

# Stylometry and Collaborative Authorship: Eddy, Lovecraft, and ‘The Loved Dead’

Alexander A. G. Gladwin, Matthew J. Lavin and Daniel M. Look  
St. Lawrence University, Canton, NY, USA

## Abstract

The authorship of the 1924 short story ‘The Loved Dead’ has been contested by family members of Clifford Martin Eddy, Jr. and Sunand Tryambak Joshi, a leading scholar on Howard Phillips Lovecraft. The authors of this article use stylometric methods to provide evidence for a claim about the authorship of the story and to analyze the nature of Eddy’s collaboration with Lovecraft. Further, we extend Rybicki, Hoover, and Kestemont’s (Collaborative authorship: Conrad, Ford, and rolling delta. *Literary and Linguistic Computing*, 2014; 29, 422–31) analysis of stylometry as it relates to collaborations in order to reveal the necessary considerations for employing a stylometric approach to authorial collaboration.

**Correspondence:** Alexander A. G. Gladwin, 753 Franklin Ave. Columbus, OH 43205 United States.  
**E-mail:** aaggladwin@gmail.com

## 1 Introduction

When ‘The Loved Dead’ was published in the May-June-July 1924 issue of *Weird Tales*—accredited to Clifford Martin Eddy, Jr., or C. M. Eddy, Jr.—controversy followed. The issue was banned in at least Indiana, if not nationwide (Joshi, 2010, p. 501); the magazine’s editor, Farnsworth Wright, would hold a wariness of stories containing socially contentious subject matter for years. ‘The Loved Dead’ is a first-person narrative of a necrophiliac who is on the run from authorities as he explains the roots of his predilection for corpses. The material proved too explicit for audiences, and Wright’s attempts to preclude another controversy caused relatively innocuous stories such as ‘In the Vault’ by Howard Phillips Lovecraft—better known as H. P. Lovecraft, who was a friend of Eddy—to be denied publication in 1925 in fear of another mishap (de Camp, 1975, p. 244).

Although controversy has continued to surround the story, the focus is no longer on the subject matter. Instead, the question of authorship has defined the discourse surrounding ‘The Loved Dead’, because Lovecraft—one of *Weird Tales*’ most popular contributors—is known to have revised it. The extent of his revisions, though, remains unclear.

## 2 Biographical and historical considerations

Because direct historical evidence is typically privileged over computational analysis, we seek to exhaust these resources before enacting a less conventional approach. However, the results are not fruitful; when we consider the fact that Lovecraft ‘revised’ the story, we find only more ambiguity. The term may imply copy-editing or

content suggestions, but Lovecraft often ghostwrote and collaborated on stories while retaining the title of ‘revisionist’. Lovecraft biographer Sunand Tryambak Joshi, or S. T. Joshi (2011), identifies the author’s frequent undermining of his input on revised stories: he and his friend, Robert H. Barlow, cowrote ‘Till A’the Seas’, but Lovecraft allowed it to be published exclusively in Barlow’s name to encourage the young author (p. 10–11). In another case, he took a two-sentence treatment by Zelia Bishop and turned it into a 29,000 word novella, ‘The Mound’ (Joshi, 2010, p. 745); he does not appear to have sought credit in its attempted publication. Thus, the term ‘revision’ is separated from its denotation and even connotations in the context of Lovecraft, and a careful suspicion of declaring the story to be Eddy’s on that basis alone is warranted. To Lovecraft, revision could mean copy-editing, minor rewrites, major rewrites, or sole writing based on ideas. Current primary and secondary claims do not clarify the extent of his contribution.

Additional historical details about the authorship complicate the question further. In terms of first-hand accounts, [Lovecraft \(1925/2005\)](#) does refer to the story in a letter to his aunt as ‘poor Eddy’s “The Loved Dead”’ (p. 252). However, Eddy and Lovecraft were friends; the latter visited the former on occasion, as both were permanent residents of Providence—excepting Lovecraft’s stint in New York City from 1924 to 1926 (Joshi, 2010, p. 466–8). If he were willing to allow his friend Barlow and numerous others to take credit for stories that he worked on, then the possibility that he would refer to the story as Eddy’s is inconclusive of its authorship.

However, in March 1935, [Lovecraft \(1935/1993\)](#) says in a letter to Robert Bloch, ‘It may interest you to know that I revised the now-notorious “Loved Dead” myself—practically re-writing the latter half’ (p. 61). In contrast, Eddy’s wife, [Muriel Eddy \(1961/1998\)](#), claims in her memoir about Lovecraft that he merely ‘read the original manuscript and touched it up in places, with my husband’s full sanction, but it was entirely the brain-child of my husband’ (p. 58–9). [Joshi \(2010\)](#) has called into question many of Muriel Eddy’s claims in this piece, citing a lack of supporting evidence and positing the possibility that

she wished to capitalize on Lovecraft’s fame (p. 464–5). There is no way to be certain in either case, as Joshi himself is merely speculating, and there is no reason to dismiss Muriel’s statement about ‘The Loved Dead’ out of hand.

Jim Dyer—Eddy’s grandson and head of Fenham Publishing, which continues to release collections of Eddy’s stories—further argues in favor of his grandfather’s nigh-complete authorship: ‘There should be no confusion regarding my Grandfather’s stories. Lovecraft and my Grandfather would read their stories aloud to each other and both would give advice and suggestions. My grandfather’s stories were written by him as any of the people who knew both Lovecraft and my Grandfather could attest to’ (personal communication, 21 February 2014). He echoes Muriel’s viewpoints, and the opinion of Eddy’s family on the matter is consistent.

The historical evidence does not provide a clear, single narrative. [Joshi \(2010\)](#) argues in favor of Lovecraft, albeit without any concrete evidence: ‘There was... in all likelihood a draft written by Eddy for this tale’, he says, ‘but the published version certainly reads as if Lovecraft had written the entire thing’ (p. 466). Joshi compares the ‘adjective-choked prose’ to Lovecraft’s contemporary short story, ‘The Hound’.<sup>1</sup> Most likely, Eddy wrote at least the plot, if not an entire draft. The uncertainty lies in how much of the printed tale comes from Eddy’s draft, and how much Lovecraft wrote or rewrote.

## 3 Primary tests

### 3.1 Background

To provide a more in-depth analysis, we look to stylometry, a study that analyzes, per [Holmes’ \(1998\)](#) definition, implicit aspects of an author’s writing style through statistical analysis. There are numerous approaches and debates within the field about the most accurate metrics—e.g. should the focus be on common words such as articles in order to measure subconscious stylistic qualities, or words that appear rarely in order to find authorial stamps—but the principle remains: authors have distinctive qualities to their writing that can be uncovered and statistically analyzed, which provides

evidence regarding the authorship of a disputed text.

We first provide a historical overview of stylometric techniques, because many have come into popularity only to be dismissed as a result of further investigation. Morton (1978) focused on words that appear only once in a text, i.e. *hapax legomena*, and their positions in sentences. Smith (1987) revealed the process' several flaws: small sample sizes, loose statistical inferences, and improper data collection. To question specific stylometric methods is important, because it pushes scholars to find more sound and rigorous ones.

Several tests have yielded useful results, chief among them being Mosteller and Wallace's (1964) testing of the *Federalist Papers*. They attempted to analyze the texts through synonym pairs, such as the selection of the word 'big' over 'large'. However, upon realizing that this method would not work for the texts at hand, they altered their study to focus on function words such as conjunctions and articles, which have little meaning on their own but reveal relationships in the structure of the sentence.

The *Federalist Papers* were written by John Jay, Alexander Hamilton, and James Madison and published under a pseudonym. Scholars and historians were certain of the authorship of seventy-three letters, but the author(s) of the remaining twelve was unclear. Mosteller and Wallace tested claims via Bayesian statistics about who of John Jay, Alexander Hamilton, and James Madison wrote the disputed papers. They discovered that the measures were consistent with Madison, which suggested a high probability that he was the author of all twelve, a claim agreed upon by historians (Juola, 2006, p. 242). This case study suggests that, if statistical rigor is maintained, reliable results can be obtained. The important step is procuring accurate measurements and tests, applying them to suitable texts, and carefully interpreting the content of the results.

There have been only a small number of stylometric analyses of texts where the authorship is collaborative. Rybicki *et al.* (2014) explore the stylistic qualities of *The Inheritors*, *Romance*, and *The Nature of a Crime*, all three of which are accredited to both Ford Madox Ford and Joseph Conrad. Their

method of testing is particularly useful for differentiating authorship in texts that were written in segments.

