Stylistic Analysis

Fotis Jannidis

https://orcid.org/0000-0001-6944-6113

Abstract In the humanities, the term *style* usually refers to a systematic choice of means of expression in a sign system, e.g. language, that is characteristic of an author or a genre or an epoch, etc. Stylometry uses these features, in the case of texts, e.g. lexis, syntax, semantics, and text structure, to attribute authorship, to profile authors, or even to assign periods and genres by means of quantitative methods such as clustering or classification. Stilometric methods were applied to religious texts very early on and, as can be seen from the history of the analysis of the Pauline epistles, reflect important stages in the development of stilometry from univariate to multivariate analysis, which today is usually carried out in a probabilistic framework with numerous test repetitions.*

Keywords Stylometry, Authorship, Genre

In many humanities disciplines, the term *style* is used to describe a use of language and other means of expression determined by an author's conscious or subconscious choice and not dictated by the content alone. Typical uses can be distinguished: style is understood as an aspect of form and refers to the entire text or even to an author's oeuvre. Style is often understood as an expression of the author's individuality, which can be seen in the selection of individual elements specific to an author. This selection might deviate systematically and characteristically from the prevailing conventions (Hermann et al. 2015). Stylistic analysis is then usually understood, especially in the study of literary texts, to indicate the identification of linguistic devices typical of the respective style, i.e., specific sentence constructions, verb constructions, adjective sequences, etc. (Leech & Short 2007).

1. Stylometry

Stylometry is the corpus-based analysis of style using quantitative methods. Compared to a qualitative description of the characteristics of a style, the aim of stylometry is much simpler: to assign a text to a group of other texts based on stylistic characteristics. Stylometry is one of the most productive fields in the Digital Humanities with a

* This chapter, including quotations in foreign languages, was translated from German by Brandon Watson.

Fotis Jannidis: Stylistic Analysis, in: Christopher A. Nunn and Frederike van Oorschot (eds.): Compendium of Computational Theology 1. Introducing Digital Humanities to Theology, Heidelberg: heiBOOKS, 2024, S. 189–202. DOI: https://doi.org/10.11588/heibooks.1521.c21948

long history (Holmes 1998) and many different approaches and applications in many academic disciplines and fields of application. Research overviews with large bibliographies can be found in Juola (2008), Stamatatos (2009), and Neal (2017).

The focus is primarily on tasks dealing with aspects of authorship:

- Authorship attribution: An anonymous text is assigned to an author from a set of possible candidates. This procedure can only work if texts exists where the authorship of the texts is undoubted.
- Authorship verification: Two texts are given; the task is to determine whether they were written by the same author.
- Author profiling, also known as *sociolinguistic profiling* (Grant 2022, 20): The task is to determine the gender, age, native language, personality traits, illnesses such as dementia (Hirst & Wei Feng 2012), or profession of the author based on one or more texts.
- Multi-authorship: In texts written by several authors, the sections should be assigned to their respective authors.
- Stylochronometry: The task is to determine the chronological order of a series of texts based on stylistic characteristics, such as determining the chronology of an author's texts (Seminck et al. 2022).

In addition to investigating these aspects of authorship, stylometry also addresses several other aspects, such as the attribution of a text to a period or genre based on its stylistic characteristics (Jannidis & Lauer 2014), the recognition of the physical writing situation (dictation, handwriting, typewriting, etc., Hoover 2021), the visibility of translators (Rybicki 2021), and character idiolects (Burrows 1980). Stylometry is also used to detect plagiarism and to reorganize texts so that their style no longer corresponds to that of the author. There is also a relatively independent sub-field of stylometry within forensic linguistics, which deals for example with the classification of blackmail letters. Forensic stylistics also uses corpora to identify author-specific usage, such as a deviation in style. However, given the limited amount of text in most cases and the parameters of the legal system, stylistics has developed its own practices (Grant 2022).

Authorship attribution, which is one of the most frequently processed tasks, takes place in very different contexts. In the simplest case, it involves assigning an unknown longer text to one of two authors from whom numerous longer texts of the same genre and from the same period are available. However, variations of these factors quickly lead to much more complex research groupings. For example, there could be a high number of possible authors. The question of whether the author of the text in question is certainly among the candidates *(closed set)* or not *(open set)* is decisive for the methods. In addition, often only texts from other genres, other time periods, or in different languages are available. At times, there are only short texts available, which complicates the statistical analysis.

