Benjamin Krautter

Über die Attribution hinaus

Forschungsperspektiven der Stilometrie als Anwendungsfeld in der Literaturwissenschaft

Abstract: In computational literary studies, stylometrics are most commonly used for authorship attribution studies. This article argues that there is a potential use for stylometrics in literary studies beyond the mere detection of authorship. To show this, several current research approaches are analysed and portrayed. However, the utilization of stylometrics bear limitations, such as its relative nature – the comparisons are only valid within the analyzed corpus – and its susceptibility for >cherry picking<.

1 Einführung

Die Stilometrie, die auf der Basis von Wortfrequenzlisten relative Ähnlichkeiten innerhalb eines gewählten Textkorpus abbildet, ist ein potentielles Anwendungsfeld quantitativer Methoden in der Literaturwissenschaft. Traditionelles Einsatzgebiet der Stilometrie ist die Autorschaftsattribution. Ob stilometrische Analysen literaturwissenschaftliche Fragestellungen auch über die Zuschreibung von Autorschaft hinaus bereichern oder sogar anleiten können, soll nachfolgend anhand ausgewählter Forschungsbeiträge diskutiert werden.

Gabriel Viehhauser definiert Stilometrie im Rahmen der Digital Humanities als »die Anwendung quantitativer Methoden zur Erfassung und Klassifizierung stilistischer Merkmale von Texten«.¹ Er markiert damit die Verbindung von quantitativen Verfahren, die digital realisiert werden, und ihrem Zielgegenstand, den stilistischen Textmerkmalen. Ob der ohnehin schwer zu fassende Begriff >Stil« eine gute Wahl ist, um quantitative Methoden wie das Auszählen und Verglei-

¹ Gabriel Viehhauser: »Historische Stilometrie? Methodische Vorschläge für eine Annäherung textanalytischer Zugänge an die mediävistische Textualitätsdebatte«, in: *Grenzen und Möglichkeiten der Digital Humanities. Sonderband der Zeitschrift für digitale Geisteswissenschaften* 1, hg. v. Constanze Baum und Thomas Stäcker. 2015, http://www.zfdg.de/sb001_009 (30.10.2017).

^{© 2018} Benjamin Krautter, publiziert von De Gruyter.

chen von Wortlisten zu beschreiben, bleibt jedoch zu diskutieren; zumal die Vorstellung, jeder Autor habe einen eindeutigen stilistischen Fingerabdruck, falsch zu sein scheint und der Autorstilk eher auf vielen kleinen Signalen beruhe.²

Die Stilometrie ist gleichfalls keine genuin digitale Methode der Textanalyse. Sie geht aus der Analyse des Sprachstils mit Hilfe der Statistik hervor und datiert bis ins 19. Jahrhundert zurück.3 Als einer ihrer Vorreiter gilt der Mathematiker Augustus de Morgan, der 1851 anregte, statistische Messwerte wie die durchschnittliche Wortlänge als Kriterium zur Bestimmung der Autorschaft anonymer Texte heranzuziehen:

Specifically, de Morgan concentrated on the number of letters per word and suspected that the average length of words in different Epistles by St. Paul might shed some light on the question of authorship; generalizing his ideas, he assumed that the average word lengths in two texts, written by one and the same author, though on different subjects, should be more similar to each other than in two texts written by two different individuals on one and the same subject.4

Andere Wissenschaftler wie Thomas Corwin Mendenhall oder Wincenty Lutosławski folgen de Morgans Beispiel mit eigenen stilometrischen Arbeiten. Sie

² Vgl. Fotis Jannidis: »Der Autor ganz nah. Autorstil in Stilistik und Stilometrie«, in: Theorien und Praktiken der Autorschaft, hg. v. Matthias Schaffrick und Marcus Willand. Berlin, Boston 2014, S. 169-195, hier 183, John F. Burrows: »Delta«: a Measure of Stylistic Difference and a Guide to Likely Authorship«, in: Literary and Linguistic Computing 17.3 (2002), S. 267–287, hier 268. und Patrick Juola: »Authorship Attribution«, in: Foundations and Trends in Information Retrieval 1.3 (2006), S. 233–334, hier 239.

³ Vgl. Juhan Tuldava: »Stylistics, Author Identification«, in: Quantitative Linguistics, hg. v. Gabriel Altmann, Reinhard Köhler und Rajmund G. Piotrowski. Berlin 2005, S. 368-387, hier 370. Vgl. Fotis Jannidis und Gerhard Lauer: »Burrows's Delta and Its Use in German Literary History«, in: Distant Readings. Topologies of German Culture in the Long Nineteenth Century, hg. v. Matt Erlin und Lynne Tatlock. Rochester 2014, S. 29-54, hier 31. Einen Abriss der Geschichte der Digital Humanities – anfangs Humanities Computing –, deren Ursprünge zumeist mit Roberto Busa verbunden werden, gibt Manfred Thaller. Vgl. Manfred Thaller: »Geschichte der Digitial Humanities«, in: Digital Humanities. Eine Einführung, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein. Stuttgart 2017, S. 3–12.

⁴ Peter Grzybek: »History and Methodology of Word Length Studies«, in: Contributions to the Science of Text and Language, hg. v. Peter Grzybek. Dordrecht 2007, S. 15-90, hier 15. Jannidis und Lauer folgen der Darstellung von Grzybek. Dazu Jannidis, Lauer: »Burrows's Delta and Its Use in German Literary History«, S. 31. Vgl. auch Christof Schöch: »Corneille, Molière et les autres. Stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik«, in: Literaturwissenschaft im digitalen Medienwandel. Beihefte zu Philologie im Netz 7 (2014), S. 130–157, hier 133 und Juola: »Authorship Attribution«, S. 240.

untersuchen etwa die Chronologie von Platons Schriften oder die lange Zeit lebhaft diskutierte Kontroverse um die Textzugehörigkeit bei William Shakespeare und Christopher Marlowe. Die Zuordnung einzelner literarischer Texte zu einem Autor versuchen Mendenhall und Lutosławski durch messbare stilistische Veränderungen innerhalb der Textkorpora zu bestimmen. Gleiches gilt für die zeitliche Abfolge, in der die Texte verfasst wurden.⁵

Mendenhalls Idee, »peculiarities of style in composition«6 über das Auszählen von Wortlängen graphisch darstellbar zu machen, erscheint kaum gealtert.⁷ Die Entwicklungslinie fortschreitend, identifiziert Efstathios Stamatatos die 1964 von Frederick Mosteller und David L. Wallace veröffentlichte Studie Inference and Disputed Authorship: The Federalist⁸ als »undoubtedly the most influential work in authorship attribution«9. Die Analyse von Mosteller und Wallace basiert auf einem kleinen Korpus von Funktionswörtern und prägt die Forschung bis in die 1990er Jahre.10

Die ursprüngliche Idee der Stilometrie, die statistisch gestützte Analyse stilistischer Merkmale von Texten, bleibt auch in ihrer computergestützen Form deutlich erkennnbar. Maciej Eder, Jan Rybicki und Mike Kestemont verdeutlichen dies in ihrem Aufsatz »Stylometry with R: A Package for Computational Text Analysis«: »Stylometry deals with the relationship between the writing style in

⁵ Vgl. Jannidis, Lauer: "Burrows's Delta and Its Use in German Literary History", S. 31, 54. Mit dem Titel von einem seiner Aufsätze prägt Lutosławski zudem den Begriff ›Stilometrie‹. Vgl. Wincenty Lutosławski: »Principes de stylométrie«, in: Revue des études grecques 41 (1890), S. 61–80. Vgl. Wincenty Lutosławski: The Origin and Growth of Plato's Logic. With an Account of Plato's Style and of the Chronology of his Writings. London, New York, Bombay 1897. Vgl. auch Grzybek: »History and Methodology of Word Length Studies«, S. 15-17. Die Debatte um Shakespeare und Marlowe ist auch mit Blick auf die Stilometrie noch immer aktuell: »Interestingly, the methods of stylometry are also actively applied in the Humanities, where multiple historic authorship problems in literary studies still seek a definitive solution - the notorious Shakespeare-Marlowe controversy is perhaps the best example in this respect.« Maciej Eder, Mike Kestemont und Jan Rybicki: »Stylometry with R: A Package for Computational Text Analysis«, in: The R Journal 8.1 (2016), S. 107–121, hier 107.

⁶ Thomas Corwin Mendenhall: »A Mechanical Solution of a Literary Problem«, in: The Popular Science Monthly 60 (1901), S. 97-105, hier 97.

⁷ Die Unterscheidung von Wortlängen erscheint heutzutage jedoch veraltet. Juola hält 2006 mit Bezug auf eine Studie von M. W. A. Smith fest: »[A]verage word length is neither stable within a single author, nor does it distinguish between authors. « Juola: »Authorship Attribution «, S. 240.

⁸ Vgl. Frederick Mosteller und David L. Wallce: Inference and Disputed Authorship: The Federalist. Reading, MA 1964.

⁹ Efstathios Stamatatos: »A Survey of Modern Authorship Attribution Methods«, in: American Society for Information Science and Technology 60.3 (2009), S. 538–556, hier 538.

¹⁰ Vgl. ebd.

texts and meta-data about those texts (such as date, genre, gender, authorship).«11 Zu einer vergleichbaren Definition gelangt Christof Schöch, der Stilometrie als »computergestütztes Verfahren der quantitativen Erhebung stilistischer Merkmale für die Klassifikation von Texten«12 definiert. Eine Verengung in der Wahrnehmung der Stilometrie ist dennoch greifbar: die Verengung auf computergestützte Verfahren. Dieser Umstand ist einerseits schon dem Kontext der Definition von Viehhauser zu entnehmen: der Rahmung innerhalb der Disziplin Digital Humanities. Schöch verweist andererseits explizit auf die mindestens unterstützende Funktion des Computers. Dasselbe gilt für Kestemont, der die Rolle des Computers gleich mehrfach betont: »Computational authorship attribution is a popular application in current stylometry, the computational study of writing style.«13

2 Perspektiven für die Literaturwissenschaft

Auch wenn derzeit Verwendungsperspektiven über die Attribution von Autorschaft hinaus sichtbar werden, etwa für die Literaturgeschichtsschreibung oder die Gattungstheorie, bleibt die Autorschaftsattribution doch der zentrale Untersuchungsgegenstand stilometrischer Analysen. 14 Ihre Relevanz zeigt sich exemplarisch an den neuesten Ergebnissen von David Hoover, Jan Rybicki, Justin Anthony Stover und anderen. 15 Spätestens John Burrows' Aufsatz »> Delta« a Measure of Stylistics Difference and a Guide to Likely Authorship«, der von Stefan

¹¹ Eder, Kestemont, Rybicki: »Stylometry with R«, S. 107.

