

Methoden-Werkstatt

Axel Philipps

Text Mining-Verfahren als Herausforderung für die rekonstruktive Sozialforschung

<https://doi.org/10.1515/sosi-2018-0017>

Zusammenfassung: In den letzten Jahren nehmen die Anwendungsvarianten computergestützter Verfahren für die qualitative Sozialforschung zu. Vor allem das Text Mining verspricht eine automatisierte Strukturierung größerer Textmengen. Während Einzelne darin ein großes Potential für die rekonstruktive Sozialforschung sehen, fehlt weitgehend unter den rekonstruktiv Forschenden eine Auseinandersetzung mit diesem Thema. Vor dem Hintergrund von Einzelfallanalysen mag dies berechtigt sein, aber bei der zeitgleich stattfindenden digitalen Durchdringung der empirischen Sozialforschung wird eine Beschäftigung damit unausweichlich. Wir brauchen Antworten, wie webbasierte Inhalte medien-spezifisch zu interpretieren sind oder worin sich die rekonstruktive Sozialforschung von den Digital Humanities unterscheidet – wo bereits das Zusammenspiel von computergestützten Analysen und hermeneutischen Verfahren erprobt wird. Der Beitrag bietet daher einen ersten Überblick zu Text Mining-Verfahren, diskutiert die Sicht ihrer Vertreterinnen und Vertreter auf die qualitative Sozialforschung und nennt erste Überlegungen, wie und unter welchen Bedingungen der Einsatz von automatisierten Verfahren für rekonstruktive Herangehensweisen sinnvoll sein könnte.

Schlüsselwörter: Computergestützte Analysen, Text Mining-Verfahren, qualitative Sozialforschung, rekonstruktive Sozialforschung, Inhaltsanalysen

1 Einleitung

In der empirischen Sozialforschung sind computergestützte Analysen ein fester Bestandteil der Statistik. Aber auch die qualitative Sozialforschung erfolgt teilweise computergestützt, zumindest wenn es darum geht, Texte zu verwalten, darin zu suchen und die Struktur der gebildeten Codes zu visualisieren (vgl. Friese 2014; Kuckartz 2010; Kuckartz; Grunenberg; Dresing 2007).¹ So bieten die Basisausführungen von Softwarelösungen wie MaxQDA oder Atlas.ti² Hilfestellungen, ohne jedoch die qualitativen Daten selbst zu analysieren. Eine Erweiterung der Bearbeitungsmöglichkeiten versprechen in diesem Zusammenhang Text Mining-Verfahren, welche automatisch größere Textmengen strukturieren und damit einen Teil der Analyse übernehmen. Sie unterscheiden sich darin von Visualisierungsfunktionen wie dem Code-Relation-Browser (MaxQDA) oder dem Cooccurrence Explorer (Atlas.ti), welche Verknüpfungen von Codes, aber nicht Strukturen auf der Ebene von Wortformen veranschaulichen.

Daraus werden eine Reihe von Argumenten für den Einsatz computergestützter Analysewerkzeuge in der qualitativen Sozialforschung abgeleitet³ (etwa Bail 2014; Evans; Aceves 2016; Goldberg 2015; McFarland; Lewis; Goldberg 2015; Mohr; Bogdanov 2013; Stulpe; Lemke 2015; Wagner-Pacifici u.a. 2015; Wiedemann 2013, 2016; Wiedemann; Lemke 2015). Die genannten Autorinnen und Autoren betonen, dass sich mit der Digitalisierung der Bestand elektronisch auswertbarer Datenbestände erstens ausgeweitet habe und darauf mit adäquaten Analyseverfahren zu reagieren sei. Zweitens seien die automatischen Verfahren geeignet, um unstrukturierte, nichtreaktive Daten auszuwerten. Und drittens ermöglichen die computergestützten Verfahren, Muster und Regelmäßigkeiten in semantischen Gebrauchsweisen zu entdecken. So könnten inhaltliche Differenzierungen ebenso wie semantische Strukturen offengelegt werden, die ansonsten in großen Textmengen untergingen und sich manuell nur mit

1 In diesem Beitrag geht es ausschließlich um Textanalyseverfahren. Unberücksichtigt bleiben andere Medien wie Bilder, Audioaufzeichnungen, Filme, Videoclips und ihre computergestützte Analyse. Forschungspraktisch lassen sich audio-visuelle Materialien beispielsweise mit MaxQDA, Atlas.ti oder TrAVis (Jost u.a. 2013) kodieren.

2 Der Beitrag geht nur auf MaxQDA und Atlas.ti näher ein. Die Ausführungen gelten in Grundzügen aber auch für Angebote wie NVivo, QDA Miner, The Ethnograph, Kwalitan, f4analyse und andere.

3 Der Fokus liegt auf rekonstruktiven Verfahren im Gegensatz zu vergleichbaren Diskussionen zu automatischen Inhaltsanalysen etwa in den Kommunikationswissenschaften (Keyling 2014; Scharkow 2013).

einem enormen Aufwand darstellen ließen. Die Text Mining-Verfahren würden dazu sowohl Häufigkeiten und Gewichtungen von Begriffen und Wortgruppen als auch unterschiedliche Grade von Begriffskontexten einbeziehen.

Auf den ersten Blick scheinen diese Möglichkeiten des Text Minings – Texte zu analysieren, mit nichtreaktiven Daten zu arbeiten sowie Muster und semantische Strukturen zu entdecken – an das qualitative Forschungsverständnis angeschlossen. Genauer betrachtet gilt dies aber nur für einen sehr begrenzten Bereich der qualitativen Sozialforschung. So überrascht es wenig, dass computergestützte Verfahren in erster Linie bei qualitativen Inhaltsanalysen (Kuckartz 2010) oder in der Diskursforschung⁴ (Diaz-Bone; Weischer 2014) zum Einsatz kommen. Ungeachtet dessen wird mit den Text Mining-Verfahren auch der Anspruch verbunden (siehe etwa Wiedemann; Lemke 2015), dass sie hermeneutische Interpretationen in der empirischen Sozialforschung komplementär ergänzen können.

Der Beitrag argumentiert, dass die vermutete Anschlussfähigkeit auf Missverständnisse zwischen den Leistungen von Text Mining-Verfahren und dem, was qualitative Sozialforschung auszeichnet, zurückgehen. In den Verknüpfungen beider Aspekte wird nämlich häufig letztere auf Inhalts- und Bedeutungsanalysen reduziert. Dieses Verständnis vernachlässigt im Allgemeinen die Vielfalt an Methoden und Datenarten in der qualitativen Sozialforschung sowie im Speziellen (zumindest im deutschsprachigen Raum) die Differenz zwischen inhaltsanalytischen und rekonstruktiven Textanalyseverfahren.⁵ Entgegen einer Zurückweisung von Text Mining-Verfahren sucht der Beitrag jedoch nach einigen Schnittmengen zur qualitativen Sozialforschung, um einerseits potentielle Probleme zu benennen und andererseits Möglichkeiten auszuloten, wie sich Text Mining-Verfahren produktiv speziell in rekonstruktiven Textanalysen einsetzen lassen.

Der Beitrag untergliedert sich in sechs Abschnitte. Einer kurzen Einführung und Übersicht zu Text Mining-Verfahren folgt eine Darstellung der Sichtweise von Text Mining-Vertreterinnen und Vertreter auf die qualitative Sozialforschung. Es wird deutlich, dass sie „qualitativ“ mit „Inhalts- und Bedeutungsanalysen“ gleichsetzen. Sie vernachlässigen damit konstitutive Grenzziehungen

⁴ In der Diskursforschung existiert ein stark ausdifferenziertes Repertoire an analytischen Verfahren (Keller 2007; Keller u.a. 2011). Die Vielschichtigkeit an Diskursanalysen kann hier nicht behandelt werden. Der Beitrag konzentriert sich daher auf qualitative Inhaltsanalysen.

