

Artikel aus:
Zeitschrift für digitale Geisteswissenschaften

Titel:
Die Digitalisierung des goldenen Zeitalters – stilometrische Autorschaftsattribution am Beispiel des Quijote

Autor/in:
Nanette Rißler-Pipka

Kontakt:
rissler@romanistik.uni-siegen.de

Institution:
Universität Siegen

GND:
[130085154](#)

ORCID:
[0000-0002-0719-9003](#)

DOI des Artikels:
[10.17175/2018_004_v2](#)

Nachweis im OPAC der Herzog August Bibliothek:
[101141175X](#)

Erstveröffentlichung:
28.09.2018

Version 2.0:
14.11.2022

Lizenz:

Sofern nicht anders angegeben 

Medienlizenzen:
Medienrechte liegen bei den Autor*Innen

Letzte Überprüfung aller Verweise:
12.10.2022

GND-Verschlagwortung:
[Literaturwissenschaft](#) | [Digital Humanities](#) | [Romanistik](#) | [Geschichte 1450-1650](#) |

Zitierweise:
Nanette Rißler-Pipka: Die Digitalisierung des goldenen Zeitalters – stilometrische Autorschaftsattribution am Beispiel des Quijote. In: Zeitschrift für digitale Geisteswissenschaften. Erstveröffentlichung vom 28.09.2018. Version 2.0 vom 14.11.2022. Wolfenbüttel 2022. PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/2018_004_v2](#).

Version 2.0 vom 14.11.2022:
Es wurden folgende Änderungen vorgenommen: Inhaltliche Anpassungen, wie sie von den Gutachten angemerkt worden sind. Aktualisierung und Ergänzung der bibliografischen Angaben. Formale Korrekturen.

Nanette Reißler-Pipka

Die Digitalisierung des goldenen Zeitalters – stilometrische Autorschaftsattributions am Beispiel des Quijote

Abstracts

Spaniens Goldenes Zeitalter scheint wie geschaffen für die quantitative Textanalyse. Gibt es doch zahlreiche Texte und verschiedene Editionen in digitalisierter Form in Virtuellen Bibliotheken wie die bekannte BVMC (Biblioteca Virtual Miguel de Cervantes). Allerdings braucht man für stilometrische Analysen nicht nur den Text selbst, sondern auch korrekte und vollständige Metadaten. Da schon die zeitgenössischen Editionen des Goldenen Zeitalters nur schwer bezüglich der Textqualität zu überprüfen sind, werden zusätzliche Informationen (die sogenannten Metadaten) zu Text, Autor, Entstehungsdatum und -ort umso wichtiger. Im Folgenden wird die verfügbare Textgrundlage für stilometrische Analysen problematisiert und am Beispiel von Autorschaftsattributionsen für den apokryphen zweiten Band des Quijote von Avellaneda (Pseudonym) angewendet. Das Ziel ist dabei nicht die wahre Identität Avellanedas aufzudecken, sondern die Methode mit verschiedenen Parametern (cosine Delta, eingeschränkte Wortliste, Culling, rolling Delta) zu testen und die Ergebnisse im Kontext anderer Methoden zur Autorschaftsattributions sowie im Kontext der traditionellen Literaturgeschichte und -wissenschaft zu diskutieren.

For quantitative literary analysis, the Spanish Golden Age seems to be a perfect field of research. Many texts are provided by Virtual Libraries like the most known BVMC (Biblioteca Virtual Miguel de Cervantes) in various editions. Though for textual analysis like stylometry not only the text itself, but also the metadata should be complete and correct. As the contemporary editions of the Golden Age are themselves difficult to judge regarding the textual originality, additional information (usually called metadata) about the origin of text, author, date, place is needed. The following paper discusses the problem of existing textual basis for stylometric analysis using the example of authorship attribution for the apocryphal second volume of the Quijote by Avellaneda (pseudonym). The aim is not to discover the real identity of the author Avellaneda, but to test the method with various parameters (cosine Delta, selected wordlist, culling, rolling Delta) and to discuss the results in the context of other existing authorship attributional methods and in the context of traditional working literary history and studies.

1. Einleitung

Spaniens goldenes Zeitalter (Siglo de Oro) bietet sich literaturgeschichtlich betrachtet in doppelter Weise für die Digitalisierung an. Zum einen gilt es als wichtigste literarische, kulturelle Epoche Spaniens und hat neben dem klassischen Theater (comedia, teatro clásico), der Lyrik (Góngora) bis hin zum Roman (vom Lazarillo zum Quijote) die Weltliteratur geprägt. Zum anderen sind die Texte des 16. und 17. Jahrhunderts nicht mehr von Urheberrechten betroffen, die es im Falle einer digitalen Veröffentlichung zu beachten gäbe. Außerdem waren die Jahre zwischen 2005–2016 von zahlreichen 400-jährigen Jahrestagen vom ersten *Quijote* bis hin zu Cervantes' Todestag bestimmt und haben so Fördermittel auch für Digitalisierungsprojekte gebracht.

Doch in welcher Form liegen die Klassiker der spanischen Literatur vor? Treffen wir nur auf Digitalisate oder auch auf echte Editionen? Wird die zeitgenössische Editionsproblematik zu Beginn des 17. Jahrhunderts und früher in den digitalen Editionen mitreflektiert und nachvollzogen? Ist es möglich, sich ein verlässliches Vergleichskorpus¹ in ausreichender Anzahl für stilometrische Untersuchungen zu erstellen? Wie steht es um die angebotenen Metadaten, sind sie ebenso zahlreich, korrekt und wiederverwertbar wie diejenigen einer gedruckten kritischen Edition?² Diese Fragen gilt es zu Beginn einer Debatte um Autorschaft und Stilometrie zu klären, da die Texte die Basis der statistischen Untersuchung bilden.

In einem zweiten Schritt wird exemplarisch für andere zahlreiche ungeklärte Autorschaften im Siglo de Oro der Fall des falschen oder apokryphen zweiten Teils des *Quijote* verhandelt, der von einem gewissen Alonso Fernández de Avellaneda geschrieben und der, kurz bevor Cervantes seinen eigenen zweiten Teil herausbringen konnte, veröffentlicht wurde. Der Name Avellaneda ist offensichtlich ein Pseudonym wie Cervantes bereits im Prolog zu seinem Quijote II beklagt.³ Selbst über den Druckort und die Erstausgaben gibt es bis heute nur vage Vorstellungen und es ranken sich darum viele Expert*innendiskussionen.⁴

Diese offene Frage um die Identität Avellanedas hat innerhalb der 400 Jahre Quijote-Forschung zu zahlreichen Thesen geführt, die aktuell durch die Verfügbarkeit digitalisierter Primärtexte offenbar angeheizt, umso zahlreicher vertreten sind. Es klafft jedoch eine erstaunliche Lücke zwischen historisch-philologischen oder auch pseudo-statistischen Analysen auf der einen Seite und den innerhalb der DH und der forensischen Linguistik seit einigen Jahren etablierten Stilometrie. Beide Seiten scheinen sich unbewusst oder bewusst zu ignorieren. Weder die bekannten DH-Experten Patrick Juola und Christopher Coufal wenden sich 2010 mit ihrer aus hispanistischer Sicht unglaublichen These, dass Cervantes nicht der Autor der letzten 69 (von 74) Kapiteln des

¹ Vgl. die Veröffentlichung der hier verwendeten Textsammlung im [DARIAH-DE Repository](#).

² Vgl. dazu die Kritik an der Biblioteca Virtual Miguel Cervantes von Fradejas Rueda 2016.

³ Vgl. Cervantes 2005 [1615], S. 26.

⁴ Vgl. Blasco 2007, S. XIV sowie Gómez Canseco 2016.

2. Bandes des Quijote sei, an die Cervantes-Experten, sondern stellen ihre stilometrische Untersuchung auf der DH-Konferenz in London eher als Test ihres Tools (JGAAP) vor, ohne weitere Öffentlichkeit zu suchen;⁵ noch kümmert sich die Hispanistik um diese vermutlich als abseitig empfundene These aus der Stilometrie. Die These von Coufal / Juola widerspricht zudem der gängigen Meinung, dass Cervantes ab dem 60. Kapitel des 2. Bandes des Quijote seinen Stil ändere, weil er ab diesem Zeitpunkt von der Existenz des Avellaneda-Buches erfuhr.⁶ Gerade dieser Stilwechsel innerhalb des zweiten Quijote-Bandes von Cervantes wird auch in einer kaum beachteten Analyse des venezolanischen Mathematikers Freddy López Quintero 2011 verhandelt, der zwar kollaborative Autorschaft ins Spiel bringt, aber das Avellaneda-Buch gar nicht einbezieht.

Insgesamt soll dieser Beitrag dazu dienen, beide Seiten, die Hispanistik und DH, in Dialog zu bringen.⁷ Ausgehend vom Konferenzpaper »Der falsche Quijote. Autorschaftsattributions für spanische Prosa der frühen Neuzeit« (DHd2016, Leipzig) konzentriert sich die Studie im Folgenden auf die Schwierigkeiten einer stilometrischen Analyse im Kontext des Siglo de Oro sowie auf die Frage eines thematischen Signals (hier: Quijote), das stärker ist als die Signale »Autorschaft« und »Gattung« und möglicherweise die Gruppierung beeinflusst.

2. Digitale Textgrundlage für spanische Literatur der frühen Neuzeit – Verfügbarkeit und Qualität von Text und Edition

Der Aufbau eines umfangreichen, repräsentativen und qualitativ verlässlichen Korpus ist für die Stilometrie, aber auch für viele andere textbasierte Verfahren der Digital Humanities unerlässlich. Allein die Menge der Texte, die aus der Zeit des 16.–17. Jahrhunderts digitalisiert vorliegt, ist erstaunlich und scheint vieles in diesem Bereich möglich werden zu lassen, was vor den Zeiten der Digitalisierung undenkbar erschien. Die wichtigste virtuelle Bibliothek für spanischsprachige Texte ist zweifelsohne die [Biblioteca Virtual Miguel Cervantes](#) (BVMC), die durch eine [private Stiftung](#) finanziert wird, deren Vorsitz der Nobelpreisträger Mario Vargas Llosa innehat. Über den Aufbau der Bibliothek lässt sich streiten, der Umfang der Bücher jedoch, die aus dem Siglo de Oro stammen, ist immens. Genaue Zahlen sind wegen stetiger Aktualisierung und fehlender Aufschlüsselung zwischen Sekundär- und Primärliteratur zwar nicht zu erhalten, aber gerade die spanischen Klassiker sind oft gleich in mehreren Ausgaben vertreten. Aus urheberrechtlichen Gründen müssen allerdings zumeist Herausgaben genommen werden, an denen auch die Herausgeber keine Rechte mehr am Buch haben. Ältere Ausgaben sind aber zumeist schwieriger zu scannen und / oder entsprechen nicht dem modernen Spanisch. Der Linguist und Experte für digitale Editionen José Manuel Fradejas Rueda konnte der BVMC nachweisen, dass im Fall des mittelalterlichen Textes *Libro de la caza de las aves* (14. Jh.) von Pero López de Ayala eine Edition von 1969 statt der angegebenen von 1876 verwendet wurde.⁸ Solche Enthüllungen scheinen aber keinen Einfluss auf die in ihrer Bedeutung für die Wissenschaft wachsende Stiftung BVMC zu haben. Von Editionen, die dazu gemacht werden, um damit im literaturwissenschaftlich quantitativen Sinn oder auch mit Annotationstools oder den Metadaten der Texte zu arbeiten, ist leider auch die spanische Literatur der frühen Neuzeit weit entfernt. Es gibt aktuell im spanischsprachigen Raum kein Repositorium wie Textgrid, das verlässliche Daten und Metadaten in wiederverwertbaren Formaten und eindeutig identifizierbar (DOI) anbietet (auch im Textgrid-Repository können sich allerdings Fehler einschleichen, vgl. [Abbildung 2](#)). Einen Überblick bezüglich der sehr verschiedenen Portale, die spanischsprachige Textressourcen ermöglichen, bietet José Calvo Tello in [GitHub](#). Lediglich die Real Academia Española hat mit [CORDE](#) (Corpus Diacrónico del Español) ein zumindest durchsuchbares Korpus im repräsentativen Umfang für die Zeitspanne von 1500–1700. Da kein Zugriff auf die Volltexte möglich ist, kann damit allerdings für stilometrische Untersuchungen nicht gearbeitet werden. CORDE spielt aber bei fast allen aktuellen Autorschaftsattributions eine Rolle und kann bei solchen, die Konkordanzen nicht mithilfe des Korpus abgeglichen haben, leicht zur Falsifizierung eingesetzt werden. Die Schwierigkeiten bezüglich der Erstellung eines Korpus für das Siglo de Oro aber auch der Digitalisierung selbst wurden jüngst von Javier Rosa Pérez dargelegt.⁹

3. Der falsche Quijote von Avellaneda: ein Literaturstreit mit Cervantes

Schon zu Cervantes' Zeiten begann die Diskussion um Text, Autor, Eigentum in einer diffizilen Gemengelage von Druckgenehmigung, Editionen, Zensur und Plagiat. War der Lazarillo de Tormes (1554) noch als Erfahrungsbericht des gleichnamigen Ich-Erzählers lesbar,¹⁰ so entspann sich kurz danach schon die Frage um die Autorschaft¹¹ und der beliebte Text

⁵ Vgl. Coufal / Juola 2010. Laut dieser Studie müsste Cervantes demnach nur die ersten 5 Kapitel des Quijote II selbst geschrieben haben.

