

PROBLEMATISCHE AUTORSCHAFT.

ZU ERSTEN ANWENDUNGSVERSUCHEN STILOMETRISCHER METHODEN IM RAHMEN DER EDITION DER DEUTSCHEN POLITISCHEN SCHRIFTEN ZINGGREFS

FRIEDRICH MICHAEL DIMPEL, VICTORIA GUTSCHE, RENÉ WUNDKE

1 Einleitung

Sicherlich kann mit Verve darüber gestritten werden, was die Digitale Editionswissenschaft leistet und wo ihre Grenzen liegen. Man könnte sogar behaupten, dass bei einer so jungen Wissenschaft der Streit um den Gegenstand essentiell für ihr Überleben und für ihre Verankerung im akademischen Feld ist. So ist es zwar hilfreich, den Geltungsbereich nicht allzu groß zu setzen, eine Einschränkung möglicher Einsatzmöglichkeiten der Digitalen Editionswissenschaft erscheint indes ebenso unangebracht.

Der vorliegende Beitrag versucht, eine solche Einsatzmöglichkeit am Rande der üblicherweise im Fokus liegenden Gegenstände zu beschreiben. Denn hier sollen nicht eine Online-Edition und ihre Markierungs- und Darstellungsverfahren vorgestellt werden, sondern die Leistungsmöglichkeit der Digitalen Editionswissenschaft im Bereich der Buchedition exemplarisch angezeigt werden. Das heißt, im Folgenden soll erläutert werden, wie digital erhobene Daten, die im Sinne der DFG-Richtlinien auch digital gespeichert werden,¹ zur Begründung und Gestaltung einer großen, in Buchform vorgelegten Werkedition der Frühen Neuzeit eingesetzt werden und Folgeuntersuchungen inspirieren können. Insofern versteht sich der vorliegende Beitrag auch als Vorschlag, die Digitale Editionswissenschaft nicht auf die Produktion von Online-Editionen zu begrenzen.²

1 Vgl. die Leitlinien zum Umgang mit Forschungsdaten der Deutschen Forschungsgemeinschaft, in: http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf [Abruf am 04.05.2016].

2 Vgl. auch die grundsätzlichen Überlegungen zum Verhältnis von Editorik und digitalen Methoden in: Daniel Apollon, Claire Bélisle, Philippe Régnier: Introduction: As Texts Become Digital, in: Digital Critical Editions, hg. von Daniel Apollon, Claire Bélisle, Philippe Régnier, Urbana, Chicago, Springfield 2014 (Topics in the Digital Humanities), S. 1-32. Dass digitale Editionswissenschaften nicht allein im Erstellen von Online-Editionen aufgehen, zeigt beispielhaft auch der Beitrag von Paolo Trovato: Everything You Always Wanted to Know about Lachmann's Method. A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Strukturalism, Cladistics, and Copy-Text, Padova 2014, der den Einsatz digitaler Methoden im Rahmen von Editionen beleuchtet.

Zentral für jede Werkedition ist zunächst die Klärung des jeweiligen auktorialen Status der in Frage kommenden Texte, entscheidet dieser doch über die Aufnahme der Texte und damit über die Gestalt der Edition. Während dies recht unproblematisch erscheint, wenn der Verfasser sich auf dem Titelblatt oder an anderem Ort zu erkennen gibt oder sein Pseudonym leicht auflösbar ist, so gestaltet sich die Lage bei anonym erschienenen Schriften mitunter durchaus komplex. Im Folgenden wird der Versuch vorgestellt, anhand der politischen Schriften von Julius Wilhelm Zingref (1591–1635)³ diese zentrale Frage nach der Autorschaft in Ergänzung von konventionellen philologischen Argumenten mittels digitaler Methoden, genauer der Stilometrie, zu untersuchen. Auf diese Weise soll gezeigt werden, wie Editionen von der Kombination traditioneller philologischer Untersuchungen mit Methoden der Digital Humanities profitieren können und in welchem Rahmen digitale Methoden konkret im Kontext von Print-Editionen eingesetzt werden können.⁴ So können stilometrische Untersuchungen nicht nur einen Beitrag zur Frage nach der Aufnahme bestimmter Texte leisten; darüber hinaus sind auch Erkenntnisse etwa zu Textsorten-Ordnungen zu erwarten, die ihrerseits in Einleitung und Kommentar einfließen.