⁵ Der Beitrag geht an dieser Stelle nicht auf andere Erhebungs- und Auswertungsverfahren der qualitativen Sozialforschung ein (etwa Ethnografie, Film- oder Bildanalysen). Der Fokus liegt mit dem Text Mining auf textanalytischen Verfahren.

zwischen der quantitativen, der qualitativen und der rekonstruktiven Sozialforschung, so dass bisher keine Notwendigkeit bestand, das Verhältnis des Text Minings zu rekonstruktiven Herangehensweisen zu problematisieren. Vor diesem Hintergrund diskutiert ein eigener Abschnitt Einsatzmöglichkeiten von Text Mining-Verfahren für die rekonstruktive Sozialforschung im Rahmen der Text- und Datenerschließung. Das Fazit fasst die Darlegung zusammen und gibt einen vorläufigen Ausblick.

2 Was leisten Text Mining-Verfahren?

Text Mining-Verfahren strukturieren unstrukturierte Daten. Während sich bei einem Text oder einer Textpassage die inhaltliche Struktur relativ leicht erkennen lässt, liegt der Inhalt in großen Mengen an Texten unstrukturiert vor. Vor allem bei umfangreichen und verschiedenartigen Texten kommen Text Mining-Verfahren zum Einsatz, um auf der Basis statistischer Merkmale von Sprache in digital vorliegenden Texten neue sowie relevante sachliche und inhaltliche Zusammenhänge zu extrahieren. Heyer und Kollegen sprechen daher von computergestützten Verfahren „für die semantische Analyse von Texten“ oder „die automatische bzw. semi-automatische Strukturierung von Texten, insbesondere sehr großen Mengen von Texten“ (Heyer u.a. 2006: 3). Mit dem Text Mining sind dabei in erster Linie statistische und musterbasierte Verfahren gemeint: Einerseits werden Texte auf Grundlage sprachstatistischer Gesetzmäßigkeiten analysiert, andererseits orientiert sich die Software an bestimmten Mustern im Sprachgebrauch. Sie dienen dabei primär explorativen Analysen, das heißt, mit ihnen lassen sich beispielsweise spezifische Ausdrücke oder Wortgruppen identifizieren, semantische Relationen zwischen Ausdrücken berechnen, Ähnlichkeiten zwischen Begriffen aufzeigen oder vergleichbare Dokumente finden.

Grundsätzlich lässt sich eine Vielfalt an Text Mining-Verfahren unterscheiden (Heyer u.a. 2006; Ignatow; Mihalcea 2016; Wiedemann 2016). Sie alle dienen dazu, große Textmengen zu strukturieren. Strukturieren heißt in diesem Zusammenhang vor allem, Wörter inhaltlich zu gruppieren oder Relationen zwischen Wortformen zu identifizieren. Die automatischen Analysewerkzeuge folgen dazu strukturalistischen Annahmen über die Sprache und gehen von Wortformen als kleinste bedeutungstragende Einheiten aus. Auf dieser Basis werden mithilfe von Sprachstatistiken und sprachlichen Beziehungsmustern automatisch potentiell divergierende Inhalte und Themen in größeren Textmengen ausgewiesen. Exemplarisch werden als Text Mining-Verfahren für die Sozialwissenschaften vor allem Frequenzanalysen, Topic-Modelle und Kook-

kurrenzanalysen genannt (Blei 2012; DiMaggio 2015; Mohr; Bogdanov 2013; Lemke; Wiedemann 2015; Wiedmann 2013). Frequenzanalysen zählen die relative und absolute Häufigkeit von Wörtern synchron als auch diachron in einem Textkorpus.⁶ Eine synchrone Analyse von Worthäufigkeiten legt beispielsweise zentrale Schlüsselbegriffe offen, insbesondere wenn Stoppwörter (Artikel, häufig auftretende Hilfsverben etc.) ausgeklammert werden. Im Zeitverlauf lässt sich hingegen das Auftreten eines Wortes oder einer Phrase visualisieren, so dass Konjunkturen in Debatten identifiziert werden können. Im Sammelband von Lemke und Wiedemann (2015) dienen Frequenzanalysen wiederkehrend für eine (zeitliche) Datenstrukturierung, um daran weitere Analyseschritte anzuschließen. Topic-Modelle berechnen hingegen die wahrscheinlichkeitsbasierte Verteilung von Themen in einem Dokument oder in größeren Textkorpora. „Topic modeling algorithms are statistical methods that analyze the words of the original texts to discover the themes that run through them, how those themes are connected to each other, and how they change over time“ (Blei 2012: 77f.). Diese Modelle gehen davon aus, dass ein Dokument stets aus einer Kombination mehrerer Themen besteht, welche in einem Text zu unterschiedlichen Anteilen enthalten sind. Innerhalb eines Themas werden den Worten unterschiedliche Gewichtungen zugewiesen und so die Wahrscheinlichkeit des Auftretens einer vorab festgelegten Zahl an Themen in der gesamten Textkollektion berechnet. Der Vorgang ist unüberwacht, das heißt, die Analyse ist rein datengetrieben und kann jedes Mal anders ausfallen. DiMaggio und Kollegen (2013) etwa haben mit einem Topic-Modell verschiedene Themen in fünf US-Printmedien hinsichtlich der Förderung von Kunst über einen Zeitraum von 1986 bis 1997 identifiziert. Die Themen unterschieden Kunstgattungen (Film, Theater, klassische Musik etc.), Kunstinstitutionen sowie Debatten um die Ausweitung bzw. Kürzung von Kunstförderprogrammen. In einer Studie zur Repräsentation von Street Art auf der webbasierten Plattform Flickr wurde wiederum ein Topic-Modell eingesetzt, um die Aussagekraft des vorab definierten Kategoriensystems mit den automatisch generierten Themenclustern zu erhöhen (Phillips; Zerr; Herder 2017). Kookkurrenzanalysen ermöglichen wiederum, signifikante Verwendungszusammenhänge eines Wortes mit anderen zu identifizieren (Lemke 2014; Stulpe; Lemke 2015). Dazu werden jene Worte in einem Netzwerk visualisiert, die häufig, aber sprachstatistisch nicht rein zufällig miteinander in Textkollektionen auftreten. Aus solchen Zusammenhängen werden schließlich

⁶ Die Option, Worthäufigkeiten in Textdokumenten auszuzählen, findet sich beispielsweise bei Atlas.ti, den erweiterten MaxQDA-Versionen (Plus und Analytics Pro) oder bei WordStat im QDA Miner.

Rückschlüsse auf die inhaltlichen Spezifikationen der analysierten Dokumente gezogen. Bei Förster (2015) beispielsweise zeigen Kookkurrenzanalysen in Printmedien für zwei Zeiträume⁷ verschiedene semantische Beziehungsgeflechte. Während in beiden Zeiträumen mit dem Begriff „Folter“ Wörter wie „Mord“, „systematisch“ und „Misshandlungen“ signifikant häufig zusammen auftreten, zeigt sich im Vergleich eine Verschiebung nach den Anschlägen am 11. September 2001 in den USA. Davor wurde „Folter“ ebenso häufig mit „Todesstrafe“ und „Vergewalt“(igung) zusammen genannt. Danach tritt das Wort „Folter“ signifikant häufig gemeinsam mit „Verbot“ oder „unmenschliche Behandlungen“ auf. Förster wählte anhand dieser Ergebnisse schließlich Printmedienbeiträge aus diesen Zeiträumen aus, um sie inhaltsanalytisch zusammenzufassen und deren Bedeutungsgehalte zu explizieren.

