Attribution Accuracy Verification by Comparing Stylometric Conclusions Against Patterns in the Publishing Credits

Anna Faktorovich
Anaphora Literary Press
1108 W 3rd Street, Quanah TX 79252, USA
E-mail: director@anaphoraliterary.com

Abstract:

The analysis of bibliographic textual information must be an essential step in any stylometric authorship attribution study. This article describes patterns among the credits given on 18th and 19th century British title-pages to printers and booksellers. These patterns are contrasted against the stylometric findings of a new method that was applied to 634 texts from these centuries. The biographies of these printers and booksellers are explored for relevant evidence. And the history of publishing is summarized to explain the significance of the findings.

Keywords: bibliography, stylometry, computational-linguistics, attribution, authorship, publishing, printing, bookselling

Across the different computational-linguistics attribution studies, such as Hugh Craig's,[1] none go beyond listing the tested currently accepted bylines and book titles. And in larger corpuses, such as Zhao and Zobel's,[2] even the titles might not be included, as it is assumed that readers can figure out what 42 texts have been assigned to a byline such as "Shakespeare". This absence can be blamed on the disconnection between the computer scientists designing and applying programs to test stylometric features to reach quantifiable attribution conclusions, and the linguists and literature scholars who later write introductions to canonical books using these bylines. The recording and analysis of bibliographic data, together with the history of how bylines first came to be assigned to texts must be a required step in any authorship attribution methodology. Both the "Christopher Marlowe" and the "Jane Austen" bylines were first-assigned by publishers without any linguistic evidence to previously anonymous texts, after their deaths, centuries apart, in 1593 and 1817. Such profiteering posthumous or ghostly byline assignments are not exceptions, but rather engulf more than these percentages of the tested corpuses (not counting initialed or otherwise cryptically-bylined texts): 24% of the Renaissance, 14% of the 18th century, and 11% of the 19th century. None of these previously posthumously or post-first-publication assigned bylines can be counted as "undisputed", since a dispute of these texts' anonymity first added their bylines. Some bylines have been changed several times over the centuries. For example, poems that carried the "*Ignoto*" anonymous byline in "Francis Davison's" *Poetical Rhapsody* (1611), and *English Helicon* (1600), were also published in the Renaissance under

the "Christopher Marlowe"-byline, and then in the 19th century they were re-assigned to "Sir Walter Raleigh" in Egerton Brydges' *The Poems of Sir Walter Raleigh: Now First Collected* (1814).

The history of authorship in Britain cannot be separated from the history of publishing and the profiteering tactics of publishers, stage managers and other beneficiaries from this trade in words. British publishers or their editors have made most of the first byline assignments to anonymous texts, including adding the "Austen" byline posthumously. Since the publishers were forming contracts with authors or their ghostwriters, they are indeed the business entity most capable of explaining authorial trends. This article is a fragment out of the first volume, *Introduction to Stylometry for Attribution*, out of a new series that re-attributes texts from the 18th and 19th centuries. Previously, I published the 20-volume British Renaissance Re-Attribution and Modernization series (BRRAM).[3] In both studies I use a new stylometric attribution method that combines 23-27 different types of linguistic tests for punctuation, parts-of-speech and lexical elements to cluster texts by their linguistic-signatures. BRRAM's analysis of the Renaissance corpus concluded that 2 out of the 6 ghostwriters also held publishing monopolies: William Byrd in music books in Britain, and Richard Verstegan in exiles' Catholic books. When European governments granted legal publishing monopolies to individuals, they invited them to restrict access to the press, and to manipulate these markets to their financial benefit. The small ghostwriting workshops I uncovered in my attribution studies of Britain between the Renaissance and 1899 are intricately linked to the publishers of these works. Thus, my quantitative linguistic attributions are strengthened when they agree with the conclusions that can be drawn from patterns in their bibliographic information.

To unify the quantitative attributions to this bibliographic information, this is a fragment out of the section that analyzes the patterns in the "Publication Data" tab of the spreadsheets of 18th and 19th century corpuses. The columns of greatest interest are the publishers or printers, the booksellers, the theaters, and the history of byline assignments. This section only addresses patterns among the printers and booksellers listed on title-pages.

The patterns of highest interest are those that show exclusivity between a linguistic-group or a ghostwriter and one of these bibliographic elements. One such revealing pattern is that all dramas in the 18th century corpus fall into the linguistic B-group. The B-group must have been correctly grouped for there to be such generic sorting precision. After a pattern of this type is observed in this bibliographic summary, the documentary evidence that explain its deeper significance will be explored in later biographical and historical sections of this study.

Printers, Booksellers and Publishers

In the 18th and 19th centuries, some of the best-known "authors" were also credited as their own publishers. These author-publishers include "Samuel Richardson", "Charles Dickens", "Walter Scott" and "William Blake"; they printed, sold, advertised, registered, and sponsored the releases of their own works, and these self-publications and self-promotions achieved greater fame for them than for the majority of authors who used otherwise-bylined publishers. The line between authors

and publishers was also blurred when publishers, such as "Charles Gildon" and "Edmund Curll", assigned their bylines to prefaces, letters and other content. Some of the same publishers who first assigned puffed bylines, such as "Austen", were also advertising themselves as authors. Given these overlaps, it would be absurd to research authorship attributions without considering the likelihood that at least some of the profitable publishers were themselves the underlying ghostwriters behind linguistically-matching texts from multiple different bylines.