¹² Schöch: Corneille, »Corneille, Molière et les autres«, S. 133.

¹³ Mike Kestemont: »Function Words in Authorship Attribution. From Black Magic to Theory?«, in: Proceedings of the Third Workshop on Computational Linguistics for Literature (2014), S. 59-66, hier 59. Eder, Kestemont und Rybicki formulieren ganz ähnlich: »Authorship attribution plays a prominent role in the nascent field of stylometry, or the computational analysis of writing style.« Eder, Kestemont, Rybicki: »Stylometry with R«, S. 107, vgl. auch 108.

¹⁴ Vgl. Eder, Kestemont, Rybicki: »Stylometry with R«, S. 108 und Schöch: »Corneille, Molière et les autres«, S. 133. Weniger festgelegt auf die Attribution formulieren Jeroen Deploige, Mike Kestemont und Sara Moens: »Today's stylometry has become an umbrella term for a still growing number of techniques for authorship analysis.« Jeroen Deploige, Mike Kestemont und Sara Moens: »Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux«, in: Digital Scholarship in the Humanities 30.2 (2015), S. 199-

¹⁵ Hoover, Kestemont und Rybicki nutzen Rolling Delta als Methode, um stilometrische Signale von zwei oder mehr Autoren in kollektiv geschriebenen Texten aufzudecken. Sie untersuchen

Evert u. a. als »groundbraking«16 bezeichnet wurde, initiierte auch eine weitreichende Methodendiskussion.17

Texte von Joseph Conrad und Ford Madox Ford und weisen den beiden Schriftstellern plausibel einzelne Textteile zu. Vgl. David Hoover, Mike Kestemont und Jan Rybicki: »Collaborative Authorship: Conrad, Ford and Rolling Delta«, in: Literary and Linguistic Computing 29.3 (2014), S. 422-431. Justin Stover, Mike Kestemont, Yaron Winter und Moshe Koppel erproben Verfahren der Autorschaftsverifikation. Im Gegensatz zur Attribution legen keine philologischen Vorannahmen nahe, dass der Autor des Textes tatsächlich in der Kandidatenliste vorhanden ist. Sie ordnen dem afrikanischen Autor Apuleius einen kürzlich entdeckten lateinischen Text zu, der als einer der weltweit ersten Romane gilt. Vgl. Mike Kestemont, Moshe Koppel, Justin Anthony Stover und Yaron Winter: »Computational Authorship Verification Method Attributes a New Work to a Major 2nd Century African Author«, in: Journal of the Association for Information Science and Technology 67.1 (2016), S. 239-242. Zusammen mit Folgert Karsdorp und Walter Daelemans authentifizieren Kestemont, Stover und Koppel das Corpus Caesarianum. Sie untersuchen die Rolle, die Cäsars General Aulus Hirtius bei der Verfassung des Werks innehatte. Hirtius behauptet, das Corpus Caesarianum selbst mitgestaltet zu haben. Der Aufsatz legitimiert diese Behauptung. Vgl. Walter Daelemans, Folgert Karsdorp, Mike Kestemont, Moshe Koppel und Justin Stover: »Authenticating the Writings of Julius Caesar«, in: Expert Systems With Applications 63 (2016), S. 86-96. Jeroen Deploige und Sara Moens widmen sich zusammen mit Kestemont einer stilometrischen Analyse der Texte von Hildegard von Bingen und ihrem Schreiber Guibert de Gembloux. Sie kommen zum Schluss, dass Guibert die diktierten Texte stilistisch soweit bearbeitet haben muss, dass computergestützte Analysen sie ihm zuschreiben. Vgl. Deploige, Kestemont, Moens: »Collaborative Authorship in the Twelfth Century«, S. 199-224.

16 Stefan Evert, Fotis Jannidis, Steffen Pielström, Thomas Proisl, Christof Schöch, Thorsten Vitt: Explaining Delta, or: How Do Distance Measures for Authorship Attribution Work? Abstract, http://www.stefan-evert.de/PUB/EvertProislEtc2015_abstract.pdf (30.01.2017). Dazu auch Jannidis, Lauer: »Burrows's Delta and Its Use in German Literary History«, S. 32: »Since John F. Burrows introduced the Delta measure in his groundbreaking paper of 2002, it has become a common standard in stylometry for establishing the relative stylistic difference between two or more texts«.

17 Kestemont geht in seinem Aufsatz »Function Words in Authorship Attribution. From Black Magic to Theory?« der noch immer strittigen Frage nach, warum gerade Funktionswörter ein guter Indikator für Autorstil und das Gelingen stilometrischer Analysen seien. Vgl. Kestemont: »Function Words in Authorship Attribution«, S. 59-66. David Hoover testet die Effektivität von Burrows's Delta größtenteils hinsichtlich der zu wählenden Zahl an >most frequent words< und konstatiert, dass die Akkuratesse zumindest bis zu einer Anzahl von 600 betrachteten Wörtern zunimmt. Vgl. David Hoover: »Testing Burrows's Delta«, in: Literary and Linguistic Computing 19.4 (2004), S. 453–475. Ein ähnliches Anliegen verfolgt Efstathios Stamatatos mit seinem Überblick über automatische Ansätze der Autorschaftszuschreibung. Seinen Fokus legt Stamatatos auf rechentechnische Voraussetzungen und Konfigurationen. Vgl. Stamatatos: »A Survey of Modern Authorship Attribution Methods«, S. 538–556. In eine andere Richtung argumentieren Argamon Shlomo, Moshe Koppel und Jonathan Schler. »[R]eal-life authorship attribution problems are rarely as elegant as straightforward text categorization problems, in which we have a small closed set of candidate authors and essentially unlimited training text for each. A number of Einen weit über wissenschaftliche Debatten hinaus bekannt gewordenen Fall erfolgreicher Attribution stellt die Zuordnung von Joanne K. Rowling zu ihrem unter dem Pseudonym Robert Gallbraith veröffentlichten Roman *The Cuckoo's Calling* dar. Durch einen anonymen Tipp angestoßen, untersuchten Peter Millican und Patrick Juola den Roman mit stilometrischen Verfahren und kamen unabhängig voneinander zu dem Schluss, dass Joanne K. Rowling die wahrscheinlichste Autorin sei. In seinem Aufsatz »The Rowling Case: A Proposed Standard Analytical Protocol for Authorship Questions« zeichnet Juola sein Vorgehen nach. Zugleich argumentiert er für explizite Standards, denn Verfahren der Autorschaftszuschreibung folgten zwar impliziten Vorannahmen, diese müssten aber reflektiert und formalisiert werden. 19

Gleichzeitig entsteht aus literaturwissenschaftlicher Warte bisweilen der Eindruck, stilometrische Verfahren wären ausschließlich für die Autorschaftsattribution zu gebrauchen. Das Interesse an einer plausiblen Zuschreibung von Autor zu Text oder Text zu Autor verfügt zwar über einen intrinsischen Mehrwert, den

varieties of attribution problems fall short of this ideal«. Sie führen drei Beispiele an, die nicht in das für Autorschaftsattribution prototypische Muster einer geschlossenen Reihe möglicher Autoren fallen. Dieses Muster wird oftmals für methodologische Überlegungen angenommen, etwa wenn Kestemont die Rolle von Funktionswörtern in stilometrischen Analysen für verschiedene Sprachen reflektiert. Den drei Beispielen lassen sich drei Problemen zuordnen: Profiling, Nadel im Heuhaufen und Verifizierung. Entweder gibt es keine Eingrenzung der möglichen Kandidaten, eine große Zahl von Kandidaten, für die es gleichzeitig nur einen geringen Textumfang gibt, oder es gibt einen Kandidaten, aber keine Eingrenzung auf weitere »Verdächtige«. Shlomo Argamon, Moshe Koppel und Jonathan Schler: »Computational Methods in Authorship Attribution«, in: *Journal of the American Society for Information Science and Technology* 60.1 (2009), S. 9–26, hier 9.

¹⁸ Vgl. Fotis Jannidis: »Der Autor ganz nah. Autorstil in Stilistik und Stilometrie«, in: *Theorien und Praktiken der Autorschaft*, hg. v. Matthias Schaffrick und Marcus Willand. Berlin, Boston 2014, S. 169–195, hier S. 169f. Vgl. auch Eder, Rybicki, Kestemont: »Stylometry with R«, S. 107. Eder u. a. weißen hier zugleich auf eine weitere prominente Debatte hin, die die Rollen von Harper Lee und ihrem Herausgeber in der Originalversion von *To Kill a Mocking Bird* diskutiert.

¹⁹ Vgl. Patrick Juola: »The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions«, in: *Digital Scholarship in the Humanities* 30.1 (2015), S. 100–113.

noch sollte eine literaturwissenschaftliche Arbeit hermeneutische Fragestellungen, ²⁰ die an die Attribution anknüpfen, nicht vernachlässigen. Gabriel Viehhauser erkennt zutreffend, dass die traditionelle Literaturwissenschaft²¹ kaum bereit sein werde, sich auf quantifizierende Verfahrensweisen einzulassen, wenn diese keinen Anschluss an relevante literaturwissenschaftliche Probleme suche.²² Damit seien Studien zur Autorschaftsattribution mit Hilfe computergestützter Verfahren keinesfalls unter Generalverdacht gestellt oder mit der Frage »So What?«23 gleichgültig bei Seite geschoben. Gleiches gilt für die noch weiter gefasste Kritik von Stanley Fish und Adam Kirsch, Fish polemisiert gegen die Digital Humanities aus traditioneller literaturwissenschaftlicher Position:

But whatever vision of the digital humanities is proclaimed, it will have little place for the likes of me and for the kind of criticism I practice: a criticism that narrows meaning to the significances designed by an author, a criticism that generalizes from a text as small as half a line, a criticism that insists on the distinction between the true and the false, between what is relevant and what is noise, between what is serious and what is mere play.²⁴

²⁰ Ich fasse Hermeneutik hier als ein begründetes Verstehen und Interpretieren von Texten. Vgl. Klaus Weimar: »[Art.] Hermeneutik«, in: Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte, hg. v. Harald Fricke. Berlin, New York 2007, S. 25–29, hier 25. Vgl. zur Beliebigkeit von Interpretationen Lutz Danneberg: »Philosophische und methodische Hermeneutik«, in: Philosophia Naturalis 32.2 (1995), S. 249-269, hier 255-264.

