Jacob P. B. Mortensen / Erich B. Pracht /
Christian H. Vrangbæk / Eva E. H. Vrangbæk (eds)

Style and Computation

Traditional and Computational Analyses of Style in Ancient Greek and Latin Texts

Studia Aarhusiana Neotestamentica Vol. 13

V&R


V&R

Studia Aarhusiana Neotestamentica (SANt)

Volume 13

Jacob P. B. Mortensen, Erich B. Pracht,
Christian H. Vrangbæk,
and Eva E. H. Vrangbæk (eds)

Style and Computation

Traditional and Computational Analyses of Style in Ancient Greek and Latin Texts

VANDENHOECK & RUPRECHT

ISSN 2364-2165
ISSN 2364-2157 (digital)
ISBN 978-3-525-56105-8
ISBN 978-3-647-56105-9 (PDF)
ISBN 978-3-666-56105-4 (eLibrary)
DOI https://doi.org/10.13109/9783666561054

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data available online: https://dnb.dnb.de.

Vandenhoeck & Ruprecht and Brill Deutschland GmbH are part of De Gruyter Brill.
www.degruyterbrill.com

Questions about General Product Safety Regulation: productsafety@degruyterbrill.com

Printing: CPI books GmbH, Leck

Table of Contents

Conclusions and Perspectives

Jacob P. B. Mortensen, Erich B. Pracht, Christian H. Vrangbæk,
Eva Elisabeth H. Vrangbæk

Style and Computation

Introduction

1 New Testament Studies, Comparative Literature, Computational Methodologies, and Style

In the spring of 2024, the *Computing Antiquity* research project at Aarhus University hosted a conference at Sandbjerg Estate in Southern Denmark, entitled *Style and Computation*. As a research initiative rooted in New Testament studies, we sought to explore the concept of style from both qualitative and quantitative perspectives, combining computational and algorithmic tools with traditional close-reading methods. Our objectives were therefore twofold: first, to examine stylistic features in New Testament texts through the complementary lenses of traditional and computational analysis; and second, to gain insights from, and collaborate with, specialists in computational and stylometric methods working across other areas of literary studies. The outcome was a genuinely interdisciplinary conference, in which scholars of the New Testament and related disciplines worked alongside practitioners in data science. To ensure a strong disciplinary grounding, we invited scholars employing traditional hermeneutic and close-reading approaches, experts in modern and contemporary literature, and specialists in computational and algorithmic methodologies.

Our aim was to draw on developments in other text-focused disciplines, where computational and algorithmic applications have advanced significantly beyond the boundaries of New Testament studies. The field of Comparative Literature, in particular, has been a pioneer in this respect, with computational and algorithmic methods explored extensively over the past two decades. This accelerated uptake in Comparative Literature is due in part to the greater availability of digital resources for modern and contemporary texts in living languages—especially English,[1] which constitutes the majority of source material in digital corpus linguistics—a factor less pronounced for ancient Greek and Latin, notwithstanding notable resources

1 Roopika Risam has pointed out the Anglophone focus of the field of digital humanities, as noted by David Bamann et al.: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.neh.gov/sites/default/files/inline-files/FOIA%2021-09%20Regents%20of%20the%20University%20of%20California%2C%20Berkeley.pdf: p. 2 (downloaded 10–08–2025).

such as Perseus, First1KGreek, and various commercial digital editions of biblical texts. Initiatives by Google, Project Gutenberg, and others have likewise fuelled the rapid proliferation of computational approaches in the humanities, placing Comparative Literature at the forefront in the adoption of these methods.[2]

Another factor contributing to the slower adoption of computational methods in New Testament studies, compared with other fields, is the enduring influence of literary methodologies on the discipline over the past century. Since the emergence of *New Criticism* in the early twentieth century and the subsequent diversification of literary approaches, Comparative Literature has exerted a marked influence on New Testament scholarship. This is evident in the development of now well-established methodologies such as narrative criticism, reader-response criticism, structuralism, deconstruction, gender hermeneutics, and disability studies within the field. In the same way, we contend that New Testament studies stand to benefit from adopting and integrating analytical methods from Comparative Literature, particularly in the domain of computational and algorithmic approaches to stylistic analysis. To be clear, we do not claim that computational or algorithmic methods are inherently more scientific or "better" than traditional hermeneutical approaches. Rather, such methods open new avenues for inquiry and knowledge production, but prompts, coding, and binary classification—all integral to computational work—are not substitutes for traditional exegetical methods and close reading. Nevertheless, we hope that in the future, computational and algorithmic approaches will contribute to the development of new areas of research and inspire fresh lines of questioning.

Prior to the conference, we resolved to explore the relationship between computational methodologies and traditional hermeneutical approaches through a single unifying theme: style. We selected style as the governing topic because this area of inquiry already has a long-standing presence in New Testament studies (*Stilkritik*), while also constituting a major focus in computational research (stylometry). Stylometry—the quantitative analysis of style at the textual level—is by definition empirical, and these two characteristics make it a suitable means for New Testament scholars both to investigate style and to assess the possibilities and limitations of integrating computational methods into their work. At the same time, style is an inherently challenging category: it permeates scholarship across many areas of literary study, yet there is little consensus regarding its definition or interpretive significance. Without claiming that one approach is inherently superior to another, we examine a range of conceptions and applications of stylistic analysis, employing both qualitative and quantitative methods, and reflecting on what "style" signifies across different fields and in relation to diverse research questions.

2 For some of the newest developments, see the contributions in Lauren Tilton, David Mimno and Jessica Marie Johnson, *Computational Humanities* (Minneapolis, MN: University of Minnesota Press, 2024).

The conference proved highly successful: it fostered new friendships, established collaborative partnerships, generated fresh insights, and enabled genuine interdisciplinarity both within and across sessions. In the final evaluation, participants expressed a clear desire for future events of this kind, together with a strong interest in continued exchange and collaboration. In what follows, we seek to acknowledge all areas of participation by outlining both ancient and modern approaches to style, while also introducing computational approaches to stylometry.

