Attributing the Bixby Letter using n-gram tracing

Grieve, Jack; Emily, Chiang; Clarke, Isobelle; Gideon, Hannah; Heini, Annina; Nini, Andrea; Waibel, Emily

DOI:
10.1093/llc/fqy042

Document Version
Peer reviewed version

Citation for published version (Harvard):
Grieve, J, Emily, C, Clarke, I, Gideon, H, Heini, A, Nini, A & Waibel, E 2019, 'Attributing the *Bixby Letter* using n-gram tracing', *Digital Scholarship in the Humanities*, vol. 34, no. 3, pp. 493–512.
https://doi.org/10.1093/llc/fqy042

Attributing the *Bixby Letter* using n-gram tracing

Jack Grieve[1], Emily Chiang[2], Isobelle Clarke[1], Hannah Gideon[2], Annina Heini[2], Andrea Nini[3] and Emily Waibel[2]

[1] University of Birmingham

[2] Aston University

[3] University of Manchester

Submitted to *Digital Scholarship in the Humanities*, 26 May 2017

Minor revisions requested by *Digital Scholarship in the Humanities*, 10 April 2018

Revised version submitted to *Digital Scholarship in the Humanities*, 7 June 2018

Minor revisions requested by *Digital Scholarship in the Humanities*, 1 August 2018

Final revised version submitted to *Digital Scholarship in the Humanities*, 3 August 2018

Accepted by *Digital Scholarship in the Humanities*, 8 August 2018

Abstract

There is a long-standing debate around the authorship of the *Bixby Letter*, one of the most famous pieces of correspondence in American history. Despite being signed by President Abraham Lincoln, some historians have claimed that its true author was John Hay, Lincoln's personal secretary. Analyses of the letter have been inconclusive in part because the text totals only 139 words and is thus far too short to be attributed using standard methods. To test whether Lincoln or Hay wrote this letter, we therefore introduce and apply a new technique for attributing short texts called *n-gram tracing*. After demonstrating that our method can distinguish between the known writings of Lincoln and Hay with a very high degree of accuracy, we use it to attribute the *Bixby Letter*, concluding that the text was authored by John Hay – rewriting this one episode in the history of the United States and offering a solution to one of the most persistent problems in authorship attribution.

Keywords: American History, Authorship Attribution, Computational Social Science, Corpus Linguistics, Forensic Linguistics, John Hay, Abraham Lincoln, Stylistics, Stylometry

Acknowledgements

We would like to thank Michael Burlingame, Mariá Csemezová, Tim Grant, Cristina Greco, Liubov Green, Krzys Kredens, Olu Popoola, Maria Tagtalidou, and David Wright and for their comments on this paper and their assistance with this project.

Attributing the Bixby Letter using n-gram tracing

1. Introduction

On the 21st of November 1864, only five months before he was assassinated, Abraham Lincoln, the 16th President of the United States, sent a short letter of condolence to Lydia Bixby of Boston, a widow whose five sons were believed to have died in the Civil War. The original letter was lost, but the Adjutant General of Massachusetts, who had requested the letter from the Department of War on the widow's behalf, also sent a copy to the *Boston Evening Transcript*, who published the letter on the 24th of November (see Table 1). The *Bixby Letter* would go on to become one of America's most famous pieces of correspondence, praised for its sentiment and style and counted among Lincoln's greatest texts along with the *Gettysburg Address,* the *Second Inaugural Address,* and the *Emancipation Proclamation*. The authorship of the letter, however, has long been the subject of debate, with some historians arguing that its true author was John Hay – Lincoln's young assistant and the future Secretary of State under William McKinley and Theodore Roosevelt.

Table 1 The *Bixby Letter* (*Boston Evening Transcript,* 25 November 1864)

EXECUTIVE MANSION,
WASHINGTON, NOV. 21, 1864.
Dear Madam,—

I have been shown in the files of the War Department a statement of the Adjutant General of Massachusetts, that you are the mother of five sons who have died gloriously on the field of battle.

I feel how weak and fruitless must be any words of mine which should attempt to beguile you from the grief of a loss so overwhelming. But I cannot refrain from tendering to you the consolation that may be found in the thanks of the Republic they died to save.

I pray that our Heavenly Father may assuage the anguish of your bereavement, and leave you only the cherished memory of the loved and lost, and the solemn pride that must be yours, to have laid so costly a sacrifice upon the altar of Freedom.

Yours, very sincerely and respectfully,

MRS. BIXBY. *A. LINCOLN.*

A wide range of external evidence has been presented in favour of both Lincoln (e.g. Barton, 1926; Basler, 1953; Randall & Current, 1955; Bullard, 1946, 1951; Emerson, 2006, 2008) and Hay (e.g. Butler, 1940; Wakefield, 1948; Burlingame, 1995, 1999). Hay is generally acknowledged to have written much of Lincoln's correspondence, as this was the task for which he was hired by John George Nicolay, Lincoln's other personal secretary, after Lincoln had secured the Republican presidential nomination in May 1860 (Kushner, 1974). Furthermore, several reliable sources – including Nicholas Murray Butler, the president of Columbia University, and Spencer Eddy, Hay's personal secretary later in life – claimed that Hay had confided in them that he had written the letter. In addition, Hay kept scrapbooks containing extensive records of his achievements, which included the *Bixby Letter*, as well as references to many texts he had certainly written, including his 1883 novel *The Bread Winners* and a series of letters sent to newspapers across the country in support of Lincoln, both of which were initially published anonymously (Kushner & Hummel, 1977). Alternatively, aside from the fact that the letter bears his name, perhaps the most convincing evidence that Lincoln wrote the *Bixby Letter* is that Hay never publicly took credit for its authorship, although he did take credit for other letters sent by the President. Hay and Nicolay even attributed the letter to Lincoln in their biography of the President (1890) and Hay's children said that their father never claimed authorship in private. Furthermore, although Hay authored much of Lincoln's correspondence at that time, Lincoln did write some letters, including letters of condolence, and he might have been especially likely to have written this letter, as he had lost three sons himself. His one surviving son, Robert Todd Lincoln, who was Hay's close friend, also asserted that his father had written the *Bixby Letter* and that Hay had confirmed as much to him personally.

In addition to external evidence, internal evidence related to the style of the *Bixby Letter* has been presented in support of both Lincoln and Hay. In 1943, Basler remarked on the quality of the letter and its similarity to Lincoln's style (Burlingame, 1995); ten years later,

he included the letter in his *Collected Works of Abraham Lincoln*. Similarly, Bullard (1946) argued that the letter was generally a better match for Lincoln's style than Hay's. A more thorough analysis was presented by Nickell (1989), who identified several distinctive words, phrases, and rhythms in the letter, for which he could only find analogues in Lincoln's writings, including the use of alliteration and the word 'tender'. Nickell also argued that Lincoln wrote in a more traditional and formal style, whereas the younger Hay wrote in a more contemporary and informal style. For example, Nickell claimed that the use of the word 'beguile' in the letter is used with its traditional sense of 'diverting', as opposed to the more modern sense of 'enticing', which is how Hay used the word in a letter Nickell quotes. Burlingame (1999), however, who has been one of the strongest proponents of Hay's authorship, found that Hay used 'beguile' at least 30 times in his writings, including in a collection of unpublished letters, while he could find no record of Lincoln ever having used the word. Burlingame (1995) also argued that various other words were indicative of Hay, including 'gloriously', 'cherish', 'republic', and 'Heavenly Father'.

The stylistic evidence is far from definitive. Burlingame and others have claimed that more passages in the *Bixby Letter* resemble Hay's known writings, while Nickell and others have claimed that more resemble Lincoln's. Emerson (2006: 2) dismissed this type of internal evidence outright, stating that 'one can find as many arguments in favour of Lincoln's literary style as one can find for Hay's.' Developing objective methods for attributing authorship, however, is the focus of considerable research in *stylometry* (Koppel et al., 2009; Stamatatos, 2009), where questioned documents are attributed, for example, by comparing the frequencies of common words or common word and character sequences in the text to their frequencies in writing samples from each possible author. The *Bixby Letter* has never been subjected to thorough stylometric analysis, at least in part, because it only contains 139 words; short texts are difficult to attribute using stylometric techniques because the relative frequencies of linguistic features in a text can only be trusted to approximate

their values in an author's writings more generally if that text is long enough to contain numerous tokens of those features. For example, the word 'beguile' occurs once in the *Bixby Letter*, but we should not assume its author used this word on average about once every 139 words. Similarly, the word 'by' does not occur in the letter, but we should not assume its author never used this word at all.