⁶ Vgl. Strosetzki 1991, S. 93; Ehrlicher 2008a, S. 42ff.; Blasco 2007, S. XVII; Gómez Canseco 2008, S. 33ff.

⁷ In diese Richtungen gingen auch die Bemühungen von Eder, Jannidis, Rybicki, Schöch, Dalen-Oskam im Stilometrie Panel »Literary Concepts: The Past and the Future« der DH2016, Kraków, vgl. Eder et al. 2016a.

⁸ Vgl. Fradejas Rueda 2016, S. 199ff.

⁹ Vgl. Rosa Pérez 2016, S. 115ff.

¹⁰ Vgl. Spitzer 1959, S. 103: Spitzer argumentiert ebenso wie später Barthes (in seinem Aufsatz zum Tod des Autors, vgl. Barthes 1968), dass der mittelalterliche Ich-Erzähler nicht zwischen Autor, Erzählung und Erfahrung unterscheidet: »And we must assume that the mediaval public saw in the »poetic« a representation of mankind, that it was interested only in this representative rôle of the poet.«

¹¹ Vgl. Rico 2010, S. 31–44.

wurde nicht nur in der neu entstandenen Gattung des Schelmenromans fortgeführt, sondern auch bewusst plagiiert. Auch kurz vor dem Fall des falschen Quijote (1614) musste sich bereits Mateo Alemán mit einem ungebetenen Nachfolgeschreiber befassen, der nur ein Jahr vor Erscheinen seines eigenen zweiten Teils des Guzmán de Alfarache einen unter Pseudonym veröffentlichten apokryphen Guzmán II publizierte.¹² Entsprechend erzürnt über den Nachahmer, geißelt Alemán diesen mit Satire im Prolog seines eigenen zweiten Teils.¹³

Das erscheint alles abenteuerlich, wenn man bedenkt, dass von 1487 bis ins 19. Jahrhundert hinein die spanische Inquisition und Krone den Buchmarkt vollständig kontrollierte. Wie können Zweifel über Autorschaften entstehen, wenn jeder Text, der veröffentlicht werden sollte, zunächst von der Druckgenehmigungsbehörde (staatlich) und danach noch von der Inquisition (kirchlich) bewilligt werden musste? Ab 1558 waren anonyme Bücher ebenso verboten wie ausländische, kritische oder im Ausland gedruckte spanische Bücher.¹⁴ Dennoch verzeichnet CORDE für den Zeitraum 1558–1650 noch 18 anonyme Werke im Korpus für spanische Narrativa (insgesamt 142), auch wenn es im Vergleich dazu von 1500–1558 verhältnismäßig viel mehr sind, nämlich 30 anonyme Prosawerke von nur 70, die in CORDE aufgeführt sind. Angesichts der überschaubaren Anzahl an Akteuren im literarischen Feld des Siglo de Oro erscheint die These von Luis Gómez Canseco wahrscheinlich, dass Cervantes seinen Plagiator kannte.¹⁵ Woher hätte dieser die Druckgenehmigung erhalten sollen, wenn nicht sein Name verifiziert werden konnte? War der in den Prologen der jeweiligen Bücher ausgetragene Streit demnach nur eine ›burla‹, ein Spaß im Sinne des Karnevals? Cervantes selbst weist in seinem Prolog zum Quijote II darauf hin: Der Autor des falschen *Quijote* versteckte sich hinter einem falschen Namen und erdachte sich ein Vaterland.¹⁶ Umgekehrt lästert Avellaneda in seinem Prolog zum Quijote II-A,¹⁷ dass Cervantes selbst wohl zu alt für einen eigenen zweiten Teil gewesen sei und auch durch eine Kriegsverletzung um eine Hand ärmer. Daher könne er hier ohne Scham als Autor einspringen.¹⁸ Daraufhin kontert Cervantes wiederum, dass man ein Buch ja nicht mit der linken Hand schreibe, sondern mit dem Kopf und er sozusagen im Folgenden eigenen zweiten Teil des Quijote zeige, welcher der »richtige« und »wahre« Quijote sei.¹⁹ Auch im Text selbst erfolgt vor allem bei Cervantes eine Reflexion über Autorschaft und Originalität, die sich ohne den apokryphen Part von Avellaneda so vielleicht nicht ergeben hätte und die sehr zur Modernität und zum Witz des in manchen Teilen auch ausufernden zweiten Bandes beiträgt. So ›klaut‹ auch Cervantes bei Avellaneda, wenn er dessen Figur des Álvaro de Tarfe im eigenen Buch auftreten lässt und ihn dann mit den ›echten‹ Quijote und Sancho Panza konfrontiert, um den Ritter daraufhin schwören zu lassen, nur diese beiden Figuren im Cervantes-Buch seien wahrhaftig der berühmte Quijote und sein Knappe Sancho Panza, während die beiden Figuren, die er ›zuvor‹ im Avellaneda Buch getroffen habe, Fälschungen seien. Diese Kommunikation zwischen den Büchern zeugt von komplexen Erzählebenen und einem intellektuellen Witz, der typisch für Cervantes, aber auch typisch für die Zeit ist. In der Sekundärliteratur wird oft betont, dass der Quijote II-A nicht in der gleichen Weise witzig, intellektuell anspruchsvoll und eher angepasst an die Bedürfnisse der Zensurbehörden sei wie derjenige von Cervantes. Vor allem stilistisch sei die Sprache, die Avellaneda benutze, nicht auf dem gleichen Niveau und mit derselben Raffinesse vorgebracht wie bei Cervantes. Auch die Überzeichnung der Sancho Panza-Figur bei Avellaneda als bössartiger Menschenfeind stehe einem sich weiterentwickelnden, seinen Herren spiegelnden Sancho bei Cervantes gegenüber.²⁰ Die eigene Lektüre beider Bücher ist dabei immer von solchen Vorverurteilungen bestimmt und daher wäre eine stilistische Untersuchung, die nicht liest sondern rechnet, möglicherweise eine Alternative.

Indiz für eine gemeinsam erdachte ›burla‹, d. h. die Inszenierung eines Autorschafts- und Plagiatsstreits zur Belustigung und Täuschung (›engaño‹) des zeitgenössischen Publikums, das diese Art von Späßen gekannt und erwartet hat, ist außerdem Cervantes' Aufforderung am Ende des ersten Teils, die Geschichte weiter zu dichten. Er nutzt dazu eine Referenz auf Ariosts Orlando furioso (1516, dt. Der rasende Roland), dessen Ritterfigur in seiner Verrücktheit (furioso, rasend) neben dem *Amadís de Gaula* (14. Jh.) Vorbild für die Quijote-Figur war. Cervantes zitiert im *Quijote I* aus dem italienischen Original: »Forsi altro canterà con miglior plectio« – »Den wohl ein anderer singt in vollern Tönen«²¹ und beendet mit diesen letzten Worten das Buch. Das gleiche Zitat aus dem Orlando furioso findet sich in spanischer Übersetzung und um einen Vers erweitert im ersten Kapitel des Quijote II von Cervantes: »Y cómo del Catay recibió el cetro, quizá otro cantará con mejor plecto.« – »Und wie sie drauf sich ließ zu Catai krönen / Singt wohl ein anderer einst in kühnern Tönen.«²² Die Fortsetzung der Geschichte wird also nicht nur begrüßt, sondern auch noch qualitativ höherwertig von einem anderen Autor in Aussicht gestellt (›miglior plectio‹, ›mejor plecto‹, ›toller‹, ›kühner‹). Avellaneda nimmt diesen Wink auf und verwendet die Figur des rasenden Roland im Quijote II-A als Überschrift für

¹² Vgl. dazu Ehrlicher 2008a.

¹³ Alemáns Worte in der Widmung an Don Juan de Mendoza klingen kriegerisch und beginnen mit einer »Antwort« auf den gemeinen Diebstahl: »Ya es conocida la razón que tengo en responder por mi causa en el desafío que me hizo sin ella el que sacó la segunda parte de mi Guzmán de Alfarache«, vgl. Alemán 2000.

¹⁴ Vgl. Delgado 2006, S. 465.

¹⁵ Vgl. Gómez Canseco 2014, S. 11.

¹⁶ Vgl. Cervantes 2005 [1615], S. 26.

¹⁷ Zur einfachen Unterscheidung wird das Werk von Avellaneda im Folgenden durch *Quijote II-A* abgekürzt. Vgl. Avellaneda 2011 [1614].

¹⁸ Avellaneda 2011 [1614], S. 106.

¹⁹ Cervantes 2005 [1615], S. 25–26.

²⁰ Vgl. dazu die ausführliche Darlegung des aktuellen Standes in der Sekundärliteratur bei Alvarez 2014.

²¹ Vgl. Cervantes 2011 [1605], S. 664; deutsche Übersetzung von Ludwig Tieck im TextGrid Repository.

²² Vgl. Cervantes 2005 [1615], S. 42; deutsche Übersetzung von Ludwig Tieck im TextGrid Repository - interessant ist hier die Variation von »tollern« zu »kühnern« in der deutschen Übersetzung.

sein sechstes Kapitel: »De la no menos estraña que peligrosa batalla que nuestro caballero tuvo con una guarda de un melonar que él pensaba ser Roldán el Furioso«. ²³ Das Verweis- und Verschleierungsspiel der intertextuellen Bezüge fordert den Geist (»ingenio«, »agudeza«) des Lesers heraus, um den Spaß, den der oder die Autor/en mit ihm treiben, aufzudecken (»desengaño«). Dieser Witz, der sich ja im Täuschungsspiel, dem Quijote nun seitens seiner Leser*innen aus dem ersten Teil zum Opfer fällt (vgl. die lange Episode beim Herzog und der Herzogin), widerspiegelt, ist dem zeitgenössischen Publikum nicht nur im Kontext des Quijote bekannt und trug sicher entscheidend zu seinem Erfolg bei. ²⁴

Der aktuelle Boom in der Suche nach unbekannten, anonymen oder apokryphen Autoren hängt zwar eng mit den Entwicklungen im Bereich der Stilometrie und der DH zusammen, aber merkwürdigerweise verwenden die wenigsten Autorschaftsdetektiv*innen diese Methode. Daher lohnt sich zunächst ein Blick auf die in diesem Bereich aktuell angewendeten Methoden und Konzepte, bevor eigene Experimente mit Stilometrie und Cervantes / Avellaneda folgen.

