



Adjectives and adverbs as stylometric analysis parameters

Eugenia Lukin¹ · James Cooper Roberts¹ · David Berdik² · Eliana Mugar¹ · Patrick Juola³

Received: 8 February 2022 / Accepted: 9 January 2023 / Published online: 22 May 2023

© The Author(s) 2023

Abstract

The present study considers the role of adjectives and adverbs in stylometric analysis and authorship attribution. Adjectives and adverbs allow both for variations in placement and order (adverbs) and variations in type (adjectives). This preliminary study examines a collection of 25 English-language blogs taken from the Schler Blog corpus, and the Project Gutenberg corpus with specific emphasis on 3 works. Within the blog corpora, the first and last 100 lines were extracted for the purpose of analysis. Project Gutenberg corpora were used in full. All texts were processed and part-of-speech tagged using the Python NLTK package. All adverbs were classified as sentence-initial, preverbal, interverbal, postverbal, sentence-final, or none-of-the-above. The adjectives were classified into types according to the universal English type hierarchy (Cambridge Dictionary Online, 2021; Annear, 1964) manually by one of the authors. Ambiguous adjectives were classified according to their context. For the adverbs, the initial samples were paired and used as training data to attribute the final samples. This resulted in 600 trials under each of five experimental conditions. We were able to attribute authorship with an average accuracy of 9.7% greater than chance across all five conditions. Confirmatory experiments are ongoing with a larger sample of English-language blogs. This strongly suggests that adverbial placement is a useful and novel idiolectal variable for authorship attribution (Juola et al., 2021). For the adjective, differences were found in the type of adjective used by each author. Percent use of each type varied based upon individual preference and subject-matter (e.g. *Moby Dick* had a large number of adjectives related to size and color). While adverbial order and placement are highly variable, adjectives are subject to rigid restrictions that are not violated across texts and authors. Stylometric differences in adjective use generally involve the type and category of adjectives preferred by the author. Future investigation will focus, likewise, on whether adverbial variation is similarly analyzable by type and category of adverb.

✉ Eugenia Lukin
citizengenia@gmail.com

Keywords Stylometry · Authorship attribution · Linguistic features · Individual language variation · Syntax

1 Introduction

Authorship attribution is the task of determining the identity, or demographic characteristics, of an author, from the material they wrote. The problem of attributing a specific text to a specific author, or distinguishing between several texts and their authors, is pertinent in historical documentation (Holmes, 1998; Binongo, 2003; Zhao & Zobel, 2007; Tyrkkö, 2013), criminal justice and forensics, detecting plagiarism and more (van Halteren, 2004).

Earliest authorship attribution approaches and attempts to identify authors relied mainly on stylistic features such as humor, sentence complexity, word choice and descriptiveness (Zhao & Zobel, 2007). Scholars of literature spent a great deal of time on the “style” of a particular author, and were able to detect, through literary criticism tools or simple instinct, the signature of an author’s work.

However, in order to attain truly reliable results, authorship attribution has to utilize statistical and computational approaches. Computers simply do the work much better than humans ever could. In such a circumstance, the problem of quantifying features for authorship identification and deciding which features to use, becomes paramount. What does it mean to operationalize style? How do we transfer the human ability to detect differences into a reliable, statistically robust computer program?

This problem of which features to choose – the problem of stylometry – is a long-standing one. From the onset, the features that stylometric analyses needed to select had to be ‘salient, structural, frequent, easily quantifiable, and relatively immune from conscious control.’ (Bailey in Holmes, 1998). Robust stylometric approaches rely on the fact that a work, whether it be an email or a classical novel, is a series of countable words and morphemes, some of which are extremely prevalent, such as the *n*-grams /the/ and /ing/.

Modern approaches to authorship attribution additionally rely on a large variety of stylometric features, grouped into the stylistic word length, phrase length, distribution of function words, digit frequency, number of short words, etc.) and content features (POS tagging, vocabulary choice, lexical features, character *n*-grams, word *n*-grams, etc.) (see e.g.: Tanguy et al., 2011; Savoy, 2012; Wu et al., 2021; Zhao & Zobel 2007; Sundararajan & Woodard, 2018). A few approaches venture beyond the lexical and look at syntactic features (Wu et al., 2021; Varela et al., 2010) by building complex Multi-Channel Self-Attention networks, and some approaches dedicate special attention to specific parts of speech and lexical entries such as verbs (Varela et al., 2010), function words, Segarra et al. (2015), thematic vocabulary (Savoy, 2012) and other linguistic features (Tanguy et al., 2011). Most authorship attribution and stylometry, however, is still focused on the lexical features mentioned above, in large part due to their robustness and reliability.