The problem of text length has received considerable attention in stylometry, with Stamatatos (2009: 553) calling it 'the most important' methodological issue in the field. Eder (2015) conducted the most thorough assessment of the effect of questioned document length in authorship attribution and recommended a minimum length of 5,000 words; this is a very conservative limit, at least in part because his tests involved between 6 and 21 possible authors, as opposed to the basic problem of 2 authors, which requires less data. Alternatively, many studies have been able to successfully attribute texts of around 1,000 (e.g. Stamatatos et al., 2001; Burrows, 2002; Juola, 2006; Stamatatos, 2009) or 500 words (e.g. Gamon, 2004; Grieve, 2007; Koppel, Schler & Argamon, 2011). Few studies have attributed shorter texts, although some promising results have been achieved in the 200- to 500-word range (e.g. Forsyth & Holmes, 1996; Koppel et al., 2011), especially based on the frequencies of relatively common parts-of-speech (e.g. Chaski, 2005; Hirst & Feiguina, 2007). The attribution of texts shorter than 200 words has received very little attention, limited mostly to a small number of recent studies of Twitter data. Most notably, Layton et al. (2010) were able to attribute posts based primarily on references to usernames, while Schwartz et al. (2013) were able to attribute posts based on character and word sequences that are used by only one author in their corpus. Although both methods worked well for classifying posts that contained these features, a substantial proportion of posts resisted attribution. Better results were achieved by Brocardo et al. (2013), who proposed a method for short-text *authorship verification* – which involves testing whether an author wrote a text, as opposed to *authorship attribution*, which involves selecting the most likely author from a

132 set of candidates, as in the case of the *Bixby Letter*. Their method is based on the number of
133 character sequences in the questioned document that also occur in the known writings of an
134 author. Crucially, all three of these studies measured the presence and absence of linguistic
135 features as opposed to their relative frequencies, whose value is limited in short texts.
136 Totalling only 139 words, the *Bixby Letter* is far too short to be attributed using
137 standard stylometric techniques. Short documents, however, are common in a forensic
138 context (Coulthard, 2004; Coulthard et al., 2017). For example, the mean length of texts
139 received by the German Federal Criminal Police Office between 2002 and 2005 was 248
140 words, with two thirds of incriminating texts containing fewer than 200 words (Ehrhardt,
141 2007). A common method for attributing texts of any length in forensic stylistics is to
142 manually identify features of interest in the questioned document and to then search for
143 those features in the possible author writing samples to see if they are used predominantly
144 by one suspect (e.g. McMenamin, 1993, 2002). This approach is based on the reasonable
145 assumption that the repetition of features across texts is evidence of shared authorship (see
146 Coulthard, 2004). Still feature selection is usually left to the judgment of the forensic linguist,
147 limiting the reliability of this approach in practice, although forensic linguists have recently
148 begun to apply more objective selection criteria (e.g. Wright 2017). Most notably, in terms of
149 short texts, Grant (2013) attributed a series of text messages in a murder investigation
150 through a systematic analysis of the occurrence of creative spellings (see also MacLeod &
151 Grant, 2012; Silva et al. 2011). Similarly, Nini (2018) measured the similarity of short letters
152 connected to the Jack the Ripper case based on shared word sequences. Once again, like
153 the stylometric research on short texts reviewed above, these studies focus on the
154 occurrence of features as opposed to their relative frequencies.
155 Because no generally applicable method for attributing short texts exists in
156 stylometry or forensic stylistics, in this paper, we attribute the *Bixby Letter* by applying a new
157 quantitative approach to short-text authorship attribution that we call *n-gram tracing*, which

builds on recent research in both fields. Our method involves first extracting all sequences of linguistic forms (i.e. characters and words) that occur in the questioned document and then finding the possible author who uses the highest percentage of these forms. In the remainder of this paper, we describe our process of data collection, introduce and exemplify n-gram tracing through the analysis of the *Gettysburg Address*, test the method on the known writings of Abraham Lincoln and John Hay, and use the method to attribute the *Bixby Letter,* showing that the text is far more likely to have been written by Hay. Finally, we conclude this paper by considering the historical, methodological, and theoretical significance of our study.

2. Data

For years, historians believed the original *Bixby Letter* was held in the collection of Brasenose College in Oxford, but in 1925 an investigation by the *New York Times* revealed that the College had no record of ever possessing the document (Emerson, 2006). A futile search for the letter ensued, but eventually it was accepted that the original must have been lost. Some historians even speculated that the letter had been destroyed by the Widow Bixby – a woman of purportedly dubious character, who had in fact lost two as opposed to five sons in the Civil War, and who was rumoured to have been a brothel owner and a Confederate sympathiser (Burlingame, 1999). Because there is no original, different versions of the letter are in circulation today. Variation between these versions is minimal – often relating to punctuation and spacing, especially in the salutation and valediction as opposed to the body of the letter – but there are some disagreements in the main text, most notably involving ‘any word of mine’ vs. ‘any words of mine’ and ‘tendering you’ vs. ‘tendering to you’. Given these inconsistencies, it is necessary to select a specific version of the *Bixby Letter* to attribute. We chose to analyse the version printed in *Boston Evening Transcript*, because it is the first known copy of the letter and because the original is

accessible online[1] (see Table 1). In our analysis, we focused on the main body of the letter, which contains 3 paragraphs, 4 sentences, and 139 words.

To compile a corpus of Lincoln's writings, we downloaded a digitised version of Balser's 1953 *The Collected Works of Abraham Lincoln*, which is provided online by The Abraham Lincoln Association through the University of Michigan Library[2]. The collection contains over 6,500 texts, including letters, bills, notes, notices, petitions, speeches, receipts, and resolutions, dated between the 26th of May 1830 and the 14th of April 1865. The collection is divided into 8 volumes and organised chronologically, aside from Volume 1, which contains some of Lincoln's most important writings. After downloading the documents individually, we inspected each by hand, as they often contain information in addition to the main text, including dates, place names, notes, and annotations by the editors. Close reading of these annotations also revealed that a number of texts were only co-authored or signed by Lincoln. Any document for which we had any doubt that Lincoln was the primary author was therefore removed from the corpus, including the *Bixby Letter*, leaving 5,601 documents totalling approximately 650,000 words. These documents were then semi-automatically cleaned to remove text that was not part of the main body, including salutations and valedictions from letters. In addition, because Hay became Lincoln's personal secretary following his presidential nomination by the Republican Party on the 18th of May 1860, we removed all texts from that date onward as they were potentially written by Hay. The final Lincoln corpus used to attribute the *Bixby Letter* therefore only contains texts written by Lincoln up to this date, totalling 1,085 texts and 400,747 words, with texts ranging in length from 5 to 17,003 words and with a median length of 125 words. Notably, average text length rises from around 100 words in Lincoln's complete corpus to 350 words in Lincoln's early corpus because the complete corpus includes a large number of telegraphs and short letters from his time in office.

[1] http://news.google.com/newspapers?nid=sArNgO4T4MoC&dat=18641125
[2] http://quod.lib.umich.edu/l/lincoln/

To compile a corpus of Hay's writings, we downloaded a digitized version of Volume I[3] and II[4] of *The Life and Letters of John Hay,* edited by William Roscoe Thayer, which was originally published in 1915. The collection is organised chronologically, and includes letters, prose, poems, and diary entries spanning Hay's entire life. The collection does not contain a copy of the *Bixby Letter*. As opposed to the Lincoln collection, where each text could be downloaded individually, the Hay texts were grouped into chapters, interspersed with extensive commentary from the editor, as well as extracts from texts written by other authors. After downloading the chapters, we therefore carefully inspected each file by hand and manually divided the text into individual documents. Documents of unclear provenance or that were co-authored by others were excluded from the corpus. In addition, we obtained other texts written by Hay from Project Gutenberg, including short stories[5], poems[6], a 1901 novel (*The Bread Winners*)[7]*,* and a 1903 collection of essays (*Castilian Days*)[8]. We divided the two book-length texts into chapters. In total, the Hay corpus contains 577 texts totalling 261,126 words, with texts ranging in length from 9 to 8,954 words and a median of 159 words per text.

3. N-gram Tracing

In forensic linguistics, short texts are often attributed by manually selecting linguistic features from the questioned document that appear to be relatively distinctive or rare and by then searching for these forms in the writing samples of each possible author. Although this method is logical and is regularly applied in casework, there are at least three potential issues with its application. First, it is unclear how to select an exhaustive or at least an unbiased feature set, as the debate around the style of the *Bixby Letter* illustrates: different

[3] http://archive.org/stream/lifeandlettersof007751mbp/lifeandlettersof007751mbp_djvu.txt
[4] http://archive.org/stream/lifelettersofjoh02inthay/lifelettersofjoh02inthay_djvu.txt
[5] http://www.gutenberg.org/cache/epub/11392/pg11392.txt
[6] http://www.gutenberg.org/cache/epub/6062/pg6062.txt
[7] http://www.gutenberg.org/cache/epub/16321/pg16321.txt
[8] http://www.gutenberg.org/cache/epub/7470/pg7470.txt

analysts can identify different sets of seemingly distinctive features and consequently come to different attributions of the same questioned document. Second, it is unclear how to control for variation in the amount of material in the possible author writing samples, which often varies tremendously, as is the case here: if more text is available for one of the possible authors, then the forms extracted from the questioned document have an increased chance of being found in that author's sample regardless of authorship. Third, it is unclear how to judge whether differences in the use of forms in the possible author writing samples are sufficient in the aggregate to attribute the questioned document: because this approach relies on the judgment of the analyst and therefore cannot be consistently or mechanically applied, it is difficult to systematically evaluate the reliability of such methods.

Based on this general approach to forensic authorship analysis, but keeping these three limitations in mind, we have developed a new method for attributing short texts in a replicable manner that we refer to as *n-gram tracing*. The method takes the n-gram as its unit of analysis, where an n-gram is defined a sequence of one or more linguistic forms (e.g. 1-grams, 2-grams) at any level of linguistic analysis (e.g. words, characters). For example, n-grams of various types extracted from the first line of the *Bixby Letter* are presented in Table 2. The basic idea behind n-gram tracing is to calculate the percentage of n-grams that occur in a questioned document that also occur at least once in a possible author writing sample. This process is repeated for each possible author and the text is then attributed to the possible author whose writing sample contains the highest percentage of the n-grams from the questioned document.