A few term definitions are necessary before exploring the bibliographic patterns in publishing credits. The title-page of 18th and 19th century books typically listed two different entity types, in place of a modern *publisher*. The bookseller was the owner of a bookshop where books were "sold at..." The printer was the owner of a printshop that books were "printed by..." Stationer was a more general term for employees in the book and periodical design and production industries, including not only the pressing of printed books, but also the papermaking, packaging, and design. There were licenses that booksellers and printers had to obtain to operate, but those listed as booksellers and printers were likely to have apprentices or employees who performed the manual labor. The Licensing Act lapsed in 1695, and this lifted previous limits on the number of presses that could operate in England, and the number of apprentices one master could sponsor.[4] The Stationers' Company held a pre-publication monopoly over the publishing industry in Britain until its grip was loosened by the Copyright Act of 1710; this Act was assisted by a propaganda campaign, attributed mostly to "Daniel Defoe", in books such as the initially-anonymous and printer/bookseller-lacking *An Essay on the Regulation of the Press* (1704). *Essay* argued that licensors had held power to bar a publication license to an "author" whose political beliefs they disliked; it recommended that the credited "author" in the byline should be held responsible as the "seller" post-publication for the contents of the book, unless the names of the "Printer or Bookseller" is also "affixed to the book". Thus, "no book can be published, but there will be somebody to answer for it. Whoever puts a false name, to forfeit... etc."[5] The last phrase specifies the trick the anonymous author of this *Essay* was anticipating to avoid blame for books' content: the utilization of "false" or fictitious authorial, bookselling and printing bylines, or using "strawmen" or "fronts" to credit unknowing parties to protect the underlying ghostwriters and ghost-publishers profiting from the larger market-monopolization. Before this Act, stationers were the main licensors who could refuse publication rights; while after the Act, publishers merely had to continue paying them to register copyright claims, while the censorship role shifted more firmly to post-publication oversight by the Attorney-General, Parliament and the courts, who could raise objections over issues such as sedition and libel. The term publisher generally referred to a book's sponsor, who paid for the printing and the copyright registration, or had invested in owning a share of a publishing business. If there was a plagiarism, byline, or copyright dispute, it was typically the publisher who took the matter to court. The publisher's name typically appears in the Stationers' Register of new books, as the publisher would have paid the publication fee.

"James Lackington's" (1746-1815) *Memoirs* offer a helpful introduction to these terms and the

tricks of the publishing industry. "Lackington" describes himself as a "present Bookseller in Chiswell-Street, Moorfields, London". A similar publishing credit is given in the tested G-group "Thomas King"-bylined *Check on Uncharitableness* (1791), which is also described as "Printed for the Author" and "sold" at, among other places, "Chiswell-Street". "Lackington" was born too late to have been the ghostwriter behind the G-group, but otherwise he seems to be a great candidate as a potential ghostwriter-publisher behind this group. "Lackington" writes in a letter addressed to potential authors: "A gentleman a few years since showed a manuscript to a publisher, which he refused to purchase, but offered to be the publisher if the gentleman would print it, etc. at his own expense, which he readily agreed to do." The author is said to have asked to sponsor the printing of as many books as there are people in Great Britain, but the publisher talked him down to 1,250, of which only 100 copies happened to have sold. Most of the books were "consigned to oblivion, through the... mismanagement of publishers". In this context, the "publisher" is merely an agent who connects an author to various other entities in the publishing business; this publisher takes money from the self-sponsoring author, keeps some of these funds, and uses the rest to pay fees to the book's printer, for the license to the registration office, and to a periodical to run a small advertisement for the book. This industrious publisher makes a large profit from performing these few administrative tasks because the author would either be rejected if he attempted to purchase these services themselves, or the author is misled into believing there is something special the publisher can do to generate sales. The trick to the publishers' profitability is to find "authors" willing to sponsor as many printed copies as possible, so that the publisher, printer, and all other collaborating parties maximize their profits. A publisher or a printer can even "purchase" 100 copies themselves, if this convinces the author a publication at least gave them some public visibility or fame. Labeling 100 out of 1,250 copies as "sold" would only discount by a small percentage the author's enormous publication bill. "Lackington" exemplifies how he and other British publishers puffed or sold this publishing process, while hiding its underhanded corruption. Booksellers unanimously refused to contract directly with authors. "Authors who become their own publishers" are said to "seldom or never answer" because a self-publisher cannot sell their books through booksellers or bookstores, and must instead sell them in their own private house, to which the public is unwilling to go. Booksellers are said to avoid self-published authors because such publications "abridge" their "usual profits", when contrasted with them not needing to give a share to the author when the author sells the copyright to the book to them; this sale of the copyright to a party other than the author seems to have been the main function of a publisher. "The more liberality authors exercise towards the trade, the greater will be their profits in the end."[6] In other words, only authors who entirely sponsor their publication, and do not contract for a percentage from the sale are welcomed by this publishing enterprise. In contrast, all professional-authors who hope to make a living from writing, or selling their writerly labor to publishers would be rejected from direct publication in this system. "Lackington's" testimony clarifies the barriers to entry to self-published authors from selling directly to booksellers, and to professional-authors from selling either directly to buyers, or to non-profit-driven publishers who

paid for pre-publication writerly-output. These barriers forced professional authors, who relied on a wage for their writerly labor, to instead sell ghostwriting services to those who wanted to pay for authorial fame through undeserved authorship bylines.

There was a steep decline in the percentage of British books' title-pages that listed printers between the Renaissance's 256 out of 303 texts (85%), to the 18th century's 77 out of 311 texts (25%). Then, this rate of printer-listings remained steady at 26% in the 19th century corpus, with 83 out of 323 texts. This rapid change in printer-listing preferences must have been tied to the end of the Stationers' monopoly and the Copyright Act of 1710. If Stationers no longer held a monopoly, they were no longer able to endorse unlimited quantities of fictitious or "false" printer bylines. When Parliament and other governing bodies gained oversight over printers, it became more advantageous to leave out the printer's name all together. A few printers could continue to monopolize the market, but now mostly under the guise of anonymity, instead of relying on false or strawman printer bylines. The growth of the power of the publisher as an intermediary between the printer and "author" also minimized the need for printers to advertise their individual services in their books. The subterfuge around printer bylines was not merely to protect individual printers from libel, sedition, and plagiarism charges, but also to protect the illusion of competition in an industry that was in fact controlled by very few players.

Another shift between the Renaissance and 18th century's printer credits was the spread of openly fictitious business names (known as trade or doing-business-as names), such as Apollo Press (Edinburgh printer listed in 1782), Literary Press (mostly a self-publishing effort of "Rev. Dr. John Trusler" at "No. 62, Wardour-Street, Soho", 1780s-1800), and Logographic Press (1784-1813). Though back in the Renaissance, there were also a few examples of trade printer credits to Eliot's Court Press, Oxford University Press and Printers to the University of Cambridge. The standard personal bylines of the stationers or licensed printers remained as the most popular type of printer-credit across the 18th and 19th centuries. The underhanded multi-byline printing monopolization of these early centuries appears to have been brought into the open during the 20-21st century mergers among publishers that resulted in only the fictitiously-named Big Five publishers monopolizing the world's book trade. In the US, monopolization statistics were cited in a failed 2021-3 lawsuit filed by book buyers against Amazon (in control of 50% of all books sold) and the Big Five publishers (controlling over 80% of the trade book market and over 40% of the overall publishing).[7] This shift to open corporate monopolization across most industries took place by the end of the 19th century, when it was noticed that John D. Rockefeller's fictitiously-named Standard Oil Company controlled 90-95% of U.S. oil refineries.[8] One of the reasons Rockefeller managed to gain such a large market share before this monopoly was broken up was because he used a corporate name, and not his own name, despite his controlling ownership of this entity. Thus, markets can be controlled merely by manipulating the names of printers; for example, the illusion of market-diversity can be generated by calling a single monopolizing publisher's output not by its solitary corporate name, but instead by the distinct names of dozens of its "imprints".