Die politischen Schriften Zingrefs erscheinen unter dieser Perspektive besonders ergiebig. So ist nicht nur die Autorschaft der anonym erschienenen Schriften umstritten, sondern sie bilden darüber hinaus ein Ensemble unterschiedlicher literarischer Kleinformen: Neben der ›Haidelberga‹, einem Prosatext zu einem großformatigen Kupferstich, finden sich durchaus heterogene Flugschriften, darunter die Hybridform ›Quotlibetisches Weltkefig‹, wie auch das Gedicht ›Vermahnung zur Dapfferkeit‹. Darüber hinaus ist nur für den letzten Text die Autorschaft Zingrefs unbestritten, bei allen anderen Texten kann die Verfasserschaft nur mit guten Gründen angenommen werden. Als problematisch erweisen sich dabei insbesondere die ›Newen Zeitungen‹:⁵ Im Medium der Flugschrift werden hier mit der Konjunktion ›daß‹ eingeleitete Feststellungen über Politisches, Konfessionelles, Juristisches etc. aneinandergereiht. Diese recht simplen Aneinanderreihungen sind es, die zunächst aus philologischer Sicht gegen eine Verfasserschaft Zingrefs sprechen.

3 Historisch-kritische Edition und Kommentar der Werke von Julius Wilhelm Zingref, Bd. 5.2: Deutsche politische Schriften, hg. von Werner Wilhelm Schnabel, Dirk Niefanger und Theodor Verweyen. Im Rahmen der Gesamtausgabe sind bisher erschienen: Bd. 2: Emblemata ethico-politica, hg. von Dieter Mertens und Theodor Verweyen. 2 Bde., Tübingen 1993; Bd. 3: Facetiae Pennalium, hg. von Dieter Mertens und Theodor Verweyen. Tübingen 1978; Bd. 4: Apophthegmata teutsch, hg. von Theodor Verweyen, Dieter Mertens und Werner Wilhelm Schnabel, 2 Bde., Berlin/Boston 2011.

4 Eine Online-Edition der politischen Schriften Zingrefs ist nicht vorgesehen, um die Kontinuität der Werkausgabe zu erhalten, da die bisher erschienenen Bände ausschließlich als Print-Edition veröffentlicht wurden. Zudem wird eine Online-Edition von Verlag, Reihen- und Werkherausgeber nicht gewünscht.

5 *Newe Zeitungen*, Von unterschiedlichen Orten: Das ist, Die alte Warheit, mit einem neuen Titul, [o.O.] 1619; *Continuatio der Newen Zeitungen Von unterschiedlichen orten: Das ist, Die alte Warheit mit einem neuen Titul. Vermehrt vnd auch verbessert*, [o.O.] 1620; *Warhafft Newe Zeitungen/ Von vnterschiedlichen Orten vnd Landen. Das ist: Die alte Warheit mit eim neuen Titul. Vermehrt und auch verbessert*, [o.O.] 1620; *Noua Nova Antiqua Continuationis der Newen zeitungen, Von vnterschiedlichen Orten: Das ist: Die alte Warheit mit eim neuen Titul. Vermehrt vnd auch verbessert. Editio tertia*, [o.O.] 1621.

Ähnliche Formulierungen sowie Parallelen inhaltlicher Art zum ›Weltkefig‹, eine ähnliche Anlage der ›Newen Zeitungen‹ wie die ›Apophtegmata‹ und ›Facetien‹ sowie eine eher propfälzische und proböhmische Position lassen andererseits eine Beteiligung Zinggrefs wahrscheinlich erscheinen.⁶ Vor diesem Hintergrund sowie der problematischen Textsortenzuordnung bieten sich die ›Newen Zeitungen‹ als Anwendungsfall für stilometrische Methoden bei der Autorschaftszuweisung frühneuzeitlicher Texte an. Vor dem Versuch, Thesen zur Autorschaft der ›Newen Zeitungen‹ mittels stilometrischer Methoden abzuweisen oder zu erhärten, ist jedoch zunächst zu ermitteln, ob sich erzählende Texte des 17. Jahrhunderts für stilometrische Analysen überhaupt eignen. So muss geprüft werden, ob man bei der deutschsprachigen Literatur der Frühen Neuzeit überhaupt von einem Autorstil ausgehen kann oder ob nicht vielmehr gattungs- und textsortenspezifische Signale die Kategorie Autor überlagern. Darüber hinaus sind weitere Fragen zu klären, wie der Einfluss unterschiedlicher Schreibungen oder anderssprachlicher Passagen in den Texten.

2 Zur Anwendung stilometrischer Verfahren bei Prosatexten aus dem 17. Jahrhundert

Der Begriff ›Stilometrie‹ bezeichnet zunächst die quantitative Erhebung stilistischer Merkmale, die der Klassifikation von Texten dient.⁷ Eingesetzt werden stilometrische Verfahren insbesondere bei der Attribuierung von Autorschaft, aber auch im Rahmen der forensischen Linguistik, der chronologischen Einordnung von Texten ins Œuvre eines Autors sowie bei Fragen nach dem Verhältnis von Autor, Gattung und Literaturgeschichte.⁸ Erhoben werden dabei bestimmte Textmerkmale, etwa die Häufigkeit von Most-Frequent-Words oder Funktionswörtern, Wort- und Satzlängen, Häufigkeit von Trigrammen oder syntaktische Merkmale,⁹ um die relative Nähe oder Distanz von Texten zu bestimmen und die Texte auf dieser Grundlage hinsichtlich Autor, Gattung und Geschlecht des Autors oder auch hinsichtlich der Epochenzugehörigkeit zu klassifizieren.

