

Einleitung: Quantitative Ansätze in den Literatur- und Geisteswissenschaften

Die Rede von digitalen Verfahren, die die geisteswissenschaftliche Forschung renovieren werden, ist derzeit ubiquitär, trübt aber den Blick dafür, dass quantitative Ansätze geisteswissenschaftlicher Forschung nicht neu sind, sondern auf eine lange Geschichte zurückblicken können. Denn der *zählende* Umgang mit Texten ist keinesfalls erst durch die ›digitale Revolution‹ der Geisteswissenschaften denkbar geworden. Vielmehr wird schon seit ungefähr 200 Jahren das, was vom späten Wilhelm Dilthey als Gegenstand der verstehenden und qualitativ orientierten Geisteswissenschaften bestimmt wurde, zum ›messbaren‹ Objekt erklärt.

Seit Beginn des 19. Jahrhunderts werden quantifizierende Verfahren angewandt, um Sprache und literarische Texte zu beschreiben, zu analysieren und zu interpretieren. Bis in die ersten Jahre des 21. Jahrhunderts sind solche Ansätze – beispielsweise aus den informellen Gruppen um Wilhelm Fucks oder Max Bense – in der Literaturwissenschaft stark unterrepräsentiert; im Zuge der Digital Humanities gewinnen sie jedoch wieder stark an Bedeutung. Diese Entwicklung fortschreibend, entwerfen auch die einzelnen Beiträge dieses Bandes historisch und systematisch reflektierte Perspektiven für eine auch, aber nicht ausschließlich, in den Digital Humanities beheimatete Quantitative Literatur- und Geisteswissenschaft und diskutieren ihr Potential in theoretischer und praktischer Hinsicht.

Um die Mitte des 19. Jahrhunderts unternehmen Wissenschaftler in zahlreichen Ländern der Welt und unabhängig voneinander den Versuch, literarische und sprachliche Parameter mit zählenden, messenden und rechnenden Methoden zu bestimmen und zu deuten. Auffallend dabei ist – und dies gilt bis ins frühe 20. Jahrhundert –, dass diese Vorstöße nicht in erster Linie von Philologen unternommen werden, sondern von Physikern, Mathematikern, Meteorologen, Philosophen oder Psychologen. Gegen Ende des 19. Jahrhunderts lässt sich vor allem in Russland und Europa ein genuines Interesse an einem quantifizierenden Umgang mit Sprache und Literatur beobachten.¹ Das erste Frequenzwörterbuch des Deutschen erschien 1898. Friedrich Wilhelm Kaeding, der Herausgeber, wurde

¹ Peter Grzybek, Emmerich Kelih: »Empirische Textsemiotik und quantitative Text-Typologie«, in: *Text & Reality. Text & Wirklichkeit*, hg. v. Jeff Bernard, Jurij Fikfak und Peter Grzybek. Ljubljana, Wien, Graz 2005, S. 95–120.

von fast 1000 freiwilligen Helfern unterstützt, um die von ihm erfassten 11 Mio. Wörter zu systematisieren.² Ohne Zweifel ist diese Zahl beeindruckend, doch vor dem Hintergrund der zeitgenössischen Verfügbarkeit und Möglichkeit der Verarbeitung von Daten verliert sie sogleich an Imposanz.

115 Jahre nach der Herausgabe des Frequenzwörterbuchs, im April 2013, verkündete Google Books, 30 Mio. Bücher gescannt zu haben; das hauseigene Textanalysetool Ngram-Viewer erlaubt es derzeit, 5,2 Mio. davon zu durchsuchen – das entspricht etwa 500 Mrd. Wörtern, von denen 37 Mrd. der deutschen Sprache zugehörig sind.³ Die 5,2 Mio. Bücher entsprechen etwa 4% aller jemals gedruckten Bücher. Lesen kann diese Menge natürlich niemand, aber neben dem häufig referierten Problem der beschränkten Leseleistung eines Menschen muss perspektivisch gerade die steigende Menge der Buchproduktion⁴ als Argument für eine Anpassung der Literaturwissenschaft an den Medienwandel der Gesellschaft hin zum Digitalen genannt werden. Wurden von 1740 bis 1900 etwa 32.000 Romane im englischsprachigen Raum publiziert, erschienen allein in Deutschland seit 2005 Jahr für Jahr etwa 75.000 bis 85.000 Bücher, bei einem recht stabilen Belletristik-Anteil von gut 30%.⁵ Zwar existiert ein Gutteil dieser Bücher ›nur digital, doch der Blick in den Bestand einer einzigen Bibliothek vermag die Notwendigkeit effizienter Digitalisierungstechniken aufzuzeigen: Die Württembergische Landesbibliothek nennt als Bestand 3,7 Millionen gedruckte Bücher, 15.420 Handschriften, 7.087 Inkunabeln und 180.439 Autographen. Es benötigte die Lebenszeit von über 7.000 Wissenschaftlern, um diesen Bestand einmal komplett zu lesen – etwa um nach einem bestimmten Ausdruck zu fahnden. Im Vergleich dazu benötigte Google Books am 20. Dezember 2017 genau 0,54 Sekunden, um in

