

Stylometry

Stylometry is the application of the study of linguistic style, usually to written language.^[1] It has also been applied successfully to music,^[2] paintings,^[3] and chess.^[4]

Stylometry is often used to attribute authorship to anonymous or disputed documents.^[5] It has legal as well as academic and literary applications, ranging from the question of the authorship of Shakespeare's works to forensic linguistics and has methodological similarities with the analysis of text readability.

Stylometry may be used to unmask pseudonymous or anonymous authors, or to reveal some information about the author short of a full identification. Authors may use adversarial stylometry to resist this identification by eliminating their own stylistic characteristics without changing the meaningful content of their communications. It can defeat analyses that do not account for its possibility, but the ultimate effectiveness of stylometry in an adversarial environment is uncertain: stylometric identification may not be reliable, but nor can non-identification be guaranteed; adversarial stylometry's practice itself may be detectable.

History

Stylometry grew out of earlier techniques of analyzing texts for evidence of authenticity, author identity, and other questions.

The modern practice of the discipline received publicity from the study of authorship problems in English Renaissance drama. Researchers and readers observed that some playwrights of the era had distinctive patterns of language preferences, and attempted to use those patterns to identify authors of uncertain or collaborative works. Early efforts were not always successful: in 1901, one researcher attempted to use John Fletcher's preference for "'em", the contractional form of "them", as a marker to distinguish between Fletcher and Philip Massinger in their collaborations—but he mistakenly employed an edition of Massinger's works in which the editor had expanded all instances of "'em" to "them".^[6]

The basics of stylometry were established by Polish philosopher Wincenty Lutosławski in *Principes de stylométrie* (1890). Lutosławski used this method to develop a chronology of Plato's Dialogues.^[7]

The development of computers and their capacities for analyzing large quantities of data enhanced this type of effort by orders of magnitude. The great capacity of computers for data analysis, however, did not guarantee good quality output. During the early 1960s, Rev. A. Q. Morton produced a computer analysis of the fourteen Epistles of the New Testament attributed to St. Paul, which indicated that six different authors had written that body of work. A check of his method, applied to the works of James Joyce, gave the result that *Ulysses*, Joyce's multi-perspective, multi-style novel, was composed by five separate individuals, none of whom apparently had any part in the crafting of Joyce's first novel, *A Portrait of the Artist as a Young Man*.^[8]

In time, however, and with practice, researchers and scholars have refined their methods, to yield better results. One notable early success was the resolution of disputed authorship of twelve of *The Federalist Papers* by Frederick Mosteller and David Wallace.^[9] While there are still questions concerning initial assumptions and methods (and, perhaps, always will be), few now dispute the basic premise that linguistic analysis of written texts can produce valuable information and insight. (Indeed, this was apparent even

before the advent of computers: the successful application of a textual/linguistic analysis to the Fletcher canon by Cyrus Hoy and others yielded clear results during the late 1950s and early 1960s.)

Applications

Applications of stylometry include literary studies, historical studies, social studies, information retrieval, and many forensic cases and studies.^{[10][11]} Recently, long-standing debates about anonymous medieval Icelandic sagas have been advanced through its utilisation.^{[12][13][14]} It can also be applied to computer code^[15] and intrinsic plagiarism detection, which is to detect plagiarism based on the writing style changes within the document.^[16] Stylometry can also be used to predict whether someone is a native or non native English speaker by their typing speed.^[17]

Stylometry as a method is vulnerable to the distortion of text during revision.^[18] There is also the case of the author adopting different styles in the course of his career as was demonstrated in the case of Plato, who chose different stylistic policies such as those adopted for the early and middle dialogues addressing the Socratic problem.^[19]

Features

Textual features of interest for authorship attribution are on the one hand computing occurrences of idiosyncratic expressions or constructions (e.g. checking for how the author uses interpunction or how often the author uses agentless passive constructions) and on the other hand similar to those used for readability analysis such as measures of lexical variation and syntactic variation.^[20] Since authors often have preferences for certain topics, research experiments in authorship attribution mostly remove content words such as nouns, adjectives, and verbs from the feature set, only retaining structural elements of the text to avoid overfitting their models to topic rather than author characteristics.^{[21][22]} Stylistic features are often computed as averages over a text or over the entire collected works of an author, yielding measures such as average word length or average sentence length. This enables a model to identify authors who have a clear preference for wordy or terse sentences but hides variation: an author with a mix of long and short sentences will have the same average as an author with consistent mid-length sentences. To capture such variation, some experiments use sequences or patterns over observations rather than average observed frequencies, noting e.g. that an author shows a preference for a certain stress or emphasis pattern,^{[23][24]} or that an author tends to follow a sequence of long sentences with a short one.^{[25][26]}

One of the first approaches to authorship identification, by Mendenhall, can be said to aggregate its observations without averaging them.^[27]

More recent authorship attribution models use vector space models to automatically capture what is specific to an author's style, but they also rely on judicious feature engineering for the same reasons as more traditional models.^{[28][29]}

Adversarial stylometry

Adversarial stylometry is the practice of altering writing style to reduce the potential for stylometry to discover the author's identity or their characteristics.^[30] This task is also known as authorship obfuscation or authorship anonymisation. Stylometry poses a significant privacy challenge in its ability to unmask anonymous authors or to link pseudonyms to an author's other identities,^[31] which, for example, creates difficulties for whistleblowers,^[32] activists,^[33] and hoaxers and fraudsters.^[34] The privacy risk is expected to grow as machine learning techniques and text corpora develop.^[35]

All adversarial stylometry shares the core idea of faithfully paraphrasing the source text so that the meaning is unchanged but the stylistic signals are obscured.^{[36][37]} Such a faithful paraphrase is an adversarial example for a stylometric classifier.^[38] Several broad approaches to this exist, with some overlap: *imitation*, substituting the author's own style for another's; *translation*, applying machine translation with the hope that this eliminates characteristic style in the source text; and *obfuscation*, deliberately modifying a text's style to make it not resemble the author's own.^[36]

Manually obscuring style is possible, but laborious;^[39] in some circumstances, it is preferable or necessary.^[40] Automated tooling, either semi- or fully-automatic, could assist an author.^[39] How best to perform the task and the design of such tools is an open research question.^{[41][35]} While some approaches have been shown to be able to defeat particular stylometric analyses,^[42] particularly those that do not account for the potential of adversariality,^[43] establishing safety in the face of unknown analyses is an issue.^[44] Ensuring the faithfulness of the paraphrase is a critical challenge for automated tools.^[35]

It is uncertain if the practice of adversarial stylometry is detectable in itself. Some studies have found that particular methods produced signals in the output text, but a stylometrist who is uncertain of what methods may have been used may not be able to reliably detect them.^[35]

Current research

Modern stylometry uses computers for statistical analysis, and artificial intelligence and access to the growing corpus of texts available via the Internet.^[45] Software systems such as Signature^[46] (freeware produced by Peter Millican of Oxford University), JGAAP^[47] (the Java Graphical Authorship Attribution Program—freeware produced by Dr Patrick Juola of Duquesne University), stylo^{[48][49]} (an open-source R package for a variety of stylometric analyses, including authorship attribution, developed by Maciej Eder, Jan Rybicki and Mike Kestemont) and Stylene^[50] for Dutch (online freeware by Prof Walter Daelemans of University of Antwerp and Dr Véronique Hoste of University of Ghent) make its use increasingly practicable, even for the non-expert.

