Nini, A., Wood, C., Bowles, H. (2025). The Dickens signal: Investigating the authorship of "The Two Brothers", a ghost story transcribed from shorthand. *Textus, English Studies in Italy*, 3, 25-52.

The Dickens signal - investigating the authorship of "The Two Brothers", a ghost story transcribed from shorthand

Andrea Nini, Claire Wood and Hugo Bowles[1]

Abstract

Our investigation into what we call the "Dickens signal" brings together digital humanities (DH) and a new forensic linguistic methodology in order to verify whether Charles Dickens is the author of a dictated ghost story - "The Two Brothers" - which was transcribed from the shorthand of his pupil, Arthur Stone, as part of the Dickens Code project. We ask two questions: is there a Dickens signal in the "The Two Brothers" story and is it strong enough to identify Dickens as the author? We first set out the literary and historical background of the "The Two Brothers" text and a similar predecessor in Dickens's "To Be Read at Dusk". We then describe the forensic methodology and the process of construction and analysis of the corpora in detail. In the final part of the article, we present the quantitative results of the analysis and make the case for a Dickens signal in "The Two Brothers", taking account of literary and historical information alongside the results of the authorship analysis. In the conclusions, we discuss the interdisciplinary study of the idea of "Dickensiness" and whether technological advances in authorship verification have widened the gap between the disciplines of stylometry and stylistics.

1. Introduction

As digital humanities (DH) has continued to evolve, there has been significant debate about how to define the field (see, for example, Nyhan, Terras, and Vanhoutte 2013). Generally speaking, DH involves the use of computational tools and approaches in humanities disciplines, enabling new lines of enquiry – often dependent on large datasets – to be pursued. Our investigation into what we call the "Dickens signal" adopts just such an approach, bringing together DH and a new forensic linguistic

[1] This article was designed collectively. CW and HB wrote sections 1 and 2, AN wrote sections 3 and 4. We all contributed to section 5 and HB wrote the Conclusions.

methodology in order to verify whether Charles Dickens is the author of a dictated ghost story - "The Two Brothers" - which was transcribed from the shorthand of his pupil, Arthur Stone, as part of the Dickens Code project[2].

In traditional work on authorship, researchers in stylistics and stylometry have been comfortable working with combinations of visible *signs*, such as letters and words. In this article, we instead investigate the idea of a dynamic *signal* as an identifier of authorship. In our search for a signal for Charles Dickens, we use a new method that leverages grammatical language models, estimated on the probability of the use of function words in context. This has proved to be not just an interesting test of authorship, but also a new qualitative exploration of this new forensic methodology.

We will be asking two questions: is there a Dickens signal in the "The Two Brothers" story and is it strong enough to identify Dickens as the author? Along the way, we will also be testing the reliability of the methods we use to answer those questions. We begin by describing the background to the project (section 2), with details of the original shorthand manuscripts, the transcribed texts and their possible sources, and the dictation process that produced them. In section 3, we outline the discipline of authorship verification and what we mean by the Dickens signal, describing the verification methodology used to identify the author of the dictated "The Two Brothers" story, as well as the data and corpora involved. We present the quantitative results of the analysis in section 4 and make the case for a Dickens signal, taking account of literary and historical information alongside the grammatical data (section 5).

2 Background to the study

In this section, we describe Arthur Stone's shorthand lessons with Dickens, the modes of dictation used by Dickens, and the texts which Stone transcribed into shorthand, with their possible literary influences. The relevance of this contextual background for possible Dickens authorship of the dictated text will be examined in section 5.

2.1 Arthur Stone's shorthand notebook and texts

Stone's shorthand lessons with Charles Dickens began in November 1859 and continued for at least two months. They took place at Dickens's house in Tavistock Square or at the offices of *All the Year Round* in Wellington

[2] The Dickens Code which was AHRC-funded from 2021–23 (ref: AH/T009144/1), sets out to bring Charles Dickens's stenography to wider academic and public attention and to decipher previously untranscribed shorthand manuscripts. For details, please see https://dickenscode.org/.

We are grateful to the Free Library of Philadelphia for permission to use the images for Figures 1 (ref: cdc5890009_08) and 2 (ref: cdc5890010_20). Figure 3 is courtesy of the Special Collections department at the University of Leicester and published under Creative Commons license https://creativecommons.org/publicdomain/mark/1.0/.

Street. Arthur's notebook of the lessons, which was acquired by the Free Library of Philadelphia in the 1950's, was found to contain a short preface in his own hand as follows:

> This paper book was made up by the late Charles Dickens at the end of the year 1859 when he was kind enough to give me lessons in shorthand. A large part of it is in my writing - probably from his dictation - but the "combinations" were written by him. His handwriting will easily be recognised.

What Stone calls a "paper book" is actually a set of five booklets, whose contents are a mixture of dense shorthand script, isolated lines of corrections, notes about characters and grammatical rules in longhand, lines of repeated characters and random shorthand scribbles. The booklets are not in any chronological or logical order. It seems likely that Stone had several shorthand booklets on the go at the same time and would pick one at random for a particular lesson; he may also have used individual sheets of paper which he collected together into individual booklets.

There are two shorthand versions of the "The Two Brothers" story in the notebook - one by Stone and one by Dickens. Stone probably used Dickens's copy to check his shorthand mistakes (Bowles, 2019). Figure 1 shows the title of Dickens's version written in his longhand, with the shorthand of the first line underneath:

FIGURE 1
"The Two Brothers" - story title in Charles Dickens's longhand and first line of his shorthand

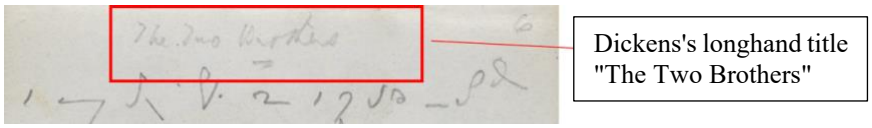
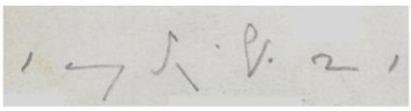
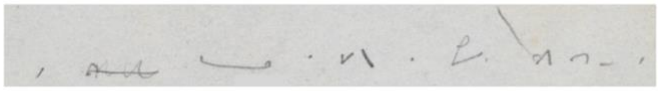


Figure 2 below shows the difference between Dickens's stenographic style and Arthur Stone's. Here it is easy to distinguish Dickens's clean, fluent shorthand (the line at the top) from Stone's more hesitant script (the line underneath).

