *Human Life*	Jeffrey	Jeffrey (100%)		
Wordsworth's *White Doe*	Jeffrey	Jeffrey (78%)	Brougham (8%)	Palgrave (8%)
Coleridge's *Christabel*	unknown*	Jeffrey (63%)	Moore (28%)	Brougham (8%)

With the articles of certain attribution, the classifier correctly identifies the author in four out of four cases, with a degree of success varying from 100% to 78% of the tests. With those where attribution in the *Wellesley Index* is founded on less secure evidence, results are more varied. The review of Forsyth's *Remarks on Italy* is attributed in 100% of the tests to Jeffrey, and not to its alleged author Brougham. However, when investigated, the *Wellesley* attribution appears more conditional, based on the article presenting itself as a continuation of another written by Brougham, and on its containing one single word that is characteristic of his style. Neither is conclusive proof of Brougham's authorship. Indeed, coupled with such a consistent stylometric attribution to Jeffrey, they indicate that the latter's possible involvement should be further investigated.

The other short article where the classifier encounters no success is the article on the "Restoration of Poland," which the *Wellesley* attributes to Brougham based on a list of his own attributions that he compiled in 1855, over fifty years after the beginning of the

Edinburgh Review, and that "must be used with great caution."[60] In this case, the classifier seems truly unable to indicate a most likely author, with Brougham and Jeffrey both identified in 36% of the tests and Mackintosh close behind in 26%. The article could be a case of multiple authorship, or have been produced by an author not included in the training set.

Overall, when Hazlitt's short articles are taken into account, the classifier achieves an encouraging degree of success, with seven correct attributions, one possible disattribution, and only one case of apparent confusion. The short length of the review of *Christabel* appears, therefore, not to be sufficient cause to judge its stylometric attribution impossible or unreliable.

There remains the possibility of multiple authorship at play in the article. The results for the review of *Christabel* are most similar to those of Hazlitt's review of Hunt's *Story of Rimini*, where multiple authorship is a very strong likelihood. Indeed, there is even the possibility of the author of the *Christabel* review having deliberately obfuscated his style to escape detection. Recent studies have demonstrated that it is possible for even non-professional writers to deliberately imitate the style of another or to alter their own in such a way as to make stylometric analysis unreliable.[61] While such a possibility cannot be fully discounted, a more likely scenario, given the editorial practices of the *Edinburgh Review*, is that editor Francis Jeffrey may have altered the text he received. Jeffrey is known to have applied numerous "retrenchments and verbal alterations," to Hazlitt's articles in at least two other occasions,[62] and to have extended this practice to all *Edinburgh* contributors.[63] Depending

---

[60] "The Edinburgh Review, 1802-1900." *Wellesley Index to Victorian Periodicals.* http://wellesley.chadwyck.co.uk.libezproxy.open.ac.uk/toc/toc.do?id=JID-ER&divLevel=1&action=new&queryId=#scroll

[61] Patrick Juola, and Darren Vescovi. "Empirical Evaluation of Authorship Obfuscation Using JGAAP." In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security.* AISec '10. New York, NY, USA: ACM, 2010: 14–18 and Michael Brennan, Sadia Afroz, and Rachel Greenstadt. "Adversarial Stylometry: Circumventing Authorship Recognition to Preserve Privacy and Anonymity." *ACM Trans. Inf. Syst. Secur.* 15, no. 3 (November 2012): 12:1–12:22.

[62] See William Hazlitt to Francis Jeffrey, 20 April 1815 in *The Letters of William Hazlitt* ed. Herschel Moreland Sikes, William Hallard Bonner and Gerald Lahey (London and Basingstoke: Macmillan, 1979), 140.

[63] John Leonard Clive. *Scotch Reviewers. The Edinburgh Review, 1802-1815.* London, Faber and

on the extent of his participation, it could be argued that any or all of the reviews in the *Edinburgh* have actually two authors. In addition, it is possible that others among the *Edinburgh*'s inner core of contributors may have participated in editorial activities. Mike Kestemont, Maciej Eder and Jan Rybicki are developing an extension of the Stylo package called "Rolling Delta" specifically to test for multiple authorship and have used it with some success in both research and pedagogical contexts.[64] When further developed, employing it for testing the review of *Christabel* could allow us to research this important hypothesis.