The publications that contained Lovecraft and Eddy's stories, such as *Weird Tales*, provide numerous authors and stories that can serve as test cases for collaborative authorship due to the literary context of 'hack writing'—a context that adds weight to our inquiry into Lovecraft as revisionist. Usage of the term 'hack', from hackney, to describe a laborer or 'drudge' dates back to the 18th century, but the concept of literary hackwork has its roots in the history of popular periodicals, especially London's Grub Street (Cox and Mowatt, 2014). In the context of the 19th- and early-20th-century United States, literary hackwork was likewise linked to mass-market periodicals and pulp publications. Because the second half of the 19th century saw the rise of the modern profession of authorship and with it a professional discourse of trade books and periodicals, there exist numerous articles defining hackwork and discussing its merits. These sources tend to agree that a hack 'writes for pay, and, if he were not paid, could not write' (Lang, 1896, p. 15). To succeed, a hack had to be 'a sort of all-trades' jack' (Reeve, 1910, p. 55). In comparison with popular notions of authorial identity and celebrity, literary hackwork occurred behind the scenes and carried with it a decreased emphasis on receiving credit. Broadly speaking, hackwork included various levels of unnamed collaborative labor, including ghostwriting, revision, line editing, and adaptation.

Lovecraft in particular is a part of this context, although he does differ from the prescribed image of a hack writer. He often edited and ghostwrote for money, including a short story published in the same issue of *Weird Tales* as 'The Loved Dead': 'Imprisoned with the Pharaohs', originally titled 'Under the Pyramids', which was accredited to famous magician Harry Houdini; Lovecraft received \$100 for his work (Joshi, 2010, p. 498–9). However, Lovecraft does not fit the image of a hack writer in many ways, namely that he was against the idea of commercial writing (Joshi, 2010, p. 297), favoring the image of the amateur who wrote out of personal passion and expression. Still, he did revise, ghostwrite, and line edit—at times for friends, at times

for money due to his financial destitution—which places him in this tradition.

Although Rybicki, Hoover, and Kestmont have begun to study literary collaboration using computational tools, further investigation into the range of literary collaborations of the modern period is called for, an investigation for which hack writing is well suited. The kinds of collaborations we seek to study are notoriously opaque but tremendously common, and crucial to the production of popular and literary texts. Successfully using authorship attribution techniques with these kinds of collaborations would have numerous implications for the study of literature and literary history. Potential fruits include: a better understanding of how and when hackwork took place; better information about the extent of collaboration needed to affect an authorial signal in a text; and, more aspirationally, methods for detecting ghostwritten texts in a field of candidate texts.

### 3.2 Lexical richness

We initially attempted to measure variables of lexical richness, which has a turbulent history in stylometry and has largely faded from popular use. Juola (2006) argues that such measurements have not ‘been demonstrated to be sufficiently distinguishing or sufficiently accurate’, although he does admit that they cannot be dismissed outright with specific counterexamples (p. 240–1). Scholars have found stability with certain variables, including Look’s (2012) study of the works of Robert E. Howard, which argues for the stability of the Type-Token Ratio given certain constraints; Holmes’ (1992) study of authorship in Mormon scripture, which combines variables including the *hapax dislegomena* ratio  $H$ , Honoré’s  $R$ , and Yule’s  $K$ ; and Grieve’s (2007) multifaceted approach to lexical richness, which utilizes the aforementioned and numerous other variables to predict the authorship of a text, achieving a success rate ranging from 59 to 77% when there are two potential authors.<sup>2</sup> However, Tweedie and Baayen (1998) provide evidence toward the instability of  $R$  and  $K$  in relation to the total token count of a text,  $N$ . Its general usage appears to be heavily predicated on having particularly apt test subjects, and even then, a level of certainty reaching even 77% could be seen as too

low for predicting the author of a single disputed text.

We measured  $T$ ,  $H$ ,  $R$ , and  $K$  and performed principal component analysis (PCA) to condense and visualize the data,<sup>3</sup> because lexical richness can provide insight into the style of a text, but our results were too invariable and muddled to use in our test case. However, it can be useful for stylistic analysis of collaborative authorship given particularly apt test cases, so we leave this information for those wishing to perform further testing on other subjects.

### 3.3 Latent features

Measurements of lexical richness can be useful, but we prefer elements of style that can more consistently allow us to distinguish between authors, and such elements have been found in the study of the parts of language that often fail to attract significant attention: function words. Included in this umbrella term are conjunctions, articles, prepositions, particles, and auxiliary verbs. The principle behind measuring function word usage is that it reflects an author’s structural stylistic impulse. Thus, measuring the frequencies of the most common function words and words in general can provide information about an author on a level that cannot be easily imitated.

We will discuss two particularly useful tools, both suggested by Burrows, that focus on function words and common words across texts in general in order to reveal the relationships among corpora.<sup>4</sup>

#### 3.3.1 *Function words*

In the first method, Burrows (1989) outlines the process of taking function word frequencies and applying PCA by analyzing the speech of certain characters in the works of Jane Austen, and then comparing the works of different authors. Binongo (2003) hones the technique in a process that will be outlined here. His focus is the 15th book in the *Oz* series, *The Royal Book of Oz*, which was published 2 years after the death of L. Frank Baum, the undisputed author of the first fourteen books. The controversy is due to a statement in the 15th book’s original publication that claims the text was written by Baum and only ‘enlarged and edited’ by Ruth

Plumly Thompson (as cited in Binongo, 2003, p. 10), who would go on to write the 16th to 33rd entries. However, this claim has been disputed. The most recent edition from 2001 recognizes Thompson's authorship, stirring the controversy.

Binongo performs an authorship attribution test distilled from Burrows' technique. He takes the first thirty-three books in the *Oz* series—including fourteen that are undisputedly by Baum, eighteen undisputedly by Thompson, and *The Royal Book of Oz*—and treats these books as a single corpus. He removes non-prose sections for ease of study. Then, he obtains the proportion of each word in the corpus—i.e. for every distinct word, he measures the number of occurrences in the text, say  $w$ , and calculates  $w / N$ . From this list, he records the top fifty function words, excluding auxiliary verbs and pronouns due to the former's multiplicity of inflections and the latter's dependence on factors such as point of view.

Binongo then subdivides each text into blocks of 5,000 words in order to see variations within a book, creating 223 text blocks; next, he calculates the proportions of the top fifty function words within each 5,000 word subdivision. He then utilizes PCA to distill the information from these measurements into easily visualized data.

Binongo's results are convincing (Fig. 1). When he visualizes the data, works by Baum and those by Thompson are divided along the  $x$ -axis. Binongo notes that even the 14th *Oz* book, *Glinda of Oz*—which was edited by Baum's son from a rough draft—falls distinctly in the cluster of Baum's works. *The Royal Book of Oz*, however, clusters with Thompson's work, lending evidence to the claim that she is the likely author of the text. The fact that the revised version of Baum's work is notably similar to his other works is useful because the disputed text we will be considering, 'The Loved Dead', is possibly just a light revision by Lovecraft. In theory, if he had no greater hand in the published version, we should obtain similar results in favor of Eddy.

This process provides visualizable data on the differences in style between two authors by tracking subconscious stylistic qualities. The utility is clear and will provide information about the disputed

text that we will consider, because a light revision by one author of another's text will not likely change an author's basic grammatical structures.

### 3.3.2 Burrows' Delta

The second method outlined by Burrows (2003) measures style based on common words—including non-function words—and does not privilege visualizations; rather, it yields a single value called Delta that reflects the similarity of subsets of a corpus to a key subset.<sup>5</sup> In our study, that will mean treating 'The Loved Dead' as our key subset, and comparing stories by Lovecraft and Eddy to it. The main set will consist of all of these stories.