4. Die Autorschaftsattributions im Vergleich

Im spanischsprachigen Raum gibt es für die großen anonymen Klassiker wie den Lazarillo de Tormes, aber auch für Avellanedas *Quijote* seit etwa den 2000er Jahren eine beachtliche Anzahl von Autorschaftsattributions, die zum Teil von Wissenschaftler*innen durchgeführt werden, die sich im Laufe der Zeit selbst widersprechen. Methodisch ist von philologischer Arbeit an Editionen (Vergleich verschiedener Erstausgaben und vermeintlicher Dokumentenfunde) über hermeneutische Argumentationen bis hin zu pseudostatistischen Untersuchungen eine große Bandbreite abgedeckt. Die Attributionen werden oft über die Zeitschrift *Lemir* lanciert oder von einflussreichen Wissenschaftler*innen direkt mit einer neuen Edition des Werkes inklusive neuer Autornamen herausgebracht. ²⁵ Die in Spanien gängige Praxis häufiger Neu-Editionen der Klassiker führt nicht zwingend zu mehr Wettbewerb und mehr Qualität, sondern zu einer unübersichtlichen Anzahl von Textfassungen und einem Mangel an editorischer Sorgfalt. Oft sind es die gleichen Forschenden, die sich zum Teil mit detektivischer Leidenschaft »ihrem« unbekannten Autor zuwenden und mit dieser Attribution auch gleich eine Edition des Werkes verbinden. ²⁶ So schlägt allein Luis Madrigal in den Jahren 2003–2014 zwei verschiedene Autoren für den *Lazarillo* vor und zwei verschiedene für Avellaneda. 2014 hat Madrigal seine Aussage von 2003 und 2009 widerrufen, dass man den Stil eines Autors so gut wie seine DNA bestimmen könne. ²⁷ Er ging z. B. davon aus, dass bestimmte Ortsnamen oder auch einzelne Phrasen den Autorstil kennzeichnen. Das sind für sich allein genommen schon recht ungewöhnliche Thesen, die jegliche Stildiskussionen der vergangenen 50 Jahre ebenso ignorieren wie aktuelle digitale Methoden. Stil ist eine komplexe Zusammenkunft von zählbaren und durch Interpretation bestimmten Elementen. Die literaturwissenschaftliche und linguistische Stildiskussion erstreckt sich nicht ohne Grund über verzweigte Theorien, die zwar vor allem in der literaturwissenschaftlichen Forschung kaum noch Anwendung finden, aber aktuell durch die quantitative Textanalyse wieder diskutiert werden. ²⁸

Wie gehen nun die »Autorschaftsdetektiv*innen« vor? ²⁹ In der überwiegenden Mehrzahl wird mit Konkordanzen gearbeitet, wobei entweder die Häufigkeit oder die Seltenheit entscheidend ist. Einfacher ist es natürlich bestimmte Ausdrücke zu finden, die angeblich nur in dem fraglichen anonymen oder apokryphen Werk vorkommen – dann aber passenderweise auch nur noch in dem einen oder mehreren Werken des favorisierten Autors. In dieser Weise vergleicht z. B. in der jüngsten Edition des Quijote von Avellaneda der Herausgeber López-Vázquez 15 Substantive in jeweils einem Werk von 6 verschiedenen Autoren, außerdem noch 10 Adjektive in der gleichen Auswahl. Daraus dann abzuleiten: »Los resultados son más claros, de las 15 palabras en Tirso aparece ninguna ...«, ³⁰ ist schon wissenschaftlich bedenklich, um nicht zu sagen pseudostatistisch. Ganz zu schweigen von der unterschiedlichen Bedeutung, die ein Wort wie »artificio« (das sich in der Liste von López-Vázquez befindet) in den Werken erhalten kann. ³¹ Ähnlich funktionieren sogenannten Passagenvergleiche, d. h. kleine Textabschnitte werden direkt miteinander

²³ Vgl. Avellaneda 2011, dt.: »Von dem nicht weniger ungewöhnlichen wie gefährlichen Kampf, den unser Ritter mit der Wache eines Melonenfeldes austrug, die er für den rasenden Roland hielt« (Übersetzung Nanette Rißler-Pipka).

²⁴ Dass es dabei gerade um Verstellung, Vorspiegelung und besonderes Talent bei derselben geht, zeigt folgendes Zitat aus Cervantes' *Quijote II*: »Reventaban de risa con estas cosas los duques, como aquellos que habían tomado el pulso a tal aventura, y alababan entre sí la agudeza y disimulación de la Trifaldi« (Cervantes 2005 [1615], S. 348) – »Über alle diese Reden starben die Herzoge fast vor Lachen, da sie das Abenteuer angeordnet hatten und im stillen die Klugheit und Verstellung der Dreischleppina bewunderten.« Deutsche Übersetzung von Ludwig Tieck im [TextGrid Repository](#).

²⁵ Vgl. dazu den von Ehrlicher und anderen kritisieren Fall der Autorschaftsattributions zugunsten von Alfonso de Valdés durch Rosa Navarro Durán: Ehrlicher 2008b.

²⁶ Vgl. Navarro Durán 2003; Suárez Figaredo 2011; Suárez Figaredo 2014; López-Vázquez 2011a.

²⁷ Vgl. Attributionen für den *Lazarillo*: Madrigal 2003 (Cervantes de Salazar), Madrigal 2008 (Arce de Otálora), Madrigal 2014 (erneut Arce de Otálora). Für Avellaneda schlägt er sowohl Tirso de Molina als auch Lope de Vega vor (vgl. Madrigal 2009). Zur Behauptung, der Stil sei wie die DNA eines Menschen zu entschlüsseln, vgl. Madrigal 2009, S. 193f. Zum Widerruf vgl. Madrigal 2014, S. 90.

²⁸ Vgl. Herrmann et al. 2015.

²⁹ Vgl. de la Rosa / Suárez 2016, S. 385–388: Die Autoren stellen ebenfalls heraus, dass in den meisten Fällen Konkordanzen (KWIC, key words in context) und biografische Details den Ausschlag bei Autorschaftsattributions geben, für den Fall des *Lazarillo* findet sich hier außerdem eine vollständige Liste der in der Forschung von 1607–2014 angeführten Autorschaftskandidaten.

³⁰ López-Vázquez 2011b, S. 29; dt.: »Die Ergebnisse sind eindeutig. Von den 15 Wörtern erscheint keines bei Tirso ... « (Übersetzung Nanette Rißler-Pipka).

³¹ Vgl. zur unterschiedlichen Verwendung von »artificio« in den drei Quijote-Bänden eine Detailanalyse von Ehrlicher 2016, passim.

verglichen und – wie erwartet – finden sich in genau diesen Passagen zahlreiche Ähnlichkeiten. Mit ganz anderer Zielrichtung als López-Vázquez stellt in diesem Zusammenhang Martín Jiménez die beiden zweiten Teile des Quijote von Cervantes und Avellaneda gegenüber (vgl. Abbildung 1).

INICIO DEL QUIJOTE DE AVELLANEDA (1614)	INICIO DE LA SEGUNDA PARTE DEL QUIJOTE DE CERVANTES (1615)
<p>El sabio Alisolan, historiador no menos moderno que verdadero, dice que, siendo expelidos los moros agarenos de Aragón, de cuya nación él decendía, entre ciertos anales de historias halló escrita en arábigo la tercera salida que hizo del lugar del Argamesilla el invicto hidalgo don Quijote de la Mancha para ir a una justas que se hacían en la insigne ciudad de Zaragoza, y dice desta manera:</p> <p>Después de haber sido llevado don Quijote por el cura y el barbero y la hermosa Dorotea a su lugar en una jaula, con Sancho Panza, su escudero, fue metido en un aposento con una gruesa y pesada cadena al pie, adonde, no con pequeño regalo de pistos y cosas conservativas y sustanciales, le volvieron poco a poco a su natural juicio.</p>	<p>Cuenta Cide Hamete Benegeli, en la segunda parte desta historia y tercera salida de don Quijote, que el cura y el barbero se estuvieron casi un mes sin verle, por no renovarle y traerle a la memoria las cosas pasadas; pero no por esto dejaron de visitar a su sobrina y a su ama, encargándolas tuviesen cuenta con regalarle, dándole a comer cosas confortativas y apropiadas para el corazón y el cerebro, de donde procedía, según buen discurso, toda su mala ventura. Las cuales dijeron que así lo hacían, y lo harían, con la voluntad y cuidado posible, porque echaban de ver que su señor por momentos iba dando muestras de estar en su entero juicio; de lo cual recibieron los dos gran contento, por parecerles que habían acertado en haberle traído encantado en el carro de los bueyes...</p>

Abb. 1: Visualisierung in Jiménez 2016, S. 34.

Martín Jiménez' These lautet, dass Cervantes von seinem apokryphen Nachahmer abgeschrieben habe und nicht umgekehrt. Cervantes habe das Werk Avellanedas vor Erscheinungsdatum gekannt und konnte sich daher daran orientieren. Dies soll der obige Vergleich der Eingangspassagen beider Werke belegen. Erstens ist es jedoch folgerichtig, dass beide zweite Teile des Quijote ähnlich beginnen, weil sie beide die Ankündigung aus dem ersten Teil aufnehmen (müssen) und den »dritten Ausritt« (tercera salida) des Quijote ankündigen. Auch der Erzähler (Aisolan bei Avellaneda, Hamete Benegeli bei Cervantes) wird von beiden ähnlich gewählt und bezieht sich auf den vorangegangenen ersten Teil. Es ist leicht ersichtlich, dass solche Vergleiche einen sehr großen Interpretationsspielraum lassen, der ohne Zweifel in jeder Werkanalyse enthalten ist. Wissenschaftlich bedenklich wird es allerdings, wenn mit Zahlen und Worten als Fakten operiert wird, deren Interpretation als logische oder mathematische Wahrheit präsentiert wird (z. B. im Fall von López-Vázquez). Zum Teil finden sich auch schlichte Falschangaben, die leicht überprüft werden könnten, aufgrund der schieren Masse von Konkordanzen wird sich aber kaum jemand die Mühe machen.³²

Auch vermeintliche Dokumentenfunde sind nach über 400 Jahren zum einen unwahrscheinlich und zum anderen müssen auch diese mit entsprechender wissenschaftlicher Präzision untersucht werden. Die Dokumentenfunde, die Mercedes Agulló anführt, um Hurtado de Mendoza als Autor des Lazarillo zu definieren, sind allerdings wenig präzise und ihre darauf basierenden weitreichenden Schlussfolgerungen sind in der Hispanistik nicht auf Anerkennung gestoßen.³³ Für die Literaturwissenschaft, die mit den Texten und Metadaten (wozu auch der Autorname gehört) arbeiten möchte, ist die Arbeit der Editionsphilolog*innen immer eine geschätzte und verlässliche Prädisposition der eigenen Arbeit gewesen. Gerade die Digitalisierung ganzer Bibliotheken sollte die Sorgfalt in diesem Bereich erhöhen, da man auf diese Weise Metadaten und Texte in großer Anzahl auswerten kann, um daraus Schlüsse zu ziehen, die sich auf ein großes Korpus beziehen – die Überprüfung der einzelnen Daten und Texte kann in diesem Umfang nicht mehr erfolgen. Leider finden fragwürdige Autorschaftsattributions immer wieder Eingang in seriöse digitale Bibliotheken. So wird z. B. bis heute Hurtado de Mendoza als Autor des Lazarillo im Textgrid-Repository geführt (vgl. Abbildung 2), obwohl im entsprechenden Werk selbst darauf hingewiesen wird, dass dessen Autorschaft umstritten ist und in der Hispanistik das Werk weiterhin unter anonymem Verfasser geführt wird.³⁴ Für Avellaneda und den falschen Quijote ist das Problem nicht in gleicher Weise vorhanden, weil hier zumindest ein Name, wenn auch ein Pseudonym, eingetragen werden kann und so keine Leerstelle im Bibliothekskatalog entsteht, die mit fragwürdigen Kandidat*innen gefüllt werden könnte.

³² Vgl. den Nachweis einer nicht in der gleichen Weise nachvollziehbaren CORDE-Suche im Fall von Madrigals *Lazarillo*-Autorschaftsattributions in Reißler-Pipka 2016, S. 329–332.

³³ Vgl. Agulló 2010 und Agulló 2011, passim. Vgl. auch Ehrlicher 2008b.