Academic venues and events

Stylometric methods are used for several academic topics, as an application of linguistics, lexicography, or literary study,^[1] in conjunction with natural language processing and machine learning, and applied to plagiarism detection, authorship analysis, or information retrieval.^[45]

Forensic linguistics

The International Association of Forensic Linguists (IAFL) organises the Biennial Conference of the International Association of Forensic Linguists (13th edition in 2016 in Porto) and publishes The International Journal of Speech, Language and the Law with forensic stylistics as one of its central topics.

AAAI

The Association for the Advancement of Artificial Intelligence (AAAI) has hosted several events on subjective and stylistic analysis of text.^{[51][52][53]}

PAN

PAN workshops (originally, plagiarism analysis, authorship identification, and near-duplicate detection, later more generally workshop on uncovering plagiarism, authorship, and social software misuse)

organised since 2007 mainly in conjunction with information access conferences such as ACM SIGIR, FIRE, and CLEF. PAN formulates shared challenge tasks for plagiarism detection,^[54] authorship identification,^[55] author gender identification,^[56] author profiling,^[57] vandalism detection,^[58] and other related text analysis tasks, many of which hinge on stylometry.

Case studies of interest

- In 1439, Lorenzo Valla showed that the Donation of Constantine was a forgery, an argument based partly on a comparison of the Latin with that used in authentic 4th-century documents.
- In 1952, the Swedish priest Dick Helander was elected bishop of Strängnäs. The campaign was competitive and Helander was accused of writing a series of a hundred-some anonymous libelous letters about other candidates to the electorate of the bishopric of Strängnäs. Helander was first convicted of writing the letters and lost his position as bishop but later partially exonerated. The letters were studied using a number of stylometric measures (and also typewriter characteristics) and the various court cases and further examinations, many contracted by Helander himself during the years until his death in 1978, discussed stylometric method and its value as evidence in some detail.^{[59][60]}
- In 1975, after Ronald Reagan had served as governor of California, he began giving weekly radio commentaries syndicated to hundreds of stations. After his personal notes were made public on his 90th birthday in 2001, a study used stylostatistical methods to determine which of those talks were written by him and which were written by various aides.^[61]
- In 1996, the stylometric analysis of the controversial, pseudonymously authored book *Primary Colors*, performed by Vassar College professor Donald Foster^[62] brought the topic to the attention of a wider audience after correctly identifying the author as Joe Klein. (This case was resolved only after a handwriting analysis confirmed the authorship.)
- In 1996, stylometric methods were used to compare the Unabomber manifesto with letters written by one of the suspects, Theodore Kaczynski, which resulted in Kaczynski's apprehension and later conviction.^[63]
- In April 2015, researchers using stylometry techniques identified a play, *Double Falsehood*, as being the work of William Shakespeare.^{[64][65]} Researchers analyzed 54 plays by Shakespeare and John Fletcher, and compared average sentence length, studied the use of unusual words and quantified the complexity and psychological valence of their language.
- In 2016, MacDonald P. Jackson, Emeritus Professor of English at the University of Auckland, New Zealand and a Fellow of the Royal Society of New Zealand, who had spent his entire academic career analyzing authorship attribution, wrote a book titled *Who Wrote "The Night Before Christmas"?: Analyzing the Clement Clarke Moore Vs. Henry Livingston Question*,^[66] in which he evaluates the opposing arguments and, for the first time, uses the author-attribution techniques of modern computational stylistics to examine the long-standing controversy. Jackson employs a range of tests and introduces a new one, statistical analysis of phonemes; he concludes that Livingston is the true author of the classic work.
- In 2017, Simon Fuller and James O'Sullivan published a study claiming that bestselling author James Patterson does not do any writing in his apparently co-authored novels.^{[67][68][69]} According to O'Sullivan, his collaboration with former U.S. president Bill Clinton, *The President is Missing*, is an exception to this rule.^[70]
- In 2017, a group of linguists, computer scientists, and scholars analysed the authorship of Elena Ferrante. Based on a corpus created at University of Padua containing 150 novels written by 40

- authors, they analyzed Ferrante's style based on seven of her novels. They were able to compare her writing style with 39 other novelists using, for example, *stylo*.^[48] The conclusion was the same for all of them: Domenico Starnone is the secret author of Elena Ferrante.^[71]
- In 2018, Mark Glickman, a senior lecturer in statistics at Harvard University, worked with Ryan Song, a former statistics student at Harvard, and Jason Brown, a professor at Dalhousie University in Nova Scotia, applying stylometry to find that, most likely, The Beatles' song "In My Life" was composed by John Lennon, but with a 50% chance that Paul McCartney wrote the middle eight.^{[72][73]}
 - In 2019, the ETSO project: Stylometry applied to the Spanish Golden Age Theater,^[74] directed by Álvaro Cuéllar González and Germán Vega García-Luengos (University of Valladolid) managed to gather 3000 plays of the Spanish Golden Age. After applying stylometrical analysis, the attribution of *Mujeres y criados* to Lope de Vega^{[75][76]} was ratified, and an authorship problem was detected in *La monja alférez*, a play attributed to Pérez de Montalbán which, thanks to these analyzes and through historical and philology research, was eventually attributed to Juan Ruiz de Alarcón.^{[77][78][79][80]} In 2023, the same project found Lope de Vega as the author of *La francesa Laura* (The Frenchwoman Laura), despite the manuscript was written years after his death.^[81] The comedy was classified as a late work of Lope de Vega and dated from 1628 to 1630, as its flattering treatment of France could be attributed to the momentary good relationship between Spain and France during the Thirty Years' War, having England as a common enemy.^[82] In this analysis, the 500 most frequent words of the text under investigation are compared with the 500 of the rest of the works. In the case of *La francesa Laura*, the finding detected that the 100 works with which it was closest were almost all by Lope de Vega. Machine learning methods, such as support vector machine analysis, were also conducted with a large range of parameters. The traditional philological analysis on the authorship of works has confirmed the investigations of stylometry and artificial intelligence.^[83]
 - In 2020, Rachel McCarthy and James O'Sullivan argued that Emily Brontë is the true author of *Wuthering Heights*, ending speculation by some critics that the novel might have been written by one of her siblings, specifically either Branwell or Charlotte.^[84]
 - In 2020, Hartmut Ilseemann used Rolling Delta and Rolling Classify from the R Stylo program suite to show that the Marlowe corpus is stylistically inhomogeneous, and that the author of the two *Tamburlaines* was hardly present in the remaining official corpus of Marlowe.^{[85][86][87]}
 - In 2022, the Italian scholars Simone Rebora and Massimo Salgaro showed, using John F. Burrows' "Delta distance" method, that Felix Salten is the most probable author of the anonymous novel *Josefine Mutzenbacher* from 1906, the final pages excluded.^[88]
 - In 2023, the Swedish journalist Lapo Lappin claimed that two crime novels by the Swedish author Camilla Läckberg may be the work of a ghost writer, presumably her editor Pascal Engman. This claim was first denied by the author and her spokesperson,^[89] but later Läckberg admitted that she and Pascal Engman work very closely together and he edits her texts.^[90]

Data and methods

Since stylometry has both descriptive use cases, used to characterise the content of a collection, and identificatory use cases, e.g. identifying authors or categories of texts, the methods used to analyse the data and features above range from those built to classify items into sets or to distribute items in a space of feature variation. Most methods are statistical in nature, such as cluster analysis and discriminant analysis, are typically based on philological data and features, and are fruitful application domains for modern machine learning methods.