Out of the 77 texts with listed printer bylines in the 18th century corpus, only 6 printers released 2 texts, while the other 65 printers only published 1 of these canonical texts. It is hardly believable that any printer could have maintained a printing career by only publishing a single major publication. Either most of these printers were also publishing the books that did not list any printers, or a smaller group of printers were using some of the other pseudonymous bylines. Meanwhile, only 1 of these 6 same-printer pairs has 2 texts that both fall in the same linguistic group. Both texts in this pair are dramatic comedies from the B-group that were printed by "J. [John] Almon" (1737-1805) and sold at "No. 183, Fleet-Street"; they include: "Leonard MacNally's" *Robin Hood; Or, Sherwood Forest: A Comic Opera* (Covent Garden Theatre, 1784) and "Frederick Pilon's" *The Fair American: A Comic Opera* (Theatre Royal Drury Lane, 1785, 1782). While 2 publications are hardly genre-defining, this slight pattern deserves some speculation regarding this printer's potential implication in ghostwriting. Though these questions are truncated by the simple fact that John Almon was born too late to have ghostwritten the B-group. Despite this setback, further research into Almon is warranted, as he is best-known as the bookseller who was convicted of seditious libel for publishing the *Letters of Junius* (1772); this was a set of anonymous letters critical of King George III's administration. No punishment, such as a jail term, was recorded for Almon, so he might have volunteered to only publish the government's propaganda going forward in exchange for this leniency. While this pair of "Almon"-printed plays was staged at 2 different "Theatre-Royals", there is a clear pattern between these two and one more B-group play with a unique printer-credit. Only this one, earlier, 18th century play, "Moses Mendez's" *Robin Hood: A New Musical Entertainment* (1751), is credited as printed by the "Theatre-Royal" itself. This indicates that it is likely that the Theatre-Royal initially experimented with crediting itself as the self-printer, before turning to sponsoring the publications of their plays under the credits of other printers, such as "Almon". This pattern is a strong indicator that the monopolization of the B-group by a single ghostwriter was interlinked to the British monarchy's monopolization of the theatrical industry through strict licensing laws. With as few as two theaters licensed to stage plays in London across the 18th century, and both being described as propagandistic branches of the monarchy ("Theatre-Royal"), it was far easier to fully control the theatrical output by only hiring one ghostwriter to deliver the desired messages. Only 16 dramas were tested in the 19th century corpus, and these were split between 2 linguistic groups: A2 and C2. These groups are labeled as A and C in the 19th century spreadsheet, but they (and the other groups) are labeled as A2 and C2 here to distinguish them from the 18th century's A and C groups. A2 includes "Charles Millward's" *The Grand Annual Comic Christmas Pantomime Entitled, Ride a Cock Horse to Banbury Cross* (1874); this play happened to be printed by "Theatre Royal Printing Works". Most of the other 19th dramas do not list any printer. Thus, this restatement that even a 19th century play's publication was sponsored by the "Theatre Royal" confirms the monopoly this entity continued to hold over Britain's dramatic output until at least the end of the 19th century.

In the 19th century corpus, there are significantly more printer names credited with more than

1 of these canonical publications. This also means that there are fewer unique printer bylines. Out of the 85 texts with named printers, there are only 51 different printer names (including both individual bylines and business names). There was blatant consolidation of the printing industry as it shrunk from 71 named canonical printers in the 18th century to only 51 in the 19th century. In contrast with the 2-book maximum from printers in the 18th corpus, the 19th century credits most printers with *at least* 2 releases. And a few 19th century printers were especially industrious. The most productive 19th century printers are Ballantyne & Co. (called by a few variants of this name across the century) and Spottiswoode & Co. 5 out of 7 titles that Ballantyne published in London and Edinburgh are poetry collections from the B2-group.

Some of these printing credits are misleading, as different company names can refer to the same printing-business owner. An important example that deserves scrutiny is the two titles (in the G2 and J2 groups) that are credited as printed by Sir Walter Scott's Press. Walter Scott (1771-1832) was one of the founding investors in both the press that carried his byline, and in Ballantyne & Co. Scott could not have been the underlying ghostwriter of any of the 3 linguistic groups that matched books assigned to his byline because he died too early for these groups' publishing timelines. And Ballantyne & Co.'s name came from its other two initiators, James (1772-1833) and John Ballantyne (1774-1821). Both brothers were born early enough to have written the texts in the C2-group (that includes "Scott's" other Waverley novel, *Rob Roy*), but they died too early to have played this ghostwriting role. And they are chronologically entirely unlikely to have been ghostwriters behind either the B2 or K2 groups. James had started his career as a licensed-printer of "Scott's" early works, while John acted as the manager of the Ballantyne & Co. entity that Scott formed in 1808. One of the great publishing mysteries of this period is how the *Waverley* series could have had record-breaking sales, and yet John Ballantyne managed to have continuous financial troubles, before he died as a bankrupt. The idea that this series was a bestseller was puffed by J. G. Lockhart's potentially entirely fictitious posthumous *Memoirs of the Life of Sir Walter Scott* (1837), which reported that James Ballantyne had first preemptively printed the first *Waverley* novel, and then convinced Archibald Constable, the "publisher", to pay £700 for the copyright to it (a flat advance fee, to be enlarged with half of the profits, if they ended up being higher than this amount), before this novel went on to sell the entire first 1,000-copy print-run in 5 weeks, and then the second and third editions of 3,000 copies in the next 4 months.[9] If these sales statistics had been accurate and similar success was seen in the following *Waverley* novels; then, the 1825 banking crisis should not have triggered Ballantyne and Scott's bankruptcy, as they should not have needed loans, if sales exceeded expenses. It is perhaps more likely that Scott had been taking out personal loans beyond his means through this publishing business, and this debt was heightened by the crisis, and the pressure of repayment was the reason Scott agreed to allow his byline to be added to the *Waverley* novels for the first time in the 1829 edition, rightly hoping that his aristocratic and judicial status would help generate new sales. Previously, "Scott" had only sponsored a poetic byline credit on *The Lady of the Lake* (1810). The changes between the different versions of the Ballantyne & Co., Ballantyne Press:

Ballantyne, Hanson and Co., Ballantyne, Hansom & Co. and other name variants, and the alternative use of Sir Walter Scott's Press might have been indicators of these undercurrents of financial fraud tied to these businesses. A successful publishing company, without such underlying problems and in cases where new partners have not been added, would have naturally wanted to retain the same name, as it gained recognition across that century.