One further mention should be given to the other case in which the classifier almost univocally rejects an existing attribution: the review of Jorgenson's *Travels*, published in the *Edinburgh* in 1817 and attributed by the *Wellesley Index* to Moore on the basis of a payment received from Jeffrey for an article on an "M. de J."[65] The classifier attributes it in 98% of the tests to Brougham, and in only 2% to Moore. At 5,230, this is the only article above 3,000 words in length not to be confidently attributed to its traditional author. The two other texts by Moore in the test set, "Irish Novels" (4,170 words) and "French Official Life" (3,168) are attributed correctly in 93% and 98% of the tests respectively. Palgrave's "Helga", at 4,625 words, is classified correctly in 95% of the tests. The long articles by Jeffrey, Brougham and Hazlitt are all attributed to them in 100% of the tests. Given the circumstantial nature of the evidence of Moore's authorship, the review of Jorgenson's *Travels* requires further scrutiny in light of the stylometric evidence pointing to a disattribution to Moore. If the stylometric evidence is correct, there is then the possibility that another unidentified article by Moore is hidden in the *Edinburgh Review*.

Conclusion

Fabers, 1957, 56-57.

[64] Jan Rybicki, M. Kestemont and D. Hoover. "Collaborative authorship: Conrad, Ford and Rolling Delta." Digital Humanities 2013: Conference Abstracts. Lincoln: University of Nebraska-Lincoln. http://dh2013.unl.edu/abstracts/ab-121.html. Maciej Eder, Jan Rybicki, and Mike Kestemont, "Testing Rolling Delta," accessed January 21, 2014, https://sites.google.com/site/computationalstylistics/projects/testing-rolling-delta

[65] Jeffrey to Moore, 30 May 1818. Thomas Moore. *Memoirs, Journals and Correspondence of Thomas Moore*. Edited by John Russell. Vol. 2. 8 vols. (London: Longman, Brown, Greene, and Longmans, 1853), 138.

What have we learned from the experiments and analysis that we have conducted so far?

We have not yet discovered any evidence that could lead to a convincing attribution of the authorship of the *Christabel* review. Consequently, we are not in a position to suggest that this controversial intervention in the reception history of the Romantics came from the pen of Thomas Moore. Nor, however, have we dispensed with Moore's candidacy. Under certain conditions, as the results of Support Vector Machine analysis show, he remains a plausible candidate for authorship. That method, in fact, identifies Moore as a more likely author than Hazlitt, the other major suspect in the field, and this discovery is a corrective to the balance of recent critical opinion that has favoured Hazlitt.

Consequently, those questions which represented the prospect of modest but meaningful advancements in Irish Studies remain unresolved. Some further detail on these topics will underline the value of continuing to seek answers. Firstly, the issue with Lord Byron concerns dismissive remarks made in the review. A friend and confidante of Moore for five years, Byron was conflated with Lord Edward Fitzgerald in the undesirable character of Lord Glenarvon in Lady Caroline Lamb's *Glenarvon* (1816). Moore wrote an unpublished review of *Glenarvon* for the *Edinburgh* (sometime during June 1816), ridiculing the novel and (presumably) defending Byron.[66] Is it possible that, in the same year, he would write of Byron in the *Christabel* review: "Great as the noble bard's merits undoubtedly are in poetry, some of his latest *publications* dispose us to distrust his authority, where the question is what ought to meet the public eye; and the works before us afford an additional proof, that his judgment on such matters is not absolutely to be relied on"?[67] If we admit the possibility of multiple authorship, it is possible that this comment might have originated from the *Edinburgh*, instead of from the review's primary author. The journal's staff included persons, like Henry Brougham, with a well-publicised animosity towards Byron. However, if a

---

[66] Moore, *The Letters of Thomas Moore*, 1:401. Francis Jeffrey to John Allen, 25 June 1816 (quoted in Stanley Jones. *Hazlitt. A Life*, 224n), stated that he had received "a very clever and severe review of it [*Glenarvon*] from Tommy Moore [. . .] full of contempt and ridicule" and that he was "very much tempted to insert it" but had decided with his collaborators not to review *Glenarvon* at all. Duncan Wu also emphasises the temporal coincidences between the publication of *Christabel* and *Glenarvon* (Wu, *William Hazlitt*, 189).