Table 2 N-gram examples from the first sentence of the *Bixby Letter*

Level	Length	Example
Word	1	*i, have, been, shown, in, the, files, of, war, …, field, battle*
	2	*I have, have been, been shown, shown in, …, of battle*
	3	*I have been, have been shown, …, field of battle*
Character	1	*i, _, h, a, v, e, b, n, s, o, w, t, …, c, y*
	2	*i_, _h, ha, av, ve, e_, _b, be, …, ba, tl*
	3	*i_h, _ha, hav, ave, _be, bee, …, ttl, tle*

Our method is grounded in two key insights. The first is that we extract the complete set of n-grams that occur in the questioned document, so as to obtain a broad and unbiased feature set. The second is that we only consider the presence or absence of these n-grams in the questioned document and the possible author writing samples, as opposed to their relative frequencies, so as to avoid examining relative frequencies in a very short text. Instead, we measure the percentage of the n-gram types found in the questioned document that also occur at least once in equal-sized samples of texts drawn from each possible author writing sample. Specifically, for each possible author, a random sample of texts is analysed that is roughly equal in length to the total number of words in the possible author writing sample with the fewest words. The author who uses a higher percentage of the n-grams in these comparable samples – or equivalently the author that uses a larger number of unique n-grams – is then selected as the most likely author of the questioned document.

To summarise, our algorithm for conducting a basic n-gram tracing analysis for authorship attribution involves the following four steps:

1. Extract all n-grams of a particular length and level from the questioned document.
2. Take a random sample of texts of equal size from each possible author writing sample.
3. Measure the percentage of n-gram types found in the questioned document that also occur at least once in each possible author writing sample.

4. Attribute the questioned document to the possible author who uses the highest percentage of these n-grams.

In general, n-gram tracing should be run across as many different types of n-grams as possible, including both word and character-level n-grams up to a length where only a small number of n-grams are occurring in the possible author writing samples. In addition, the analysis can be repeated for different random samples of texts, allowing for the average percentages of n-grams seen to be calculated and compared.

More formally, n-gram tracing involves measuring and comparing the similarity between the set of n-grams occurring in a questioned document and the set of n-grams occurring in each possible author writing sample. Specifically, we use the *Overlap Coefficient* (Vijaymeena & Kavitha, 2016; Oakes, 2014), which measures the similarity between two sets (X, Y) by dividing size of the intersection of those two sets (i.e. the number of shared elements) by the size of the smaller set (i.e. the total number of elements):

$$\frac{|X \cap Y|}{min(|X|, |Y|)}$$

In the context of n-gram tracing, this amounts to dividing the number of linguistic features, in our case n-grams, shared by the questioned document (Q) and a possible author writing sample (A) by the number of features in the questioned document, which should always be considerably smaller than in the number of features in the possible author writing sample.

$$\frac{|Q \cap A|}{|Q|}$$

This process is then repeated for all possible authors, using comparable writing samples, and the questioned document is then attributed to the possible author with the highest Overlap Coefficient.

Although the Overlap Coefficient is rarely used in stylometry (although see Brocardo et al., 2013), the closely related *Jaccard Index*, which uses the size of the union of the two sets as the denominator as opposed to the size of the smaller set, has been applied in

numerous recent authorship studies especially by forensic linguists (e.g. Grant, 2013; Wright, 2017; Nini, 2018). We prefer the Overlap Coefficient primarily because it provides a more meaningful metric of stylistic difference, directly measuring the percentage of the features in the questioned document that also occur in the possible author writing sample. Alternatively, the Jaccard Index measures the percentage of features shared by the questioned document and the possible author writing sample, which is less interpretable, as writing samples are usually far longer than questioned documents.

The results of n-gram tracing can also be visualised by calculating the cumulative percentage of n-grams seen as texts are drawn at random from each possible author's writing sample and by plotting these percentages against the total number of words in these texts. In this way, it is possible to graph how the percentage of n-grams seen increases for each possible author as the amount of data seen increases. To ensure the results are not dependent on the random sampling of texts, this analysis can be repeated several times on many different random sequences of texts and the average cumulative percentages of n-grams seen can then be calculated and plotted at regular intervals of total words seen (e.g. up to 5,000 words, up to 10,000 words, etc.). In general, these traces will rise rapidly at first and often overlap, but as more texts are analysed, the traces will flatten out, as fewer new n-grams will be encountered (see Zipf, 1935), and a clear and consistent distinction between the authors should become apparent. In essence, the basic n-gram tracing algorithm described above involves comparing the traces for each of the possible authors at the point when the curve for the author with the smallest writing sample is exhausted; however, plotting these values across sample sizes provides additional information about the use of the set of n-grams in the possible author corpora. Most important, inspecting these graphs allows for the definitiveness of the attribution to be judged, both by comparing the degree of difference between the possible authors and the consistency of the analysis as more data is analysed.

328 Although n-gram tracing was inspired by the qualitative approach to authorship
329 analysis commonly applied in forensic linguistic casework, it also builds on recent
330 quantitative research in stylometry and forensic linguistics. The multivariate analysis of word
331 and character n-grams, as broadly defined here, is the standard approach in stylometry (e.g.
332 Kešelj et al., 2003; Grieve, 2007; Luyckx & Daelemans, 2008), but the more distinctive
333 aspect of our approach is that we only consider the presence and absence of these features
334 rather than their relative frequencies. A similar approach has been taken in a small number
335 of recent studies (e.g. Brocardo et al., 2013; Grant, 2013; Schwartz et al., 2013; Wright,
336 2017; Nini, 2018). Our method is most similar to the approach for short-text authorship
337 verification proposed in Brocado et al. (2013), which is based on the analysis of the
338 occurrence of all 3-5 alphabetic character n-grams in the questioned document using the
339 Overlap Coefficient. The main difference between these two techniques are that our method
340 is designed for attribution as opposed to verification and is based on a much larger and
341 more principled feature set, including both word and character-level n-grams. Our method is
342 also similar to the approach for authorship attribution proposed in Wright (2017), where the
343 occurrence of all 2-6 word n-grams in the questioned document and the possible author
344 writing samples are compared using the Jaccard Index (see also Johnson & Wright, 2014).
345 The main differences between these two techniques are that our method is designed
346 especially for short texts, controls for the size of the possible author writing sample, is based
347 on the Overlap Coefficient as opposed to the Jaccard Index, and is based on a much larger
348 feature space. In addition, our approach to visualisation is entirely new.
349
350 **4. Demonstration: *Gettysburg Address***
351 To illustrate how n-gram tracing works, we present an analysis of the *Gettysburg Address*,
352 which was delivered by Abraham Lincoln on the 19th of November 1863 at the site of the
353 bloodiest battle of the Civil War. We selected this text because it is one of Lincoln's most

famous texts, drafts prove it was written by Lincoln, and it is a relatively short text (272 words) that postdates May 1860, like the *Bixby Letter*. There are five final versions of the *Gettysburg Address* written in Lincoln's hand, which differ slightly from each other. In this case, we chose to analyse the *Bliss Copy*, as it is generally considered the standard – the only version signed and dated by Lincoln and the version etched into the Lincoln Memorial. We then compared the *Gettysburg Address* to the texts in our Hay and Lincoln corpora using a series of n-gram tracing analyses.

We began by extracting all 2-word n-grams from the *Gettysburg Address*, of which there are 239 distinct types when we ignore case and punctuation and prohibit n-grams from spanning sentences. For example, the first 2-word n-gram in the *Address* is 'four score', while the last is 'the earth'. We then measured the percentage of these 2-word n-grams in the complete Hay corpus (261,126 total words) and in a random sample of texts drawn from the Lincoln corpus totalling 260,954 words. We found that Hay used 55% of the n-grams, whereas Lincoln used 60% (64% of the n-grams occur in Lincoln's complete 400,747 word corpus). Because the 2-word n-gram overlap with the Lincoln corpus is greater, this analysis correctly attributes the *Gettysburg Address* to Lincoln. We also repeated the 2-word n-gram tracing analysis for Lincoln with 50 different random samples of his texts, which agreed with our first analysis, with a mean percentage of n-grams seen at 260,000 words of 60%.

To visualise the 2-word n-gram analysis, we first extracted a random sequence of texts from each possible author corpus and computed the cumulative percentage of the 239 2-word n-grams that had been seen as each additional text was added to the analysis. We then plotted these cumulative percentages of n-grams seen against the total number of words seen, as presented in Figure 1. The figure contains two traces: the longer line on top plots the percentage of the 239 n-grams seen for Lincoln, which reaches 64% at 400,000 words, while the shorter line below plots the same value for Hay, which reaches 55% at 260,000 words. Individual texts are marked with a cross. Notably, both traces are monotonic

because adding new texts can only result in new n-grams being seen. Furthermore, both traces show plateaus because at times numerous texts are added to the analyses that do not contain any new n-grams. As the basic analysis found, the trace for Lincoln is higher at the point where Hay's trace ends around 260,000 words, but the visualisation offers further support for this attribution by showing that there is a clear and consistent difference in the percentage of n-grams used by the two authors after approximately 100,000 words from each had been seen.

We also extracted 50 random sequences of texts for each author and plotted the cumulative percentage of the 239 2-word n-grams that were seen as each additional text was added to the analysis. All 100 traces are presented together in Figure 2 in the same way as Figure 1, except that marks for individual texts have been omitted for clarity. Although each trace takes a different path, Lincoln always outstrip Hay over time, confirming that the attribution does not depend substantially on the randomisation procedure. In addition to presenting 100 traces on the same graph, we reduced the 50 traces for each author to a single aggregated trace by taking the average cumulative percentage of n-grams seen across all analyses every 5,000 words. The results of this analysis are presented in the second cell of Figure 3, which shows the same overall pattern as Figures 1 and 2, with Lincoln once again clearly using a higher percentage of the 2-word n-grams in the *Gettysburg Address* than Hay.