³⁴ Im Vorwort zur TextGrid-Ausgabe heißt es: »Es handelt sich hier um ein anonymes Werk verschiedener Autoren, das u. a. auch Hurtado de Mendoza zugeschrieben wird.« (vgl. *Lazarillo de Tormes im TextGrid Repository*). Auch in der BVMC gibt es den entsprechenden Eintrag einer Ausgabe, die Hurtado de Mendoza als Autor nennt, aber es gibt daneben **andere Ausgaben**, die keine Autorinformation enthalten.

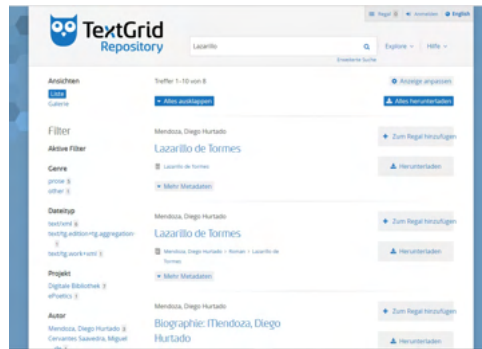


Abb. 2: Fensteransicht der Suchabfrage »Lazarillo« in TextGrid Repository.

Die Folgen solcher nicht sicher belegbaren Autorschaftsattributionsen sind immens, wie man am TextGrid-Beispiel sieht. Sämtliche Bibliotheksnutzer*innen, die sich zuvor nicht näher mit dem Werk auseinander gesetzt haben, müssen annehmen, Hurtado de Mendoza sei der Autor des *Lazarillo de Tormes*. Die Falschanagabe setzt sich unweigerlich in Werken fort, in denen diese Ausgabe zitiert und möglicherweise noch zusätzlich die hier angegebene Biografie von Mendoza genutzt wird, um Rückschlüsse auf das Werk, das ihm zugeschrieben wurde, zu ziehen. Denn dieser Autornamen oder auch ein anderer nimmt unweigerlich Einfluss auf die Lektüre des Werkes selbst, unabhängig davon, ob sich der*die Rezipient*in die Theorien vom »Tod des Autors« (Barthes, Foucault) zu eigen macht oder nicht.³⁵

5. Stilometrie mit R für spanische Literatur der frühen Neuzeit: Korpus, Methode, Ergebnisse

Die Frage ist nun, warum man den zahlreichen Autorschaftsattributionsen eine weitere hinzufügen sollte, die ebenso wenig stichhaltige Beweise liefern kann wie die vorigen. Es geht zum einen darum, eine Methode zu betrachten, deren Funktionalität in mathematisch-statistischer Hinsicht weit intensiver untersucht wurde und zahlreich in anderen Kontexten erprobt ist. Zur wissenschaftlichen Diskussion der Methode gehört aber auch, dass daraus eben kein Wahrheitsanspruch abgeleitet wird, sondern eine kritisch abwägende Analyse der Ergebnisse, die am Ende offen bleiben muss. Jedoch kann eine solche Analyse eine Revision der vorliegenden Autorschaftsattributionsen leisten sowie eine neue Ausrichtung der literaturwissenschaftlichen Kanondebatte um die stilistischen Merkmale des vermeintlich gefälschten Quijote. Zur Weiterentwicklung der Methode selbst kann hier nur insofern beigetragen werden, als bisher ausgeschlossen wurde, dass ein thematisches Signal die Gruppierung wesentlich beeinflussen könnte, dies zwar im Fall der drei Quijotes überlegenswert erscheint, aber hier widerlegt werden kann. Außerdem liegen für spanischsprachige Texte noch wenige Untersuchungen vor, die **Stilometrie mit Stylo für R** anwenden.³⁶ Eine Beweisführung im mathematischen Sinne kann aus literaturwissenschaftlichem Kontext heraus diesbezüglich nicht erfolgen. Die Funktionalität der Methode muss an dieser Stelle nicht erneut erklärt werden.³⁷ Es lässt sich zwar nach wie vor darüber streiten, ob die Verteilung von Häufigkeiten in der Verwendung der MFW (Most frequent words) tatsächlich das entscheidende Merkmal zur Unterscheidung von Autorenstilen sein kann, aber die Untersuchungen sollen hier schlicht eine Diskussionsebene eröffnen, die einen Vergleich zu den bisherigen, auf hermeneutischem oder auch dem oben erwähnten pseudostatistischen Weg ermittelten Ergebnissen anbietet.

Was nun die Frage nach der Autorschaft des »falschen« Quijote angeht, so wurden für die Identität Avellanedas bisher vor allem Lope de Vega, Suárez Figueroa, Ginés de Pasamonte und zuletzt Baltasar Navarrete (alias Úbeda als Autor der *Pícara Justina*) als Kandidaten angeführt.

Um ein Vergleichskorpus für die drei Quijotes zusammenzustellen, das sowohl bezüglich der Zeit, der Gattung und auch des Umfangs ähnlich ist, wurden aus den frei zugänglichen virtuellen Bibliotheken Wikisource, BVMC, Project Gutenberg sowie pdf-Editionen von Enrique Suárez Figaredo die entsprechenden Werke im Textformat gespeichert und eine Tabelle erstellt, die Metadaten wie Edition, Ressource, Sprachform (altspanisch oder neuspanisch) sowie die Anzahl der Wörter und die Spezifizierung der Gattung oder Subgattung enthält (zusätzlich zu den ohnehin unerlässlichen Daten wie Autornamen, Titel, Jahr). Dabei zeigte sich, dass im Umfang des Quijote, d. h. über 100.000 Wörter, nur noch 7 weitere Texte zum Vergleich in Frage kämen: Alemáns *Guzmán de Alfarache*, Suárez de Figueroas historischer Roman *El Pasajero*, Úbedas *Pícara Justina* (dessen Autorschaft

³⁵ Vgl. zu dieser Debatte ausführlicher Jannidis 1999 oder Schaffrick / Willand 2014.

³⁶ Vgl. dazu Calvo Tello 2015 und Calvo Tello 2016. Hier wurde R in der Version 3.2.3 und die Version 0.6.3 des *stylo* packages für R verwendet, vgl. Eder et al. 2013 und Craig et al. 2014 sowie The R Project for Statistical Computing 2015.

³⁷ Vgl. für den spanischsprachigen Kontext Calvo Tello 2015 und Calvo Tello 2016 sowie de la Rosa 2016; für den grundlegenden allgemeinen Kontext vgl. Eder et al. 2017 und Evert et al. 2016. Für eine stilometrische Untersuchung mittelalterlicher Texte vgl. Viehhauser 2017.

ebenfalls umstritten ist),³⁸ mit Espinels Vida del escudero Marcos de Obregón und Alcalá Yañez' Alonso, mozo de muchos amos zwei weitere Pícaro-Romane sowie de Silvas Fortführung der Celestina: Segunda Celestina (Tragikomödie) und schließlich die Crónica de la Nueva España von Cervantes de Salazar. Richtet man sich nur nach dem Umfang der Werke, ergeben sich Probleme der Gattungseinheitlichkeit sowie der Einzeltexte, die prinzipiell keinen Partner finden können (vgl. Abbildung 3). Da Untersuchungen vorliegen, die ein verlässliches Ergebnis auch bei Werken unterschiedlicher Länge nachweisen,³⁹ wurden auch hier unterschiedliche Größen zugelassen. Wie die ersten Ergebnisse zeigen, funktioniert die Zuordnung der Texte nach Autoren trotzdem einwandfrei (vgl. Abbildung 3, Abbildung 4 und Abbildung 5). Die Problematik einer Edition der »vielen Hände« bleibt davon allerdings unberührt, da bereits die Erstaussagen dieses Problem aufweisen.⁴⁰ Auch ist nicht gesichert, ob die von den virtuellen Bibliotheken zur Verfügung gestellten Metadaten korrekt sind (vgl. Anm. 7). Doch selbst wenn wir all diese Unwägbarkeiten berücksichtigen, sind die Ergebnisse aufgrund der Anzahl der verglichenen Texte und Wörter statistisch dennoch relevant, weil es sich kaum auswirkt, ob ein oder mehrere wenige Wörter von 5.000 mit dem Originalmanuskript oder anderen Editionen des gleichen Werkes unterschiedlich sind und so z. B. »mismo« oder entsprechend der älteren Form »mesmo« geschrieben wird. Gerade an diesem Beispiel lässt sich die sprachliche Einheitlichkeit des Korpus' überprüfen. Das Funktionswort »mismo / mesmo« (»gleich«) gehört gewöhnlich zu den 150 am häufigsten verwendeten Wörtern (MFW) in den untersuchten Texten.⁴¹ Solche verschiedenen Schreibformen, die zu unterschiedlichen Treffern bezüglich der MFW führen, werden in der Fachdiskussion als »Noise« bezeichnet, der z. B. im Fall von mittelalterlicher und frühneuzeitlicher Literatur mit sehr vielen unterschiedlichen Schreibweisen recht hoch ist und dennoch die Funktionalität der Stilometrie nicht entscheidend beeinflusst.⁴²

Für die Auswertungen wurde das von Jannidis et al. getestete cosine Delta verwendet,⁴³ das über die Serie von Experimenten 100–5.000 MFW konstant gute Ergebnisse brachte. Als Ausgangssituation wurden die Romane von Cervantes und die beiden Teile des Guzmán de Alfarache von Mateo Alemán mit einem Umfang etwa 100.000 Wörtern verglichen (Abbildung 3).

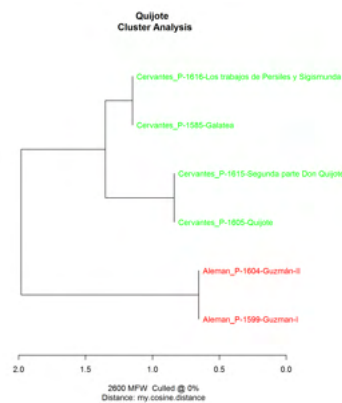


Abb. 3: Dendrogramm, cosine Delta, 2.600 MFW (von 200–4800 MFW alle konstant). [Rißker-Pipka 2016]

Fügt man nach bewährtem Verfahren danach zunächst diejenigen Werke hinzu, die zwar unterschiedlicher Länge, aber gleicher Gattung sind und von denen zumindest zwei von einem Autor im Korpus vorhanden sind, zeigt sich, dass zum einen das Korpus nicht entsprechend viele Texte von allen möglichen Autorschaftskandidaten hergibt und zum anderen aber, dass die Gruppierungen nach wie vor gut funktionieren (Abbildung 4).

³⁸ Vgl. Strosetzki 1991, S. 112.

³⁹ Vgl. Eder 2010.

⁴⁰ Vgl. dazu auch Blasco / Ruiz Urbón 2009, S. 44, die darauf hinweisen, dass aufgrund dieser historischen Textsituation stilometrische Untersuchungen am Textmaterial des Siglo de Oro ohnehin nicht gesichert sein könnten.

⁴¹ Dabei ist »mesmo« natürlich viel weniger gebräuchlich als die bei den meisten Editionen umgewandelte modernere Form »mismo«. Letztere belegt im Korpus, je nach Zusammensetzung z. B. Platz 131, während »mesmo« gleichzeitig auf Platz 649 landet. Interessant ist, dass im Quijote II-A und im Quijote II beide Formen auftauchen, während im Quijote I nur »mismo« verwendet wird. Alle drei Texte entstammen der BVMC und wurden von Florencio Sevilla Arroyo ediert: *Edición de Florencio Sevilla Arroyo in der BVMC*.

⁴² Vgl. dazu Evert et al. 2016, S. 64ff. und allgemein Eder 2013, passim.

⁴³ Vgl. Jannidis et al. 2015.