This paper looks at two stylometric parameters that have, so far, received relatively little attention; adverbs and adjectives. While amounts of adjectives and adverbs have been utilized as part of general *n*-gram approaches, little has been done to examine

their properties and to utilize these specific properties in authorship attribution. To that end, we have looked at two types of corpora; the Schler Blog Corpus, a collection of English Language blogs, and the Project Gutenberg Corpus, with a specific focus on the novels of Herman Melville, Jane Austen, and G. K. Chesterton.

2 Adverbs

Adverbs in English can appear in a variety of positions in a clause. In the example case below, each sentence is at least borderline acceptable and has equivalent semantic content.

1. *Passionately*, she kissed her husband.
2. She *passionately* kissed her husband
3. ?She kissed *passionately* her husband.
4. She kissed her husband *passionately*.

Given that there is no discernibly significant difference between the meanings of (1)–(4), a speaker or writer has the liberty to choose where they will place an adverb. From this liberty, one could hypothetically develop personal preferences for one position over another. These preferences may subsequently carry authorial information and have the potential for stylometric applications.

We hypothesize that adverb placement analysis could be used for the purposes of authorship attribution. By comparing average percentages of adverb positions of a known document to an anonymous document, we should be able to accurately posit the author of the anonymous document.

2.1 Constraints on placement

Of course, we are not insinuating that a given author will use adverbs in a single position in a sentence and nowhere else. For one, it is not the case that every adverb is grammatically pronounceable in any position in a sentence. As illustrated in Table 1¹ below, some adverbs—particularly adverbs of manner, such as *beautifully*, *quietly*, *aggressively*, *carefully*, and so on—can occupy a wide variety of positions in a sentence. Other adverbs, like *nearly*, are only pronounceable immediately before or after the verb phrase, while adverbs like *daily* are pronounceable only at the end of the sentence.

This variation in placeability may stem from the adverb's argued status as a waste-bin taxon (Carter & McCarthy, 2017). Scholars from many different schools of thought have investigated the scope, semantics, and syntactic constraints of adverbs (Ernst, 2002; Cinque, 1999). To the extent of our knowledge, however, there is not

¹In this paper, we employ the abbreviations TP and VP for *tense phrase* and *verb phrase*, respectively. Such abbreviations are standard in the syntax-semantics literature. For the purposes of this study, tense phrases are interpreted as the syntactic constituent containing the subject, verb, and object(s) (if applicable). They may be alternatively referred to as *inflectional phrases*. Verb phrases are interpreted as the syntactic constituent containing the verb and its object(s) (if applicable).

Table 1 Legal and illegal adverb placements in English

pre-TP	*Nearly, Takako ate the wrong curry.	Usually, Paul draws trees.	Carefully, Max painted the ceramic pot.	*Daily, the student eats ramen.
pre-VP	Takako nearly ate the wrong curry.	Paul usually draws trees.	Max carefully painted the ceramic pot.	*The student daily eats ramen.
post-VP	Takako ate the wrong curry, nearly.	Paul draws trees usually.	Max painted the ceramic pot carefully.	The student eats ramen daily.

yet a comprehensive taxonomy of adverbs based on their placeability in the surface representation of a sentence.² For the purposes of this article, we will bear in mind that some adverbs have a wider range of possible placements than others (Table 2).

2.2 Methodology

We began with a corpus of 25 English-language blogs taken from the Schler Blog Corpus (Schler et al., 2006), which were then processed to extract the first and last 100 sentences of each. Each sentence was part-of-speech tagged using the Python NLTK package (Steven et al., 2009); objects identified as adverbs were sorted into one of six categories based on their position relative to every non-adverb. In a scenario where the first verb in a sentence is α and the last verb in a sentence is β ...

1. an adverb that is before the first non-adverb is *sentence-initial*
2. an adverb that is between the first non-adverb but before α is *preverbal*
3. an adverb that is between α and β is *interverbal*
4. an adverb that is between β and the last non-adverb in the sentence is *postverbal*
5. an adverb that is after the last non-adverb is *sentence-final*
6. an adverb that is not truthfully described by any of the above is *other*

The initial samples were paired and used as training data to attribute the final samples. This resulted in a series of 600 triangle tests, with one true author and one distractor author. Five different experimental conditions were used to make the determination: raw adverbial counts, adverbial counts vectorized by type, total data, and the first two metrics again with normalized adverbial counts.