[67] Review of *Christabel: Kubla Khan, a Vision; The Pains of Sleep*, by Samuel Taylor Coleridge: 58-9.

comment so at variance with his own opinions had been inserted into Moore's article, would

he have continued to write for the *Edinburgh Review*, and to value its opinions?

Secondly, another node in this interesting network of associations is that Moore would later

write biographies of both the figures attacked in *Glenarvon*, Byron in 1830 and Lord Edward

Fitzgerald in 1831. Indeed, 1816 is an important year in Moore's career as a biographer. Not

only did it mark Byron's departure from England, but Moore's first biographical subject

(1825), Richard Brinsley Sheridan, died on 7 July 1816. In the context of his career as a

biographer, and particularly for the sake of understanding his usually honest and forthright

relationship with Byron, it is important to discover how the *Christabel* review figures in the

events of that summer.

Thirdly, this work aims to clarify the important question of the extent of Moore's writing for

the *Edinburgh*. In the course of seeking an attribution for the *Christabel* review, our analysis

cast serious doubts about the attribution of the review of Jorgensen's *Travels* to Moore. The

possible attribution of the article to Brougham and consequent uncertainty about the identity

of Moore's *Edinburgh* article on "M. de J." reveals as many as three possible changes to the

Moore *Edinburgh* canon, representing a potentially significant revision to his corpus of

writings.

Our work to this point has revealed a number of possibilities about the text: there is an

increasingly plausible chance that the review is a case of multiple authorship. The elusive

nature of the review's style, and its evasion of consistent identification with any one of the

candidate authors, may well be a result of the article being edited by Jeffrey or another

member of the *Edinburgh*'s editorial staff. Kestemont, Eder and Rybicki's work on "Rolling

Delta" is a very welcome development, not least because it may help in untangling the

intractable textual web of the *Christabel* review, and shedding consequent light on the tense

personal relationships that heighten the intrigue of the article's authorship.[68] In more general

terms, this tool would aid authorship attribution in coming to terms with the reality that

---

[68] The nature of Coleridge and Hazlitt's acquaintance is particularly volatile and bitter. See Beer,
"Coleridge, Hazlitt, and 'Christabel.'"

multiple authorship, "the collaborative authorship of writings that we routinely consider the work of a single author,"[69] is more common than literary studies allows.

Future directions for our study will focus on the two main hypotheses opened up by the stylometric evidence we have gathered so far: that of multiple authorship, and of the possible involvement of an author not included in the training set. The first hypothesis will be tested through the application of "Rolling Delta" to the *Christabel* review and to those other articles, such as Hazlitt's review of *The Story of Rimini*, where there is evidence of multiple authorship. The second hypothesis will be examined through the inclusion of those articles by less frequent contributors originally excluded from the training set, such as Douglas Kinnaird, John Cam Hobhouse and Sir Walter Scott, and of *Edinburgh Review* founder Sydney Smith, usually a prolific contributor but otherwise absent during 1814-1818. Duncan Wu's contention that the *Christabel* review differs from Hazlitt's usual style because of a lack of intervention by Jeffrey[70] will also be explored by including in the training set some of his critical essays that were not published in the *Edinburgh Review*, such as those he collected in *The Round Table* (1817). Different measures of stylometric analysis, such as n-grams, part-of-speech tags and syntactic analysis will also be tested, now that the apparently promising technique of most frequent words has proven inconclusive.