FIGURE 2
The first line of "The Two Brothers" in Dickens's shorthand (above) and Stone's shorthand (below)



I once heard a story when I



I once heard a story when I

The quality of Stone's shorthand varies across the booklets. Sometimes it is closer to Dickens's with extremely clear markings, suggesting that Stone had been able to take his time with the shorthand and was perhaps transliterating a written text rather than taking it down from an oral dictation in real time. On other occasions, his script is very hesitant with unclear marking, suggesting that he was writing under pressure. As Figure 2 plainly shows, Stone's shorthand for "The Two Brothers" is of very poor quality, which strongly suggests that he was being dictated to and that the shorthand version of "The Two Brothers" was created via dictation. One of the questions this article seeks to answer is a) whether Dickens was just a dictating voice who was reading someone else's story aloud to Stone, i.e. an "animator" but not an "author" in Goffman's terms (1981), (b) whether he was in fact the author, either by animating his own story through oral improvisation or by reading aloud a text that he had previously written, or (c) whether he was neither of the two.

2.2 Dictation: reading aloud or improvisation?

In Stone's preface, the phrase “probably from his dictation” implies that at least some of his shorthand lessons involved Dickens dictating to him and that some of his shorthand notes will be records of this dictation. An act of dictation involves either reading aloud from a written text or speaking spontaneously (improvisation) or a mixture of the two. A strong argument in favour of the idea that Stone's lessons with Dickens involved *improvised dictation* is that Dickens was preparing Stone for parliamentary reporting; improvised speech would be a much more realistic way of practicing verbatim reporting in real time than reading aloud. Dickens also enjoyed improvisation. When he gave speeches, he tended not to use notes or to write them out beforehand. His public readings manager, George Dolby, recalled that he would improvise his speeches around a mental map of talking points arranged like the spokes of a wheel "with himself at the hub" (Dolby 2012: 274). When Dickens taught shorthand to his son Henry in the 1860's, he would improvise comic speeches which "were of the character you would expect from a street tub orator or from a speaker on the hustings or a parody of orations in the House of Commons" (Collins 1981: 164). This suggests that Dickens seems to have had a theatrical leaning in favour of improvisation.

The passages about shorthand dictation in *David Copperfield* paint an ambiguous picture. Some include descriptions of reading aloud. David's shorthand learning is assisted by his friend Traddles, who would read speeches aloud "with the assistance of Enfield's Speaker or a volume of parliamentary orations .... with his finger in the page to keep the place" (p.552) and "at a pace, and with occasional stoppages, adapted to my weakness" (p.552). However, there is also a suggestion of improvisation - Traddles would get carried away and "work himself into the most violent heats, and deliver the most withering denunciations of the profligacy and corruption of my aunt and Mr. Dick" (p.552). The ambiguity of whether, for Stone, dictation involved reading aloud from a book or improvising, or a mixture of the two, is nicely captured in Figure 3, which shows Phiz's illustration of the scene for *David Copperfield*.

FIGURE 3
Hablot K. Browne, 'Phiz', *Traddles Makes A Figure in Parliament and I Report Him* (May 1850)

On this literary and historical evidence, then, it seems likely that whenever Dickens was dictating to Stone, he would either be improvising the story or *reading and improvising at the same time*.

2.3 Identifying literary sources of the transcribed texts

Having transcribed eleven different texts in Stone's booklets (for the transcription process, see Bowles and Wood, 2024; for full transcriptions of all the texts, see Wood and Bowles, 2023), we found that Stone had taken down five of the texts almost verbatim from the philosophical works of Rev. Sydney Smith, as shown in table 1.

TABLE 1
Transcribed texts with fully identified written sources

Title in longhand	*Identified written sources*
"Sydney Smith"	Sydney Smith's *Elementary Sketches of Moral Philosophy*, Lecture IX, "On the Conduct of the Understanding"
"Sunday Night 5th February 1860"	Sydney Smith's *Elementary Sketches of Moral Philosophy*, Lecture XIV, "On the Beautiful – Part II"
None	Three essays by Sydney Smith: "Spring Guns and Man Traps"; "On the Beautiful"; "I Would Not Live Always"

Sydney Smith, an early 19th century philosopher, was one of Dickens's favourite writers. The two men corresponded and Smith's *Elementary Sketches of Moral Philosophy* was a book that Dickens always kept with him and recommended to his guests at Gad's Hill Place (Cobbett Barker 1937; Drew 2003: 165). So, it is highly likely that Dickens would have dictated passages from the book during his lessons with Stone or given the book to him to use on his own for transcription practice.

Five of the shorthand texts that we transcribed were critiques and descriptions with no identifiable written source. However, two of them show literary influences, which are set out below in table 2:

TABLE 2
Literary influences of five transcribed critiques and descriptions

Longhand title	*Contents*	*Literary influences*
"Nelson"	An alternative account of Nelson's final hours at the Battle of Trafalgar	Robert Southey's *Life of Nelson*
"Travelling"	A philosophical meditation on the benefits of travel to a receptive mind	None identified
"Didactic"	A critique of Catholicism	Charles Dickens's *Pictures from Italy*
None	A critique of hereditary privilege	None identified

Can we detect a Dickensian influence in these texts? In "Nelson", what the text describes as a "picture of the events of the day" resembles several aspects of Robert Southey's *Life of Nelson* and adopts similar phrasing, without being a word for word copy. The text contains an aside about "any really or pretendedly choice rum offered in the market" being called "The Admiral". This is a reference to the preservation of Nelson's corpse in a

barrel of spirits on the way back to England and is in keeping with Dickens's sometimes macabre sense of humour (see Carey 1973). Other texts reflect sentiments that Dickens expressed elsewhere, such as "Didactic", which resembles passages from *Pictures from Italy*. The exercises vary in tone, content, and style, but often end in sententiae, such as the reflection in "Travelling" about that "kind of man who, when he sees a remarkable thing, thinks not of the thing itself but of what he shall/should say of it", which may suggest oral dictation.