Während die genannten Text Mining-Verfahren Textkollektionen inhaltlich explorieren, geht es bei automatischen Klassifikationsverfahren und dem Active Learning bereits stärker um computergestützte quantitative Auswertungen auf der Basis qualitativ gebildeter Kategorien (Stulpe; Lemke 2015). Der qualitativen Textanalyse kommt bei diesen Verfahren die Aufgabe zu, trennscharfe und eindeutige Kategorien zu bilden. Sie helfen, größere Textmengen automatisch zu differenzieren. Darüber hinaus kommen in den Digital Humanities auch computergestützte Verfahren zur Anwendung, die automatisch Gesprächs- und Erzählstrukturen (etwa Reiter; Frank; Hellwig 2014) oder diskursspezifische Überzeugungsschritte (etwa Biber; Connor; Upton 2007) analysieren und visualisieren. Davon abgesehen wird die Relevanz von Mining-Verfahren für die qualitative Sozialforschung insgesamt eher in ihrem explorativen Potential gesehen. Der nächste Abschnitt geht darauf näher ein.

3 Der Anschluss von Text Mining-Verfahren an die qualitative Sozialforschung

Eine rege Debatte zur Anwendung von Text Mining-Verfahren in den Sozialwissenschaften pendelt zwischen Belegen für ihre Produktivität (Ahonen 2015; Dumm; Niekler 2015; Evans; Aceves 2016; Mohr; Bogdanov 2013) und Vorbehalten gegenüber ihrer Nutzung. Hinsichtlich der Schwierigkeiten verweisen Bubenhofer und Scharloth (2015) beispielsweise auf einen unterkomplexen Text-

⁷ Der erste Zeitraum geht vom 1. September 1990 bis zum 10. September 2001 und der zweite vom 11. September 2001 bis zum 10. September 2012.

begriff im Zusammenhang mit Text Mining-Verfahren. Sie würden versäumen, „Texte als komplexe Gewebe zu operationalisieren“ (Bubenhof; Scharloth 2015: 13). Des Weiteren kennzeichne die Verfahren einen Mangel an Theorie und methodischen Standards (Cecire 2011), der Einsatz von intransparenten Algorithmen (Bubenhof; Scharloth 2015; González-Bailón u.a. 2014), unklare Funktionsweisen von Visualisierungen (Bubenhof; Scharloth 2015; Kath; Schaal; Dumm 2015) oder der noch ungeklärte ethische Umgang mit digitalen Daten (Boyd; Crawford 2012; Bubenhof; Scharloth 2015).

Abgesehen von diesen Schwierigkeiten wird für eine produktive Nutzung von Text Mining-Verfahren in den Sozialwissenschaften allgemein und in der qualitativen Forschung im Besonderen geworben. Die entsprechenden und teilweise nicht ganz unproblematischen Argumentationen betonen vor allem Überschneidungen zwischen den Einsatzfeldern und Vorgehensweisen des Text Minings und der qualitativen Sozialforschung. Zu den Argumenten zählt erstens die induktive Analyseperspektive. Auch bei den Text Mining-Verfahren gehe es nicht primär um eine Hypothesenprüfung, sondern um die Entdeckung von Mustern und Sinnzusammenhängen in Texten. Es würde sich um ein „excellent tool for discovery“ (Brett 2012: 14) handeln, „[to] identify novel pattern and relations“ (McFarland u.a. 2015: 1). So schreiben auch Bubenhof und Scharloth: „Die algorithmische Vorstrukturierung der Daten erlaubt es, Zusammenhänge in den Daten zu finden, die vorher unbekannt waren und zu neuen Hypothesen führen“ (2015: 2f.). Darin wird eine Analogie zur gegenstandsbezogenen Generierung von Annahmen und Theorien der Grounded Theory (Glaser; Strauss 1967) gesehen (siehe auch Evans; Aceves 2016). Im Vordergrund stünden die Texte sowie die darin offengelegten Muster und Zusammenhänge. In dieser Sichtweise werden aber auch erste Verkürzungen deutlich. Erstens erklärt sie die gegenstandsbezogene und von Interpretationen vorangetriebene Theoriebildung der Grounded Theory zum zentralen Grundverständnis der qualitativen Sozialforschung, ohne die Vielfalt und divergenten Erkenntnisansprüche qualitativer Methoden bei unterschiedlichen Fragestellungen zu berücksichtigen. So verfolgen einige qualitative Verfahren im Gegensatz zur Grounded Theory auch andere Ziele. Beispielsweise beschränkt sich die zusammenfassende Inhaltsanalyse (Mayring 2010) darauf, einen Text mit den Worten der Forschenden zu komprimieren oder das thematische Kodieren (Flick u.a. 2013) dient dazu, die thematische Struktur von verbalisierten Sichtweisen für vorab festgelegte Gruppen herauszuarbeiten. Zweitens beziehen sich die Text Mining-Verfahren ausschließlich auf Common Sense-Bedeutungen in den Texten, also auf „meaning (as manifest in text and image)“ (McFarland u.a. 2015: 20). Vertreterinnen und Vertreter von Text Mining-Verfahren setzen folglich die qualitative

Sozialforschung mit der Analyse von Textinhalten und Themen (frames, topics) gleich. Wenn die Rede von qualitativen Methoden der Soziologie ist, sind in der Regel qualitative Inhaltsanalysen gemeint (Bail 2014; Ignatow 2016; Mohr; Bogdanov 2013; Wiedemann 2013). Die Gleichsetzung spitzt sich etwa bei Dumm und Niekler (2015: 102) in der Annahme zu, dass qualitativ-orientiert Forschende die qualitative Inhaltsanalyse bevorzugen würden. Diese Charakterisierung steht jedoch im Widerspruch zur Breite etablierter qualitativer Vorgehensweisen (siehe etwa Denzin; Lincoln 2011; Flick u.a. 2013 oder Hitzler; Honer 1997). Drittens gehen Verfechterinnen und Verfechter der Text Mining-Verfahren davon aus, dass mit dem Vorrang qualitativer Inhaltsanalysen in der qualitativen Sozialforschung ein verstärktes Interesse an Themengruppierungen und semantischen Bedeutungsanalysen in Texten bestehe (Bail 2014; Stulpe; Lemke 2015; Wiedemann 2013) – was wiederum eine besondere Stärke von Text Mining-Verfahren sei. So würden Topic-Modelle mittels wahrscheinlichkeitsbasierter Wortverteilungen Texte inhaltlich differenzieren und dafür charakterisierende Wörter herausstellen. Es ist daher auch wenig verwunderlich, dass die bisherige Forschung viertens mit Text Mining-Verfahren vor allem Themen differenzierte und identifizierte (etwa DiMaggio; Nag; Blei 2013; Hopkins 2013; Ignatow; Williams 2011) oder in Kombination mit qualitativen Inhaltsanalysen arbeitete (siehe dazu einige Beiträge in Lemke; Wiedemann 2015). Die Konzentration auf die inhaltliche Bedeutungsebene von Texten dokumentiert sich fünftens auch in der Rezeption rekonstruktiver Verfahren der Textinterpretation. So gehen Wiedemann und Lemke (2015) beispielsweise auf die objektive Hermeneutik ein; wobei sie betonen, dass Text Mining-Verfahren ungeeignet sind, latente Sinngehalte wie in einer sequenzanalytischen Vorgehensweise zu rekonstruieren. Sie stellen aber in Aussicht, dass rekonstruktive Verfahren komplementär zu Text Mining-Verfahren eingesetzt werden könnten. So würde dann zum beobachteten manifesten Inhalt „im Zuge eines rekonstruktiven Analyseschritts ein latenter Sinngehalt“ ergänzend herausgearbeitet werden (Wiedemann; Lemke 2015: 415). In ihrem Beispiel komme in der Äußerung „gegen den Mindestlohn“ eine Ablehnung explizit zum Ausdruck. Die latente Ablehnung des Mindestlohns im arbeitnehmerfreundlichen Argument „Tarifautonomie nicht beschädigen“ sei jedoch erst zu rekonstruieren, um sie anschließend als manifesten Ausdruck zu messen.