More broadly, this project has highlighted some interesting points about the nature of text, and foregrounded some important issues about the apparently subjective and objective characters of different kinds of scholarly evidence. Humanities data is not like scientific data. In many cases, it consists of historical documentary records, rather than data gathered from observation or experiment. Not only this, but its inheritance from an agent other than the investigator makes the task of establishing its veracity more difficult.[71] Judgement is inescapable at every stage of the process described above, but in particular, the preparation

---

[69] Jack Stillinger, *Multiple Authorship and the Myth of Solitary Genius* (New York ; Oxford: Oxford University Press, 1991), 22.
[70] See Hazlitt, New Writings, 204.
[71] Christine L. Borgman, "The Digital Future Is Now: A Call to Action for the Humanities."*Digital Humanities Quarterly* 3, no. 4 (Autumn 2009): paragraph 33.
http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html

of the corpus allows for the presence of multiple variables: from the choice of authors, to the exclusion of certain topics, to the determination of chronological limits. Even decisions about the texts themselves—whether all quotations should be removed to preserve the authorial fingerprint; if apparent misspellings should be corrected—highlight the fact that text is not like other kinds of data. Text and its interpretation are influenced by genre, gender, and time, and language bears a polysemic burden absent from numbers. While statistical methods like those we have employed result in a form of quantifiable evidence not available to traditional attribution methods, these results alone will no more resolve this puzzle than the knowledge that Hazlitt bore a grudge against Coleridge.

We share an expectation with John Burrows: that our work will not yield finality but rather open further questions.[72] Without the evidence to locate a work at a particular desk in a particular room, the hypotheses of the attributionist are always uncertain and subject to revision. Our work on the *Christabel* review has revealed, among other things, that there are many good reasons why no attempt at attribution has been satisfactory to this date. Nor, we suspect, is this attempt. The puzzle remains, and attempts to solve it will continue, somewhat comforted by the knowledge that "statistical analysis deals in probabilities and not in certainties."[73]

---

[72] John Burrows. "Questions of Authorship": 26.
[73] *Ibid.*

Bibliography

Barthes, Roland. *S/Z*. Translated by Richard Miller. New York: Hill and Wang, 1974.

Beer, John. "Coleridge, Hazlitt, and 'Christabel.'" *Review of English Studies* 37, no. 145
    (1986): 40–54.

Benatti, Francesca. "Joining the Press-Gang: Thomas Moore and the Edinburgh Review." In
    Thomas Moore: Texts, Contexts, Hypertext, edited by Francesca Benatti, Sean Ryder,
    and Justin Tonra, 167–85. Oxford: Peter Lang, 2013.

Benatti, Francesca, Sean Ryder, and Justin Tonra, eds. *Thomas Moore: Texts, Contexts,*
    Hypertext. Oxford: Peter Lang, 2013.

Borgman, Christine L. "The Digital Future Is Now: A Call to Action for the Humanities."
    Digital Humanities Quarterly 3, no. 4 (Fall 2009).
    http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Brennan, Michael, Sadia Afroz, and Rachel Greenstadt. "Adversarial Stylometry:
    Circumventing Authorship Recognition to Preserve Privacy and Anonymity." *ACM*
    *Trans. Inf. Syst. Secur.*15, no. 3 (November 2012): 12:1–12:22.
    doi:10.1145/2382448.2382450.

Burrows, John. "All the Way Through: Testing for Authorship in Different Frequency Strata."
    Literary and Linguistic Computing 22, no. 1 (April 1, 2007): 27–47.
    doi:10.1093/llc/fqi067.

———. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Lit*
    Linguist Computing 17, no. 3 (September 1, 2002): 267–287. doi:10.1093/llc/17.3.267.

———. "Questions of Authorship: Attribution and Beyond: A Lecture Delivered on the
    Occasion of the Roberto Busa Award ACH-ALLC 2001, New York." *Computers and*
    the Humanities 37, no. 1 (February 2003): 5–32. doi:10.1023/A:1021814530952.

Byron, George Gordon. *The Works of Lord Byron: With His Letters and Journals, and His*
    Life. Edited by Thomas Moore. 16 vols. London: John Murray, 1832.

Carver, P. L. "The Authorship of a Review of 'Christabel' Attributed to Hazlitt." *Journal of*

English and Germanic Philology 29, no. 4 (1930): 562–78.

Chandler, James. *England in 1819: The Politics of Literary Culture and the Case of Romantic Historicism*. Chicago ; London: University of Chicago Press, 1998.

Clive, John Leonard. *Scotch Reviewers. The Edinburgh Review, 1802-1815*. London: Faber and Faber, 1957.

Coburn, Kathleen H. "Who Killed Christabel?" *TLS* (May 20, 1965): 397.