Whereas in the past, stylometry emphasized the rarest or most striking elements of a text, contemporary techniques can isolate identifying patterns even in common parts of speech. Most systems are based on lexical statistics, i.e. using the frequencies of words and terms in the text to characterise the text (or its author). In this context, unlike for information retrieval, the observed occurrence patterns of the most common words are more interesting than the topical terms which are less frequent.^{[91][92]}

The primary stylometric method is the writer invariant: a property held in common by all texts, or at least all texts long enough to admit of analysis yielding statistically significant results, written by a given author. An example of a writer invariant is frequency of function words used by the writer.

In one such method, the text is analyzed to find the 50 most common words. The text is then divided into 5,000 word chunks and each of the chunks is analyzed to find the frequency of those 50 words in that chunk. This generates a unique 50-number identifier for each chunk. These numbers place each chunk of text into a point in a 50-dimensional space. This 50-dimensional space is flattened into a plane using principal components analysis (PCA). This results in a display of points that correspond to an author's style. If two literary works are placed on the same plane, the resulting pattern may show if both works were by the same author or different authors.

Gaussian statistics

Stylometric data are distributed according to the Zipf–Mandelbrot law. The distribution is extremely spiky and leptokurtic, the reason why researchers could not use statistics to solve e.g. authorship attribution problems. Nevertheless, usage of Gaussian statistics is perfectly possible by applying data transformation.^[93]

Neural networks

Neural networks, a special case of statistical machine learning methods, have been used to analyze authorship of texts. Texts of undisputed authorship are used to train a neural network by processes such as backpropagation, such that training error is calculated and used to update the process to increase accuracy. Through a process akin to non-linear regression, the network gains the ability to generalize its recognition ability to new texts to which it has not yet been exposed, classifying them to a stated degree of confidence. Such techniques were applied to the long-standing claims of collaboration of Shakespeare with his contemporaries John Fletcher and Christopher Marlowe,^{[94][95]} and confirmed the opinion, based on more conventional scholarship, that such collaboration had indeed occurred.

A 1999 study showed that a neural network program reached 70% accuracy in determining the authorship of poems it had not yet analyzed. This study from Vrije Universiteit examined identification of poems by three Dutch authors using only letter sequences such as "den".^[96]

A study used deep belief networks (DBN) for authorship verification model applicable for continuous authentication (CA).^[97]

One problem with this method of analysis is that the network can become biased based on its training set, possibly selecting authors the network has analyzed more often.^[96]

Genetic algorithms

The genetic algorithm is another machine learning technique used for stylometry. This involves a method that starts with a set of rules. An example rule might be, "If *but* appears more than 1.7 times in every thousand words, then the text is author X". The program is presented with text and uses the rules to determine authorship. The rules are tested against a set of known texts and each rule is given a fitness score. The 50 rules with the lowest scores are not used. The remaining 50 rules are given small changes

and 50 new rules are introduced. This is repeated until the evolved rules attribute the texts correctly.

Rare pairs

One method for identifying style is termed "rare pairs" and relies upon individual habits of collocation. The use of certain words may, for a particular author, be associated idiosyncratically with the use of other, predictable words.

Authorship attribution in instant messaging

The diffusion of the internet has shifted the authorship attribution attention towards online texts (web pages, blogs, etc.) electronic messages (e-mails, tweets, posts, etc.), and other types of written information that are far shorter than an average book, much less formal and more diverse in terms of expressive elements such as colors, layout, fonts, graphics, emoticons, etc. Efforts to take into account such aspects at the level of both structure and syntax were reported in.^[98] In addition, content-specific and idiosyncratic cues (e.g., topic models and grammar checking tools) were introduced to unveil deliberate stylistic choices.^[99]

Standard stylometric features have been employed to categorize the content of a chat by instant messaging,^[100] or the behavior of the participants,^[101] but attempts of identifying chat participants are still few and early. Furthermore, the similarity between spoken conversations and chat interactions has been neglected while being a major difference between chat data and any other type of written information.

See also

- Data re-identification
- Digital watermarking
- Linguistics and the Book of Mormon § Stylometry
- Moshe Koppel
- Quantitative linguistics
- Steganography
- Writeprint

Notes

1. Argamon, Shlomo, Kevin Burns, and Shlomo Dubnov, eds. The structure of style: algorithmic approaches to understanding manner and meaning. Springer Science & Business Media, 2010.
2. Westcott, Richard (15 June 2006). "Making hit music into a science" (<http://news.bbc.co.uk/1/hi/5083986.stm?ls>). *BBC News*.
3. Sethi, Ricky (2016-06-07). "Using computers to better understand art" (<https://theconversation.com/using-computers-to-better-understand-art-56887>). *The Conversation*. Retrieved 2021-12-01.
4. McIlroy-Young, Reid; Wang, Yu; Sen, Siddhartha; Kleinberg, Jon; Anderson, Ashton (2021). *Detecting Individual Decision-Making Style: Exploring Behavioral Stylometry in Chess* (<https://openreview.net/forum?id=9RFFgpQAOzk>). 35th Conference on Neural Information Processing Systems.

