

# Stilistische Analyse

Fotis Jannidis

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

**Abstract** Der Begriff *Stil* bezeichnet in den Geisteswissenschaften zumeist eine systematische Wahl der Ausdrucksmittel in einem Zeichensystem, z. B. der Sprache, die charakteristisch für eine\*n Autor\*in oder ein Genre oder eine Epoche etc. sind. Die Stilometrie verwendet diese Merkmale, also im Fall von Texten etwa Lexik, Syntax, Semantik und Textstruktur, um mittels quantitativer Verfahren wie Clustering oder Klassifikation Autorschaftsattributionsen, Autorschaftsprofiling oder auch Epochen- und Gattungszuordnungen vorzunehmen. Stilometrische Verfahren wurden bereits sehr früh auf religiöse Texte angewendet und spiegeln, wie man an der Geschichte der Analyse der Paulinischen Briefe sehen kann, wichtige Stationen der Entwicklung der Stilometrie von der univariaten zur multivariaten Analyse, die heute zumeist in einem probabilistischen Framework mit zahlreichen Versuchswiederholungen unternommen wird.

**Keywords** Stilometrie, Autorschaft, Genre

In vielen geisteswissenschaftlichen Disziplinen wird der Begriff *Stil* verwendet, um eine Verwendung von Sprache und anderen Ausdrucksmitteln zu bezeichnen, die durch die bewusste oder unbewusste Wahl der Autor\*innen bestimmt ist und nicht durch den Inhalt allein vorgegeben ist. Dabei lassen sich typische Verwendungsweisen unterscheiden: Stil wird als Gestaltaspekt verstanden und bezieht sich auf den ganzen Text oder sogar auf das Oeuvre der jeweiligen Autor\*innen. Stil wird dabei oft als Ausdruck der Individualität der Autor\*innen verstanden, die sich in der für das schreibende Individuum typischen Auswahl an Elementen zeigt. Diese Auswahl weicht systematisch und in charakteristischer Weise von den jeweils herrschenden Konventionen ab oder eben nicht (Hermann et al. 2015). Unter stilistischer Analyse versteht man dann zumeist, vor allem in der Untersuchung literarischer Texte, eine Ermittlung der für den jeweiligen Stil typischen sprachlichen Mittel, also spezifische Satzkonstruktionen, Verbkonstruktionen, Adjektivreihungen u. a. m. (Leech & Short 2007).

## 1. Stilometrie

Verglichen mit diesem Ziel einer qualitativen Beschreibung der Eigenheiten eines Stils beschäftigt sich die Stilometrie, also die korpusbasierte Analyse von Stil mit quantitativen Verfahren, zumeist mit sehr viel einfacheren Aufgaben, nämlich der Zuordnung eines Textes zu einer Gruppe von anderen Texten aufgrund stilistischer Eigenschaften. Die Stilometrie ist eines der produktivsten Felder der digitalen Geisteswissenschaften mit einer langen Geschichte (Holmes 1998) und vielen unterschiedlichen Ansätzen und Anwendungen in vielen wissenschaftlichen Disziplinen und Anwendungsfeldern. Forschungsüberblicke mit reichhaltigen Bibliographien finden sich in Juola (2008), Stamatatos (2009) und Neal (2017).

Im Zentrum stehen dabei vor allem Aufgaben, die Aspekte von Autorschaft behandeln:

- Autorschaftsattribuion. Ein Text mit unbekannter Autorschaft wird jemandem aus einer Menge von möglichen Kandidat\*innen zugewiesen, von denen Texte vorliegen, deren Autorschaft unzweifelhaft ist.
- Die Autorschaftsverifizierung. Gegeben sind zwei Texte; die Aufgabe ist es, festzustellen, ob dieselbe Autorschaft vorliegt.
- Autor-Profilung, auch *sociolinguistic profiling* (Grant 2022, 20). Hierbei besteht die Aufgabe darin, aufgrund von einem oder mehreren Texten das Geschlecht, das Alter, die Muttersprache, Persönlichkeitseigenschaften, Erkrankungen wie Demenz (Hirst & Wei Feng 2012) oder den Beruf des\*der Autors\*in zu bestimmen.
- Multiautorschaft. In Texten, die von mehreren Autor\*innen verfasst wurden, sollen die Abschnitte ihren jeweiligen Autor\*innen zugeordnet werden.
- Stylochronometry. Ziel ist es, die zeitliche Anordnung einer Reihe von Texten aufgrund stilistischer Eigenschaften zu bestimmen, etwa um die Chronologie der Texte einer Person festzulegen (Seminck et al. 2022).