Da es sich um ein sog. unüberwachtes Klassifizierungsverfahren handelt, sind die Zielkategorien oder Gruppen nicht vorgegeben, sondern ergeben sich aus der jeweils

6 Anders Julius Opel, Adolf Cohn: Der dreißigjährige Krieg. Eine Sammlung von historischen Gedichten und Prosadarstellungen, Halle 1862, S. 483–486, die den Straßburger Ratskonsulenten Theophil Dachtler als möglichen Autor der ›Newen Zeitungen‹ nennen.

7 Vgl. Fotis Jannidis: Methoden der computergestützten Textanalyse, in: Methoden der literatur- und kulturwissenschaftlichen Textanalyse. Ansätze – Grundlagen – Modellanalysen, hg. von Vera Nünning, Ansgar Nünning, Stuttgart 2010, S. 109–132.

8 Christof Schöch: Corneille, Molière et les autres. Stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik, in: PhiN-Beiheft 7 (2014), S. 130–157, hier S. 133. Vgl. einflussreich auch Fotis Jannidis: Der Autor ganz nah – Autorstil in Stilistik und Stilometrie, in: Theorien und Praktiken der Autorschaft, hg. von Matthias Schaffrick, Marcus Willand, Berlin 2014, S. 169–195.

9 Vgl. Jannidis [Anm. 7], S. 117.

festgestellten Ähnlichkeit der Texte. Dass sich die Autorkategorie nicht immer isolieren lässt, ist gerade eines der Probleme – oder auch eines der Ergebnisse – der aktuellen methodischen Debatten in diesem Feld.¹⁰

Vor diesem Hintergrund ging es zunächst darum, zu überprüfen, ob sich die Autorkategorie bei frühneuzeitlichen deutschen Texten isolieren lässt und damit die Voraussetzung für die Autorschaftszuweisung gegeben ist.¹¹

Burrows' Delta ist mittlerweile als Autorschaftstest etabliert.¹² Validierungsstudien führen vor, dass Delta für germanische Sprachen sehr gute Ergebnisse liefert.¹³ Erzielt wurden diese guten Ergebnisse allerdings mit normalisierten Texten bzw. mit Texten aus der Moderne, bei denen Abweichungen von einer Standardgraphie nicht zu einem Problem werden. Im Falle der frühneuzeitlichen Prosatexte liegt jedoch eine zum Teil erhebliche graphematische Varianz vor. Beispielhaft sei auf die unterschiedliche Schreibung von ›und‹ verwiesen: Neben ›und‹ finden sich ›vnd‹, ›vnnd‹, ›vñ‹, ›unnd‹ sowie ›uñ‹. Da ›und‹ zu den häufigsten Wörtern in deutschsprachigen Texten zählt, wirken sich Schreibungsvarianten unmittelbar auf die Ergebnisse aus, sodass die Texte dahingehend normalisiert werden müssen.¹⁴ Darüber hinaus enthalten zahlreiche Prosatexte mitunter umfangreiche Anmerkungen innerhalb der Texte, die hier verwendeten Abkürzungen können ebenfalls die Ergebnisse beeinflussen. So weist beispielsweise Johannes Praeto-

10 Schöch [Anm. 8], S. 134.

11 Christof Schöch konnte anhand einer stilometrischen Analyse von 51 französischen Dramen des 17. Jahrhunderts zeigen, dass die »Unterscheidung von Prosa und Vers offensichtlich eine massiv differenzierende Kategorie ist, die die Autorenkategorie klar überlagert.« Schöch [Anm. 8], S. 145. Daher wurden bei der Zusammenstellung des Korpus ausschließlich Prosatexte berücksichtigt. Inwiefern Schöch's Ergebnisse tatsächlich auf deutschsprachige Prosatexte des 17. Jahrhunderts übertragbar sind, gilt es jedoch in weiteren Untersuchungen zu klären.

12 John F. Burrows: ›Delta‹. A Measure of Stylistic Difference and a Guide to Likely Authorship, in: LLC 17 (2002), S. 267–287.