² Friedrich Wilhelm Kaeding: *Häufigkeitswörterbuch der deutschen Sprache. Festgestellt durch einen Arbeitsausschuß der deutschen Stenographiesysteme*. Berlin 1898; Toni Bernhart: »Von Aalschwanzspekulanten bis Abendrotlicht. Buchstäbliche Materialität und Pathos im ›Häufigkeitswörterbuch der deutschen Sprache‹ von Friedrich Wilhelm Kaeding«, in: *Ethos und Pathos der Geisteswissenschaften. Konfigurationen der wissenschaftlichen Persona seit 1750*, hg. v. Ralf Klausnitzer, Carlos Spoerhase und Dirk Werle. Berlin, Boston 2015, S. 165–189.

³ Jean-Baptiste Michel (u. a.): »Quantitative Analysis of Culture Using Millions of Digitized Books«, in: *Science* (2011) H. 331, S. 176–182.

⁴ Marcel Lepper: »Big Data, Global Villages«, in: *Philological Encounters* 1 (2016), S. 131–162.

⁵ Die Zahlen stammen vom Börsenverein des Deutschen Buchhandels: http://www.boersenverein.de/sixcms/media.php/976/Titelproduktion_Erst_und_Neuauflage_final.pdf und http://www.boersenverein.de/sixcms/media.php/1117/Tab.4_BuBiZ_2017.pdf (beide 24.10.2017).

den bereits digitalisierten Beständen 202.000.000 Treffer für das Suchwort »Lesen« zu finden.⁶

Auf den ersten Blick scheint sich hieraus vor allem ein arbeitsökonomisches Argument ableiten zu lassen; immerhin könnten zeitintensive Archiv- und Bibliotheksbesuche bald der Vergangenheit angehören. Tatsächlich erstreckt sich der durch die Textdigitalisierung ermöglichte Wandel aber auch auf die inhaltliche Ebene der Texterschließung und Textanalyse. Denn mit dem Textkorpus wächst potentiell auch der Geltungsbereich literaturwissenschaftlicher Aussagen. Diese häufig als ›demokratisierend‹ oder ›entkanonisierend‹ lancierte Entwicklung der zunehmend ungerichteten und auf Masse abzielenden Digitalisierung gedruckter Wissensbestände durch Großkonzerne und öffentliche bestandshaltende Institutionen ist jedoch nicht ganz unkritisch zu betrachten. Sie stellt uns vor eine stattliche Reihe ökonomischer, juristische, politischer und schließlich auch technischer und wissenschaftlicher Probleme und Herausforderungen. Die verschiedenen wissenschaftlichen Disziplinen müssen dabei je eigene Wege im Umgang mit den digitalen Sammlungen finden.

Für die Geistes- und Literaturwissenschaft ist das Novum quantitativer Ansätze jedoch nicht nur die Frage nach dem *Was*, sondern insbesondere auch die Frage nach dem *Wie*, also nach der *Methodik* des Umgangs mit den textuellen Artefakten. Hier schließt der Band sowohl an Diskussionen der Methodologie als auch an zeitgenössische Diskussionen über spezifisch literaturwissenschaftliches Arbeiten an, die unter dem Stichwort *Praxeologie der Literaturwissenschaft* verhandelt werden.⁷ Hinzu kommt die Beobachtung, dass quantitative Verfahren in den Literatur- und Geisteswissenschaften lange Zeit unter erheblichem Legiti-

⁶ <http://www.google.de/search?q=lesen&btnG=Nach+B%C3%BCchern+suchen&tbo=bks&tbo=1&hl=de> (20.12.2017).