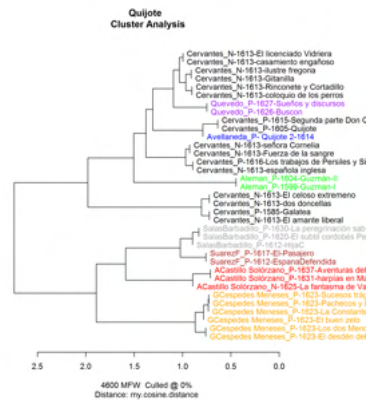


Abb. 4: Dendrogramm: cosine Delta, 4.600 MFW (16 Texte von Cervantes, 2 von Quevedo, 2 von Alemán, 3 von Salas Barbadillo, 2 von Suárez Figueroa, 3 von Castillo Solórzano, 6 von Céspedes y Meneses = 34 Texte). [Rißker-Pipka 2016]

Als einziger anonymer Text wurde außerdem Avellaneda Quijote II-A hinzugefügt. Er wird nicht zu dem Kandidaten für die Autorschaft Suárez de Figueroa gruppiert, sondern scheint mit den beiden Quijote-Bänden von Cervantes eine eigene »Gruppe« zu formen. Ebenso wie die Werke von Alemán und Quevedo das narrative Werk von Cervantes hier unterteilen, bilden die drei Quijotes einen eigenen Ast auf dem Dendrogramm. Interessant ist auch die Aufteilung zwischen den als Klassikern der spanischen Literaturgeschichte bekannten Autoren (Cervantes, Quevedo, Alemán) und den weniger bekannten Autoren (Salas Barbadillo, Suárez de Figueroa, Castillo Solórzano und Céspedes y Meneses). Legt man den Gattungsbegriff hier streng aus, sind auch die Werke *España defendida* (»poema heroico«, ein episches Heldengedicht) von Suárez de Figueroa und *La peregrinación sabia* (»fábula en prosa«, eine Fabel) von Salas Barbadillo als Abweichungen von der narrativen Gattung zu sehen. Trotz dieser Abweichungen werden diese Texte hier korrekt nach dem Autorschaftssignal gruppiert.

Fasst man jedoch die Frage nach Gattung und Umfang der Werke als Vergleichbarkeitsindiz enger und verzichtet dafür auf die Bedingung, dass mindestens zwei Texte eines Autors vorhanden sein müssen, erhält man zwar bedeutend weniger Texte, aber auch ein stimmiges, interessantes Ergebnis (Abbildung 5).

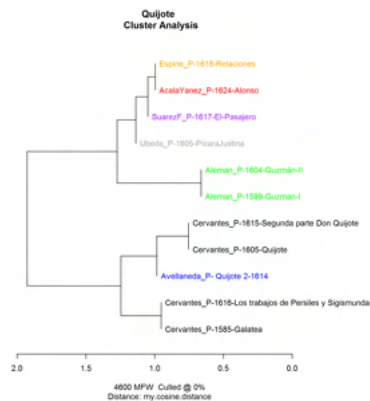


Abb. 5: Dendrogramm, cosine Delta, 4.600 MFW (alle Texte mind. 100.000 Wörter: 11 Texte). [Rißker-Pipka 2016]

Sämtliche Einzeltexte außer Avellaneda versammeln sich hier auf einem Ast. Dies ist besonders bemerkenswert, weil mit Úbeda (alias Baltasar Navarrete) und Suárez de Figueroa zwei Autorschaftskandidaten für Avellaneda im Korpus sind. Mit 11 Texten ist das Korpus zwar recht klein, aber da alle sehr umfangreich sind, gleicht sich dies wieder aus (1 Roman entspricht ca. 10 Novellen). Deutlich wird aber auch, dass nicht etwa die beiden Quijote II und Quijote II-A unmittelbar auf einem Ast gruppiert werden, sondern Quijote I und Quijote II von Cervantes.

Fügt man nun beide Experimente zusammen, wird das Ergebnis erstaunlicherweise noch stimmiger (Abbildung 6).

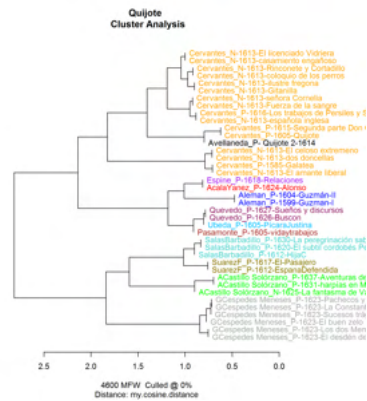


Abb. 6: Dendrogramm, cosine Delta, 4.600 MFW (39 Texte). [Riřker-Pipka 2016]

Obwohl mit Pasamonte hier ein weiterer Kandidat für die Autorschaft Avellanedas hinzugekommen ist, bleibt die Gruppierung im Grunde mit den obigen beiden Experimenten identisch: Quevedo und Alemán verschwinden aus dem Cervantes-Block und gruppieren sich in den Block der Einzeltexte (bilden aber nach wie vor autorschaftsgemäß einen Ast). Suárez de Figueroas Pasajero wandert wieder zu seinem passenden Text des gleichen Autors – trotz der Gattungsdifferenz. Avellaneda bleibt im Quijote-Block.

Schwieriger wird es bei Texten, die stilistisch gar nicht zu den anderen passen und auf diese Weise das Gesamtgefüge stören. Ein Beispiel dafür wäre die Crónica de la Nueva España von Cervantes de Salazar oder La mosquée (ein episches Gedicht) von Villaviciosa (vgl. Abbildung 7).

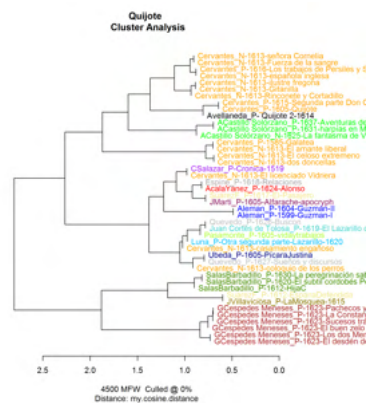


Abb. 7: Dendrogramm, cosine Delta, 4.500 MFW (44 Texte). [Riřker-Pipka 2016]

Während zuvor Gattungsdifferenzen und auch Einzeltexte noch kaum Auswirkungen hatten, zerstört die Hinzufügung von hier insgesamt 5 neuen Texten das Gesamtgefüge und ergibt auch literarhistorisch betrachtet keinen Sinn mehr. Die vorher korrekt zusammen gruppierten Werke von Suárez de Figueroa, Quevedo und Cervantes werden auseinander gerissen, und man mag stilistisch und hermeneutisch betrachtet beim besten Willen keine Verbindung zwischen der Crónica de la Nueva España und der Novelle El licenciado Vidriera von Cervantes entdecken.

Bei Einzeltexten, wie es auch der Quijote II-A ist, besteht generell in der Gruppierung im Dendrogramm (die statistisch gesehen nach dem nearest neighbour-Prinzip verläuft) die Gefahr, dass relativ weit voneinander entfernte Texte auf einem Ast gruppiert werden, weil sie schlichtweg von den anderen Texten noch weiter entfernt sind und eigentlich gar nicht in die Auswahl passen. Eine Überprüfung dieser Distanzen und der möglicherweise unverhältnismäßigen Gruppierung ist aber durch einen Blick auf die Distanzwerte in der von Stylo erstellten Tabelle »frequencies analyzed« möglich. Hier muss man sich nicht allein auf die Visualisierung verlassen, sondern kann die Werte dahinter einsehen.

Nimmt man folgerichtig die beiden Texte mit der größten stilistischen Abweichung wieder heraus (Crónica de la Nueva España und La mosquée) und fügt statt dessen drei der wenigen narrativen Texte von Lope de Vega hinzu (um den letzten Kandidaten für die Person Avellaneda dabei zu haben), ergibt sich trotz der großen Bandbreite zwischen unterschiedlicher Länge und Subgattung wieder ein nachvollziehbares Clustering (Abbildung 8).

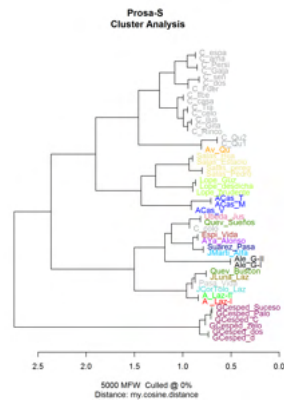


Abb. 8: Dendrogramm, cosine Delta, 5.000 MFW (49 Texte, abgekürzte Titel). [Rißker-Pipka 2016]

Zwar sind mit dem Lazarillo (I+II) zwei weitere anonyme Werke und mit der Novelle Tía fingida von Cervantes ein umstrittenes Werk hinzugekommen, aber die Zuordnung ist außer bei Quevedo und der pikaresken Novelle Coloquio de los perros von Cervantes weiterhin korrekt. Interessant ist hier, dass sich der pikareske Block zwischen den späteren Pícaro-Romanen um den Guzmán de Alfarache und den frühen um den Lazarillo in zwei Äste aufteilt, die dementsprechend auch die beiden Texte Quevedos trennen.

Zur Darstellung von Distanzen im Raum eignet sich allerdings auch die Netzwerkvisualisierung mithilfe von Gephi, die Stylo durch die Ausgabe einer Kantentabelle ermöglicht. Die Übersichtlichkeit ist bei dem recht groß gewählten Korpus zwar nicht mehr vollständig gegeben, aber man erkennt deutlich, dass die Verbindung im Cervantes-Werk (inkl. Avellaneda) und den pikaresken Romanen jeweils besonders stark ist (hier gekennzeichnet durch Dicke und Farbtintensität der Kanten).



Abb. 9: Netzwerkvisualisierung mit Gephi (vgl. Bastian et al. 2009). [Rißker-Pipka 2016]

Die meisten Einzeltexte scheinen aber in diesem Korpus nicht zum Problem zu werden, weil sie zusammen gruppiert werden. Gattungshistorisch interessant ist dabei, dass es die pikaresken Romane sind, die allesamt konstant einen Ast des Dendrogramms für sich beanspruchen, obwohl es sich zumeist um Einzeltexte handelt. Vermutlich ist allein die Präsenz eines autobiographisch berichtenden Ich-Erzählers derart stilprägend, dass es sich auf die statistische Verteilung des MFW auswirkt.

Das narrative Werk von Cervantes wird unabhängig von der zum Teil sehr unterschiedlichen Länge der Texte zwischen den großen Romanen (ca. 100.000 Wörter) und den Novellen (ca. 5.000 Wörter) recht einheitlich in den verschiedenen Parameterversuchen zusammen gruppiert. Nur die Novelle Coloquio de los perros verirrt sich bei manchen Korpuskonstellationen in den pikaresken Block (vgl. Abbildung 8 und vorige). Doch auch diese Gruppierung passt prinzipiell zur literaturgeschichtlichen Debatte, gilt doch gerade diese Novelle Cervantes' als besonders nah an der Pikareske. Wenngleich nach dieser Logik zumindest auch die Novelle Rinconete y Cortadillo in diese Gattung der Pikareske passen müsste.⁴⁴

Alle vier Autorschaftskandidaten für den »falschen« Quijote (Lope de Vega, Suárez Figueroa, Ginés de Pasamonte und Baltasar Navarrete (alias Úbeda als Autor der Pícaro Justina) sind mit mindestens einem Werk im Korpus vertreten, werden aber nicht mit Avellaneda zusammen gruppiert. Dagegen könnte man nun einwenden, dass die Beispieltex te schlicht stilistisch so unterschiedlich sind, weil sie nicht der gleichen Gattung angehören oder weil der Autor gerade für den Quijote II-A den Stil von Cervantes perfekt nachahmte. Die Tatsache, dass der Quijote II-A immer mit den beiden Quijotes von Cervantes gruppiert wird, lässt leicht die Vermutung zu, dass es sich hier doch um eine thematisch bedingte Gruppierung handeln muss.

⁴⁴ Vgl. Strosetzki 1991, S. 114.

Dieser letzten Frage kann man auf verschiedene Art und Weise nachgehen. Ein erster leichter Ansatz wäre es, die Hauptthemenwörter wie »Quijote«, »Sancho«, »Caballero«, die zum Teil auch in den MFW des Gesamtkorpus einen hohen Stellenwert einnehmen, schlicht aus der Wortliste zu streichen. Das führt allerdings erwartungsgemäß nicht zu einer Veränderung des Ergebnisses (vgl. Abbildung 10 und Abbildung 11). Für eine bessere Übersicht wurden hier die Einzeltexte wieder entfernt. Eine Änderung erfolgt nur zwischen 500 und 600 MFW, darüber ist alles wie bisher.