The quantitative analysis of style uses text features to group texts according to authorship *(cluster analysis)* or to classify them *(supervised learning).* These features can be divided into four groups (Neal 2017, 11):

- Lexical, such as word frequency, word n-grams, number of letters, or average word length
- Syntax, such as frequency of punctuation marks, sentence length, complexity of sentence structure
- Semantics, such as synonyms and topic distribution
- Text Structure, such as length of paragraphs, paragraph indentations, and font

While the step in the quantitative analysis of style concentrated on rare words and sentence constructions as individualizing features, it has long been common practice to consider the distributions in the use of the *most frequent* features as particularly discriminative. Research has shown that there is no clearly defined set of features that leads to the best results in each of the tasks and in each of the many possible interrelations. In the beginnings of stylometry, a single feature was collected and compared; however, since the 1980s, the more common practice has been to use several features at once.

The *Shared Tasks* for stylometry from PAN[1] have been available since 2011. *Shared Tasks* are an established method in computational linguistics for summarizing and further developing research knowledge on a problem. For this purpose, a data set is published together with a task that is to be solved using the data. For example, one of the tasks in 2019 was to identify authors of fanfiction from a specific fandom; however, only their works on other fandoms were known *(cross-domain),* and the task was formulated as an *open-set attribution.* The submitted solutions are then evaluated using data that the participants have not yet seen, and the result and the data sets are published so that later studies can compare their results directly with those of the competition. These PAN *Shared Tasks* have furthered standardization and promoted knowledge about the methodological state of stylometry. They remain cutting edge in the field.

In the context of stylometry, the specific linguistic form of a text is the result of several factors. The factors include: The theme of the text, the general language usage at the time, which is often abbreviated as period style, the linguistic features typical of the text type (genre style), and the author-specific linguistic choices (authorial style). The aim of a stylometric study is to make a classification based on linguistic characteristics that do not essentially depend on the topic, thus considering authorship attribution, and to control other factors like epoch and genre. The factors authorship, genre, and epoch/period have a simultaneous influence on the style; accordingly,

1 See https://pan.webis.de (Accessed: 16 June 2024).

research designs must control the respective non-questionable aspects as possible confounding variables. In the case of authorship attribution, one can control by including texts from the same period and genre in the corpus.

After compiling all relevant texts in digital form, the N features deemed relevant, e.g., words, punctuation marks, letter n-grams etc., are extracted from all texts, where N stands for the number of features. Each text is then represented as a vector with the N values of these features. Geometrically, each of these vectors can be seen as a point in an N-dimensional space. Distance measures then allow one to determine the distance between these points, each of which represents a text. Let's assume the simplest constellation. There is a text X with unknown authorship and two groups of texts, G_A and G_B, where the first is by author A and the second by author B, and we know that one of the two is the author of X. We now select the text features so that the texts have a relatively small distance within the group and a clear distance between the groups. If we now add the text X in question, we will likely observe that the text has a relatively small distance to one group and a relatively large distance to the other – if the above-mentioned conditions are met, that the texts are of the same genre and from the same period, and that the text in question really is by one of the two authors.

However, a decision in favor of a certain set of features that optimally differentiates the texts of the candidates can only be made if it is certain that only one of the candidates can be the author *(closed set).* In most cases, however, there is no certainty; it could also be another author who is not among the candidates *(open set).* In this case, however, the text features cannot be optimally selected. From a practical research perspective, the procedure of selecting features that are considered optimal for a certain data set and only one evaluation method has also not been successful as the results are hardly verifiable and comparable. For this reason, newer approaches are based not only on working with a specific set of characteristics and then carrying out a distance measurement; rather, the measurement is repeated with a different selection of features (Juloa 2015).

In their methodological study on authorship attribution under *open set* conditions, Sperling et al. have shown that Stephen King can be clearly identified as the author of the novels published under the name Richard Bachmann (Sperling et al. 2023). To do this, they collected a large number of characteristics (120,000) based on the standard imposter method (Koppel & Winter 2014) and repeatedly drew a smaller sample (10,000). They then use these 10,000 sample features to measure the distance between an excerpt of a novel by Bachmann and randomly selected excerpts from novels by four authors of horror novels including King. This process is repeated thousands of times for each excerpt of a novel by Bachmann. Each of these measurements results in different distances between the candidate authors and Bachmann, which can be noted as a ranking. It turns out that King ranks first more often than the other authors, i.e., is closest to the Bachmann text.

In analyses of this kind, as is already known from the long tradition of *Information Retrieval*, the determination of the distance measure, with which the distances

between the vectors are measured, is decisive. A particularly important measure of this type is *Burrows Delta* (Burrows 2002), which has proven to be particularly robust (Evert et al. 2017). In the case of very large amounts of text, supervised machine learning methods have also proven to be very good (Savoy 2020).