13 David L. Hoover: Delta Prime? in: LLC 19 (2004), S. 477–495; Shlomo Argamon: Interpreting Burrows's Delta: Geometric and Probabilistic Foundations, in: LLC 23 (2008), S. 131–147; Maciej Eder, Jan Rybicki: Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words? in: LLC 26 (2011), S. 315–321; Maciej Eder: Mind Your Corpus: systematic errors in authorship attribution, in: LLC 28 (2013a), S. 603–614; zu Support Vector Machines vgl. auch Maciej Eder: Does size matter? Authorship attribution, small samples, big problem, in: LLC Advanced Access 29, doi:10.1093/lc/fqt066 (2013b), S. 1–16, hier S. 9f. Für das Neuhochdeutsche 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. von Matt Erlin, Lynne Tatlock, New York 2014, S. 29–54; Stefan Evert, Thomas Proisl, Fotis Jannidis, Steffen Pielström, Christof Schöch, Thorsten Vitt: Towards a better understanding of Burrows's Delta in literary authorship attribution, in: North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015) doi: 10.5281/zenodo.18177 (2015) und Stefan Evert, Fotis Jannidis, Steffen Pielström, Isabella Reger, Christof Schöch, Thorsten Vitt: Burrows' Delta verstehen, in: Konferenzabstracts DHd 2016. Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma, hg. von Elisabeth Burr, Leipzig 2016, S. 62–65, diskutieren weitere Abstandsmaße, erfolgreich hat sich insbesondere der Cosinus-Delta erwiesen: Hier bleiben insbesondere bei langen Vektoren die Erkennungsquoten länger stabil. Da hier mit teils kurzen Texten und daher mit Vektoren bis 800 Most-Frequent-Words gearbeitet wird, kommt hier nur Burrows' Delta zur Anwendung.

14 Bei der Zusammenstellung des Korpus wurde auf digitale Textsammlungen (textgridrep.org, gutenberg.spiegel.de, zeno.org) zurückgegriffen.

rius' ›Anthropodemus Plutonicus‹ die Abkürzung ›lib.‹ rund 270mal, den Hinweis ›Vide‹ 310mal auf. Inwiefern dies die Ergebnisse verschlechtert, muss im Einzelnen getestet werden. Insbesondere in den Anmerkungen, im Proömium, aber auch eingeschoben in den Text finden sich darüber hinaus – zum Teil in erheblichem Umfang – fremdsprachige Passagen, in erster Linie lateinische, die ebenfalls die Ergebnisse beeinflussen können.

Vor diesem Hintergrund wurde per Skript eine Teilnormalisierung häufiger Varianten, z. B. von ›und‹, vorgenommen, Silbentrennungen wurden bereinigt und Sonderzeichen entfernt. Umfangreichere fremdsprachige Passagen (insbesondere lateinische) wurden manuell entfernt. Erste Versuche deuten an, dass diese Teilnormalisierung durchaus zu besseren Ergebnissen führen kann. So wird anstelle von Erkennungsquoten von 60–70% nunmehr eine Quote von etwas unter 90% erzielt.¹⁵

Auch wenn normalisierte Texte bei Verfahren, die auf den Most-Frequent-Words (MFWs) beruhen, stets vorzuziehen sind, hat doch Eder gezeigt, dass die Erkennungsquoten relativ stabil bleiben, selbst wenn man dem Testsample willkürlich Fehler hinzufügt: So wird die Erkennungsquote im Neuhochdeutschen bei einem zufälligen Buchstabentausch von 12% bei 100–400 MFWs kaum schlechter.¹⁶ Untersuchungen für das Mittelhochdeutsche haben ähnliche Ergebnisse erbracht. Bei einem Test – mit kleinem Korpus, daher nicht repräsentativ – lässt sich auch für nicht-normalisierte mittelhochdeutsche Texte eine Erkennungsquote von ca. 95% ermitteln.¹⁷

Im vorliegenden Fall soll zuerst geprüft werden, inwieweit Burrows' Delta für nicht-normalisierte deutsche Texte des 17. Jahrhunderts anwendbar ist. Erst dann kann erwo-gen werden, ob das Testverfahren dafür in Betracht kommt, einen Beitrag zu der Frage zu leisten, ob die ›Newen Zeitungen‹ und der ›Weltkefig‹ Zingref zuzuordnen sind.¹⁸ Da das Untersuchungskorpus relativ klein ist, hat diese Prüfung nur den Status einer explorativen Vorstudie.¹⁹

Das Korpus setzt sich zusammen aus insgesamt 29 Prosatexten des 17. und – im Falle von Wickram – des 16. Jahrhunderts, deren Autorschaft gesichert ist. Sie gehören

15 Teilweise geraten hochfrequente lateinische Wörter in die Liste der untersuchten Most-Frequent-Words, teilweise ist der Umfang lateinischer Passagen so groß, dass dadurch die Mittelwerte des deutschen Vokabulars spürbar sinken.