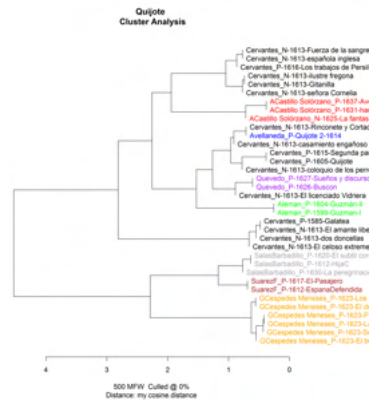


Abb. 10: Dendrogramme mit cosine Delta, 500 und 600 MFW: nur bis 500 MFW wird der *Quijote II-A* nicht direkt mit den beiden *Quijotes* von Cervantes gruppiert. [Riðsker-Pipka 2016] 2016

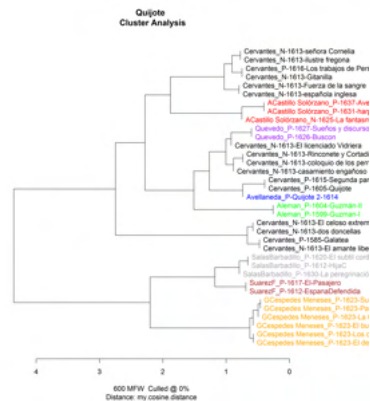


Abb. 11: Dendrogramme mit cosine Delta, 500 und 600 MFW: nur bis 500 MFW wird der *Quijote II-A* nicht direkt mit den beiden *Quijotes* von Cervantes gruppiert. [Riðsker-Pipka 2016] 2016

Auch wenn man »Culling« auf z. B. 20 einstellt, d. h. dass nur solche Wörter in die Berechnung einbezogen werden, die in 20% aller Texte des Gesamtkorpus vorkommen, verändert sich das Ergebnis nicht (vgl. Abbildung 12).⁴⁵ Auf diese Weise finden Wörter, die nur in z. B. drei Werken (wie den Quijote-Bänden) vorkommen, weil es beispielsweise die Namen der Hauptfiguren sind, keine Beachtung. So ist wesentlich unabhängiger von bestimmten selbst ausgewählten Wörtern (wie oben »Quijote« etc.) eine statistisch verwertbare Einschränkung der Wortliste gewährleistet. Doch wenn selbst dieses Feature keine Veränderung der Ergebnisse hinsichtlich inhaltlicher Gruppierung ergibt, kann man diesen Hintergrund mit einiger Sicherheit ausschließen.

⁴⁵ Vgl. Eder et al. 2017, S. 13. Man könnte den Wert, wenn man weiterhin bis zu 5.000 Wörter vergleichen möchte, nicht sehr viel höher als 20 ansetzen, weil bei 100 alle Wörter der Wortliste in allen Texten des Korpus vorhanden sein müssten. Auf diese Weise würde man kaum noch eine 5.000 Wörter lange Liste zusammenstellen können.



Abb. 12: Bootstrap Consensus Tree, cosine Delta, 1.000-4.000 MFV, Culled 20%. [Rißker-Pipka 2016]

Der Bootstrap Consensus Tree bietet hier die Möglichkeit noch mehr Texte in das Korpus aufzunehmen, die, wie man in Abbildung 12 sehr schön sehen kann, nach Gattung und Autor sortiert werden. So erkennt man klare Trennungen zwischen dramatischen, lyrischen und narrativen Texten, und Autoren, die in allen drei Gattungen vorhanden sind (wie z. B. Lope de Vega), werden zumindest innerhalb des Gattungssasts korrekt gruppiert.

Um vollständig sicher zu gehen, dass hier keine inhaltlich-thematische Beeinflussung der Gruppierung vorliegt, kann man auch einfach die beiden *Quijote*-Bände von Cervantes aus dem Gesamtkorpus entfernen und dann schauen, wohin Avellaneda mit seinem falschen Quijote gruppiert wird (Abbildung 13). Auch hier wird Avellaneda in den Cervantes-Block eingeordnet, obwohl sich auch nach wie vor die Novelle Colloquio de los perros von Cervantes in den Pícaro-Block gruppiert.

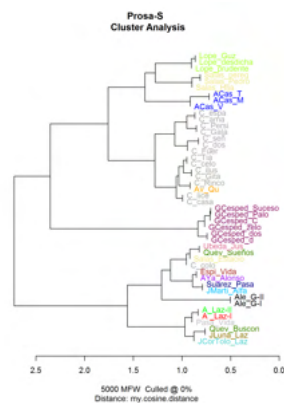


Abb. 13: Dendrogramm, cosine Delta, 5.000 MFV (49 Texte, abgekürzte Titel). [Rißker-Pipka 2016]

Es ändert sich trotz der Entfernung der direkten Partner des Quijote II-A im Vergleich zur ähnlichen Konstellation aus Abbildung 8 nichts in Bezug auf Avellanedas Gruppierung. Avellaneda scheint seinen Stil erfolgreich demjenigen von Cervantes anzupassen.

Daraus zu schließen, dass Cervantes und Avellaneda ein und dieselbe Person sind, wäre verfrüht. Zumal Cervantes' Werk selbst nicht vollständig in einen Block gruppiert wird und auch nicht dasjenige von Quevedo oder Salas Barbadillo. Offensichtlich ist daher nicht allein die Autorschaft entscheidendes Signal, sondern gerade in diesem Korpus des Siglo de Oro auch die Subgattung entscheidend. Interessant ist auch, dass Avellanedas Quijote II-A mit über 100.000 Wörter nicht zu Cervantes' Persiles oder Galatea gruppiert wird, die als einzige aus dem Werk von Cervantes etwa ebenso lang wären.

Wenn die Zusammensetzung des Korpus also insgesamt schwierig ist, lohnt sich vielleicht ein Blick ins Detail der drei oder zwei Quijotes. Angesichts des großen Umfangs könnte man auch aus jedem Kapitel einen eigenen Text machen, der immer noch ca. 4.000 Wörter hätte. Würde sich die These einer sehr großen Ähnlichkeit des Stils bzw. eines nahezu identischen Stils bestätigen, müssten die Kapitel von Quijote II-A und Cervantes' Quijote durchmischt werden. Sie werden jedoch nach Autor korrekt gruppiert, bis auf das Kapitel 5 von Cervantes und 35 von Avellaneda, die als einzige zusammen eingeteilt werden (Abbildung 14).

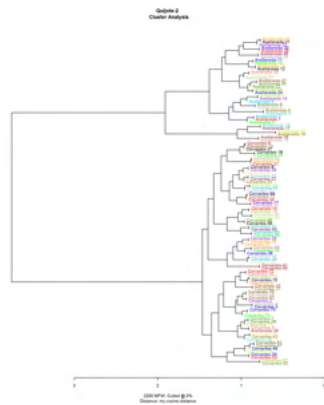


Abb. 14: Dendrogramm, cosine Delta, 2.200 MFV (Kapitel von *Quijote II-A* und *Quijote II*). [Rißker-Pipka 2016]

Die klare Trennung beider Bücher nach Kapiteln deutet auch nicht auf ein Abschreiben des einen oder anderen Autors hin,⁴⁶ sondern auf eine durchaus sichtbare stilistische Trennung. Kollaborative Autorschaft oder ein Stilwechsel mitten im Buch wie er von einigen in Bezug auf Cervantes angenommen wird,⁴⁷ lässt sich auf diese Weise jedoch nicht widerlegen.

6. Die Möglichkeit einer kollaborativen Autorschaft: Test des rolling delta für den Quijote II von Cervantes

Wenn Cervantes und Avellaneda auch nicht ein und dieselbe Person sind, so scheint es doch möglich, dass beide an den Büchern zusammengearbeitet haben oder jeweils sehr gut den Stil des anderen nachahmen konnten. Daher bietet sich abschließend die stilometrische Untersuchung des »rolling delta« an, um eine mögliche kollaborative Autorschaft am Quijote II nachzuweisen – oder mindestens die Teile des Quijote II aufzudecken, die sich besonders dem Quijote II-A annähern. Rolling delta bedeutet, dass der Gesamttext automatisch in Abschnitte eingeteilt wird und die Delta-Berechnung sozusagen abschnittsweise über die Texte verläuft. Es geht darum, stilistische Veränderungen oder Wechsel in einem bestimmten Text herauszufinden.⁴⁸ Angewendet auf den Quijote II von Cervantes würde das auch bedeuten, die These, dass Cervantes ab Kapitel 60 seinen Stil grundlegend wegen der Kenntnis über den apokryphen Quijote ändere oder sogar nur die ersten fünf Kapitel selbst geschrieben habe⁴⁹, überprüfen zu können.

Daher wurde hier als zu testender Text der Quijote II von Cervantes zugrunde gelegt (secondary set) und mit Avellanedas Quijote II-A sowie mit Quijote I und dem Persiles von Cervantes und danach mit Quevedos Buscón verglichen (primary set). Die Ergebnisse führen leider nicht sehr viel weiter, was die Frage nach einer kollaborativen Autorschaft betrifft (vgl. Abbildung 15, Abbildung 16 und Abbildung 17).

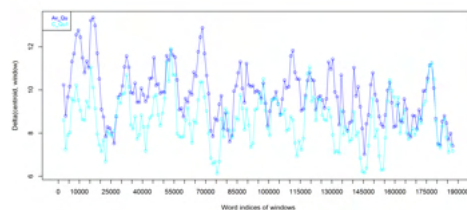


Abb. 15: Rolling Delta, Windowsize 5.000 (Oben: Av_Qu = Quijote II-A; C_Qu1 = Quijote I; Mitte: + C_Persiles; Unten: + Quev_Buscon). [Rißker-Pipka 2016]

⁴⁶ Vgl. Martín Jiménez 2016, S. 34 (vgl. auch Abbildung 1).

⁴⁷ Vgl. Coufal / Juola 2010 oder Strosetzki 1991, S. 93; Ehrlicher 2008a, S. 42ff.; Blasco 2007, S. XVII; Gómez Canseco 2008, S. 33ff.

⁴⁸ Vgl. Eder et al. 2016b, S. 118ff.; Eder 2016, Gladwin et al. 2015, S. 14f.

⁴⁹ Vgl. Strosetzki 1991, S. 93; Ehrlicher 2008b, S. 42ff. Blasco 2007, S. XVII. Gómez Canseco 2008, S. 33ff. zur These des Stilbruchs ab Kap. 60 und Coufal / Juola 2010, S. 2 zur These, dass 69 von 74 Kapiteln nicht aus der Feder von Cervantes stammen.

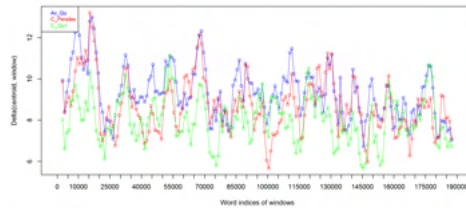


Abb. 16: Rolling Delta, Windowsize 5.000 (Oben: Av_Qu = Quijote II-A; C_Qu1 = Quijote I; Mitte: + C_Persiles; Unten: + Quev_Buscon). [Rißker-Pipka 2016]

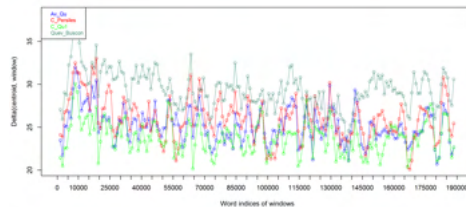


Abb. 17: Rolling Delta, Windowsize 5.000 (Oben: Av_Qu = Quijote II-A; C_Qu1 = Quijote I; Mitte: + C_Persiles; Unten: + Quev_Buscon). [Rißker-Pipka 2016]

Es gibt keinen sogenannten »shift« im Stil des Quijote II, der sich beispielsweise mit dem Quijote II-A abwechseln würde. In der Grafik sehen wir nur die im Verhältnis zum secondary set (Quijote II) getesteten Werke. Ihre stilistische Nähe zum fraglichen Quijote II zeigt sich durch die Nähe der Linie zur x-Achse. Weiter unten angesiedelte Texte sind dem fraglichen Text also am nächsten. Nun sieht man hier erstens, dass die beiden Texte von Cervantes seinem Quijote II doch noch etwas näher sind als der Quijote II-A und dass die Linien quasi parallel verlaufen und sich kaum überschneiden. Letzteres hätte für eine kollaborative Autorschaft gesprochen, weil dann ein anderer Autor an dieser Stelle des Buches sozusagen übernommen und für den ursprünglichen Autor ein Stück eingefügt hätte, um dann später wieder aus dem Buch zu verschwinden (erneute Linienkreuzung). Dennoch sieht man auch wie nahe sich die beiden Linien von Quijote I und Quijote II-A kommen und der Persiles von Cervantes zum Teil stilistisch näher am Quijote II ist als der Quijote I. Tatsächlich einschätzen kann man die stilistische Nähe zwischen den Texten von Cervantes und demjenigen von Avellaneda aber erst, wenn man sich den Vergleich zu Quevedos Buscón ansieht. Diese Linie (Abbildung 17) weicht deutlich von den unteren drei ab und hat, wie erwartet, deutlich mehr Abstand zum Quijote II als die anderen drei Texte.