²¹ Wie genau traditionelle Literaturwissenschaft oder traditionelle Methoden der Literaturwissenschaft zu fassen sind, erläutert Viehhauser nicht. Gleiches gilt für die später aufgeführte Kritik von Stanley Fish, der die Richtung des interpretativen Vorgehens zur Unterscheidung heranzieht. Vgl.: Stanley Fish: »Mind Your P's and B's: The Digital Humanities and Interpretation«, in: The New York Times (23.01.2012), dann: https://opinionator.blogs.nytimes.com/2012/01/23/ mind-your-ps-and-bs-the-digital-humanities-and-interpretation (28.10.2017). Adam Kirsch sieht in der Philologie die traditionellste Art textbasierter Wissenschaft. Vgl. Adam Kirsch: Technology Is Taking Over English Departments. The False Promise of the Digital Humanities, 02.05.2014, https://newrepublic.com/article/117428/limits-digital-humanities-adam-kirsch (27.10.2017).

²² Vgl. Viehhauser: »Historische Stilometrie?«, o. S.

^{23 »}So What?« ist der Titel eines Blogeintrages von Matthew L. Jockers, in dem er die gleichgültige Haltung des Journalisten Dana Mackenzie seinen quantitativen und empirischen Methoden gegenüber aufgreift. Daran anschließend argumentiert er, warum die Bestätigung bisheriger Annahmen aus einer neuen Perspektive sinnvoll ist, auch wenn das Ergebnis (scheinbar) das gleiche bleibt. Vgl. Matthew L. Jockers: So What?, 07.05.2014, http://www.matthewjockers.net/ 2014/05/07/so-what/ (27.10.2017).

²⁴ Fish: »Mind Your P's and B's«, o. S. Fish macht vor allem eine Beobachtung hinsichtlich des interpretativen Vorgehens stark, wodurch sich Literaturwissenschaft von Digital Humanities unterscheide: »The direction of my inferences is critical: first the interpretative hypothesis and then

Aus einer anderen Blickrichtung und noch schärfer greift Kirsch die Digital Humanities an: Traditionelle Geisteswissenschaften müssten sich gar gegen die Gefahr zur Wehr setzen, die von den Digitial Humanities ausgehe.²⁵ »Does the digital component of digital humanities give us new ways to think, or only ways to illustrate what we already know?«, fragt Kirsch und gibt die Antwort selbst, indem er quantitativ gestützte Ergebnisse als »fancy reiterations of conventional wisdom« fasst.26 Warum die Bestätigung anerkannter Thesen durch eine neue Methode nicht negativ zu bewerten sei, zeigt indes die Reaktion von Jockers: »[C]omputational work in the humanities can be simply about testing, rejecting, or reconfirming, what we think we already know. And I think that is a good thing!«27 Die Beweisführung in geisteswissenschaftlichen Fragestellungen sei niemals komplett abgeschlossen. Neue Hinweise, die bereits ausgeführte Thesen stützen, haben genau deshalb einen Mehrwert. Nicht umsonst formulieren Literaturwissenschaftler häufig den Wunsch der Annäherung an ein Desiderat.

Sowohl zur Kritik von Fish und Kirsch als auch auf die Antwort von Jockers passend, entgegnen Jeroen Deploige, Mike Kestemont und Sara Moens mit Blick auf das Problem gemeinschaftlicher Autorschaft im Korpus von Hildegard von Bingen: »As neither traditional stylistic analysis nor contextual historical research has so far been able to resolve the problem, we will approach this issue through a stylometric analysis.«28 Weder philologische noch kontextgestützte historische Arbeit könne bisher zufriedenstellend klären, ob die beiden Texte Visio de Sancto Martino und Visio ad Guibertum missa Hildegard von Bingen oder ihrem Schreiber Guibert de Gembloux zuzuschreiben sind.29 Einer der Gründe dafür könnte die spezifische Schreibkultur des Mittelalters sein:

Writing in the Middle Ages meant entering into a dialogue with a long line of predecessors, whether through citations, paraphrasing, or allusions. In the actual process of literary composition too, medieval authors only seldom worked alone. A >new< text could be the result

the formal pattern, which attains the status of noticeability only because an interpretation already in place is picking it out. The direction is the reverse in the digital humanities: first you run the numbers, and then you see if they prompt an interpretative hypothesis. The method, if it can be called that, is dictated by the capability of the tool.« Die Generalisierung von Fish halte ich an dieser Stelle für überspitzt.

²⁵ Vgl. Adam Kirsch: »Technology Is Taking Over English Departments«, o. S.

²⁶ Ebd.

²⁷ Vgl. Jockers: »So What?«, o. S. Vgl. auch nachfolgend.

²⁸ Deploige, Kestemont, Moens: "Collaborative Authorship in the Twelfth Century", S. 200. Computergestützte Analyseverfahren sind hier bereits im Begriff Stilometrie mitgedacht.

²⁹ Vgl. ebd.

of drafts on wax tablets copied by professional scribes, of processes of dictation and subsequent correction, etc.30

Deploige, Kestemont und Moens zeigen in ihrem Aufsatz, dass eine auktoriale Unterscheidung im beschriebenen Korpus möglich ist. Stilometrische Methoden erlauben es, die Differenzen im Schreibstil von Hildegard von Bingen und Guibert de Gembloux zu modellieren. Davon ausgehend spricht vieles dafür, dass es mit Hilfe der Stilometrie möglich sei, zumindest oberflächliche Eingriffe von Schreibern in den ursprünglichen Text auszuklammern.³¹ Zugleich stützen die Ergebnisse eine der im Vorfeld philologisch erarbeiteten Vermutungen, nämlich dass Guibert de Gembloux die Texte zumindest sehr tiefgreifend überarbeitet habe.³² Die empirische Plausibilisierung macht deutlich, dass philologische Problemstellungen und computergestützte quantitative Methoden durchaus das Potential einer ertragreichen Zusammenarbeit haben.

3 Methodologische Implikationen und philologische Vorannahmen: Stilometrie und **Textinterpretation**

Doch Fragestellungen, die über die reine Autorschaftsattribution hinausreichen, sind selten oder werden lediglich in einem Ausblick angerissen.33 Für die Interpretation literarischer Texte ist die Zuschreibung des Autors letztlich nur eine

³⁰ Ebd.

³¹ Vgl. ebd., S. 219.

³² Vgl. ebd., S. 220f.

³³ Einen ersten Feldversuch einer stilometrisch gestützten deutschen Literaturgeschichtsschreibung unternehmen Jannidis und Lauer. Ihre Aussagen über ein Gelingen sind vorsichtig optimistisch: »For the first time, a quantitative approach can work on large corpora and we expect that literary historians will appreciate these new research possibilities«. Jannidis, Lauer: »Burrows's Delta and Its Use in German Literary History«, S. 49. Martin Becker u. a. versuchen sich an einer engeren Gattungsklassifikation des deutschen Romans vom 16. bis zum 20. Jahrhundert, die über den Kanon hinausgehen und eine Perspektive für eine Gattungsgeschichtsschreibung etwa hinsichtlich narrativer Techniken und Themen bieten soll. Vgl. Martin Becker, Lena Hettinger, Andreas Hotho, Fotis Jannidis und Isabella Reger: »Genre Classification on German Novels«, in: Proceedings of the 12th International Workshop on Text-based Information Retrieval (2015), http://www.uni-weimar.de/medien/webis/events/tir-15/tir15-papers-final/Het tinger2015-tir-paper.pdf (31.10.2017, 15.41 Uhr). Einen kleinen Ausblick zur Relevanz stilometri-

Vorstufe, an die es mit hermeneutischen Fragestellungen anzuknüpfen gilt. Als Beispiel dafür lohnt ein systematischer Blick auf den Aufsatz »Collaborative Authorship: Conrad, Ford and Rolling Delta«.

Rolling Delta ist eine Variation herkömmlicher stilometrischer Clusteranalvsen, die besonders geeignet ist, um kollektive Autorschaft zu identifizieren. Die Ähnlichkeitssignale werden dementsprechend nicht hierarchisch, sondern in einem Verlaufsdiagramm abgebildet. Ähnlichkeiten und Differenzen ermittelt Rollin Delta in Abhängigkeit von einem Referenztext.³⁴ Die computergestützte Stilometrie besitzt eine Vielzahl frei wählbarer Parameter, die ihr einerseits eine große Freiheit einräumen, andererseits aber eine exakte Dokumentation und eine systematische Überprüfung der Resultate erfordern. Das gilt insbesondere für offene Forschungsfragen, die im Gegensatz zur Autorschaftsattribution nur wenige Hinweise liefern, ob die Berechnungen nachvollziehbare Ergebnisse erzielen. Evident ist dann vielfach das Ergebnis, das die eigene These stützt.35 Auf diese Weise wird die Ausgangsthese zum Gradmesser der Funktion stilometrischer Analysen. Man spricht hierbei vom sogenannten > Cherry-Picking (-Problem. 36

Die drei wichtigsten Parameter, die es auszuwählen und zu begründen gilt, sind das Textkorpus, das Distanzmaß und die Anzahl der zur Berechnung herangezogenen >most frequent words (mfw). Es gibt noch keinen Konsens darüber, welche Distanzmaße und Parameter die Funktion stilometrischer Analysen bestmöglich absichern. Theoretische und empirische Argumente divergieren oftmals.³⁷ Mehr noch wird in der Forschung häufig die Ansicht vertreten, dass jedes

scher Methoden für die Literaturwissenschaft liefert Schöch im Fazit seiner Studie Corneille, Molière et les autres: »Ebenso interessant scheint mir, die Ergebnisse computergestützter Klassifikationsverfahren (insbesondere, wenn der Faktor des individuellen Autors ausgeklammert werden könnte), mit etablierten Gattungs- und insbesondere Untergattungs-Klassifikationen zu vergleichen. Hier wird es dann auch für die Literaturgeschichtsschreibung interessant: Gerhard Lauer und Fotis Jannidis haben kürzlich bereits festgestellt, dass es beispielsweise um 1800 eine Gruppe weiblicher Autoren gibt, deren Texte von stilometrischen Verfahren nicht mit einer der etablierten Autorengruppen der Zeit assoziiert werden, sondern eine eigene Gruppe bilden, die bisher von der Literaturgeschichte gar nicht als solche wahrgenommen worden ist.« Schöch: »Corneille, Molière et les autres«, S. 152.