5. Chen, Hsinchun; Yang, Christopher C.; Chau, Michael; Li, Shu-Hsing (2009). *Intelligence and Security Informatics: Pacific Asia Workshop, PAISI 2009, Bangkok, Thailand, April 27, 2009. Proceedings*. Berlin: Springer Science & Business Media. p. 15. ISBN 9783642013928.
6. Samuel Schoenbaum, *Internal evidence and Elizabethan dramatic authorship; an essay in literary history and method*, p. 171.
7. Lutoslawski, W. (1898). "Principes de stylométrie appliqués à la chronologie des œuvres de Platon". *Revue des Études Grecques*. **11** (41): 61–81. doi:10.3406/reg.1898.5847 (<https://doi.org/10.3406%2Freg.1898.5847>). ISSN 0035-2039 (<https://search.worldcat.org/issn/0035-2039>).
8. Samuel Schoenbaum, *Internal evidence and Elizabethan dramatic authorship; an essay in literary history and method*, p. 196.
9. F. Mosteller & D. Wallace (1964). *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley.
10. Chaski, Carole (2012). Solan, Lawrence M; Tiersma, Peter M (eds.). *Author Identification in the Forensic Setting* (<http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199572120.001.0001/oxfordhb-9780199572120-e-36>). Oxford University Press. doi:10.1093/oxfordhb/9780199572120.001.0001 (<https://doi.org/10.1093%2Foxfordhb%2F9780199572120.001.0001>). ISBN 9780199572120. {{cite book}}: |journal= ignored (help)
11. Chaski, Carole (22 December 2005). Wecht, Cyril H.; Rago, John T. (eds.). *Forensic Science and Law: Investigative Applications in Criminal, Civil and Family Justice* (<https://books.google.com/books?id=oIjLBQAAQBAJ>). CRC Press. ISBN 978-1-4200-5811-6.
12. Michael MacPherson and Yoav Tirosh (2020). "A Stylometric Analysis of Ljósvetninga saga" (<http://www.academia.edu/44830058>). *Gripla*. **31**: 7–41.
13. Haukur Thorgeirsson (2018). "How similar are Heimskringla and Egils saga? An application of Burrows' delta to Icelandic texts" (<https://www.researchgate.net/publication/323901992>). *European Journal of Scandinavian Studies*. **48** (1): 1–18. doi:10.1515/ejss-2018-0001 (<https://doi.org/10.1515%2Fejss-2018-0001>).
14. Sigurður Ingibergur Björnsson, Steingrímur Páll Kárasón, and Jón Karl Helgason (2021). "Stylometry and the Faded Fingerprints of Saga Authors" (<https://doi.org/10.1515%2F9783110725339-005>). In *Search of the Culprit: Aspects of Medieval Authorship, edited by Lukas Rösli and Stefanie Gropper*: 97–122. doi:10.1515/9783110725339-005 (<https://doi.org/10.1515%2F9783110725339-005>). ISBN 9783110725339.
15. Claburn, Thomas (March 16, 2018). "FYI: AI tools can unmask anonymous coders from their binary executables" (https://www.theregister.co.uk/2018/03/16/identifying_anonymous_programmers/). *The Register*. Retrieved August 2, 2018.
16. Bensalem, Imene; Rosso, Paolo; Chikhi, Salim (2019). "On the use of character n-grams as the only intrinsic evidence of plagiarism". *Language Resources and Evaluation*. **53** (3): 363–396. doi:10.1007/s10579-019-09444-w (<https://doi.org/10.1007%2Fs10579-019-09444-w>). hdl:10251/159151 (<https://hdl.handle.net/10251%2F159151>). S2CID 86630897 (<https://api.semanticscholar.org/CorpusID:86630897>).
17. Brizan, David (October 2015). "Utilizing linguistically enhanced keystroke dynamics to predict typist cognition and demographics". *International Journal of Human-Computer Studies*. **82**: 57–68. doi:10.1016/j.ijhcs.2015.04.005 (<https://doi.org/10.1016%2Fj.ijhcs.2015.04.005>).
18. Alican, Necip Fikri (2012). *Rethinking Plato: A Cartesian Quest for the Real Plato*. Amsterdam: Rodopi. p. 183. ISBN 9789042035379.

19. Rowe, Christopher (2000). *The Cambridge History of Greek and Roman Political Thought* (<https://books.google.com/books?id=UmQJZPvu6uYC&q=stylometry+weakness&pg=PA160>). Cambridge, UK: Cambridge University Press. p. 160. ISBN 0521481368.
20. Stamatatos, Efstathios (2009). "A survey of modern authorship attribution methods". *JASIST*. **60** (3): 538–556. doi:10.1002/asi.21001 (<https://doi.org/10.1002%2Fasi.21001>). S2CID 6231242 (<https://api.semanticscholar.org/CorpusID:6231242>).
21. Stamatatos, Efstathios (2018). "Masking topic-related information to enhance authorship attribution". *JASIS*. **69** (3).
22. Karlgren, Jussi; Esposito, Lewis; Gratton, Chantal; Kanerva, Pentti (2018). "Authorship Profiling Without Using Topical Information". *CLEF Working Notes*. CEUR-WS.
23. Corbara, Silvia; Moreo, Alejandro; Sebastiani, Fabrizio (2022). "Syllabic quantity patterns as rhythmic features for Latin authorship attribution" (<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.24660>). *JASIST*. **74**: 128–141. arXiv:2110.14203 (<https://arxiv.org/abs/2110.14203>). doi:10.1002/asi.24660 (<https://doi.org/10.1002%2Fasi.24660>). S2CID 239998537 (<https://api.semanticscholar.org/CorpusID:239998537>).
24. Corbara, Silvia; Chulvi, Berta; Rosso, Paolo; Moreo, Alejandro (2022). "Rhythmic and Psycholinguistic Features for Authorship Tasks in the Spanish Parliament: Evaluation and Analysis" (https://doi.org/10.1007/978-3-031-13643-6_6). *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. CLEF. Springer. pp. 79–92. doi:10.1007/978-3-031-13643-6_6 (https://doi.org/10.1007%2F978-3-031-13643-6_6).
25. Karlgren, Jussi; Eriksson, Gunnar (2007). "Authors, Genre, and Linguistic Convention". *SIGIR Workshop on Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection*. SIGIR. PAN.
26. Eriksson, Linda (2014). *Sequential Aggregation of Textual Features for Domain Independent Author Identification* (MSc). KTH Royal Institute of Technology.
27. Mendenhall, T C (1887). "The characteristic curves of composition" (<https://zenodo.org/record/1448355>). *Science*. **9** (214S): 237–246. doi:10.1126/science.ns-9.214S.237 (<https://doi.org/10.1126%2Fscience.ns-9.214S.237>). PMID 17736020 (<https://pubmed.ncbi.nlm.nih.gov/17736020>).
28. Chen, Beichen (2021). *Embeddings for Book Similarities* (<https://www.diva-portal.org/smash/get/diva2:1601084/FULLTEXT01.pdf>) (PDF) (MSc). KTH Royal Institute of Technology.
29. Stamatatos, Efstathios; Kestemont, Mike; Kredens, Krzysztof; Pezik, Piotr; Heini, Annina (2022). "Overview of the Authorship Verification Task at PAN 2022". In Faggioli; Ferro; Hanbury; Potthast (eds.). *CLEF 2022 Labs and Workshops, Notebook Papers* (https://pan.webis.de/publications.html#stamatatos_2022). CEUR-WS. Retrieved September 6, 2022.
30. Neal et al. 2018, p. 5.
31. Gröndahl & Asokan 2020a, p. 3.
32. Kacmarcik & Gamon 2006, p. 444.
33. Mahmood et al. 2019, p. 54.
34. Afroz, Brennan & Greenstadt 2012, p. 461.
35. Gröndahl & Asokan 2020a, p. 28.
36. Neal et al. 2018, p. 6.
37. Potthast, Hagen & Stein 2016, p. 10.
38. Saedi & Dras 2020, p. 181.
39. Gröndahl & Asokan 2020a, p. 21–22.
40. Wang, Juola & Riddell 2022, p. 2.