2.4 The dictated story of "The Two Brothers" and the tale of two brothers in *To Be Read at Dusk*

The final two shorthand texts that we transcribed are two short ghost stories, whose contents and literary influences are shown in table 3 below:

TABLE 3
Two transcribed ghost stories: contents and literary influences

Longhand title	*Contents*	*Literary influences*
"The Two Brothers"	A ghost story about two brothers, in which the apparition of one appears to the other shortly before death	Charles Dickens's *To Be Read at Dusk*
"Anecdote"	A ghost story about a student prank and a phantom in green	None identified

Our study focuses on the first of these dictated stories - "The Two Brothers". It has a much stronger literary connection with Dickens than the critiques and descriptions and a more complex publishing history. The plot of the "The Two Brothers" story is contained in a longer story by Dickens called *To Be Read at Dusk* (henceforth *Dusk* which was first published in a literary annual called *The Keepsake* in 1852. In *Dusk*, the ghost story of two brothers is the second of two tales exchanged by couriers gathered outside the convent of the Great St Bernard in Switzerland.

The full texts of the *Dusk* tale that was published in *The Keepsake* in 1852 and the one dictated to Stone in 1859 are shown in the Appendix. The contours of the two stories are the same: a dying man appears to his brother as an apparition, with this supernatural sighting confirmed on attendance at the deathbed, when the dying brother announces "you have seen me before, to-night". However, there are also a number of differences. In the dictated 1859 version, the frame story is different: instead of five couriers, an unidentified storyteller describes hearing a story "from the mouth of a deceased judge" who had "shared of the gains attending a will made by one of the wards". There are also differences in some of the specific details: the location of the two brothers changes, from two London/Essex based

locations in 1852 (Mr James lives in Poland Street; Mr John resides in Epping Forest), to Slough and London in the 1859 version. In 1852, the brothers are named twins; in 1859, the brothers are 'old bachelor[s]' and only one is named, with the name of the surviving brother reversed in the two versions. In contrast with the 1852 story, told by a German courier called Wilhelm who is directly involved in the events and overheard by the frame tale's first-person narrator, the narrator of the 1859 story relates the tale, but does so second-hand.

We now turn to the question of authorship and whether the dictated 1859 version of "The Two Brothers" can be attributed to Dickens.

3. Authorship verification of the dictated "The Two Brothers" story

3.1 Stylometry and comparative authorship analysis

The discipline of *comparative authorship analysis*, or just *authorship analysis*, involves analysing a set of texts to answer questions about authorship by using linguistic analysis (Grant 2022). The term *stylometry* is often used as a synonym, even though it generally refers to the quantitative analysis of style. In the case of the "The Two Brothers" story, we are dealing with *authorship verification* - a specific case in which only one candidate author (Dickens) is available and the authorship problem is to determine the likelihood that a questioned text or set of texts was or was not authored by this candidate (Koppel et al. 2012; Halvani 2021).

3.2 What is an authorial signal?

The most reliable methods for authorship verification are based either on word frequency, such as *Delta* (Burrows 2002; Ishihara 2021), or on the frequency of sequences of 4/5 characters, called *n-grams*. The best-known method using this type of feature is the *Impostors Method* (Koppel and Winter 2014; Potha and Stamatatos 2020), which relies on building a reference corpus of similar texts, called *impostors,* in order to determine whether the linguistic similarities between the questioned text and the candidate is significantly greater than chance when compared with the impostors.

These methods exploit the fact that individuals tend to use different *function words* at different rates. Words in every language can be classified as *function* or *content* words. The former is a closed set of words that just have a grammatical function, while the latter is an open set of words that convey meaning. Typically, function words in a language are words like articles, pronouns or prepositions, while content words are nouns, verbs, adjectives and adverbs. However, there is no hard boundary between the two categories, with some very common verbs such as *go, make* or adverbs like *really* or *very* often having properties of both function and content words; indeed the function words of a language tend to develop from content words of this kind, losing their meaning over time.

Authorship analysis and stylometry, ever since the pioneering work of Mosteller and Wallace (1963) and later by Burrows (2002), have long demonstrated that there is a significant amount of information about authorship in function words. Although content words may also contain authorship information, we do not know enough about their discriminatory potential, whereas we do know that function words by themselves are excellent authorship discriminators. The reason for the effectiveness of function words as indicators of authorship has been a matter of debate (Kestemont 2014). Nini (2023) has recently proposed that function words distinguish authors because the frequency of their usage approximates to the author's grammar. Cognitive linguistic theory (Croft and Cruse 2004; Divjak and Dąbrowska 2019) predicts that a mental grammar is unique to a specific author at a certain point in time. If this is correct, then an author's *signal* is constituted by the idiosyncratic co-selection of grammatical constructions that make up the author's unique grammar. A signal can thus be regarded as an author's *statistical linguistic identity* embedded in the text that they authored.

3.3 Author verification methodology - *LambdaG*

In an attempt to measure the authorial signal more directly, Nini et al. (2025) have introduced a new method for authorship verification called *LambdaG*. This method is based on first constructing a grammar of the candidate author and then a set of grammars from a reference population of different authors. In line with Cognitive Linguistic theories, these grammars are constructed by estimating the probability of the usage of a function word in its context, where the context is the set of words *preceding* the function word; this is a similar mechanism to the one used in modern Large Language Models, such as the one powering *ChatGPT*.

The method consists in calculating an overall score that quantifies the authorship evidence, called λ_G , which is positive if it is more likely that the text was produced by the grammar of the candidate author and negative if it is more likely to have been produced by the grammars of the reference population. The size of the λ_G score expresses the strength of this likelihood.

Nini et al. (n.d.) have found that *LambdaG* outperforms not only the *Impostors Method* but also more sophisticated methods for authorship variation based on Large Language Models or neural networks. The advantage of *LambdaG* is also that this method is more amenable to interpretation because an λ_G score can be calculated for *every single word* in the questioned text, thus allowing the creation of text heatmaps that can guide the analyst towards the parts of the text that support one or the other conclusion. *LambdaG* is therefore able to analyse the unique linguistic signal of an author such as Dickens and to do so more transparently.

3.4 Corpus construction and data analysis

This section outlines the construction of the corpora to which *LambdaG* was applied. For reasons of space, it omits the most technical details about the mechanisms of *LambdaG*, which can be found in Nini et al. (n.d.).