2.3 Results

In this experiment, 300 correct attributions were expected by chance. As can be seen in Fig. 1, every experimental condition performed above that value. Specifically, we were able to attribute authorship with an average accuracy of 9.7% greater than chance across all conditions. At maximum, we attributed authorship with an accuracy 12.8% greater than chance.

²We were informed through word-of-mouth that Prof. Evelien Keizer of the University of Vienna is studying adverb placement. However, as of this article's writing, she has not yet published anything reporting her findings.

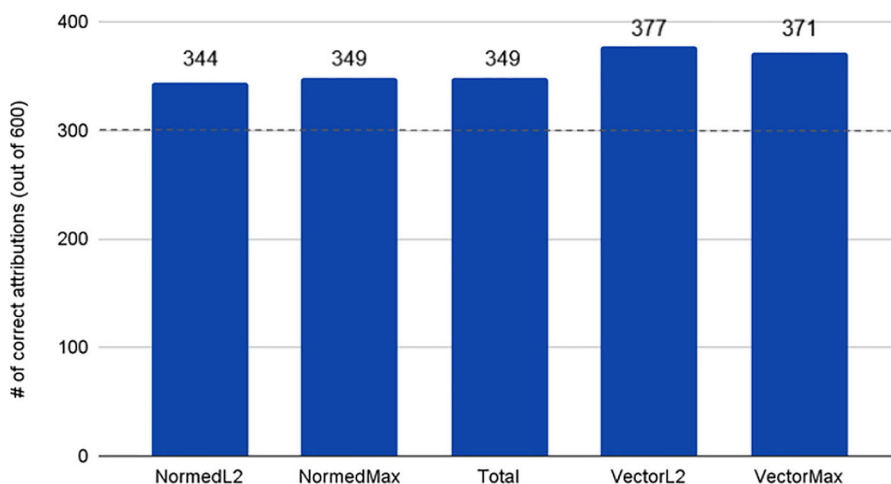
Table 2 Examples of each category described in the list above

sentence-initial	<i>Realistically</i> , I doubt I could finish this paper by Tuesday.	post-verbal	The doctor's advice was <i>incredibly</i> helpful.
preverbal	Erin <i>carefully</i> removed the plant from its pot.	sentence-final	John is suspended from the library <i>indefinitely</i> .
interverbal	He swims <i>quickly</i> and runs slowly.	other	<i>Really?</i>

However, this method of analysis seems to be most effective when raw adverb counts are used. Using normalized counts (i.e., the number of adverbs in each category are divided by the total number of adverbs) loses a noticeable amount of information and is less accurate. This suggests that comparing the frequency of adverbs is useful for stylometry as well. Nonetheless, the data suggests that individual variations in adverbial placement are distinctive. By extension, this variable has stylometric potential and is useful for authorship attribution.

2.4 A D-structure approach?

We based our analysis on adverbs as they appeared relative to verbs and non-adverbs in a given sentence. In other words, we used surface representations as opposed to deep structures (henceforth D-structure(s)) that represent the underlying syntax and semantics of an expression (Chomsky, 1971). Analyzing the placement of adverbs in this hypothesized D-structure may reveal a similar pattern in regards to authorial information. For example, one author may adjoin adverbs to a verb phrase (as in (1)) more frequently than they do to an inflectional phrase (as in (2)). Alternatively, they

**Fig. 1** Correct attributions by experimental condition

may place adverbs to the right of a phrase (as in (3)) more often than they do the left of a phrase (as in (4)).

1. Alice *tenderly* hugs Laina.
2. *Tenderly*, Alice hugs Laina.
3. I have work today *actually*.
4. *Actually*, I have work today.