2 What is Style?

Style may be defined as the manner or technique by which an author communicates a message. It concerns the way a text presents itself—the form and shape of a work as distinct from its content—although, in various respects, style is inseparable from content. In an article on style in Hebrews and John, Dan Nässelqvist observes that style "concerns the expression of ideas, how a message is conveyed."[3] The style of a text is thus manifested through its sounds, visual impression, choice of words, sentence structures, figures, and content, thereby creating a distinctive voice, tone, atmosphere, and overall effect. Put differently, style may be regarded as the "how" of a text, in contrast to the "what." The "what" refers to the content; the "how" to the manner or technique by which that content is presented. By way of example, a detective novel typically concerns the solving of a murder mystery—this pertains to the content or "what" of the narrative. If, however, the novel is designated as a *noir* detective novel, the label refers to a specific mode, technique, or style through which the story is told. In such cases, *noir* generally signals a pronounced darkness in the development of the plot, which might include, among other features, a rain-slick back alley, a victimised and self-destructive protagonist, and pervasive corruption and violence. This particular *noir* style imparts a dark and brooding atmosphere, tone, and mood to the story.

If an author were to remove all overt stylistic features in an attempt to produce a text without style, this would nonetheless constitute a stylistic choice, one that inevitably affects the meaning of the text. Even a shopping list displays a certain style, however minimal, though it may seek to reduce stylistic elements in favour of a sole focus on content. Style can never be annulled or erased entirely, yet some texts lend themselves more readily to stylistic analysis than others.

When discussed in a narrower sense, style can be difficult to define or to confine to a sharply delineated and closed set of features. In a literary context, style is closely associated with other domains such as rhetoric and genre, and it should always be

3 Dan Nässelqvist, "Stylistic Levels in Hebrews 1.1–4 and John 1.1–18," *Journal for the Study of the New Testament* 35, no. 1 (2012): 3.

considered in relation to content. Nevertheless, style is neither synonymous with rhetoric nor with genre, though it may be regarded as an element of both. Rhetoric may, in broad terms, be described as the combination of arrangement or structure, argumentation, aesthetics, and style, with the explicit purpose of persuasion. Genre may, likewise in broad terms, be described as the combination of form or style, content, and purpose in the overall communication of a text, shaped in the interplay between readerly expectations and authorial constraints.[4] Style is thus an inescapable component of literary analysis, but it overlaps with—and should be examined in close connection to—concepts such as rhetoric, genre, authorship, historical context, and literary movement or period.

When examined in depth, style often proves elusive and resistant to precise definition. It can be regarded as the *x*-factor of a text or an author: that indefinable quality which constitutes its inimitable individuality, making the work stand out as uniquely itself because of a particular style. From this perspective, style assumes a more abstract and expressive character—an attitude, tone, gesture, or voice that carries a quality of surplus or excess. In this sense, style becomes something difficult to define with exactness or to capture exhaustively.

Style occupies a space between the measurable and the ineffable. It can be approached through concrete, analysable features—specific words, morphological choices, syntactic structures, and rhetorical tropes—yet it also encompasses more elusive qualities: the *x*-factor, individuality, distinctive signature, and atmosphere that make a text recognisably itself. It is at once a matter of authorial intention and of forces that lie beyond conscious control, such as education, upbringing, historical period, and cultural milieu. Style is therefore simultaneously formal and expressive, shaped by technique yet saturated with context. This dual nature explains why it resists any single definition and why it requires multiple, complementary approaches, whether drawn from traditional hermeneutics or from computational analysis.

3 Levels of Style

Style is a complex system comprising an array of features of many different kinds. Such stylistic features take multiple forms and operate on a range of levels. In this volume, we conceptualise five distinct levels of style, which we present in this section in order from the smallest to the largest scale. While our examples are neces-

4 For definitions of genre, see, for instance, Tzvetan Todorov, *Genres in Discourse*, trans. C. Porter, (Cambridge: Cambridge University Press, 1990); Alastair Fowler, *Kinds of Literature: An Introduction to the Theory of Genres and Modes* (Oxford: Clarendon Press, 1982); M. Bakhtin, *Speech Genres and Other Late Essays*, trans. Vern W. McGee, (Austin: University of Texas, 1986); Michail Bakhtin, *The Dialogic Imagination: Four Essays*, trans. Caryl Emerson and Michael Holquest, (Austin: University of Texas Press, 1981); Jacques Derrida, "The Law of Genre," *Critical Inquiry* 7, no. 1 (1980).

sarily selective, we contend that this five-level paradigm is applicable to all types of literature, both ancient and modern.

Level 1

The most basic level of style is that of sound and typeface, referring to the design of individual letters in terms of size and shape. In other words, this level concerns the materiality and aesthetics of a text. For example, an author's decision to use bold type or italicise a word is a stylistic choice. Likewise, choosing to centre the text rather than begin each line on the left or right margin constitutes a stylistic decision. An author's choice to omit full stops, as James Joyce does in the final chapter of *Ulysses*, creates a "run-on" style that significantly shapes the reading experience. Interpretive decisions at this level can have major consequences for how a text is read. For instance, should the so-called "Christ Hymn" in Philippians 2:6–11 be understood as prose or as poetry? While the ancient manuscripts offer no guidance, the 28th edition of Nestle-Aland presents the passage as centred text, thereby indicating that it is poetry rather than prose. In such cases, the visual presentation of a text influences its tone, mood, and atmosphere, and consequently affects how readers interpret it. Although New Testament scholars today generally consult the text through typeset copies of the 28th edition of Nestle-Aland, stylistic features at level 1 were already significant at the time of composition. Paul, for instance—who typically employed an amanuensis when dictating his letters (Rom 16:24)—concludes his letter to the Galatians with a personal signature, drawing attention to the distinctive shape of his own handwriting: "See what large letters I use as I write with my own hand!" (Gal 6:11). This personal touch may have served as a mark of authenticity (cf. Phlm 19; Col 4:18) or, given the angry tone sustained throughout Galatians, as a visual display of anger, analogous to the way people today "shout" in emails by using capital letters. Since the original manuscripts of the New Testament writings no longer survive, many level 1 stylistic features valued by the authors are lost to us. Nevertheless, aspects of their materiality remain visible in the history of manuscript transmission, and many of these reflect interpretative decisions about how the texts were intended to be read and heard.