There is a wider stylistic diversity among Spottiswoode's 7 London printings: 3 C2-group novels, 2 rhetorical texts form I2, 1 poetry book from B2, and 1 philosophy release from J2. Spottiswoode is less commonly discussed in modern scholarship, but this printer had the more distinguished role of serving as the King's Printer, starting in 1739, when it was founded by William Strahan (-1785). Strahan was listed as the printer of one of those mentioned rare pairs of texts in the 18th century corpus that were created by the same printer; these are two poetry collection from the E and I groups: "Christopher Smart's" *Poems on Several Occasions* (1752) and "John" and "Charles Wesley's" *Hymns and Sacred Poems* (1739). The company's name changed to Spottiswoode when descendants of the original owners took over the business in 1819, who had married into this "Spottiswoode" surname. The two especially popular printing enterprises detailed here were so intertwined that they were once called Spottiswoode, Ballantyne & Co. Such collaboration is not likely to have been accidental, but rather coincided with their shared owners' or patrons' hidden manipulation of the greater publishing industry.

Just as in the 18th century, in the 19th century, there is only a single pair of texts from the same printer that falls into the same linguistic-group, in the I2-group's "Henry Alford's" *Biblical Revision* (1870) and "Robert Williams Buchanan's" *Fleshly School of Poetry* (1872) that were printed by "Virtue and Co. City Road". The presence of only a single book pair per canonical century-long corpus that matches in both their ghostwriter and printer hints that the ghostwriting workshops might have deliberately avoided such specialization to obfuscate administrative efforts to identify them. Meanwhile, Virtue and Co. was founded in the 1820s in London by George Virtue (1794-1868), then it operated under James Sprent Virtue (1829-1892) from 1855, and later offshoots of it were started, including Virtue Brothers & Company, and Arthur Hall, Virtue & Co. The printing exclusivity between Virtue and Co. and the I2-group is somewhat interrupted by the listing of a slightly differently spelled, Virtue & Co., as the bookseller of the A2-group's "Anthony Trollope's" *Phineas Finn* (1869). All three of the tested Virtue titles were printed during James Sprent Virtue's control of this company; he was a licensed stationer, who had opened new American branches even prior to taking the helm. James sold their *St Paul's Magazine* in 1869, so it appears this business was short on funds between 1869-72. Other significant events in these years including Samuel Jones Spalding (1820-1892; preacher) joining the partnership in their "City Road" office in 1871, and they published the "Charles Knight"-edited *The Pictorial Edition of the Works of Shakespeare* (1870). In summary, it appears that James Virtue contracted to publish an A2-group novel when the firm was desperate for funds, and as a preacher called Samuel Spalding joined the firm, he seems to have invited the I2-ghostwriter to publish differently-bylined theological non-fiction texts with them.

There are many printers in both 18th and 19th century corpuses that have the same surnames

(such as Ballantyne or Virtue), but different first names, as family members took over family-run printing businesses. For example, there were several Scotts operating as booksellers and printers over the centuries. Back in the 17th century, "William Saywell's" *Evangelical and Catholic Unity* (1682) is described as "Printed for Robert Scott", referring to a bookseller with a shop in Little Britain. Then, in the 19th century, "Henry James Byron's" *Cinderella* (1860?) from the A2-group is described as printed by "Thomas Scott, Warwick Court, Holborn" in London. Families who owned multiple different types of printing, bookselling and publishing businesses could have had significant power over a market, without being regulated as a market-dominating single unit.

When the entire 18th century corpus' cities and signatures are compared, there are no obvious positive correlations between any ghostwriter's preference for only single city or the reverse. When one considers only same-byline pairs, many of them are split between both linguistic groups and cities, especially London versus Dublin. There seems to have been a negative correlation in a conscious or subconscious effort to avoid clustering any given ghostwriter's output in any specific city. It might have been easier to implicate a specific ghostwriter in byline-identity or publishing fraud if all their texts were published in a single jurisdiction. This pattern is most pronounced between the three tested "Henry Fielding"-bylined texts, as the two dramas are in the B-group and were sold in London, while the third novel is in the F-group and was sold in Dublin. Another example is that "George Berkeley's" C-group text was sold in London, while his G-group text was sold in Dublin. Additionally, "Elizabeth Inchbald's" novel is in the D-group and was sold in Dublin, while her drama is in the B-group and was sold in London. This pattern is not entirely consistent, as there is also an example of a same-byline, same-group, and different-city combination is the two "Miles Peter Andrew"-bylined B-group dramas, one of which was sold in Dublin in 1780 and the other in London in 1795. This data does point to the pattern that the B-ghostwriter strongly preferred selling books in London, with at least one same-byline exception. This is a logical conclusion as the B-ghostwriter was responsible for nearly all of 18th century's dramatic output, with many of his plays published or sponsored by the Theatre Royal in London; thus, he would have had to be stationed in London and would have had connections to publishers there.

These geographic patterns are missing in the 19th century corpus because printers had formed regular relationships with international booksellers; for example, "Charles James Lever's" two novels both fall into the same G2-group, though one was only sold in London and the other in London and Dublin; alternatively, "James Clarence Mangan's" two texts fall into two different groups, and both were sold both in London and Dublin.

The location and ownership of a bookshop was essential for buyers to find these establishments. Thus, there were significantly more booksellers than printers listed in books' title-pages. Booksellers were listed in 274 tested texts in the 18th century, and 309 in the 19th century. In both centuries, most of the booksellers sold multiple titles out of these canonical corpuses. 160 different bookseller bylines (counting different first names separately) appeared in the 18th century, and 145 in the 19th century. Between these two centuries, there was an increase in the percentage of texts listing

booksellers, but a decrease in the number of different bookseller bylines. There was consolidation among booksellers, as some shops grew and spread to new locations, while others went out of business.