⁷ Vgl. Steffen Martus und Carlos Spoerhase: »Praxeologie der Literaturwissenschaft«, in: *Geschichte der Germanistik* (2009) 35/36, S. 89–96; Marcus Willand: »Replik: Steffen Martus und Carlos Spoerhase: Praxeologie der Literaturwissenschaft«, in: *Aussiger Beiträge* 5 (2011), S. 180–184; Carlos Spoerhase: »Big Humanities. ›Größe‹ und ›Großforschung‹ als Kategorien geisteswissenschaftlicher Selbstbeobachtung«, in: *Geschichte der Germanistik* 37/38 (2010), S. 9–27; Steffen Martus, Carlos Spoerhase: »Die Quellen der Praxis. Probleme einer historischen Praxeologie der Philologie. Einleitung«, in: *Zeitschrift für Germanistik* 23 (2013) H. 2, S. 221–225, *Theorien, Methoden und Praktiken des Interpretierens*, hg. v. Andrea Albrecht, Lutz Danneberg, Olav Kraemer und Carlos Spoerhase. Berlin 2015; vgl. auch den Beitrag von Friederike Schruhl in diesem Band.

mationsdruck standen. Vor diesem Hintergrund wird es aufschlussreich sein, gerade eine genuin quantitative Methodengeschichte für die Geisteswissenschaften zu rekonstruieren, an die auch in der Zukunft angeschlossen werden kann.⁸

Bei genauerer Betrachtung zeigt sich, dass in den gegenwärtigen Literatur- und Geisteswissenschaften quantitative Ansätze breit vertreten sind. Doch es ist kaum möglich, einen gemeinsamen Nenner dieser Forschung auszumachen. Dieses Desiderat erkannte der Wissenschaftsrat bereits 2012 in den Sozial- und Verhaltenswissenschaften und sprach folgende Empfehlung zur »Weiterentwicklung der wissenschaftlichen Informationsstrukturen in Deutschland bis 2020« aus:

Zur verbesserten Integration von quantitativen und qualitativen Forschungsansätzen in den Sozial- und Verhaltenswissenschaften wird der DFG und dem BMBF empfohlen, sich abzustimmen und ein Programm auszuschreiben, das Modellprojekte in diesem Bereich fördert. Das Programm sollte sich an Wissenschaftlerinnen und Wissenschaftler richten. Gegenstand der Ausschreibung sollte einerseits die Bearbeitung einer fachwissenschaftlichen Fragestellung mittels einer Verbindung quantitativer und qualitativer Forschungsdaten und -methoden sein. Darüber hinaus sollten entsprechende Projekte andererseits auch zur Weiterentwicklung von Standards und Methoden für die Langzeitverfügbarmachung von qualitativen Forschungsdaten beitragen.⁹

Diese Forderung einer Interaktion und Vermittlung quantitativer und qualitativer Forschungsansätze beschränkt sich nicht auf die Sozialwissenschaften, sondern scheint sich zuletzt und im Vergleich zu früheren Versuchen dieser Art nun längerfristig und über Disziplinengrenzen hinweg auch in einer Institutionalisierung der Digital Humanities niederzuschlagen. So sind seither, in den fünf Jahren

⁸ Ein fachgeschichtlich interessantes Beispiel aus der zweiten Hälfte des 20. Jahrhunderts ist gerade das Verschwinden quantitativer Verfahren aus der Literaturwissenschaft, das sich mit einem Verweis auf die Geschichte der Linguistik erklären lässt. Diese hatte ab den 1970er Jahren, auch im Zuge der Emanzipation von der Literaturwissenschaft, deren quantitative Verfahren inkorporiert und weiterentwickelt. So ist in der Sprachwissenschaft *Quantitative Linguistik* seit Jahrzehnten ein etabliertes Spezialgebiet mit ausgeprägter Theoriebildung, eigenen Zeitschriften, Lehrbüchern und Lehrstühlen.

⁹ Wissenschaftsrat: »Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020« (13.07.2012), S. 58. <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf> (10.08.2013)

nach der Empfehlung des Wissenschaftsrates, etwa 50 Digital Humanities-Professuren im deutschsprachigen Raum ausgeschrieben worden.¹⁰ »Interdisziplinarität« in den DH bedeutet nicht nur, in den D- und H-Disziplinen gemeinsame Forschungsfragen oder Forschungsgegenstände zu entwickeln, sondern vielmehr auch, zwischen disziplinspezifischen Formen der *Modellierung* und *Operationalisierung* von Forschungsfragen auf theoretischer und methodischer Ebene zu vermitteln. Diese Vermittlung findet dabei sowohl zwischen den Forschungspositionen selbst als auch zwischen den sie charakterisierenden Theorien, Methoden und Praktiken statt.