Level 2

The second level of style is the level of words. This concerns the types of words an author employs—whether long or short, old or new, simple or complex, elevated or colloquial. A text whose style is characterised by a high frequency of adjectives may be described as descriptive, while one that makes extensive use of coordinating conjunctions can be said to exhibit a paratactic style. Since an author's choice of vocabulary directly shapes the presentation of the text, words—whether considered in isolation or in combination—constitute critically important stylistic features.

Word choice can reflect an author's conscious decisions and can be approached prescriptively. In *The Elements of Style*—arguably the most well-known style guide for formal American English—William Strunk Jr. and E. B. White advise authors to omit needless words, to place emphatic words at the end of a sentence, and to avoid unclear synonyms (e. g., "dependable" is not a satisfactory substitute for "trustworthy" or "reliable").[5] Similarly, Stephen King, one of the most prolific novelists of recent decades, discusses many aspects of word choice in his memoir on the craft of writing. For example, when advising aspiring novelists, King argues that adverbs (which he refers to as "'-ly' words") should be avoided wherever possible, as they tend to slow the narrative.[6] In many cases, a more precise verb ("She hurried") is preferable to a verb–adverb combination ("She went quickly") in conveying the author's intention. Word choice may also serve to emphasise particular concepts. For example, in the New Testament, the book of Hebrews stands as a statistical outlier among a wide range of Greek texts owing to its high frequency of the terms αἷμα ("blood") and ἱερεύς ("priest"). Here, the frequent use of αἷμα and ἱερεύς constitutes a distinctive feature of the book's style, most likely reflecting the author's thematic focus.

In addition to deliberate choices, word selection can also reflect an author's unconscious stylistic habits. Indeed, studies in authorship attribution operate on the premise that every author leaves behind a unique stylistic fingerprint when composing a text. Traditionally, markers of an author's individuality have been identified in the ways they use common function words such as conjunctions and prepositions, which are therefore often subjected to quantitative analysis.[7] A striking example is provided by studies showing that Jane Austen's style is remarkably consistent across her novels, with her stylistic fingerprint including, for instance, a frequent use of feminine pronouns.[8] In what is perhaps the most influential and widely publicised computational study of authorship attribution, Frederick Mosteller and David L. Wallace identified the stylistic fingerprint of James Madison (as distinct from Alexander Hamilton or John Jay) in a set of political essays of disputed authorship, basing their conclusion on Madison's characteristic use of common function words.[9]

Indeed, methods of forensic authorship attribution—and the underlying assumption that an author invariably leaves behind a stylistic fingerprint—were developed in large part within the context of Pauline studies, in response to Enlightenment-era

5 William Strunk and E. B. White, *The Elements of Style* (New York: Longman, 1999).

6 Stephen King, *On Writing: A Memoir of the Craft* (New York: Scribner, 2000).

7 In recent years, other factors, such as education, social position, familial relationships, literary inspiration, historical context, and gender, have been shown to influence style. For discussion, see Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History* (Baltimore: University of Illinois Press, 2013), 63–104.

8 For discussion, see Jockers, *Macroanalysis*.

9 Frederick Mosteller and David L. Wallace, *Inference and Disputed Authorship: The Federalist* (Reading, Mass., Palo Alto, London: Addison-Wesley, 1964).

German scholarship that questioned the authenticity of certain letters within the thirteen-letter corpus. Beginning with Percival Neale Harrison in 1921, biblical scholars have employed quantitative approaches to uncover the unconscious writing habits of the "authentic" Paul. This enterprise, however, has met with scepticism. Kenneth Neumann observes that stylistic variation within the work of a single author is common, and is therefore pessimistic about the effectiveness of stylometry in forensic authorship attribution.[10] Paul Robertson is sharply critical of "bottom-up" approaches to stylometry, arguing that common function words reveal nothing about an author's unconscious writing habits.[11] Nevertheless, despite such critiques, the notion that unconscious word choice reflects individual stylistic patterns remains a cornerstone of quantitative analyses of style, including in New Testament studies.

Word choice can also be connected to questions of literary relationship—that is, whether an author employs his own words or borrows vocabulary from another author or text. In New Testament studies, this perspective is particularly prominent in the redaction criticism of the Synoptic Gospels. Scholars, for example, may argue that Luke borrowed particular words from Mark and then supplemented them with additional material of his own. Word choice in relation to borrowing is likewise central to one of the most debated issues concerning the book of Acts: whether Luke knew and made use of Paul's letters. Many scholars investigating this question compile lists of linguistic parallels between Paul's letters and Acts, assessing whether Luke employs Pauline vocabulary in a distinctively Pauline manner. In borrowing words from another author or text, an author may consciously or unconsciously evoke the atmosphere or tone of the source material, as Luke does in the opening chapters of his Gospel by adopting language familiar from Israel's scriptures in Greek (LXX), thereby shaping how the text is perceived by the reader.

Particularly in computational analyses of style, scholars regard word groupings as stylistically relevant textual features. Although the units of analysis are small and do not constitute complete sentences (see level 3 below), such groupings reveal patterns of language use that are difficult to detect in large corpora without the aid of corpus-linguistic software. Word groupings may consist of unbroken sequences—known as "clusters" or "n-grams"—such as the fixed expression "COVID-19 pandemic" in modern usage or the divine title "Jesus Christ" originating in early Christianity. While these cannot be analysed at the level of grammar, they nonetheless represent steps towards meaning-making that go beyond single lexical items.[12]

---

10 Kenneth J. Neumann, "The Authenticity of the Pauline Epistles in the Light of Stylostatistical Analysis" (Scholars Press, 1990), 214.

11 Paul Robertson, *Statistical Approaches to Paul's Letters* (Leiden/Boston: Brill, 2025).

12 For a study that showcases the interpretive potential of collocations, see Donald E. Hardy, "Collocational Analysis as a Stylistic Discovery Procedure: The Case of Flannery O'Connor's Eyes," *Style* 38, no. 4 (2004). For an overview of statistical association measures, see Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing* (Cambridge, Mass: MIT Press, 2019).