The authorial subset that yields the smallest Delta score will be the least unlike the key text in terms of common word usage. Burrows emphasizes using the phrase 'least unlike' in order to clarify that the author with the lowest Delta score does not necessarily have a similar style to the author of the disputed text, but rather is closer in style than any other author with a higher Delta score. Fortunately for us, this will not be an issue, as 'The Loved Dead' is all but certainly the work of Lovecraft, Eddy, or both. The purpose of using this test in conjunction with function word PCA is that it provides another way to measure style as seen

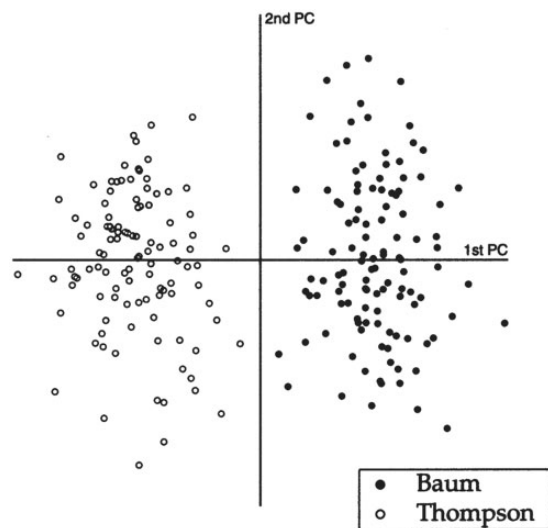


Fig. 1 Binongo's results for function word PCA for Baum and Thompson



in common words—and the results for both tests should thus coincide with each other—but with a potentially different word set.

### 3.3.3 Rolling Delta

Rybicki *et al.* (2014) use Rolling Delta in their study of collaborative authorship, which applies the same basic technique as Burrows' Delta, but provides several Delta scores for a single test text by setting two numbers—a 'rolling window' and a step size—and then 'rolling' through the text. The window is the number of words that will be considered when calculating Delta, and the step size is the increment by which the index of the first word for the window increases. So, for example, if the window size were set to 3,000 and the step size 500, then that test would start with a window containing the first 3,000 words of the text, evaluate the Delta scores for the authorial subsets, then measure another 3,000 word window starting with the 501st word, followed by a 3,000 word window starting with the 1,001st, etc., until a window contained the last word of the text.

There are two other differences between Burrows' Delta and Rolling Delta. First, while the Delta score is weighted by the number of top words considered, Rolling Delta is not. So, if we were to measure the frequencies of the thirty top words, and one author differed from the usage in the test text by 1 standard deviation each time—and, in the case of Rolling Delta, for each rolling window—then the Burrows' Delta score would be 1, but the Rolling Delta scores would be 30. Second, Rolling Delta uses 'culling', which is the process of removing words that do not appear in a certain percentage of the text. If the culling value is thirty, then only words that appear in at least 30% of the texts will be considered when calculating the Rolling Delta scores. This process is useful, in particular for larger texts, because it removes words that are used often, but only in a designated percentage of samples.

## 4 Text selection and natural language processing

Before any tests can be run, the data must be standardized. We choose to work only with prose because

that is the genre of the disputed text. We want to have samples that are of the same language form as each other and as the test text to avoid unforeseen variables. Moreover, because 'The Loved Dead' is fiction, we have decided to exclude essays and letters in case they interfere with the language structures that we are attempting to unearth.<sup>6</sup> Non-English language sections, epigraphs, and chapter numberings (e.g. 'II' or 'Chapter 2') have been removed.

Because we only have twelve Eddy stories to work with (Table 1), we create a subset of Lovecraft's horror prose that contains twelve tales contemporary to 'The Loved Dead' (Table 2).<sup>7</sup>

Contractions are not expanded in the Delta tests as a result of Hoover's (2004) findings that doing so actually lowers accuracy, and are similarly not expanded for the function word PCA test. Spellings are not standardized in order to avoid altering the data. Thus, even sections of the stories written in dialect, e.g. Zadok Allen's monologue in *The Shadow Over Innsmouth*, are left intact, as they represent specific word choices by the authors.

The Lovecraft set is tested against Eddy in relation to not only 'The Loved Dead', but also a text that we know to be entirely by Lovecraft: 'The Mound', which is not in our test sets due to its technical designation as a collaboration between him and Zelia Bishop.<sup>8</sup> Testing 'The Mound' provides base cases that should yield strong results in favor of Lovecraft if the tests are accurately capturing style.

We perform tokenization in Python using the Natural Language Tool Kit's (NLTK) tokenizer, which creates a list wherein each item is a word or piece of punctuation (<http://www.nltk.org/>).<sup>9</sup> All of the tests and calculations for this paper excluding Rolling Delta—which was performed using the Stylo package for R programming language, developed by Eder, Rybicki, and Kestemont (<https://sites.google.com/site/computationalstylistics/stylo>)—are written in Python to optimize performance and collection.<sup>10</sup>

## 5 Validating our approach

To validate our method, we perform our tests in two separate scenarios. First, we run our tests on authors

**Table 1** Eddy's stories and token counts

Title	Token count
An Arbiter of Destiny	2,698
Arhl-A of the Caves	3,601
Ashes	3,182
The Better Choice	3,654
The Cur	2,883
Deaf, Dumb, and Blind	4,640
Eterna	3,278
The Ghost-Eater	3,842
Red Cap of the Mara	4,294
Sign of the Dragon	22,561
Souls and Heels	5,484
With Weapons of Stone	3,498

**Table 2** Lovecraft stories and token count

Title	Token Count
The Festival	3,611
Herbert West-Reanimator	11,894
The Hound	2,942
Hypnos	2,745
The Lurking Fear	8,050
The Music of Erich Zann	3,433
The Nameless City	4,926
The Outsider	2,556
The Picture in the House	3,317
The Rats in the Walls	7,819
The Temple	5,337
The Terrible Old Man	1,126

that worked during the same time and/or in the same genre as Lovecraft to ensure that we can differentiate authors that share those qualities; second, we run our tests on established revisions/collaborations to see how our test results reflect the mixed authorship.

## 5.1 Differentiating Lovecraft from contemporaries

To explore how our tests interpret results for authors working in similar time periods or genres, we choose to test the Lovecraft set against a selection of texts by three authors: Booth Tarkington, who wrote during approximately the same period as Lovecraft, but not in the horror, fantasy, or weird fiction genres; Edgar Allan Poe, whose horror stories greatly influenced Lovecraft (Joshi, 2011, p. 44), but who wrote approximately a century earlier; and

Arthur Machen, who wrote before and during Lovecraft's lifetime and worked in the horror, fantasy, and weird fiction genres. For each author, we select a subset of his/her bibliography (Table 3), as well as one story that will act as a test text in the same way that we will use 'The Mound'. For Tarkington, we use *The Magnificent Ambersons*; for Poe, 'The Fall of the House of Usher', and for Machen, 'The White People'.

### 5.1.1 Function word PCA

We note that only twenty-four top words could be tested due to our smaller sample size and the constraints of our method of PCA.<sup>11</sup> Thus, a lack of clarity in our visualizations could indicate an actual lack of distinction between the two authors in terms of function word usage, or that we are not measuring enough top words. This bolsters the importance of using Burrows' Delta, as we can select any number of top words regardless of the number of samples.

The Lovecraft and Tarkington sets cluster distinctly across the first PC with no overlap (Fig. 2). Both 'The Mound' and *The Magnificent Ambersons* cluster with their respective authors.

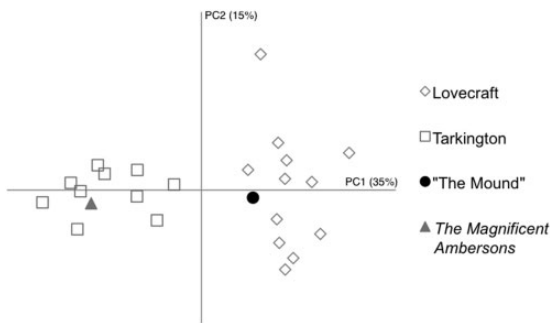
For the comparison between the Lovecraft and Poe sets, the test does not differentiate the stories as clearly (Fig. 3); 'The Mound' and 'The Fall of the House of Usher' appear between the two clusters. This discrepancy shows why it is especially important to follow up on unclear results with other tests in order to check whether the lack of separation is due to the constraints of the test or the texts themselves.

The Lovecraft and Machen sets differentiate clearly (Fig. 4). 'The Mound' and 'The White People' are substantially closer in terms of function word usage to the Lovecraft set and Machen set, respectively.

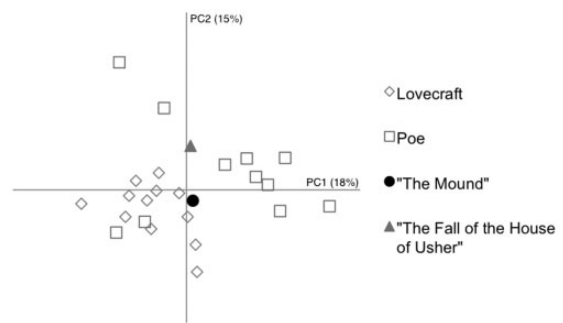
Function word PCA does differentiate the authors distinctly in most cases, with few cases of 'mis-attribution'. The tests do not always demarcate the differences clearly enough, which shows that we should look more closely at the comparison either by increasing the number of top word counts measured or looking to other tests, namely Burrows' Delta.