³⁴ Vgl. Hoover, Kestemont, Rybicki: »Collaborative Authorship«, S. 425f.

³⁵ Vgl. Maciej Eder: »Computational Sylistics and Biblical Translation: How Reliable Can a Dendrogram Be?«, in: The Translator and the Computer, hg. v. Tadeusz Piotrowski und Łukasz Grabowski. Breslau 2013, S. 155–170, hier 159, 161 und Juola: »The Rowling Case«, S. 102.

³⁶ Vgl. Juola: »The Rowling Case«, S. 110 und Eder, Maciej: »Visualization in Sylometry: Cluster Analysis Using Networks«, in: Digital Scholarship in the Humanities 32.1 (2017), S. 50–64, hier 52. 37 Vgl. Stefan Evert, Fotis Jannidis, Steffen Pielström, Thomas Proisl, Isabella Reger, Christoph

Korpus in Abhängigkeit der Forschungsfrage ganz eigene Einstellungen benötige.³⁸ Den Analysen fehlt also noch ein »generally accepted theoretical framework«.39

Hoover, Kestemont und Rybicki entscheiden sich für Burrows's Delta als Distanzmaß, das seit 2002 als ein wichtiger Standard der Stilometrie gilt. 40 Das Distanzmaß gewichtet die normalisierten Wortfrequenzen und berechnet ein sogenanntes Delta, das die Ähnlichkeit der Texte relativ zum gewählten Textkorpus anzeigt. 41 Die Anzahl der Wörter, die für die Berechnung herangezogen werden, beschränken Hoover, Kestemont und Rybicki auf die 1000 > most frequent words <. Das scheint für das Distanzmaß Burrows's Delta durchaus eine sinnvoll gewählte Größe zu sein, wie Stefan Evert u. a. erörtern: Burrows's Delta performe am besten bei Umfängen zwischen 1000 und 1500 >most frequent words<, 42 Eine Begründung über Vergleichsdaten anderer Wortumfänge wäre gleichwohl dennoch wünschenswert, da schon kleinste Abstufungen zu deutlich divergierenden Resultaten führen können. 43 So sind es vor allem die impliziten Standards der Autorschaftsattribution, auf die Hoover, Kestemont und Rybicki vertrauen.

Schöch und Thorsten Vitt: »Understanding and Explaining Delta Measures for Authorship Attribution«, in: Digital Scholarship in the Humanities (2017), S. 1–13, hier 11f.

³⁸ Vgl. Schöch: »Corneille, Molière et les autres«, S. 142.

³⁹ Kestemont: »Function Words in Authorship Attribution«, S. 59.

⁴⁰ Vgl. Jannidis, Lauer: »Burrows's Delta and Its Use in German Literary History«, S. 32.

⁴¹ Für eine sehr genaue Aufarbeitung von Burrows's vgl. Jannidis: »Der Autor ganz nah«, S. 183-189.

⁴² Vgl. Evert, u. a.: »Understanding and Explaining Delta Measures for Authorship Attribution«, S. 5.

⁴³ Dazu Eder: »Visualization in Sylometry«, S. 54: »Cluster analysis using [...] 136 MFWs [...] reveals a perfect authorial recognition, but when 137 MFWs are used, the cluster for Joseph Conrad is split into two parts and remains detached (along many other substantial rearrangements of the corpus) until the same corpus is assessed at 969 MFWs«.

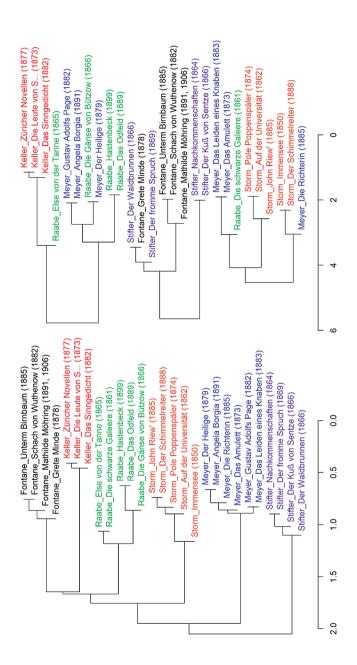


Abb. 1: Hierarchische Clusteranalyse realistischer Erzählungen. 1000 mfw, kein Culling, Ward-Clustering. Links: Burrows's Delta; rechts: Euclidean Distance.

Abbildung 1 zeigt zwei gegenübergestellte Clusteranalysen, die sich lediglich im Distanzmaß unterscheiden und die Problematik des »Cherry-Pickings« explizieren. Die stilometrische Analyse selbst gibt keine Hinweise auf die Plausibilität der hierarchischen Darstellung. Erst der Abgleich mit literaturgeschichtlich gesichertem Wissen erlaubt es, die linke der beiden Grafiken als nachvollziehbarere zu bewerten. Da die Autoren der realistischen Erzähltexte bekannt sind und das Ziel der stilometrischen Analyse die Wiedergabe von Autorsignalen ist, wird für dieses Textkorpus deutlich, dass Burrows's Delta (links) bessere Ergebnisse erzielt als das euklidische Distanzmaß (rechts).

Auch die Auswahl des Textkorpus ist zentral für die Plausibilität der Ergebnisse, wie Juola pointiert darlegt: »No matter how large the control sample, Marlowe would still probably have written Shakespeare's plays if the alternatives were Maureen Dowd, Paul Krugman, and Mark Bittman (all bloggers for the New York Times)«.44 Wichtig ist es deshalb ein Vergleichskorpus zu finden, das mit dem zu untersuchenden Text beziehungsweise dessen Urheber in Entstehungszeit, Gattung, Sprache, Region und Geschlecht größtmöglich übereinstimmt.

Entsprechend philologisch begründet leiten Hoover, Kestemont und Rybicki in ihre Untersuchung kollektiver Autorschaft ein: »The story of the notable literary collaboration between Joseph Conrad (1857-1924) and Ford Madox Ford (1873–1939) dates back at least to September 1898, the date of the first meeting between the two writers. «45 Sie ziehen sodann historische Kontexte heran, die die Schreib- und Lebenssituation der beiden Autoren Joseph Conrad und Ford Madox Ford näher situieren. 46 Das Textmaterial, das den Ausgangspunkt für die Überlegung einer stilometrischen Analyse darstellt, beschreiben Hoover, Kestemont und Rybicki wie folgt:

[T]he duo managed to produce three acknowledged collaborative novels: The Inheritors (1901), Romance (1903), and The Nature of a Crime (1909, 1924). And their story is made even more interesting by various authorship claims by Ford, including those concerning a fragment of Nostromo (1904) and the dramatization of Conrad's To-morrow (1901-2), One Day More (1904). Of course, physical evidence about authorship is complicated by the fact that Ford (and others, including John Galsworthy) took Conrad's dictation when he was sick or indisposed or could not make a deadline; hence pages have been preserved in Ford's and others' handwriting of works of otherwise unquestioned Conrad's authorship.⁴⁷

⁴⁴ Juola: »The Rowling Case«, S. 106.

⁴⁵ Hoover, Kestemont, Rybicki: »Collaborative Authorship«, S. 422.

⁴⁶ Vgl. ebd., S. 422f.

⁴⁷ Ebd., S. 423.

Den ausschlaggebenden Antrieb für die Analyse liefern Selbstaussagen von Ford und Conrad hinsichtlich ihrer Zusammenarbeit, die sich etwa auf Schreibanteile an verschiedenen Texten oder ihre jeweiligen Rollen als Autor, Ideengeber oder Lektor beziehen. Fords genaue Beteiligung an Nostromo etwa wird von Xavier Brice als widersprüchlich geschildert. 48. Zwei Textausschnitte, der erste aus einem Brief Fords an George T. Keating, 49 der zweite aus Return to Yesterday: Reminiscences of James, Conrad & Crane, 50 dienen als Belegstellen. Der Widerspruch entsteht allerdings auch durch eine verkürzte Darstellung, die Fords frühere Aussage aus dem Zusammenhang herauslöst, wodurch eine weitaus größere Kluft zwischen den zwei Passagen entsteht.⁵¹ In seinem Brief an Keating formuliert Ford 1923 (1925):52

The circumstances in which I wrote that small portion of NOSTROMO [...] are very simple and explicable. [...] But to argue from that that I had any large share in Conrad's writing would be absurd. I may have written ten thousand words that I remember and could place my finger on fairly substantial passages and perhaps another twenty thousand that I remember only faintly and should find difficult to trace.⁵³

Diesen schon gekürzten Textausschnitt verknappen Hoover, Kestemont und Rybicki weiter. Sie beschränken sich dabei auf den zweiten mit »I may [...]« eingeleiteten Abschnitt, in dem Ford die eigene quantitative Schreibleistung herausstellt. Ohne die Einschränkung, dass seine Beiträge keinen großen Anteil an Conrads Text hätten, bleibt mit insgesamt 30000 Wörtern, die Ford beigesteuert haben will, - darunter wesentliche Textpassagen - der Eindruck zurück, dass Nostromo zu den kollektiv verfassten Texten zu zählen sei und Ford entsprechend einen erheblichen Teil dazu beigetragen habe. Was Brice als »air of arrogant flippancy«⁵⁴ bezeichnet, ist gleichwohl im gesamten Zitat Fords zu vernehmen.