Word groupings may also include collocations, which in corpus linguistics refers to statistically significant associations between words within a text. Collocations represent a more advanced form of feature selection than unbroken sequences, as they are based on which words tend to co-occur in the immediate context of a given "query word." For example, in Paul's letters, the query word πίστις forms collocations with seven nodes, listed here in order of the strength of their association: ἀκοή, δικαιόω, ἀβραάμ, δικαιοσύνη, ἐκ, νόμος, διά, ἐγώ. In other words, it is characteristic of Paul's style that when he uses the word πίστις, the seven words listed above tend to occur in close proximity to that query token. The incorporation of clusters or n-grams and collocations into stylistic analysis means that, particularly in corpus linguistics, there is effectively no upper limit to the combinations and sequences that can be examined. Consequently, "style" in modern computational studies is often defined in very broad terms, as exemplified by the formulation of J. Berenike Hermann, Karina van Dalen-Oskam, and Christof Schöch: style is "a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively."[13]

Level 3

The third level of style is the level of sentences. While words begin to acquire specific meaning when grouped into clusters or collocations (see above), language truly finds its full expression in the form of sentences. In this regard, Paul Ricoeur observes that a sentence is a synthetic unit in the sense that it constitutes a totality greater than the sum of its constituent parts.[14] The third level of style, therefore, concerns variation in the construction of entire sentences, including their morphology and syntax.

Morphology and syntax concern the way words are formed and how they relate to one another within a sentence. In terms of variation, a sentence may be ordered paratactically or hypotactically, and it may also be polysyndetic or asyndetic. Syntactical choices—such as whether to employ many infinitives or a high number of present participles—affect both the atmosphere and the meaning of a text. The same holds true when an author makes frequent use of imperatives, thereby imparting a commanding or exhortative tone.

Analysis at the level of sentences offers insight into a text's structural and rhetorical complexity. This includes the syntactic arrangement, length, and type of sentences—whether simple, compound, complex, or complex-compound—and

13 J. Berenike Herrmann, Karina van Dalen-Oskam and Christof Schöch, "Revisiting Style, a Key Concept in Literary Studies," *Journal of Literary Theory* 9, no. 1 (2015).

14 Paul Ricoeur, *Interpretation Theory: Discourse and the Surplus of Meaning* (Fort Worth, Texas: The Texas Christian University Press, 1976).

their functions within a broader context (for example, declarative, interrogative, imperative, or exclamatory). Short, simple sentences may generate a sense of urgency or directness, whereas longer, more complex sentences can convey a reflective, nuanced, intellectual, or even strained tone. The choice between parataxis and hypotaxis also shapes the flow of communication. Parataxis can create immediacy and equality between ideas, but it may also produce a tone reminiscent of children's language, lacking refinement and the selective ordering of important information. By contrast, hypotaxis can indicate hierarchy or causality between ideas or concepts, and it is often employed to good effect in discursive or philosophical writing.

Furthermore, sentence-level analysis entails identifying patterns of repetition, parallelism, and deviation, all of which contribute to emphasis and thematic development. The choice between active and passive voice likewise reflects stylistic decisions concerning agency and focus. Sentence construction also influences the pacing of a text: rapid sequences of short sentences can quicken the narrative, whereas long, flowing sentences can slow it down, creating space for reflection or description. Attention to these syntactic elements allows for a more comprehensive interpretation of a text's meaning, tone, and intended effect on the reader.

Level 4

The fourth level of style is that of tropes and figures. Stylistic features such as metaphor, irony, comparison, personification, symbol, euphemism, pleonasm, chiasm, and exclamation are regarded as forms of stylistic ornamentation. In other words, level 4 features are typically viewed as literary art, whereby an author ornaments the text with stylistic devices to shape its presentation. Tropes and figures allow authors to make points without speaking literally, employing figurative or metaphorical language to say one thing by means of another. The manner in which a point is expressed thus reflects specific stylistic decisions. In New Testament studies, stylistic ornamentation is sometimes connected to an author's communicative intentions. For example, Sigurvin Lárus Jónsson argues that the author of the book of James adopts an ornamented, elevated style to establish and assert authority over the wealthy members of his audience, whom he admonishes to accept the poor.[15] In this way, stylistic analysis at the level of tropes, figures, and ornamentation contributes to understanding the atmosphere, intention, and interpretive possibilities of a text.

Stylistic studies addressing level 4 features are generally conducted through close-reading methods. On this point, J. Berenike Hermann and colleagues observe

15 Sigurvin Lárus Jónsson, *James among the Classicists: Reading the Letter of James in Light of Ancient Literary Criticism* (Göttingen: Vandenhoeck & Ruprecht, 2021).

that, in computational studies, "most style markers have so far been relatively simple in nature. Among such features are frequencies and frequency distributions of characters, words, lemmata, word classes or syntactical structures, taken by themselves or in sequences (n-grams)."[16] Within our framework, such features would typically be situated between levels 1 and 3, each representing a bottom-up approach. In the first three levels, analysis proceeds from the specific to the general; in other words, larger insights are derived from stylistic micro-patterns as the basic units of analysis. Hermann, however, argues that even in computational analysis "it seems both useful and possible to add the levels beyond the sentence among the style markers."[17] As examples, they cite "stylistically relevant phenomena such as metaphors or irony which, despite some recent advances, are currently analysed in a qualitative rather than quantitative paradigm and remain a challenge to formal modelling, reliable automatic identification and computational assessment."[18] In other words, while bottom-up stylistic analysis is useful and relatively straightforward to perform computationally, it offers only a partial picture of an author's or text's style, as it neglects important "higher-level" non-linguistic elements.[19] As computational approaches continue to develop, we may anticipate algorithms capable of distinguishing between different tropes—for example, identifying metaphor as distinct from metonymy, and metonymy as distinct from irony.

In New Testament studies, Paul Robertson is, to our knowledge, the only scholar to have undertaken computational analysis at the fourth level of style. Building on a 2016 monograph that identifies text-features characteristic of Paul's letters and of other ethical-philosophical literature in the Greco-Roman world, Robertson applies stylometric methods to Paul's use of pathos, metaphor, irony or satire, oppositions, choices, and related devices, employing established techniques such as cosine similarity and k-means clustering.[20] In doing so, Robertson is the first New Testament scholar to conduct a top-down stylometric analysis—that is, beginning with scholarly informed observations and moving "down" to the micro-features of the text. The chief advantage of this approach is that it demonstrates that stylistic features at level 4 can be analysed quantitatively. Robertson, however, catalogued all stylistic features in Paul's letters manually, a process he considers essential to stylometric analysis, but which nonetheless opens the way for future New Testament scholarship to identify level 4 features using machine-learning methods.