**Table 3** Tarkington, Poe, and Machen stories

Tarkington		Poe		Machen	
Alice Adams	Harlequin and Columbine	The Black Cat	The Masque of the Red Death	The Angels of Mons	The Hill of Dreams
The Beautiful Lady	Penrod	The Cask of Amontillado	The Murders in the Rue Morgue	Far Off Things	The Inmost Light
The Conquest of Canaan	Penrod and Sam	A Descent into the Maelström	The Pit and the Pendulum	A Fragment of Life	The Secret Glory
The Flirt	Seventeen	The Facts in the Case of M. Valdemar	The Purloined Letter	The Great God Pan	The Terror
Gentle Julia	The Turmoil	Hop-Frog	The Tell-Tale Heart	The Great Return	The Three Imposters
The Gentleman from Indiana		Ligela		Hieroglyphics: A Note upon Ecstasy in Literature	



**Fig. 2** Function word PCA results for Lovecraft and Tarkington

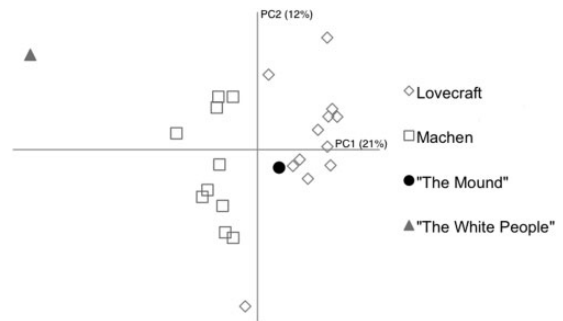


**Fig. 3** Function word PCA results for Lovecraft and Poe

### 5.1.2 Burrows' Delta

Burrows' Delta provides the most accurate results in terms of predicting the authorship of our test texts (Table 4).<sup>12</sup> Although a Delta score for an authorial subset on its own can reveal how much or little that author's usage of common words resembles that of the test text's, the number we will use in comparing authors is the difference between the Delta scores. Because a lower Delta score means an author's style is 'less unlike' that of the test text, the difference between Delta scores reveals the extent to which one author is stylistically closer.

Our results for the comparison of the Lovecraft and Tarkington sets yield smaller Delta scores for Lovecraft when 'The Mound' is the test text, and similarly smaller Delta scores for Tarkington when



**Fig. 4** Function word PCA results for Lovecraft and Machen

*The Magnificent Ambersons* is the test text. The differences range from an absolute value of 0.5 to over 0.8. Our results for Machen show that Lovecraft's style is closer to that of 'The Mound' than



**Table 4** Burrows' Delta results for Lovecraft, Tarkington, Poe, and Machen

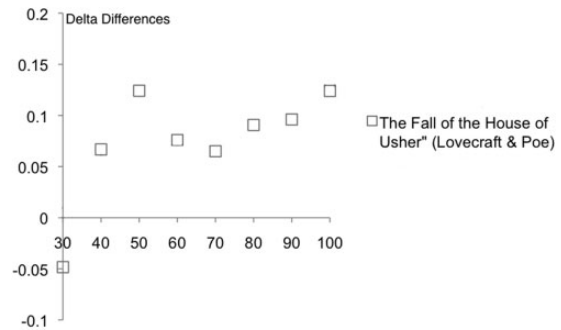
Test	Author 1	Delta 1	Author 2	Delta 2	Diff
'The Mound'	Lovecraft	0.48089	Tarkington	0.98876	0.50787
<i>The Magnificent Ambersons</i>		1.15699		0.29169	0.865306
'The Mound'		0.41838	Poe	0.67259	0.25421
'The Fall of the House of Usher'		0.61295		0.66131	0.04836
'The Mound'		0.4799	Machen	0.69305	0.21315
'The White People'		1.41344		1.22569	0.18775

Machen's, and Machen's is closer to that of 'The White People' than Lovecraft's.

Poe, in contrast, produces similar difficulties to those seen with function word PCA. Although 'The Mound' is closer to the Lovecraft set, 'The Fall of the House of Usher' produces a difference in Delta scores of approximately 0.05 in favor of Lovecraft. Thus, we have a 'false positive', or a result that inaccurately indicates the authorship of a disputed text. To test that Delta is actually unable to differentiate the style of the two authors clearly, though, we return to Burrows' suggestion about the number of top words chosen: although thirty is often enough to differentiate, it is on the low end of the recommended values.<sup>13</sup> Thus, one way we can test our result is to perform Delta tests for various top word counts. Therefore, we perform Delta tests with word counts ranging from 30 to 100 in increments of ten words. The results clearly show that usage of common words in 'The Fall of the House of Usher' is closer to that of the Poe set (Fig. 5). In fact, the more top words we use, the clearer that result becomes. This suggests that if our results for an unknown text yield differences in Delta scores that are below 0.1, then we should further investigate the scores by testing with more top words to ensure that the result for thirty top words is not a misrepresentation of the actual stylistic qualities being measured. As a result, we will consider an author's style to be notably closer to that of the test text if the difference between the authors' Delta scores is 0.1 or greater.

### 5.1.3 Rolling Delta

The results for Rolling Delta are similar to those for Burrows' Delta.<sup>14</sup> Rolling Delta is useful for examining the stylistic qualities of several sections of the test text, which informs collaborative authorship attribution by potentially revealing segmentation. Our

**Fig. 5** Delta differences for Lovecraft and Poe

test texts, however, are not only unsegmented, but also non-collaborative, which means our results should resemble those we found for Burrows' Delta.

In the comparisons between the Tarkington and Lovecraft sets, we see similar results to the ones from Burrows' Delta; for 'The Mound', with a window size of 100% of the windows yield smaller Delta scores for the Lovecraft set, with an average difference of 3.871. For Lovecraft, the Delta scores range from 3.242 to 11.320, and for Tarkington, they range from 8.627 to 17.138. When we use *The Magnificent Ambersons* as a test text, we see similarly clear results: 100% of the windows yield smaller Delta scores for Tarkington, with an average difference of 6.078 in favor of Tarkington. The values for Tarkington range from 5.221 to 11.483, and for Lovecraft, they range from 9.945 to 17.915.

The Rolling Delta results for comparing Poe and the Lovecraft sets similarly reflect our results for Burrows' Delta. When we use 'The Mound' as the test text, the Lovecraft set has smaller Delta scores for 100% of the windows, with an average differences of 3.358 in favor of Lovecraft. The values for Lovecraft range from 3.367 to 12.183, and for Poe,

from 7.441 to 14.635. With ‘The Fall of the House of Usher’—for which we run all Rolling Delta tests with a window size of 1,500 and step size of 100—the results are not as decisive, but still indicate that Poe’s style is predominant: 72.4% of the windows yield smaller scores for Poe, with an average difference for those windows of 0.898. For the windows that favor Lovecraft, the average difference is 0.302. The Delta scores for Poe range from 4.039 to 8.986, and for Lovecraft they range from 3.74544 to 10.27372.

Although there is some misattribution here for the windows in ‘The Fall of the House of Usher’, we remind the reader that these scores are unweighted. Thus, for thirty top words, the average differences of 0.898 and 0.302 are weighted to 0.030 and 0.010. These small values indicate that the test text does not as clearly differentiate the style of the potential authors, and should be considered with the results from the other tests, or further testing.

For Machen, we again see reflections of our results for Burrows’ Delta. For ‘The Mound’, the Lovecraft set has smaller Delta scores for 80.8% of the windows, with an average difference for those windows of 1.302 in favor of Lovecraft. For the windows that yield smaller Delta scores for Machen, the average difference is 0.315. The values for Lovecraft range from 2.926 to 8.724, and for Machen, from 5.238 to 10.012. For ‘The White People’, 87.5% of the windows have smaller Delta scores for Machen, albeit with a smaller average difference of 0.618. For the windows in favor of Lovecraft, the average difference is 0.293. The values for Machen range from 7.443 to 15.978, and for Lovecraft, from 7.616 to 16.095.