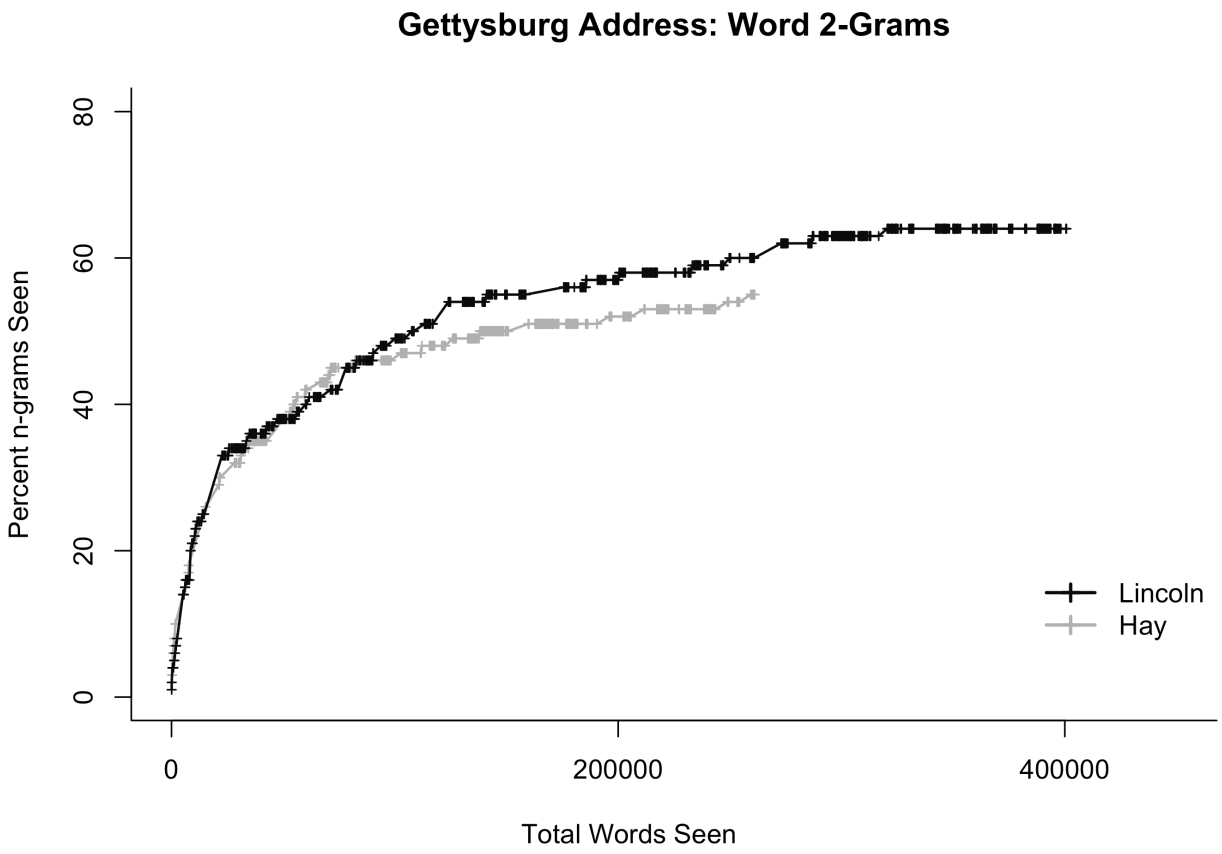
In addition to 2-word n-grams, we also analysed 1-, 3- and 4-word n-grams, based on the average percentage of n-grams seen in 50 random 260,000-word samples of texts. The analysis was only run up to 4-word n-grams because from this point onward the Hay corpus contains none of the n-grams found in the *Gettysburg Address*. The 3- and 4-word n-gram analyses also correctly attributed the *Gettysburg Address* to Lincoln: 18% of 3-grams for Lincoln vs. 14% for Hay and 2% of 4-grams for Lincoln vs. 0% for Hay. The 1-word n-gram analysis, however, incorrectly attributed the *Gettysburg Address* to Hay. Figure 3

406 presents the aggregated n-gram traces for all analyses. Notably, the 2-, 3- and 4-word n-
407 gram analyses, which correctly attributed the document to Lincoln, appear to be far more
408 definitive than the incorrect 1-word n-gram analysis.

409 Finally, we analysed 1- to 20-character n-grams, where an n-gram could be
410 composed of any case-insensitive sequence of characters, including not only letters and
411 numbers, but punctuation marks and spaces, allowing word boundaries to be preserved,
412 although once again we did not allow n-grams to span sentences. This analysis was run for
413 n-grams of up to 20 characters in length because after this point the Hay corpus contains
414 none of the n-grams found in the *Gettysburg Address*. From 3-character n-grams onward the
415 analysis correctly attributes the document to Lincoln; the 1- and 2-character n-gram
416 analyses were inconclusive as both authors use 100% of these n-grams by 260,000 words.
417 The first 15 analyses are visualised in Figure 4, showing that the attribution becomes
418 especially clear from 7-characters onward and that the 1- and 2-character analyses both
419 reach 100% of n-grams seen almost immediately.

420 N-gram tracing therefore correctly identifies Lincoln as the author of the *Gettysburg*
421 *Address*. Overall, 21 of the 24 analyses we ran attributed the document to Lincoln, while in 2
422 of the remaining 3 cases, the analysis is inconclusive. The only analysis that incorrectly
423 attributes the *Address* to Hay is based on 1-word n-grams. To assess the degree to which
424 such misattributions affect the ability of n-gram tracing to distinguish between Lincoln and
425 Hay, we conducted a systematic evaluation of the method on the known writings of these
426 two authors.

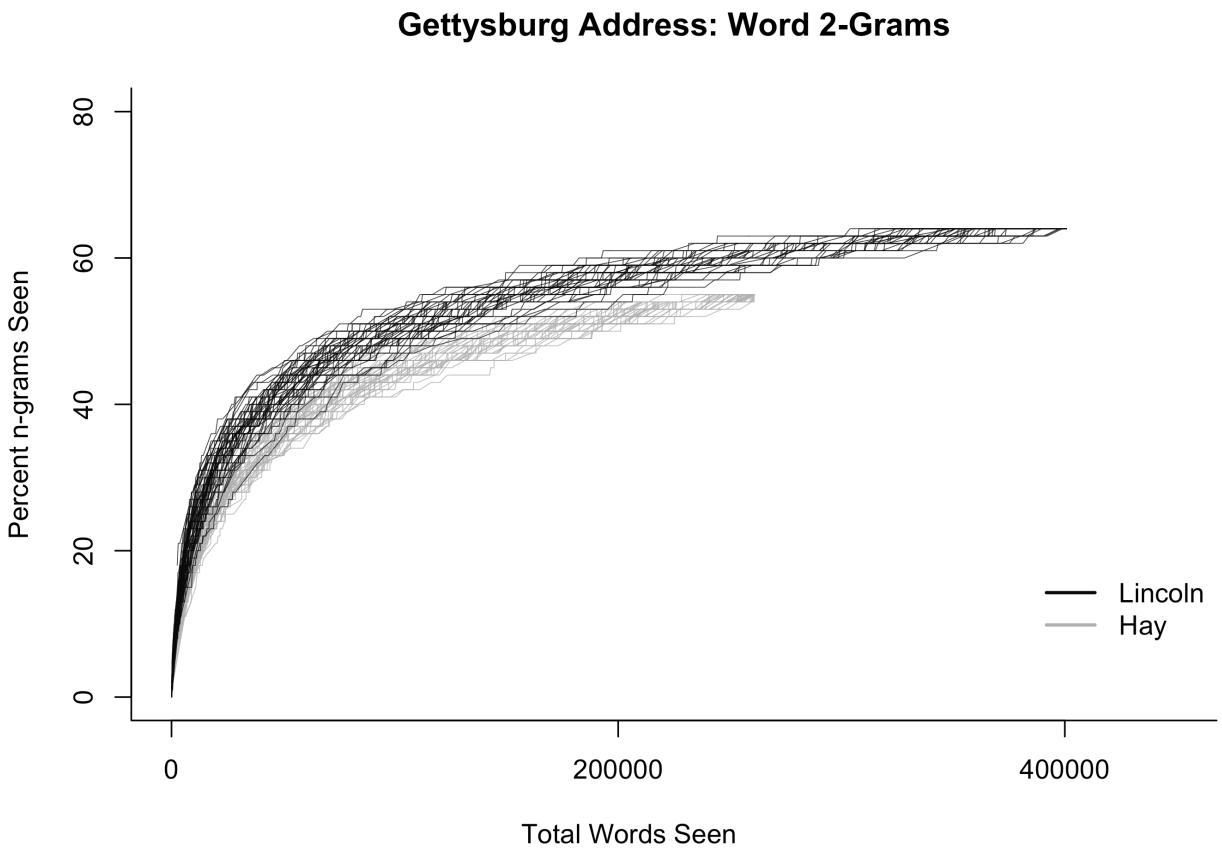
427 **Figure 1** **One *Gettysburg Address* 2-word n-gram traces**