Neben der Untersuchung dieser Aspekte von Autorschaft gibt es in der Stilometrie noch eine Reihe weiterer Fragestellungen, die aber sehr viel weniger intensiv erforscht werden. Dazu gehören die Zuweisung eines Textes aufgrund seiner stilistischen Eigenschaften zu einer Epoche oder einem Genre (Jannidis & Lauer 2014), die Erkennung der physischen Schreibsituation (Diktat, Handschrift, Schreibmaschine usw. in Hoover 2021), die Sichtbarkeit von Übersetzer\*innen (Rybicki 2021) und Figurenideolekte (Burrows 1987). Stilometrie wird außerdem auch zur Entdeckung von Plagiaten eingesetzt sowie um Texte so umzugestalten, dass ihr Stil nicht mehr dem des\*der Autor\*in entspricht. Es gibt als relativ eigenständiges Teilfeld außerdem die Stilometrie innerhalb der forensischen Linguistik, die sich z. B. mit der Zuordnung von Erpressungsbriefen beschäftigt. Die forensische Stilistik verwendet ebenfalls Korpora, etwa um autor\*innenspezifische Verwendungsweisen im Sinne einer

Abweichungsstilistik zu identifizieren, hat aber aufgrund der zumeist sehr kleinen Textmenge und den Rahmenbedingungen des Rechtssystems doch recht eigene Arbeitsweisen (Grant 2022).

Die Autorschaftsattributions, die sicherlich eine der am häufigsten bearbeiteten Aufgaben darstellt, geschieht in ganz unterschiedlichen Konstellationen. Im einfachsten Fall geht es darum, einen unbekanntem längeren Text einem\*r von zwei Autor\*innen zuzuweisen, von denen zahlreiche längere Texte des gleichen Genres und aus der gleichen Zeit wie der fragliche Text vorliegen. Variationen dieser Faktoren führen aber schnell zu sehr viel komplexeren Untersuchungskonstellationen. So kann etwa die Zahl der möglichen Autor\*innen sehr groß sein. Die Frage, ob sich der\*die Autor\*in des fraglichen Texts sicher unter den Kandidat\*innen befindet (*closed set*) oder nicht (*open set*), ist entscheidend für die verwendeten Methoden. Häufig liegen zudem nur Texte aus anderen Genres bzw. anderen Zeiten oder in unterschiedlichen Sprachen vor. Nicht zuletzt liegen oft nur kurze oder sogar sehr kurze Texte vor, was die Auswertung mit statistischen Verfahren sehr erschwert.

Die quantitative Analyse von Stil verwendet Textmerkmale, um Texte nach Autorschaft zu gruppieren (*cluster analysis*) oder zu klassifizieren (überwachtes maschinelles Lernen). Diese Merkmale kann man in vier Gruppen teilen (Neal 2017, 11):

- Lexik, z. B. Worthäufigkeiten, Wort-N-Gramme, Anzahl von Buchstaben, durchschnittliche Wortlänge.
- Syntax, z. B. Häufigkeit von Satzzeichen, Satzlänge, Komplexität des Satzbaus.
- Semantik, z. B. Synonyme, Topic-Verteilungen.
- Textstruktur, z. B. Länge von Absätzen, Absatzeinzüge, Schriftart.

Haben sich die ersten Überlegungen zur quantitativen Analyse von Stil noch auf die seltenen Wörter und Satzkonstruktionen als individualisierende Merkmale konzentriert, ist es schon lange üblich, die Verteilungen in der Verwendung der *häufigsten* Merkmale für besonders diskriminativ zu erachten. Die Forschung der letzten Jahrzehnte hat gezeigt, dass es keine eindeutig definierte Menge von Merkmalen gibt, die bei jeder der oben genannten Aufgaben und in jeder der vielen möglichen Konstellationen zu den besten Ergebnissen führt. In den frühen Jahren der Stilometrie wurde jeweils ein einzelnes Merkmal erhoben und verglichen, während es seit den 1980er Jahren üblich wurde, mehrere Merkmale auf einmal heranzuziehen.

Seit 2011 gibt es die *Shared Tasks* zur Stilometrie von PAN.<sup>1</sup> Shared Tasks sind ein in der Computerlinguistik etabliertes Verfahren, das Wissen der Forschung zu einem Problem zusammenzufassen und weiterzuentwickeln. Dazu wird ein Datenset zusammen mit einer Aufgabe publiziert, die anhand der Daten zu lösen ist. So lautete eine der Aufgaben des Jahrs 2019, Autor\*innen von Fanfiction eines bestimmten Fandoms zu identifizieren; bekannt waren aber nur ihre Werke zu anderen Fandoms

1 S. <https://pan.webis.de>, zuletzt aufgerufen am 16.06.2024.

(*cross-domain*), außerdem war die Aufgabe als *open-set attribution* formuliert. Die eingereichten Lösungen werden dann anhand von Daten, die die Teilnehmenden noch nicht gesehen haben, bewertet und das Ergebnis sowie die Datensets publiziert, sodass spätere Studien ihre Ergebnisse direkt mit denen des Wettbewerbs vergleichen können. Diese *Shared Tasks* von PAN haben viel dazu beigetragen, das Wissen über den methodischen Stand der Stilometrie zu vereinheitlichen und voranzutreiben und zeigen heute zumeist den Stand der Forschung auf.