4 Zur Grundannahme der rekonstruktiven Sozialforschung

Vor allem an diesem Verständnis von rekonstruktiven Verfahren wird das Grundproblem der Anschlussfähigkeit von Text Mining-Verfahren an die hermeneutisch orientierte Sozialforschung im Weiteren beleuchtet. Wie deutlich geworden ist, verstehen die angeführten Vertreterinnen und Vertreter die qualitative Sozialforschung vorrangig als zusammenfassende oder explizierende Analyse von Inhalten und Bedeutungen. Während sich die qualitative Inhaltsanalyse vor allem mit manifesten Bedeutungen auseinandersetzt, würden rekonstruktive Verfahren verborgene Mehrfachbedeutungen aufdecken. Bei der Rekonstruktion „latenter Sinnstrukturen“ mithilfe der objektiven Hermeneutik geht es jedoch nicht primär um weitere Bedeutungen, sondern um die Freilegung bedeutungsgenerierender Regeln (Oevermann 2001). Ebenso rekonstruieren die Narrationsanalyse (Schütze 1983; Küsters 2009) oder die dokumentarische Methode (Bohnsack 2014; Bohnsack; Nentwig-Gesemann; Nohl 2013) in erster Linie die Strukturierung von Äußerungen und Handlungen und nicht verschiedene Bedeutungsinhalte. Dazu ist für die rekonstruktiven Verfahren kennzeichnend, dass sie über die Interpretation der Inhalte (WAS) hinausgehen und über die Art und Weise, wie Themen und Darstellungen bearbeitet und strukturiert werden (WIE), auf latent wirkende Strukturierungsmuster schließen. Ein zentraler Arbeitsschritt ist die Sequenzanalyse. Ein Text wird in Sequenzen aufgelöst, um sich gedankenexperimentell zu fragen, welche möglichen sinnvollen Anschlüsse eine Sequenz eröffnet. Diese Überlegungen werden im Verlauf einer Interpretationssitzung als Lesarten festgehalten, um im Fortgang der Sequenzanalyse nachzuvollziehen, welche konkreten Anschlussäußerungen im Material gewählt wurden.⁸ Die realisierte Bearbeitung eines Themas gibt damit Aufschluss über eine bestimmte typische Strukturierungsweise im Denken und Handeln der Beforschten. Von einem Strukturierungsmuster oder einer latenten Sinnstruktur wird jedoch erst gesprochen, wenn Themen oder Situationen (in einem Interview oder Protokoll) wiederkehrend und regelhaft von den Beforschten auf die gleiche Weise strukturiert wurden. Anschauliche Beispiele für rekonstruktive Vorgehensweisen finden sich für die objektive

⁸ Die dokumentarische Methode unterscheidet sich von der objektiven Hermeneutik in diesem Interpretationsschritt darin, dass sie neben gedankenexperimentell entwickelten Lesarten auch andere empirische Fälle heranzieht, um weitere sinnvolle Anschlüsse zu berücksichtigen (Bohnsack 2014; Nohl 2013).

Hermeneutik etwa in Wernet (2000), für die Narrationsanalyse in Küsters (2009), für die dokumentarische Methode in Bohnsack (2014) und Nohl (2013) oder für diese und andere Verfahren in Przyborski und Wohlrab-Sahr (2014).

In diesem Zusammenhang lässt sich noch eine weitere Schwierigkeit bei der Verknüpfung von Text Mining-Verfahren mit rekonstruktiven Herangehensweisen benennen. Die unhinterfragte besondere Stärke des Text Minings besteht in der automatisierten Strukturierung größerer Textmengen. Stulpe und Lemke (2015) gehen sogar weiter, wenn sie argumentieren, dass sich damit die Aussagekraft sozialwissenschaftlicher Textanalysen erhöhen lässt. „Je größer der Umfang des Analysematerials, desto größer die Reichweite und Validität der daraus ableitbaren Aussagen über die soziale Wirklichkeit“ (Stulpe; Lemke 2015: 31). Diese Schlussfolgerung kann für Inhalts- und teilweise für Diskursanalysen Geltung beanspruchen, aber nicht für rekonstruktive Verfahren. Latente Sinnstrukturen werden anhand von Regelmäßigkeiten in der Herstellung eines Themas in einem Interview oder Text rekonstruiert und teilweise durch Vergleiche mit anderen Textquellen spezifiziert. Die Plausibilität der Rekonstruktion steigt jedoch nicht mit der Ausweitung des Analysematerials.

Zusammengefasst haben die Vertreterinnen und Vertreter von Text Mining-Verfahren zugespitzt formuliert eine unpräzise bzw. unvollständige Vorstellung von der qualitativen Sozialforschung (insbesondere in Deutschland). In der Reduktion der qualitativen Methoden auf die Inhaltsanalyse werden viele der qualitativ Forschenden nicht adressiert. Letzteren geht es in der Regel um mehr als das Zusammenfassen und Explizieren von Inhalten und Bedeutungen. Vor allem die Verfahren der rekonstruktiven Textinterpretation haben sich damit auseinandergesetzt, wie sich anhand von Texten Erkenntnisse über die soziale Wirklichkeit generieren lassen (Bohnsack 2014; Schütze 1983; Przyborski 2004). Es muss daher nicht mehr gezeigt werden, dass sich in Texten etwas zum „text-jenseitigen Erkenntnisgegenstand finden lässt“ (Stulpe; Lemke 2015: 19). Texte sind als Erkenntnisquelle für sozial wirksame Strukturierungen in Deutungs- und Handlungsweisen ein fester Bestandteil rekonstruktiver Verfahren (etwa in der Narrationsanalyse, dokumentarische Methode, objektive Hermeneutik).

Neben der unzureichenden Durchdringung der textanalytischen qualitativen Methoden sehe ich ein Hauptproblem für die Einbindung von Text Mining-Verfahren in die rekonstruktive Sozialforschung darin, dass die ausgewiesene inhaltsanalytische Stärke solcher Verfahren allein für hermeneutische Herangehensweisen zu wenig ist. Der Verweis auf diesen Vorzug wird sogar zum größten Hindernis, wenn man berücksichtigt, dass sich innerhalb der qualitativen Sozialforschung im Besonderen die rekonstruktiv Forschenden von der qualitativen Inhaltsanalyse abgrenzen.

5 Qualitative Inhaltsanalyse und rekonstruktive Sozialforschung

Die qualitative Inhaltsanalyse (vor allem Mayring 2010) wird angewendet, um Textinhalte zusammenzufassen, die Bedeutung von Begriffen zu explizieren oder um Codes und Kategorien zu bilden. Sie wird daher in der Überblicksliteratur zu qualitativen Methoden behandelt (etwa Flick u.a. 2013; Lamnek 2010). In Methodenbüchern mit einer rekonstruktiven Ausrichtung sind qualitative Inhaltsanalysen als Auswertungsverfahren hingegen nicht vertreten (siehe bei Bohnsack 2014; Hitzler; Honer 1997; Kleemann; Krähnke; Matuschek 2013; Przyborski; Wohlrab-Sahr 2014; zumindest als Gegenhorizont bei Rosenthal 2014), was mit einer anderen methodologischen Grenzziehung einhergeht. Während erstgenannte Methodenbücher zwischen den quantitativen und qualitativen Methoden unterscheiden und damit grob die deduktiven Analysen großer numerischer Datenmengen auf der einen Seite und die Interpretation von einzelnen Beobachtungen und Texten auf der anderen Seite meinen, sehen die Vertreterinnen und Vertreter rekonstruktiver Methoden den Unterschied in der Vorgehensweise. Demnach ziehen quantitativ Forschende Theorien und Vorwissen heran, um ihren Untersuchungsgegenstand strukturiert zu beobachten. Rekonstruktiv Forschende hingegen lassen sich offen auf ihren Forschungsgegenstand ein und arbeiten heraus, wie die Beobachteten ihre Äußerungen und Handlungen strukturieren.⁹ So begründen Przyborski und Wohlrab-Sahr ihre Auslassung der qualitativen Inhaltsanalyse in ihrem Methodenlehrbuch wie folgt: „Die qualitative Inhaltsanalyse [...] klassifiziert u. E. eher als dass sie Sinnstrukturen rekonstruiert, sie ist nicht in der Lage bzw. nicht darauf angelegt, implizite Bedeutungen, wie sie in der Art und Weise einer Formulierung oder einer Interaktionssequenz zum Ausdruck kommen können, zu erfassen“ (Przyborski; Wohlrab-Sahr 2014: 189, Hervorhebung im Original). Vor diesem Hintergrund wird deutlich, dass qualitative Inhaltsanalysen zwar zu den qualitativen Methoden gehören, da sie mit kleinen Textbeständen operieren und inhaltliche Verdichtungen oder theoretisch informierte Kategorien produzieren. Sie zählen aber nicht zu den rekonstruktiven Verfahren.

