CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The sociology of style: Writing and influence within literary families
Authors	O'Sullivan, James
Publication date	2021-10-29
Original Citation	O'Sullivan, J. (2021) 'The sociology of style: Writing and influence within literary families', Poetics. doi: 10.1016/j.poetic.2021.101620
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1016/j.poetic.2021.101620
Rights	© 2021, Elsevier B.V. All rights reserved. This manuscript version is made available under the CC BY-NC-ND 4.0 license. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-04 12:18:09
Item downloaded from	https://hdl.handle.net/10468/12344



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

The Sociology of Style:
Writing and Influence within Literary Families

https://doi.org/10.1016/j.poetic.2021.101620

"What is in a name?" Such is the question posed by Patrick McAleer in his exploration of one of popular literature's most prolific households: Stephen, Tabitha, Joseph[1] and Owen King (McAleer 2011, 9). The Kings are what one might call a "literary family", a phenomenon which emerged a good many centuries ago when cohorts of relatives and close friends began forming creative cooperatives that would prove hugely influential to the artistic production of their individual constituents (Krawczyk 2009; Kenny 2020). Literary families tell us much about the social production of art, about hierarchy and heredity (Kenny 2020) and the role that social status plays in the cultural sphere. To answer McAleer's question, there probably is quite a bit in a name.

There is a world of research and ideas from a range of disciplines engaged with the biopsychosocial model that might have relevance to an examination of literary families, but the exact focus of this brief essay is the specific role that *style* can play in assessing the sociology of art. A substantial contribution to knowledge on the connections and tensions between nature, nurture and aesthetics is not presented here,[2] but it is possible to examine if established social hierarchies have real, measurable—in the quantitative sense—impacts on how a person writes. It has been several decades since Janet Wolff argued that art is a consequence of "real, historical factors" (1984, 1). She and other advocates for a sociology of art[3] have often been criticised for lending too much credence to social determinism, a debate which, while possibly tired, is not entirely resolved. Social production has everything to do with environment, and so if one's socio-cultural space does impact their writing, or specifically, determine to some degree their literary style, members of the same space should cluster together in a stylometric analysis. Such is the question that this brief experiment in the style of literary families looks to answer: do members of the same literary family share stylistic fingerprints such that, when examined quantitatively, they consistently cluster as a unit?

The idea for this study was prompted during a prior project, in which a stylometric cluster analysis of the Brontë siblings showed that "the Brontës had a significant influence on each other's style, such that, at a macro-level, they gather together as a unique cluster" (McCarthy and O'Sullivan 2020). In a slight deviation from the central focus of the Brontë study, writings by each of the siblings—Charlotte, Branwell, Emily and Anne—were tested alongside a comparative sample of other British authors. The Brontë family formed their own stylistic cluster (see Fig. 1), which might be seen as quantitative evidence for something which many Brontë scholars have long contended, that the siblings formed a creative collaboration which greatly influenced their respective oeuvres (Van Der Meer 2008; Malfait and Demoor 2015; Braxton 2019; Butcher 2019).