41. Neal et al. 2018, p. 27.
42. Brennan, Afroz & Greenstadt 2012, p. 2.
43. Zhai et al. 2022, p. 7373.
44. Emmery, Kádár & Chrupała 2021, p. 2388-2389.
45. Argamon, Shlomo, Jussi Karlgren, and James G. Shanahan. Stylistic analysis of text for information access. Papers from the workshop held in conjunction with the 28th Annual International ACM Conference on Research and Development in Information Retrieval, August 13–19, 2005, Salvador, Bahia, Brazil. Swedish institute of computer science, 2005.
46. "The Signature Stylometric System" (<http://www.philocomp.net/texts/signature.htm>). PhiloComp. Retrieved 2014-01-03.
47. "JGAAP" (<http://www.jgaap.com>). JGAAP. 2012-09-04. Retrieved 2012-10-15.
48. "The stylo for R package" (<https://web.archive.org/web/20141221100741/https://sites.google.com/site/computationalstylistics/stylo>). Computational Stylistics Group. 2014-10-24. Archived from the original (<https://sites.google.com/site/computationalstylistics/stylo>) on 2014-12-21. Retrieved 2014-10-24.
49. Eder, Maciej; Rybicki, Jan; Kestemont, Mike (2016). "Stylometry with R: a package for computational text analysis" (<https://journal.r-project.org/archive/2016-1/eder-rybicki-kestemont.pdf>) (PDF). *R Journal*. **8** (1): 107–121. doi:10.32614/RJ-2016-007 (<https://doi.org/10.32614/RJ-2016-007>).
50. Daelemans, Walter & Hoste, Véronique (2013). *STYLENE: an Environment for Stylometry and Readability Research for Dutch* (<http://stylene.be>) (Technical report). CLiPS Technical Report Series. ISSN 2033-3544 (<https://search.worldcat.org/issn/2033-3544>).
51. Yan Qu, James G. Shanahan, and Janyce Wiebe. "Exploring attitude and affect in text: Theories and applications." AAAI Spring Symposium Technical report SS-04-07. AAAI Press, Menlo Park, CA. 2004.
52. Jussi Karlgren, Björn Gambäck, and Pentti Kanerva. "Acquiring (and Using) Linguistic (and World) Knowledge for Information Access." (2002). AAAI Spring Symposium. Technical report SS-02-09. AAAI Press, Menlo Park, CA. 2002.
53. Shlomo Argamon, Shlomo Dubnov, and Julie Jupp. "Style and Meaning in Language, Art, Music, and Design" (2004). AAAI Fall Symposium. Technical report FS-04-07.
54. Potthast, Martin, Benno Stein, Alberto Barrón-Cedeño, and Paolo Rosso. "An evaluation framework for plagiarism detection." In Proceedings of the 23rd international conference on computational linguistics: Posters, pp. 997–1005. Association for Computational Linguistics, 2010.
55. Stamatatos, Efsthios, Walter Daelemans, Ben Verhoeven, Patrick Juola, Aurelio López-López, Martin Potthast, and Benno Stein. "Overview of the Author Identification Task at PAN 2014." In CLEF (Working Notes), pp. 877–897. 2014.
56. Rangel, Francisco, Paolo Rosso, Martin Potthast, and Benno Stein. "Overview of the 5th author profiling task at pan 2017: Gender and language variety identification in twitter." Working Notes Papers of the CLEF (2017).
57. Rangel Pardo, Francisco Manuel, Fabio Celli, Paolo Rosso, Martin Potthast, Benno Stein, and Walter Daelemans. "Overview of the 3rd Author Profiling Task at PAN 2015." In CLEF 2015 Evaluation Labs and Workshop Working Notes Papers, pp. 1–8. 2015.
58. Potthast, Martin, Benno Stein, and Teresa Holfeld. "Overview of the 1st International Competition on Wikipedia Vandalism Detection." In CLEF (Notebook Papers/LABs/Workshops). 2010.

59. Text processing text analysis and generation – text typology and attribution. Proceedings of Nobel symposium 51. Edited by Sture Allén. Stockholm: Almqvist & Wiksell international 1982. Data linguistica, 16. Nobel symposium, 51. ISBN 91-22-00594-3
60. Karlgren, Jussi (2003). "Helander: An Authorship Attribution Case" (<https://jussikarlgren.wordpress.com/2003/01/01/helander-an-authorship-attribution-case/>). Retrieved 4 October 2017.
61. Airoidi, Edoardo M.; Fienberg, Stephen E.; Skinner, Kiron K. (July 2007). "Whose Ideas? Whose Words? Authorship of Ronald Reagan's Radio Addresses" (http://www.stat.columbia.edu/~gelman/stuff_for_blog/Airoidi_PS_Final.pdf) (PDF). *PS: Political Science & Politics*. **40** (3): 501–506. CiteSeerX 10.1.1.190.5798 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.190.5798>). doi:10.1017/S1049096507070874 (<https://doi.org/10.1017%2FS1049096507070874>). S2CID 18730541 (<https://api.semanticscholar.org/CorpusID:18730541>).
62. *Author Unknown* by Gavin McNett Salon November 2, 2000 (http://www.salon.com/2000/11/02/foster_5/)
63. Belluck, Pam (April 10, 1996). "In Unabom Case, Pain for Suspect's Family" (<https://www.nytimes.com/1996/04/10/us/in-unabom-case-pain-for-suspect-s-family.html>). *The New York Times*. Archived (<https://web.archive.org/web/20170810092506/http://www.nytimes.com/1996/04/10/us/in-unabom-case-pain-for-suspect-s-family.html>) from the original on August 10, 2017. Retrieved July 5, 2008.
64. "Study finds a disputed Shakespeare play bears the master's mark" (<https://www.latimes.com/science/sciencenow/la-sci-sn-shakespeare-play-linguistic-analysis-20150410-story.html>). *Los Angeles Times*. 2015-04-10. Retrieved 2015-04-13.
65. Boyd, Ryan L.; Pennebaker, James W. (2015). "Did Shakespeare Write Double Falsehood? Identifying Individuals by Creating Psychological Signatures With Text Analysis" (<https://journals.sagepub.com/doi/full/10.1177/0956797614566658>). *Psychological Science*. **26** (5): 570–582. doi:10.1177/0956797614566658 (<https://doi.org/10.1177%2F0956797614566658>). PMID 25854277 (<https://pubmed.ncbi.nlm.nih.gov/25854277>). S2CID 13022405 (<https://api.semanticscholar.org/CorpusID:13022405>).
66. Jackson, MacDonald P (April 27, 2016). *Who Wrote "The Night Before Christmas"? Analyzing the Clement Clarke Moore Vs. Henry Livingston Question*. McFarland & Co. ISBN 978-1476664439.
67. Fuller, Simon; O'Sullivan, James (2017). "Structure over Style: Collaborative Authorship and the Revival of Literary Capitalism" (<http://www.digitalhumanities.org/dhq/vol/11/1/000286/000286.html>). *Digital Humanities Quarterly*. **11** (1). Retrieved April 20, 2017.
68. Lane, Anthony (June 18, 2018). "Bill Clinton and James Patterson's Concussive Collaboration" (<https://www.newyorker.com/magazine/2018/06/18/bill-clinton-and-james-pattersons-concussive-collaboration>). *The New Yorker*. Retrieved 2018-06-07.
69. "Why you don't need to write much to be the world's bestselling author" (<https://theconversation.com/why-you-dont-need-to-write-much-to-be-the-worlds-bestselling-author-75261>). *The Conversation*. April 3, 2017. Retrieved April 20, 2017.
70. O'Sullivan, James (2018-06-07). "Bill Clinton and James Patterson are co-authors – but who did the writing?" (<https://www.theguardian.com/books/booksblog/2018/jun/07/bill-clinton-james-patterson-on-the-president-is-missing-co-authors>). *The Guardian*. Retrieved 2018-06-07.
71. Savoy, Jacques (2018). "Is Starnone really the author behind Ferrante?" (<https://academic.oup.com/dsh/article/33/4/902/5001585>). *Digital Scholarship in the Humanities*. **33** (4): 902–918. doi:10.1093/llc/fqy016 (<https://doi.org/10.1093%2Fllc%2Ffqy016>).
72. Reuell, Peter: "You say John, I say Paul. But what does stylometry say?" (<https://news.harvard.edu/gazette/story/2018/09/harvard-statistician-examines-beatles-mystery/>)