The booksellers with the most different titles on sale dominated the marketplace, and thus are likely to have had an impact on the ghostwriting workshops. The booksellers with several titles from the 19th century corpus include Richard Bentley and Son, Blackwood and Sons, Chapman and Hall, Harper & Brothers, Longman (with Longman's varied co-booksellers), Macmillan and Co., John Murray, and Elder Smith and Co. In the 18th century, the booksellers with the most titles are T. Cadell, J./R. Dodsley, J. Johnson, and A. Millar.

As "James Lackington" explained, authors had a choice between selling their book themselves at their private residence, or employing a bookseller. There was a significant drop in the percentage of titles that listed the author as the bookseller between the 18th and 19th centuries. In the 18th century, 32 titles described themselves as printed "for the author" on title-pages, while in the 19th century this phrase only appeared in 3 titles. These are cases where the author acts as the bookseller, or sells the title at their own private residence, or at a specified business where the author might or might not plan to be present. One telling description of who is selling the book is in the G-group's "Thomas Woolston's" *Third Discourse on the Miracles of Our Savior* (1728): "Printed for the Author, and Sold by him next Door below the Star in Aldermanbury, and by the Booksellers of London and Westminster". This description signifies that the author has sponsored the printing and copyrighting of this book, and is both selling it himself at a named address, and has also commissioned booksellers in two cities to sell it for him. Commonly, books printed "for the author" name a bookseller who is selling these copies, such as S. Colbert or A. Millar. An example of a rare title that invites buyers to go to the author's house include the C-group's "William Hogarth's" *Analysis of Beauty* (1753): "printed by J. Reeves for the author, and sold by him at his house in Leicester-Fields". 3 of these "printed for the author" titles are especially cryptic, as they do not name an address either for a bookseller or the author. Such titles could only have been sold in private directly to consumers, as buyers could not otherwise locate them. These titles without purchasing locations include: C-group's "Henry Bourne's" *Antiquitates Vulgares* (1725), D-group's "Henry Brooke's" *Fool of Quality* (1765-70), and E-group's "Stephen Duck's" *Poems on Several Occasions* (1736). The motive for these titles' lack of confessed selling locations is obvious from their titles, as they refer to vulgarities and fools. The contents of these titles risked offending obscenity or libel laws, and thus they threatened with incriminating anybody who openly sold them.

In addition to not listing a bookseller, "Duck's" text also did not list a printer; but these omissions are not due to any censorable content, but rather exemplify a private printing; this collection was sold to a predetermined list of subscribers, including Jonathan Swift, Alexander Pope, and the Prince of Wales. This private printing is worthy of a digression to consider the type of "author" who could negotiate such self-advertising. With only a primary education, and without a firm birth-year, an "agricultural laborer" called "Stephen Duck" (1705?-1756) is claimed, by the *Dictionary of National Biography, 1885-1900*, to have educated himself in the poetic genre by

purchasing books, such as *Paradise Lost*. In the 18th century, a two-volume edition of *The Works of Andrew Marvel* (1772) cost 6s., "Zachary Grey's two-volume edition of Bulter's *Hudibras* (1744) 15s, Thomas Newton's *Paradise Regained* of 1750 10s. in sheets. Theobald's 1733 Shakespeare cost his subscribers 2 guineas [£2 2s.] apiece".[10] Duck's weekly salary as a laborer was 4s. 6d. (£11 14s annually), so purchasing a two-volume set would have cost more than a week's labor. There were no libraries where he might have accessed these books for free. Supposedly, some clergy encouraged him to write and publish *Poems on Several Subjects* (1730), which saw 10 editions in its first year, and was rebutted with a parody called *The Thresher's Miscellany* (1730) under the pseudonym "Arthur Duck". Seemingly bolstered by this sudden fame, "Duck" read this work on September 11, 1730 before Queen Caroline. After a delay of three years, in April 1733, Caroline "allowed him" an annual salary of £30 as the "yeoman of the guard". Then, without any clerical training, he was ordained as a priest in 1746, and then became the preacher at Kew Chapel in 1751, and at the rectory of Byfleet, Surrey in 1752, before ending it all by committing suicide.[11] There is a logical progression between Duck's labor in the fields and Queen Anne's award of a "yeoman" title. On the other hand, "Duck's" authorship of poetry is just as unlikely as his sudden appointment to a priesthood. The publication of the "Arthur Duck"-bylined *Thresher's* parody simultaneously with the "serious" "Stephen Duck"-bylined *Poems* explains that both were designed to ridicule the absurdity of the common practice in this period of assigning bylines to absurdly unlikely pseudonyms or strawmen, who might have been either extremely impoverished and uneducated, as with "Sarah Bulter", or overly-titled, as with "Isaac Bickerstaff Esquire". This farce about the misuse of pseudonyms must have been especially influential for the sponsor of *Poems'* 1736 edition to refrain from listing both a printer and a bookseller, and only distributing it privately to the Workshop's insiders, like "Swift" and "Pope", who were implicating in satirizing and puffing "Duck". "Duck"-bylined verse was clearly a ghostwriting-career-shaping text for the 18th century Workshop, and this significance was only revealed by noticing its unusual publishing credit. The commonality of such discoveries during this process is why all attribution studies must scrutinize bibliographic patterns in tandem with selecting a corpus, and then before reaching final attribution conclusions.

Bookselling Mysteries of the 18th Century G-Group

In both the 18th and 19th centuries, there was only a single printer whose pair of published books fell into the same linguistic-group. In contrast, just in the 18th century corpus, there are 9 booksellers with a pair or more titles that all match a single linguistic-group. Still more significantly, 5 of these booksellers' books all matched the G-group. In this section, researching these booksellers and publications explains the motives or causes for this pattern, and helps to establish Samuel Richardson as the most likely ghostwriter of the G-group. Richardson was extremely productive as a printer, bookseller, and author under his own byline, and this review indicates that he covered the potentially-censorable or otherwise overflowing output that he could not release under his own byline by engaging in pseudonymous publishing and authorship. This section does not dive far enough into the archives

to determine with certainty that these booksellers' bylines were pseudonyms, but it reveals enough to conclude this is likely because of the relative absence of biographical materials about them.

All 3 of Richardson's tested novels had some matches to the G-group, but 1 of them, *Sir Charles Grandison* (1753), was a closer match to the D-group. Richardson listed himself as the printer for all 3 novels, as he self-printed his publications. However, he only listed himself additionally as their bookseller on only the two firm G-group matches, *Pamela* (1742) and *Clarissa* (1748). No booksellers are listed on the title-page of *Grandison*, which is also divergent because it was released in Dublin (instead of London).