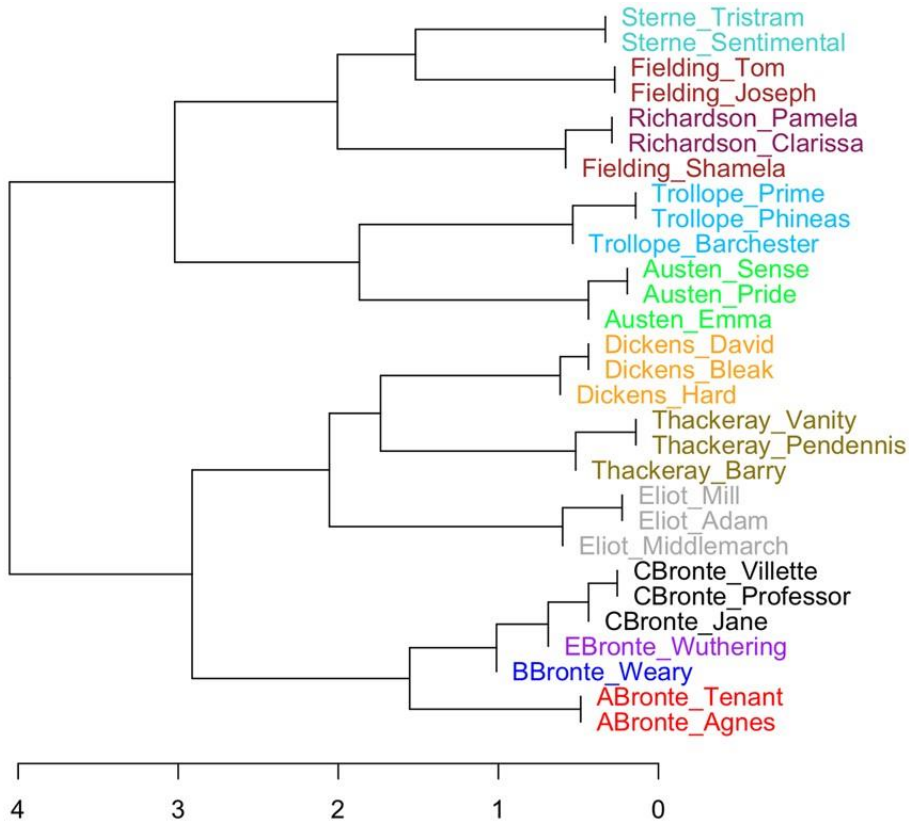


Figure 1. A stylometric cluster analysis showing that the Brontë siblings, comparative to other British authors, write with similar stylistic fingerprints (McCarthy and O'Sullivan 2020)

And yet, it is possible that the Brontës might only be exceptional when compared with authors who are not part of a family unit, and indeed, when analysed alongside other literary families, could simply represent what is the norm for authors raised within a consistent family unit. As indicated in a note from the aforementioned paper: "A further interesting study would be to test the degree to which this trend is replicated across other literary families, or if there is in fact something very particular about the ways in which the Brontës developed something of a shared style" (McCarthy and O'Sullivan 2020).

This rudimental experiment responds to that note, using stylometry as a means of indicating, at least foundationally, if shared social contexts are so influential that they cause authors to develop authorial signatures so similar that they cluster together when measured quantitatively, results which speak to decades of sociological debate on art and literature.

Data, Methods & Findings

For readers who might be unfamiliar with the technique, stylometry is a computer-assisted method which determines how an author tends to write by counting the frequency of a particular feature—usually words—in a sample of their texts. Having "profiled" authorial style in such a fashion, statistical clustering can be used to show stylistic proximity and distance between authors and texts.[4] This analysis was conducted with Stylo (Eder, Kestemont, and Rybicki 2013), using Classic Delta (Burrows 2002)[5] and the 100 most frequent words.[6] The study is confined to English language[7] novel-length fiction,[8] with only blood relatives included.[9] Collating a corpus was challenging, as suitable texts could not be obtained for several other high-profile literary families,[10] but it was nonetheless possible to gather suitable texts from across six distinct households,

enough to conduct a germinal evaluation of the hypothesis. The final corpus is comprised of Kingsley and Martin Amis (father-son); Anne, Charlotte, and Emily Brontë (sisters); William Godwin, Mary Wollstonecraft, and Mary Shelley (father-mother-daughter); A. S. Byatt and Margaret Drabble (sisters); W. Somerset and Robin Maugham (uncle-nephew); John le Carré and Nick Harkaway (father-son).[11]

It would appear from this initial analysis that there is in fact nothing at all exceptional about the similarities in style shared by the Brontë siblings,[12] and that most literary families cluster together when tested with stylometry (see Fig. 1).[13] The familial units consistently cluster in close proximity,[14] which suggests—acknowledging the limitations in this dataset—that members of the same family typically tend to write in similar styles.[15]