Im Kontext der Stilometrie wird die spezifische sprachliche Gestalt eines Textes als Ergebnis einer Reihe von Faktoren betrachtet, zu denen mindestens die folgenden gehören: Das Thema des Texts, die allgemeine Sprachverwendung der Zeit, die manchmal etwas verkürzend als Epochenstil bezeichnet wird, die textsortentypischen sprachlichen Merkmale (Genrestil) und die autor\*innenspezifische Gestaltung der sprachlichen Freiräume (Autor\*innenstil). Ziel einer stilometrischen Untersuchung ist es, aufgrund sprachlicher Merkmale, die nicht wesentlich vom Thema abhängen, eine Zuordnung im Sinne des Erkenntnisinteresses vorzunehmen, also z. B. Autorschafts-attribution, und dabei die anderen Faktoren zu kontrollieren, also z. B. Epoche und Genre. Die Faktoren Autorschaft, Genre und Epoche/Zeit haben gleichzeitig Einfluss auf den Stil; entsprechend müssen Untersuchungsdesigns die jeweils nicht-fraglichen Aspekte als mögliche Störvariablen kontrollieren. Eine solche Kontrolle kann etwa im Fall einer Autorschafts-attribution dadurch geschehen, dass nur Texte der gleichen Zeit und des gleichen Genres in das Untersuchungskorpus aufgenommen werden.

Nach der Zusammenstellung aller relevanten Texte in digitaler Form werden die für relevant erachteten  $N$  Merkmale, z. B. Wörter, Satzzeichen, Buchstaben-Ngramme usw., aus allen Texten extrahiert, wobei  $N$  für die Anzahl der Merkmale steht. Jeder Text wird dann als ein Vektor mit den  $N$  Werten dieser Merkmale repräsentiert. Geometrisch gesehen kann man jeden dieser Vektoren als einen Punkt in einem  $N$ -dimensionalen Raum auffassen. Abstandsmaße erlauben es dann, die Distanz zwischen diesen Punkten zu bestimmen, die jeweils einen Text repräsentieren. Gehen wir von der einfachsten Konstellation aus, also dass es einen Text  $X$  mit unbekannter Autorschaft gibt sowie zwei Gruppen von Texten,  $G_A$  und  $G_B$ , wobei die erste von Autorin  $A$  und die zweite vom Autor  $B$  stammt, und dass wir wissen, dass  $X$  von  $A$  oder  $B$  verfasst wurde. Man wählt nun die Textmerkmale so, dass die Texte einen relativ kleinen Abstand innerhalb der Gruppe und einen deutlichen zwischen den Gruppen aufweisen. Wenn wir nun den fraglichen Text  $X$  hinzufügen, dann ist zumeist zu beobachten, dass der Text einen relativ geringen Abstand zur einen Gruppe und einen relativ großen Abstand zur anderen aufweist – wenn die oben genannten Rahmenbedingungen eingehalten werden, dass es sich um Texte der gleichen Gattung und der gleichen Zeit handelt, und der fragliche Text wirklich von einem\*r der beiden Autor\*innen stammt.

Eine Entscheidung für eine bestimmte Menge von Merkmalen, die die Texte der Kandidat\*innen optimal differenziert, kann aber nur gefällt werden, wenn fest-

steht, dass als Autor\*in nur eine\*r der beiden Kandidat\*innen in Frage kommt (*closed set*). Aber in den meisten Fällen gibt es diese klare Sicherheit nicht, vielmehr könnte auch jemand der\*die Autor\*in sein, der nicht unter den Kandidat\*innen ist (*open set*). Dann aber kann man die Textmerkmale gar nicht optimal wählen. Auch aus forschungspraktischer Perspektive hat sich das Vorgehen, für ein bestimmtes Datenset als optimal erachtete Merkmale und nur eine Auswertungsmethode zu wählen, nicht bewährt, da die Ergebnisse kaum überprüfbar und vergleichbar sind. Deshalb setzen neuere Ansätze darauf, nicht nur mit einer bestimmten Menge von Merkmalen zu arbeiten und dann eine Abstandsmessung durchzuführen. Vielmehr wird die Messung mit jeweils einer anderen Auswahl aus den Merkmalen wiederholt (Juloa 2015).

Sperling et al. zeigen z. B. in ihrer methodischen Studie zur Autorschaftsattribu- tion unter *open set*-Bedingungen, dass Stephen King auf diese Weise sehr eindeutig als Verfasser der Romane identifiziert werden kann, die unter dem Namen Richard Bachmann publiziert worden sind (Sperling et al. 2023). Dazu erheben sie in Anlehnung an die als Standard erachtete Imposter-Methode (Koppel & Winter 2014) eine große Menge von Merkmalen (120 000) und ziehen wiederholt daraus eine kleinere Stichprobe (10 000). Diese 10 000 Stichproben-Merkmale verwenden sie, um den Abstand zwischen einem Ausschnitt eines Romans von Bachmann und zufällig gewählten Ausschnitten aus Romanen von vier Autoren von Horrorromanen einschließlich King zu messen. Das wird für jeden Ausschnitt eines Romans von Bachmann tausendmal wiederholt. Jede dieser Messungen ergibt unterschiedliche Abstände der Kandidaten-Autoren zu Bachmann, die man als Rangfolge notieren kann, und tatsächlich zeigt sich, dass King sehr viel häufiger als die anderen Autoren den ersten Rang hat, also am nächsten am Bachmann-Text ist.