---

16 Berenike Herrmann, van Dalen-Oskam and Schöch, "Revisiting Style," 45.

17 Berenike Herrmann, van Dalen-Oskam and Schöch, "Revisiting Style," 46.

18 Berenike Herrmann, van Dalen-Oskam and Schöch, "Revisiting Style," 46.

19 For this criticism, see Matthew Brook O'Donnell, *Corpus Linguistics & the Greek of the New Testament* (Sheffield: Phoenix Press, 2005), 85–101. See also James A. Libby, "The Pauline Canon Sung in a Linguistic Key: Visualizing New Testament Text Proximity by Linguistic Structure, System, and Strata," *BAGL* 5 (2016).

20 Robertson, *Statistical Approaches*.

Level 5

Another top-down approach to style, generally confined to the qualitative tradition, is the level of discursive mode and point of view. In the New Testament, for example, the author of Acts occasionally departs from the third-person perspective to write in the first person. In these "we"-passages (Acts 16:10–17; 20:5–15; 21:1–18; 27:1–28:16), Luke writes as though he were a participant in the events he narrates. Scholars have proposed a variety of explanations for the presence of these passages, but for our purposes the essential point is that this shift in perspective may be regarded as an element of style. Another New Testament example is found in the book of Hebrews, whose author famously alternates between exposition and exhortation. This alternation has generated substantial scholarly discussion of the book's structure,[21] yet here, too, the key point is that the interplay between logical argumentation and hortatory sections can be understood as a matter of style. As noted above, level 5 stylistic features are typically examined qualitatively. However, Hermann and colleagues suggest that it would be valuable to investigate such features—such as paragraph length—by computational means as well.[22]

Summarising chart

Stylistic level	Feature types	Directionality
Level 1	Font, type, lettering	Bottom-up
Level 2	Words and word-groupings	Bottom-up
Level 3	Sentences	Bottom-up
Level 4	Tropes and figures	Top-down
Level 5	Discursive mode and point of view	Top-down

4 Ancient Theories of Style: Continuity and Contrasts

When working with ancient texts such as the New Testament or the scriptures of Israel, it is essential to recognise their historical, ideological, and socio-cultural embeddedness. Applying ancient definitions of style to such texts allows for a historically contextualised understanding, grounded in the rhetorical traditions of

21 E.g. George H. Guthrie, *The Structure of Hebrews: A Text-Linguistic Analysis* (Leiden: E. J. Brill, 1994); Cynthia Long Westfall, *A Discourse Analysis of the Letter to the Hebrews: The Relationship between Form and Meaning* (London: T&T Clark, 2005).

22 Berenike Herrmann, van Dalen-Oskam and Schöch, "Revisiting Style," 46.

their own time. Ancient theories defined style in terms specific to their cultural and intellectual contexts, framing it in relation to rhetorical aims—whether persuasive, judicial, epideictic, or otherwise.[23]

By contrast, modern definitions of style, shaped by linguistic and literary theories such as structuralism, post-structuralism, semiotics, or performance criticism, often prioritise originality, subversion, or the relationship between language and identity. While these approaches can yield valuable insights, they may obscure features central to ancient writing. What may appear formulaic or repetitive by modern standards could, in its original context, be the deliberate use of valued rhetorical devices such as anaphora or parallelism to achieve elegance or persuasive force. Without awareness of this historical framework, scholars risk imposing anachronistic criteria on ancient works, misreading their stylistic functions and rhetorical effectiveness. A historically informed approach thus enables us to assess ancient artistry and communicative strategies on their own terms, preserving the intentions of authors and the expectations of their audiences.

Style was a prominent topic in ancient literary theory. In rhetorical and literary handbooks roughly contemporaneous with the New Testament, theorists prescribe what they regard as best practice. These works focus less on the idiosyncrasies of individual authors and more on archetypes, *exempla*, trends, and broadly accepted standards of good (or poor) style. Beginning with Theophrastus, the most widespread approach divided style into three levels: plain, middle, and grand.[24] This tripartite scheme is most fully articulated in Roman rhetorical treatises by Cicero and Quintilian, and Dionysius of Halicarnassus also touches on it (see further down). Ancient theorists generally considered all three levels acceptable, though individual preferences existed, and all agreed that style should be adapted to circumstance.

A key factor in selecting an appropriate style was the nature of the content. This reflects the fluid boundary between style and content: the two often blend to the point where the distinction is blurred. Demetrius emphasises that an author must "preserve propriety, whatever the subject; in other words, use the relevant style—slight for slight themes, grand for grand themes."[25] As Nässelqvist summarises, "the levels of style combine content, function, and shape. These features should always, following the virtue of propriety, be congruous. A low, or simple, matter is dressed in a plain style and the text is then delivered accordingly; the content and function determine what stylistic level to use, which in turn provides the shape."[26]

---

23 The ancient Greek and Roman words for style are λέξις, φράσις, *dictio*, and *elocutio* (among others).

24 No direct statement of Theophrastus on the tripartite division of style (plain, middle, grand) survives, yet later authors preserve testimonia which attest to his treatment of the subject, see Dionysius of Halicarnassus, *On the Arrangement of Words*, (Oxford: Clarendon, 1973).

25 Demetrius, *On Style* 120 (Innes/Roberts, LCL).

26 Dan Nässelqvist, "Stylistic Levels in Hebrews 1.1–4 and John 1.1–18," *Journal for the Study of the New Testament* 35, no. 1 (2012): 36.

Although ancient theories emerged in distinct rhetorical, philosophical, and literary contexts, they offer valuable points of comparison with modern approaches, including the five-level model as presented above.

Aristotle's treatment of style in *Rhetoric* (3.1–12) situates style alongside arrangement as one of the core elements of rhetoric.[27] For Aristotle, style concerns clarity, appropriateness, and a certain propriety of expression, with attention to rhythm and metaphor as tools for persuasion. His view already combines what in our model would span several levels: the choice of words (level 2), the arrangement of sentences (level 3), and the deployment of tropes such as metaphor (level 4). Aristotle's emphasis on functional clarity also anticipates modern debates about the relationship between style and communicative purpose.