7. Schlussfolgerungen: Wie können die Ergebnisse der stilometrischen Untersuchung in die Fachdiskussion rückgeführt werden?

Es konnte gezeigt werden, dass mit relativ einfachen Mitteln wie der Verwendung des stylo Packages für R die stilometrische Methode angewendet und verifiziert werden konnte. Wichtig ist es im Methodenvergleich die zahlreichen pseudostatistischen Autorschaftsattributionen zu widerlegen, damit nicht jegliche Form von scheinbar quantitativer und digitaler Methodik als unseriös abgestempelt wird.

Abseits der eigentlichen Autorschaftsfrage konnten zudem Themen diskutiert werden, die stilistische Ähnlichkeiten und literaturgeschichtliche Einteilungen betreffen und damit für die Literaturwissenschaft insgesamt interessant sind. Wie vielfältig neue Perspektiven geschaffen werden, zeigen die verschiedenen Experimente. Dabei geht es eben nicht darum, zu bestätigen, was die Literaturgeschichte schon seit Jahren proklamiert, sondern zum einen festgefahrene Diskussionen wie diejenige um die stilistischen Merkmale und die damit verbundene Debatte um literarische Qualität des Nachahmers von Cervantes im Vergleich zum Original aufzubrechen und zum anderen die technisch-mathematische Diskussion der Stilometrie an die Lektüre des Textes sinnvoll wieder anzubinden. Auch der editionsgeschichtliche und zeitgeschichtliche Hintergrund trägt dazu bei, gleichzeitig die Textgrundlage generell (unabhängig von der Qualität der digitalen Edition) und die Lektüregewohnheiten der Zeit einzubeziehen. Der Autor ist im Siglo de Oro noch eine junge Erfindung und Spielball des geistreichen Scherzes im Quijote und anderen Werken. Gerade aus diesem Grund müssen auch abseitig erscheinende Überlegungen wie eine Zusammenarbeit von Cervantes und Avellaneda in Betracht gezogen werden, ohne dies als bewiesenen Fakt zu präsentieren, wie die Ergebnisse in den zahlreichen Autorschaftsattributionen angepriesen werden. Dem Versuch einer kollaborativen Arbeit widersprechen die hier durchgeführten Experimente mit dem Rolling Delta. Wenn Cervantes und Avellaneda demnach stilistisch dennoch sehr ähnlich sind, bleibt

die Frage offen, ob sich Cervantes nicht selbst verdoppelt hat und eine geniale „burla“ (einen karnevalistischen Spaß) mit dem Publikum treibt. Stilometrie bleibt zum überwiegenden Teil eine Interpretationsaufgabe, die wie jede hermeneutische Analyse mithilfe von verschiedenen Indizien argumentiert, aber nur überzeugen und nicht letztgültig beweisen kann.

Bibliographische Angaben

Mercedes Agulló y Cobo: A vueltas con el autor del «Lazarillo». Madrid 2010. [\[Nachweis im GVK\]](#)

Mercedes Agulló y Cobo: Mercedes. A vueltas con el autor del Lazarillo. Un par de vueltas más. PDF. [\[online\]](#) In: Lemir 15 (2011), S. 217–34. [\[online\]](#)

David Alvarez Roblin: De l'imposture à la création. Le Guzmán et le Quichote apocryphes. Madrid 2014. [\[Nachweis im GVK\]](#)

Alonso Fernández de Avellaneda: Segundo tomo del ingenioso hidalgo Don Quijote de la Mancha. Primärtext von 1614. In: Alonso Fernández de Avellaneda: El Quijote apócrifo. Hg. von Alfredo Rodríguez López-Vázquez. Madrid 2011, S. 101–556. [\[Nachweis im GVK\]](#)

Roland Barthes: La mort de l'auteur. In: Manteia 1968, S. 12–17. [\[Nachweis im GVK\]](#)

Javier Blasco: Introducción. In: Segundo tomo del ingenioso hidalgo Don Quijote de la Mancha. Hg. von Javier Blasco. Madrid 2007, S. XIII–LXXXIX. [\[Nachweis im GVK\]](#)

Javier Blasco / Cristina Ruiz Urbón: Evaluación y cuantificación de algunas técnicas de 'atribución de autoría' en textos españoles. DOI: [10.24197/cel.0.2009.27-47](#) In: Castilla: Estudios de Literatura 0 (2009), S. 27–47. DOI: [10.24197/cel.0.2009](#)

José Calvo Tello / Christof Schöch / Nanette Rißler-Pipka / Tobias Kraft: Humanidades Digitales y estudios hispánicos en Alemania. In: Voz y Letra 26 (2015), H. 1, S. 45–61. [\[Nachweis im GVK\]](#)

José Calvo Tello: Entendiendo Delta desde las Humanidades. [\[online\]](#) In: Caracteres: estudios culturales y críticos de la esfera digital 5 (2016), H. 1, S. 140–176. [\[online\]](#)

Miguel de Cervantes [1605]: El Ingenioso Hidalgo Don Quijote de la Mancha. Hg. von John Jay Allen. 2 Bde. 31. Auflage. Madrid 2011. Bd. I. Siehe auch [\[Nachweis im GVK\]](#) oder [\[Nachweis im GVK\]](#)

Miguel de Cervantes [1615]: El Ingenioso Hidalgo Don Quijote de la Mancha. Hg. von John Jay Allen. 2 Bde. 25. Auflage. Madrid 2005. Bd. II. [\[Nachweis im GVK\]](#)

Christopher Coufal / Patrick Juola: Authorship Discontinuities of El Ingenioso Hidalgo don Quijote de la Mancha as detected by Mixture-of-Experts. [\[online\]](#) In: Digital Humanities 2010. Conference abstracts King's College London. (DH 2010, London, 07.-10.07.2010) London 2010. [\[online\]](#) [\[Nachweis im GVK\]](#)

Hugh Craig / Maciej Eder / Mike Kestemont / Jan Rybicki / Christof Schöch: Validating Computational Stylistics in Literary Interpretation. [\[online\]](#) In: Digital Humanities 2014: Book of Abstracts. (DH 2014, Lausanne, 08-12.07.2014) Lausanne 2014, S. 136–39. PDF. [\[Nachweis im GVK\]](#)

Mariano Delgado: Spanische Inquisition und Buchzensur. In: Stimmen der Zeit 224 (2006), H. 7, S. 461–474. [\[Nachweis im GVK\]](#)

Maciej Eder: Does size matter? Authorship attribution, short samples, big problem. [\[online\]](#) In: Digital Humanities 2010. Conference Abstracts. (DH 2010, London, 07.10.07.2010) London 2010, S. 132–135. PDF. [\[online\]](#) [\[Nachweis im GVK\]](#)

Maciej Eder: Mind Your Corpus: systematic errors in authorship attribution. DOI: [10.1093/llc/fqt039](#) In: Literary and Linguistic Computing 28 (2013), H. 4, S. 603–614. [\[online\]](#)

Maciej Eder / Mike Kestemont / Jan Rybicki: Stylometry with R: a suite of tools. [\[online\]](#) In: Digital Humanities 2013. Conference Abstracts. (DH 2013, Lincoln, NE., 16.-19.07.2013) Lincoln, NE. 2013, S. 487–89. PDF. [\[online\]](#) [\[Nachweis im GVK\]](#)

Maciej Eder: Rolling stylometry. DOI: [http://dx.doi.org/10.1093/llc/fqv010](#) In: Digital Scholarship in the Humanities 31 (2016), H. 3, S. 457–469. [\[online\]](#)

Maciej Eder / Karina van Dalen-Oskam / Fotis Jannidis / Jan Rybicki / Christof Schöch (2016a): Literary Concepts: The Past and the Future. [\[online\]](#) In: Digital Humanities 2016. Conference Abstracts. (DH 2016, Krakau, 11.-16.07.2016) Krakau 2016, S. 63–65. PDF. [\[online\]](#)

Maciej Eder / Mike Kestemont / Jan Rybicki (2016b): Stylometry with R: A package for computational text analysis. PDF. [\[online\]](#) In: R Journal (2016), H. 1, S. 107–121. [\[online\]](#)

Maciej Eder / Jan Rybicki / Mike Kestemont (2017): 'Stylo': a package for stylometric analyses. PDF. [\[online\]](#) In: Computational Stylistics Group. HOWTO. Krakau/Antwerpen 2017. [\[online\]](#)

Hanno Ehrlicher (2008a): »Der andere Autor im eigenen Werk. Medialisierte Autorschaft bei Mateo Alemán und Miguel de Cervantes«. In: Automedialität. Hg. von Jörg Dünne / Christian Moser. München 2008, S. 27–52. [\[Nachweis im GVK\]](#)

Hanno Ehrlicher (2008b): »Das aufgegebene Anonymat. Kritische Anmerkungen zu einer philologischen Kanonrevision aus aktuellem Anlass«. [\[online\]](#) In: PhiN. Philologie im Netz 46 (2008), S. 1–13. [\[online\]](#)

Hanno Ehrlicher: »La artificiosidad aumentada. Avellaneda como catalizador de la narrativa del Quijote. In: El otro Don Quijote. La continuación de Fernández de Avellaneda y sus efectos. Hg. von Hanno Ehrlicher. Augsburg 2016, S. 55–74. URN: [urn:nbn:de:bvb:384-opus4-37049](#)

Stefan Evert / Fotis Jannidis / Friedrich Michael Dimpel / Christof Schöch / Steffen Pielström / Thorsten Vitt / Isabella Reger / Andreas Büttner / Thomas Proisl: »Delta« in der stilometrischen Autorschaftsattributions. [\[online\]](#) In: Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. DHd 2016. Konferenzabstracts. Hg. von Elisabeth Burr. (DHd 2016, Leipzig, 07.-12.03.2016) Duisburg 2016, S. 61–74. PDF. [\[online\]](#) [\[Nachweis im GVK\]](#)

José Manuel Fradejas Rueda: Lingüística forense y crítica textual. El caso Ayala–Cervantes. DOI: [10.5944/signa.vol25.2016.16927](#) In: Signa: Revista de la Asociación Española de Semiótica 25 (2016), S. 193–220. DOI: [10.5944/signa.vol25.2016](#)

Alexander A. G. Gladwin / Matthew J. Lavin / Daniel M. Look: Stylometry and Collaborative Authorship: Eddy, Lovecraft, and »The Loved Dead«. DOI: [10.1093/llc/fqv026](#) In: Digital Scholarship in the Humanities 32 (2015), H. 1, S. 123–140. [\[online\]](#)

Luis Gómez Canseco: 1614: Cervantes escribe otro »Quijote«. PDF. [\[online\]](#) In: Tus obras los rincones de la tierra descubren: actas del VI congreso internacional de la Asociación de Cervantistas. Hg. von Alexia Dotras Bravo / José Manuel Lucía Megías / Elisabet Magro García / José Montero Reguera. (Congreso internacional de la Asociación de Cervantistas: 6, Alcalá de Henares, 13.-16.12.2006) Alcalá de Henares 2008, S. 29–44. [\[online\]](#) [\[Nachweis im GVK\]](#)