Coleridge, Samuel Taylor. *Biographia Literaria; Or, Biographical Sketches of My Literary Life and Opinions*. 2 vols. London: Rest Fenner, 1817.

———. *Christabel: Kubla Khan, a Vision; The Pains of Sleep.* London: John Murray, 1816.

Dowden, Wilfred S. "Thomas Moore and the Review of 'Christabel.'" *Modern Philology* 60, no. 1 (1962): 47–50.

Eder, Maciej. "Computational Stylistics and Biblical Translation: How Reliable Can a Dendrogram Be?" In *The Translator and the Computer*, 155–170. Philological School of Higher Education, Wroclaw: WSF Press, 2013.

———. "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." *Literary and Linguistic Computing* (November 14, 2013): 1-16. doi:10.1093/llc/fqt066.

———. "Mind Your Corpus: Systematic Errors in Authorship Attribution." *Literary and Linguistic Computing* 28, no. 4 (December 1, 2013): 603–614. doi:10.1093/llc/fqt039.

Eder, Maciej, and Jan Rybicki. "Do Birds of a Feather Really Flock Together, or How to Choose Training Samples for Authorship Attribution." *Literary and Linguistic Computing* 28, no. 2 (June 1, 2013): 229–236. doi:10.1093/llc/fqs036.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. "Testing Rolling Delta." Accessed January 21, 2014. https://sites.google.com/site/computationalstylistics/projects/testing-rolling-delta

Grieve, Jack. "Quantitative Authorship Attribution: An Evaluation of Techniques." *Literary and Linguistic Computing* 22, no. 3 (September 1, 2007): 251–270. doi:10.1093/llc/fqm020.

Hazlitt, William. "Coleridge's Lay-Sermon," Review of *The Statesman's Manual*, by Samuel

Taylor Coleridge. *Edinburgh Review* 27 (1816): 444–459.

———. "Coleridge's Literary Life," Review of *Biographia Literaria*, by Samuel Taylor

Coleridge. Edited by Francis Jeffrey. *Edinburgh Review* 28 (1817): 488–515.

———. *The Letters of William Hazlitt*. Edited by Gerald B. Lahey. New York University

Press, 1978.

———. *New Writings of William Hazlitt*. Edited by Duncan Wu. Oxford; New York: Oxford

University Press, 2007.

———. Review of *Christabel: Kubla Khan, a Vision; The Pains of Sleep*, by Samuel Taylor

Coleridge. *The Examiner* no. 440 (June 2, 1816): 348–49.

———. "Sismondi's Literature of the South," Review of *De la Litterature du Midi de l'Europe*,

by J. C. L. Simonde de Sismondi. *Edinburgh Review* 25 (1815): 31–63.

———. "Standard Novels and Romances," Review of *The Wanderer: or, Female Difficulties*,

by Madame d'Arblay. *Edinburgh Review* 24 (1815): 320–338.

———. *A View of the English Stage: Or, a Series of Dramatic Criticisms*. London: John

Warren, 1821.

Holmes, David I., Lesley J. Gordon, and Christine Wilson. "A Widow and Her Soldier:

Stylometry and the American Civil War." *Literary and Linguistic Computing* 16, no. 4

(November 1, 2001): 403–420. doi:10.1093/llc/16.4.403.

Hoover, David L. "Frequent Word Sequences and Statistical Stylistics." *Literary and

Linguistic Computing* 17, no. 2 (June 1, 2002): 157–180. doi:10.1093/llc/17.2.157.

———. "Testing Burrows's Delta." *Literary and Linguistic Computing* 19, no. 4 (November 1,

2004): 453–475. doi:10.1093/llc/19.4.453.

Hutchinson, Thomas. "Coleridge's Christabel." *Notes and Queries* X. 9th (1903): 170–2.

Jockers, Matthew L. "Testing Authorship in the Personal Writings of Joseph Smith Using

NSC Classification." *Literary and Linguistic Computing* 28, no. 3 (September 1, 2013):

371–381. doi:10.1093/llc/fqs041.

Jockers, Matthew L., and Daniela M. Witten. "A Comparative Study of Machine Learning

Methods for Authorship Attribution." *Literary and Linguistic Computing* 25, no. 2 (June

1, 2010): 215–223. doi:10.1093/llc/fqq001.