16 Eder [2013a, Anm. 13], S. 612f.

17 Vgl. Friedrich Michael Dimpel: Burrows' Delta im Mittelalter: Wilde Graphien und metrische Analysedaten, in: Konferenzabstracts DHd 2016. Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma, hg. von Elisabeth Burr, Leipzig 2016, S. 65–70 (auch zur Publikation in erweiterter Form vorgesehen). Hier wurde ebenfalls eine Teilnormalisierung vorgenommen, indem skriptgesteuert alle Sonderzeichen in Zeichen konvertiert wurden, die im neuhochdeutschen Alphabet enthalten sind.

18 Im Gegensatz zu den ›Newen Zeitungen‹ erscheint die Verfasserschaft Zingrefs für den ›Weltkefig‹ recht wahrscheinlich. Vgl. Emil Weller: Der Dichter Zingref als Verfasser des Welt- und Hummel-Käfigs, in: Anzeiger für Kunde der deutschen Vorzeit, N.F. 4 (1856), Sp. 297–300 und Franz Schnorr von Carolsfeld: Julius Wilhelm Zingrefs Leben und Schriften, in: Archiv für Literaturgeschichte 8 (1879), S. 1–58, 446–490.

19 Evert/Proisl/annidis/Pielström/Schöch/Vitt [Anm. 11] testen jeweils drei Texte von 25 Autoren.

durchaus unterschiedlichen Textsorten an: Satirische und biblische Romane, in sich überaus diverse Historien-, Schwank- und Exempelsammlungen, Predigtsammlungen, Narrensatiren, Apophthegmata und Fazetien.²⁰ Die Heterogenität der Textsorten ist zum einen auf die vorhandenen digitalisierten Bestände zurückzuführen,²¹ zum anderen dem Versuch geschuldet, das Verhältnis von Autorschaft und Textsorte zu prüfen. Aufgrund des kleinen Korpus sowie der nicht immer eindeutigen Textsortenzuordnung²² können hier jedoch nur erste Hinweise zum Verhältnis der Kategorien gegeben werden. So ergab eine erste Analyse der Texte mit dem ›stylo‹-Paket für die Skript-Sprache R, das von Maciej Eder, Jan Rybicki und Mike Kestemont entwickelt wurde,²³ dass die Autorkategorie anscheinend die Kategorie Textsorte überlagert, werden doch alle Texte unabhängig von der Textsorte korrekt zugeordnet. Eine Ausnahme bilden jedoch die Texte Grimmelshausens. Hier werden die kürzeren Texte (›Galgen-Männlin‹, ›Stolzer Melcher‹, ›Bart-Krieg‹) getrennt von den satirischen Romanen eingeordnet. Dies kann an den Texten selbst liegen, möglicherweise zeichnen sie sich durch einen eher am Medium Flugschrift orientierten Stil aus, oder die Textlänge bestimmt in diesem Fall das Clustering.²⁴

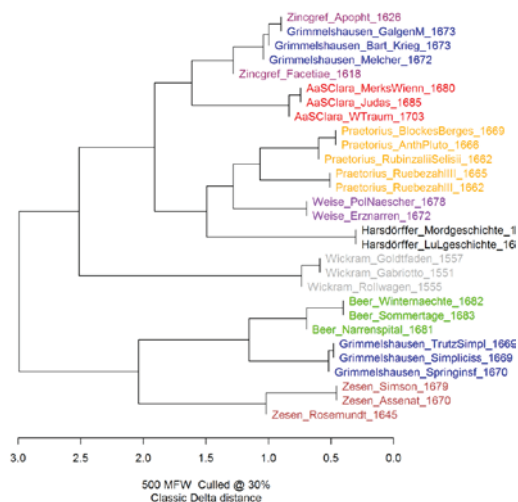
20 Abraham a Sancta Clara: Merks Wienn Das ist deß wütenden Todts ein vmbständige Beschreibung [...], Wien 1680; ders.: Judas der Erzschemel, Salzburg 1685 [ND Passau 1834]; ders.: Wunderlicher Traum von einem Narren-Nest [...], Salzburg 1703 [ND Stuttgart 1991]; Johann Beer: Das Narrenspital, [o. O.] 1681 [ND Hamburg 1957]; ders.: Teutsche Winter-Nächte [...], Nürnberg 1682 [ND Frankfurt a.M. 1963]; ders.: Die kurzweiligen Sommer-Täge, [o. O.] 1683 [ND Frankfurt a.M. 1963]; Hans Jakob Christoffel von Grimmelshausen: Trutz Simplex [...], Nürnberg 1670 [ND Stuttgart 1921]; ders.: Der seltsame Springinsfeld [...], Nürnberg 1670 [ND Stuttgart 1921]; ders.: Der abentheuerliche Simplicissimus Teutsch [...], [Nürnberg] 1669 [ND Stuttgart 1921]; ders.: Bart-Krieg [...], [o. O.] 1673 [ND Frankfurt a.M. 1997]; ders.: Simplicissimi Galgen-Männlin [...], [o. O.] [1673] [ND Frankfurt a.M. 1997]; ders.: Der stolze Melcher [...], [o. O.] [1672] [ND Frankfurt a.M. 1997]; Johannes Prätorius: Blockes-Berges Verrichtung [...], Leipzig/ Frankfurt 1669; ders.: Anthropodemus Plutonicus [...], Magdeburg 1666/67; ders.: Daemonologia Rvbinzalii Silesii [...], Leipzig 1662; ders.: Des Rübezahls Anderen [...] Theils, Leipzig 1662; ders.: Rübezahls Dritter [...] Theil, Leipzig 1665; Christian Weise: Kurtzer Bericht vom politischen Näscher [...] Leipzig 1678; ders.: Die drei ärgsten Erznarren in der ganzen Welt [...], Leipzig 1672 [ND Saale 1878]; Georg Wickram: Der Goldtfaden, Straßburg 1557 [ND Tübingen 1903]; ders.: Rollwagenbüchlein, Straßburg 1555 [ND Tübingen 1903]; ders.: Gabriotto und Reinhart, Straßburg 1551 [ND Tübingen 1903]; Philipp von Zesen: Assenat, Amsterdam 1670; ders.: Simson, Nürnberg 1679 [ND Berlin/ New York 1970]; ders.: Adriatische Rosemundt, Amsterdam 1645 [ND Halle a.d.S. 1899]; Julius Wilhelm Zinggraf: Der Teutschen scharpsinnige kluge Sprüch/ Apophthegmata genant [...], Straßburg 1628 [ND Tübingen 2008]; ders.: Facetiae Pennalium [...], [o. O.] 1618 [ND Tübingen 1978].