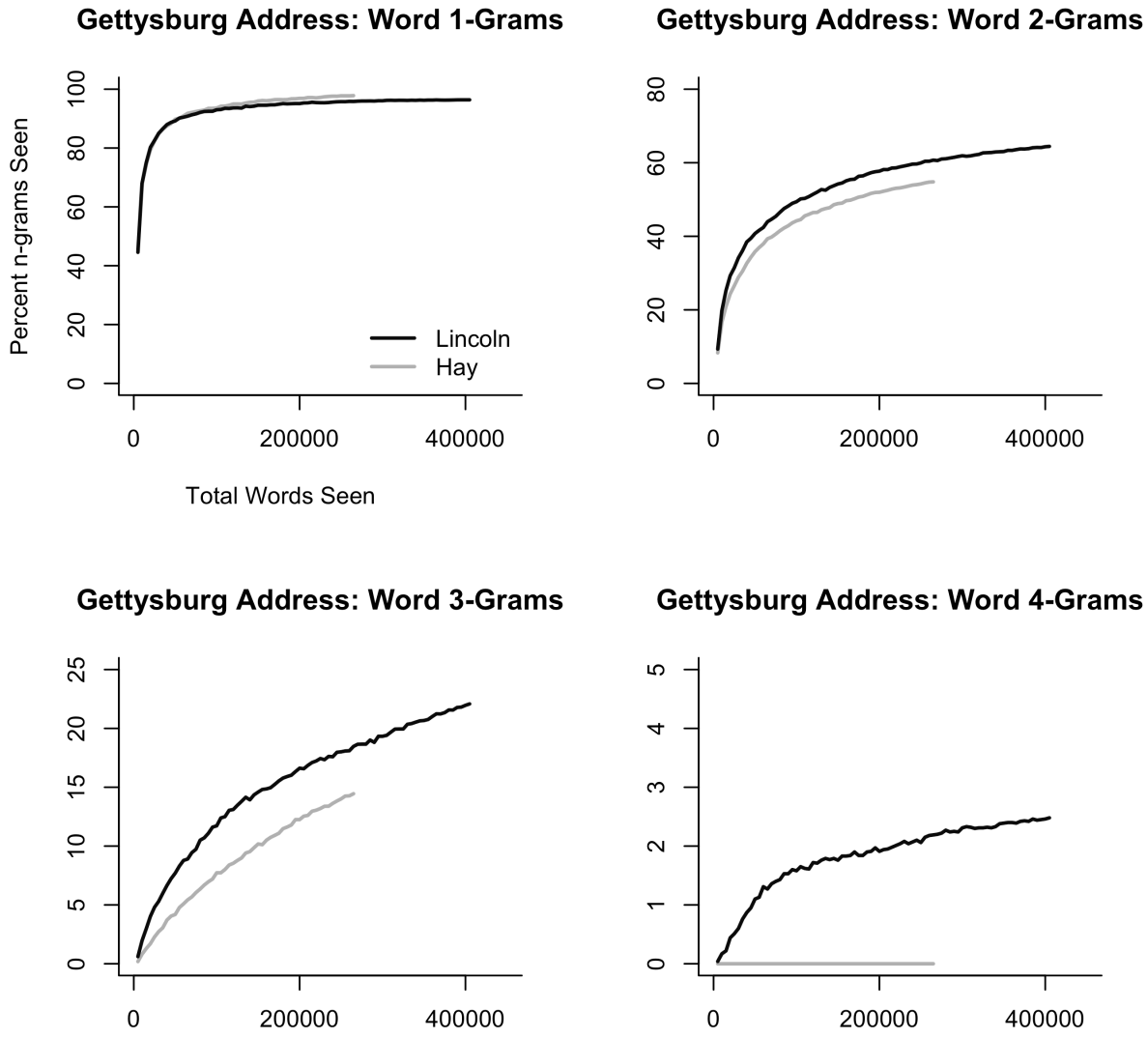
428

429

Figure 2 **Fifty *Gettysburg Address* 2-word n-gram traces**



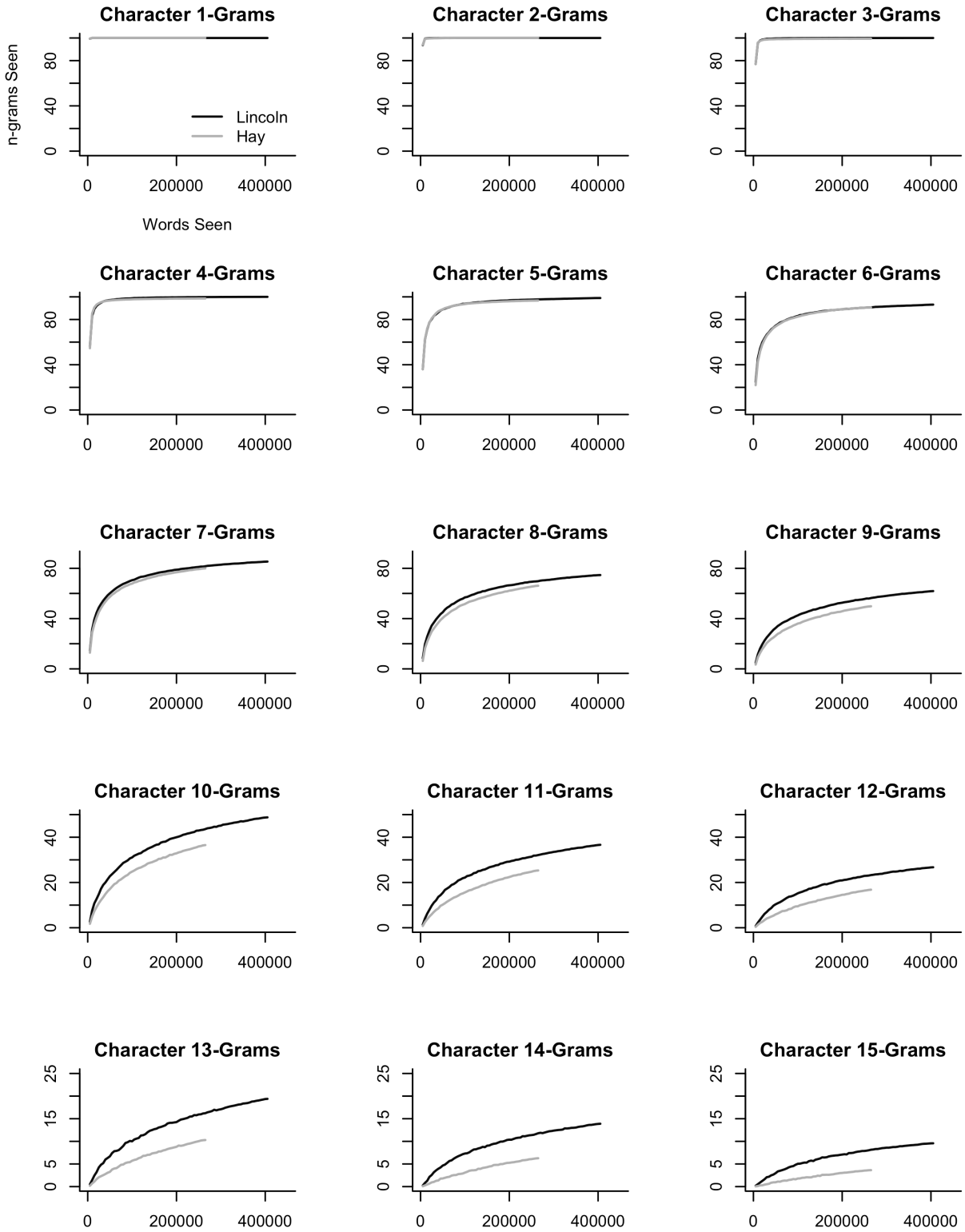
433 **Figure 3** ***Gettysburg Address*** **word-level aggregated n-gram traces**



434

435

436 **Figure 4** ***Gettysburg Address*** **character-level aggregated n-gram traces**



437

438

439

5. Evaluation

Before any method for authorship attribution can be used to resolve a case of disputed authorship, it must be shown that the method can distinguish between the writings of the possible authors under consideration with a reasonable degree of accuracy. If the method can correctly classify the known writings of those authors, then it can be used to attribute the questioned document, assuming its true author is one of the authors under consideration. This is the approach taken here: in this section, we show that n-gram tracing is capable of distinguishing between the writings of Lincoln and Hay with a very high degree of accuracy; in the next section, we use n-gram tracing to attribute the *Bixby Letter*. We do not assess or assume the general applicability of n-gram tracing. This is the subject of future research, but it is not a prerequisite for the application of a method to a specific case of disputed authorship (see Grant 2013).

To evaluate the suitability of n-gram tracing for attributing the *Bixby Letter*, we used our method to attribute each text in our corpus of possible authors following a leave-one-out approach to cross-validation (Zhang & Yang, 2015). In other words, we removed each of the 1,662 texts from our corpus one at a time (1,085 for Lincoln, 577 for Hay), and then attributed that text by comparing it to the remaining texts in the corpus using n-gram tracing. For each text, we compared 25 different n-gram types, including 1- to 5-word and 1- to 20-character n-grams, aggregating each analysis over 10 randomised sequences of texts per author, selecting the author who used the higher percentage of n-grams at 260,000 words.

We measured the accuracy of our attributions in various ways. For each n-gram type and for each author, we calculated both the *recall* (i.e. the percentage of texts written by that author that were attributed to him) and the *precision* (i.e. the percentage of texts attributed to that author that were written by him), in addition to a summary F_1 score, which is essentially an average of precision and recall. For each n-gram type, we also calculated the percentage of texts attributed correctly across the entire analysis, although this overall measure of

accuracy is imbalanced, as there are nearly twice as many Lincoln texts than Hay texts in the corpus. Across all analyses, we counted ties, where Lincoln and Hay had the same percentage of n-grams seen at 260,000 words (often 0% or 100%), as incorrect attributions for both authors. In addition, we measured the accuracy of two aggregated analyses, where we selected the author returned by the majority of a series of the best performing word- and character-level analyses.