In the Hellenistic period, Dionysius of Halicarnassus developed a more explicitly literary conception of style. In works such as *On Imitation*, *The Art of Rhetoric*, *On Thucydides* and *On the Arrangement of Words*, he classifies styles into categories—plain, middle, and grand—each associated with particular effects and audience expectations.[28] He also links stylistic effect to rhythm, euphony, and syntactical arrangement, concepts that overlap with our sentence-level concerns (level 3) as well as the use of figures and ornamentation (level 4). His approach suggests that style is both an aesthetic and a functional phenomenon, embedded in genre conventions and rhetorical aims.

Longinus' *On the Sublime* moves beyond technical classification to explore style as a vehicle for emotional and intellectual elevation. For Longinus, the sublime arises from a fusion of innate genius and learned technique, producing a tone and atmosphere that can overwhelm or transport the audience.[29] This anticipates modern notions of style as the *x*-factor (as discussed in our introduction)—a quality that operates at the upper levels of our model, especially in relation to discursive mode and authorial point of view (level 5).

Roman rhetorical theory largely adopted and adapted Greek principles. Quintilian, in his *Institutio Oratoria*, treats style (*elocutio*) as inseparable from invention and arrangement, but subdivides it into *virtutes* (clarity, correctness, ornament, appropriateness) and links these to the moral character of the orator.[30] His discussion of tropes and figures, including metaphor, irony, and chiasm,[31] aligns closely with our level 4, while his attention to sentence rhythm and variation in periodic structure parallels our level 3. Cicero's works—particularly *Brutus* and *Orator*—elaborate extensively

27 Aristotle, *The "Art" of Rhetoric*, ed. and trans. Freese/Striker, LCL (Cambridge: Harvard University Press, 2020).

28 Halicarnassus, *Arrangement of Words* 321–42.

29 Longinus, *On the Sublime*, 466 ff., (Oxford: Clarendon, 1973).

30 Quintilian, *Institutio Oratoria*, 8.1.1–3, ed. and trans. Butler, LCL (London: William Heinemann, 1966).

31 Quintilian, *Inst. Or.* 9.1 (Butler, LCL).

on the principles that would later be codified by Quintilian, developing a vision of style (*elocutio*) as inseparable from the broader aims of rhetoric. For Cicero, style is not merely a matter of ornament but a vehicle for expressing the orator's ethos and for shaping persuasion through appropriateness (*decorum*), rhythm, and variation.[32] He links stylistic choice closely to the orator's moral character, the demands of the audience, and the specific rhetorical situation, anticipating Quintilian's later emphasis on the "good man speaking well" (*vir bonus dicendi peritus*).[33]

From a modern perspective, two contrasts are especially significant. First, ancient theories tend to treat style as inseparable from rhetoric and persuasion; stylistic choice is almost always evaluated in terms of its effectiveness in influencing an audience. This contrasts with many modern approaches, especially in computational stylistics, where style can be treated as a measurable property of texts independent of rhetorical situation. Second, ancient discussions rarely separate the levels of style analytically in the way our model does. For Aristotle or Quintilian, word choice, sentence arrangement, and figurative language are integrated aspects of a unified rhetorical act, whereas our five-level model allows for targeted analysis at each level—whether through close reading or algorithmic processing. Nevertheless, there are clear continuities. Ancient authors were acutely aware of how stylistic features operate across multiple scales, from the form of individual words (level 2) to the orchestration of whole discourses (level 5). They recognised that style could be both consciously crafted and unconsciously revealing, and that it could signal genre, social identity, and authorial intent. These insights resonate with modern hermeneutical approaches and, increasingly, with computational methods capable of identifying patterns—such as recurrent metaphors or preferred syntactical structures—that ancient critics described qualitatively.

By placing our model in dialogue with these earlier theories, we underscore that the study of style has always involved negotiating between formal description and interpretive judgement. The ancient emphasis on function, appropriateness, and the interplay of aesthetic and persuasive goals offers a reminder that style is never purely decorative, but central to meaning-making. While our five-level paradigm separates the analysis into discrete layers, it remains in continuity with a tradition that understood style as multi-layered, purposeful, and inseparable from the act of communication.

Ancient literary theorists can offer valuable tools for interpreting the style(s) of New Testament authors. By employing the stylistic and rhetorical categories articulated by figures such as Dionysius of Halicarnassus, we gain an *emic* perspective

32 Cicero, *De Oratore*, 3.210–13 Sutton/Rackham, LCL (Harvard: Cambridge University Press. See also J. G. F. Powell, "Cicero's Style," in *Cambridge Companion to Cicero*, ed. Catherine Steel (Cambridge: Cambridge University Press, 2013).

33 Cicero, *De Or.* 1.217–18 (Sutton/Rackham, LCL). Quintilian, *Inst. Or.* 12.1.1 (Butler, LCL).

on ancient writing techniques, enabling us to interpret stylistic elements in the New Testament through concepts and terminology that circulated in the ancient world.[34] The New Testament authors would have encountered these categories—directly or indirectly—through their educational formation;[35] understanding the principles they were taught increases our capacity to discern their communicative strategies and, in some cases, even the underlying aims of their messages. At the same time, certain limitations must be acknowledged. Ancient rhetorical handbooks were composed in specific historical contexts and for particular audiences, and the New Testament writers were not their intended users. Cicero, for instance, theorises style with a view toward oral delivery in contexts such as the Senate, popular assemblies, or law courts—settings quite removed from, say, the apostle Paul writing letters to religious communities. Nevertheless, despite such differences in purpose and audience, the New Testament authors were not merely passive participants in the stylistic conventions of the Greek-speaking world; they also contributed to, adapted, and reshaped those conventions, adding their own voices to the diverse stylistic landscape of antiquity.

5 Style as a Key to Interpretation

When conceived in material terms—as the words, sentences, and tropes or figures that make up a text—style may appear concrete and straightforward. On this level, it is indeed relatively easy to analyse, since it can be observed directly in lexical choices, syntactic patterns, and rhetorical devices. Yet the mere identification of such features is of limited interest in itself. Style only becomes critically significant when description is coupled with interpretation, and when questions of meaning, effect, and significance are brought into play. At this point, stylistic features cease to be (mere) aesthetic ornamentation and instead become constitutive elements in our perception and understanding of texts.