Bei Analysen dieser Art, so ist bereits aus der langen Tradition des *Information Retrieval* bekannt, ist die Bestimmung des Abstandsmaßes, mit dem die Distanzen zwischen den Vektoren gemessen werden, entscheidend. Ein wichtiges Maß dieser Art ist *Burrows Delta* (Burrows 2002), das sich als besonders robust erwiesen hat (Evert et al. 2017). Im Fall von sehr großen Textmengen haben sich auch Verfahren des überwachten maschinellen Lernens als sehr gut erwiesen (Savoy 2020).

Die Annahme, dass jede\*r Autor\*in, tatsächlich jeder Mensch, wenn er oder sie Sprache verwendet, eine einzigartige Weise der Sprachverwendung aufweist, oder in linguistischer Terminologie: einen Ideolekt hat, gehört zu den Grundthesen der Stilometrie (Nini 2023). Allerdings ist diese Annahme nicht notwendig, um die Arbeitsweise der Stilometrie zu begründen. Dafür reicht es aus, anzunehmen, dass es diskriminative Merkmale für eine gegebene Menge von Autor\*innen gibt. Der erste Punkt ist eine schwer beweisbare Annahme, der zweite Punkt ist für viele Konstellationen inzwischen empirisch gezeigt worden, woraus aber noch kein Anspruch auf Allgemeingültigkeit abzuleiten ist. Insgesamt sind die Befunde der Stilometrie, nämlich dass Texte eines\*r Autors\*in nachweisbare Gemeinsamkeiten aufweisen, ein Problem für Theorien, die Individualität als bloßes Diskursphänomen oder als Emergenzphäno-

men von Interaktionen modellieren (Jannidis 2014, Grand & McLeod 2018). Allerdings ist die manchmal in populären Kontexten verwendete Metapher vom *linguistischen Fingerabdruck* irreführend. Einerseits entwickelt sich der Ideolekt über die Zeit, wenn auch nur langsam, zum anderen wird für die Bearbeitung von stilometrischen Aufgaben eine korpus- und aufgabenspezifische Auswahl von linguistischen Merkmalen verwendet und nicht ein sich stets gleichbleibender Identitätsmarker. Insgesamt ist die Stilometrie noch eine weitgehend von Einzeluntersuchungen getriebene Wissenschaft, die aber in den letzten zehn Jahren zunehmend nach einer theoretischen und methodischen Fundierung strebt (z. B. Juola 2015, Nini 2023).

## 2. Anwendungen der Stilometrie auf religiöse Texte

1851 spekuliert der Mathematiker Augustus de Morgan, der heute als einer der Begründer der formalen Logik gilt, über die Möglichkeit, den Brief der Hebräer von den Paulusbriefen aufgrund der durchschnittlichen Länge der Wörter in den Briefen zu unterscheiden (De Morgan 1882, 215 f.). Diese Überlegung gilt heute als einer der Ursprünge der Stilometrie (Holmes 1998, 112). Stilometrische Verfahren sind dann auch früh auf biblische Texte angewandt worden, was von Anfang an mit sehr hohen Hoffnungen verbunden war. So schreibt Radday in einer Studie 1973: „Certain problems that have vexed Biblical scholars for centuries can at least be approached and possibly solved once for all by quantification.“ (Radday 1973, 273). Besonders im Fall der Forschungskontroversen über die Autorschaft bieten sich stilometrische Verfahren an, um neben den historischen und sprachgeschichtlichen Informationen eine weitere Informationsquelle in die Urteilsfindung einzubinden (Oakes 2014). Ein knapper Überblick über die Frage nach der Autorschaft der Paulusbriefe zeigt, wie lange stilometrische Verfahren bereits hier eingesetzt werden und wie sich die Verfahren über die Zeit verändert haben.

Die Frage, ob die vierzehn paulinischen Briefe wirklich alle vom historischen Paulus geschrieben worden sind oder ob einige von einem oder mehreren anderen Autoren stammen, insbesondere ob die beiden Korintherbriefe, der Galaterbrief und der Römerbrief tatsächlich von einem Autor verfasst wurden, gehört nicht nur wegen der Überlegungen von De Morgan zu den ältesten Themen der theologischen Stilometrie. Schon seit dem 18. Jahrhundert werden diese Autorschaftsfragen diskutiert, nicht zuletzt weil die Theolog\*innen deutliche stilistische Unterschiede wahrnahmen. Die breitgefächerte Forschung zu diesem Thema hat teilweise auch einfache quantitative Argumente mit Bezug auf die Frequenz eines Wortes beim Textvergleich verwendet. Dieser Vergleich wurde durch die von vielen genutzte Arbeit von Robert Morgenthaler deutlich erleichtert, in der die Häufigkeiten aller Worte in den Büchern des Neuen Testaments aufgelistet werden (Morgenthaler 1958). Ein Beispiel dafür ist etwa die Arbeit von Bujard zum Kolosserbrief. Bujard verwendet nur die einfachen

Zählwerte (oft auf der Basis von Morgenthaler) und macht auch keinen statistischen Test, um zu sehen, ob die beobachteten Unterschiede signifikant sind (Bujard 1973). Der Theologe Andrew Q. Morton brach in zweierlei Weise mit dieser etablierten Arbeitsweise: er verwendete den Computer und er griff auf statistische Verfahren zurück, um die Zuverlässigkeit seiner Vergleiche zu stützen.<sup>2</sup> Morton erregte 1963 Aufsehen, als er in einem Zeitungsartikel behauptete, dass seine stilometrischen Untersuchungen gezeigt hätten, dass nur vier Briefe vom gleichen Verfasser stammen würden (Korinther 1 u. 2, Galater, Römer), während die historische Theologie seiner Zeit davon ausging, dass auch der erste Brief an die Thessalonicher, der Brief an die Philipper und der Brief an Philemon dem historischen Paulus zuzuschreiben sind. Die heftige Kritik veranlasste ihn dazu, seine Methode und sein Ergebnis ausführlich darzustellen (Morton & McLeman 1966).