73. Glickman, Mark; Brown, Jason; Song, Ryan (2019). "(A) Data in the Life: Authorship Attribution in Lennon-McCartney Songs" (<https://doi.org/10.1162%2F99608f92.130f856e>). *Harvard Data Science Review*. 1 (1). arXiv:1906.05427 (<https://arxiv.org/abs/1906.05427>). doi:10.1162/99608f92.130f856e (<https://doi.org/10.1162%2F99608f92.130f856e>). S2CID 189762434 (<https://api.semanticscholar.org/CorpusID:189762434>).
74. The ETSO project. (<http://etso.es/>)
75. "Un monstruo de la naturaleza llamado Lope" (https://www.abc.es/cultura/abci-monstruo-naturaleza-llamado-lope-201811280249_noticia.html) [A monster of nature called Lope]. *abc* (in Spanish). 2018-11-28. Retrieved 2019-08-11.
76. "Rastreadores digitales en el Siglo de Oro" (<https://www.elnortedecastilla.es/culturas/teatro/rastradores-digitales-siglo-20181223133815-nt.html>) [Digital trackers in the Golden Age]. *El Norte de Castilla* (in Spanish). 2018-12-23. Retrieved 2019-08-11.
77. Real, La Tribuna de Ciudad (2019-07-09). "Juan Ruiz de Alarcón aumenta su obra cinco siglos después" (<https://www.latribunadeciudadreal.es/noticia/Z6F846F17-0F70-FEF3-449E11A2A1C6BD36/201907/Juan-Ruiz-de-Alarcon-aumenta-su-obra-cinco-siglos-despues>) [Juan Ruiz de Alarcón increases his work five centuries after]. *La Tribuna de Ciudad Real* (in Spanish). Retrieved 2019-08-11.
78. Migueláñez, Daniel (28 July 2019). "El Holmes de la filología" (https://web.archive.org/web/20200718220716/http://www.psoechamberi.com/esp/tags/suplemento/2019/n6_julio28/08_right_daniel_Miguela%C3%B1ez_01.html). *PSOE Chamberí*. No. 6. p. 8. Archived from the original (http://www.psoechamberi.com/esp/tags/suplemento/2019/n6_julio28/08_right_daniel_Miguela%C3%B1ez_01.html) on 2020-07-18. Retrieved 2019-08-11.
79. "Sor Juana Inés centró las 42 Jornadas de Teatro Clásico" (<https://www.lanzadigital.com/cultura/sor-juana-ines-centro-las-42-jornadas-de-teatro-clasico/>). *Lanza Digital* (in European Spanish). 2019-07-14. Retrieved 2019-08-11.
80. "'La monja alférez' ya no es de Pérez de Montalbán, sino de Ruiz de Alarcón" (<https://www.elnortedecastilla.es/culturas/teatro/monja-alferez-perez-20190710071933-nt.html>) ['La monja alférez' is no longer by Pérez de Montalbán, but by Ruiz de Alarcón]. *El Norte de Castilla* (in Spanish). 2019-07-10. Retrieved 2019-08-11.
81. "Artificial intelligence helps find prominent Spanish playwright Lope de Vega as the author of a play from a manuscript written years after his death" (<https://www.newsendip.com/artificial-intelligence-helps-find-prominent-spanish-playwright-lope-de-vega-as-the-author-of-a-play-from-a-manuscript-written-years-after-his-death/>). *newsendip.com*. 31 January 2023. Retrieved 8 February 2023.
82. Jones, Sam (5 February 2023). "Artificial intelligence uncovers lost work by titan of Spain's 'Golden Age' " (<https://www.theguardian.com/world/2023/feb/05/artificial-intelligence-uncovers-lost-work-by-titan-of-spains-golden-age>). *The Guardian*. Retrieved 8 February 2023.
83. Morales, Manuel (2023-01-31). "La inteligencia artificial atribuye a Lope de Vega una obra anónima del fondo de manuscritos de la Biblioteca Nacional" (<https://elpais.com/cultura/2023-01-31/la-inteligencia-artificial-atribuye-a-lope-de-vega-una-obra-anonima-del-fondo-de-manuscritos-de-la-biblioteca-nacional.html>) [Artificial intelligence attributes an anonymous work from the National Library's manuscript collection to Lope de Vega]. *El País* (in Spanish). Retrieved 2023-02-08.