Four corpora were constructed in all. Two were corpora of novels - one of Dickens's novels and one of twenty 19th century authors' novels that were comparable to Dickens's. The other two corpora were of short ghost stories - one of Dickens's stories and the other of 19th century authors' ghost stories. The full lists of authors and stories are shown in the Appendix. Table 4 below offers a breakdown of the number of texts and word count of the corpora by genre and authorship:

TABLE 4
Summary statistics of the four corpora by genre and authorship

Genre	*Authorship*	*N. texts*	*N. words*
Novel	Dickens	12	3,503,401
Novel	Non-Dickens	22	4,622,377
Ghost story	Dickens	10	47,345
Ghost story	Non-Dickens	10	63,379

The four corpora and the two texts to be analysed - the dictated "The Two Brothers" story (henceforth *dictated TB*) and the original tale of two brothers in *To Be Read at Dusk* (henceforth *Dusk TB*) - were then prepared for application of *LambdaG*.

First, all punctuation was removed from the corpus because decisions on punctuation may have been editorial not authorial and because the dictated shorthand did not mark punctuation. Texts were processed using an algorithm called *POSnoise* (Halvani and Graner 2021), which replaces content words with their grammatical tags (Noun, Verb, Adjective, Adverb, and Proper Noun) while leaving function words in place. For example, the first nine words of dictated TB would be tagged as follows

Word	*I*	*once*	*heard*	*a*	*story*	*when*	*I*	*was*	*living*
Tag	i	once	heard	a	N	when	i	was	living

Before the analysis of the questioned texts could take place, two steps are necessary: the creation of a *calibration* dataset and *validation*. The λ_G scores are uncalibrated and do not reflect the real weight of the evidence unless they are calibrated. The calibration dataset is used to turn these λ_G scores into calibrated λ_G scores, which we call Λ_G. The validation phase is simply a test that the method works in these specific circumstances.

First, equally sized samples of the dataset were created so that each sample contained the same number of words as in the analysed texts (*dictated TB* or *Dusk TB*). The *calibration dataset* was created using these samples. To build this calibration dataset, fifty random samples of Dickens novels and fifty random samples of non-Dickens novels of the same size

as the analysed text were selected as 'fake' questioned samples and λ_G was then calculated using a 1-million-word random sample of remaining Dickens novels as the known data and all the remaining non-Dickens novels as the reference corpus.

For the *validation* phase, the method was tested on "fake" questioned samples from the ghost stories. The λ_G score for each ghost story sample was calculated using a random 1 million words of Dickens novels as known data and the full set of non-Dickens novels as reference data. These λ_G scores were then calibrated into Λ_G using the calibration dataset. The purpose of this phase was both to evaluate how well the method generalises to unseen texts and how well the reference data, which is composed of novels, can be used for ghost stories.

3.5 Scoring and analysing the *Dusk TB* and *dictated TB* texts

After calibration, the Λ_G for the real *Dusk TB* and *dictated TB* texts was calculated using the same procedure as the validation phase. Because of the randomness in the sampling, each analysis was repeated 20 times and any of the results below reports average values and 95% confidence intervals, which give an estimation of uncertainty.

The analysis produces a score for *each word of the analysed texts*, as shown below for each of the words in our example line:

Word	*I*	*once*	*heard*	*a*	*story*	*when*	*I*	*was*	*living*
Tag	I	once	heard	a	N	when	I	was	living
λ_G	≈ 0	-1.4	1.7	≈ 0	≈ 0	≈ 0	≈ 0	0.2	-1

The performance of the method was evaluated using a metric called C_{llr}, which is typically used in forensic science to evaluate the performance of a method based on a log-likelihood ratio (Ramos et al. 2013; Ishihara 2021). The C_{llr} assesses the confidence with which the prediction is being made.

4. Results

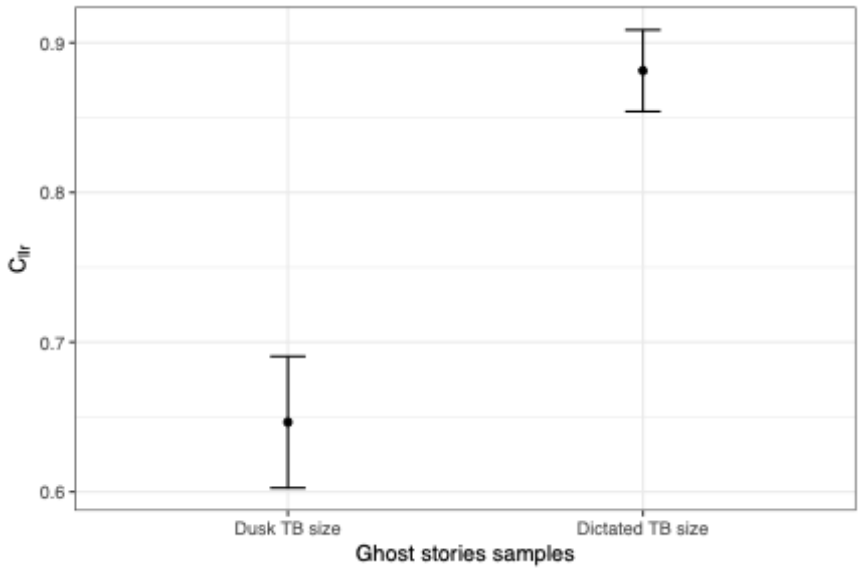
In this section we look at the results of the statistical analysis, viewed as quantitative graphs (4.1) and more qualitative shades of "Dickensiness" (4.2).

4.1 Statistical results

Validation analysis - sample corpora of fake ghost stories

The results of the validation analysis are shown in Figure 4 below. The figure shows two bars with dots in the middle. The dot indicates the mean C_{llr} score and the bar represents the 95% confidence interval. The higher the score for each fake story sample, the less reliable the method is for samples of the size of the corresponding analysed text.