⁴⁸ Xavier Brice: »Ford Madox Ford and the Composition of Nostromo«, in: The Conradian 29.2 (2004), S. 75-95.

⁴⁹ Brief zitiert nach John Hope Morey: Joseph Conrad and Ford Madox Ford: A Study in Collaboration. Ithaca, NY 1960, S. 120f.

⁵⁰ Ford, Ford Madox: Return to Yesterday: Reminiscences of James, Conrad & Crane (1931). New York 1972, S. 189f.

⁵¹ Vgl. das nachfolgende Zitat mit Hoover, Kestemont, Rybicki: »Collaborative Authorship«, S. 424.

⁵² Die Jahreszahl ist nicht gesichert. Es könnte auch 1925 gewesen sein. Vgl. Brice: »Ford Madox Ford and the Composition of *Nostromo*«, S. 79.

⁵³ Brief zitiert nach Morey: *Joseph Conrad and Ford Madox Ford*, S. 120f.

⁵⁴ Brice: »Ford Madox Ford and the Composition of *Nostromo*«, S. 80.

Indem Ford sein Zutun relativiert, erscheint jedoch zumindest die Arroganz gedämpft. Den Widerspruch zur zweiten, späteren Aussage Fords, den Hoover, Kestemont und Rybicki erkennen wollen, konstruieren sie also letztlich selbst. Acht (oder sechs) Jahre später äußert sich Ford nämlich folgendermaßen:

What I actually wrote into Conrad's books was by no means great in bulk and was usually done when he was too ill to write himself and had to catch up with serial publication. [...] But, indeed, the importance of the passages I did write was so negligible, and they themselves were so frequently emended out of sight that they could not make as much difference to the completion and glory of his prose as three drops of water poured into a butt of Malmsev.55

Der Vergleich dreier Tropfen Wasser in einem großen Fass Malvasier mit der Bedeutung der Passagen, die Ford zu Conrads Buch beigesteuert habe, zeigt eindrücklich, was Brice mit »Ford is careful to play down his collaborative role«56 meint. Eine gewisse kritische Distanz sollte bei solchen Selbstaussagen⁵⁷ – zu welchem Zweck schreibt Ford seine Memoiren? – zwar immer mitschwingen, der postulierte Widerspruch der beiden Textpassagen aber, und das ist ausschlaggebend für die daran anschließende stilometrische Studie von Hoover, Kestemont und Rybicki, scheint nur eingeschränkt gültig zu sein. Denn auf diese Weise schreiben sie implizit ein zusätzliches Spannungsmoment in die Fragestellung ein: Ist Nostromo als kollektiver Text zu werten? Und welche Äußerung Fords ist letztlich glaubwürdiger? Die eigentlich exoterisch formulierte Fragestellung ist dagegen weit weniger festgelegt: »This study tries to find the drops of Ford's water in Conrad's Malmsey, or vice versa«.58

Um mit der Hilfe stilometrischer Analysen tatsächlich forschungsrelevante Ergebnisse zu erzielen, sind also zwei Punkte entscheidend. Einerseits gilt es, die Auswahl der Parameter zu dokumentieren und zu begründen – auch, um die Reproduzierbarkeit der Ergebnisse zu gewährleisten. Andererseits ist die exakte philologische Vorarbeit zentral, um ertragreiche Fragestellungen zu generieren und diese auf ein gut gewähltes Textkorpus anzuwenden. Was sind aber die Ergebnisse, die die computergestützten stilometrischen Analysen bei Hoover, Kestemont und Rybicki abbilden? Die drei kollektiv geschriebenen Texte von

⁵⁵ Ford: Return to Yesterday, S. 189f.

⁵⁶ Brice: »Ford Madox Ford and the Composition of *Nostromo*«, S. 80.

⁵⁷ Das zeigt Brice etwa an der von Ford aufgerufenen Wortzahl, die selbst im für Ford bestmöglichen Auslegungsfall vorhandener und möglicherweise fehlender Manuskripte nicht über 5000 Wörtern liegen dürfte. Vgl. ebd. Brice kommt sogar zu dem Schluss, dass Conrad diese Ford diktiert habe. Vgl. ebd., S. 92.

⁵⁸ Hoover, Kestemont, Rybicki: »Collaborative Authorship«, S. 424.

Conrad und Ford gruppieren sich nicht einheitlich. Während sich *Romance* zu Texten von Conrad ordnet, finden sich *The Inheritors* und *Nature of a Crime* im Umfeld von Fords Texten. Die nähere Analyse lässt plausibel beobachten, dass *The Inheritors* zum größten Teil, wie vorab angenommen, ⁵⁹ von Ford geschrieben ist.

Ford's style [...] dominates over that of Conrad's – with some exceptions. There are several short glimpses of Conradian style in the middle of Chapter 4, reminiscent of *Chance* [...]; a section at the onset of Chapter 11 [...]. The only more extensive fragment in *The Inheritors* with a stronger Conradian signal combines Chapters 16 and 17 [...]. ⁶⁰

Mit ihren Rolling Delta-Diagrammen können Hoover, Kestemont und Rybicki relative stilistische Ähnlichkeiten der beiden Autoren im Verhältnis zu *The Inheritors* abbilden. Da die horizontale Achse die lineare Abfolge des Textes in 5000 Wortschritten wiedergibt, können stilistische Übereinstimmungen verschiedenen Kapiteln zugeordnet werden.

Die detaillierten Analysen zu *Nature of a Crime* und *Romance* erlauben es, einzelne Textpassagen aufgrund quantitativ erfasster Signale je einem der beiden Autoren zuzuordnen. So überwiegt etwa in *Romance* das Autorsignal Joseph Conrads. ⁶¹ Bei *Nostromo* ergeben sich jedoch keine stilometrisch gestützten Evidenzen, die nahe legen würden, dass Ford signifikant zum Text beigetragen habe. Das ist letztlich keine Überraschung, da die Hypothese zu großen Teilen auf einer fehlerhaften Auslegung von Fords Selbstaussagen gründet. Hoover, Kestemont und Rybicki konstatieren dazu abschließend:

Fig. 5 tests the hypothesis that Ford did indeed contribute in any significant way to Conrad's Nostromo but provides little to support it: Conrad's style dominates Ford's throughout. This time, contrarily to the diagrams for The Inheritors and The Nature of a Crime, and in a much stronger way that in Romance, the signal of almost all Conradian texts is stronger than that of almost all texts by Ford throughout the course of the novel [...]. ⁶²

Mit den Resultaten liefert die Studie nach eigener Aussage »a confirmation of the usual (if uncertain) consensus about the proportions of the styles of both writers in their three collaborations«.⁶³

⁵⁹ Vgl. ebd., S. 423. Ford habe den größten Teil selbst geschrieben, jedoch ausführlich mit Conrad besprochen.

⁶⁰ Ebd., S. 427f.

⁶¹ Vgl. Hoover, Kestemont, Rybicki: »Collaborative Authorship«, S. 428.

⁶² Ebd., S. 429.

⁶³ Ebd.

Hoover, Kestemont und Rybicki ergänzen die Debatte um Ford und Conrad somit – ganz im Sinne von Matthew Jockers – um eine neue, quantitative Blickrichtung. Der intrinsische Wert ihrer Erkenntnisse ergibt sich aus den zusätzlichen Evidenzen, die philologischen Argumenten beigefügt werden. Eine hermeneutische Fragestellung, die dezidiert an die stilometrische Analyse anschließt, die sie im besten Fall sogar anleitet, liegt jedoch nicht vor. Aus Sicht der Literaturwissenschaft müssten die Ergebnisse einer stilometrischen Analyse gleichwohl nicht nur dargelegt, sondern ebenfalls interpretiert werden. Der Aufsatz »Collaborative Authorship: Conrad, Ford and Rolling Delta« ist dafür aus mehreren Gründen ein hervorragendes Beispiel. Erstens existiert die aufgegriffene Forschungsfrage nach dem Verhältnis kollektiver Autorschaft in den Texten von Ford und Conrad in der Anglistik schon lange. Brices Aufsatz »Ford Madox Ford and the Composition of Nostromo« steht dafür exemplarisch ein. ⁶⁴ Hoover, Kestemont und Rybicki greifen entsprechend eine philologische Fragestellung auf und versuchen der Debatte mit Hilfe quantitativer Methoden neue Impulse zu verleihen. Fishs generalisierter Vorwurf, dass Interpretationsthesen im Feld computergestützter Textanalyseverfahren nur zufällig durch die Spielerei mit Datenmengen zu finden sind, läuft damit ins Leere. 65 Stilometrische Analysen, das zeigt der behandelte Aufsatz mit seinen Ergebnissen, lassen sich somit erfolgreich an entsprechend gebaute philologische Probleme andocken. Von entscheidendem Interesse ist indes aber der umgekehrte Fall: Wie können hermeneutische Fragestellungen wiederum an die stilometrische Analyse anknüpfen? Erneut verhilft das betrachtete Beispiel zu einer Annäherung: Die Belege für Conrads editorische Eingriffe in *The Inheritors* und *The Nature of Crime* und die Entkräftung der Selbstaussage Fords, er habe 30000 Wörter zu Nostromo beigesteuert, sind Bestätigung der getesteten Thesen. Sie liefern dadurch aber zugleich interpretatorische Leerstellen. Zum Verständnis hilft ein Blick auf die Perspektive, aus der die jeweilige Untersuchung durchgeführt wurde. Hoover, Kestemont und Rybicki greifen die Debatte um die Zusammenarbeit von Conrad und Ford auf. In diesen werden Texte und Kontexte – das können biographische Notizen wie in Fords Memoiren, aber auch Briefe, Vergleichstexte und im Fall von Nostromo die 16 handschriftlich

⁶⁴ Siehe auch Raymond Brebach: »Conrad, Ford and the ›Romance‹ Poem« in: Modern Philology 81.2 (1983), S. 169–172.

⁶⁵ Selbst die durch >Spielerei < erreichte Wahl des Forschungsobjektes wäre indes vertretbar. Vgl. dazu Max Weber: Wissenschaft als Beruf, in: Max Werber. Schriften 1894-1922, hg. v. Dirk Kaesler, Stuttgart 2002, S. 474-511, hier 483. Gemäß Weber verdankt die Wissenschaft viele ihrer besten Problemstellung Dilettanten, denen eine feste Arbeitsmethode fehle und die die Tragweite ihres Einfalls nicht nachvollziehen könnten.