Eine angemessene Würdigung der Arbeit von Morton steht noch aus, da der Blick auf seine Leistung durch sein späteres Wirken überschattet ist. Seine neue Cusum-Technik sollte eine eindeutige Autorschaftsattribuierung auch von kurzen Texten erlauben und wurde in den 1990er Jahren in Großbritannien auch vor Gericht genutzt. Doch dem öffentlichen Erfolg standen die Zweifel von Fachkolleg\*innen gegenüber, die das statistische Verfahren und seine Anwendung durch Morton mit guten Gründen für höchst unzuverlässig hielten (Holmes 1998, 114). Diese Perspektive färbt bis heute den Blick auch auf die frühen Arbeiten von Morton.

In seiner Untersuchung von 1966 verwendet er im Wesentlichen zwei Textmerkmale: die durchschnittliche Satzlänge und häufige Wörter, sogenannte Funktionswörter. Er schließt die Artikel, die am häufigsten vorkommen, aus, da sie seiner Meinung nach zu sehr vom Thema des Textes abhängen, und konzentriert sich auf *kai* (und), die Partikel *de*, *en* (in), *einai* (sein) und *autos*. Zuerst stellt er anhand eines Korpus von altgriechischen Texten fest, ob diese Funktionswörter überhaupt dazu geeignet sind, zwischen Autoren eines Textes zu unterscheiden, indem er die Texte in gleich große Segmente teilt, die Häufigkeit des jeweiligen Funktionsworts erhebt und dann mit  $\chi^2$  die Wahrscheinlichkeit dafür berechnet, ob die beobachtete Differenz zwischen den Segmenten zufällig entstanden ist oder nicht – im letzteren Fall würde das für zwei Autoren sprechen. Alle Tests bestätigen das Resultat: 1 Kor, 2 Kor, Gal und Röm stammen von einem Autor. Philemon ist aufgrund seiner Kürze ein Problem, aber die Daten sprechen ebenfalls für die Zugehörigkeit zu dieser Gruppe. Mortons Ergebnisse bestätigen Thesen, die bereits im 19. Jahrhundert von Friedrich Christian Baur formuliert wurden (Neumann 1990, 2 ff.). Eine wesentliche Schwäche seiner Vorgehensweise ist der Umstand, dass jede Untersuchung lediglich eine Variable, z. B. die Häufigkeit von *kai*, einbezieht, während sich spätestens seit den 1980er Jahren multivariate Analysen in der oben beschriebenen Art als Standard durchgesetzt haben.

2 Neumann listet einige nicht beachtete stilometrische Arbeiten, die schon vor Morton mit statistischen Tests arbeiteten (Neumann 1990, Chap. II).



Neumann, dessen Arbeit erst 1990 erscheint, aber wohl 1980 abgeschlossen ist, erhebt 617 Merkmale, zu denen lexikalische, syntaktische u. a. m. gehören. Er macht eine ausführliche Voruntersuchung, inwieweit diese Merkmale in einer Diskriminanzanalyse ein Textkorpus zu 100 % richtig teilt und wählt dann für seine Untersuchung der Paulusbriefe nur die vier Merkmale (Wortlänge, unbestimmte Pronomina, Wörter, die mit dem Buchstaben *tau* beginnen, und die Position des ersten Substantivs im Satz) aus, die das geleistet haben. Sein Ergebnis lautet, dass die umstrittenen Briefe – Kol, Eph, 2 Thess – Paulus zugeschrieben werden. Die Anforderung, dass die Merkmale die Daten zu 100 % richtig teilen sollen, würde man heute als *overfitting* bezeichnen, da die Merkmalsauswahl für den Text-Datensatz genau passte, aber es unklar bleibt, wie gut sie für den eigentlichen Untersuchungsgegenstand passen.

Der Philosoph Antony Kenny, der 1982 eine Einführung in die Stilometrie und deren Statistik geschrieben hat (Kenny 1982), publizierte vier Jahre später seine einschlägigen Arbeiten zum Neuen Testament (Kenny 1986), in denen er sich auch mit dem Problem der Paulusbriefe beschäftigt. Er wählt dazu die meisten aus einem Set von 99 Merkmalen aus, die er für seine Arbeit entwickelt hat; dazu gehören Konjunktionen, Partikel, Präpositionen, die Fälle von Substantiven, Pronomina und Adjektiven u. a. m. zur Differenzierung der Texte. Kenny analysiert jedes dieser Merkmale einzeln und testet dann darauf, ob die Häufigkeiten sich in den fraglichen Texten signifikant unterscheiden. Er interpretiert seine Resultate, dass einige der Merkmale bei einigen Briefen eine andere Autorschaft zu indizieren scheinen, sehr zurückhaltend: Die beobachtete Varianz in zwölf der Briefe könnte auch dadurch erklärt werden, dass sie das Werk eines einzigen, extrem vielseitigen Autors seien. Mealand weist in seiner Rezension von Kenny daraufhin, dass das Ergebnis nicht unwesentlich von seiner Aufgabenstellung, nämlich ob ein Text besonders abweicht, abhängig ist, und zeigt, dass man die Korrelationen, die Kenny auflistet, auch so auswerten kann, dass die enge Verwandtschaft von Röm, 1 Kor, 2 Kor und Gal deutlich wird (Mealand 1988).