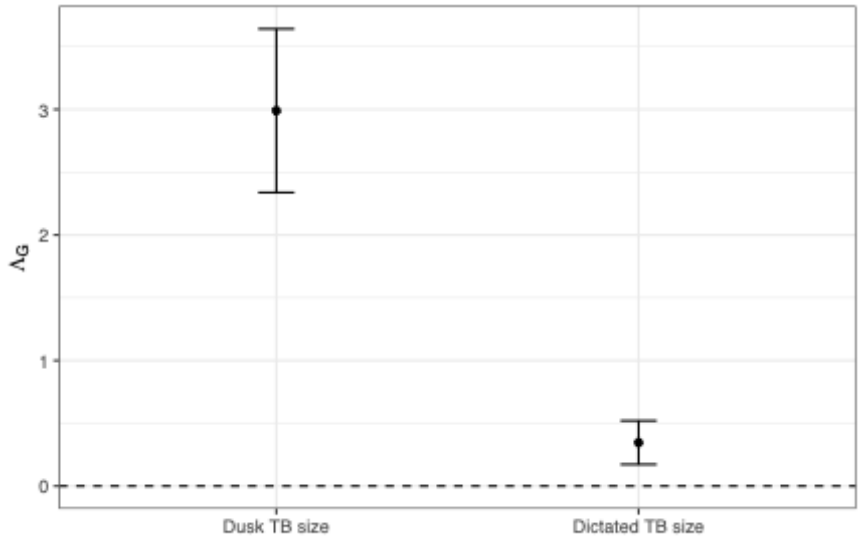
FIGURE 4
Results of validation analysis - mean C_{llr} and 95% confidence intervals



For the story samples of the size of *Dusk TB* (1275 words), the C_{llr} score is between 0.7 and 0.6. This shows that a sample of the size of *Dusk TB* poses no problems for a method such as *LambdaG*. For shorter samples of the size of *dictated TB* (374 words), however, the mean C_{llr} is 0.88. The comparatively poor performance of this sample is due to the shorter length of *dictated TB*. The method is more tentative with a sample as short as 374 words.

Λ_G analysis
The results for the analysis of the real *Dusk TB* and *dictated TB* texts are shown in Figure 5. The horizontal dashed line at the bottom of the graph marks the cut-off point. Anything above the dashed line, i.e. a score of 1, 2 or 3 on the vertical λG axis, indicates support for Dickens as the author of the story. Anything below the dashed line, i.e. less than zero, indicates support for not-Dickens. A score of 0 or near zero indicates that the result is inconclusive.

FIGURE 5
Values of Λ_G for *Dusk TB* and *dictated TB*



The results show that scoring for both texts is above the dashed zero line, i.e. the scores identify Dickens as the author for both texts. However, there is a marked difference in the strength of that identification. *Dusk TB*, which we know has been authored by Dickens, has a strong mean Λ_G score of nearly 3. In contrast, the *dictated TB* score is only 0.3. This means that the analysis supports a Dickens authorship for *dictated TB*, but only weakly.

4.2 Shades of "Dickensiness" in *Dusk TB* and *dictated TB*

For a more qualitative analysis of the texts, the statistical data which produced the graphs in Figure 2 can be highlighted in terms of relative "Dickensiness" and "not-Dickensisess". Returning to our example line, the first words of the *dictated TB* story would be highlighted as follows:

Word	I	once	heard	a	story	when	I	was	living
Tag	I	once	heard	a	N	when	I	was	living
λ_G	≈ 0	-1.4	1.7	≈ 0	≈ 0	≈ 0	≈ 0	0.2	-1
Tag	I	*once*	**heard**	a	N	when	I	was	*living*

The bold type of "heard" matches a strong positive score, i.e. "Dickensy", while the italicization of "once" and "living" matches a negative score, i.e. "not-Dickensy".

One can also shade these words to mark their degree of Dickensiness or not-Dickensiness. Below we have shaded the first three lines of *dictated TB* - the tags N, V, J, B, P indicate, respectively, Noun, Verb, Adjective, Adverb, and Proper Noun.

i *once* **heard** a N when I was *living* *in P* from the N of a J N which V my N very B he

told it by the V N of a N N with a J N V **outside** I ought to V that he had been N *in* an N that **afterwards** V of the N V the will made by one of the N this is the N there were two J N N of *whom* **one** *at P* and the other at P they were in the N of V each other for a

The different degrees of shading show different degrees of Dickensiness or not-Dickensiness. The dark grey shading of the word "one" is strongly Dickensy, the medium grey of "heard" and "outside" are medium-Dickensy and the pale grey "afterwards" is mildly Dickensy, while, for the not-Dickensy words, "once", "living" and "in" are medium not-Dickensy and "in P", "whom" and "at P" are mildly not-Dickensy.

One can use the shading to look for words of potential stylistic interest. In carrying out this search, two things should be remembered. First, the degrees of shading are impressionistic and do not capture the comprehensiveness or accuracy of the scoring; every single token in the text is given a score so even unshaded words have a Dickensiness rating and contributes to the overall λ_G value assigned to the text. Secondly, the scoring and colouring of a word is the consequence of an analysis of that word *in its immediately preceding context*, which for this analysis was set to *up to five preceding words*. So, a shaded word would be one that is more or less likely to be used by Dickens in that context compared to the reference population.

As an example of the kind of results that might be interesting from a qualitative perspective, let us look at the words "seen" and "myself", which were both scored as "strongly Dickensy", in their preceding verbal context:

Preceding verbal context				shaded word
I	have	just	now	**seen**
ADJECTIVE	NOUN	out	of	**myself**

The strongly Dickensy use of *seen* in the context of the phrase *just now seen* is confirmed to be a usage that is exclusive to Dickens, who uses it once each in *Little Dorrit (LD)*, *The Mystery of Edwin Drood (MED)* and *Barnaby Rudge (BR)*; in *LD* and *MED*, the matching phrase is even longer: *I have just now seen*. The *out of myself* pattern is used four times by Dickens in *The Old Curiosity Shop (OCS)*, *David Copperfield (DC)*, *Great Expectations* (*GE*) and *Martin Chuzzlewit (MC)* and only twice by Victorian authors, Gaskell and Eliot. In *MC* and *GE*, Dickens uses the grammatical pattern adjective + noun + *out of myself*, which matches the token in *Dusk TB* even more closely.

However, although the data scores are able to point to Dickensy or non-Dickensy grammatical patterns, they do not seem to provide us with stylistic information that relates to Dickens as a writer. Sequences which derive their Dickensiness or not-Dickensiness from function words, like "out of myself", are so unusual that they are hard to comment on stylistically. Stylistics prefers to look at the patterning of words using content words as a starting point. This underscores the point made in section 3 that the term "stylometry" may be misleading because it actually

says very little about literary style and is perhaps best referred to as "author verification".

5. Is Dickens the author of the dictated "The Two Brothers" story?

The forensic linguistic results in section 3, like any forensic evidence, should be interpreted within its context and considering material, literary and historical evidence.