Taken in whole, Rolling Delta always favors the proper author, and the cases of mis-attribution do not yield large differences in Delta scores. However, Rolling Delta does get results that correctly identify the true author for test texts that are as small as 4,000 words. Thus, it is sufficiently reliable and will help us test for a segmented collaboration ‘The Loved Dead’.

## 5.2 Determining authorship in an established collaboration

To test an established collaboration, we return to the field of ‘hack writing’, in this case focusing on stories featuring the character Conan the Barbarian. Starting in 1950 and ending in 1954, Gnome Press published five books containing Robert E. Howard’s original Conan stories. In 1955, a sixth book, *Tales of Conan*, was published. However, the stories contained in the sixth book were not original Conan stories written by Howard. Rather, L. Sprague de Camp took four existing non-Conan stories by Howard, recast them as Conan stories, and changed their titles, resulting in ‘*The Road of the Eagles*’, *The Flame Knife*, ‘*Hawks over Shem*’, and *The Blood-Stained God*.

As these works were originally penned by Howard and then heavily edited by de Camp—including changes to the setting, time period, and main characters—these stories are an example of a specific type of revision, one that borders on post-humous collaboration. Given our interest in the nature of the revision versus collaboration, these texts will be compared to the nineteen Conan stories by Howard as well as sixteen non-Conan stories by de Camp (Table 5). For the collaborations, we will

**Table 5** Howard and de Camp stories

Howard		de Camp	
Beyond the Black River	Red Nails	The Ameba	The Hostage of Zir
Black Colossus	Rogues in the House	The Clocks of Iraz	The Inspector’s Teeth
The Devil in Iron	The Scarlet Citadel	The Command	Little Green Men from Afar
Gods of the North	Shadows in the Moonlight	The Emperor’s Fan	The Merman
Hour of the dragon	Shadows in Zamboula	The Gnarly Man	Nothing in the Rules
Jewels of Gwahalur	The Slithering Shadow	The Goblin Tower	Reward of Virtue
The People of the Black Circle	The Tower of the Elephant	The Guided Man	Two Yards of Dragon
The Phoenix on the Sword	The Vale of Lost Women	The Hardwood Pile	The Unbeheaded King
The Pool of the Black One	A Witch Shall Be Born		
Queen of the Black Coast			

use ‘*The Road of the Eagles*’, *The Flame Knife*, and ‘*Hawks over Shem*’ as test texts.

### 5.2.1 Function word PCA

Function word PCA clearly divides the authors across the first principal component, and the test texts all cluster with Howard (Fig. 6).<sup>15</sup> The analysis of function words reflects that heavily edited texts may retain stylistic similarities to the original author, namely in the use of function words; despite de Camp’s heavy editing of the texts, these three test texts clearly cluster with the Howard samples.

### 5.2.2 Burrows’ Delta

Burrows’ Delta corroborates our results from Function Word PCA, with all three test texts yielding smaller Delta scores for Howard than de Camp (Table 6). This reflects the fact that, in terms of common word usage, the test texts resemble Howard more closely than they do de Camp. Based on these tests, we can infer that even a heavily edited text can still bear stylistic similarities to the original author. Notably, both author sets yield small Delta scores, less than 0.5 for all of them, and Howard’s Delta scores for two of the stories is smaller by a margin greater than 0.2. These results imply that both author sets have word usage closer to the test texts than Tarkington, Poe, and Machen do to ‘*The Mound*’, as well as Lovecraft to those authors’ respective texts.

### 5.2.3 Rolling Delta

The results for Rolling Delta similarly reflect the clear stylistic resemblance of the common word usage in the test texts to that of the works of Howard. For ‘*Hawks Over Shem*’, 81.1% of the windows yield smaller Delta scores for Howard, with an average difference of 2.937, whereas the average difference for the windows in favor of de Camp is 1.559. The Delta scores for Howard range from 5.430 to 11.871 and from 6.207 to 17.053 for de Camp.

For *The Flame Knife*, 66.7% of the windows favor Howard with an average difference of 2.691. The windows in favor of de Camp have an average difference of 1.606. For Howard, the values range from

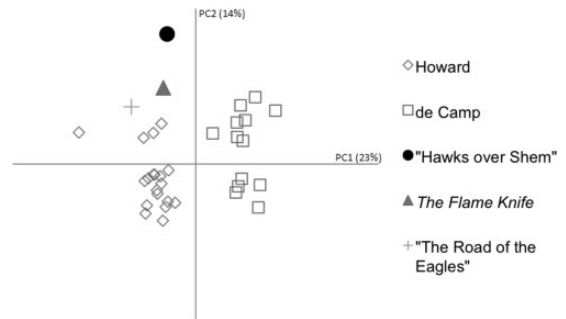


Fig. 6 Function word PCA results for Howard and de Camp

6.638 to 14.814, and from 5.652 to 16.188 for de Camp.

For ‘*The Road of the Eagles*’, 71.7% of the windows favor Howard with an average difference of 3.609, compared to the average difference of 3.444 for the windows in favor of de Camp. For Howard, the values range from 5.646 to 13.813, and from 5.430 to 18.023 for de Camp.

For each text, several windows of the text are stylistically similar to de Camp; this could stem from the authors’ varying word usage, or reflect sections more heavily rewritten by de Camp. Regardless, the results indicate that even though heavy editing can affect the common words usage to resemble the style of the editing author, the style of the original author can remain predominant.

## 6 Results for ‘*The Loved Dead*’

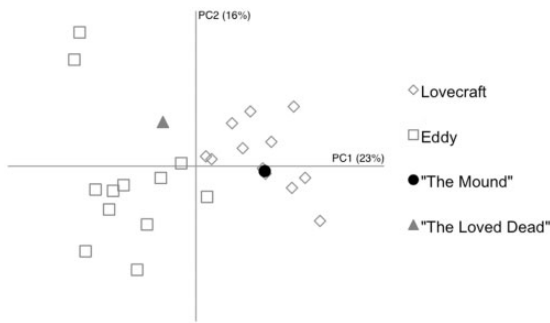
### 6.1 Function word PCA results

The Lovecraft and Eddy sets differentiate with minor overlap, and ‘*The Mound*’ is correctly clustered with Lovecraft’s texts; ‘*The Loved Dead*’ clusters toward the rightmost edge of the Eddy set, although we must again note that only 25 words were tested here for the reasons mentioned in Note 11 (Fig. 7).

These results would imply that ‘*The Loved Dead*’ is closer in style to Eddy in terms of common function word usage, although not as drastically as ‘*The Mound*’ is to Lovecraft. These results could further imply that this test struggles to differentiate the

**Table 6** Burrows’ Delta results for Howard and de Camp

Test	Author 1	Delta 1	Author 2	Delta 2	Diff
‘Hawks over Shem’	Howard	0.25829	de Camp	0.33691	0.07862
<i>The Flame Knife</i>		0.23253		0.46937	0.23684
‘The Road of the Eagles’		0.17150		0.37429	0.20279



**Fig. 7** Function word PCA results for Lovecraft and Eddy

authors, although the ability to accurately categorize ‘The Mound’ so strongly would provide counter-evidence to that claim. Still, another test of common word usage that approaches the measurements differently and considers a less selective set is useful in corroborating or contradicting these findings, and so we look to Burrows’ Delta.

### 6.2 Burrows’ Delta results

Preliminary, Burrows’ Delta tests on ‘The Mound’ with thirty top words and no subdivision clearly distinguish the two authors: Lovecraft’s set has a Delta score of 0.37722, and Eddy’s of 0.73271 (Table 7). The test captures large differences that indicate Lovecraft is stylistically closer in terms of common (function and non-function) word usage. The difference of 0.355 between the two authors is notable for Delta scores, and the low score corroborates the picture painted here that Lovecraft is stylistically closer to the author of ‘The Mound’ than Eddy.

When ‘The Loved Dead’ is the test text, the results are less staggering: Lovecraft’s set has a Delta score of 1.03259, and Eddy’s, 0.96894, resulting in a difference of 0.063651 in favor of the Eddy set.

Nonetheless, the results show that Eddy is closer in common word usage than Lovecraft.