The assumption that every author, in fact every person, when using language, has a unique way of using language, or in linguistic terminology an idiolect, is one of the basic theses of stylometry (Nini 2023). However, this assumption is not necessary to justify the way stylometry functions. It is sufficient to assume that there are discriminative features for a given set of authors. The first point is a difficult assumption to prove; the second point has now been empirically demonstrated for many constellations, but no claim to general validity can be derived. Overall, the findings of stylometry, namely that an author's texts exhibit demonstrable commonalities, are a problem for theories that model individuality as a mere discourse phenomenon or as an emergent phenomenon of interactions (Jannidis 2014, Grand & McLeod 2018). However, the metaphor of the *linguistic fingerprint* sometimes used in popular contexts is misleading. On the one hand, the idiolect develops over time, however slowly, and on the other hand, a corpus-specific and task-specific selection of linguistic features is used for the processing of stylometric tasks and not a constant identity marker. Overall, stylometry is still largely a science driven by individual studies, but in the last ten years it has increasingly been striving for a theoretical and methodological foundation (cf. Juola 2015, Nini 2023).

2. Using Stylistic Analysis for Religious Texts

In 1851, the mathematician Augustus de Morgan, who is now regarded as one of the founders of formal logic, speculated on the possibility of distinguishing the letter of Hebrews from the letters of Paul based on the average word length of the letters (De Morgan 1882, 215f.). This hypothesis is still regarded as one of the origins of stylometry (Holmes 1998, 112). Stylometric methods were then applied to biblical texts at an early stage, which showed a promising trajectory for the methods from the outset. As Radday wrote in a 1973 study, "certain problems that have vexed Biblical scholars for centuries can at least be approached and possibly solved once and for all by quantification" (Radday 1973, 273). Particularly in the case of controversies about authorship, stylometric methods are a good way of integrating a further source of information into the judgment process alongside historical and linguistic information (Oakes 2014). A brief overview of the question of the authorship of the Pauline letters shows how stylometric methods have been used and how the methods have changed over time.

The question of whether the fourteen Pauline letters were really written by the historical Paul or whether some were written by one or more other authors, whether

the two letters to the Corinthians, Galatians, and Romans were actually written by one author, is one of the oldest topics of theological stylometry. Authorship questions have been discussed since the 18th century, not least because theologians perceived clear stylistic differences. Some of the research on this topic has also used simple quantitative arguments relating to the frequency of a word when comparing texts. Robert Morgenthaler's work has promulgated this comparison, listing the frequencies of all the words in the books of the New Testament (Morgenthaler 1958). One example can be found in Bujard's work on Colossians. Bujard only uses the simple counts (often based on Morgenthaler) and does not carry out a statistical test to see whether the observed differences are significant (Bujard 1973). The theologian Andrew Q. Morton broke with this established way of working in two ways: he used the computer, and he used statistical methods to support the reliability of his comparisons.[2] Morton's 1963 newspaper article was controversial because he claimed, based on his stylometric studies, that only four letters were written by the same author (1. and 2. Corinthians, Galatians, and Romans), whereas the historical scholars of his time assumed that the first letter to the Thessalonians, Philippians, and Philemon were also written by the historical Paul. The fierce criticism prompted him to present his method and results in detail (Morton & McLeman 1966).

Morton's work has yet to be fully appreciated, as the view of his achievement is overshadowed by his later work. His *Cumulative Sum* (CUSUM) technique was supposed to show clear authorship attribution even of short texts and was also used in court in Great Britain in the 1990s. However, the public success was countered by the doubts of specialist colleagues, who considered the statistical procedure and its application by Morton to be highly unreliable, and with good reason (Holmes 1998, 114). This perspective still distorts the reception of Morton's early work.

In his 1966 study, Morton essentially used two text features: the average sentence length and frequent words, so-called function words. He excludes the articles, which occur most frequently, since they depend too much on the subject of the text, and concentrates on *kai* (and), the particles *de, en* (in), *einai* (to be), and *autos.* First, he uses a corpus of ancient Greek texts to determine whether these function words are suitable for distinguishing between authors of a text by dividing the texts into segments of equal size, recording the frequency of the respective function word, and then using Chi^2 to calculate the probability of whether the observed difference between the segments is coincidental – in the latter case, this would speak in favor of two authors. All tests confirm the result: 1. Corinthians, 2. Corinthians, Galatians and Romans were written by one author. Philemon is problematic given its brevity, but the data also speak in favor of it belonging to this group. Morton's results confirm theses that were already formulated in the 19th century by F.C. Baur (Neumann 1990, 2ff.). A major

2 Neumann lists stylometric works that were not considered, yet worked with statistical analysis even before Morton (Neumann 1990, Chap. II).

weakness of his approach is the fact that each study only includes one variable, such as the frequency of *kai*, whereas multivariate analyses of the kind described above have become the standard since the 1980s at the latest.