Die Untersuchungen in den 1990er Jahren verwenden multivariate statistische Verfahren. Greenwood gruppiert die Texte aufgrund der 10, 20 und 30 häufigsten Wörter und stellt fest, dass die Cluster sehr ähnlich sind (Greenwood 1993). Er verwendet außerdem ein Verfahren, das nicht nur Cluster bildet, sondern auch erlaubt festzustellen, ob sich die Cluster im hochdimensionalen Raum überschneiden, was dafür sprechen würde, dass es sich nicht wirklich um separierbare Einheiten handelt. Er kommt zu dem Ergebnis, dass dies der Fall ist, wenn sich die Struktur auch nicht in der PCA erhält, einem Verfahren zur Dimensionsreduktion, das die hochdimensionalen Daten zweidimensional darstellt. Anders ausgedrückt: Es gibt eine Informationsstruktur, die eine klare Separierung der Texte ermöglicht, aber sie ist nur im hochdimensionalen Raum sichtbar. Insgesamt sieht er in seinen Ergebnissen aber eine Stärkung der These Mortons, dass es gerade die Funktionswörter sind, die eine Autorschaftsattributions ermöglichen, eine These „not comfortably assimilated within the spirit of classical scholarship“ (217), obwohl sie eigentlich zur gleichen Zeit etwa durch die Arbeiten von Burrows in der Literaturwissenschaft etabliert wurde.



Ledger (1995) teilt die Texte in 1000 Wort-Abschnitte und erhebt für jeden Abschnitt 29 Merkmale, die sich zuvor bei der Analyse von Plato bewährt haben, 19 davon sind der Anteil von Worten mit einem bestimmten Buchstaben, neun mit dem Anteil von Worten mit einem bestimmten Endbuchstaben und das Standardmaß für die lexikalische Komplexität, Type-Token-Ratio. Er verwendet ebenfalls PCA und untersucht die so entstandenen Gruppierungen. Er erkennt ein zentrales Cluster mit 1 Kor, 2 Kor, Gal, Phil, 2 Thess und Röm sowie eine zweite Gruppe mit Kol, Eph, Hebr, die restlichen lassen sich nicht eindeutig zuordnen. Interessant sind seine Versuche, die Merkmale zu identifizieren, die etwa zur Sonderstellung von Hebr führen und diese wiederum auf die Texte zu beziehen.

Mealand verwendet in seiner Studie, die neben den Paulusbriefen auch weitere Briefe des NT untersucht, wie Neumann eine Diskriminanzanalyse sowie eine Faktorenanalyse zur Dimensionsreduktion. Grundlage sind 25 Merkmale, u. a. die häufigsten Wörter, die sich in früheren literarischen Studien und bei Neumann als unterscheidend erwiesen haben. Seine Ergebnisse bestätigen einige Erwartungen, aber verlaufen in anderen Punkten sehr entgegen dem, was die theologische Forschung sonst annimmt. Im Fall der Paulusbriefe wird der erste Abschnitt von Röm mit Kol und Eph gruppiert, allerdings zeigen sich nicht alle Ergebnisse in allen Untersuchungen.

Eine der neuesten Studien (Savoy 2019) überprüft drei Hypothesen, nämlich dass nur die bekannten vier Briefe von Paulus stammen, dass die üblichen sieben Briefe von Paulus stammen oder dass zehn der Briefe von Paulus stammen, aber 1 Tim, 2 Tim, Tit und Hebr nicht. Aufgrund aller Wörter, die häufiger als zweimal vorkommen, werden drei Abstandsmaße auf die 21 Texte (Paulusbriefe + weitere Briefe der Bibel) angewandt und das Ergebnis hierarchisch gruppiert in einem Dendogramm dargestellt. Auch bei einem Wechsel der Distanzmaße sehen wir bei den paulinischen Briefen eine Bestätigung der vier Briefe-Hypothese – und dass Kol und Eph von einem Autor stammen. Mit der sogenannten Imposter-Methode (Koppel & Winter 2014), die heute als der höchste Standard stilometrischer Technik gilt, da sie wie oben beschrieben in zahlreichen Iterationen die Abstandmaße aufgrund immer neuer Teilmengen der Merkmale erhebt, überprüft Savoy seine Ergebnisse. Einerseits bestätigt sich, dass Kol und Eph vom gleichen Autor stammen und auch, dass die üblichen vier eine Gruppe bilden, aber diese Gruppe hat eine deutliche Verbindung zu 1 Thess und 2 Thess sowie zu Phil. Das geht also einerseits über die These hinaus, dass nur vier Briefe von Paulus stammen, bestätigt aber nicht die übliche These, welche sieben Briefe von Paulus stammen. Der Autor kommt damit auch zu dem negativen Schluss, dass er keine der drei Hypothesen bestätigen kann.