This is evident in a wide range of literary works, from canonical novels to more experimental prose. The atmosphere or feel of a text—often a cumulative product of numerous stylistic decisions—can be an interpretive clue in its own right. Mary Shelley's *Frankenstein*, for instance, employs the hallmarks of Gothic style: richly

34 For the distinction between emic and etic, see Clifford Geertz, *The Interpretation of Cultures: Selected Essays* (London: Fontana Press, 1993); Clifford Geertz, *Local Knowledge: Further Essays In Interpretive Anthropology* (New York: Basic Books, 2008). For the importance of emic categories in the analysis of New Testament texts, see Michael W. Martin and Mikeal C. Parsons, *Ancient Rhetoric and the New Testament: The Influence of Elementary Greek Composition* (Waco: Baylor University Press, 2018), 75.

35 See e. g. Steve Reece, *The Formal Education of the Author of Luke-Acts* (London: Bloomsbury, 2022); Martin and Parsons, *Ancient Rhetoric*.

detailed description, intense dramatic suspense, and a heightened emotional register. These features are not incidental decoration but directly serve the novel's fundamental engagement with questions about human nature and the ethical boundaries of creation. Jonathan Swift's *A Modest Proposal* offers a different kind of example, where the stylistic register itself becomes a vehicle for irony and critique. Swift's matter-of-fact, objective tone stands in stark dissonance to the grotesque suggestion that Ireland's economic problems might be solved by eating its children. The very incongruity between style and content intensifies the satire, sharpening his indictment of British policy towards Ireland and signalling to the reader how the text ought to be unlocked.

Stylistic form can also direct a reader's attention towards certain themes or motifs. The minimalist prose of Ernest Hemingway, for example, is marked by its sparseness and restraint, forcing the reader to attend to subtext and implication. In *The Old Man and the Sea*, such simplicity foregrounds the universal themes of struggle, dignity, and endurance without encumbering them with superfluous detail. At the same time, style functions as a marker of social and cultural identity. Dialect, sociolect, and idiolect can locate a character within a particular class, region, or historical moment. Mark Twain's *Adventures of Huckleberry Finn* draws extensively on dialect and idiomatic speech, lending authenticity to its characters while also framing a commentary on racial and social divisions in nineteenth-century America.

Figurative language is another crucial dimension of style. In Shakespeare's plays, metaphors and similes are not mere poetic embellishments; they are deeply enmeshed with character psychology and thematic development. *Hamlet* offers a case in point: the soliloquy "To be, or not to be" distils the prince's existential crisis, but its resonance is amplified by a network of metaphors and similes throughout the play that sustain its dualities and ambiguities. Such continuity between different levels of style, content, and character underscores the integrated nature of Shakespeare's craft.

In short, the interpretative significance of style cannot be overstated. It shapes our aesthetic experience of a text while also serving as a primary key to its deeper themes, symbolic structures, and the author's intentions. Stylistic analysis connects the micro-level—the morphology of a historical present, or the syntactic rhythm of parataxis—with the macro-level of a text's worldview and rhetorical force. As the preceding sections have shown, ancient and modern theories alike emphasise this interplay between the tangible features of style and the broader frameworks of meaning they support. The chapters that follow in this volume build on this dual perspective, exploring how close reading methods and computational approaches can detect, quantify, and model such stylistic patterns, thereby extending traditional literary criticism into new analytical domains.

Building on these reflections on the interpretative significance of style, it is now necessary to turn to the practical question of how stylistic features can be systematically detected, measured, and modelled. As discussed above in the section on ancient the-

ories of style, classical rhetoricians were deeply concerned with the identification, classification, and evaluation of stylistic features. While their approaches relied on qualitative observation and rhetorical training, computational methods offer new possibilities for extending such analyses to large-scale corpora with a high degree of precision and replicability. The following overview introduces a range of computational techniques, many of which are applied in the contributions to this volume. It is not intended as an exhaustive technical manual, but rather as an extended glossary of key approaches, outlining their principles, typical applications, and potential for integration. Several of these methods overlap or can be combined within a single analysis, and many have evolved into hybrid forms drawing on multiple computational traditions. By situating these techniques in relation to the study of style—both ancient and modern—the aim is to provide the reader with a conceptual and practical framework for engaging with the computational analyses that follow.

6 Computational Methods for the Analysis of Style

This section presents a concise survey of computational methods relevant to the analysis of style, organised according to their most common areas of application. Rather than offering a technical manual, it serves as a reference guide to key techniques, providing readers with a sense of their scope, underlying principles, and potential for combination. Several of the approaches described here are not mutually exclusive: tools developed within one methodological tradition can be integrated into another, as when natural language processing techniques such as topic modelling are incorporated into a stylometric study. Similarly, within a single framework such as stylometry, both supervised and unsupervised approaches can be employed, depending on the aims and design of the analysis.

a Stylometry

Stylometry is the quantitative study of linguistic style, employing statistical and computational techniques to calculate and analyse patterns in written texts. Stylometry is rooted in efforts to resolve authorship disputes, such as distinguishing several authors in one piece of writing, and stylometry has evolved into an interdisciplinary field combining linguistics, computer science, and literary studies.[36]

36 Richard Forsyth, "Stylistic Structures: a Computational Approach to Text Classification" (PhD diss., University of Nottingham, (1996); Schlomo Argamon and Moshe Koppel, "The Rest of the Story: Finding Meaning in Stylistic Variation," in *The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning*, ed. Shlomo Argamon, Kevin Burns, and Shlomo Dubnov, (Berlin, Heidelberg: Springer, 2010), 79–112; Berenike Herrmann, van Dalen-Oskam and Schöch, "Revisiting Style," 37–38. For stylometric approaches in this volume see the contributions by Pracht and van Nes.