Neumann, whose work was not published until 1990, yet was probably completed in 1980, collects 617 features, including lexical, syntactic, and other features. He carries out a detailed preliminary investigation into the extent to which these features divide a text corpus 100 % correctly in a discriminant analysis and then selects only the four features (word length, indefinite pronouns, words beginning with the letter *tau*, and the position of the first noun in the sentence) for his study of the Pauline epistles. His conclusion is that the disputed letters – Colossians, Ephesians, 2. Thessalonians – are attributed to Paul. The requirement that the features should divide the data 100 % correctly would currently be labeled as overfitting, as the feature selection for the textual data set fit exactly, yet it remains unclear how well they fit the actual object of study.

The philosopher Antony Kenny, who wrote an introduction to stylometry and its statistics in 1982 (Kenny 1982), published his relevant work on the New Testament four years later (Kenny 1986), in which he also deals with the problem of Paul's letters. He selects most of them from a set of 99 features that he developed for his work; these include conjunctions, particles, prepositions, the cases of nouns, pronouns, and adjectives to differentiate the texts. Kenny analyzes each of these features individually and then tests whether the frequencies differ significantly in the texts in question. His results are presented with caution. He concludes that some of the features in some letters seem to indicate a different authorship: the variance observed in twelve of the letters could also be explained by the fact that they are the work of a single, extremely versatile author. In his review of Kenny, Mealand claims that the result is not insignificantly dependent on his task, namely whether a text is particularly divergent, and shows that the correlations Kenny lists can also be evaluated so that the close relationship between Romans, 1. Corinthians, 2. Corinthians, and Galatians becomes clear (Mealand 1988).

The studies in the 1990s use multivariate statistical methods. Greenwood groups the texts based on the 10, 20, and 30 most frequent words and finds that the clusters are very similar (Greenwood 1993). He also uses a method that not only forms clusters, but also allows him to determine whether the clusters overlap in high-dimensional space, which would indicate that they are not separable units. He draws this conclusion even if the structure is not preserved in the PCA, a dimension reduction method that represents the high-dimensional data in two dimensions. In other words, there is an information structure that enables a clear separation of the texts, but it is only visible in high-dimensional space. Overall, however, his results show a strengthening of Morton's thesis, revealing that the function words make authorship attribution possible, a thesis "not comfortably assimilated within the spirit of classical scholarship" (217), despite being established in literary studies at the same time by the work of Burrows.

Ledger (1995) divides the texts into 1000-word sections and for each section collects 29 features that have previously proved useful in Plato's analysis, 19 of which are the proportion of words with a particular letter, 9 with the proportion of words with a particular final letter, and the standard measure of lexical complexity, type-token ratio. He also uses PCA and examines the resulting groupings. He recognizes a central cluster with 1. Corinthians, 2. Corinthians, Galatians, Philippians, 2. Thessalonians, and Romans as well as a second group with Colossians, Ephesians, and Hebrews; the others cannot be clearly assigned. His attempts to identify the characteristics that lead to the special position of Hebrews and to relate these in turn to the texts are noteworthy.

Like Neumann, Mealand uses a discriminant and factor analysis to reduce dimensions in his study, which also examines other NT letters in addition to the Pauline letters. This is based on 25 features, including the most frequent words, which have been shown to be distinctive in earlier literary studies and in Neumann's work. His results confirm some expectations, but in other respects run counter to what theological research assumes. In the case of the Pauline epistles, the first section of Romans is grouped with Colossians and Ephesians, but not all results are evident in the studies.

One of the most recent studies (Savoy 2019) tests three hypotheses: 1) that only the known four letters are from Paul; 2) that the usual seven letters are from Paul; and 3) that ten of the letters are from Paul, while 1. Timothy, 2. Timothy, Titus, and Heb are not Pauline. Based on all words that occur more than twice, three distance measures are applied to the 21 texts (Pauline letters + other letters of the Bible) and the result is shown hierarchically grouped in a dendogram. Even when changing the distance measures, there is confirmation of the four-letter hypothesis in the Pauline letters – Colossians and Ephesians were written by one author. Savoy checks his results using the so-called Imposter method (Koppel & Winter 2014), which is now considered the highest standard of stylometric technique, as it raises the distance measures in numerous iterations based on new subsets of the features, as described above. On the one hand, it is confirmed that Colossians and Ephesians were written by the same author and that the usual four form a group, but this group has a clear connection to 1. and 2. Thessalonians as well as to Philippians. On the one hand, this result goes beyond the thesis that only four letters were written by Paul, but it does not confirm the usual thesis that seven letters were written by Paul. Savoy also concludes negatively that none of the three hypotheses can be confirmed.