First, Dickens's connection to the dictated "The Two Brothers" story is clear because, as section 2 has shown, he wrote his own version of "The Two Brothers" in shorthand and Stone took possession of it in 1859. Although Dickens is clearly involved in the lesson, the mere existence of a shorthand version by Dickens does not necessarily imply that Dickens authored "The Two Brothers", or that he dictated it. However, it is unlikely that a third party would have read out the story to Dickens and Stone together because Dickens is not recorded as having a copy of *The Keepsake* in his library[3]. Although we cannot rule out there being a manuscript version among his papers, it seems unlikely that Dickens or a third party would have had the original story to hand.

Secondly, the framing of "The Two Brothers" as a story heard from a "deceased judge" who "had been counsel in an action that afterwards shared of the gains attending a will made by one of the wards" is extremely curious. It is distinctly possible that the frame includes the words "judge", "counsel", "action", "will", and "wards" for a pedagogical reason, namely that it is being improvised to give Arthur extra practice in writing legal terminology in shorthand. Dickens, having worked as a legal clerk in his youth and as a parliamentary reporter until he was 24 years of age, knew the terminology well and enjoyed comic improvisation when giving shorthand lessons. The random connections between the legal words in the frame, as well as the absurdity of hearing a story from a deceased judge, resemble the mock parliamentary speeches that were typical of Dickens's shorthand practice with his son Henry in the 1860's (see section 2).

Thirdly, there is the thorny issue of the memorability of the original *Dusk* tale over time. We should remember that the tale of the two brothers is the second, and much shorter, of the two tales in *Dusk*. There is no historical evidence from the publishing history of *Dusk* that this short, secondary tale has had any particular public or critical response. Given this lack of resonance, would a third party (i.e. not Dickens) have been able to remember the plot of this obscure tale seven years after its publication? Moreover, the phrases "passed on to the end of the room" and "you have seen me before tonight and you know it" are repeated word for word in both stories. If a third party (i.e. not Dickens) was dictating the story, would he or she have been able to have exact recall of those phrases seven years after reading them? The historical evidence is that Dickens would certainly have been able to do so. In his 1867 Pocket Diary, he scribbled some lines

[3] See Dickens Library Online <URL>.

in shorthand quoting from his story *The Holly Tree*, which he had published twelve years previously. The shorthand shows a great deal of verbatim recall of the original text of *The Holly Tree*, even though he is copying from memory twelve years later (see Bowles, 2018).

On the other hand, if Dickens was dictating, although he would certainly have been able to recall the original story plot and some of its phrasing, it is perhaps less clear why he would do so. Ghosts were likely in Dickens's mind at the time of the lessons, given that at the end of 1859 Dickens was working on *The Haunted House* as the extra Christmas number of *All the Year Round*. However, according to Ruth Glancy (1987: 40), *Dusk* was pirated by *Harper*'s in the United States on its release, but thereafter not reprinted until 1882 so, with no copy in his library, why this particular story of brotherly haunting should recur to Dickens for this particular dictation is a matter for debate.

6. Conclusions

Dickens's famous ghost story *The Signalman*, published in 1866, is the story of a railway signalman, haunted by a ghostly figure whose appearance always precedes a tragic accident, including ultimately the signalman's own death. An important theme of the story is the unsettling effect of new, machine-based signalling technology. The signalman is anxious about the uncertainty of his mechanical role and fearful of its consequences. Our interdisciplinary digital exploration of Dickens's authorial identity by technological means, via an obscure writing system (19th century shorthand) and an unusual oral medium (dictation), speaks to a similarly uncertain relationship.

We have found in this study that our dictated "The Two Brothers" story has a weak Dickens signal which, when viewed in its literary and historical context, is perhaps not as weak as it appears. The variable and dynamic nature of the signal has proved to be an advantage from our interdisciplinary perspective because our perception of its strength can be modified by literary and historical considerations. However, the grammatical underpinnings of the signal present a challenge to our ideas about the individuality of an author. Is "Dickensiness" just a configuration of Dickens's individual grammar at a particular moment? Dickens famously claimed that he was "inimitable", but if an individual's lexical choices are guided by that individual's ever-changing cognitive grammar, then are we not all as inimitable as each other? Like Dickens's signalman's reaction to his new machinery, when we find a new technology telling us that an author as stylistically interesting as Dickens can be identified entirely by particular grammatical patterns, it is disquieting.

Perhaps, though, it does not have to be quite so unsettling. The technological progress epitomised in Nini et al's new methodology for author verification highlights the ever-widening chasm between what was once called stylometry and the domain of stylistics. Yet while the increasing depth of technological analysis widens the rift between the

identity of an author and our ability to describe his or her style, and does so in an often bewildering way, it also clarifies the respective roles of the "author verifier" and the stylistician. Just as a fingerprint can identify, but not describe or define, who we are, we can clarify the identity of an author like Dickens without necessarily having to describe what it is that makes him one.

REFERENCES

BOWLES, HUGO and WOOD, CLAIRE, 2024, "Decoding Dickens: Social Stenography and the 'Tavistock' letter", in H. Boedekker and K. Minot McCay (eds), *New Approaches to Shorthand: Studies of a Writing Technology*, De Gruyter, Berlin, pp. 243-273. Available at https://doi.org/10.1515/9783111382692-009

BOWLES, HUGO, 2018, "Dickens's shorthand manuscripts", *Dickens Quarterly*. *35*(1), pp. 5-24. Available at https://dx.doi.org/10.1353/dqt.2018.0001

BOWLES, HUGO, 2019, *Dickens and the stenographic mind*, O.U.P., Oxford.

BURROWS, JOHN, 2002, "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship", *Literary and Linguistic Computing* 17 (3), pp. 267–87.

CAREY, JOHN, 1973, *The Violent Effigy: A Study of Dickens' Imagination*, Faber and Faber, London.

COBBETT BARKER, W., 1937, "Letters of Sydney Smith to Charles Dickens", *Dickensian*, vol. 33, issue 242, pp. 91–92.

COLLINS, PHILIP, 1981, *Dickens: Interviews and Recollections* vol. 1, Macmillan, London.

CROFT, WILLIAM and CRUSE, ALAN, 2004, *Cognitive Linguistics*, Cambridge University Press, Cambridge.