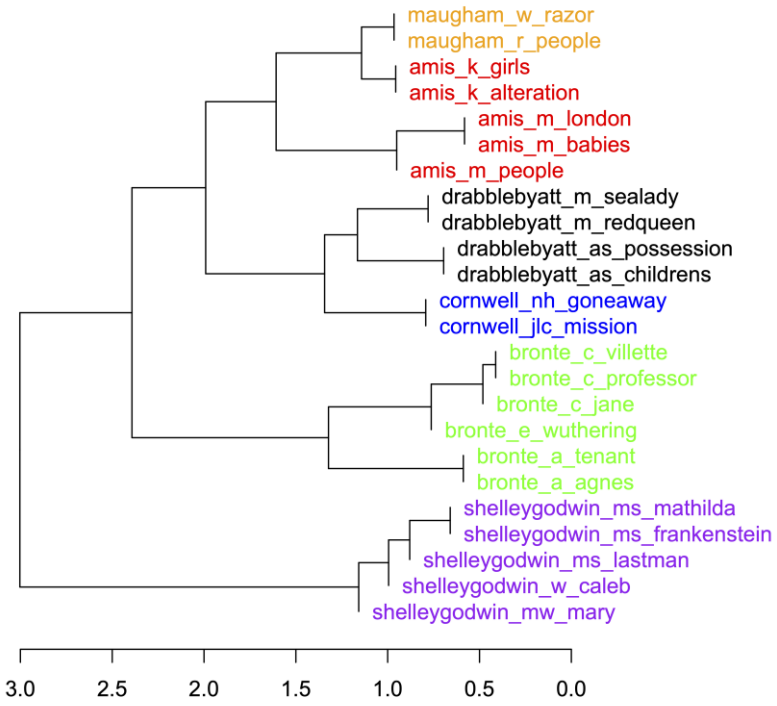


Figure 2. A stylometric cluster analysis of several literary families, all of which cluster in relatively close proximity, showing authorial signatures tend to be similar across family units

These results might be seen as further, quantitative evidence of the significance of the creative collaborations seen in families such as the Brontës: it is easy to appreciate how the stylistic tendencies of one family member might influence another, or indeed, how direct editorial interventions between relatives might account for the familial clustering seen in this study. Viewed from this perspective, these findings are quite obvious: families like the Brontës worked in close proximity, so it is perhaps unsurprising that they would cluster together. Still, the consistent clustering across all families suggests that there is, despite the amount of scholarship on the matter, nothing particularly special, at least not stylistically, about the cooperation between the Brontës, and that a shared literary style is to be expected among relatives.

These results might also be taken as evidence for social determinism, a more elusive phenomenon than creative collaboration among kin. Perhaps it is typical for literary families to form the same type of creative collaboration that was seen with the Brontës, or perhaps social experiences play a quantifiable role in shaping how authors write. Environment obviously contributes to someone's writing, but it remains

fascinating to see how such determines—to such a measurable degree—similarity in style across individuals.

Further evidence that there is something exceptional in how social experience determines literary style can be found in the Godwin-Wollstonecraft-Shelley cluster: Mary Shelley never knew her mother, and yet, the two still share stylistic proximity. One might mistakenly conclude—as I almost did—that this clustering is a consequence of William Godwin's influence on both his wife and daughter, but Mary Wollstonecraft's only novel was published *before* her courtship with William Godwin began. It is possible that William Godwin acted as something of a conduit between his wife and daughter, transferring any stylistic influence he picked up from the former into the development of the latter, but it is remarkable to think that such influence would be quantitatively present when mother and daughter did not directly interact. Or perhaps Wollstonecraft's novel was a source of stylistic inspiration for both her husband and daughter. As Mary Wollstonecraft could not have had direct editorial influence over her daughter, and it seems unlikely—but not impossible—that any influence she had over Godwin could have been so potent it would transfer from him to the next generation, we are left with social experience as the only consistent force shaping the style of Mary Wollstonecraft, William Godwin, and Mary Shelley.[16]

Conclusion

Measuring style using quantitative methods is often an exercise in measuring influence, because while we use the results to inform wider interpretations, the central task of stylometry is to establish whose voice a piece of writing is most like. While an author's own stylistic signal should dominate a text, in comparative studies they will always be shown to be stylistically more similar to one figure than they are another. There is a hierarchy of reasons for this, including, for example, genre and chronology, but influence is also a major reason for authorial signals appearing to be similar.

This analysis had two possible outcomes: either the majority of literary families included in the study would *not* cluster, suggesting that there is something exceptional about those families like the Brontës that have been previously shown to possess stylistic affinities (McCarthy and O'Sullivan 2020); or the majority of literary families included in the study would cluster, suggesting that there is nothing exceptional about those families that do form distinct groupings, because such is the norm. The latter outcome, which has turned out to be the case, tells us something significant about the social production of art in relation to literary style. At a trivial level, these results might be seen as quantitative proof—if such was needed—that the idea of naturally occurring genius is indeed a romantic myth and that great artists are as much a product of their social, economic and cultural environments as anyone. Less trivially, it is significant that members of the same literary families, be it through creative collaborations or shared social contexts or both, develop writing styles that are quantitatively similar. It has long been established that socialisation plays a role in the production of art, but it is interesting to see that influence extend so explicitly to something as nuanced as literary style.

In showing that familial influence is both substantial and consistent, these results confirm the role played by social privilege in art: where you come from matters greatly in how you write. This experiment shows that stylometry is a useful tool in modelling the social production of writing and it should play a role in future studies of literary families.