Those critical of the stylometric studies of Paul's letters point out that the different quantity and type of quotations from other texts could falsify the values in the case of lexis (however, some stylometricians remove the quotations from the texts), that the subsequent insertion of punctuation marks could make the use of sentence length as a feature problematic, and also that the variance could be explained by the use of different scribes. Overall, it has become clear that the stylometric studies of Paul's letters are an accurate reflection of the development of the method: from

simple frequency values to univariate methods and multivariate statistics, even the criticism of the method is indicative in many ways.

Perhaps apart from the question of Q, a presumed source for the Gospels of Matthew and Luke (Poirier 2008, Oakes 2014, 153ff., Mealand 2011), few other problems in Christian theology have attracted as much attention from stylometrics as the Pauline epistles. Nevertheless, many questions concerning the Old and New Testaments have been addressed using stylometric methods, such as whether the first and second letters of Peter really have different authors (House 2002), or whether the doubted unity of the book of Isaiah can be illuminated more precisely using stylometrics (Radday 1973).

The stylometric analysis of religious texts has also dealt with a series of questions beyond the texts of the Bible, which can only be mentioned here. For example, possible authors of the Book of Mormon have been identified (Holmes 1992, Jockers et al. 2008), there have been studies examining whether the author of the Koran is identical to the author of the Hadith, those investigating the reception tradition of the sayings and actions of the Prophet Mohammed (Sayoud 2012), studies looking at whether the Nahj Al-Balagha, which is attributed to Imam Ali, can be attributed to two authors (Sarwar & Mohamed 2022), and some studies have analyzed early Buddhist literature that came from India and was translated into Chinese (Hung et al. 2010).

Literature Cited

Bujard, W. (1973). *Stilanalytische Untersuchungen zum Kolosserbrief.* Göttingen: Vandenhoek & Ruprecht.

Burrows, J.F. (1987). *Computation into Criticism. A Study of Jane Austen's Novels and an Experiment in Method.* Oxford: Clarendon Press.

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

De Morgan, S.E. (1882). *Memoir of Augustus De Morgan.* London: Longmans, Green. URL: http://archive.org/details/memoirofaugustusoodemouoft (Accessed: 16 June 2024).

Evert, S., Proisl, Th., Jannidis, F., Reger, I., Pielström, S., Schöch, Ch., & Vitt, Th. (2017). Understanding and explaining Delta measures for authorship attribution, *Digital Scholarship in the Humanities,* 32(2), 4–16.

Grant, T. (2022). *The Idea of Progress in Forensic Authorship Analysis.* Cambridge: Cambridge University Press.

Greenwood, H.H. (1993). St Paul Revisited. Word Clusters in Multidimensional Space, *Literary and Linguistic Computing,* 8(4), 211–219. DOI: https://doi.org/10.1093/llc/8.4.211 (Accessed: 16 June 2024).

Hirst, G., & Wei Feng, V. (2012). Changes in Style in Authors with Alzheimer's Disease, *English Studies,* 93(3), 357–370. DOI: https://doi.org/10.1080/0013838X.2012.668789 (Accessed: 16 June 2024).

Holmes, D. (1992). A Stylometric Analysis of Mormon Scripture and Related Texts, *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 155(1), 91–120.

Id. (1998). The Evolution of Stylometry in Humanities Scholarship, *Literary and Linguistic Computing,* 13 (3), 111–117.

Hoover, D.L. (2021). *Modes of Composition and the Durability of Style in Literature.* New York/London: Routledge.

House, M.A. (2002). *Stylometry and the Authorship of Second Peter.* Pasadena, California: Proquest.

Hung, J.-J., Bingenheimer, M., & Wiles, S. (2010). Quantitative evidence for a hypothesis regarding the attribution of early Buddhist translations, *Literary and Linguistic Computing,* 25(1), 119–134.

Jannidis, F. (2014). Der Autor ganz nah. Autorstil in Stilistik und Stilometrie. In M. Schaffrick & M. Willand (Eds.), *Theorien und Praktiken der Autorschaft* (pp. 169–195). Berlin: de Gruyter.

Id., & Lauer, G. (2014). Burrows Delta and its Use in German Literary History. In M. Erlin & L. Tatlock (Eds.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century* (pp. 29–54). Rochester: Camden House.

Jockers, M.L., Witten, D.M., & Criddle, C.S. (2008). Reassessing authorship of the Book of Mormon using delta and nearest shrunken centroid classification*, *Literary and Linguistic Computing,* 23(4), 465–491.

Juola, P. (2008). Authorship Attribution, *Foundations and Trends in Information Retrieval,* 1(3), 233–334.