überlieferten Textblätter, die Ford geschrieben hat, sein⁶⁶ – einer Mikroanalyse, dem >close reading<, unterzogen. Das mit quantitativen Methoden erarbeitete Ergebnis zeichnet wiederum ein >big picture<. Der Vergleich zieht das Erzähltextkorpus von Ford und Conrad zu Rate und stützt sich auf eine stilometrische Makroanalyse. 67 Die Leerstellen, die die Resultate der Makroanalyse durch ihre Belege erzeugen, sind aber mikroanalytischer Natur. Das heißt, dass es gerade die >close reading<-Methoden traditionell gearteter Literaturwissenschaft sein müssten, die das volle Potential der Makroanalyse überhaupt erst entfalten. Die Eigenart der postulierten Leerstellen ist der Grund für diesen Schluss.

Was sind nun diese interpretationsbedürftigen Leerstellen? Durch die nachvollziehbare Zuordnung der Texte zu einem der beiden Autoren, ja sogar die Zuordnung einzelner Teilkapitel zum je dominanten Schreibstil, ergeben sich für die hermeneutische Textarbeit neue Grundlagen, die weitere Fragen anregen. Da die Aufarbeitung der Ford- beziehungsweise Conrad-Forschung nicht Ziel dieses Aufsatzes sein kann, sind die anschließenden Folgerungen als hypothetisches Gedankenexperiment zu werten.

Gibt es erstens historische Indizien oder Belege – diese können biographischer Natur sein, aber etwa auch verschiedene Stufen der Textgenese umfassen -, die erst durch das Vorwissen einer erfolgreichen Attribution einen Mehrwert generieren und so der Textinterpretation beigestellt werden können? Lassen sich zweitens Fords handschriftlich überlieferte Textabschnitte in Nostromo den stilometrischen Signalen zuordnen? Können also stilistische Ausschläge in den Diagrammen plausibel philologisch kontextualisiert und interpretiert werden? Sind durch die Eingrenzung auf gewisse Textabschnitte drittens auch Brüche im ›close

⁶⁶ Brice rechnet in seinem Aufsatz vor, wie die 16 überlieferten Blätter zustande kommen. Die Seitenzahlen machen deutlich, dass insgesamt von 23 Blättern in Fords Handschrift auszugehen ist. Vgl. Brice: »Ford Madox Ford and the Composition of Nostromo«, S. 75–79.

⁶⁷ An dieser Stelle hätte, anknüpfend an den Begriff >close reading <, ebenso von >distant reading« die Rede sein können. Da ich die von Franco Moretti unter diesem Begriff mitgeprägte Polemik nicht mittragen möchte, verzichte ich hier auf die Nennung. Vgl. Franco Moretti: »Conjectures on World Literature. My Mission: To Say it More Simply than I Understand it. Schönberg, Moses and Aaron«, in: New Left Review 1 (2000), S. 54-68, hier 57. Was Moretti als >distant reading (bezeichnet, nennt Jockers) macroanalysis (. Den Begriff entlehnt er der Ökonomie: »By way of an analogy, we might think about interpretive close-readings as corresponding to microeconomics while quantitative macroanalysis corresponds to macroeconomics. [...] I prefer the term macroanalysis over distant-reading. The former term places the emphasis on the quantifiable methodology over the more interpretive practice of reading « Matthew L. Jockers: On Distant Reading and Macroanalysis, 01.07.2011, http://www.matthewjockers.net/2011/07/01/on-dis tant-reading-and-macroanalysis/ (06.04.2017). Vgl. auch Jockers: Macroanalysis, S. 24–26.

reading wahrnehmbar, die einerseits noch nicht wahrgenommen wurden, andererseits einen Autorwechsel wahrscheinlich machen? Und was ändert sich abschließend für die Interpretation der Texte, wenn wir die einzelnen Textteile den Autoren zuordnen können? Ließe sich eine für Ford und Conrad differenzierte Autorintention rekonstruieren und beurteilen? Wäre das überhaupt sinnvoll?

Teilbereiche dieser aufgeführten Fragestellungen ließen sich wiederum durch digitale Methoden der Textanalyse ergänzen. Ein realisierbarer Ansatz könnte sein, Jockers' Idee einer quantitativen Bestimmung für Stimmungs- und Handlungsbögen des englischsprachigen Romans anzuwenden.⁶⁸ Aus seiner Stimmungs- und Handlungsanalyse von 41383 Romanen – Handlung fasst er für diese Untersuchung mit den Worten Genettes heuristisch als ›discours‹ - leitet Jockers sechs beziehungsweise sieben Handlungsarchetypen ab, die er als Stimmungskurven repräsentiert. Dafür nutzt er eine semantische Wortfeldanalyse, die letztlich durch eine graduelle Einordnung einzelner Begriffe als >positiv‹ oder >negativ« eine Kurve erzeugt. Potentiell ließe sich durch die so geartete quantitative Erfassung der Romane von Ford und Conrad eine Differenzbildung hinsichtlich der Stimmung einzelner Textabschnitte vornehmen, die dann vor dem Hintergrund der stilometrischen Analyse zu bewerten und einzuordnen sind.

4 Forschungsperspektiven

Das exemplarisch aufgeführte und skizzierte Beispiel der stilometrischen Analyse von Hoover, Kestemont und Rybicki sollte dazu dienen, das hermeneutische Anschlusspotential stilometrischer Analysen über die Autorschaftsattribution hinaus darzulegen. Ich habe das als Andocken hermeneutischer Fragestellung an die im Vorfeld durchgeführte Attribution gefasst und als eine Option für die weiterführende und auf quantitativen Analysen basierende Textinterpretation benannt. Maßgeblich dafür ist die Operationalisierung literarischer Konzepte. Hoover, Kestemont und Rybicki schlagen diese Richtung in Teilen ein: Sie operationalisieren das Konzept kollektiver Autorschaft. Die automatisch aus der Prob-

⁶⁸ Das von Jockers zur Verfügung gestellte Syuzhet Package für R ist ein interessantes Tool zur >sentiment<-Bestimmung für quantitativ arbeitende Literaturwissenschaftler. Vgl. Jockers, Matthew L.: Revealing Sentiment and Plot Arcs with the Syuzhet Package, 02.02.2015, http://www.matthewjockers.net/2015/02/02/syuzhet/ (06.02.2017) und Matthew L. Jockers: The Rest of the Story, 25.02.2015, http://www.matthewjockers.net/2015/02/25/the-rest-of-the-story/ (30.10.2017).

lemannäherung entstehenden Implikationen betrachten sie aber nicht. Sie ziehen weder konkrete Rückschlüsse noch benennen sie Forschungsperspektiven. Die Annäherung ist also rein technisch oder, mit literaturwissenschaftlichem Vokabular beschrieben: rein methodisch.69

Eine weitere Möglichkeit, die darauf abzielt, stilometrische Analysen für die Interpretation literarischer gewinnbringend einzusetzen, ergibt sich durch eine methodische Umkehr in der Anwendung stilometrischer Analysen. Ist davon auszugehen, dass computergestützte stilometrische Analysen eine ausreichend hohe Trefferquote bei der Attribution von Text zu Autor oder von Text zu Epoche haben – und damit implizit von Autor zu Epoche –, kann dieses Wissen eingesetzt werden, um Ziel und Ausgangspunkt der Analyse zu tauschen. Ausgangspunkt der bisher betrachteten Studien, die sich stilometrischer Analysen zur Autorschaftszuschreibung bedienen, ist ein Text, dessen Autor mit herkömmlichen philologischen Methoden nicht eindeutig zu bestimmen ist. Ziel war es, den Text einem Autor aus einer Reihe zuvor erörterter Kandidaten der Wahrscheinlichkeit nach zuzuschreiben. Als Methode wurden dafür stilometrische Analysen genutzt.

Was passiert aber, wenn die Ausgangssituation gesichert ist, die Analyse dementsprechend bereits vermeintlich gesichertes Wissen untersucht? Einerseits lässt sich auf diese Weise die Funktion stilometrischer Analysen, etwa die Frage nach der optimalen Wahl verschiedener Parameter, überprüfen. Andererseits sind es gerade die potentiellen Dissonanzen einer solchen Analyse, die interessante Forschungsperspektiven aufwerfen könnten. Wie ist etwa die Epochenzuordnung eines Autors auf Grundlage relativer stilometrischer Abgrenzung zu bewerten? Relevanz hat diese Frage gerade dann, wenn die Analyse bis dato gängige Ex-post-Zuschreibungen nicht bestätigt. Gleiches gilt für aktuell ungeklärte oder kontroverse Zuschreibungen – an dieser Stelle wohlgemerkt keine Zuschreibung von Text zu Autor -, die durch stilometrische Clusteranalysen mit andersgearteten Argumenten aus neuer Blickrichtung betrachtet werden können. Ein prominentes Beispiel sind die Texte Heinrich von Kleists, der mit ihrer bestürzenden Andersartigkeit und Modernität die Grenzen seiner Zeit weit überschritten habe:70

⁶⁹ Vgl. Thaller: »Digitial Humanities als Wissenschaft«, S. 13f. Werkzeug und Methode fallen hier jedoch zusammen. Generell beschreibt Thaller quantitative Verfahren als methodenneutral, die Forschungsgegenstände auch unter unveränderter Methodologie intensiver und gründlicher untersuchen könnten. Vgl. ebd., S. 15, 18.

⁷⁰ Vgl. Marie Haller-Nevermann und Dieter Rehwinkel: »Einleitung. Kleist – ein moderner Aufklärer?«, in: Kleist – ein moderner Aufklärer?, hg. v. dens. Göttingen 2005, S. 9–12, hier 9f.