This study does not address a whole range of contextual considerations, such as the fact that having a famous author for a mother or father will obviously, particularly now in the age of literary capitalism, help an aspiring writer get published. But whether one's parents have clout in the literary market is only a bonus when the ability to write in

a certain way has already been partly dictated by social conditions. Style as constructed by most computer-assisted analyses can be seen as an objective measure of literary quality (see Herrmann, van Dalen-Oskam, and Schöch 2015). Thus, if you are from a literary family you are more likely to have a style that is seen as "quality".

There is much more work, computer-assisted and otherwise, to be done on literary families if we are to fully appreciate *why* their constituents share similar styles, and indeed, a first step towards this would be an expanded dataset which includes a more inclusive range of genres, forms, and other literary contexts. Should such an expanded study confirm these results, then a potential next step would be to identify and expand upon any anomalies, so that we might better appreciate literary families as social constructs by critiquing relatives which do not possess shared authorial fingerprints. Indeed, loosely attributing the results of this present to study to shared social contexts is only the first step in many towards a better understanding of style as a consequence of sociological factors and the measurable influence of creative networks of practice on authorship. But this short experiment nonetheless indicates that style is socially influenced in a very substantial sense, and evidence of such is beneficial to our understanding of style and influence.

The processes of socialisation are nuanced, and it is far beyond the scope of this research note to establish with any real certainty why such stylistic proximities occur in any instance; the causalities hinted at in this paper are only illustrative of what might be more substantively developed in future work on the association between literary style and social construction.

Notes

[1] Joseph Hillstrom King writes under the pen name, Joe Hill.
[2] I understand that a more thorough examination of this topic is currently being prepared by Terence Murphy, Yonsei University.
[3] See, most notably, Arnold Hauser (2011).
[4] There are many studies to which readers interested in developing a better understanding of stylometry as a critical method might turn (Hoover 2007; 2009; O'Sullivan et al. 2018)
[5] I must admit to a modicum of cherry-picking here, in that I repeated the experiment using Cosine distance (Evert et al. 2017), and retrieved a slightly different result. The vast majority of clusters remained consistent across both measures, so I opted for Delta, seeing as it is still the dominant metric used in the field, and the results lent themselves to a clearer explanation. In no way did the results derived from Cosine change my interpretation of findings, so I am comfortable admitting to this critical selection.
[6] I have always aligned with the view that stylometry is best practiced using small samples of high-frequency words, because doing so usually means that the test is being conducted on words which are "especially resistant to intentional authorial manipulation" (Hoover 2009, 35) and thus suited to measuring "style" as it is perceived for the purposes of authorship attribution. I am particularly convinced of the efficacy of this approach in the context of literary families, wherein using a small sample of high frequency words, typically function words, avoids content words being introduced. Considering the shared social contexts of families, it is particularly important that content words are avoided in this instance, as it is probable that many family members might write about the same thing. Remember, we are to measuring *how* our candidates write, not what they write about.
[7] Translated texts were excluded to ensure that translator signals did not interfere.
[8] I have observed in previous experiments that poetry and fiction cannot be reliably compared in a stylometric cluster analysis.

[9] This seems like the easiest way to determine what constitutes a "family unit" without venturing into moral or sociological arguments that have little to do with this study.
[10] With thanks to the following colleagues for suggesting literary families: Mary Morrissy, Laurel Ryan, Chuck Jones, Mary O'Connell, Miranda Corcoran, Frankie Gaffney, Karl Parkinson, Madeleine D'Arcy, Jack Murray, Rachel McGovern, Mark Bernstein, Alex Davis, Valerie Cotter, and Dean Irvine.
[11] Texts were also available for J. R. R. Tolkien and Simon Tolkien (grandfather-grandson), but they were excluded from the study in an attempt to reduce interference from genre signals. When only *The Hobbit* and *Final Witness*/*The Stepmother* were included, J. R. R. and Simon clustered together but caused some small upset to the other clusters, while including their complete oeuvres caused them to separate, and again, interfered with other clusters that otherwise remained consistent. Removing these candidates might be seen as a further act of cherry-picking (see Note 5), but considering the J. R. R. novels are quintessentially fantasy, and Simon's equally typical of the crime / courtroom genre, I feel their exclusion is warranted on the grounds that they introduce matters of style related to genre that only serve to confuse. This matter is more thoroughly addressed in "Who Wrote *Wuthering Heights*?" (McCarthy and O'Sullivan 2020), which provides an example of stylometry being conducted on a "less-than-ideal" dataset. In the *Wuthering Heights* paper, the content of letters and personal diary entries are used as writing samples for Emily Brontë, but this is because she only wrote one novel, and even at that, was largely done for illustrative purposes. The limitations in such an approach are discussed in the 2020 study. Generally, where possible, researchers conducting tests using stylometry should reduce any signals—such as genre—which might not relate directly to the question being posed.
[12] This of course does not invalidate the main findings from this previous study cited (McCarthy and O'Sullivan 2020), which is primarily about the authorship of *Wuthering Heights*, rather than the Brontë's creative collaborations.
[13] The naming convention on the dendrogram should be relatively intuitive: labels begin with family names or some composite of such, with the forename initial of each author being present after the first underscore, followed by a shorthand version of the specific text. The family name Cornwell is used for John le Carré and Nick Harkaway.
[14] For those wondering, clusters are determined by proximity on the "dendrogram" displayed in Figure 1; the further one has to traverse a line between two texts, the more dissimilar their style.
[15] Kingsley Amis is closer to W. Somerset and Robin Maugham than he is to his son Martin but is still very close to the latter. I will leave it to other, more qualified, scholars to comment on the possible significance of this finding.
[16] One could use the Godwin-Wollstonecraft-Shelley case as evidence that the idea of "natural genius" is not as outdated as we like to think, that one's genetics play a part in how they write, and Mary Shelley, being the daughter of Mary Wollstonecraft, was naturally predisposed to sharing her mother's aesthetic sentiments. This seems, to my mind at least, a far less convincing explanation than social determinism provides, but it is worth noting for those who still see some merit to the nature versus nurture debate.