Id. (2015). The Rowling Case. A Proposed Standard Analytic Protocol for Authorship Questions, *Digital Scholarship in the Humanities,* 30(1), 100–113.

Kenny, A. (1982). *Computation of Style.* Oxford/Frankfurt: Pergamon Press.

Id. (1986). *A Stylometric Study of the New Testament.* Oxford: Clarendon Press.

Koppel, M., & Winter, Y. (2014). Determining If Two Documents Are Written by the Same Author, *Journal of the Association for Information Science and Technology,* 65(1), 178–187.

Ledger, G. (1995). An Exploration of Differences in the Pauline Epistles using Multivariate Statistical Analysis, *Literary and Linguistic Computing,* 10(2), 85–97.

Leech, G., & Short, M. (2007). *Style in Fiction. A Linguistic Introduction to English Fictional Prose.* 2nd ed. New York: Routledge.

Mealand, D.L. (1988). *Review of A Stylometric Study of the New Testament,* von A. Kenny. *The Journal of Theological Studies,* 39(1), 194–196.

Id. (1996). The Extent of the Pauline Corpus. A Multivariate Approach, *Journal for the Study of the New Testament,* 18(59), 61–92.

Id. (2011). Is There Stylometric Evidence for Q?, *New Testament Studies,* 57(4), 483–507.
Morgenthaler, R. (1958). *Statistik des neutestamentlichen Wortschatzes.* Zürich/Frankfurt a.M.: Gotthelf-Verlag.
Morton, A.Q., & McLeman, J. (1966). *Paul, the Man and the Myth. A Study in the Authorship of Greek Prose.* London: Hodder and Stoughton.
Neal, T., Sundarajan, K., Fatima, A., Yan, Y., Xiang, Y, & Woddard, D. (2017). Surveying Stylometry Techniques and Applications, ACM Computing Surveys, 50(6.86), 1–36. DOI: https://doi.org/10.1145/3132039 (Accessed: 16 June 2024).
Neumann, K.J. (1990). *The Authenticity of the Pauline Epistles in the Light of Stylostatistical Analysis.* Atlanta, Ga.: Scholars Press.
Nini, A. (2023). *A Theory of Linguistic Individuality for Authorship Analysis.* Cambridge: Cambridge University Press.
Oakes, M.P. (2014). Stylometric Analysis of Religious Texts. In id. (Ed.), *Literary Detective Work on the Computer* (pp. 149–206). Amsterdam: John Benjamins.
Poirier, J.C. (2008). Statistical Studies of the Verbal Agreements and Their Impact on the Synoptic Problem, *Currents in Biblical Research,* 7(1), 68–123.
Radday, Y.Th. (1973). *The Unity of Isaiah in the Light of Statistical Linguistics.* Hildesheim: H.A. Gerstenberg.
Rybicki, J. (2021). A Third Glance at a Stylometric Map of Native and Translated Literature in Polish. In M. Heydel & Z. Ziemann (Eds.), *Retracing the History of Literary Translation in Poland* (pp. 247–261). New York/London: Routledge.
Sarwar, R., & Mohamed, E. (2022). Author verification of Nahj Al-Balagha, *Digital Scholarship in the Humanities,* 37(4), 1210–1222.
Savoy, J. (2019). Authorship of Pauline Epistles Revisited, *Journal of the Association for Information Science and Technology,* 70(10), 1089–1097.
Id. (2020), *Machine Learning Methods for Stylometry. Authorship Attribution and Author Profiling.* Cham: Springer. DOI: https://doi.org/10.1007/978-3-030-53360-1 (Accessed: 16 June 2024).
Sayoud, H. (2012). Author discrimination between the Holy Quran and Prophet's statements, *Literary and Linguistic Computing,* 27(4), 427–444.
Seminck, O., Gambette, Ph., Legallois, D., & Poibeau, Th. (2022). The Evolution of the Idiolect over the Lifetime. A Quantitative and Qualitative Study of French 19th Century Literature, *Journal of Cultural Analytics,* 7(3), no pag. DOI: https://doi.org/10.22148/001c.37588 (Accessed: 16 June 2024).
Sperling, D.H.M., Kestemont, M., & Neyt, V. (2024). The Authorship of Stephen King's Books Written Under the Pseudonym "Richard Bachman". A Stylometric Analysis, *Journal of Computational Literary Studies,* 2(1), no pag. DOI: https://doi.org/10.5281/zenodo.8397200 (Accessed: 16 June 2024).
Stamatatos, E. (2009). A Survey of Modern Authorship Attribution Methods, *Journal of the American Society for Information Science and Technology,* 60(3), 538–556.

Literature Related

Abakuks, A. (2015). *The Synoptic Problem and Statistics.* London/New York: CRC Press.