Als Übertrag für die quantifizierende Perspektive und gleichsam als leitmotivische Bedingung für die in diesem Band versammelten Beiträge ist auf diesem Wege die Prämissen eines *theoriegeleiteten Umgangs mit Daten* eingeführt. Die damit explizierte Korrektiv-Funktion theoretisch entworferne Modelle durch die Forderung ihrer praktischen Anwendbarkeit auf Daten funktioniert ebenso *vice versa* als Korrektur eines unsystematischen Umgangs mit Daten durch Theorie. Der geisteswissenschaftlich-interdisziplinär ausgerichtete Band mit Fokus auf die literaturwissenschaftliche Forschung soll als Plattform dieses Dialogs dienen und die Wissenschaftsfähigkeit der disziplinspezifischen Modellierungen auf den Prüfstand setzen. Dabei sollen – idealiter ausgerichtet auf die brisante Frage nach den Möglichkeiten einer interdisziplinären Standardisierung quantitativer Verfahren¹¹ – in einem ersten Schritt quantitative Verfahrensmodelle gesichtet, Verfahrenskataloge erstellt und gegebenenfalls die Grundlagen für Empfehlungen von *best-practice*-Modellen als Antwort auf spezifische Fachfragen vorbereitet werden.

Der Band fußt auf der Tagung »Scientia Quantitatis. Quantitative Literaturwissenschaft in systematischer und historischer Perspektive«, die im Oktober 2014 von den Abteilungen für Neuere deutsche Literatur des Instituts für Literaturwissenschaft, dem Stuttgart Research Center for Text Studies der Universität Stuttgart und dem Deutschen Literaturarchiv Marbach mit Unterstützung der

10 Vgl. Patrick Sahle: »Zur Professoralisierung der Digital Humanities«, <http://dhd-blog.org/?p=6174> (20.12.2017). Teilweise wurden Professuren mehrfach ausgeschrieben; d. h. die Zahl der tatsächlichen Professuren ist etwas niedriger.

11 Vgl. das Sonderheft des *Journal of Literary Theory* 5 (2011) H. 2, dessen Beiträger auf die Frage antworten, ob (und wenn ja, welche) Standards und Normen im Feld der *literary studies* festzumachen seien.

VolkswagenStiftung im Schloss Herrenhausen in Hannover stattfand.¹² Die Referentinnen und Referenten der Tagung haben ihre Beiträge für den Druck überarbeitet; weitere Beiträge kamen auf Einladung der Herausgeberinnen und Herausgeber hinzu und erweitern diesen Band.

Unterschiedliche Ordnungsprinzipien böten sich an, um die Beiträge in diesem Band zu gruppieren. Eine Unterscheidung nach stärker systematischer oder stärker historischer Perspektivierung wäre denkbar, eine Unterteilung in Beiträge, die quantitative Ansätze eher praktisch anwenden oder eher methoden-theoretisch oder wissenschaftsgeschichtlich reflektieren; schließlich wäre auch eine chronologische Reihung hinsichtlich der Gegenstände möglich, die von den Beiträgen als Forschungsobjekt in den Blick genommen werden. Denkbar wäre auch eine Gliederung nach den Fachgebieten Computerlinguistik und Linguistik, Literaturwissenschaft, Archiv- und Handschriftenkunde, Bildungsforschung, Wissenschaftsgeschichte und Geographie. Weil jedoch sämtliche Beiträge mindestens auf zwei der genannten Bereiche ausgreifen, erschien uns eine vergleichsweise offene Gliederung nach den zwei Teilen »Applikationen« und »Reflexionen« angemessen.

Der erste Teil »Applikationen« wird eröffnet durch den Beitrag von Jonas Kuhn, der am Beispiel von Abenteuerromanen das textanalytische Potential computerlinguistischer Verfahren exploriert und in der Form eines Werkstattberichts langjährige Kooperationserfahrungen zwischen Computerlinguistik und Literaturwissenschaft reflektiert. Nils Reiter und Marcus Willand nehmen dramatische Texte in den Blick, die – anders als erzählende Texte – erst allmählich zu einem Forschungsgegenstand quantitativer und computergestützter Analysemethoden werden. Besondere Aufmerksamkeit richten sie dabei auf die Operationalisierung literaturwissenschaftlicher und dramenpoetologischer Fragestellungen. Dramenanalytisch ist auch der Beitrag von Christof Schöch, der das Distanzmaß Zeta nach John Borrows verwendet, um die Gattungen Komödie, Tragödie und Tragikomödie messend miteinander zu vergleichen und so auf methodischer Ebene Zeta weiterzuentwickeln. Der poetischen Zusammenarbeit von Bertolt