While not technically impossible, this approach was avoided for several reasons. For one, there are scenarios where an adverb-bearing sentence is ambiguous between two D-structures. In (3) above, it is unclear whether *actually* adjoins to the verb phrase or the inflectional phrase. Furthermore, formal analyses of both individual adverbs and the category as a whole are remarkably complex, if not subjects of debate. As summarized by Ernst (2002), “nobody seems to know exactly what to do with adverbs.” Cinque (1999), for example, suggests that the adjoin function could be done away with entirely under the schema for adverbs they posited. Conflicting theories aside, it seems unrealistic to generate syntax trees for thousands of sentences in a corpus that may contain typographical errors or irregular structures not recognized by an automatic sentence parser. While a D-structure approach to this question is worth revisiting for a future study, we limited our analysis to surface representations.

3 Adjectives

When a noun is preceded by more than one adjective in English, the adjectives have a canonical, internal order. It’s possible to talk about the violent big green monster from outer space, but it would be extremely odd to talk about the green big violent monster from outer space. The Cambridge English Dictionary (Cambridge Dictionary Online, 2021) defines the order of adjectives in the English language as:

Opinion > Size > Physical Quality > Shape > Age > Color > Origin > Material > Type > Purpose

Each adjective is classified into one of the aforementioned categories, as shown (Table 3). Note, however, that polysemic adjectives can be classified into more than one category. For instance, the adjective *great* can either denote size as in ‘the great white whale’, quality, as in ‘the great person’ or number, as in ‘the great majority’.

Notably, while variations in classification exist (origin, material and type are often lumped together under the broad category of proper adjective), the order of adjectives is not generally disputed. Though syntactic arguments for adjective order have been made (Rosato, 2013) it seems that actual constraints on the order of adjectives as seen here are semantic and pragmatic, adjectives located closer to the noun are seen as more essentially tied to the noun, or as more necessary (Rosato, 2013; Wulff, 2003).

We examined two potential factors that may differentiate between authors for the purpose of authorship attribution. The first was variability in the canonical order of adjectives – whether it would be possible to find exceptions or deviations correlating with demographic characteristics or period of writing. The second was adjective distribution and proportion of different categories of adjectives in a work.

Table 3 Examples of adjective categories

Category	Examples
Opinion	great, poor, unusual
Size	big, small, tiny
Physical Quality	rough, smooth, clean
Age	young, old
Color	white, black, blue, red
Origin	English, American, Japanese
Material	metal, plastic, glass
Type	bread-like, U-shaped
Purpose	cleaning, cooking, working

3.1 Methodology

We began by analyzing the large corpus of classical works available on Project Gutenberg, and tagging all adjectives with the NLTK speech tagger. The number of adjectives in the works can be seen in Table 4 below:

We then proceeded to use the NLTK POS-tagger to tag pairs of the form [adjective-adjective] noun, where a noun was preceded by more than two adjectives, the pairs were tagged separately and the adjectives overlapped. The adjective pairs were then analyzed manually in G. K. Chesterton's *Father Brown*, Jane Austen's *Emma* and Herman Melville's *Moby Dick*. Word frequencies were then extracted in order to allow us to analyze adjectives by category, and several sample categories chosen for analysis.

Additionally, we calculated word frequencies for the corpus, and manually assembled them into the adjective categories listed above for the works of Jane Austen and Herman Melville. Relative distribution of adjectives was calculated for several of the works.

Table 4 Number of tagged adjectives in Project Gutenberg corpus, by author and work

Author	Work	No. of Adjs.
Austen	Emma	10873
Austen	Persuasion	5695
Austen	Sense & Sensibility	7549
Bryant	Stories	3193
Burgess	The Adventures of Buster Bear	1090
Carrol	Alice in Wonderland	1501
Chesterton	The Ball and the Cross	6415
Chesterton	The Adventures of Father Brown	5716
Chesterton	The Man Who Was Thursday	4502
Melville	Moby Dick	18398
Whitman	Leaves of Grass	10461

3.2 Results

The first notable result was that manual analysis of adjective pairs did not show divergence in adjective order. All adjective pairs and triads adhered to the general order of opinion > age > size > color. This held true across author genders (male and female), across genres (Austen's romance, Chesterton's mystery and Melville's nautical work) and across period (*Emma*, published in 1815; *Moby Dick*, published in 1851; *Father Brown*, published in 1910-1936).

Frequency analysis and grouping shows differing distributions of categories of adjectives by author. Whereas the Opinion adjectival category is the most numerous for all works, the percent share it has of the total number of adjectives differs markedly. This is true for all other categories of adjectives, as well (Table 5).