References

Braxton, Kimberley Jayne. 2019. 'The Brontës' Writing Community: Family, Partnership and Creative Collaboration'. Doctoral, Keele University. United Kingdom. http://eprints.keele.ac.uk/6083/.

Burrows, John. 2002. '"Delta": A Measure of Stylistic Difference and a Guide to Likely Authorship'. *Literary and Linguistic Computing* 17 (3): 267–87.
Butcher, Emma. 2019. *The Brontës and War: Fantasy and Conflict in Charlotte and Branwell Brontë's Youthful Writings*. Palgrave Macmillan.
Eder, Maciej, Mike Kestemont, and Jan Rybicki. 2013. 'Stylometry with R: A Suite of Tools'. In *Digital Humanities 2013: Conference Abstracts*, 487–89. University of Nebraska–Lincoln.
Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. 2017. 'Understanding and Explaining Delta Measures for Authorship Attribution'. *Digital Scholarship in the Humanities* 32 (suppl_2): ii4–16. https://doi.org/10.1093/llc/fqx023.
Hauser, Arnold. 2011. *The Sociology of Art*. Translated by Kenneth J. Northcott. Routledge.
Herrmann, J. Berenike, Karina van Dalen-Oskam, and Christof Schöch. 2015. 'Revisiting Style, a Key Concept in Literary Studies'. *Journal of Literary Theory* 9 (1): 25–52.
Hoover, David L. 2007. 'Corpus Stylistics, Stylometry, and the Styles of Henry James'. *Style* 41 (2): 174–203.
———. 2009. 'Word Frequency, Statistical Stylistics and Authorship Attribution'. In *What's in a Word-List? Investigating Word Frequency and Keyword Extraction*, edited by Dawn Archer, 35–51. Ashgate.
Kenny, Neil. 2020. *Born to Write: Literary Families and Social Hierarchy in Early Modern France*. Oxford University Press.
Krawczyk, Scott, ed. 2009. *Romantic Literary Families*. Nineteenth-Century Major Lives and Letters. Palgrave Macmillan.
Malfait, Olivia, and Marysa Demoor. 2015. 'Sibling Collaboration and Literary After-Life: The Case of the Brontës'. *Brontë Studies* 40 (3): 187–200. https://doi.org/10.1179/1474893215Z.000000000148.
McAleer, Patrick. 2011. *The Writing Family of Stephen King: A Critical Study of the Fiction of Tabitha King, Joe Hill and Owen King*. McFarland & Company.
McCarthy, Rachel, and James O'Sullivan. 2020. 'Who Wrote Wuthering Heights?' *Digital Scholarship in the Humanities*. https://doi.org/10.1093/llc/fqaa031.
O'Sullivan, James, Katarzyna Bazarnik, Maciej Eder, and Jan Rybicki. 2018. 'Measuring Joycean Influences on Flann O'Brien'. *Digital Studies/Le Champ Numérique* 8 (1). https://doi.org/10.16995/dscn.288.
Van Der Meer, Carolyne. 2008. 'Branwell Brontë'. In *The Literary Encyclopedia*. https://www.litencyc.com/php/speople.php?rec=true&UID=581.
Wolff, Janet. 1984. *The Social Production of Art*. New York University Press.