DICKENS, CHARLES, [1850] 2004, *David Copperfield*, edited by Jeremy Tambling, Oxford University Press, Oxford.

DICKENS, CHARLES, [1852] 2016, *To Be Read At Dusk*, in D. A. Thomas (ed.), *Selected Short Fiction*, Penguin, London, pp. 66-77.

[DICKENS, CHARLES], 1855, *The Holly Tree Inn*. Household Words, December 12.

DICKENS, CHARLES, [1866] 1985, *No.1 Branch Line. The Signalman*, in D. A. Thomas (ed.), *Selected Short Fiction*, Penguin, London, pp. 78-90.

DIVJAK, DAGMAR and DABROWSKA, EWA, 2019, *Cognitive Linguistics - Foundations of Language*, De Gruyter Mouton, Berlin-Boston.

DOLBY, GEORGE, [1885] 2012, *Charles Dickens as I Knew Him: The Story of the Reading Tours in Great Britain and America*, C.U.P., Cambridge.

DREW, JOHN, 2003, *Dickens the Journalist*, Palgrave Macmillan, Basingstoke.

GLANCY, RUTH, 1987, "To Be Read at Dusk", *Dickensian*, vol. 83, issue 411, pp. 40–47.

GOFFMAN, ERVING, 1981, *Forms of Talk*, University of Pennsylvania Press, Philadelphia.

GRANT, TIM, 2022, *The Idea of Progress in Forensic Authorship Analysis*, C.U.P., Cambridge. Available at https://www.cambridge.org/core/elements/idea-of-progress-in-forensic-authorship-analysis/6A4F7668B4831CCD7DBF74DECA3EBA06

HALVANI, OREN and GRANER, LUKAS, "POSNoise: An effective countermeasure against topic biases in authorship analysis", in *Proceedings of the 16th International Conference on Availability, Reliability and Security*, ACM Digital Library, article 47, 2021, pp. 1-12. Available at https://doi.org/10.1145/3465481.3470050

HALVANI, OREN, 2021, "Practice-Oriented Authorship Verification", PhD dissertation, Technische Universität Darmstadt, Darmstadt.

ISHIHARA, SHUNICHI, 2021, "Score-Based Likelihood Ratios for Linguistic Text Evidence with a Bag-of-Words Model", *Forensic Science International* 327, 110980. Available at https://doi.org/10.1016/j.forsciint.2021.110980

KESTEMONT, MIKE, "Function Words in Authorship Attribution from Black Magic to Theory?", in *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLF)*, Association for Computational Linguistics, Gothenburg, 2014, pp. 59-66.

KOPPEL, MOSHE, and WINTER, YARON, 2014, "Determining If Two Documents Are Written by the Same Author", *Journal of the Association for Information Science and Technology* 65 (1), pp. 178–87. Available at https://doi.org/10.1002/asi.22954

KOPPEL, MOSHE, SCHLER, JONATHAN, ARGAMON, SHLOMO, and WINTER, YARON, 2012, "The 'Fundamental Problem' of Authorship Attribution", *English Studies* 93 (3), pp. 284–91.

MOSTELLER, FREDERICK and WALLACE, David L., 1963, "Inference in an Authorship Problem", *Journal of the American Statistical Association* 58, pp. 275–309. Available at https://doi.org/10.2307/2283270.

NINI, ANDREA, 2023, *A Theory of Linguistic Individuality for Authorship Analysis*, C.U.P., Cambridge,

NINI, ANDREA, HALVANI OREN, GRANER, LUKAS, GHERARDI, VALERIO, and ISHIHARA; SHUNICHI, 2025, "Grammar as a Behavioral Biometric: Using Cognitively Motivated Grammar Models for Authorship Verification." Available at https://doi.org/10.48550/arXiv.2403.08462

NYHAN, JULIANNE, TERRAS MELISSA, and VANHOUTTE EDWARD, 2013, "Introduction", in M. Terras, J. Nyhan, and E. Vanhoutte (eds.), *Defining Digital Humanities: A Reader*, Ashgate, Farnham, pp. 17–26.

POTHA, NEKTARIA, and STAMATATOS EFSTATHIOS, 2020, "Improved Algorithms for Extrinsic Author Verification", *Knowledge and Information Systems* 62 (5), pp. 1903–21. Available at https://doi.org/10.1007/s10115-019-01408-4

RAMOS, DANIEL, GONZALEZ-RODRIGUEZ, JOAQUIN, ZADORA GRZEGORZ, and AITKEN; COLIN, 2013, "Information-Theoretical Assessment of the Performance of Likelihood Ratio Computation Methods", *Journal of Forensic Sciences* 58 (6), pp. 1503–18. Available at https://doi.org/10.1111/1556-4029.12233

SMITH, SYDNEY, 1850, *Elementary Sketches of Moral Philosophy Delivered at the Royal Institution in the Years 1804, 1805, 1806,* Longman, London.

THE SHORTHAND NOTEBOOKS OF CHARLES DICKENS AND ARTHUR P. STONE. ca. 1859, Manuscripts, Free Library of Philadelphia, Philadelphia, PA

WOOD, CLAIRE, and BOWLES, HUGO, 2023-, "Deciphered Shorthand", online exhibition. Available at https://dickenscode.omeka.net/exhibits/show/deciphered-shorthand

APPENDIX A

1. Full text of "The Two Brothers" (374 words)

I once heard a story when I was living in Scotland from the mouth of a deceased judge which struck my imagination very forcibly. He told it by the dying light of a wood fire with a high wind roaring outside. I ought to premise that he had been counsel in an action that afterwards shared of the gains attending a will made by one of the wards. This is the story:

There were two old bachelor brothers of whom one reposed at Slough and the other at London. They were in the habit of visiting each other for a week or so at a time. It happened that the Slough brother, being on one of these visits to the London brother, felt rather unwell said 'I will go home. My old housekeeper [will help] me. I shall soon be all right and will then come back. Once again they parted. A few nights afterwards, the London brother being in bed and finding himself uncomfortable and unable to sleep, lighted a candle and composed himself to read. Suddenly the door of his room opened and an appearance of his brother, dressed in white, pale and looking very attentively at him. He spoke to the figure but it made no answer. So it passed on to the end of the room turned and went out at the door. The London brother's immediate impression was that his mind was ill, had some fullness of blood in the region of the brain and had best be bled. So he put on his dressing gown to call up his servant. He heard the sound of a carriage in the street which came over and stopped at his door. On his opening the window and asking who was there, he was answered "A messenger to take you to your brother who is very ill [and] supposed to be dying." Of course he accompanied the messenger instantly. When he arrived in his brother's room, there were several people present. On sight of them all, the brother rose in his bed with difficulty and spoke these words: "John, you have seen me before tonight and you know it!" With that, he fell back on his pillow and never spoke another word.