As shown in Fig. 2, opinion adjectives constitute approximately 48% of all adjectives used, whereas color constitutes less than 1%. The work of Herman Melville, on the other hand, shows a very different adjectival distribution, as shown in Fig. 3 below.

In Melville's work, Opinion adjectives constitute 28% of total adjectives, whereas age (3%), size (8%) and color (3%) constitute a significantly greater share of adjectives compared to Austen. Adjectives marked as Other were either ambiguous, unique, or belonging to the category Number and therefore potentially constituting classifiers (Table 6).

4 Discussion

Our results point to the fact that looking in detail at specific parts of speech, such as adverbs and adjectives, yields fruitful approaches to authorship attribution. Authors differ in their choice of adverb placement, their choice of adjective categories, and their choice of adjective and adverb frequency, just as they differ in their choices of nouns and verbs.

It is important, however, to note that not all features of these parts of speech are created equal; what features we look at will matter. For instance, all authors examined used adjective order in the same way; as a canonical feature of the language, it is not

Table 5 Counted number of adjectives in Jane Austen's *Emma*

Category	Number
Color	25
Size	611
Age	375
Opinion	5186
Origin	3
Shape	0
Other	4673
Total No. of Adjs.	10873

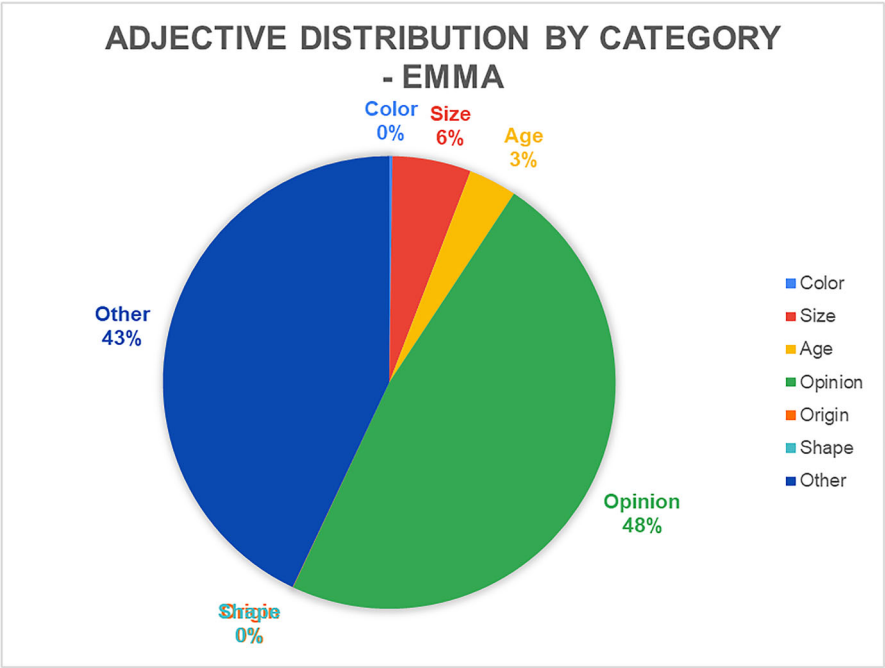


Fig. 2 Distribution of Adjectives by category in Jane Austen’s Emma

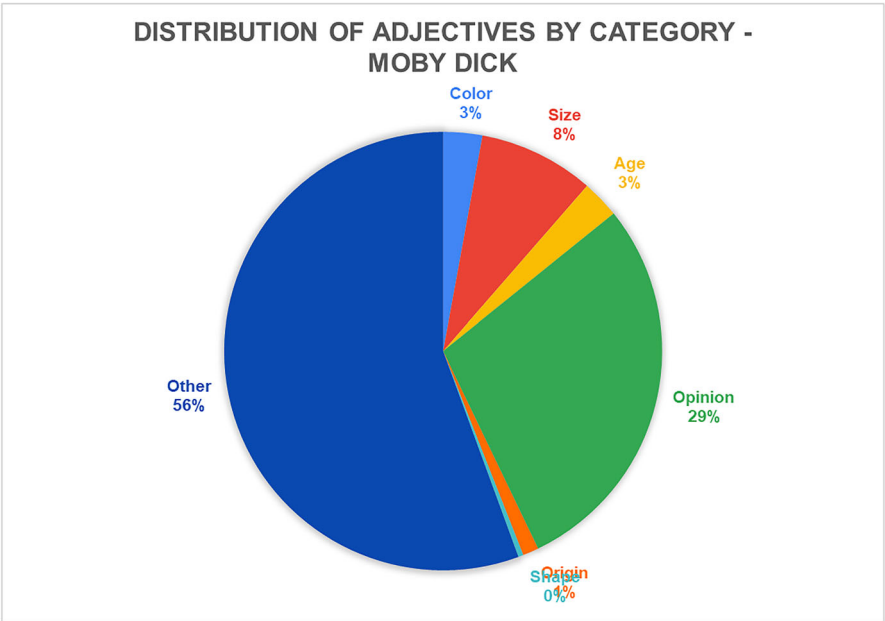


Fig. 3 Distribution of Adjectives by category in Herman Melville’s Moby-Dick

Table 6 Counted number of adjectives in Herman Melville's *Moby Dick*

Category	Number
Color	530
Size	1570
Age	514
Opinion	5267
Origin	217
Shape	68
Other	10222
Total No. of Adjs.	18388

much more variable than the English-language requirement that sentences contain a pronoun and a verb.

In the case of adjectives, the variability seen in the results when examining adjective categories and word choice stems from two obvious sources; the first one is that adjectives are determined by the choice of subject and the descriptive, pragmatic necessity. So, in a book about the sea and a great white whale, we will see a preponderance of adjectives of color, and in a work concerning families and love such as Austen's *Emma*, we will see many adjectives pertaining to individual quality and age. The second determiner of the adjectives used in a work is, of course, individual preference. The most frequently utilized adjective in Jane Austen's *Emma* is 'little' (347 instances out of 10873 total adjectives, 1:31 adjectives), whereas the most frequent adjective in Herman Melville's *Moby Dick* is 'old' (429 instances out of a total of 18398 adjectives.). Melville also used the adjective 'little' but only in 239 instances out of 18398 (a ratio of only 1:76 adjectives).