Kritische Stimmen zu den stilometrischen Untersuchungen der Paulusbriefe weisen darauf hin, dass die unterschiedliche Menge und Art der Zitate aus anderen Texten die Werte im Fall der Lexik verfälschen könnten (allerdings entfernen einige Stilometriker die Zitate aus den Texten), dass die nachträgliche Einfügung von Satzzeichen die Verwendung von Satzlänge als Merkmal problematisch machen könnte, und auch, dass die Varianz sich durch die Verwendung unterschiedlicher Schreiber

erklären ließe. Insgesamt ist deutlich geworden, dass die stilometrischen Untersuchungen der Paulusbriefe ein ziemlich getreuer Spiegel der Entwicklung der Methode sind. Von einfachen Häufigkeitswerten über univariate Methoden bis hin zur multivariaten Statistik. Und selbst die Kritik der Methode ist in vielerlei Hinsicht repräsentativ.

Vielleicht mit Ausnahme der Frage nach Q, einer vermuteten Quelle für die Evangelien von Matthäus und Lukas (Poirier 2008, Oakes 2014, 153 ff., Mealand 2011), hat kaum ein anderes Problem der christlichen Theologie so viel Aufmerksamkeit der Stilometrie auf sich gezogen wie die Paulusbriefe. Aber dennoch sind viele Fragen bezüglich des Alten und des Neuen Testaments mit stilometrischen Methoden bearbeitet worden, etwa ob der erste und der zweite Brief des Petrus wirklich unterschiedliche Verfasser haben (House 2002), oder ob man die bezweifelte Einheit des Buchs Jesaja auf diese Weise genauer beleuchten kann (Radday 1973).

Die stilometrische Analyse religiöser Texte hat sich außerdem mit einer Fülle von Fragen jenseits der Texte der Bibel beschäftigt, die hier nur angedeutet werden kann. So hat man mögliche Autoren des Buch Mormons ermittelt (Holmes 1992, Jockers et al. 2008), untersucht, ob der Autor des Korans identisch ist mit dem Autor des Hadith, den Überlieferungen der Aussprüche und Handlungen des Propheten Mohammed (Sayoud 2012), ist der Frage nachgegangen, ob das Nahj Al-Balagha, das dem Imam Ali zugeschrieben wird, zwei Autoren zugeschrieben werden kann (Sarwar & Mohamed 2022) und hat frühe buddhistische Literatur analysiert, die aus Indien kam und ins Chinesische übersetzt wurde (Hung et al. 2010).

## Verwendete Literatur

- Bujard, W. (1973). *Stilanalytische Untersuchungen zum Kolosserbrief*. Göttingen: Vandenhoeck & Ruprecht.
- Burrows, J. F. (1987). *Computation into Criticism. A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.
- Ders. (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> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.1093/llc/8.4.211> (zuletzt aufgerufen am 16.06.2024).
- Hirst, G., & Wei Feng, V. (2012). Changes in Style in Authors with Alzheimer's Disease, *English Studies*, 93(3), 357–70. <https://doi.org/10.1080/0013838X.2012.668789> (zuletzt aufgerufen am 16.06.2024).
- 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.
- Ders. (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), 119–34.
- Jannidis, F. (2014). Der Autor ganz nah. Autorstil in Stilistik und Stilometrie. In M. Schaffrick & M. Willand (Hrsg.), *Theorien und Praktiken der Autorschaft* (S. 169–195). Berlin: de Gruyter.
- Ders., & Lauer, G. (2014). Burrows Delta and its Use in German Literary History. In M. Erlin & L. Tatlock (Hrsg.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century* (S. 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.
- Ders. (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.
- Ders. (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 edition. 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.

- Ders. (1996). The Extent of the Pauline Corpus. A Multivariate Approach, *Journal for the Study of the New Testament*, 18(59), 61–92.
- Ders. (2011). Is There Stylometric Evidence for Q?, *New Testament Studies*, 57(4), 483–507.
- Modrall Sperling, D. H., Kestemont, M., & Neyt, V. (2024). The Authorship of Stephen King's Books Written Under the Pseudonym „Richard Bachman“. A Stylo-metric Analysis, *Journal of Computational Literary Studies*, 2(1), o. S. <https://doi.org/10.5281/zenodo.8397200> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.1145/3132039> (zuletzt aufgerufen am 16.06.2024).
- Neumann, K. J. (1990). *The Authenticity of the Pauline Epistles in the Light of Stylo-statistical 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 ders. (Hrsg.), *Literary Detective Work on the Computer* (S. 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 (Hrsg.), *Retracing the History of Literary Translation in Poland* (S. 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.
- Ders. (2020), *Machine Learning Methods for Stylometry. Authorship Attribution and Author Profiling*. Cham: Springer. <https://doi.org/10.1007/978-3-030-53360-1> (zuletzt aufgerufen am 16.06.2024).
- 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), o. S. <https://doi.org/10.22148/001c.37588> (zuletzt aufgerufen am 16.06.2024).