Adams, L.L., & Rencher, A.C. (1974). A Computer Analysis of the Isaiah Authorship Problem, *BYU Studies Quarterly,* 15(1), 95–102.

Adams, L.L. (1984). A Scientific Analysis of Isaiah Authorship. In M.S. Nyman & Ch.D. Tate Jr. (Eds.), *Isaiah and the Prophets. Inspired Voices from the Old Testament* (pp. 151–164). Provo, UT: Religious Studies Center, Brigham Young University. URL: https://rsc.byu.edu/isaiah-prophets/scientific-analysis-isaiah-authorship (Accessed: 16 June 2024).

Alqurneh, A., Mustapha, A., Murad, M.A.A., & Sharef, N.M. (2016). Stylometric model for detecting oath expressions. A case study for Quranic texts, *Digital Scholarship in the Humanities,* 31(1), 1–20. DOI: https://doi.org/10.1093/llc/fqu038 (Accessed: 16 June 2024).

Barr, G.K. (2003). Two Styles in the New Testament Epistles, *Literary and Linguistic Computing,* 18(3), 235–248. DOI: https://doi.org/10.1093/llc/18.3.235 (Accessed: 16 June 2024).

Denaux, A. (2006). Style and Stylistics, with special reference to Luke, *Filologia neotestamentaria,* 19(37–38), 31–51.

Eder, M. (2013). Computational stylistics and Biblical translation. How reliable can a dendrogram be. In T. Piotrowski & L. Grabowski (Eds.), *The translator and the computer* (pp. 155–170). Wroclaw: WSF.

Elewa, A. (2019). Authorship verification of disputed Hadiths in Sahih al-Bukhari and Muslim, *Digital Scholarship in the Humanities,* 34(2), 261–276. DOI: https://doi.org/10.1093/llc/fqy036 (Accessed: 16 June 2024).

Göransson, E., Maurits, L., Dahlmann, B., Åkerman Sarkisian, K., Rubenson, S., & Dunn, M. (2023). Improved distance measures for 'fixed-content miscellanies'. An adaptation for the collections of sayings of the desert fathers and mothers, *Digital Scholarship in the Humanities,* 38(1), 127–150. DOI: https://doi.org/10.1093/llc/fqac025 (Accessed: 16 June 2024).

Grant, T., & MacLeod, N. (2018). Resources and constraints in linguistic identity performance. A theory of authorship, *Language and Law,* 5(1), 80–96.

Greenwood, H.H. (1995). Common Word Frequencies and Authorship in Luke's Gospel and Acts, *Literary and Linguistic Computing,* 10(3), 183–187. DOI: https://doi.org/10.1093/llc/10.3.183 (Accessed: 16 June 2024).

Herrmann, J.B., van Dalen-Oskam, K., & Schöch, Ch. (2015). Revisiting Style, a Key Concept in Literary Studies, *Journal of Literary Theory,* 9(1), 25–52. DOI: https://doi.org/10.1515/jlt-2015-0003 (Accessed: 16 June 2024).

Kilpatrick, G.D. (1988). Review of A Stylometric Study of the New Testament, *Novum Testamentum,* 30(4), 373–375. DOI: https://doi.org/10.2307/1560626 (Accessed: 16 June 2024).

Kinzig, Wolfram. (1990). *In Search of Asterius. Studies on the Authorship of the Homilies on the Psalms.* Göttingen: Vandenhoeck u. Ruprecht [= *Forschungen zur Kirchen- und Dogmengeschichte,* 47].

Klaussner, C., & Vogel, C. (2015). Stylochronometry. Timeline Prediction in Stylometric Analysis. In M. Bramer & M. Petridis (Eds.), *Research and Development in Intelligent Systems XXXII* (pp. 91–106). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-25032-8_6 (Accessed: 16 June 2024).

Kocher, M., & Savoy, J. (2019). Evaluation of text representation schemes and distance measures for authorship linking, *Digital Scholarship in the Humanities,* 34(1), 189–207. DOI: https://doi.org/10.1093/llc/fqy013 (Accessed: 16 June 2024).

Koppel, M., Schler, J., & Argamon, Sh. (2011). Authorship Attribution in the Wild, *Language Resources and Evaluation,* 45(1), 83–94. DOI: https://doi.org/10.1007/s10579-009-9111-2 (Accessed: 16 June 2024).

Eid., & Winter, Y. (2012). The "Fundamental Problem" of Authorship Attribution, *English Studies,* 93, 284–291. DOI: https://doi.org/10.1080/0013838X.2012.668794 (Accessed: 16 June 2024).