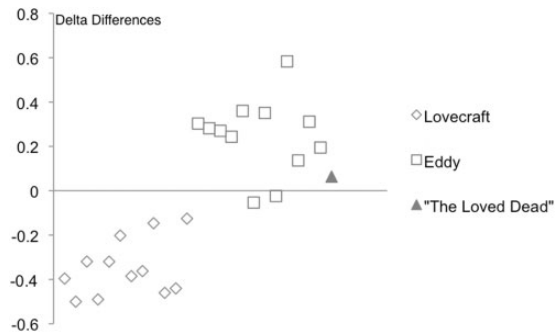
A possibility for the lack of such large differences as those seen in ‘The Mound’ is that ‘The Loved Dead’ is a significantly shorter text—nearly one-seventh of the Lovecraft story’s token count. Thus, in order to find out how this difference compares to other Delta scores, we perform Burrows’ Delta with each of the Lovecraft and Eddy stories as the test text one at a time, tested against the Lovecraft and Eddy subsets. In each case, we remove the test text from the respective author subset. This allows us to see how many false positives are found in stories for which we know the author, and find the range for the differences in Delta scores.

We visualize the results for these tests by plotting the difference between Delta scores for each graph, with a negative difference reflecting a smaller score for Lovecraft, and a positive difference reflecting a smaller Delta score for Eddy. Thus, a story that reads closer to Eddy will have a positive difference, while a story that reads closer to Lovecraft will have a negative difference. We find that there are no false positives for Lovecraft, but two for Eddy: ‘The Ghost-Eater’ and ‘Deaf, Dumb, and Blind’ (Fig. 8).

The differences in Delta scores for these stories are smaller than most of the other tests, which implies that the common word usage is not nearly as distinguishing as it is for the other texts. To further investigate the false positives, we reapply the method we used when Delta scores for Lovecraft and Poe tested against ‘The Fall of the House of Usher’ yielded a false positive: we test Delta scores for word counts ranging from 30 to 100 in increments of ten. When we perform these tests, we see only confirmation that the styles of the stories are closer to Lovecraft, rather than the clear picture we saw with ‘The Fall of the House of Usher’ (Fig. 9). ‘Deaf, Dumb, and Blind’ in particular consistently

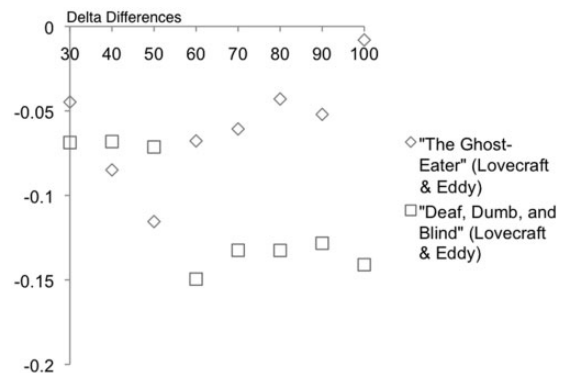
**Table 7** Burrows' Delta results for Lovecraft and Eddy

Test	Author 1	Delta 1	Author 2	Delta 2	Diff
'The Mound'	Lovecraft	0.37722	Eddy	0.73271	0.35549
'The Loved Dead'		1.03259		0.96894	0.063651

**Fig. 8** Delta differences for Lovecraft and Eddy

yields a stronger score for Lovecraft by a margin greater than 0.1.

'The Ghost-Eater' and 'Deaf, Dumb, and Blind' are the only two stories that Delta fails to distinguish in our testing. The data for the function word PCA graphs reveal that the two rightmost Eddy values—i.e. those closest to the Lovecraft cluster, which implies they are less distinctly similar to Eddy in terms of function word usage—are 'The Ghost-Eater' and 'Deaf, Dumb, and Blind' (Fig. 7). This is notable because these are two stories on which Lovecraft is also known to have done revision work. In terms of 'The Ghost-Eater', Lovecraft (as cited in Derleth, 1966) claims in a 1923 letter to Muriel Eddy that he 'made two or three minor revisions', which Joshi (2010) finds agreeable, claiming that he 'cannot detect much actual Lovecraft prose, unless he was deliberately altering his style' (p. 466). In regard to 'Deaf, Dumb, and Blind', there is little in terms of Lovecraft's own claims, although Joshi adds that—while Eddy only implies that Lovecraft fixed up the last paragraph—'in truth, the entire tale was probably revised, although again Eddy very likely had prepared a draft' (p. 467). Although the margin for Delta scores with 'The Ghost-Eater' are relatively small, the differences for 'Deaf, Dumb, and Blind' are large

**Fig. 9** Delta differences for Lovecraft and Eddy

enough to support the claim that Lovecraft's style is as present—if not more present—than Eddy's.

Both stories are actually closer to Lovecraft's style in terms of common word usage than 'The Loved Dead', and are the only two of the Eddy subset for which that is true. Thus, we have inadvertently found that these three stories are the most difficult to differentiate stylistically between Lovecraft and Eddy; therefore, the stories have measurable similarities to the former's style. When we consider the Delta scores that strongly favored Howard over de Camp, the scores for 'The Loved Dead', 'The Ghost-Eater', and 'Deaf, Dumb, and Blind' could provide evidence toward claims that Lovecraft revised the stories substantially, because even de Camp's extensive rewriting did not make his common word usage more predominant in the texts. This would imply that he rewrote at least portions of the story, and the small difference in scores could imply that he did indeed rewrite the second half.

When we apply the varying word count methodology to Delta tests for 'The Mound' and 'The Loved Dead', we find a corroboration of our previous findings for the two stories, namely, 'The Mound' is definitively closer in style to the Lovecraft set, and 'The Loved Dead' is not



significantly differentiated, as the author with the smaller Delta score changes, given the number of top words considered, and the differences are within 0.1 of each other (Fig. 10).

### 6.3 Rolling Delta results

Our Rolling Delta results largely reflect our Burrows' Delta results.<sup>16</sup> For 'The Mound', 94.7% of the windows have smaller Delta scores for the Lovecraft set when compared to Eddy, with an average difference in Delta scores of 1.791. For the windows that favor Eddy, the average difference is 0.186. The Delta scores for Lovecraft range from 3.301 to 10.301, and for Eddy they range from 4.530 to 12.008. For 'The Loved Dead', interestingly, Eddy yields smaller Delta scores for 100% of the windows, with an average difference of 1.996. The values range from 5.624 to 7.771 for Eddy, and from 7.724 to 9.838 for Lovecraft.

To further investigate these results, we look into the effects of culling. Because 'The Loved Dead' is such a small text, and we are already working with shorter and fewer texts, the effects of culling can be severe. Even common words might not show up in a smaller text, and thus would be removed from the set due to our decision to set the culling rate to 100. When the culling rate is 100, we do not control the number of top words considered, which means the count can be below thirty. We will consider no culling with 30, 50, and 100 top words.

When we perform Rolling Delta with thirty top words (Fig. 11), our results seem in line with our findings in Burrows' Delta. Sixty percent of the windows favor Lovecraft, with an average difference of 0.435, compared to the average difference of 0.311 for the windows that favor Eddy. The values for Lovecraft range from 14.016 to 15.669, and from 13.933 to 16.770 for Eddy. The windows are almost evenly split in terms of which author has a smaller Delta score, which reflects the fact that the test can not clearly differentiate the style of the author. However, the windows with the largest difference in Delta scores are the last two, which due to the window size encapsulate the entire second half of the story. The differences for these windows are 1.249 and 1.470, which are greater than the average differences by a factor of approximately 4.

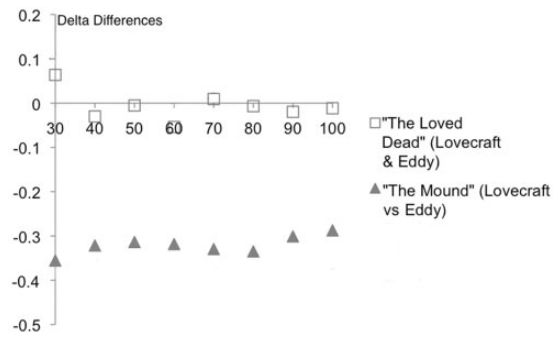


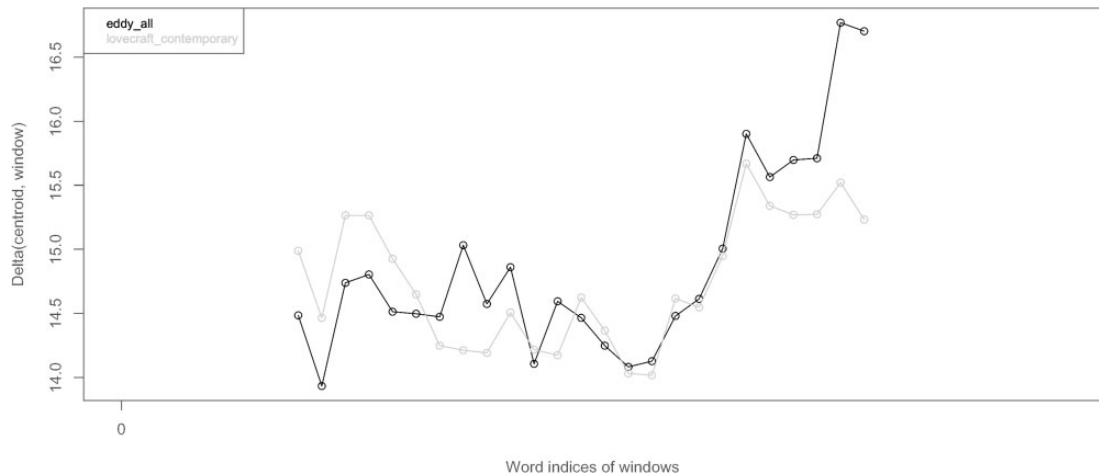
Fig. 10 Delta differences for Lovecraft and Eddy