We found tracing character-level n-grams to be an especially good way to attribute the writings of Lincoln and Hay (Table 3). Overall, all analyses based on between 5- and 10-grams achieved F_1 scores ≥ 0.95 for both authors, with the best results obtained using 7- and 8-grams. In addition, when we selected the author chosen by a majority of the analyses based on between 4- and 10-grams (i.e. the author returned by at least 4 of these 7 analyses), we correctly identified the author of all 1,662 texts. These results clearly attest to the power of n-gram tracing for distinguishing between this set of possible authors and are especially remarkable given the brevity of many of the texts, a majority of which contain fewer than 200 words and 10% of which contain no more than 50 words.

We also found tracing word-level n-grams to be good way to attribute the writings of Lincoln and Hay (Table 4), although it was not as accurate as the character-level analysis. Overall, analyses based on between 1- and 3-grams achieved F_1 scores ≥ 0.90 for both authors, with the best results obtained using 2-grams. In addition, when we selected the author chosen by a majority of the analyses based on between 1- and 3-word n-grams (i.e. the author returned by at least 2 of these 3 analyses), we achieved F_1 scores ≥ 0.95 for both authors.

Table 3 **Character n-gram Evaluation results**

	Hay			Lincoln			
n	**Rec**	**Pre**	**F_1**	**Rec**	**Pre**	**F_1**	**Acc**
1	.43	.96	.59	.12	.99	.21	.23
2	.62	.93	.74	.56	.95	.70	.58
3	.93	.86	.89	.80	.98	.88	.85
4	.98	.91	.94	.93	.99	.96	.95
5	.99	.91	.95	.94	1	.97	.96
6	.99	.93	.96	.96	.99	.97	.97
7	.97	.96	.96	.98	.98	.98	.98
8	.95	.98	.96	.99	.98	.98	.98
9	.94	.98	.96	.99	.97	.98	.97
10	.92	.99	.95	.99	.96	.97	.97
11	.91	.98	.94	.99	.95	.97	.96
12	.89	.98	.93	.99	.94	.96	.96
13	.86	.98	.92	.99	.93	.96	.94
14	.83	.97	.89	.99	.92	.95	.93
15	.79	.97	.87	.99	.90	.94	.92
16	.77	.97	.86	.98	.90	.94	.91
17	.72	.97	.83	.98	.88	.93	.89
18	.68	.95	.79	.96	.89	.92	.86
19	.63	.92	.75	.94	.88	.91	.83
20	.58	.90	.71	.92	.88	.90	.80
4-10	1	1	1	1	1	1	1

Table 4 **Word n-gram Evaluation results**

	Hay			Lincoln			
n	**Rec**	**Pre**	**F_1**	**Rec**	**Pre**	**F_1**	**Acc**
1	.96	.91	.93	.93	.98	.95	.94
2	.91	.97	.94	.99	.96	.97	.96
3	.85	.97	.91	.98	.93	.95	.93
4	.69	.94	.80	.94	.90	.92	.85
5	.41	.83	.55	.82	.89	.85	.68
1-3	.93	.98	.95	.99	.97	.98	.97

In addition to identifying the most reliable n-gram types upon which to base our attribution of the *Bixby Letter*, it is important to consider why our analyses of other n-gram types were less accurate. Analyses based on 1- and 2-character n-grams are problematic because these features are far too common in the corpus of possible authors, resulting in a large number of 100% ties, as reflected by the low recall scores for both authors. We therefore excluded 1- and 2- character n-grams from our main analysis of the *Bixby Letter*. Alternatively, analyses based on the longest word and character n-grams are problematic because these features are far too uncommon in the corpus of possible authors. For example, it is entirely possible that only one 5-word n-gram in a questioned document will reoccur anywhere in the corpus of possible authors; in such cases, the attribution will be driven entirely by this one text, potentially leading to unreliable results. We therefore restricted our main analysis of the *Bixby Letter* to n-gram types where at least 5% of the n-grams found in the letter are also found in the writings of Lincoln or Hay

We also considered how the performance of n-gram tracing was affected by text length by comparing the length of texts that were successfully and unsuccessfully attributed by each analysis using a series of Wilcoxon signed-rank tests. All n-gram tracing analyses for each author were found to be less successful on shorter texts ($p < 0.001$). For example, the median length of Hay's texts that were successfully attributed by the 7-character n-gram analysis was 160 words, whereas the median length of texts that were unsuccessfully attributed was 115 words. Similarly, the median length of Lincoln's texts that were successfully attributed was 127 words, whereas the median length of texts that were unsuccessfully attributed was 70 words. Despite these differences, n-gram tracing still attributes very short texts written by Lincoln and Hay with a very high degree of accuracy, as our evaluation has shown. For example, attributing texts containing fewer than 100 words using a 7-character n-gram analysis still achieves 0.94 recall for Hay (vs. 0.98 recall for Hay's texts that contain 100 words or more) and 0.96 recall for Lincoln (vs. 0.99 recall for

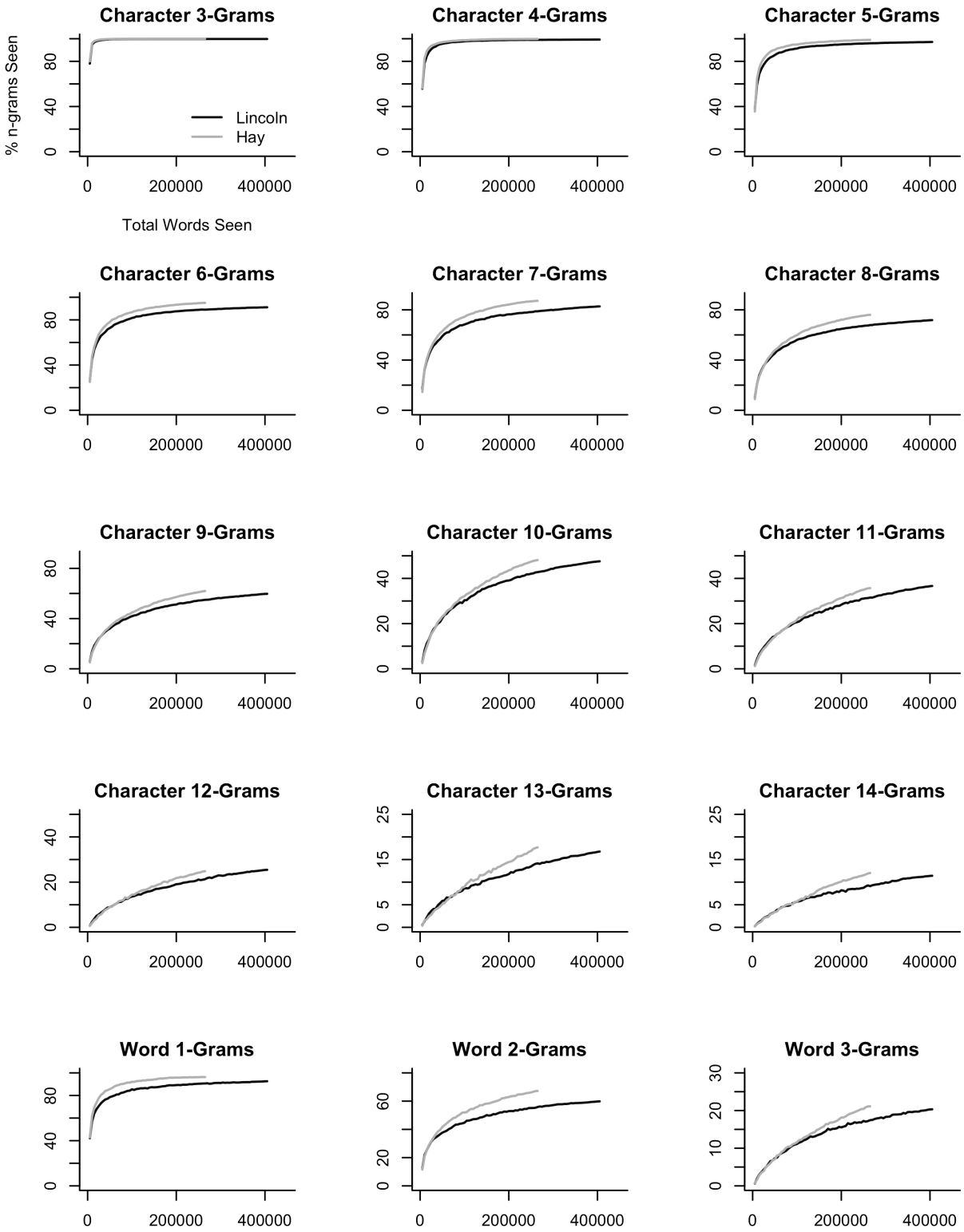
Lincoln's texts that contain 100 words or more). Furthermore, by this standard, the *Bixby Letter* is a relatively long text.

In summary, we found that n-gram tracing, based on a range of different n-gram types, is able to distinguish between the known writings of Lincoln and Hay with a very high degree of accuracy, including texts containing fewer than 100 words. We found that the analysis of 4- to 12-character n-grams and 1- to 3-word n-grams was especially useful for distinguishing between Lincoln and Hay. We also found that selecting the author chosen by the majority of the 4- to 10-character analyses attributed all 1,662 texts in our corpus of possible authors perfectly. Based on the results of our evaluation, we are therefore confident using n-gram tracing to investigate whether Lincoln or Hay is more likely to have written the *Bixby Letter*.