21 Da von jedem Autor mindestens zwei Prosatexte vorliegen müssen, reduziert sich die Zahl in Frage kommender Autoren und Texte erheblich.

22 Vgl. beispielhaft a Sancta Claras ›Judas der Erzschemel‹ [Anm. 20]. Hier handelt es sich weniger um eine Lebensbeschreibung des Judas als um eine Kompilation von kurzen Berichten über Judas Leben, Exempeln, moralisierenden Ermahnungen und Exkursen zu verschiedensten Gegenständen.

23 Maciej Eder, Mike Kestemont, Jan Rybicki: Stylometry with R: a suite of tools, in: Digital Humanities 2013: Conference Abstracts, University of Nebraska–Lincoln, NE, pp. 487–489; Maciej Eder, Jan Rybicki, Mike Kestemont: stylo R package, in: <https://sites.google.com/site/computationalstylistics/stylo> [Abruf 31.03.2016].

24 Getestet wurde mit einer Wortliste von 500 Wörtern, 30% Culling-Rate sowie dem Distanzmaß ›Classic Delta‹. Die Zuordnung blieb auch bei einer mäßigen Manipulation der Parameter (z. B. längere oder kürzere Wortliste, andere Culling-Rate) stabil.



Für den nächsten Test wurde das Korpus aufgeteilt: Neun Texte wurden als Validierungskorpus verwendet. In ein Ratekorpus wurden 20 Texte derselben Autoren gegeben (von einigen Autoren sind mehrere Texte verfügbar). In einem Perl-Skript wurde ein Delta-Test implementiert, der zu jedem Text aus dem Ratekorpus das Delta-Distanzmaß zu allen Texten im Validierungskorpus berechnet. Wenn das niedrigste Distanzmaß das Distanzmaß des richtigen Autors ist, gilt der Text als korrekt zugeordnet. Die Erkennungsquote wurde für verschiedene Parameter erhoben: in 100er Schritten von 100 bis 800 Most-Frequent-Words.

Während Burrows zunächst der Ansicht war, Delta sei ab einer Textlänge von 1.500 Wörtern anwendbar, hat Eder gezeigt, dass im Englischen erst ab 5.000 Wörtern sehr gute Ergebnisse möglich werden.²⁵ Texte, die kürzer als 5.000 Wörter waren, wurden daher nicht untersucht.