When we perform Rolling Delta with 50 (Fig. 12) and 100 (Fig. 13) top words, the Lovecraft set garners a smaller Delta score for every window, with respective average differences of 1.036 and 1.448. The values for the tests with fifty words range from 16.738 to 18.735 for Lovecraft and from 17.217 to 20.263 for Eddy. The values for the tests with 100 words range from 21.848 to 23.864 for Lovecraft and from 23.278 to 26.088 for Eddy. In both cases, the last two windows yield the largest difference in Delta scores. For Rolling Delta with fifty top words, the differences for the last two windows are 2.537 and 2.828—approximately twice the average difference—and for 100 top words, the differences for the last two windows are 2.406 and 2.732—again approximately twice the average difference.

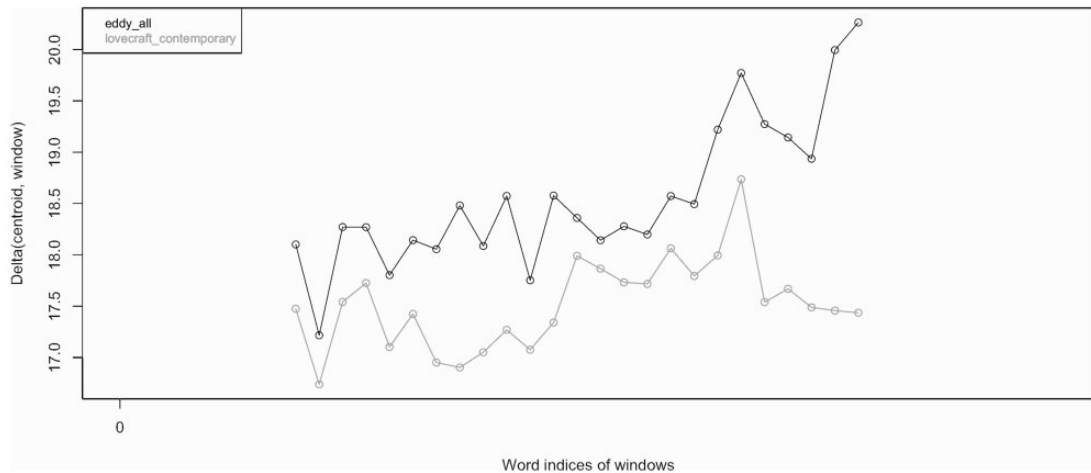
## 7 Conclusion

### 7.1 Interpretation of results

Our tests consistently reveal that 'The Loved Dead' does not bear an overwhelming stylistic similarity to Lovecraft or Eddy, and is consistent with the conjecture that the story is a collaboration. We performed lexical richness PCA to see if we could detect stylistic similarities in infrequently used words. However, these tests did not reveal anything regarding the authorship of 'The Loved Dead' and are known to be of questionable reliability in general. We performed function word PCA because the analysis of common word usage is more reliable, and this test indicated that 'The Loved Dead' is



**Fig. 11** Rolling Delta results for Lovecraft and Eddy



**Fig. 12** Rolling Delta results for Lovecraft and Eddy

closer in terms of function word usage to Eddy's style, although it does share similarities to Lovecraft's.

We selected Burrows' Delta because it proved to be the most reliable in our preliminary tests, and allows us to look at most common word usage and not just function word usage. Again, our results showed that 'The Loved Dead' is stylistically closer to Eddy, but also has stylistic similarities to Lovecraft, which could imply that Lovecraft's presence as author in the story is a larger revision akin to de Camp's revision of Howard stories. Further

study on the relationship between the extent of a collaboration and the margin of Delta scores is warranted. Finally, we also tested with Rolling Delta to see if 'The Loved Dead' might be a segmented collaboration, and while our results generally suggest that both authors' style is detectable in the whole story—again, implying that Lovecraft likely rewrote sections of the text—Lovecraft's is particularly noticeable in the story's second half, which corroborates his epistolary claim.

The results for all four tests, when considered in tandem, suggest that Lovecraft likely edited 'The

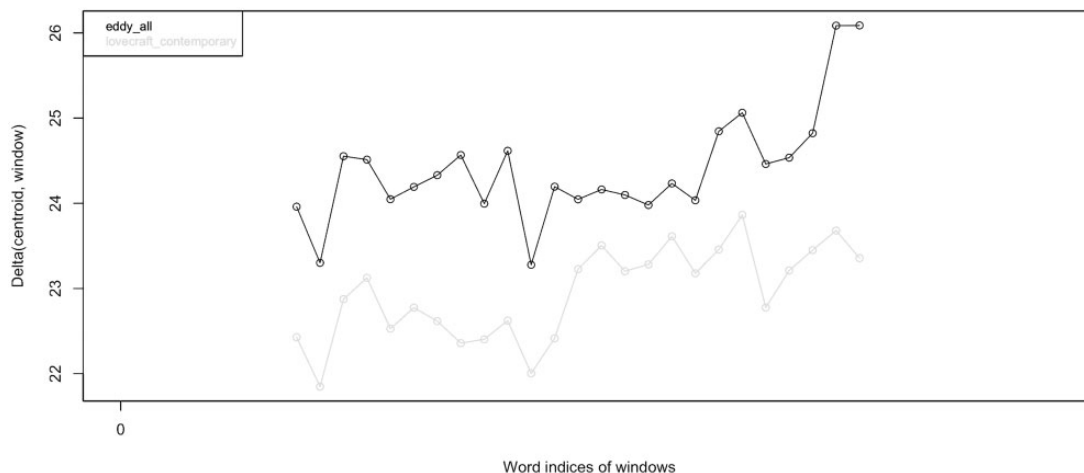


Fig. 13 Rolling Delta results for Lovecraft and Eddy

Loved Dead’, perhaps extensively, but did not write the entirety or majority of the story as it appeared in *Weird Tales*. At most, it appears that his edits focused on the second half of the story, as he stated in his letters. This claim supports a middle-ground between the claims made by Joshi and Eddy’s family. No definite claims can be made—as, we want to stress, this information is only evidence and should not be considered an endpoint for scholarship on the matter—but the evidence certainly goes against any claim that either author is solely responsible for ‘The Loved Dead’. We have found that the same is true for ‘The Ghost-Eater’ and ‘Deaf, Dumb, and Blind’.

Our set of tests allows for a deeper understanding of the extent to which multiple authors might have contributed to a text. Although the tests do require interpretive work by the person performing the tests, the use of these four tests in tandem is useful because each provides unique insights while potentially corroborating the results found in the other three. Although the results of stylometric analyses cannot demonstrate causality, mathematical methods can describe features that correlate with authorial identity, which is particularly useful when historical evidence is not present. Further, we find that ‘hack writers’ of the pulp market provide particularly fruitful test cases of collaborative authorship. Future analyses of collaborative authorship might extend our findings by applying our

methods to other collaborative case studies, comparing our methods with machine-learning methods, or studying how ‘collaborative proportion’ affects Delta scores.