6. Results

To attribute the *Bixby Letter*, we used n-gram tracing to compare all 1- to 3-word n-grams and all 3- to 16-character n-grams in the *Bixby Letter* to our Lincoln and Hay writing samples based on random samples of approximately 260,000 words. Longer n-gram types were excluded from our analysis because fewer than 5% of the n-grams were found to occur in the Hay and Lincoln corpora. Overall, all 17 of these analyses identify Hay as the author of the *Bixby Letter*. Each of these n-gram tracing analyses (excluding the 15- and 16-character n-gram analyses, which are very similar to traces for the other analyses) are also visualised in Figure 5, based on 50 random sequences of texts for each author, aggregated in increments of 5,000 words. These traces show that clear and consistent differences between Hay and Lincoln are identified by 100,000 words for all word-level analyses and for all character-level analyses from 5 characters onward. The n-gram tracing analysis therefore clearly attributes the *Bixby Letter* to John Hay, providing very strong stylistic evidence against the standard attribution of the letter to Abraham Lincoln.

546 **Figure 5** ***Bixby Letter*** **aggregated n-gram traces**



547

548 Although we excluded longer character n-grams from our main attribution, n-gram
549 tracing analyses based on these additional feature sets also attribute the *Bixby Letter* to
550 Hay, as does the 4-word n-gram analysis. The 5-word n-gram analysis, however, attributes
551 the *Bixby Letter* to Lincoln. This attribution is made because 'may be found in the' is the only
552 5-word n-gram out of the 115 unique 5-word n-grams in the *Bixby Letter* that occurs
553 anywhere in our corpus of possible authors, specifically in a single speech delivered by
554 Lincoln on the 11th of January 1837 at the Illinois State Assembly:

555 *If any gentleman be entitled to stock in the Bank, which he is kept out of possession*
556 *of by others, let him assert his right in the Supreme Court, and let him or his*
557 *antagonist, whichever* may be found in the *wrong, pay the costs of suit.*

558 This example illustrates the problem that arises when tracing very rare n-gram types: the
559 entire attribution can be based on a single phrase in a single text, leading to unreliable
560 results. In light of the preponderance of evidence for Hay, this one result should not diminish
561 our confidence in the attribution, especially because the meaning of 'found' in this passage
562 is different than in the *Bixby Letter*, where it means 'discovered' as opposed to 'judged'. In
563 fact, 'may be found in' is used twice by Hay, both times with the 'discovered' meaning, once
564 in an 1863 diary entry ('After every battle Lee may be found in his tent') and once in *Castilian*
565 *Days* ('This custom, more or less modified, may be found in most cities of Europe').

566 Finally, the n-grams in the *Bixby Letter* that are only used by Lincoln or Hay are
567 presented in Table 5, of which there are notably fewer for Lincoln despite being drawn from
568 a much larger corpus. Although their discriminatory value was found to be weaker, it is more
569 instructive to consider unique word-level n-grams rather than unique character-level n-
570 grams, because word-level n-grams are less common, more distinctive, and more
571 interpretable. Thematically, Hay's unique word sequences appear more evocative and
572 emotive than Lincoln's more mundane sequences – the types of constructions one might
573 expect to find in official letters sent from the Office of the President. For example, Hay's

574 unique n-grams often reference emotion (e.g. *anguish, grief*) and religion (e.g. *altar*, *pray*),
575 whereas Lincoln's often reference governmental bureaucracy (e.g. *war department, files*).
576 Grammatically, Hay's word sequences tend to contain more forms related to the construction
577 of complex noun phrases. For example, 66% of Hay's sequences contain nouns, compared
578 to 50% for Lincoln, and 49% of Hay's sequences contain determiners, compared to 32% for
579 Lincoln. Alternatively, Lincoln's word sequences tend to contain more forms related to the
580 construction of complex verb phrases. For example, 32% of Lincoln's sequences contain
581 verbs, compared to 14% for Hay, and 18% of Lincoln's sequences contain auxiliaries,
582 compared to 9% for Hay. Furthermore, 23% of Lincoln's sequences contain pronouns, while
583 only 9% of the Hay sequences do. Overall, these patterns imply that Hay's style tends to be
584 more formal than Lincoln's (see Biber 1988). Overall, while far from definitive, this closer
585 analysis of the tone and structure of the unique n-grams used by each author helps us
586 obtain a subtler understanding of the basic differences in style detected and revealed
587 through n-gram tracing.

588 **Table 5** ***Bixby Letter*** **unique word-level n-grams**

n	Unique Hay n-grams	Unique Lincoln n-grams
1	*adjutant, altar, anguish, beguile, costly* (5)	*bereavement, tendering* (2)
2	*a loss, altar of, anguish of, any words, been shown, consolation that, feel how, grief of, have laid, I pray, pride that, sons who, thanks of, the altar, the anguish, the cherished, the consolation, the thanks, weak and* (19)	*a sacrifice, and fruitless, cannot refrain, father may, files of, mine which, shown in, the loved, war department, yours to* (10)
3	*and the solemn, but I cannot, from the grief, gloriously on the, thanks of the, the altar of, the anguish of, the consolation that, the grief of, the thanks of, you from the* (11)	*a statement of, and leave you, and lost and, cannot refrain from, I cannot refrain, of mine which, shown in the, statement of the, the files of, the war department* (10)

589

7. Conclusion

The historical significance of our attribution is clear. The *Bixby Letter* is one of the most famous and beautiful letters in the history of the United States and, despite on-going academic debate, it has generally been attributed to Abraham Lincoln, both by historians and the media. We have demonstrated, however, that the *Bixby Letter* was far more likely to have been authored by his 26-year-old assistant, John Hay. Assuming that only these two men could have written the *Bixby Letter*, our analysis shows that John Hay was almost certainly its primary author, providing strong linguistic support for the attributions made by Burlingame (1995, 1999) and other historians based primarily on external evidence.

Although we believe that our finding should finally lead to the official reattribution of this famous letter to John Hay, it could not detract from Abraham Lincoln's record, which was built upon far greater achievements than the *Bixby Letter*. Nevertheless, this short text is of considerable cultural, historical, and literary significance, and it is therefore important that we can now finally attribute the *Bixby Letter* with confidence to its true author. This study not only rights the historical record, but it should help historians better understand the inner workings of the Lincoln White House, arguably the most important presidency in the history of the United States. In addition, this result should remind us that John Hay was a great writer and a singular statesman, whose unwillingness to take credit for such a famous letter testifies to his humility and his love for Abraham Lincoln. Our attribution might even go some way to repairing the reputation of Mrs Lydia Bixby, for even if she was a Copperhead and a procuress, it is certainly better to have torn up a letter written by a secretary than by the President.

In addition to the historical significance of this study, the method introduced in this paper for attributing short texts represents a major step forward for authorship attribution. Short text attribution is considered to be one of the most important and difficult problems in stylometry, and n-gram tracing is a powerful solution to this problem. Our method has been

used here not only to attribute the *Bixby Letter,* which contains only 139 words, but over 1,600 texts of known authorship in both the Hay and Lincoln cannon, a majority of which are shorter than 200 words and some of which are as short as 5 words. Furthermore, given that n-gram tracing successfully attributed texts from various different genres without taking this information into consideration, it appears that our method may also provide a solution to the problem of cross-genre attribution, another fundamental challenge in stylometry and forensic stylistics. Testing whether or not these types of results can be replicated over other sets of possible authors is the goal of future research, in addition to testing the maximum number of authors between which the method can distinguish and the minimum amount of data needed for each. This is the main limitation of n-gram tracing: to reliably attribute short texts, the method requires access to substantial amounts of training data for each possible author, which is not always possible in historical and forensic contexts. Nevertheless, it seems clear that the method could have resolved this case of disputed authorship based on far less data, as many of the aggregated traces presented in Figure 5 and 6 diverge by 25,000 words.

More generally, the success of our method, which is rooted in forensic authorship analysis, shows how insights from forensic linguistics can inform computational research on authorship attribution. At the same time, our results should give forensic linguists pause. This study has shown that manually selecting features, especially rare features, can lead to misleading results. For example, the unique word sequences listed in Table 3 would seem to be good markers of authorship, but this list, and the number of unique n-grams used by each author, is only informative because it is exhaustive, especially as there are almost as many unique forms for Lincoln as there are for Hay. One analyst, like Nickell, might consider the word ‘tendering’, while another analyst, like Burlingame, might consider the word ‘beguile’, and each will honestly come to a different conclusion, while an analyst who considers both forms would come to no conclusion at all. When analysing authorship, it is therefore extremely important to select a representative sample of features that is truly capable of

642 distinguishing between the authors under comparison. We have essentially taken the
643 simplest solution to this problem in this paper, attributing a text by extracting all the features
644 of a particular type that occur within it.

645 Finally, our study offers evidence in support of two theories of language use, outlined
646 in Coulthard (2004), which provide a theoretical foundation for much research in authorship
647 analysis and forensic linguistics. The first is the theory of the *uniqueness of the utterance*,
648 which claims that as sequences of words (or characters) become longer, they become less
649 likely to be repeated. This claim is supported by the results of this study, which shows that
650 the likelihood that a sequence of words or characters found in the *Bixby Letter*, or any of the
651 1,662 texts over which we evaluated our method, is repeated in the possible author writing
652 samples falls as the length of these sequences increases. In particular, n-gram tracing is
653 most successful when it focuses on n-grams of middling lengths, because sequences that
654 are too short tend to be reused by all authors, while sequences that are too long tend to be
655 reused by none. Furthermore, n-gram tracing successfully distinguishes between the
656 writings of Lincoln and Hay precisely because the likelihood of repetition falls at a slower
657 rate for the true author of these texts than for the other author. The second is the theory of
658 *idiolectal co-selection*, which states that an individual's *idiolect* – their underlying system of
659 linguistic knowledge – manifests itself during language production through the unique co-
660 selection of a variety of linguistic features. In other words, although the use of a single
661 linguistic feature is unlikely to be distinctive on its own, the co-occurrence of many features
662 will generally distinguish the linguistic output of individual authors. These co-occurrence
663 patterns are exactly the information upon which n-gram tracing is based, and our
664 unambiguous attribution of the *Bixby Letter* therefore also supports this theory of idiolectal
665 co-selection.