84. McCarthy, Rachel; O'Sullivan, James (2020). "Who wrote Wuthering Heights?" (<https://academic.oup.com/dsh/article/doi/10.1093/llc/fqaa031/5862913>). *Digital Scholarship in the Humanities*. **36** (2): 383–391. doi:10.1093/llc/fqaa031 (<https://doi.org/10.1093%2Fllc%2Ffqaa031>). hdl:10468/10194 (<https://hdl.handle.net/10468%2F10194>).
85. Ilseman, Harmut (2020) "Phantom Marlowe: Paradigmenwechsel in Autorschaftsbestimmungen des englischen Renaissancedramas". Düren: Shaker, ISBN 978-3-8440-7412-3
86. Ilseman, Harmut (2020). "The Marlowe corpus revisited" (<https://academic.oup.com/dsh/article-abstract/36/2/333/5825419>). *Digital Scholarship in the Humanities*. **36** (2): 333–360. doi:10.1093/llc/fqaa010 (<https://doi.org/10.1093%2Fllc%2Ffqaa010>).
87. Ilseman, Harmut (2021). "A brief supplement to "The Marlowe Corpus Revisited" and Phantom Marlowe" (<https://academic.oup.com/dsh/advance-article-abstract/doi/10.1093/llc/fqab078/6397020>). *Digital Scholarship in the Humanities*. **37** (2): 462–468. doi:10.1093/llc/fqab078 (<https://doi.org/10.1093%2Fllc%2Ffqab078>).
88. Rebora, Simone & Salgaro, Massimo (2022). "Is Felix Salten the Author of the Mutzenbacher Novel (1906)? Yes and no" (<https://journals.sagepub.com/doi/abs/10.1177/09639470221090384>). *Language and Literature: International Journal of Stylistics*. **31** (2): 243–264. doi:10.1177/09639470221090384 (<https://doi.org/10.1177%2F09639470221090384>). S2CID 248135373 (<https://api.semanticscholar.org/CorpusID:248135373>).
89. AI avslöjar: Läckberg har antagligen spökskrivare – skjuter ned anklagelserna. (<https://www.hbl.fi/artikel/fa306045-6387-5b08-8ef8-10c7c923acb0>) *Hufvudstadsbladet*, 27 September 2023 (in Swedish).
90. "Läckberg om rykterna: 'Han petade i meningarna' " (<http://www.hbl.fi/artikel/84301c66-05a2-51ba-8edd-1078456be4be>). *Hufvudstadsbladet* (in Swedish). Helsingfors. 21 December 2023. p. 23.
91. Biber, Douglas. Variation across speech and writing. Cambridge University Press, 1991.
92. Karlgren, Jussi; Cutting, Douglass (1994). "Recognizing text genres with simple metrics using discriminant analysis". *Proceedings of the 15th conference on Computational linguistics -*. Vol. 2. p. 1071. arXiv:cmp-lg/9410008 (<https://arxiv.org/abs/cmp-lg/9410008>). Bibcode:1994cmp.lg...10008K (<https://ui.adsabs.harvard.edu/abs/1994cmp.lg...10008K>). doi:10.3115/991250.991324 (<https://doi.org/10.3115%2F991250.991324>). S2CID 1297432 (<https://api.semanticscholar.org/CorpusID:1297432>).
93. Van Droogenbroeck F. J., "An essential rephrasing of the Zipf-Mandelbrot law to solve authorship attribution applications by Gaussian statistics" (<https://www.academia.edu/40029629>) (2019).
94. Matthews, Robert A. J.; Merriam, Thomas V. N (1993). "Neural Computation in Stylometry I: An Application to the Works of Shakespeare and Fletcher". *Literary and Linguistic Computing*. **8** (4): 203–209. doi:10.1093/llc/8.4.203 (<https://doi.org/10.1093%2Fllc%2F8.4.203>).
95. Merriam, Thomas V. N; Matthews, Robert A. J. (1994). "Neural Computation in Stylometry II: An Application to the Works of Shakespeare and Marlowe". *Literary and Linguistic Computing*. **9** (1): 1–6. doi:10.1093/llc/9.1.1 (<https://doi.org/10.1093%2Fllc%2F9.1.1>).
96. JF Hoorn; SL Frank; W Kowalczyk; F van der Ham (2012-09-03). "Neural network identification of poets using letter sequences". *Literary and Linguistic Computing*. **14** (3): 311–338. doi:10.1093/llc/14.3.311 (<https://doi.org/10.1093%2Fllc%2F14.3.311>).
97. Brocardo, ML; Traore, I; Woungang, I; Obaidat, MS (2017). "Authorship verification using deep belief network systems". *Int J Commun Syst*. **30** (12): e3259. doi:10.1002/dac.3259 (<https://doi.org/10.1002%2Fdac.3259>). S2CID 40745740 (<https://api.semanticscholar.org/CorpusID:40745740>).

98. de Vel, O.; Anderson, A.; Corney, M.; Mohay, G. (2001-12-01). "Mining e-Mail Content for Author Identification Forensics". *SIGMOD Rec.* **30** (4): 55–64. CiteSeerX 10.1.1.408.4231 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.408.4231>). doi:10.1145/604264.604272 (<https://doi.org/10.1145/604264.604272>). ISSN 0163-5808 (<https://search.worldcat.org/issn/0163-5808>). S2CID 1623521 (<https://api.semanticscholar.org/CorpusID:1623521>).
99. Argamon, Shlomo; Koppel, Moshe; Pennebaker, James W.; Schler, Jonathan (2009-02-01). "Automatically Profiling the Author of an Anonymous Text". *Commun. ACM.* **52** (2): 119–123. CiteSeerX 10.1.1.136.9952 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.9952>). doi:10.1145/1461928.1461959 (<https://doi.org/10.1145/1461928.1461959>). ISSN 0001-0782 (<https://search.worldcat.org/issn/0001-0782>). S2CID 5413411 (<https://api.semanticscholar.org/CorpusID:5413411>).
100. "Classification of Instant Messaging Communications for Forensics Analysis – TechRepublic" (<http://www.techrepublic.com/resource-library/whitepapers/classification-of-instant-messaging-communications-for-forensics-analysis/>). *TechRepublic*. Retrieved 2016-01-26.
101. Zhou, L.; Zhang, Dongsong (2004-01-01). "Can online behavior unveil deceivers? - an exploratory investigation of deception in instant messaging". *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. pp. 9 pp.–. doi:10.1109/HICSS.2004.1265079 (<https://doi.org/10.1109/HICSS.2004.1265079>). ISBN 978-0-7695-2056-8. S2CID 7154702 (<https://api.semanticscholar.org/CorpusID:7154702>).

References

- Afroz, Sadia; Brennan, Michael; Greenstadt, Rachel (2012). "Detecting Hoaxes, Frauds, and Deception in Writing Style Online". *2012 IEEE Symposium on Security and Privacy*. pp. 461–475. doi:10.1109/SP.2012.34 (<https://doi.org/10.1109/SP.2012.34>). ISBN 978-1-4673-1244-8.
- Brennan, Michael; Afroz, Sadia; Greenstadt, Rachel (2012). "Adversarial stylometry: Circumventing Authorship Recognition to Preserve Privacy and Anonymity" (https://www1.icsi.berkeley.edu/~sadia/papers/adversarial_stylometry.pdf) (PDF). *ACM Transactions on Information and System Security*. **15** (3): 1–22. doi:10.1145/2382448.2382450 (<https://doi.org/10.1145/2382448.2382450>). S2CID 16176436 (<https://api.semanticscholar.org/CorpusID:16176436>).
- Brennan, Michael Robert; Greenstadt, Rachel. "Practical Attacks Against Authorship Recognition Techniques" (<http://www.aaai.org/ocs/index.php/IAAI/IAAI09/paper/view/257>). *Innovative Applications of Artificial Intelligence*.
- Brocardo, Marcelo Luiz; Issa Traore; Sherif Saad; Isaac Woungang (2013). *Authorship Verification for Short Messages Using Stylometry*. IEEE Intl. Conference on Computer, Information and Telecommunication Systems (CITS). doi:10.1109/CITS.2013.6705711 (<https://doi.org/10.1109/CITS.2013.6705711>).
- Can, Fazli; Patton, Jon M. (2004). "Change of writing style with time". *Computers and the Humanities*. **38** (1): 61–82. CiteSeerX 10.1.1.1.8850 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1.8850>). doi:10.1023/b:chum.0000009225.28847.77 (<https://doi.org/10.1023/b:chum.0000009225.28847.77>). S2CID 38242388 (<https://api.semanticscholar.org/CorpusID:38242388>).
- Emmery, Chris; Kádár, Ákos; Chrupała, Grzegorz (2021). "Adversarial Stylometry in the Wild: Transferable Lexical Substitution Attacks on Author Profiling". *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. pp. 2388–2402. arXiv:2101.11310 (<https://arxiv.org/abs/2101.11310>). doi:10.18653/v1/2021.eacl-main.203 (h

<https://doi.org/10.18653%2Fv1%2F2021.eacl-main.203>). S2CID 231719026 (<https://api.semanticscholar.org/CorpusID:231719026>).