Stylometry is the investigation of stylistic markers or fingerprints that reveal a stylistic imprint, be it for the sake of authorship attribution of a disputed text, genre recognition or a classification task.[37] The basic idea of stylometry is to select features, normally the most frequent words (MFWs), and then calculate the distances between the value matrices that appear in the feature selection. Of these are, for example, burrows delta, support vector machines, k-nearest neighbours, character trigrams and quatrograms, or a lexicometric analysis where all words are vectorised and compared with principal component analysis.[38]

How to determine the stylistic imprint and what to select as markers has shifted over time. In early stylometry, word length and relative frequency of words were manually counted and calculated. The most common known feature selection is simply to investigate MFWs or a set of selected words (N-grams), since, surprisingly, even apparent trivial words like conjunctions and pronouns (in some analyses labelled as stop words) can be valuable markers of an author's style.[39] For example, we would regard it a stylistic marker if an author consistently used the trigram "ὁ ἀνήρ ἀγαθός" instead of "ὁ ἀνήρ ὁ ἀγαθός." As a matter of fact, some stylometric studies use character-based n-grams to analyse for genre- and author specific markers, for example to distinguish prose from verse in Ancient Greek.[40]

The computational programmes of today like AntConc or R's package Stylo allow for a faster calculation and a broader selection of features.[41] All textual features can be quantified and may contribute to the constitution of style. For this reason, style has by computational stylistics been described broadly as being "a property of texts

37 Maciej Eder, "Rolling Stylometry," *Digital Scholarship in the Humanities* 31, no. 3 (2016): 457.

38 Maciej Eder, "Does Size Matter? Authorship Attribution, Small Samples, Big Problem," *Digital Scholarship in the Humanities* 30.2 (2015): 167–82; Sophie Robert-Hayek, Jacques Istas, and Frédérique Rey, "Unravelling the Synoptic Puzzle: Stylometric Insights into Luke's Potential Use of Matthew," *Fourth Conference on Computational Humanities Research (CHR 2023)* (Paris, France: https://hal.science/hal-04827223). For Burrow's Delta see J. F. Burrows, "'Delta': a measure of stylistic difference and a guide to likely authorship," *Literary and Linguistic Computing* 17.3 (2002): 267–87. For Jaccard Similarity see Wu, Shushu, Fang Liu, and Kai Zhang, "Short Text Similarity Calculation Based on Jaccard and Semantic Mixture," 37–45 in *Bio-Inspired Computing: Theories and Applications*, Vol. 1363 of *Communications in Computer and Information Science* (Singapore: Springer, 2021).

39 Maciej Eder, "Style-Markers in Authorship Attribution: A Cross-Language Study of the Authorial Fingerprint," Studies in Polish Linguistics 6 (2011): 99–114.

40 Vanessa B. Gorman and Robert J. Gorman, "Approaching Questions of Text Reuse in Ancient Greek Using Computational Syntactic Stylometry," *Open Linguistics* 2.1 (2016); Pierre van Hecke, "Computational Stylometric Approach to the Dead Sea Scrolls: Towards a New Research Agenda: Dead Sea Discoveries," *Dead Sea Discoveries* 25.1 (2018): 57–82.

41 Lawrence Anthony, "AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit," in *IWLeL 2004: An Interactive Workshop on Language e-Learning, Waseda University* (2005), 7–13; Maciej Eder, Jan Rybicki, and Mike Kestemont, "Stylometry with R: A Package for Computational Text Analysis," *The R Journal* 8, no. 1 (2016).

constituted by an ensemble of formal features which can be observed quantitatively or qualitatively."[42] However, as Robertson notes in this volume, there has to be made a distinction between bottom-up, or unsupervised, and top-down, or supervised stylometric analyses. The difference between bottom-up and top-down is not the mathematical modelling, but rather how and what to select to model upon.[43]

If one, on the one hand, chooses a bottom-up, data-driven approach and e. g. investigate word frequencies and other raw data of the text, then we are searching for hidden patterns or author fingerprints.[44] Such approaches focus on the author's patterns and invariants in common linguistic features like function words (e. g., "the," "of") to detect any possible consistencies across an author's works. From-below thus means to let the various selected features constitute the patterns to conclude upon.

If, on the other hand, one works from-above, the approach is to pre-label sections of texts and operate with features that are selected by an informed researcher. This supervised approach applies theoretical frameworks or domain-specific knowledge to guide the selection of features. Robertson (2016) is seminal in proposing a scholarly informed, top-down approach to Paul's letters.[45] For example, scholars might prioritise syntactic structures or thematic elements identified through literary theory, or, as Robertson suggests, pre-label sections of Pauline texts to which mathematical models can then be applied. Some studies combine top-down and from-below approaches by contextualizing stylistic traits within historical, cultural, or psychological factors combining supervised and unsupervised methods.[46]

Collocation Analysis and Key Word in Context (KWIC)

When utilising collocation analysis, one investigates collocations, i. e. which words that co-locate together in a text. For example, in New Testament literature, the word "Jesus" will often co-locate with "Christ," "said" and "answered." These col-

---

42 Berenike Herrmann, van Dalen-Oskam and Schöch, "Revisiting Style," 44.

43 See the section above on "Levels of Style."

44 Eder, "Rolling Stylometry", 457–458. The selection of features requires informed selection. For a systematised survey see Nadezha Lagutina Ksenia, Elena Boychuk, Inna Vorontsova, Elena Shliakhtina, Olga Belyaeva, Ilya Paramonov, and P. G. Demidov, "A Survey on Stylometric Text Features," 184–95 in *2019 25th Conference of Open Innovations Association (FRUCT)*, 2019; Erich Benjamin Pracht and Thomas McCauley, "Style and Influence: Computing Hebrews and the Early Christian Stylistic Fingerprint," *Religions* 16.1 (2025): art. 1, p. 55.

45 Paul Robertson, *Paul's Letters and Contemporary Greco-Roman Literature: Theorizing a New Taxonomy* (Leiden: Brill, 2016); Roy Ashley and Paul Robertson, "Applying Cosine Similarity to Paul's Letters: Mathematically Modelling Formal and Stylistic Similarities," 88–117 in *New Approaches to Textual and Image Analysis in Early Jewish and Christian Studies* (Leiden: Brill, 2022).

46 K. J. Jautze, C. W. Koolen, Andreas van Cranenburgh, and H. A. de Jong, "From High Heels to Weed Attics: A Syntactic Investigation of Chick Lit and Literature," 72–81 in *Proceedings of the Workshop on Computational Linguistics for Literature* (Atlanta, GA: Association for Computational Linguistics (ACL), 2013).