666 Of course, a systematic analysis of the writings of many authors and many registers
667 is needed to demonstrate that the uniqueness of the utterance and idiolectal co-selection

hold across the population. These are research questions we are currently pursuing, but the results presented in this paper nevertheless offers initial empirical support for both of these claims. Furthermore, n-gram tracing provides a replicable technique for measuring the distinctiveness of linguistic forms and authorial styles. In addition to offering a solution to the short text attribution problem, n-gram tracing may therefore finally provide linguists with a way for judging the reality of the linguistic individual – a question of central theoretical importance not only to forensic linguistics and stylometry, but many other domains of linguistic inquiry.

References

Barton, W. E. (1926). *A Beautiful Blunder: The True Story of Lincoln's Letter to Mrs. Lydia A. Bixby*. Indianapolis, IN: Bobbs-Merrill.

Basler, R. P. (1953). *The Collected Works of Abraham Lincoln* (8 Volumes). New Brunswick, NJ: Rutgers University Press.

Brocardo, M. L., Traore, I., Saad, S., and Woungang, I. (2013). Authorship verification for short messages using stylometry. In *Proceedings of the 2013 International Conference on Computer, Information and Telecommunication Systems (CITS),* IEEE, Athens, pp. 1–6.

Bullard, F. L. (1946). *Abraham Lincoln and the Widow Bixby*. New Brunswick, NJ: Rutgers University Press.

Bullard, F. L. (1951). Again, the *Bixby Letter*, *Lincoln Herald*, **37**: 26–27.

Burlingame, M. (1995). New Light on the *Bixby Letter*, *Journal of the Abraham Lincoln Association*, **16**: 59–71.

Burlingame, M. (1999). The trouble with the *Bixby Letter*: The stirring Civil War document featured in *Saving Private Ryan* grew out of a lie and probably wasn't really written by Lincoln, *American Heritage*, **50**: 64-67.

694 **Burrows, J.** (2002). 'Delta': A measure of stylistic difference and a guide to likely authorship,
695 *Literary and Linguistic Computing*, **17**: 267–287.
696 **Butler, N. M.** (1940). *Across the Busy Years: Recollections and Reflections* (Volume 2).
697 New York: Charles Scribner's Sons.
698 **Chaski, C. E.** (2005). Who's at the keyboard? Authorship attribution in digital evidence
699 investigations. *International Journal of Digital Evidence*, **4**: 1–13.
700 **Coulthard, M.** (2004). Author identification, idiolect, and linguistic uniqueness, *Applied*
701 *Linguistics*, **25**: 431–447.
702 **Coulthard, M., Johnson, A., and Wright, D.** (2017). *An Introduction to Forensic Linguistics.*
703 London: Routledge.
704 **Eder, M.** (2015). Does size matter? Authorship attribution, small samples, big problem.
705 *Digital Scholarship in the Humanities*, **30**: 167–182.
706 **Ehrhardt, S**. (2007). Forensic linguistics at the German Bundeskriminalamt. In Grewendorf,
707 G. and Rathert, M. (eds), *Formal Linguistics and Law*. Berlin: Mouton de Gruyter.
708 **Emerson, J**. (2006). America's most famous letter. *American Heritage*, **57**: 1-5.
709 **Emerson, J**. (2008). New evidence from an ignored voice: Robert Todd Lincoln and the
710 authorship of *Bixby Letter*. *Lincoln Herald*, **110**: 86-116.
711 **Forsyth, R. S. and Holmes, D. I.** (1996). Feature-finding for text classification. *Literary &*
712 *Linguistic Computing*, **11**: 163-174.
713 **Gamon, M.** (2004). Linguistic correlates of style: Authorship classification with deep
714 linguistic analysis features. In *Proceedings of the 20th International Conference on*
715 *Computational Linguistics* (COLING), ACL, Geneva, Switzerland, pp. 611-617.
716 **Grant, T.** (2013). TXT 4N6: Method, consistency, and distinctiveness in the analysis of SMS
717 text messages, *Journal of Law and Policy*, **21**: 467–494.
718 **Grieve, J.** (2007). Quantitative authorship attribution: An evaluation of techniques, *Literary*
719 *and Linguistic Computing*, **22**: 251–270.

720 **Hirst, G. and Feiguina, O.** (2007) Bigrams of syntactic labels for authorship discrimination
721 of short texts. *Literary and Linguistic Computing,* **22**: 405–417.
722 **Johnson, A. and Wright, D.** (2014). Identifying Idiolect in Forensic Authorship Attribution:
723 An N-Gram Textbite Approach, *Language and Law/Linguagem E Direito*, **1**: 37–69.
724 **Juola, P.** (2006). Authorship attribution, *foundations and trends in information retrieval*, **1**:
725 233–334.
726 **Kešelj, V., Peng, F., Cercone, N., and Thomas, C.** (2003). N-gram-based author profiles
727 for authorship attribution. In *Proceedings of the Third Conference of the Pacific*
728 *Association for Computational Linguistics* (*PACLING 3*), Halifax, Canada, 255–264.
729 **Koppel, M., Schler, J., and Argamon, S.** (2009). Computational Methods in Authorship
730 Attribution. *JASIST*, **60**: 9–26.
731 **Koppel, M., Schler, J., and Argamon, S.** (2011). Authorship attribution in the wild.
732 *Language Resources and Evaluation*, **45**: 83–94.
733 **Kushner, H. I.** (1974). 'The Strong God Circumstance': The political career of John Hay,
734 *Journal of the Illinois State Historical Society,* **67**: 352–84.
735 **Kushner, H. I. and Hummel, S. A.** (1977). *John Milton Hay: The Union of Poetry and*
736 *Politics*. Boston, MA: Twayne Publishers.
737 **Layton, R., Watters, P., and Dazeley, R.** (2010). Authorship attribution for Twitter in 140
738 characters or less. In *Proceedings of the Second Cybercrime and Trustworthy*
739 *Computing Workshop (CTC),* Ballarat, Australia, pp. 1–8.
740 **Luyckx, K., and Daelemans, W.** (2008). Authorship attribution and verification with many
741 authors and limited data. In *Proceedings of the Twenty-Second International*
742 *Conference on Computational Linguistics (COLING 2008)*, ACL, Manchester, UK, pp.
743 513–520.

MacLeod, N. and Grant, T. (2012). Whose tweet?: Authorship analysis of micro-blogs and other short form messages. In *Proceedings of the International Association of Forensic Linguists' 10th Biennial Conference*, IAFL, Birmingham, UK, pp. 210–224.

McMenamin, G. R. (1993). *Forensic Stylistics*. Amsterdam: Elsevier.

McMenamin, G. R. (2002*). Forensic Linguistics: Advances in Forensic Stylistics*. Boca Raton, FL: CRC press.

Nickell, J. (1989). Lincoln's *Bixby Letter:* A study in authorship. *Lincoln Herald,* **91**: 135–140.

Nini, A. (2018). An authorship analysis of the Jack the Ripper letters. *Digital Scholarship in the Humanities, **qx065:*** 1–16.

Oakes, M. P. (2014). *Literary Detective Work on the Computer*. Amsterdam: John Benjamins Publishing Company.

Randall, J. G., and Current, R. N. (1955). *Lincoln the President*. New York: Dodd, Mead.

Schwartz, R., Tsur, O., Rappoport, A. and Koppel, M. (2013). Authorship Attribution of Micro-Messages. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), ACL, Seattle, USA, pp. 1880–1891.

Silva, R. S, Laboreiro, G., Sarmento, L., Grant, T., Oliveira, E. and Maia, B. (2011). 'twazn me!!! ;(' Automatic authorship analysis of micro-blogging messages. In Muñoz R., Montoyo A., and Métais E. (eds), *Natural Language Processing and Information Systems* (NLDB 2011). Berlin: Springer, pp. 161–168.

Stamatatos, E. (2009). A survey of modern authorship attribution methods, *Journal of the American Society for Information Science and Technology*, **60**: 538–556.

Stamatatos, E., Fakotakis, N., and Kokkinakis, G. (2001) Computer-based authorship attribution without lexical measures. *Computers and the Humanities,* **35**: 193–214.

Vijaymeena, M. K., & Kavitha, K. (2016). A survey on similarity measures in text mining, *Machine Learning and Applications: An International Journal*, ***3*:** 19–28.

Wakefield, S. D. (1948). *Abraham Lincoln and the Bixby Letter*. New York: Wakefield, S. D.

Wright, D. (2017). Using word n-grams to identify authors and idiolects. *International Journal of Corpus Linguistics,* **22**: 212–241.

Zhang, Y. and Yang, Y. (2015). Cross-validation for selecting a model selection procedure, *Journal of Econometrics*, **187**: 95–112.

Zipf, G. (1935). *The Psycho-biology of Language: An Introduction to Dynamic Philology*. Boston, MA: Houghton Mifflin.