During the initial chronological comparison between the publication and authors' birth and death dates within each linguistic-group, Richardson's lifespan (1689-1761) excluded him as a likely ghostwriter because the first release in the G-group is technically the "1700" publication of "Mary Astell's" *Some Reflections*. Noticing the G-group pattern among booksellers led me to reconsider this automatic rejection. If merely this one "1700" release was backdated (or falsely assigned to an earlier date to forge a book's antique status), the group's remaining earliest publication would have been the 1704 anonymous and erroneously "Defoe"-assigned *Regulation of the Press*, which advanced a case for the use of disguised authorial and publishing bylines. After an exclusion of the "1700" publication and of bylines without firmly documented birth years (such as "George Psalmanazar" (1679?-1763)), only Samuel Richardson and Walter Lynn (1677-1763) remain as the likeliest ghostwriters of the G-group. In 1704, Richardson was 15. When Richardson's nephew, Thomas Verren Richardson, first apprenticed into his uncle's business in 1732, he was 15; this is a documented fact because Richardson described it in *The Apprentice's Vade Mecum: Or, Young Man's Pocket Companion* (1734).[12] Thus, there is a strange 2-year gap between when Richardson came into the proper age to have started an apprenticeship in 1704, and when he first registered as starting an apprenticeship in 1706. Biographies of Richarson do not mention his attendance of a college, but perhaps he spent these 2 years studying in college under another name, or spent it in an unregistered ghostwriting apprenticeship, when he might have ghostwritten personal and governmental correspondences alongside with starting to publish texts such as *Regulation*. Returning to the second most likely ghostwriter of the G-group, Walter Lynn, presents a possible attribution solution. Lynn had received a BA from Cambridge in 1698, and an MB in 1704. He was elected as a "member of the Gentleman's Society at Spalding" in 1712, and then published a few medical and satirical books between 1714-26. He stopped publishing under his own byline after printing a claim to the invention of the steam-engine, in *The Case of Walter Lynn, M.B., in Relation to Diverse Undertakings of His, Particularly for the Improvement of an Engine to Raise Water by Fire, etc.*, which he says he submitted to "Sir Isaac Newton" and other scientists, but the credit for creating a functioning engine is typically granted to Thomas Newcomen in 1712. The missing detail regarding Richardson's higher schooling is neatly filled with Lynn's advanced degrees. Lynn's completion of his BA in 1698 also means that the 1700 text could have been written by him, without a need for backdating. And his completion of the MB by 1704 would explain the burst of

philosophical publications in the G-group in this year. One possibility is that Lynn decided to apprentice to a printer under the "Richardson" pseudonym in 1706, when he decided to become a professional publisher and "author" (as the Gentleman's Society described him).[13] Lynn might have decided not to use his own byline, even when he was publishing, because of the temporary imprisonment and other legal obstacles his pseudonyms, such as "Daniel Defoe", encountered early in his ghostwriting career. Even if Lynn was the authentic byline of the G-ghostwriter, far more is known about the biography of Richardson than about Lynn; assuming an intersection between these bylines in a single real individual, it is this documented information regarding Richardson's self-confessed publishing and writing activities that best helps to explain the history of British ghostwriting.

Richardson is a rare example of a printer-author from Britain's 18th century. As a documented practitioner of both trades, it is more likely that Richardson could monopolize a segment of the publishing industry as opposed to rivals who are known either as merely authors or printers or booksellers. Aside for self-printing his own novels, Richardson is credited with printing around 500 other texts. Richardson's father was implicated in the Monmouth Rebellion in 1685. Richardson attended the Christ's Hospital grammar school in London. Then, Richardson contracted under master John Wilde, a printer specializing in almanacs, into a 7-year apprenticeship between 1706-13; and it was extended for some reason through 1719. There is no accessible biography of a printer called "John Wilde". There was a John Wilde Esq, Serjeant at Law, serving in Parliament in 1640. And there was a John Wilde of Pembroke, who completed and MA on July 9, 1670 from the University of Oxford. There is a biography reported in the *Dictionary of the Printers and Booksellers* for Allington Wilde, John Wilde's son, who printed at Aldersgate Street in London between 1722-31. In 1721, Richardson's first wife became Allington Wilde's daughter, Martha.[14] Richardson's tasks as an apprentice included compositing and proofreading for John's press. At the end of the initial 7-year apprenticeship term, in 1713, Richardson claims that he became an "Overseer and Corrector of a Printing-Office", but he does not disclose this office's location.[15] This omission and the lack of documentation supporting this employment introduces the likelihood that Richardson began working as a ghost-printer, or using multiple printing pseudonyms. Richardson's first known printing location and home was at Blue Ball Court and Dorset Street. Richardson took on several apprentices starting in 1722, but none of these apprentices' names appear as master printers or booksellers in the 18th century corpus. This might suggest that they all took on pseudonyms under Richardson's instructions, as they set up presses after ending apprenticeships, or Richardson might have failed to find apprentices who were skillful enough to become independent printers. Richardson's first experiment with a major printing contract was in 1723, from Philip Wharton, 1st Dule of Wharton, for printing *The True Briton*. This Jacobite periodical was immediately censored for "common libels". Obviously, Richardson had to have had a long-standing ghostwriting and printing record prior to this newspaper to have been awarded this politically and socially valuable contract in 1723. Richardson was not personally prosecuted for the libel because he was only credited as the printer, and not as the author. Scholars such as John Carroll have speculated that Richardson probably ghostwrote some of

the *Briton*.[16] This near-imprisonment would have been yet another firm motive for Richardson to have primarily used pseudonyms for his questionable writing, printing, and bookselling.

While the lack of a birth year and of a documented origin-story largely chronologically disqualifies George Psalmanazar as the G-ghostwriter, the myriad of identity frauds he was discovered committing fits most logically with the false-byline frauds the underlying ghostwriter was committing. While Psalmanazer falsely claimed to be from *Formosa* (i.e., Taiwan, the subject of his tested 1704 "autobiography"), he might have been using pseudonyms because his authentic Jewish identity was still outlawed across Europe (just like the Renaissance Workshop's Josuah Sylvester, who was exiled in 1613 over his Judaism).[17] The only reason modern scholars generally refer to Psalmanazer as a fraudster is because he confessed to the falsity of his preceding "autobiography" in the posthumous *Memoirs of ****. Commonly known by the name of George Psalmanazar; a Reputed Native of Formosa* (1764). Such a confession is very unusual for this period, as more typically false "autobiographies" were later reclaimed as fictions (as was the case with *Robinson Crusoe*). Given the seemingly mixed biographical details that were necessary to make a professional ghostwriter between Richardson and Lynn, the confessions in "Psalmanazer's" posthumous *Memoirs* might be read as genuine posthumous confessions of byline-fraud from the G-ghostwriter (Richardson and Lynn died in 1761 and 1763, or at around the same time as Psalmanazer).