¹² Vgl. Ruth Kuntz-Brunner: »Zwischen den Zeilen. Mit Technik und Methode«, in: *Impulse. Das Wissenschaftsmagazin der VolkswagenStiftung* 1 (2015), S. 72–75; Peggy Bockwinkel: »Tagungsbericht Scientia Quantitatis – Quantitative Literaturwissenschaft in systematischer und historischer Perspektive«, 30.09.–02.10.14 in Hannover, in: *Glottotheory* 6.1 (2015), S. 229–235; Friederike Schruhl: »Scientia Quantitatis. Quantitative Literaturwissenschaft in systematischer und historischer Perspektive (Tagung in Hannover v. 30.9.–2.10.2014)«, in: *Zeitschrift für Germanistik*, NF 25.2 (2015), S. 423–424.

Brecht mit Elisabeth Hauptmann widmet sich Florian Barth in seiner stilometrischen Studie, gefolgt von Friedrich Michael Dimpel, der Textauszeichnungen in den Gattungen Märe und Novelle vornimmt und die gewonnenen Daten narratologisch evaluiert. Empirische Bildungsforschung und Literaturwissenschaft verknüpfen Angelika Zirker, Judith Glaesser, Augustin Kelava und Matthias Bauer in ihrer Studie, die am Beispiel eines Shakespeare-Sonetts Textverständnis misst. Wenn Gabriel Viehhauser mittels quantifizierender Methoden mittelhochdeutsche Tagelieder analysiert und interpretiert, leitet er mit seinem methodologisch reflektierten Rekurs auf Forschungen, die aus der Zeit des prä-digitalen Zeitalters stammen, über auf den zweiten Teil des Bandes.

Dieser zweite Teil »Reflexionen« beginnt mit einem wissenschaftsgeschichtlichen Abriss zur Verwendungsgeschichte quantitativer Methoden in den letzten zwei Jahrhunderten von Toni Bernhart. Cornelis Menke reflektiert kritisch die Vorstellung eines Quantifizierungsideals und Friederike Schruhl sieht aus praxeologischer Perspektive Forschungsaufsätze in DH-Zeitschriften mit dem Ziel, quantifizierende Interpretationspraktiken zu identifizieren. Benjamin Krautter nimmt bekannte Verfahren zur Autorschaftsattribution kritisch in den Blick, um nach stilometrischen Erweiterungsmöglichkeiten dieser wohl ältesten und prominentesten quantitativen Verfahren zu fragen. Literaturwissenschaftliche Netzwerkanalyse ist das Thema des Beitrags von Carolin Hahn, während Celia Krause und Philipp Hegel eine quantitative Kodikologie entwickeln. Ebenfalls mit Netzwerken, deren Punkten und Kanten, Grenzen und Weiten beschäftigt sich Anne Baillot. Abgeschlossen und zugleich geöffnet wird der zweite Teil durch den Beitrag von Julia Lossau, die das Fach der Geographie auf programmatische Aspekte und Verständnisweisen des Quantitativen hin durchmisst.

Wie danken den Autorinnen und Autoren für ihre Beiträge und die sehr inspirierende und produktive Zusammenarbeit. Anja-Simone Michalski danken wir für die Aufnahme unseres Bandes in das Verlagsprogramm von Walter de Gruyter und Stella Diedrich für die Begleitung des gesamten Produktionsprozesses. Den studentischen Hilfskräften Anja Braun, Martin Kuhn und Florian Barth danken wir für ihre gründliche Mitarbeit bei Korrektorat und Texteinrichtung. Dem Deutschen Literaturarchiv Marbach und dem Stuttgart Research Centre for Text Studies schulden wir Dank für die Basisfinanzierung dieser Buchveröffentlichung. Unser besonderer Dank schließlich gilt der VolkswagenStiftung, die durch ihre großzügige Mittelbewilligung diesen Band in dieser Form ermöglicht hat.

Toni Bernhart, Marcus Willand, Sandra Richter und Andrea Albrecht
Stuttgart und Heidelberg, im Dezember 2017