Wie kaum ein anderer vermochte er, die Phänomene des Lebens in ihrer Ambivalenz darzustellen, er der Aufklärungsskeptiker, der Dichter des Unbewussten, der doch in so hohem Maße ein bewusster, ein aufgeklärter Schriftsteller war, von einem nicht mehr zu überbietenden Formwillen getragen.⁷¹

Die von Marie Haller-Nevermann und Dieter Rehwinkel gezeichnete Ambivalenz findet sich ebenso in der Epochenzuschreibung Heinrich von Kleists wieder. Kleist gilt als Grenzgänger zwischen den Epochen. 72 Folgerichtig versucht etwa Anke Bennholdt-Thomsen »Kleists Standort zwischen Aufklärung und Romantik« auszumachen. Als Fragestellung formuliert sie ganz generell die »richtige Einordnung des Dichters in die Literaturgeschichte und ihre Epochen«.⁷³ Die Fragestellung lässt sich sogar auf einzelne Textgattungen und Untergattungen zuspitzen. Orientiert sich Kleist mit seinen Dramen am Sturm und Drang, orientiert er sich direkt an Shakespeare oder aber am bürgerlichen Trauerspiel? Seine Konzeption der Figurenrede scheint jedoch nicht auf eindeutige literarische Vorbilder rückführbar zu sein.74

Anders als bei den vorgestellten Studien zu Hildegard von Bingen und Guibert de Gembloux sowie Ford Madox Ford und Joseph Conrad ist das Interesse der Fragestellung in Nils Reiters und Marcus Willands Aufsatz »Geschlecht und Gattung. Digitale Analysen von Kleists >Familie Schroffenstein« durch eine offene Problemstellung geleitet. Ziel ist es nicht empirisch darzulegen, dass Der zerbrochene Krug eine Komödie oder Penthesilea eine Tragödie ist. Die im Aufsatz

⁷¹ Ebd., S. 9.

⁷² Jens Ewen schreibt dazu: »Die literaturgeschichtliche Verortung Heinrich von Kleists ist seit jeher heftig umstritten. Sein Werk, so vielgestaltig es ist, scheint sich den gängigen literaturgeschichtlichen Kategorien zu entziehen. Kleist ist weder ein richtiger Aufklärer noch ein vollgültiger Romantiker, und er lässt sich auch keiner anderen Strömung sinnvoll zuordnen.« Jens Ewen: »Periodisierungsprobleme zwischen Aufklärung und Romantik – am Beispiel Heinrich von Kleists«, in: Aufklärung und Romantik. Epochenschnittstellen, hg. v. Daniel Fulda, Sandra Kerschbaumer und Stefan Matuschek. Paderborn 2015, S. 87-100, hier 89. Dazu auch Helga Gallas: »Kleists Welt im Vergleich zur Dominanz des Imaginären in der Romantik und zur Inflation des Realen in der Postmoderne«, in: Heinrich von Kleist, hg. v. Ortrud Gutjahr. Würzburg 2008, S. 279-295, hier 279: »Mich hat immer beschäftigt, dass Kleists Werk in der Literaturgeschichte eine so offensichtliche Ausnahmestellung innehat: es passt weder zur Klassik noch zur Romantik, einerseits gilt es als äußerst modern und auf das 20. Jahrhundert Kafkas weisend, andererseits als konservativ und anti-aufklärerisch, dem 18. Jahrhundert verhaftet.«

⁷³ Anke Bennholdt-Thomsen: »Kleists Standort zwischen Aufklärung und Romantik. Ein Beitrag zur Quellenforschung«, in: Kleist – ein moderner Aufklärer?, hg. v. Marie Haller-Nevermann und Dieter Rehwinkel. Göttingen 2005, S. 13-40, hier 13.

⁷⁴ Vgl. Nils Reiter und Marcus Willand: »Geschlecht und Gattung. Digitale Analysen von Kleists >Familie Schroffenstein < .. In: Kleist-Jahrbuch (2017), S. 177–195, hier 190.

erläuterte Gattungsklassifikation wird viel eher als makroanalytischer Befund gewertet, die eine Aussage über das Gattungsbewusstsein Kleists erlaubt. 75 Dementsprechend ist auch die Intention der Analyse formuliert:

Vielmehr ermöglicht die Digitalisierung sehr großer Dramen-Korpora eine komparative Neuperspektivierung einzelner dramatischer Texte und erlaubt es, auf einem alternativen Weg bisher kaum feststellbare dramenhistorische Einflussbeziehungen offenzulegen oder aber Einflusshypothesen zu konterkarieren.⁷⁶

Einflussbeziehungen und Auswahlkriterien sind es auch, die Thomas Weitin für den Deutschen Novellenschatz nachzeichnet. Er orientiert sich dabei an der Maßgabe von Paul Heyse und Hermann Kurz, die die Textauswahl des von ihnen herausgegebenen Deutschen Novellenschatzes »nach ›Ton‹, also nach dem Stil der Texte« begründen.⁷⁷ Die Zusammenstellung des Novellenschatzes »nach den Prinzipien von Distinktivität und Relationalität« versucht Weitin mit stilometrischen Ansätzen abzubilden, 78 um an schon entwickelte literaturhistorische und -theoretische Probleme anzuschließen. 79 Mit Blick auf die literaturgeschichtliche Problemstellung der insgesamt 86 Novellen, die zwischen »massenhafter Ähnlichkeit« einerseits sowie »der stilorientierte[n] Zusammenstellung nach den Prinzipien von Distinktivität und Relationalität« andererseits verlaufen, hält Weitin das stilometrische Delta-Verfahren für historisch plausibel einsetzbar.⁸⁰

Die Heyse zuteil gewordene Charakterisierung als >Virtuose des Durchschnitts« spiegelt sich im stilometrischen Ähnlichkeitsverhältnis seiner Novelle zum übrigen Textkorpus des Novellenschatzes wider. Sowohl Heyse als auch Heinrich Zschokke und Karl Immermann seien sehr »korpusähnlich und mithin durchschnittlich auf das Ganze gesehen«. 81 Literarische Texte werden in diesem Fall nicht mehr zu Autoren oder verschiedenen Gattungen geordnet, sondern nach einem nicht weiter spezifizierten Ähnlichkeitssignal, das normalerweise von Autor- oder Gattungssignalen überlagert werden würde. Dieses Vorgehen ge-

⁷⁵ Vgl. ebd., S. 194.

⁷⁶ Ebd., S. 178

⁷⁷ Thomas Weitin: »Selektion und Distinktion. Paul Heyses und Hermann Kurz' Deutscher Novellenschatz als Archiv, Literaturgeschichte und Korpus«, in: Archiv/Fiktionen. Verfahren des Archivierens in Literatur und Kultur des langen 19. Jahrhunderts, hg v. Daniela Gertz und Nicolas Pethes. Freiburg, Berlin, Wien 2016, S. 385-408, hier 394.

⁷⁸ Ebd., S. 396.

⁷⁹ Vgl. ebd., S. 395.

⁸⁰ Ebd., S. 396. Vgl. auch S. 402.

⁸¹ Ebd. S. 403.

lingt, da nur sehr vereinzelt mehrere Novellen des gleichen Autors in den Deutschen Novellenschatz aufgenommen wurden. Hierin wird eine Limitierung stilometrischer Analysemethoden implizit nachvollziehbar. Die Resultate basieren stets auf einem Vergleichskorpus. Literaturtheoretische Fragen, die nicht explizit das Verhältnis literarischer Texte zu ihren jeweiligen Autoren, Gattungen oder Epochen zu ergründen suchen, benötigen eine dieser Limitierung geschuldete Modellierung. Letztlich versuchen auch die Experimente Weitins Gemeinsamkeiten innerhalb eines Textkorpus ausfindig zu machen. Es scheinen Gemeinsamkeiten zu sein, die etwa einen Teil der 86 Novellen als sehr korpusähnlich und somit »besonders durchschnittlich aufs Ganze gesehen« auszeichnen. 82 Dagegen sind andere, so argumentiert Weitin, dem Gros des Korpus weit weniger ähnlich, könnten aber dennoch gruppenbildend wirken.

5 Zusammenfassung

Die Ausführungen versuchen darzulegen, dass die Stilometrie auch über die Autorschaftsattribution hinaus Potentiale für einen genuin literaturwissenschaftlichen Einsatz besitzt. Die Operationalisierung literarischer Konzepte mit Hilfe der Stilometrie unterliegt jedoch einigen Einschränkungen. Denn stilometrische Analysen bilden immer relative Ähnlichkeiten oder Unterschiede ab, die in Abhängigkeit vom gewählten Textkorpus zu interpretieren sind. Die Fragestellung muss diesem relativen Umgang mit dem Untersuchungsgegenstand entsprechend gewählt sein. Neben einer hypothesengeleiteten Verwendung der Stilometrie – sei es, um bestätigende Evidenzen abzubilden oder um Reibungspunkte, die zu relevanten Fragestellungen führen, überhaupt erst zu entdecken – ist somit ein methodisch reflektiertes Vorgehen unerlässlich. Schließlich ist die Zahl der zur Berechnung herangezogenen Parameter, etwa Wortumfang oder Distanzmaß, groß und sie beeinflussen die Ergebnisse der Analyse immens. Da offene Forschungsfragen - anders als die Autorschaftsattribution, die eine Tatsachenanalyse ist – selbst nur wenige Anhaltspunkte geben, ob die Berechnungen nachvollziehbare Ergebnisse erzielen, wird vielfach das Ergebnis als evident bewertet, das die eigene These belegt und die Ausgangsthese wird zum Gradmesser der Funktion stilometrischer Analysen.⁸³ Soll die Stilometrie also nachvollziehbare Evidenzen schaffen, dürfen Studien nicht nur auf implizite Standards der Autorschaftsattribution vertrauen. Die Untersuchungen sollten stattdessen von einer Methodendiskussion begleitet werden, die die Ergebnisse kontextualisiert und ihre Reproduzierbarkeit ermöglicht.