## References

- Binongo, J. N. G.** (2003). Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution. *Chance*, 16(2): 9–17.
- Burrows, J.** (1989). ‘An ocean where each kind...’: statistical analysis and some major determinants of literary style. *Computers and the Humanities*, 23(4/5): 309–21.
- Burrows, J.** (2003). 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(1): 5–32.
- Cox, H. and Mowatt, S.** (2014). *Revolutions from Grub Street: A History of Magazine Publishing in Britain*. Oxford: Oxford UP.
- de Camp, L. S.** (1975). *Lovecraft: A Biography*. London: New English Library.
- Derleth, A. W.** (1966). [Introduction]. In Divers Hand and H. P. Lovecraft (Author), *The Dark Brotherhood and Other Pieces*. Sauk City, WI: Arkham House, pp. ix–x.
- Eddy, M.** (1998). The gentleman from Angell Street. In Cannon P. H. (ed.), *Lovecraft Remembered*. Sauk City, WI: Arkham House, pp. 49–64. (Original work published 1961)

- Grieve, J.** (2007). Quantitative authorship attribution: an evaluation of techniques. *Literary and Linguistic Computing*, 22(3): 251–70.
- Holmes, D. I.** (1992). A stylometric analysis of Mormon scripture and related texts. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155(1): 91–120.
- Holmes, D. I.** (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing*, 13(3): 111–7.
- Honorè, A.** (1979). Some simple measures of richness of vocabulary. *Association for Literary and Linguistic Computing Bulletin*, 7: 172–7.
- Hoover, D. L.** (2004). Delta prime? *Literary and Linguistic Computing*, 19(4): 477–95.
- Joshi, S. T.** (1981). A note on the texts [Foreword]. In Joshi S. T. (ed.) and H. P. Lovecraft (Author), *The Horror in the Museum*. New York: Del Rey, Reprint edn., pp. xv–xviii.
- Joshi, S. T.** (2010). *I Am Providence: The Life and Times of H.P. Lovecraft*. New York: Hippocampus.
- Joshi, S. T.** (2011). [Introduction]. In Joshi S. T. (ed.) and H. P. Lovecraft (Author), *The Crawling Chaos and Others: The Annotated Revisions and Collaborations of H. P. Lovecraft*, Vol. 1. Welches, OR: Arcane Wisdom, pp. 7–17.
- Juola, P.** (2006). Authorship attribution. *Foundations and Trends in Information Retrieval*, 1(3): 233–334.
- Kuiper, S. and Sklar, J.** (2012). Principal component analysis: stock market values. In *Practicing Statistics: Guided Investigations for the Second Course*, 1<sup>st</sup> edn. Upper Saddle River, NJ: Pearson, pp. 332–68.
- Lang, A.** (1896). In defense of the literary hack. *Current Literature*, 20: 15.
- Look, D. M.** (2012). Statistics in the Hyborian Age: an introduction to stylometry. In Prida J. (ed.), *Conan Meets the Academy: Multidisciplinary Essays on the Enduring Barbarian*. Jefferson, NC: McFarland, pp. 103–22.
- Lovecraft, H. P.** (1993). *Letters to Robert Bloch*. S. T. Joshi and D. E. Schultz (Eds.). West Warwick, RI: Necronomicon Press.
- Lovecraft, H. P.** (2005). Letter to Lillian D. Clark [13 December 1925]. In Joshi S. T. and Schultz D. E. (eds), *Letters from New York*, vol. 2. Lovecraft Letters. San Francisco: Night Shade, p. 252.
- Morton, A. Q.** (1978). *Literary Detection*. New York: Scribners.
- Mosteller, F. and Wallace, D. L.** (1964). *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley.
- Reeve, J. K.** (1910). *Practical Authorship*. Ridged, NJ: Editor Company.
- Rybicki, J., Hoover, D., and Kestemont, M.** (2014). Collaborative authorship: Conrad, Ford, and rolling delta. *Literary and Linguistic Computing*, 29: 422–31.
- Smith, M. W. A.** (1987). Hapax legomena in prescribed positions: an investigation of recent proposals to resolve problems of authorship. *Literary and Linguistic Computing*, 2: 145–52.
- Tweedie, F. J. and Baayen, R. H.** (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and The Humanities*, 32: 323–52.
- Yule, G. U.** (1944). *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

## Notes

- 1 This opinion seems to have evolved over two decades; writing three decades prior, [Joshi \(1981\)](#) states, ‘the two authors probably contributed equally’ (p. xvii).
- 2  $T = V/N$ , where  $V$  is the number of distinct words in a text and  $N$ , the total number of words.  $H = V_2/V$  where  $V_2$  is the number of words appearing exactly twice in a text.  $R = \frac{100 \log(N)}{1 - V_1/V}$  a value first suggested and defined in [Honorè \(1979\)](#).  $K = \frac{10^4 (\sum i^2 V_i - N)}{N^2}$ , where  $V_i$  is the number of words appearing  $i$  times; this value was first suggested and defined in [Yule \(1944\)](#).
- 3 The process for performing PCA is outlined by [Kuiper and Sklar \(2012\)](#).
- 4 A corpus is defined as any collection of texts. In our case, we have three subsets: ‘The Loved Dead’, Lovecraft texts, and Eddy texts.
- 5 The mathematics behind calculating Delta are outlined more completely in [Burrows \(2003\)](#), but are outlined in short here. A main set—or, to use a familiar term, corpus—is created from all of the texts under consideration. Any number of the most frequent words between 30 and 150 are found, and Burrows separates homographs such as ‘before’, which can be either a conjunction or preposition, although the practice is not necessary. He then finds the percentage that each of these words takes up for each text of the corpora, and standardizes these percentages based on the mean and standard deviation for that word across the texts. A Delta score for a subset is the average absolute value

of the average  $z$ -scores for the number of words being considered across all the texts in that subset. Thus, if we have a key text  $k$  and author subset  $a$ , the Delta score for  $a$  for a list of  $n$  words is:

$$\text{Delta}(a) = \frac{\sum_{i=1}^n |z_{k_i} - z_{a_i}|}{n}.$$

- 6 The discarded draft of *The Shadow Over Innsmouth*, a purposefully experimental piece, is not included, as it was a deliberate attempt by Lovecraft to depart from his usual style (Joshi, 2010, p. 791). ‘The Thing in the Moonlight’, which was adapted from notes in one of Lovecraft’s letters, is also not included, as he did not write the published piece (Joshi, 2010, p. 697).
- 7 We initially ran tests with this set and a larger set of all of Lovecraft’s horror stories, which excludes his prose poems, humorous stories, collaborations, ghost-written projects, and so-called ‘Dream Cycle’ stories. All of the results were largely identical and did not add insight into our testing.
- 8 ‘The Mound’ contains 29,019 tokens and ‘The Loved Dead’, 3,909.
- 9 Tokenization involves three major steps: first, we remove apostrophes, semicolons, commas, and periods. One consequence of this process is that contractions can be confused with homographs, e.g. ‘can’t’ and ‘cant’; however, this does not skew our results significantly. Second, we tokenize the texts with NLTK’s function. Third, we remove any remaining punctuation or leftover odd characters such as HTML tags from the tokenized list. The remaining list consists of words only, which allows us to perform the frequency tests.
- 10 We find the values used to calculate  $T$ ,  $H$ ,  $R$ , and  $K$ , as well as word frequencies, using methods in NLTK’s FreqDist class, which creates a frequency distribution

of the tokens in a text. We utilize SciPy’s ‘NumPy’ and ‘stats’ packages. The former allows us to store data in a matrix format for use in PCA computation, and the latter allows us to standardize values as  $z$ -scores (<http://www.scipy.org>). We perform PCA using Matplotlib’s ‘mlab’ module, which mimics MATLAB functions (<http://www.matplotlib.org>).

- 11 PCA, in our case, treats texts as rows and measurements as columns, and the tool we use to run PCA—Matplotlib’s ‘mlab’ module—requires that there be more rows than columns.
- 12 We have chosen to perform these tests using only two authors at a time in order to most closely emulate the testing procedure we will use for ‘The Loved Dead’.
- 13 We use thirty as the word count as a default because it is, generally, sufficient. It also avoids any potential complications from including too many top words, e.g. incorporating words that are used commonly in only one set. This is the reason that, when the Delta scores are close to each other, we try a range of values between 30 and 100: it gives us a complete picture of the differences in common word usages between the sets, accounting for the potential errors associated with the extremes of our word count range.
- 14 We run this test at a culling rate of 100. Unless otherwise stated, the window size is 5,000 and the step size is 1,000.
- 15 Two de Camp samples are not visible in the selected window. They carried such highly positive values for the first principal component that their inclusion made the samples on the graph more difficult to discern.
- 16 We use a window size of 1,500 words and a step size of 100 words for tests with ‘The Loved Dead’, as that is the low end of the range stated by Burrows for Delta.