2. Text of the tale of the two brothers in *To Be Read At Dusk* (1028 words)

I took an engagement once (pursued the German courier) with an English gentleman, elderly and a bachelor, to travel through my country, my Fatherland. He was a merchant who traded with my country and knew the language, but who had never been there since he was a boy—as I judge, some sixty years before.

His name was James, and he had a twin-brother John, also a bachelor. Between these brothers there was a great affection. They were in business together, at Goodman's Fields, but they did not live together. Mr. James dwelt in Poland Street, turning out of Oxford Street, London; Mr. John resided by Epping Forest.

Mr. James and I were to start for Germany in about a week. The exact day depended on business. Mr. John came to Poland Street (where I was staying in the house), to pass that week with Mr. James. But, he said to his brother on the second day, 'I don't feel very well, James. There's not much the matter with me; but I think I am a little gouty. I'll go home and put myself under the care of my old housekeeper, who understands my ways. If I get quite better, I'll come back and see you before you go. If I don't feel well enough to resume my visit where I leave it off, why *you* will come and see me before you go.' Mr. James, of course, said he would, and they shook hands—both hands, as they always did—and Mr. John ordered out his old-fashioned chariot and rumbled home.

It was on the second night after that—that is to say, the fourth in the week—when I was awoke out of my sound sleep by Mr. James coming into my bedroom in his flannel-gown, with a lighted candle. He sat upon the side of my bed, and looking at me, said:

'Wilhelm, I have reason to think I have got some strange illness upon me.'

I then perceived that there was a very unusual expression in his face.

'Wilhelm,' said he, 'I am not afraid or ashamed to tell you what I might be afraid or ashamed to tell another man. You come from a sensible country, where mysterious things are inquired into and are not settled to have been weighed and measured—or to have been unweighable and unmeasurable—or in either case to have been completely disposed of, for all time—ever so many years ago. I have just now seen the phantom of my brother.'

I confess (said the German courier) that it gave me a little tingling of the blood to hear it.

'I have just now seen,' Mr. James repeated, looking full at me, that I might see how collected he was, 'the phantom of my brother John. I was sitting up in bed, unable to sleep, when it came into my room, in a white dress, and regarding me earnestly, passed up to the end of the room, glanced at some papers on my writing-desk, turned, and, still looking earnestly at me as it passed the bed, went out at the door. Now, I am not in the least mad, and am not in the least disposed to invest that phantom with any external existence out of myself. I think it is a warning to me that I am ill; and I think I had better be bled.'

I got out of bed directly (said the German courier) and began to get on my clothes, begging him not to be alarmed, and telling him that I would go myself to the doctor. I was just ready, when we heard a loud knocking and ringing at the street door. My room being an attic at the back, and Mr. James's being the second-floor room in the front, we went down to his room, and put up the window, to see what was the matter.

'Is that Mr. James?' said a man below, falling back to the opposite side of the way to look up.

'It is,' said Mr. James, 'and you are my brother's man, Robert.'

'Yes, Sir. I am sorry to say, Sir, that Mr. John is ill. He is very bad, Sir. It is even feared that he may be lying at the point of death. He wants to see you, Sir. I have a chaise here. Pray come to him. Pray lose no time.'

Mr. James and I looked at one another. 'Wilhelm,' said he, 'this is strange. I wish you to come with me!' I helped him to dress, partly there and partly in the chaise; and no grass grew under the horses' iron shoes between Poland Street and the Forest.

Now, mind! (said the German courier) I went with Mr. James into his brother's room, and I saw and heard myself what follows.

His brother lay upon his bed, at the upper end of a long bed-chamber. His old housekeeper was there, and others were there: I think three others were there, if not four, and they had been with him since early in the afternoon. He was in white, like the figure—necessarily so, because he had his night-dress on. He looked like the figure—necessarily so, because he looked earnestly at his brother when he saw him come into the room.

But, when his brother reached the bed-side, he slowly raised himself in bed, and looking full upon him, said these words:

'James, you have seen me before, to-night—and you know it!'

And so died!

I waited, when the German courier ceased, to hear something said of this strange story. The silence was unbroken. I looked round, and the five couriers were gone: so noiselessly that the ghostly mountain might have absorbed them into its eternal snows. By this time, I was by no means in a mood to sit alone in that awful scene, with the chill air coming solemnly upon me—or, if I may tell the truth, to sit alone anywhere. So I went back into the convent-parlour, and, finding the American gentleman still disposed to relate the biography of the Honourable Ananias Dodger, heard it all out.

Appendix B Datasets

Novels corpus - n. tokens by author		
Dickens corpus	*no.novels*	*Total no. tokens*
Dickens	12	3,503,401
19th century corpus		
Collins	2	542,013
Eliot	2	608,904
Gaskell	2	404,522
Gissing	2	411,748
Jerrold	1	122,456
Lever	2	591,622
Martineau	3	455,552
Parr	1	173,785
Stowe	1	228,588
Thackeray	2	565,399
Trollope	1	233,300
Yonge	1	284,488
Total	**20**	**4,622,377**

Ghost story corpus - n.stories and tokens by author		
Dickens corpus	*no. stories*	*Total no. tokens*
Dickens	**10**	**47,345**
Victorian corpus		
Ainsworth	1	4,402
Edwards	1	2,968
Gaskell	1	11,256
Hawker	1	4,844
Le Fanu	1	10,032
Millington	1	4,653
Molesworth	1	4,845
Mulock	1	5,718
Scott	1	6,189
Wood	1	8,472
Total	**10**	**63,379**

STORIES BEING ANALYSED		*n.tokens*
Dickens	Tale of two brothers in *To Be Read at Dusk*	374
Unknown	"Two Brothers" story as dictated to Arthur Stone	1,028

Nini, A., Wood, C., Bowles, H. (2025). The Dickens signal: Investigating the authorship of “The Two Brothers”, a ghost story transcribed from shorthand. *Textus, English Studies in Italy*, 3, 25-52.