Larsen, W.A., Rencher, A.C., & Layton, T. (1980). Who Wrote the Book of Mormon? An Analysis of Wordprints, *Brigham Young University Studies,* 20(3), 225–251. URL: https://byustudies.byu.edu/article/4316 (Accessed: 16 June 2024).

Libby, J. (2015). *Disentangling Authorship and Genre in the Greek New Testament. History, Method and Praxis* [Diss.]. Hamilton: MacSphere. URL: http://hdl.handle.net/11375/24538 (Accessed: 16 June 2024).

Lichtblau, D., & Stoean, C. (2023). Chaos game representation for authorship attribution, *Artificial Intelligence,* 317(103858), no pag. DOI: https://doi.org/10.1016/j.artint.2023.103858 (Accessed: 16 June 2024).

Linmans, A.J.M. (1998). Correspondence Analysis of the Synoptic Gospels, *Literary and Linguistic Computing,* 13(1), 1–13. DOI: https://doi.org/10.1093/llc/13.1.1 (Accessed: 16 June 2024).

Mealand, D. (1989). Positional Stylometry Reassessed. Testing a Seven Epistle Theory of Pauline Authorship, *New Testament Studies,* 35(2), 266–286. DOI: https://doi.org/10.1017/S0028688500024656 (Accessed: 16 June 2024).

Id. (1995). Correspondence Analysis of Luke, *Literary and Linguistic Computing,* 10(3), 171–182. DOI: https://doi.org/10.1093/llc/10.3.171 (Accessed: 16 June 2024).

Id. (1997). Measuring genre differences in Mark with correspondence analysis, *Literary and Linguistic Computing,* 12(4), 227–245. DOI: https://doi.org/10.1093/llc/12.4.227 (Accessed: 16 June 2024).

Id. (1999). Style, genre, and authorship in acts, the Septuagint, and Hellenistic historians, *Literary and Linguistic Computing,* 14(4), 479–506. DOI: https://doi.org/10.1093/llc/14.4.479 (Accessed: 16 June 2024).

Nieboer, M.C. (1970). The Statistical Analysis of A.Q. Morton and the Authenticity of the Pauline Epistles, *Calvin Theological Journal,* 5, 64–80.

Parsons, M.C., & Gorman H.M. (2012). The Assumed Authorial Unity of Luke and Acts. A review essay, *Neotestamentica,* 46(1), 139–152. DOI: https://doi.org/10.10520/EJC121513 (Accessed: 16 June 2024).

Pruscha, H. (1998). Statistical Models for Vocabulary and Text Length with an Application to the NT Corpus, *Literary and Linguistic Computing,* 13(4), 195–198. DOI: https://doi.org/10.1093/llc/13.4.195 (Accessed: 16 June 2024).

Roper, M., Fields, P.J., & Schaalje, G.B. (2012). Stylometric Analyses of the Book of Mormon. A Short History, *Journal of Book of Mormon Studies,* 21(1), 29–45. URL: https://scholarsarchive.byu.edu/jbms/vol21/iss1/4 (Accessed: 16 June 2024).

Roy, A., & Robertson, P. (2022). Applying Cosine Similarity to Paul's Letters. Mathematically Modeling Formal and Stylistic Similarities. In G.V. Allen, S. Schulthess, P. Dilley, & P. Phillips (Eds.), *New Approaches to Textual and Image Analysis in Early Jewish and Christian Studies* (pp. 88–117). Leiden/Boston: Brill [= *Digital Biblical Studies,* 5]. DOI: https://doi.org/10.1163/9789004515116_007 (Accessed: 16 June 2024).

Royal, K. (2012). Using Objective Stylometric Techniques to Evaluate New Testament Authorship, *Journal of MultiDisciplinary Evaluation,* 8(19), 1–7. DOI: https://doi.org/10.56645/jmde.v8i19.352 (Accessed: 16 June 2024).

Satos, P., & Chrysostomos, S. (2022). Examining n-grams and Multinomial Naïve Bayes Classifier for Identifying the Author of the Text "Epistle to the Hebrews". In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence,* SCITEPRESS. Science and Technology Publications, 447–458. DOI: https://doi.org/10.5220/0010970300003116 (Accessed: 16 June 2024).

Smith, M.W.A. (1988). Pseudoscience. A Comedy of Statistical Errors, *Style,* 22(4), 650–653.

Zadok, A., Zhitomirsky-Geffet, M., Schler, J., & Katzoff, B. (2023). Comparative network analysis as a new approach to the editorship profiling task. A case study of the Mishnah and Tosefta from Rabbinic literature, *Digital Scholarship in the Humanities,* 38(4), 1720–1739. DOI: https://doi.org/10.1093/llc/fqad038 (Accessed: 16 June 2024).