If a bookseller was the G-ghostwriter, instead of one of the signed "authors"; then, among the booksellers who were exclusive to the G-group, an especially likely candidate is Daniel Browne II. The largest same-group cluster for an 18th century bookseller are the 4 G-group texts published by variants of the Brown(e) bookseller surname. "Dan. Brown, G. Strahan, W. Davis, Fran. Coggan, Bernard Lintott" sold "George Psalmanazar's" travel narrative *Formosa* (1704). "D. Browne, J. Taylor, R. Smith, F. Coggan, T. Browne" are listed as the booksellers for "John Beaumont's" theological *Treatise of Spirits* (1705). "Browne jun., S. Chapman" sold "Eliza Haywood's" novel *Fantomina* (1725). And Browne jun. and Chapman had previously worked in collaboration with "W. D. Chetwood" on selling "Haywood's" *Idalia* (1723). All, except the "Beaumont" title, were described as sold at "the Black Swan without Temple-Bar"; thus, these variant names are referring to the same booksellers, Daniel Browne I (flourished: 1662-1729; dying in 1729) and his son, Daniel Browne II (started working for the shop in 1704, took over the shop in 1729, and died in 1762). Browne I died too early to have ghostwritten the G-group. On the other hand, Browne II's lifetime is a great fit for the G-group. Given that Browne II died in 1762, it is reasonable to assume he started working in the publishing business a few years before 1704, or early enough to have ghostwritten the 1700 release in the G-group. The appearance of "Dan. Brown's" name as the bookseller of "Psalmanazar's" *Formosa* in 1704 could have been a semi-confession of Browne II's authorship of this fictitious "autobiography". A bookseller could sell his own books without appealing to an intermediary for access to the public. Scrutiny of Browne II's documented biography might help to uncover if there is any substance to these suspicions of his potential ghostwriting. These suspicions cannot be linguistically verified, since Browne did not publish any texts with his name in the authorial byline.

There is also a smaller same G-group cluster in the two titles J. or John Baker (active between 1680-1717, and residing at the work-office of Black Boy in 1710) sold "at the Black Boy in Pater-Noster-Row"; both are currently assigned to "Daniel Defoe". Though only one of them initially had "Defoe's" byline on it, *Appeal to Honor and Justice* (1715), while the previously released *What If the Queen Should Die?* (1713) was anonymous (due to its seditious content). Baker sold several untested political pamphlets during his active years, such as *The Rights of Succession to the Empire of Germany* and *A Plot Discovered; Or the Protestant Succession in Danger*. Given the similar genres between these other early 18th century titles and the 2 tested titles, it is likely that if more of these were tested, they would also fall into the G-group. The first tested publication in the G-group is from 1700, so it is likely this ghostwriter began to be an active ghostwriter of untested texts prior to 1700. The most likely ghostwriters for the G-group were born between 1677-89. However, Baker probably became active as a bookseller too early to have ghostwritten this entire group; exceptions would include absurd scenarios such as that his first 1680-95 publications were backdated, and he remained active after 1717 under a different pseudonym. Researching "John Baker's" biography is challenging because the earliest book assigned to this byline (or the "J.B." initials) was published in 1581, while the last was published in 1783. Obviously, these bylines cannot be referring to the same person. This same byline also appears between 1657 and 1716 as a publisher, as a printer between 1685-1800, and as a bookseller between 1708-16. J. Baker's earliest active year, of 1680, is set by his advertisement of *The Political State of Great Britain* in the *Post Man*; he first advertised in the *Term Catalogues* in 1709, on entering *A Vindication of the Church and Clergy of England from Some Late Reproaches*. These re-appearances signify that the British workshops reused this pseudonym across the centuries. On the other hand, it could have simply been a popular name, if families of bakers adopted it. There was also a J. Baker, the senior warden of the Company of Stationers between 1686-7, and John Baker, another London bookseller, but at Mercers' Chapel in Cheapside between 1703-25 and other places. Amidst this biographical chaos, it is impossible to confirm any facts to prove or disprove Baker's status as a potential ghostwriter.

Similarly to "Baker", the "Parker" surname appeared as a bookseller with two different first initials on rhetorical publications in the G-group: "R. Parker" "at the Unicorn under the Royal Exchange" for "Daniel Turner's" *Remarkable Case in Surgery* (1709), and decades later "W. Parker" "at the Kings-Head in St. Paul's-Church-Yard" for "John Latham's" *Attending Those Who Enter into Holy Orders* (1736). Richard Parker was active as a bookseller and publisher of dramas and rhetoric in London from 1692 (when he first registered a book in the *Term Catalogue*) to 1742. "Richard Parker" has been claimed to be the author of at least one theological text, *An Essay on the Usefulness of Oriental Learning* (1739), which was initially anonymous, with a dedication signed with the pseudonym, "Philoglottus". The identity of "W. Parker" is more mysterious, as *A Dictionary of the Printers and Booksellers Who Were at Work in England, Scotland and Ireland from 1668 to 1725* lists several first names for this surname, including Andrew, Edmund, Edward, G., and Peter, but none start with a W.[18] "W. Parker", at the same address and in the same year, was also credited with selling *A*

Collection of Anthems, As the same are now performed in his Majesty's Chapels Royal, &c. Published by the direction of the Reverend the sub-Dean of his Majesty's said Chapels Royal (1736). Given the number of Parkers operating bookshops, and these few appearances of the W. initial without a matching biography, it is likely that "W. Parker" was a pseudonym, which might have been designed to be confused with William Parks, a printer who moved between Ludlow, Hereford and Reading, and who specialized in theological books between 1719-23. Peter Parker senior had a relatively successful bookselling business in the second half of the 17th century. Peter Parker junior attempted to sell books without a license from the Company of Stationers briefly in 1707, before Peter was forced to close this business. The rest of the Parkers were only listed on a few isolated titles. Peter junior's prosecution by the Stationers explains that it is likely that a differently-named bookseller (such as Richardson) might have been using these various Parker pseudonyms to escape paying official fees to the Stationers for selling books. It is also possible that Richard Parker was the underlying ghostwriter of the G-group, who convinced members of his extended family to use their names as booksellers to keep suspicion away from himself. The most likely scenario is that one of these players (Richardson, Browne II, Parker or Baker) was an authentic individual, who was using the other bookselling pseudonyms to evade fees, prosecutions and other unfavorable outcomes. Though it is also possible that all of these were simply laborers who manually pressed books, while the design, editing and ghostwriting was the element that was monopolized between them by the G-ghostwriter (perhaps Lynn). The absence of biographical information about the Parkers or texts with their byline(s) on their title-pages prevents a firm discrediting. On the other hand, this absence alone is sufficient proof to make it unlikely that any bookseller without any writerly credits could have been the G-ghostwriter.