Stamatatos, E. (2009). A Survey of Modern Authorship Attribution Methods, *Journal of the American Society for Information Science and Technology*, 60(3), 538–556.

## Weiterführende Literatur

- 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. (Hrsg.), *Isaiah and the Prophets. Inspired Voices from the Old Testament* (S. 151–164). Provo, UT: Religious Studies Center, Brigham Young University. URL: <https://rsc.byu.edu/isaiah-prophets/scientific-analysis-isaiah-authorship> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.1093/llc/fqu038> (zuletzt aufgerufen am 16.06.2024).
- Barr, G. K. (2003). Two Styles in the New Testament Epistles, *Literary and Linguistic Computing*, 18(3), 235–248. <https://doi.org/10.1093/llc/18.3.235> (zuletzt aufgerufen am 16.06.2024).
- 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 (Hrsg.), *The translator and the computer* (S. 155–70). Wrocław: WSF.
- Elewa, A. (2019). Authorship verification of disputed Hadiths in Sahih al-Bukhari and Muslim, *Digital Scholarship in the Humanities*, 34(2), 261–276. <https://doi.org/10.1093/llc/fqy036> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.1093/llc/fqac025> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.1093/llc/10.3.183> (zuletzt aufgerufen am 16.06.2024).

- 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. <https://doi.org/10.1515/jlt-2015-0003> (zuletzt aufgerufen am 16. 06. 2024).
- Kilpatrick, G. D. (1988). Review of A Stylometric Study of the New Testament, *Novum Testamentum*, 30(4), 373–375. <https://doi.org/10.2307/1560626> (zuletzt aufgerufen am 16. 06. 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 Stylo-metric Analysis. In M. Bramer & M. Petridis (Hrsg.), *Research and Development in Intelligent Systems XXXII* (S. 91–106). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-25032-8\\_6](https://doi.org/10.1007/978-3-319-25032-8_6) (zuletzt aufgerufen am 16. 06. 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. <https://doi.org/10.1093/llc/fqy013> (zuletzt aufgerufen am 16. 06. 2024).
- Koppel, M., Schler, J., & Argamon, Sh. (2011). Authorship Attribution in the Wild, *Language Resources and Evaluation*, 45(1), 83–94. <https://doi.org/10.1007/s10579-009-9111-2> (zuletzt aufgerufen am 16. 06. 2024).
- Dies., & Winter, Y. (2012). The „Fundamental Problem“ of Authorship Attribution, *English Studies*, 93, 284–291. <https://doi.org/10.1080/0013838X.2012.668794> (zuletzt aufgerufen am 16. 06. 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> (zuletzt aufgerufen am 16. 06. 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> (zuletzt aufgerufen am 16. 06. 2024).
- Lichtblau, D., & Stoean, C. (2023). Chaos game representation for authorship attribution, *Artificial Intelligence*, 317(103858), o. S. <https://doi.org/10.1016/j.artint.2023.103858> (zuletzt aufgerufen am 16. 06. 2024).
- Linmans, A. J. M. (1998). Correspondence Analysis of the Synoptic Gospels, *Literary and Linguistic Computing*, 13(1), 1–13. <https://doi.org/10.1093/llc/13.1.1> (zuletzt aufgerufen am 16. 06. 2024).
- Mealand, D. (1989). Positional Stylometry Reassessed. Testing a Seven Epistle Theory of Pauline Authorship, *New Testament Studies*, 35(2), 266–286. <https://doi.org/10.1017/S0028688500024656> (zuletzt aufgerufen am 16. 06. 2024).
- Ders. (1995). Correspondence Analysis of Luke, *Literary and Linguistic Computing*, 10(3), 171–182. <https://doi.org/10.1093/llc/10.3.171> (zuletzt aufgerufen am 16. 06. 2024).



- Ders. (1997). Measuring genre differences in Mark with correspondence analysis, *Literary and Linguistic Computing*, 12(4), 227–245. <https://doi.org/10.1093/lc/12.4.227> (zuletzt aufgerufen am 16.06.2024).
- Ders. (1999). Style, genre, and authorship in acts, the Septuagint, and Hellenistic historians, *Literary and Linguistic Computing*, 14(4), 479–506. <https://doi.org/10.1093/lc/14.4.479> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.10520/EJC121513> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.1093/lc/13.4.195> (zuletzt aufgerufen am 16.06.2024).
- 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> (zuletzt aufgerufen am 16.06.2024).
- 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 (Hrsg.), *New Approaches to Textual and Image Analysis in Early Jewish and Christian Studies* (S. 88–117). Leiden/Boston: Brill [= *Digital Biblical Studies*, 5]. [https://doi.org/10.1163/9789004515116\\_007](https://doi.org/10.1163/9789004515116_007) (zuletzt aufgerufen am 16.06.2024).
- Royal, K. (2012). Using Objective Stylometric Techniques to Evaluate New Testament Authorship, *Journal of MultiDisciplinary Evaluation*, 8(19), 1–7. <https://doi.org/10.56645/jmde.v8i19.352> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.5220/0010970300003116> (zuletzt aufgerufen am 16.06.2024).
- 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. <https://doi.org/10.1093/lc/fqado38> (zuletzt aufgerufen am 16.06.2024).