Zudem wird von den Vertreterinnen und Vertretern der qualitativen Inhaltsanalyse selbst eine größere Nähe zu den quantitativen Methoden dargestellt. Mayring (2010) versteht beispielsweise das Verfahren explizit als Vorstufe zu quantitativen Untersuchungen. Auch Kuckartz (2014) oder Kelle (2014) ver-

⁹ Diese Unterscheidung findet sich auch in der Ethnografie nach Breidenstein u.a. (2015).

stärken die Relationierung durch den Mixed Method-Ansatz, wonach Ergebnisse der qualitativen ungebrochen in die quantitative Forschung übernommen werden können und umgekehrt. Dies gelingt aber nur, wenn qualitative Studien Standardisierungen erzeugen, die als Vorleistungen in numerisch-strukturierte Untersuchungen eingehen, wenn unverständliche Varianzen mit qualitativen Daten aufgeklärt, Einstellungen untersucht oder quantitative Ergebnisse inhaltlich angereichert werden sollen.

Für die Text Mining-Verfahren bedeutet folglich eine herausgehobene Anschlussfähigkeit an qualitative Inhalts- und Bedeutungsanalysen, dass die Analysewerkzeuge in weiten Teilen der qualitativen Sozialforschung unbeachtet bleiben. Die Text Mining-Verfahren müssten daher zeigen, wie sie sich produktiv etwa für rekonstruktive Analysen einsetzen lassen. Der folgende Abschnitt diskutiert daher mögliche Anschlüsse und macht gleichzeitig deutlich, dass eine Methodenkombination nicht komplementär ist, sondern optional. Text Mining-Verfahren können rekonstruktive Analysen nicht ergänzen und umgekehrt, aber gerade bei großen Textkollektionen können Text Mining-Verfahren Hinweise auf Texte liefern, die zu aufschlussreichen Feinanalysen führen können.

6 Verwendungsmöglichkeiten von Text Mining in der rekonstruktiven Sozialforschung

Text Mining-Verfahren könnten sich in der qualitativen Sozialforschung darauf konzentrieren, Anschlüsse zu den Inhalts- und Bedeutungsanalysen weiter auszubauen. Wie eine produktive Kombination aussehen kann, haben in jüngster Zeit etwa die Themenhefte der Zeitschriften *Poetics* (Mohr; Bogdanov 2013) und *Big Data & Society* (Wagner-Pacific; Mohr; Breiger 2015), der Reviewartikel von Evans und Aceves (2016) oder der Sammelband *Text Mining in den Sozialwissenschaften* (Lemke; Wiedemann 2015) gezeigt. Andere Bereiche der qualitativen Sozialforschung (z. B. Ethnografie, rekonstruktive Verfahren) bleiben damit aber außen vor. Dieser Abschnitt des Beitrages skizziert hingegen, wie das Text Mining auch für rekonstruktive Textanalyseverfahren hilfreich sein kann. Ich möchte an dieser Stelle zumindest darauf hinweisen, dass sie – wie Wiedemann und Lemke (2015) zu Recht bemerken – rekonstruktive Interpretationen nicht ersetzen, aber in manchen Fällen den Prozess der Text- und Datenerschließung unterstützen können.

Dafür ist relevant, dass eine gemeinsame Schnittmenge zwischen rekonstruktiven Verfahren und dem Text Mining besteht. So geht auch den rekonstruktiven Interpretationen eine Verständigung über den Inhalt und die Themen der zu analysierenden Texte voraus. Die Interpretierenden klären häufig in einem ersten Schritt, was in einem Gespräch oder in einem Text gesagt wird. So steht im Interpretationsprozess der objektiven Hermeneutik (Oevermann u.a. 1979; Wernet 2000) am Anfang ein Austausch über die verschiedenen Themen in einem Text oder Interview sowie über die Common Sense-Bedeutung einer Verhaltensweise oder Aussage,¹⁰ um daran gedankenexperimentell variierende Anschlüsse zu formulieren. Ebenso werden in der Narrationsanalyse (Schütze 1983) und Biographieforschung (Rosenthal 2014) neben den Textsorten die Themen bestimmt. Zu den ersten Schritten der Interpretation gehört, den thematischen Verlauf einer Erzählung herauszuarbeiten. Dies gilt auch für die formulierende Interpretation in der dokumentarischen Methode (Bohnsack 2014; Nohl 2013), das heißt, die Interpretierenden verständigen sich über die Common Sense-Bedeutung von Äußerungen, einer Handlungssequenz oder Erzählung. Erst im Anschluss konzentrieren sich die rekonstruktiven Interpretationen auf die Art und Weise, wie etwas getan oder gesagt wird, um über Regelmäßiges in den Herstellungsweisen auf ein Strukturierungsprinzip zu schließen. Im Unterschied zur qualitativen Inhaltsanalyse gehen rekonstruktive Interpretationen also über die zusammenfassenden und explizierenden Analysen hinaus.

Rekonstruktive Interpretationen sind demnach aber auch offen für thematische Variationen – zumindest als Ausgangspunkt für rekonstruktive Feinanalysen. So schreibt Rosenthal: „eine inhaltsanalytische Vorauswertung des vorliegenden Gesamtkorpus an Texten [...] kann zur ersten Sichtung des Materials und zur Bildung einer theoretischen Stichprobe für eine weitere Auswertung dienen“ (Rosenthal 2014: 212). Text Mining-Verfahren können insbesondere bei großen Textkollektionen¹¹ einen Beitrag leisten, da sie durch automatisch identifizierte thematische Unterschiede Ansatzpunkte und Indizien liefern, bestimmte Dokumente und Texte für Feinanalysen auszuwählen (vgl. Mohr; Bogdanov 2013; Stulpe; Lemke 2015). Die Bedeutung der computergestützten Verfahren besteht also in erster Linie darin, das Material großer Datenbestän-

10 Dieser Schritt findet sich zwar nicht in den Einführungstexten, ist aber eine unbedingte Voraussetzung, um in einer Interpretationsgruppe Geschichten zu formulieren und zu klären, ob es sich um angemessene Kontextualisierungen handelt.