This has been a brief review of the linguistic patterns among 18th and 19th century printers and booksellers, and how these patterns are useful in confirming or rejecting attribution hypotheses. A purely quantitative attribution that is based on linguistics and chronology will always miss cluster-defining biographical elements, such as the presence of confessed identity-forgers such as "Psalmanazar". And a review of the initial bylines and how eventual bylines came to be assigned is an essential step in the attribution process, to determine if a computational-linguist can place any trust in current bylines. Hopefully some of the revelations discussed in this fragment will help future attribution scholars with noticing similar publishing trickery in any corpus.

Endnotes:

1. Hugh Craig, "Authorial Attribution and Computational Stylistics: If you tell authors apart, have you learned anything about them?" *Literary and Linguistic Computing* (14:1, 1999), 103-113.
2. Ying Zhao, and Justin Zobel, "Searching with Style: Authorship Attribution in Classic Literature", *Proceedings of the 30th Australasian Computer Science Conference* (New York: ACM Press, 2007), 59-68.
3. The data from these three corpuses is available at https://github.com/faktorovich/Attribution.
4. Stephanie Fysh, *The Work(s) of Samuel Richardson* (Newark: University of Delaware Press, 1997), 32-33.
5. Daniel Defoe, *An Essay on the Regulation of the Press* (London, 1704), 15-17.
6. James Lackington, *Memoirs of the Forty-Five First Years of the Life of James Lackington* (London: Printed for the Author, and sold by all other Booksellers, 1795), 385-389.

7. Judge Valerie Figueredo, *Amazon.com, Inc. EBook Antitrust Litigation: Report and Recommendation*, U.S. District Court for the Southern District of New York (March 25, 2021, July 31, 2023). *LaFaro v. New York Cardiothoracic Group, PLLC*, 570 F.3d 471, 475 (2d Cir. 2009).
8. "Standard Oil Company and Trust", *Encyclopædia Britannica*.
9. J. G. Lockhart, *Memoirs of the Life of Sir Walter Scott*, Volume I (Philadelphia: Carey. Lea, & Blanchard, 1837), 237-238, 243.
10. Marcus Walsh, "Literary Scholarship and the Life of Editing", *Books and Their Readers in Eighteenth-Century England: New Essays*, Isabel Rivers, Ed. (London: Bloomsbury, 2003), 209-210.
11. "Duck, Stephen", *Dictionary of National Biography, 1885-1900*, Volume 16, Leslie Stephen, Ed. (New York: Macmillan and Co., 1888).
12. Fysh, *Work(s)*, 27.
13. Sidney Lee, Ed., "Lynn, Walter", *Dictionary of National Biography*, Volume 34 (New York: Macmillan and Co., 1893), 343-344.
14. Ibid., 313.
15. William M. Sale, *Samuel Richardson: Master Printer* (Ithaca: Cornell University Press, 1950).
16. John Carroll, "Review of Samuel Richardson: A Biography", *The Review of English Studies* (23:92, 1972).
17. Michael Keevak, *Pretended Asian: George Psalmanazar's Eighteenth Century Formosan Hoax* (Detroit: Wayne State University Press, 2004), 1-2.
18. Henry R. Plomer, *A Dictionary of the Printers and Booksellers Who Were at Work in England, Scotland and Ireland from 1668 to 1725* (Oxford: Bibliographical Society: Oxford University Press, 1922).

References

Britannica, Encyclopædia. "Standard Oil Company and Trust." *Encyclopædia Britannica* (2015).

Carroll, John. "Review of Samuel Richardson: A Biography." *The Review of English Studies* 23: 92 (1972): 504-508.

Craig, Hugh. "Authorial attribution and computational stylistics: If you can tell authors apart, have you learned anything about them?." *Literary and Linguistic Computing* 14.1 (1999): 103-113.

Defoe, Daniel. *An Essay on the Regulation of the Press*. London, 1704.

Fysh, Stephanie. *The Work(s) of Samuel Richardson*. Newark: University of Delaware Press, 1997.

Keevak, Michael. *Pretended Asian: George Psalmanazar's Eighteenth Century Formosan Hoax*. Detroit: Wayne State University Press, 2004.

Lackington, James. *Memoirs of the Forty-Five First Years of the Life of James Lackington*. London: Printed for the Author, and sold by all other Booksellers, 1795.

Lee, Sidney, (Ed.). *Dictionary of National Biography*, Volume 34. New York: Macmillan and Co., 1893.

Lockhart, J. G. *Memoirs of the Life of Sir Walter Scott*, Volume I, Philadelphia: Carey. Lea, & Blanchard, 1837.

Plomer, Henry R. *A Dictionary of the Printers and Booksellers Who Were at Work in England, Scotland and Ireland from 1668 to 1725*. Oxford: Bibliographical Society/Oxford University Press, 1922.

Sale, William M. *Samuel Richardson: Master Printer*. Ithaca: Cornell University Press, 1950.

Stephen, Leslie, (Ed.). *Dictionary of National Biography, 1885-1900*, Volume 16. New York: Macmillan and Co., 1888.

Walsh, Marcus. "Literary Scholarship and the Life of Editing." *Books and Their Readers in Eighteenth-Century England: New Essays*, Ed. Isabel Rivers. London: Bloomsbury, 2003. 191-215.

Zhao, Ying, and Justin Zobel. "Searching with Style: Authorship Attribution in Classic Literature." *Proceedings of the 30th Australasian Computer Science Conference*. New York: ACM Press, 2007. 59-68.