As we originally hypothesized, the data we gathered suggests that analyzing adverb placement is useful for the purpose of authorship attribution. To some extent, authors vary in where they tend to place an adverb in a sentence. However, this particular technique has the potential for refinement. It would be interesting to repeat this study excluding the most frequent placement category across authors. If there is some "marked" position to which writers of the dataset default, would excluding this category in the analysis improve accuracy? Alternatively, as natural language processing technology improves, an approach similar to the one mentioned in 2.4 may become feasible. It is possible that analyzing an adverb's position in a syntax tree will improve the rate of correct attributions in our model. Furthermore, the study is significantly limited by the size of the examined corpus. Although examination of a larger corpus is beyond the scope of the present, it may prove a fruitful avenue for further research. Regardless, the results of the current study are encouraging and warrant further investigation.

It is possible that the pattern we observed for adverbs holds for other parts of speech that have flexible placeability. Certain prepositional phrases, for example, can occur before or after a clause and mean essentially the same thing (consider *on Tuesday I walked the dog* and *I walked the dog on Tuesday*). Analogous phenomena in other languages, such as floating quantifiers in Japanese (Fukushima, 1991), may also

be subject to an author's preferences. If this is the case, they could also be analyzed for stylometric purposes.

The results of the current study dovetail neatly with the general, accepted approaches to stylometry. Adverbs and adjectives can be seen as special cases of word n-grams and POS-tagging, two approaches that have proven reliable in the past (Houvardas & Stamatatos, 2006; Koppel et al., 2009), as well as of vocabulary selection (Savoy, 2012).

5 Conclusion

Great success has been observed using previously established methods for authorship attribution. However, some counters to these techniques—principally obfuscation (Mahmood et al., 2019)—lie on the horizon. Emerging problems such as these present a difficult challenge to the field as said techniques are not designed to account for them. To handle this inevitability, we need as many tools as possible on our belt. While an analysis of adverbs or adjectives is not the figurative silver bullet that will solve all problems, the results of this study are encouraging. As the field grows, linguistic approaches to stylometry are necessary and must continue to be explored.

Data Availability The raw data generated and analysed during the adverbs portion of the current study are not publicly available due to hardware failure. The texts used in the adverb analysis of this study are available to the public at [The Blog Authorship Corpus](#). The texts used in the adjectives analysis are available to the public through [Project Gutenberg](#).