11 Bei wenigen Texten ist die Aussagekraft von Topic-Modellen beschränkt. Eine mögliche Variante zur Steigerung der Qualität extrahierter Themen zieht beispielsweise automatisch Dokumente aus externen Wissenskorpora hinzu (Tran u.a. 2013).

den zu erschließen (etwa in Archiven oder umfassenden Forschungsarbeiten). Für rekonstruktive Verfahren hat dieser Vorzug grundsätzlich keine Bedeutung, da eine nachvollziehbare und plausible Rekonstruktion eines Strukturierungsmusters bereits mit relativ kleinen Textmengen möglich ist. Es gibt jedoch einerseits Forschungsprojekte, die weit mehr Daten erheben, als sie letztlich in ihrer Auswertung berücksichtigen, und andererseits wird derzeit der Aufbau von Infrastrukturen zur Sicherung qualitativer Daten vorangetrieben (siehe etwa Medjedovic 2011). Der Ausbau wird beispielsweise dadurch verstärkt, dass Einrichtungen der Forschungsförderung wie die Deutsche Forschungsgemeinschaft (DFG) den Umgang mit den im Projekt erzielten Forschungsdaten nachfragen oder Forschungseinrichtungen wie das Soziologische Forschungsinstitut (SOFI) in Göttingen damit begonnen haben, ihre qualitativ erhobenen Interviews zu digitalisieren.¹² Unter solchen Voraussetzungen kann es schließlich auch für rekonstruktive Herangehensweisen interessant sein oder relevant werden, größere Datenbestände im Forschungsprozess zu berücksichtigen.

Grundsätzlich bleibt dabei für das methodisch-kontrollierte und nachvollziehbare Vorgehen die Auswahl kontrastierender Fälle anhand von Beobachtungen im Feld und eigenen Interpretationen eine unerlässliche Strategie (Glaser; Strauss 1967). Zusätzlich können statistisch nachgewiesene Beziehungen zwischen Wörtern oder Themenclustern in den Textkollektionen überraschende Zugänge bieten. So schreiben Mohr und Bogdanov zur Anwendung von Topic-Modellen in den Sozialwissenschaften: „hermeneutic work [...] is still required in order to read and to interpret the meanings that operate within a textual corpus“. Ungeachtet dessen würden Topic-Modelle eine Art Objektiv („lens“) bieten, „that allows researchers working on a problem to view a relevant textual corpus in a different light and at a different scale“ (Mohr; Bogdanov 2013: 560; siehe auch DiMaggio u.a. 2013: 582). Der Vorzug der automatischen Verfahren besteht also allein darin, Anregungen zu geben, auch unvorhergesehenen Pfaden zu folgen. Sie ersetzen aber keinesfalls die empirisch begründete Vorgehensweise minimaler und maximaler Kontrastierungen, denn es gibt keine Gewissheit, dass das Text Mining das Datenmaterial richtig strukturiert und wiedergibt.

Die Anwendungsmöglichkeiten von Text Mining-Verfahren für rekonstruktive Analysen gewinnen des Weiteren an Bedeutung bei zeitlich und inhaltlich aufeinander aufbauenden Textkollektionen; gerade bei Untersuchungen, die sich weniger auf die thematische Verteilung des gesamten Textkorpus konzent-

¹² Siehe BMBF-Verbundprojekt: Re-SozIT: „Gute Arbeit“ nach dem Boom (Laufzeit: 10/2012–09/2015).

rieren, sondern auf die thematische Struktur über die Zeit. Exemplarische Studien zeigen, wie Text Mining-Verfahren thematische Verschiebungen in wissenschaftlichen Zeitschriften offenlegen (Blei; Lafferty 2006; Rosvall; Bergstrom 2010). Ebenso nutzt Förster (2015) in ihrer Studie zum Thema „Folter“ die Kookkurrenzanalyse, um auf der Basis veränderter begrifflicher Netzwerke mit diesem Wort zwei Zeiträume zu unterscheiden und jeweils relevante Artikel inhaltlich zusammenzufassen. Es dürfte klar sein, dass weder das signifikante gemeinsame Auftreten von Wörtern noch die Zusammenfassung mehrerer Texte ausreichen, ein generatives Prinzip zu rekonstruieren. In großen zeitlich und sachlich zusammenhängenden Textkollektionen (etwa Briefwechseln, E-Mail-Korrespondenzen, Blog-Kommentaren) kann aber das wiederkehrende gemeinsame Auftreten von Wörtern oder umgekehrt ihr Ausbleiben Hinweise liefern, dass eine rekonstruktive Feinanalyse von Texten oder Textpassagen sinnvoll ist. Möglicherweise deuten sich in der Variation an Themen veränderte typische Denkweisen an. Natürlich verändern sich die Themen auch mit den wechselnden Akteurinnen und Akteuren, technischen Neuerungen, Krisen, Kriegen oder Naturkatastrophen. Mit Mannheim (1964, 1980) ist es aber für eine bestimmte Gruppe oder Zeit charakteristisch, Ereignisse auf eine spezifische Art und Weise (perspektivisch) wahrzunehmen und zu beurteilen, unter anderem durch die Verwendung bestimmter Begriffe oder durch Beziehungen zu übergreifenden Themen. So werden einerseits in Texten einer Epoche bestimmte Themen mit einer anderen Gewichtung und Begrifflichkeit als vorher und danach behandelt. Andererseits dokumentiert sich das Charakteristische weniger in der Themenzusammensetzung und typischen Wortkombinationen, sondern darin, wie etwas sprachlich geäußert oder visuell gestaltet wird (Mannheim 1964). Das Spektrum an Themen oder Mustern und Regelmäßigkeiten beim Gebrauch von Wörtern und Phrasen allein liefert also noch keinen Beleg für bedeutungsgenerierende Regeln oder Strukturierungsprinzipien. Dazu bedarf es rekonstruktiver Feinanalysen. Die verschiedenartigen Gewichtungen von Begriffen in unterschiedlichen Themen können aber eine Spur sein, Zugänge zu typischen Denk-, Bewertungs- und Handlungsweisen zu finden.

7 Fazit

Mit den wachsenden digitalen Daten- und Textbeständen steht auch die empirische Sozialforschung vor der Herausforderung, mit angemessenen Erhebungsmethoden und Analyseverfahren zu reagieren. Während Savage und Burrows (2007) noch dazu aufforderten, sich als Soziologie an der Entwicklung adäqua-

ter Methoden zu beteiligen, liegen heute bereits verschiedene Möglichkeiten des Data und Text Minings in großen Daten- und Textkollektionen vor. Ein nächster Schritt scheint nun zu sein, solche Verfahren in die empirische Sozialforschung einzubinden. Insbesondere die Vertreterinnen und Vertreter von Text Mining-Verfahren sehen für die qualitative Ausrichtung einen Mehrwert und bemühen sich, Anschlussmöglichkeiten aufzuzeigen.

Dieser Beitrag beschäftigte sich mit diesem Anspruch und dem darin angelegten ungenauen Verständnis qualitativer Forschung im Allgemeinen und den rekonstruktiven Verfahren im Speziellen. Als Hauptproblem wurde die Reduktion der qualitativen Sozialforschung auf Inhalts- und Bedeutungsanalysen ausgemacht. So bieten diese Verfahren zwar eine zusätzliche Option, Inhalte zu differenzieren und semantische Bedeutungsgehalte zu identifizieren, es bleibt jedoch unklar, ob und wie solche Ergebnisse etwa für rekonstruktive Verfahren förderlich sind. Im Beitrag wurde argumentiert, dass das Text Mining zumindest optional im Prozess rekonstruktiver Untersuchungen für die Text- und Datenerschließung von großen Textkollektionen eingesetzt werden könnte.

Eine solche optionale Anwendung bedarf letztlich gemeinsamer Anstrengungen aufseiten rekonstruktiv Forschender sowie der Informatik, wobei die disziplinäre Differenz zwischen Informatik und qualitativer Sozialforschung nicht zu unterschätzen ist. Bereits DiMaggio (2015) als auch Philipps (2017, 2018) verweisen auf unterschiedliche disziplinäre Herangehensweisen und Probleme in der Anwendung von Text Mining-Verfahren. Diese Kluft fällt hinsichtlich der erworbenen Fertigkeiten und Selbstverständnisse in der Informatik und der rekonstruktiv vorgehenden Sozialforschung sogar noch deutlicher aus. Die Unterschiede lassen sich nicht durch gemeinsame Gespräche ausräumen. Dafür ist vielmehr erforderlich, längerfristig miteinander zu kooperieren und voneinander zu lernen.