Jones, Stanley. *Hazlitt : a Life, from Winterslow to Frith Street*. Oxford: Clarendon Press,

1989.

Jordan, Hoover H. "Thomas Moore and the Review of 'Christabel.'" *Modern Philology* 54, no.

2 (1956): 95–105.

Juola, Patrick. "Authorship Attribution." *Foundations and Trends in Information Retrieval* 1,

no. 3 (2007): 233–334. doi:10.1561/1500000005.

Juola, Patrick, and Darren Vescovi. "Empirical Evaluation of Authorship Obfuscation Using

JGAAP." *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security*,

14–18. AISec '10. New York, NY, USA: ACM, 2010. doi:10.1145/1866423.1866427.

Kelly, Linda. *Ireland's Minstrel: A Life of Tom Moore*. London: IBTauris, 2006.

Kelly, Ronan. *Bard of Erin: The Life of Thomas Moore*. Dublin: Penguin Ireland, 2008.

"Leigh Hunt Letters." Accessed January 21, 2014.

http://digital.lib.uiowa.edu/cdm/landingpage/collection/leighhunt

Love, Harold. *Attributing Authorship: An Introduction*. Cambridge: Cambridge University

Press, 2002.

Luyckx, Kim, and Walter Daelemans. "The Effect of Author Set Size and Data Size in

Authorship Attribution." *Literary and Linguistic Computing* 26, no. 1 (April 1, 2011): 35–

55. doi:10.1093/llc/fqq013.

McGann, Jerome J. *Radiant Textuality: Literature after the World Wide Web*. New York ;

Basingstoke: Palgrave, 2001.

Moore, Jane, ed. *British Satire, 1785-1840: Satires of Thomas Moore*. Vol. 5. 5 vols.

London ; Brookfield, VT: Pickering & Chatto, 2003.

Moore, Thomas. *The Letters of Thomas Moore*. Edited by Wilfred S. Dowden. 2 vols.

Oxford: Clarendon Press, 1964.

———. *Memoirs, Journals and Correspondence of Thomas Moore*. Edited by John Russell.

8 vols. London: Longman, Brown, Greene, and Longmans, 1853-1856.

———. *Memoirs of Captain Rock*. Edited by Emer Nolan and Seamus Deane. Dublin: Field

Day, 2008.

———. *The Unpublished Letters of Thomas Moore.* Edited by Jeffery W. Vail. 2 vols. London: Pickering & Chatto, 2013.

Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism.* Urbana: University of Illinois Press, 2011.

Review of *Christabel: Kubla Khan, a Vision; The Pains of Sleep*, by Samuel Taylor Coleridge. *Edinburgh Review* 27 (1816): 58–67.

Rudman, Joseph. "The State of Non-Traditional Authorship Attribution Studies—2012: Some Problems and Solutions." *English Studies* 93, no. 3 (2012): 259–274. doi:10.1080/0013838X.2012.668785.

Rybicki, Jan, M. Kestemont and D. Hoover. "Collaborative authorship: Conrad, Ford and Rolling Delta." Digital Humanities 2013: Conference Abstracts. Lincoln: University of Nebraska-Lincoln. http://dh2013.unl.edu/abstracts/ab-121.html

Schneider, Elisabeth. "The Unknown Reviewer of Christabel: Jeffrey, Hazlitt, Tom Moore." *PMLA* 70, no. 3 (1955): 417–32.

———. "Tom Moore and the Edinburgh Review of Christabel." *PMLA* 77, no. 1 (1962): 71–76.

Shattock, Joanne. *Politics and Reviewers: The "Edinburgh" and the "Quarterly" in the Early Victorian Age.* Leicester: Leicester UP, 1989.

Stillinger, Jack. *Multiple Authorship and the Myth of Solitary Genius.* New York ; Oxford: Oxford University Press, 1991.

"The Wellesley Index to Victorian Periodicals, 1824-1900." Accessed January 21, 2014. http://wellesley.chadwyck.co.uk/home.do.

Wu, Duncan. *William Hazlitt: The First Modern Man.* Oxford and New York: Oxford University Press, 2008.