Luis Gómez Canseco: Introducción a Fernández de Avellaneda. In: Alonso Fernández de Avellaneda: El ingenioso hidalgo Don Quijote de la Mancha. Hg. von Luis Gómez Canseco. Madrid 2014, S. 9–123. [\[Nachweis im GVK\]](#)

Luis Gómez Canseco: Avellaneda en la imprenta: Tarragona. In: El otro Don Quijote. La continuación de Fernández de Avellaneda y sus efectos. Hg. von Hanno Ehrlicher. Augsburg 2016, S. 11–26. URN: [urn:nbn:de:bvb:384-opus4-37049](#)

Julia Berenike Herrmann / Karina van Dalen-Oskam / Christof Schöch: Revisiting Style, a Key Concept in Literary Studies. URN: [urn:nbn:de:0222-003072](#) In: Journal of Literary Theory 9 (2015), H. 1, S. 25–52. [\[online\]](#) [\[Nachweis im GVK\]](#)

Fotis Jannidis: Der nützliche Autor. Möglichkeiten eines Begriffs zwischen Text und historischem Kontext. [\[Nachweis im GVK\]](#) In: Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs. Hg. von Fotis Jannidis / Gerhard Lauer / Matias Martinez / Simone Winko. Berlin 1999, S. 353–390. [\[Nachweis im GVK\]](#)

Fotis Jannidis / Steffen Pielström / Christof Schöch / Thorsten Vitt: Improving Burrows' Delta – An empirical evaluation of text distance measures. [\[online\]](#) In: DH 2015. Global Digital Humanities. Conference Abstracts. Hg. von University of Western Sydney / Universität Hamburg. (DH 2015, Sydney, 29.06.-03.-07.2015) Sydney 2015. [\[online\]](#)

Alfonso Martín Jiménez: Hacen falta cuatro siglos para entender a Cervantes. Por el expresidario pinillo, natural vecino del mundo. Valladolid 2016. [\[online\]](#)

Freddy López Quintero: Diferencias internas en Don Quijote. Cambios en proporciones y cambios estructurales. PDF. [\[online\]](#) In: Lemir 15 (2011), S. 259–270. [\[online\]](#)

Alfredo Rodríguez López-Vázquez (2011a): Alonso Fernández de Avellaneda: El Quijote apócrifo. Hg. von Alfredo Rodríguez López-Vázquez. Madrid 2011. [\[Nachweis im GVK\]](#)

Alfredo Rodríguez López-Vázquez (2011b): Introducción. In: Alonso Fernández de Avellaneda: El Quijote apócrifo. Hg. von Alfredo Rodríguez López-Vázquez. Madrid 2011, S. 13–84. [\[Nachweis im GVK\]](#)

José Luis Madrigal: Cervantes De Salazar, Autor Del Lazarillo. In: Artifara 2 (2003). [\[online\]](#)

José Luis Madrigal: Tirso, Lope y el Quijote de Avellaneda. PDF. [\[online\]](#) In: Lemir 13 (2009), S. 191–250. [\[online\]](#)

José Luis Madrigal: Notas Sobre La Autoría Del Lazarillo. PDF. [\[online\]](#) In: Lemir 12 (2008), S. 137–236. [\[online\]](#)

José Luis Madrigal: De nombres y lugares: el corpus del licenciado Arce de Otálora. PDF. [\[online\]](#) In: Lemir 18 (2014), S. 89–118. [\[online\]](#)

Alfonso de Valdés: La vida de Lazarillo de Tormes, y de sus fortunas y adversidades. Introducción de Rosa Navarro Durán. Ed. y notas de Milagros Rodríguez Cáceres. Barcelona 2003. S. 9-64. [\[Nachweis im GVK\]](#)

The R Project for Statistical Computing. Hg. von R Foundation for Statistical Computing. Wien 2015. [\[online\]](#)

Francisco Rico: Introducción. In: Lazarillo de Tormes. Hg. von Francisco Rico. 21. Auflage. Madrid 2010, S. 13–127. Siehe auch [\[Nachweis im GVK\]](#) oder [\[Nachweis im GVK\]](#)

Nanette Reißler-Pipka: Digital Humanities Und Die Romanische Literaturwissenschaft Der Autorschaftsstreit um den Lazarillo de tormes. In: Romanische Forschungen 128 (2016), H. 3, S. 316–342. DOI: [/10.3196/003581216819562166](#)

Javier de la Rosa / Juan Luis Suárez: The Life of Lazarillo de Tormes and of His Machine Learning Adversities. PDF. [\[online\]](#) In: Lemir 20 (2016), S. 373–438. [\[online\]](#)

Javier de la Rosa Pérez: Making Machines Learn. Applications of Cultural Analytics to the Humanities. Electronic Thesis and Dissertation Repository. Paper 3486. London, Canada 2016. [\[online\]](#)

Matthias Schaffrck / Marcus Willand: Theorien und Praktiken der Autorschaft. Berlin 2014. [\[Nachweis im GVK\]](#)

Leo Spitzer: »Note on the Poetic and the Empirical « in Medieval Authors», In: Leo Spitzer: Romanische Literaturstudien (1935-1956). Tübingen 1959, S. 100–112. [\[Nachweis im GVK\]](#)

Christoph Strosetzki: Der Roman im Siglo de Oro. In: Geschichte der spanischen Literatur. Hg. von Christoph Strosetzki. Tübingen 1991, S. 84–118. [\[Nachweis im GVK\]](#)

Enrique Suárez Figaredo: Sobre la atribución del Quijote apócrifo a José de Villaviciosa. PDF. [\[online\]](#) In: Lemir 15 (2011), S. 135–146. [\[online\]](#)

Alonso Fernández de Avellaneda: El Quijote apócrifo. Hg. von Enrique Suárez Figaredo. PDF. [\[online\]](#) In: Lemir 18 (2014), S. 1–312. [\[online\]](#)

Gabriel Viehhauser: Digitale Gattungsgeschichten. Minnesang zwischen generischer Konstanz und Wende. DOI: [10.17175/2017_003](#) In: Zeitschrift für digitale Geisteswissenschaften (2017). DOI: [10.17175/2017_003](#)

Weiterführende Literatur

Mateo Alemán: Guzmán de Alfarache. Segunda Parte. Alicante 2000. [\[online\]](#)

Alonso Fernández de Avellaneda: Segundo tomo del ingenioso hidalgo don Quijote de la Mancha. Hg. von Luis Gómez Canseco. Madrid 2014. [\[Nachweis im GVK\]](#)

Mathieu Bastian / Sebastien Heymann / Mathieu Jacomy: Gephi: an open source software for exploring and manipulating networks. PDF. [\[online\]](#) In: Proceedings of the Third International AAAI Conference on Weblogs and Social Media. Hg. von Association for the Advancement of Artificial Intelligence. (AAAI Conference on Weblogs and Social Media: 3, San Jose, CA., 17.-20.05.2009) Menlo Park, CA. 2009. [\[online\]](#) [\[Nachweis im GVK\]](#)

Javier Blasco: La lengua de Avellaneda en el espejo de »La pícara Justina«. PDF. [\[online\]](#) In: Boletín de la Real Academia Española 85 (291–292) (2005), S. 53–109. [\[online\]](#) [\[Nachweis im GVK\]](#)

Roger Chartier: Materialidad del texto, textualidad del libro. [\[online\]](#) In: Orbis Tertius 11 (2006), H. 12. [\[online\]](#)

Antonio Cruz Casado: Revisión de una hipótesis. Juan Valladares de Valdelomar; autor del »Quijote«. PDF. [\[online\]](#) In: Tus obras los rincones de la tierra descubren. Actas del VI congreso internacional de la Asociación de Cervantistas. Hg. von Alexia Dotras Bravo / José Manuel Lucía Megías / Elisabet Magro García / José Montero Reguera. (Congreso Internacional de la Asociación de Cervantistas: 6, Alcalá de Henares, 13.-16.12.2006) Alcalá de Henares 2008, S. 269-283. [\[online\]](#) [\[Nachweis im GVK\]](#)

Alfonso Martín Jiménez: Cotejo por medios informáticos de la Vida de Pasamonte y el Quijote de Avellaneda. PDF. [\[online\]](#) In: Etiópicas 3 (2007), S. 69-131. [\[online\]](#)

Jan Rybicki / Maciej Eder: Deeper Delta across genres and languages: do we really need the most frequent words?. DOI: [10.1093/llc/fqr031](#) In: Literary and Linguistic Computing 26 (2011), H. 3, S. 315–321. [\[online\]](#) [\[Nachweis im GVK\]](#)

Nanette Reißler-Pipka: Avellaneda y los problemas de la identificación del autor. Propuestas para una investigación con nuevas herramientas digitales. In: El otro Quijote. La continuación de Avellaneda y sus efectos. Hg. von Hanno Ehrlicher. Augsburg 2016, S. 27–51. [\[online\]](#)

Antonio Sánchez Portero: El autor del »Quijote« de Avellaneda es Pedro Liñán de Ríaza, poeta de Calatayud. Edición digital por cortesía del autor para la Biblioteca Virtual Miguel de Cervantes. Alicante 2006. [\[online\]](#)

Abbildungslegenden und -nachweise

Abb. 1: Visualisierung in Jiménez 2016, S. 34.

Abb. 2: Fensteransicht der Suchabfrage »Lazarillo« in TextGrid Repository.

Abb. 3: Dendrogramm, cosine Delta, 2.600 MFW (von 200-4800 MFW alle konstant). [Rißker-Pipka 2016]

Abb. 4: Dendrogramm: cosine Delta, 4.600 MFW (16 Texte von Cervantes, 2 von Quevedo, 2 von Alemán, 3 von Salas Barbadillo, 2 von Suárez Figueroa, 3 von Castillo Solórzano, 6 von Cespedes y Meneses = 34 Texte). [Rißker-Pipka 2016]

Abb. 5: Dendrogramm, cosine Delta, 4.600 MFW (alle Texte mind. 100.000 Wörter: 11 Texte). [Rißker-Pipka 2016]

Abb. 6: Dendrogramm, cosine Delta, 4.600 MFW (39 Texte). [Rißler-Pipka 2016]

Abb. 7: Dendrogramm, cosine Delta, 4.500 MFW (44 Texte). [Rißker-Pipka 2016]

Abb. 8: Dendrogramm, cosine Delta, 5.000 MFW (49 Texte, abgekürzte Titel). [Rißker-Pipka 2016]

Abb. 9: Netzwerkvisualisierung mit Gephi (vgl. Bastian et al. 2009). [Rißker-Pipka 2016]

Abb. 10: Dendrogramme mit cosine Delta, 500 und 600 MFW: nur bis 500 MFW wird der *Quijote II-A* nicht direkt mit den beiden *Quijotes* von Cervantes gruppiert. [Rißker-Pipka 2016]

Abb. 11: Dendrogramme mit cosine Delta, 500 und 600 MFW: nur bis 500 MFW wird der *Quijote II-A* nicht direkt mit den beiden *Quijotes* von Cervantes gruppiert. [Rißker-Pipka 2016]

Abb. 12: Bootstrap Consensus Tree, cosine Delta, 1.000-4.000 MFW, Culled 20%. [Rißker-Pipka 2016]

Abb. 13: Dendrogramm, cosine Delta, 5.000 MFW (49 Texte, abgekürzte Titel). [Rißker-Pipka 2016]

Abb. 14: Dendrogramm, cosine Delta, 2.200 MFW (Kapitel von *Quijote II-A* und *Quijote I*). [Rißler-Pipka 2016]

Abb. 15: Rolling Delta, Windowsize 5.000 (Oben: Av_Qu = Quijote II-A; C_Qu1 = Quijote I; Mitte: + C_Persiles; Unten: + Quev_Buscon). [Rißler-Pipka 2016]

Abb. 16: Rolling Delta, Windowsize 5.000 (Oben: Av_Qu = Quijote II-A; C_Qu1 = Quijote I; Mitte: + C_Persiles; Unten: + Quev_Buscon). [Rißler-Pipka 2016]

Abb. 17: Rolling Delta, Windowsize 5.000 (Oben: Av_Qu = Quijote II-A; C_Qu1 = Quijote I; Mitte: + C_Persiles; Unten: + Quev_Buscon). [Rißler-Pipka 2016]