Grundsätzlich gilt, dass die rekonstruktive Sozialforschung auch ohne Text Mining relevante Resultate hervorbringt. Sowohl mit Text Mining- also auch mit Data Mining-Verfahren besteht aber auch die Chance, neue Möglichkeiten der Datenerhebung für rekonstruktive Interpretationen zu erschließen. Die empirische Grundlage rekonstruktiver Analyseverfahren sind Kommunikations- und Handlungssequenzen. Solche Daten werden bisher in der Regel durch audio(-visuelle) Aufzeichnungen natürlicher Gesprächssituationen, Gruppendiskussionen, Erzählungen oder Verhaltensweisen generiert. Einerseits lässt sich die wachsende Menge solcher Daten mit Data und Text Mining-Verfahren vorselektieren. Andererseits bieten gerade digitale Werkzeuge weitere Optionen der Datenerhebung, beispielsweise die Nutzung von Metadaten im Internet (Marres 2017; Rogers 2013). Solche Daten von Daten werden beständig beim Gebrauch

des Webs gespeichert. Unter dem Vorbehalt der Bereitschaft und Zustimmung der Nutzerinnen und Nutzer könnten digitale Werkzeuge schließlich auch entwickelt und eingesetzt werden (Bail 2017; Golder; Macy 2014), um entsprechende Daten zum Nutzungsverhalten zu erheben und für die Analyse aufzubereiten.

Letztlich wird die Zukunft zeigen, wie sich die rekonstruktive Sozialforschung zur digitalen Durchdringung der Gesellschaft verhält. Dazu braucht es nicht nur weitere Anstrengungen, etwa die Spezifik digitaler Medien in der Textproduktion (Meißner 2015) oder in hochgeladenen Bildern (Schreiber; Kramer 2016) interpretativ einzuholen, sondern auch hinsichtlich der Einbindung digitaler Werkzeuge in die rekonstruktive Sozialforschung. Wie gehen wir diese Herausforderung an? Wo können wir von den Bemühungen in den Digital Humanities profitieren? Wo liegen jedoch markante Unterschiede? Welche konkreten Ansprüche ergeben sich daraus für den sinnvollen Einsatz digitaler Instrumente wie Text Mining-Verfahren und anderer?

Literatur

- Ahonen, P. (2015): Institutionalizing Big Data methods in social and political research, in: *Big Data & Society*, July–Dec., 1-12
- Bail, C.A. (2014): The cultural environment: measuring culture with big data, in: *Theory & Society*, 43 (3-4), 465-482
- Bail, C.A. (2017): Taming Big Data Using App Technology to Study Organizational Behavior on Social Media, in: *Sociological Methods & Research*, 46 (2), 189-217
- Biber, D.; Connor, U.; Upton, T.A. (2007): *Discourse on the Move*, Amsterdam.
- Blei, D.M. (2012): Probabilistic topic models, in: *Communications of the ACM*, 44, 77-84
- Blei, D.M.; Lafferty, J.D. (2006): Dynamic topic models, in: *Proceedings of the 23rd international conference on Machine learning*, ACM, 113-120
- Bohnsack, R. (2014): *Rekonstruktive Sozialforschung. Einführung in qualitative Methoden*, 9. Aufl., Opladen und Toronto
- Bohnsack, R.; Nentwig-Gesemann, I.; Nohl, A.-M. (Hrsg.) (2013): *Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung*, 3. Aufl., Wiesbaden
- Boyd, D.; Crawford, K. (2012): Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon, in: *Information, Communication & Society*, 15 (5), 662-679
- Breidenstein, G.; Hirschauer, S.; Kalthoff, H.; Nieswand, B. (2015): *Ethnografie: Die Praxis der Feldforschung*, 2. Aufl., Konstanz und München
- Brett, M.R. (2012): Topic Modeling: A basic introduction, in: *Journal of Digital Humanities*, 2 (1), 2-16
- Bubenhofer, N.; Scharloth, J. (2015): Maschinelle Textanalyse im Zeichen von Big Data und Data-driven Turn – Überblick und Desiderate, in: *Zeitschrift für germanistische Linguistik*, 43 (1), 1-26

- Cecire, N. (2011): Introduction: Theory and the Virtues of Digital Humanities, in: *Journal of Digital Humanities*, 1 (1), 45-53
- Denzin, N.K.; Lincoln, Y.S. (2011): *The Sage handbook of qualitative research*, 4. Aufl., Thousand Oaks, London, New Dehli und Singapore
- Diaz-Bone, R.; Weischer, C. (Hrsg.) (2014): *Methoden-Lexikon für die Sozialwissenschaften*, Wiesbaden
- DiMaggio, P. (2015): Adapting computational text analysis to social science (and vice versa), in: *Big Data & Society*, December, 1-5
- DiMaggio, P.; Nag, M.; Blei, D. (2013): Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government art fundings, in: *Poetics*, 41 (6), 570-606
- Dumm, S.; Niekler, A. (2015): Methoden, Qualitätssicherung und Forschungsdesign, in: Lemke, M.; Wiedemann, G. (Hrsg.): *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, Wiesbaden, 89-116
- Evans, J.A.; Aceves, P. (2016): Machine Translation: Mining Text for Social Theory, in: *Annual Review of Sociology*, 42, 21–50
- Flick, U.; von Kardorff, E.; Keupp, H.; von Rosenstiel, L.; Wolff, S. (2013): *Handbuch qualitative Sozialforschung: Grundlagen, Konzepte, Methoden und Anwendungen*, 3. Aufl., Weinheim und Basel
- Förster, A. (2015): Der Folterdiskurs in den deutschen Printmedien, in: Lemke, M.; Wiedemann, G. (Hrsg.): *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, Wiesbaden, 139-166
- Friese, S. (2014): *Qualitative data analysis with ATLAS.ti*, London
- Glaser, B.G.; Strauss, A.L. (1967): *The discovery of ground theory: Strategies for qualitative research*, Chicago
- Goldberg, A. (2015): In defense of forensic social science, in: *Big Data & Society*, July–December, 1-3
- Golder, S.A.; Macy, M.W. (2014): Digital Footprints: Opportunities and Challenges for Online Social Research, in: *Annual Review of Sociology*, 40, 129-152
- González-Bailón, S.; Wang, N.; Rivero, A.; Borge-Holthoefer, J.; Moreno, Y. (2014): Assessing the bias in samples of large online networks, in: *Social Networks*, 38 (1), 16-27
- Heyer, G.; Quasthoff, U.; Wittig, T. (2006): *Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*, Herdecke und Bochum
- Hitzler, R.; Honer, A. (1997): *Sozialwissenschaftliche Hermeneutik. Eine Einführung*, Opladen
- Hopkins, D.J. (2013): The exaggerated life of death panels: the limits of framing effects in the 2009–2012 health care debate, Working Paper, SSRN
- Ignatow, G. (2016): Theoretical Foundations for Digital Text Analysis, in: *Journal for the Theory of Social Behaviour*, 46 (1), 104-120
- Ignatow, G.; Mihalcea, R. (2016): *Text Mining. A Guidebook for the Social Sciences*. Los Angeles u.a.
- Ignatow, G.; Williams, A.T. (2011): New media and the ‘anchor baby’ boom, in: *Journal of Computer-mediated Communication*, 17 (1), 60–76
- Jost, C.; Klug, D.; Schmidt, A.; Reautschnig, A.; Neumann-Braun, K. (2013): *Computergestützte Analyse von audiovisuellen Medienprodukten*, Wiesbaden
- Kath, R.; Schaal, G.; Dumm, S. (2015): New Visual Hermeneutics, in: *Zeitschrift für germanistische Linguistik*, 43 (1), 27–51