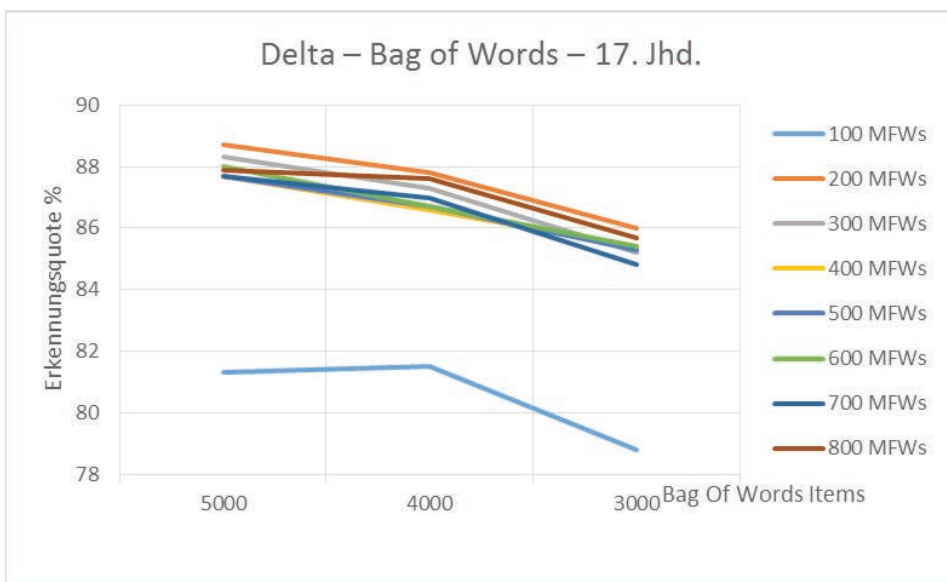
Nach Eder erzielt die Bag-of-Words-Technik, bei der nicht der ganze Text in den Test eingeht, sondern ein randomisiert zusammengestelltes Sample, das bspw. 5.000 Wörter enthält, bessere Ergebnisse als eine lineare Begrenzung.²⁶ Schon um beurteilen zu können, wie rasch die Erkennungsquote sinkt, wurden alle Tests jeweils mit Bag-of-Words in einer Länge von 3.000, 4.000 und 5.000 Wörtern durchgeführt. Da jede Bag-of-Words durch den Zufallsgenerator neu zusammengestellt wird, wurde jeder Test für jede Bag-

25 Eder [2013b, Anm. 13]; ähnlich die Ergebnisse von Dimpel [Anm. 17]. Stefan Evert und Friedrich Michael Dimpel versuchen derzeit, mittels Modifikationen im Testverfahren bessere Quoten für 3000–4000 Wörter zu erzielen.

26 Vgl. Eder [2013b, Anm. 13], S. 4; mit ähnlichem Ergebnis Dimpel [Anm. 17].

of-Words 100mal durchgeführt und der Mittelwert dieser Ergebnisse verwendet. Im Rahmen dieses Validierungstests wurden 432.000 Delta-Werte berechnet.

Hier die Ergebnisse:



Items BOW	100 MFWs	200 MFWs	300 MFWs	400 MFWs	500 MFWs	600 MFWs	700 MFWs	800 MFWs	Mittelwerte
5000	81,3	88,7	88,3	87,7	87,7	88	87,7	87,9	87,2
4000	81,5	87,8	87,3	86,6	86,7	86,7	87	87,6	86,4
3000	78,8	86	85,2	85,4	85,3	85,4	84,8	85,7	84,6
Z-max=1,64	82,3	88,5	88,2	87,8	87,3	87,4	87,8	87,2	87,1

Die letzte Tabellenzeile zeigt den Versuch einer Normalisierung der Z-Werte; hierbei ergibt sich keine weitere Verbesserung.²⁷ Gemessen an der Problemlage und der aus Zeitgründen unzureichenden Textaufbereitung geben diese Daten Anlass zur Hoffnung, dass sich bei einem größeren Korpus, besserer Textaufbereitung und bei einer Verbesserung der Skripts, die eine partielle automatische Normalisierung durchführen, zumindest nicht schlechtere Ergebnisse einstellen. Eder hat bei polnischen bzw. lateinischen Texten Erkennungsquo-

27 Zur Z-Wert-Normalisierung und der Reduktion von »Ausreißer«-Werten vgl. Evert/Jannidis/Pielström/Reger/Schöch/Vitt [Anm. 13]. Z-Werte > 1,64 wurden auf 1,7 gesetzt.

ten von etwas unter 80% bzw. 90% ermittelt.²⁸ Bemerkenswert ist die Nähe etwa zu Eders Befunden für das Lateinische: Auch hier werden die Erkennungsquoten für das Englische oder das Neuhochdeutsche nicht ganz erreicht, allerdings bleiben die Quoten wie auch hier bei kürzeren Texten bzw. Bag-of-Words stabiler als etwa beim Englischen.

Vor diesem Hintergrund sei nun ein erster Blick auf die fraglichen Zingref-Texte gewagt: Bei gleichem Validierungskorpus wurden nun jeweils ausschließlich der ›Weltkefig‹ (1632)²⁹ und die ›Newe Zeitung‹ (1621)³⁰ in das Ratekorpus gegeben.