Bibliographie

- Argamon, Shlomo, Moshe Koppel und Jonathan Schler: »Computational Methods in Authorship Attribution«, in: Journal of the American Society for Information Science and Technology 60.1 (2009), S. 9-26.
- Bennholdt-Thomsen, Anke: »Kleists Standort zwischen Aufklärung und Romantik. Ein Beitrag zur Quellenforschung«, in: Kleist - ein moderner Aufklärer?, hg. v. Marie Haller-Nevermann und Dieter Rehwinkel. Göttingen 2005, S. 13-40.
- Brebach, Raymond: »Conrad, Ford and the >Romance« Poem« in: Modern Philology 81.2 (1983), S. 169-172.
- Brice, Xavier: »Ford Madox Ford and the Composition of Nostromo«, in: The Conradian 29.2 (2004), S. 75-95.
- Burrows, John F.: »>Delta<: a Measure of Stylistic Difference and a Guide to Likely Authorship«, in: Literary and Linguistic Computing 17.3 (2002), S. 267-287.
- Daelemans, Walter, Folgert Karsdorp, Mike Kestemont, Moshe Koppel und Justin Stover: »Authenticating the Writings of Julius Caesar«, in: Expert Systems With Applications 63 (2016), S. 86-96.
- Danneberg, Lutz: »Philosophische und methodische Hermeneutik«, in: Philosophia Naturalis 32.2 (1995), S. 249-269.
- Deploige, Jeroen, Mike Kestemont und Sara Moens: »Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux«, in: Digital Scholarship in the Humanities 30.2 (2015), S. 199–224, hier S. 206.
- Eder, Maciej, Mike Kestemont und Jan Rybicki: »Stylometry with R: A Package for Computational Text Analysis«, in: The R Journal 8.1 (2016), S. 107–121.
- Eder, Maciej: »Computational Sylistics and Biblical Translation: How Reliable Can a Dendrogram Be?«, in: The Translator and the Computer, hg. v. Tadeusz Piotrowski und Łukasz Grabowski. Breslau 2013, S. 155-170.
- Eder, Maciej: »Visualization in Sylometry: Cluster Analysis Using Networks«, in: Digital Scholarship in the Humanities 32.1 (2017), S. 50-64.
- Evert, Stefan, Fotis Jannidis, Steffen Pielström, Thomas Proisl, Christof Schöch, Thorsten Vitt: Explaining Delta, or: How Do Distance Measures for Authorship Attribution Work? Abstract, http://www.stefan-evert.de/PUB/EvertProislEtc2015_abstract.pdf (30.01.2017)

⁸³ Vgl. Maciej Eder: »Computational Sylistics and Biblical Translation: How Reliable Can a Dendrogram Be?«, in: The Translator and the Computer, hg. v. Tadeusz Piotrowski und Łukasz Grabowski. Breslau 2013, S. 155–170, hier 159, 161 und Juola: »The Rowling Case«, S. 102.

- Evert, Stefan, Fotis Jannidis, Steffen Pielström, Thomas Proisl, Isabella Reger, Christoph Schöch und Thorsten Vitt: »Understanding and Explaining Delta Measures for Authorship Attribution«, in: Diaital Scholarship in the Humanities (2017), S. 1–13.
- Ewen, Jens: »Periodisierungsprobleme zwischen Aufklärung und Romantik am Beispiel Heinrich von Kleists«, in: Aufklärung und Romantik. Epochenschnittstellen, hg. v. Daniel Fulda, Sandra Kerschbaumer und Stefan Matuschek. Paderborn 2015, S. 87-100.
- Fish, Stanley: »Mind Your P's and B's: The Digital Humanities and Interpretation«, in: The New York Times (23.01.2012), dann: http://opinionator.blogs.nytimes.com/2012/01/23/mindyour-ps-and-bs-the-digital-humanities-and-interpretation (28. 10.2017).
- Ford, Ford Madox: Return to Yesterday: Reminiscences of James, Conrad & Crane (1931). New York 1972.
- Gallas, Helga: »Kleists Welt im Vergleich zur Dominanz des Imaginären in der Romantik und zur Inflation des Realen in der Postmoderne«, in: Heinrich von Kleist, hg. v. Ortrud Gutjahr. Würzburg 2008, S. 279-295.
- Grzybek, Peter: »History and Methodology of Word Length Studies«, in: Contributions to the Science of Text and Language, hg. v. Peter Grzybek. Dordrecht 2007, S. 15-90.
- Haller-Nevermann, Marie und Dieter Rehwinkel: »Einleitung. Kleist ein moderner Aufklärer?« In: Kleist – ein moderner Aufklärer?, hg. v. dens. Göttingen 2005, S. 9–12.
- Hoover, David, Mike Kestemont und Jan Rybicki: »Collaborative Authorship: Conrad, Ford and Rolling Delta«, in: Literary and Linguistic Computing 29.3 (2014), S. 422-431.
- Hoover, David: »Testing Burrows's Delta«, in: Literary and Linquistic Computing 19.4 (2004), S. 453-475.
- Jannidis, Fotis und Gerhard Lauer: »Burrows's Delta and Its Use in German Literary History«, in: Distant Readings. Topologies of German Culture in the Long Nineteenth Century, hg. v. Matt Erlin und Lynne Tatlock. Rochester 2014, S. 29-54.
- Jannidis, Fotis: »Der Autor ganz nah. Autorstil in Stilistik und Stilometrie«, in: Theorien und Praktiken der Autorschaft, hg. v. Matthias Schaffrick und Marcus Willand. Berlin, Boston 2014, S. 169-195.
- Jockers, Matthew L.: Marcroanalysis. Digital Methods & Literary History. Urbana, Chicago und Springfield 2014.
- Jockers, Matthew L.: On Distant Reading and Macroanalysis, 01.07.2011, http://www.matthew jockers.net/2011/07/01/on-distant-reading-and-macroanalysis/ (06.04.2017).
- Jockers, Matthew L.: Revealing Sentiment and Plot Arcs with the Syuzhet Package, 02.02.2015, http://www.matthewjockers.net/2015/02/02/syuzhet/ (06.02.2017).
- Jockers, Matthew L.: So What?, 07.05.2014, http://www.matthewjockers.net/2014/05/07/sowhat/ (27.10.2017).
- Jockers, Matthew L.: The Rest of the Story, 25.02.2015, http://www.matthewjockers.net/2015/ 02/25/the-rest-of-the-story/ (30.10.2017).
- Juola, Patrick: »Authorship Attribution«, in: Foundations and Trends in Information Retrieval 1.3 (2006), S. 233-334.
- Juola, Patrick: »The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions«, in: Digital Scholarship in the Humanities 30.1 (2015), S. 100–113.
- Kestemont, Mike, Moshe Koppel, Justin Anthony Stover und Yaron Winter: »Computational Authorship Verification Method Attributes a New Work to a Major 2nd Century African Author«, in: Journal of the Association for Information Science and Technology 67.1 (2016), S. 239-242.

- Kestemont, Mike: »Function Words in Authorship Attribution. From Black Magic to Theory?«, in: Proceedings of the Third Workshop on Computational Linguistics for Literature (2014), S. 59-66.
- Kirsch, Adam: Technology Is Taking Over English Departments. The False Promise of the Digital Humanities, 02.05.2014, https://newrepublic.com/article/117428/limits-digital-humanit ies-adam-kirsch (27.10.2017).
- Lutosławski, Wincenty: »Principes de stylométrie«, in: Revue des études grecques 41 (1890), S. 61-80.
- Lutosławski, Wincenty: The Origin and Growth of Plato's Logic. With an Account of Plato's Style and of the Chronology of his Writings. London, New York, Bombay 1897.
- Mendenhall, Thomas Corwin: »A Mechanical Solution of a Literary Problem«, in: The Popular Science Monthly 60 (1901), S. 97-105.
- Moretti, Franco: »Conjectures on World Literature. My Mission: To Say it More Simply than I Understand it. Schönberg, Moses and Aaron«, in: New Left Review 1 (2000), S. 54-68.
- Morey, John Hope: Joseph Conrad and Ford Madox Ford: A Study in Collaboration. Ithaca, NY 1960.
- Reiter, Nils und Marcus Willand: »Geschlecht und Gattung. Digitale Analysen von Kleists ›Familie Schroffenstein«. In: Kleist-Jahrbuch (2017), S. 177-195.
- Schöch, Christof: »Corneille, Molière et les autres. Stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik«, in: Literaturwissenschaft im digitalen Medienwandel. Beihefte zu Philologie im Netz 7 (2014), S. 130-157.
- Stamatatos, Efstathios: »A Survey of Modern Authorship Attribution Methods«, in: American Society for Information Science and Technology 60.3 (2009), S. 538-556.
- Thaller, Manfred: »Geschichte der Digitial Humanities«, in: Digital Humanities. Eine Einführung, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein. Stuttgart 2017, S. 3-12.
- Tuldava, Juhan: »Stylistics, Author Identification«, in: Quantitative Linquistics, hg. v. Gabriel Altmann, Reinhard Köhler und Rajmund G. Piotrowski. Berlin 2005, S. 368-387.
- Vgl. Becker, Martin, Lena Hettinger, Andreas Hotho, Fotis Jannidis und Isabella Reger: »Genre Classification on German Novels«, in: Proceedings of the 12th International Workshop on Text-based Information Retrieval (2015), http://www.uni-weimar.de/medien/webis/ events/tir-15/tir15-papers-final/Hettinger2015-tir-paper.pdf (31.10.2017, 15.41 Uhr).
- Viehhauser, Gabriel: »Historische Stilometrie? Methodische Vorschläge für eine Annäherung textanalytischer Zugänge an die mediävistische Textualitätsdebatte«, in: Grenzen und Möglichkeiten der Digital Humanities. Sonderband der Zeitschrift für digitale Geisteswissenschaften 1, hg. v. Constanze Baum und Thomas Stäcker. 2015, http://www.zfdg.de/ sb001_009 (30.10.2017).
- Weber, Max: Wissenschaft als Beruf, in: Max Werber. Schriften 1894-1922, hg. v. Dirk Kaesler. Stuttgart 2002, S. 474-511.
- Weimar, Klaus: »[Art.] Hermeneutik«, in: Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte, hg. v. Harald Fricke. Berlin, New York 2007, S. 25-29
- Weitin, Thomas: »Selektion und Distinktion. Paul Heyses und Hermann Kurz' Deutscher Novellenschatz als Archiv, Literaturgeschichte und Korpus«, in: Archiv/Fiktionen. Verfahren des Archivierens in Literatur und Kultur des langen 19. Jahrhunderts, hg v. Daniela Gertz und Nicolas Pethes. Freiburg, Berlin, Wien 2016, S. 385-408.