Declarations

Conflict of Interests This material is based upon work supported by the National Science Foundation under Grant No. 1814602. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Annear, S. S. (1964). The ordering of pre-nominal modifiers in English. *Project on Linguistic Analysis*, no. 8.
- Binongo, J. N. G. (2003). Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution. *Chance*, 16(2), 9–17.
- Cambridge Dictionary Online (2021). <https://dictionary.cambridge.org/us/grammar/british-grammar/adjectives-order>. Accessed 30 June 2021.

- Carter, R., & McCarthy, M. (2017). Spoken grammar: Where are we and where are we going? *Applied Linguistics*, 38(1), 1–20.
- Chomsky, N. (1971). Deep structure, surface structure and semantic interpretation.
- Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Ernst, T. B. (2002). *The syntax of adjuncts*. Cambridge: Cambridge University Press.
- Fukushima (1991). Phrase structure grammar, montague semantics, and floating quantifiers in Japanese. *Linguistics and Philosophy*, 14(6), 581–628. <https://doi.org/10.1007/BF00631961>.
- van Halteren, H. (2004). Linguistic profiling for author recognition and verification. In *Proc. of the 42nd annual meeting of the association for computational linguistics* (pp. 199–206).
- Holmes, D. I. (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing*, 13(3), 111–117. <https://doi.org/10.1093/lc/13.3.111>.
- Houvardas, J., & Stamatatos, E. (2006). N-Gram feature selection for authorship identification. *Lecture Notes in Computer Science*, 77–86. https://doi.org/10.1007/11861461_10.
- Juola, P., Berdik, D., & Roberts, J.C. (2021). Adverbial placement for stylometry [Conference presentation] Corpus Linguistics International Conference, Limerick, Ireland.
- Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for information Science and Technology*, 60(1), 9–26.
- Mahmood, A., Ahmad, F., Shafiq, Z., Srinivasan, P., & Zaffar, F. (2019). A girl has no name: Automated authorship obfuscation using Mutant-X. *Proceedings on Privacy Enhancing Technologies*, 2019(4), 54–71.
- Rosato, E. (2013). Adjective order in English: A semantic account with cross-linguistic applications (Doctoral dissertation, Carnegie Mellon University).
- Savoy, J. (2012). Authorship attribution based on specific vocabulary. *ACM Transactions on Information Systems (TOIS)*, 30(2), 1–30.
- Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. (2006). Effects of Age and Gender on Blogging in Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs. (pdf).
- Segarra, S., Eisen, M., & Ribeiro, A. (2015). Authorship attribution through function word adjacency networks. *IEEE Transactions on Signal Processing*, 63(20), 5464–5478.
- Steven, B., Loper, E., & Klein, E. (2009). *Natural language processing with python*. O'Reilly Media Inc.
- Sundararajan, K., & Woodard, D. (2018). What represents “style” in authorship attribution? In *Proceedings of the 27th international conference on computational linguistics* (pp. 2814–2822).
- Tanguy, L., Urieli, A., Calderone, B., Hathout, N., & Sajous, F. (2011). A multitude of linguistically-rich features for authorship attribution. In *PAN Lab at CLEF*.
- Tyrkkö, J. (2013). Exploring part-of-speech profiles and authorship attribution in early modern medical texts. Meaning in the history of english: words and texts in context.
- Varela, P., Justino, E., & Oliveira, L.S. (2010). Verbs and pronouns for authorship attribution. In *17th International conference on systems, signals and image processing (IWSSIP 2010)* (pp. 89–92).
- Wu, H., Zhang, Z., & Wu, Q. (2021). Exploring syntactic and semantic features for authorship attribution. *Applied Soft Computing*, 111, 107815.
- Wulff, S. (2003). A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics*, 8(2), 245–282.
- Zhao, Y., & Zobel, J. (2007). Searching with style: Authorship attribution in classic literature. In *ACM international conference proceeding series*, (Vol. 244 pp. 59–68).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Eugenia Lukin¹ · James Cooper Roberts¹ · David Berdik² · Eliana Mugar¹ · Patrick Juola³

James Cooper Roberts
jc robert@bu.edu

David Berdik
dgberdik@gmail.com

Eliana Mugar
emugar@bu.edu

Patrick Juola
juola@mathcs.duq.edu

¹ Linguistics Department, Boston University, Boston, MA, USA

² Computer Science, Duquesne University, Pittsburgh, PA, USA

³ Evaluating Variations in Language Laboratory, Duquesne University, Pittsburgh, PA, USA