MFW:	200	300	400	500	600	700	800
Erkennungsquote % des ›Weltkefig‹ zu Zingref	91	93	96	94	92	92	92
Erkennungsquote % der ›Newen Zeitung‹ zu Zingref	11	18	12	5	1	2	4
Erkennungsquote % der ›Newen Zeitung‹ zu a Sancta Clara	91	88	86	92	93	92	95

Die Erkennungsquoten informieren darüber, wie oft der Ratetext bei welchem Vektor (Zahl der MFWs) bei je neu randomisiert zusammengestellter Bag-of-Words (Länge 5.000) in 100 Iterationen das niedrigste Distanzmaß zum ›richtigen‹ Autor aufweist. Für die ›Newe Zeitung‹ (1621) erweisen sich Zingrefs ›Apophtegmata‹ im Validierungskorpus nicht als der am wenigsten unähnliche Text,³¹ sondern ›Judas der Erzscheml‹ (1685) von Abraham a Sancta Clara. Ein Vergleich von der ›Newen Zeitung‹ (1621) mit ›Judas der Erzscheml‹ zeigt, dass zwischen diesen Texten die geringste Unähnlichkeit in dem Sample mit neun Autoren besteht. Dagegen erkennt Delta unter neun möglichen Autoren in 92,9% der Fälle Zingref als richtigen Autor für den ›Weltkefig‹ bei Vektoren von 200–800 MFWs.

Diese Daten kongruieren mit vorhandenen Annahmen (im Falle des ›Weltkefigs‹) und überraschen (›Newe Zeitungen‹), ist es doch nicht möglich, dass Abraham a Sancta Clara (geb. 1644) Verfasser der ›Newen Zeitungen‹ ist. Die geringe Unähnlichkeit zwischen den Texten muss deshalb andere Ursachen haben, doch bleiben alle Erklärungsansätze vor weiteren Tests rein spekulativ. Denkbar wäre etwa, dass in diesem Fall doch die Gattungs- bzw. Textsortenkategorie, die sich z. B. in einer bestimmten Wort- und Themenwahl manifestiert, die Autorkategorie überlagert. So können beide Texte als theologisch-politische Traktate bezeichnet werden, die auch hinsichtlich ihrer Anlage durchaus

28 Eder [2013a und 2013b, Anm. 13].

29 Quotlibetisches Welt vnd Hummel Kefig [...], [o. O.] 1632.

30 Die drei früheren Fassungen der ›Newen Zeitungen‹ (1619, 1620a, 1620b) konnten aufgrund der geringen Textlänge nicht berücksichtigt werden.

31 Burrows' vermeidet es, von Ähnlichkeit zu sprechen und bevorzugt die Formulierung »less unlike«. Vgl. John F. Burrows: Questions of Authorship: Attribution and Beyond. A Lecture Delivered on the Occasion of the Roberto Busa Award, in: CHUM 37 (2003), S. 5-32, hier insbes. S. 13-15.

Ähnlichkeiten aufweisen. Hier sind weitere Tests mit erweiterten Korpora durchzuführen, die idealerweise Traktate, eventuell auch aus dem Straßburger Umfeld Zinggreffs, enthalten.³²

Grundsätzlich müssen die hier vorgestellten Ergebnisse noch mit Zurückhaltung interpretiert werden. Wichtig wäre eine Validierungsstudie mit größerem Textkorpus, eine Verbesserung der Teilnormalisierung, eine gründlichere Bereinigung von fremdsprachlichem Material und von Abkürzungen sowie der Einsatz von weiteren Techniken, die im Rahmen der Autorschaftsattribuion verwendet werden. Sofern die avisierten weiteren Validierungstests die hier vorgestellten Ergebnisse bestätigen, haben sie unmittelbaren Einfluss auf die Gestalt der Edition, bieten die Daten doch einerseits ein Argument gegen die Aufnahme der ›Newen Zeitungen‹ und andererseits ein Argument für die Aufnahme des ›Weltkefigs‹. Darüber hinaus fließen die Ergebnisse aber auch konkret in Einleitung und Kommentar der einzelnen Texte ein: So ergänzen sie nicht nur die philologischen Argumente zur Autorschaft der fraglichen Texte, sondern auch zu Schreibstrategien und Textsortenzuordnung.

32 Heranzuziehen wären in jedem Fall noch Texte von Theophil Dachtler. Vgl. Opel/Cohn [Anm. 5]. Des Weiteren wäre es prinzipiell möglich, wenn auch sehr unwahrscheinlich, dass sich Abraham a Sancta Clara am Stil der ›Newen Zeitungen‹ orientiert hat oder an einem unbekanntem dritten Autor, der seinerseits den ›Newen Zeitungen‹ nicht unähnlich schreibt.