- Keyling, T. (2014): Automatisierte Inhaltsanalyse, in: Welker, M.; Taddicken, M.; Schmidt, J.-H.; Jackob, N. (Hrsg.): Handbuch Online-Forschung. Sozialwissenschaftliche Datengewinnung und -auswertung in digitalen Netzen, Köln, 233-254
- Kelle, U. (2014): Mixed Methods, in: Baur, N.; Blasius, J. (Hrsg.): Handbuch Methoden der empirischen Sozialforschung, Wiesbaden, 153-166
- Keller, R. (2007): Diskursforschung, Wiesbaden
- Keller, R.; Hirsland, A.; Schneider, W.; Viehöver, W. (2011): Handbuch Sozialwissenschaftliche Diskursanalyse. Band 1: Theorien und Methoden, Wiesbaden
- Kleemann, F.; Krähnke, U.; Matuschek, I. (2013): Interpretative Sozialforschung. Eine praxisorientierte Einführung, 2. Aufl., Wiesbaden
- Kuckartz, U. (2014): Mixed Methods. Methodologie, Forschungsdesigns und Analyseverfahren, Wiesbaden
- Kuckartz, U. (2010): Einführung in die computergestützte Analyse qualitativer Daten, 3. Aufl., Wiesbaden
- Kuckartz, U.; Grunenberg, H.; Dresing, T. (2007): Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis, Wiesbaden
- Küsters, I. (2009): Narrative Interviews. Grundlagen und Anwendungen, Wiesbaden
- Lamnek, S. (2010): Qualitative Sozialforschung, 5. Aufl., Weinheim und Basel
- Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabási, A.-L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; Jebara, T.; King, G.; Macy, M.; Roy, D.; Van Alstyne, M. (2009): Life in the network: the coming age of computational social science, in: Science, 323 (5915), 721-723
- Lenke, M. (2014): Kookkurrenzanalyse. ePol Text Mining-Verfahren, Serie „Atomenergiediskurs“, Modul 2/5, Hamburg, Leipzig
- Lenke, M.; Wiedemann, G. (2015): Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse, Wiesbaden
- Mannheim, K. (1980): Strukturen des Denkens, Frankfurt am Main
- Mannheim, K. (1964): Beiträge zur Theorie der Weltanschauungs-Interpretation, in: ders.; Wolff, K.H. (Hrsg.): Wissenssoziologie. Auswahl aus dem Werk, Berlin, 91-154
- Marres, N. (2017): Digital Sociology. The reinvention of social research, Hoboken, New Jersey
- Mayring, P. (2010): Qualitative Inhaltsanalyse, 11. Aufl., Wiesbaden
- McFarland, D.A.; Lewis, K.; Goldberg, A. (2015): Sociology in the era of big data: The ascent of forensic social science, in: American Sociologist, DOI 10.1007/s12108-015-9291-8
- Medjedović, I. (2011): Secondary Analysis of Qualitative Interview Data: Objections and Experiences. Results of a German Feasibility Study [45 paragraphs], in: Forum Qualitative Sozialforschung, 12, Art. 10, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1103104>
- Meißner, S. (2015): Die Medialität und Technizität internetbasierter Daten, in: Schirmer, D.; Sander, N.; Wenninger, A. (Hrsg.): Die qualitative Analyse internetbasierter Daten, Wiesbaden, 33-49
- Mohr, J.W.; Bogdanov, P. (2013): Introduction – Topic models: What they are and why they matter, in: Poetics, 41 (6), 545–569
- Nohl, A.-M. (2013): Interview und dokumentarische Methode. Anleitungen für die Forschungspraxis, 4. Aufl., Wiesbaden
- Oevermann, U. (2001): Die Struktur sozialer Deutungsmuster – Versuch einer Aktualisierung, in: Sozialer Sinn, 2 (1), 35-81
- Oevermann, U.; Allert, T.; Konau, E.; Krambeck, J. (1979): Die Methodologie einer „objektiven Hermeneutik“ und ihre allgemeine forschungslogische Bedeutung in den Sozialwissen-

- schaften, in: Soeffner, H.-G. (Hrsg.): Interpretative Verfahren in den Sozial- und Textwissenschaften, Stuttgart, 352-434
- Philipps, A. (2017): Brauchen wir eine Soziologie des Digitalen?, in: *Soziologie*, 46 (4), 403-416
- Philipps, A. (2018): How to sort out uncategorisable documents for interpretive social science? On limits of currently employed text mining techniques, in: *Proceedings of the 2nd Conference on Advanced Research Methods and Analytics*, Valencia.
- Philipps, A., Zerr, S.; Herder, E. (2017): The representation of street art on Flickr. Studying reception with visual content analysis, in: *Visual Studies*, 32 (4), 382-393
- Przyborski, A. (2004): Gesprächsanalyse und dokumentarische Methode. Qualitative Auswertung von Gesprächen, Gruppendiskussionen und anderen Diskursen, Wiesbaden
- Przyborski, A.; Wohlrab-Sahr, M. (2014): *Qualitative Sozialforschung. Ein Arbeitsbuch*, 4. Aufl., München
- Reiter, N.; Frank, A.; Hellwig, O. (2014): An NLP-based cross-document approach to narrative structure discovery, in: *Literary and Linguistic Computing*, 29 (4), 583-605
- Rogers, R. (2013): *Digital Methods*, Cambridge
- Rosenthal, G. (2014). *Interpretative Sozialforschung. Eine Einführung*, 4. Aufl., Weinheim und München
- Rosvall, M.; Bergstrom, C.T. (2010): Mapping change in large networks, in: *PloS one*, 5, e8694
- Savage, M.; Burrows, R. (2007): The coming crisis of empirical sociology, in: *Sociology*, 41 (5), 885-899
- Scharkow, M. (2013): Thematic content analysis using supervised machine learning: An empirical evaluation using German online news, in: *Quality & Quantity*, 47 (2), 761-773
- Schreiber, M.; Kramer, M. (2016): „Verdammt schön“: methodologische und methodische Herausforderungen der Rekonstruktion von Bildpraktiken auf Instagram, in: *Zeitschrift für Qualitative Forschung*, 17 (1-2), 81-106
- Schütze, F. (1983): Biographieforschung und narratives Interview, in: *Neue Praxis*, 13 (3), 283-293
- Stulpe, A.; Lemke, M. (2015): Blended Reading, in: Lemke, M.; Wiedemann, G. (Hrsg.): *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, Wiesbaden, 17-61
- Tran, N.K.; Zerr, S.; Bischoff, K.; Niederée, C.; Krestel, R. (2013): Research and advanced technology for digital libraries, in: Aalberg, T.; Papatheodorou, C.; Dobрева, M.; Tsakonas, G.; Farrugia, C. (Hrsg.): *Topic cropping: leveraging latent topics for the analysis of small corpora*, Berlin und Heidelberg, 297-308
- Wagner-Pacifici, R.; Mohr, J.W.; Breiger, R.L. (2015): Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences, in: *Big Data & Society*, December, 1-11
- Wernet, A. (2000): *Einführung in die Interpretationstechnik der objektiven Hermeneutik*, Opladen
- Wiedemann, G. (2016): *Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany*, Wiesbaden
- Wiedemann, G. (2013): Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences [54 paragraphs], in: *Forum Qualitative Sozialforschung*, 14, Art. 13, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1302231>
- Wiedemann, G.; Lemke, M. (2015): Text Mining für die Analyse qualitativer Daten, in: Lemke, M.; Wiedemann, G. (Hrsg.): *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, Wiesbaden, 397-419

Wiedemann, G.; Niekler, A. (2015): Analyse qualitativer Daten mit dem „Leipzig Corpus Miner“, in: Lemke, M; Wiedemann, G. (Hrsg.): Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse, Wiesbaden, 63-88