- Gröndahl, Tommi; Asokan, N. (2020a). "Text Analysis in Adversarial Settings: Does Deception Leave a Stylistic Trace?". *ACM Computing Surveys*. **52** (3): 1–36. arXiv:1902.08939 (<https://arxiv.org/abs/1902.08939>). doi:10.1145/3310331 (<https://doi.org/10.1145%2F3310331>). S2CID 67856540 (<https://api.semanticscholar.org/CorpusID:67856540>).
- Hope, Jonathan (1994). *The Authorship of Shakespeare's Plays* (<https://archive.org/details/authorshipofshak00hope>). Cambridge: Cambridge University Press. ISBN 9780521417372.
- Hoy, Cyrus (1956–1962). "The Shares of Fletcher and His Collaborators in the Beaumont and Fletcher Canon (I-VII)". *Studies in Bibliography*. **7–15**.
- Juola, Patrick (2006). "Authorship Attribution" (<https://web.archive.org/web/20201024054632/http://www.mathcs.duq.edu/~juola/papers.d/fnt-aa.pdf>) (PDF). *Foundations and Trends in Information Retrieval*. **1** (3): 3. CiteSeerX 10.1.1.219.1605 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.219.1605>). doi:10.1561/15000000005 (<https://doi.org/10.1561%2F15000000005>). Archived from the original (<http://www.mathcs.duq.edu/~juola/papers.d/fnt-aa.pdf>) (PDF) on 2020-10-24. Retrieved 2008-11-13.
- Kacmarcik, Gary; Gamon, Michael (17 July 2006). "Obfuscating document stylometry to preserve author anonymity" (<https://aclanthology.org/P06-2058/>). *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*. pp. 444–451.
- Kenny, Anthony (1982). *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities*. Oxford: Pergamon Press.
- Mahmood, Asad; Ahmad, Faizan; Shafiq, Zubair; Srinivasan, Padmini; Zaffar, Fareed (2019). "A Girl Has No Name: Automated Authorship Obfuscation using Mutant-X" (<https://doi.org/10.2478%2Fpopets-2019-0058>). *Proceedings on Privacy Enhancing Technologies*. **2019** (4): 54–71. doi:10.2478/popets-2019-0058 (<https://doi.org/10.2478%2Fpopets-2019-0058>). S2CID 197621394 (<https://api.semanticscholar.org/CorpusID:197621394>).
- Neal, Tempestt; Sundararajan, Kalaivani; Fatima, Aneez; Yan, Yiming; Xiang, Yingfei; Woodard, Damon (2018). "Surveying Stylometry Techniques and Applications". *ACM Computing Surveys*. **50** (6): 1–36. doi:10.1145/3132039 (<https://doi.org/10.1145%2F3132039>). S2CID 21360798 (<https://api.semanticscholar.org/CorpusID:21360798>).
- Potthast, Martin; Hagen, Matthias; Stein, Benno (2016). *Author Obfuscation: Attacking the State of the Art in Authorship Verification* (<https://ceur-ws.org/Vol-1609/16090716.pdf>) (PDF). Conference and Labs of the Evaluation Forum.
- Romaine, Suzanne (1982). *Socio-Historical Linguistics*. Cambridge: Cambridge University Press.
- Saedi, Chakaveh; Dras, Mark (December 2020). "Large Scale Author Obfuscation Using Siamese Variational Auto-Encoder: The SiamAO System" (<https://aclanthology.org/2020.starsem-1.19>). *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*. pp. 179–189.
- Samuels, M. L. (1972). *Linguistic Evolution: With Special Reference to English*. Cambridge: Cambridge University Press.
- Schoenbaum, Samuel (1966). *Internal Evidence and Elizabethan Dramatic Authorship: An Essay in Literary History and Method*. Evanston, IL, USA: Northwestern University Press.
- Van Droogenbroeck, Frans J. (2016) "Handling the Zipf distribution in computerized authorship attribution (<https://www.academia.edu/24147736/>)"

- Van Droogenbroeck, Frans J. (2019) "An essential rephrasing of the Zipf-Mandelbrot law to solve authorship attribution applications by Gaussian statistics (<https://www.academia.edu/40029629/>)"
- Wang, Haining; Juola, Patrick; Riddell, Allen (2022). "Reproduction and Replication of an Adversarial Stylometry Experiment". [arXiv:2208.07395](https://arxiv.org/abs/2208.07395) (<https://arxiv.org/abs/2208.07395>). {{cite journal}}: Cite journal requires |journal= (help)
- Zenkov, Andrei V. (2018). "A Method of Text Attribution Based on the Statistics of Numerals". *Journal of Quantitative Linguistics*. **25** (3): 256–270. doi:10.1080/09296174.2017.1371915 (<https://doi.org/10.1080%2F09296174.2017.1371915>). S2CID 49692378 (<https://api.semanticscholar.org/CorpusID:49692378>).
- Zhai, Wanyue; Rusert, Jonathan; Shafiq, Zubair; Srinivasan, Padmini (2022). "A Girl Has A Name, And It's ... Adversarial Authorship Attribution for Deobfuscation". *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pp. 7372–7384. [arXiv:2203.11849](https://arxiv.org/abs/2203.11849) (<https://arxiv.org/abs/2203.11849>). doi:10.18653/v1/2022.acl-long.509 (<https://doi.org/10.18653%2Fv1%2F2022.acl-long.509>). S2CID 248780012 (<https://api.semanticscholar.org/CorpusID:248780012>).

Further reading

See also the academic journal *Literary and Linguistic Computing*, now *Digital Scholarship in the Humanities* (published by the [University of Oxford](#)) and the *Language Resources and Evaluation* journal (previously *Computers and the Humanities*).

External links

- Association for Computers and the Humanities (<http://www.ach.org/>)
 - Literary and Linguistic Computing (<https://web.archive.org/web/20050613011644/http://llc.oxfordjournals.org/>)
 - Computational Stylistics Group (<https://computationalstylistics.github.io/>)
 - *Signature* Stylometric System (<http://www.philocomp.net/texts/signature.htm>)
 - JGAAP Authorship Attribution Program (<https://github.com/evllabs/JGAAP>)
 - Uncovering the Mystery of J.K. Rowling's Latest Novel (<https://www.npr.org/2013/07/26/205794448/uncovering-the-mystery-of-j-k-rowlings-latest-novel>)
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Stylometry&oldid=1249878952>"