

Schriftenreihe des Verbundprojekts
„Postdemokratie und Neoliberalismus“

Discussion Paper Nr.4

Methoden und Gütekriterien

Computergestützte Diskurs- und Inhaltsanalysen zwischen Sozialwissenschaft und Automatischer Sprachverarbeitung

Sebastian Dumm & Andreas Niekler

www.epol-projekt.de

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

UNIVERSITÄT LEIPZIG



HELMUT SCHMIDT
UNIVERSITÄT

Universität der Bundeswehr Hamburg

ISSN 2363-6335

Zitierweise:

Dumm, Sebastian / Niekler, Andreas (2014):

Methoden und Gütekriterien

Computergestützte Diskurs- und Inhaltsanalysen zwischen Sozialwissenschaft
und Automatischer Sprachverarbeitung

Schriftenreihe des Verbundprojekts Postdemokratie und Neoliberalismus

Discussion Paper Nr. 4

Helmut-Schmidt-Universität Hamburg (UniBw) und Universität Leipzig

Zusammenfassung

Die sozialwissenschaftlichen Methoden der Diskurs- und Inhaltsanalyse sind in ihrer Verarbeitungskapazität von großen Textmengen beschränkt. Mit den Verfahren der Automatischen Sprachverarbeitung können aber auch große Textmengen zur Beantwortung sozialwissenschaftlicher Fragestellungen eingesetzt werden. Das vorliegende Paper geht erstens der Frage nach, wie die Methoden der Sozialwissenschaft mit den Verfahren der Automatischen Sprachverarbeitung kombiniert werden können, und zweitens, wie bei einer solchen Kombination eine adäquate Qualitätssicherung mit Gütekriterien vorgenommen werden kann. Zur Beantwortung werden die Grundlagen der qualitativen und quantitativen Methoden der Diskurs- und Inhaltsanalyse und des Text Mining sowie deren Gütekriterien vorgestellt. Am Forschungsdesign des Projekts ePol wird beispielhaft eine Kombination der angesprochenen Methoden und Verfahren demonstriert und deren Qualitätssicherung dargestellt. Abschließend wird ein dreistufiges Vorgehen zur Identifikation eines Forschungsdesigns in der computergestützten qualitativen und quantitativen Inhaltsanalyse vorgeschlagen.

Abstract

Methods of discourse and content analysis in Social Sciences are limited in their processing capacity regarding the amounts of text. The usage of computational methods like Natural Language Processing and Machine Learning allow for examination of Social Science questions with large digital document collections. Following this premise this paper discusses two questions. How should a researcher combine methods of Computer Science and Social Science? How to ensure appropriate quality criteria for quality management of the research process. To answer these questions the paper describes the foundations of qualitative and quantitative methods for discourse and content analysis. Furthermore, it discusses relevant methods of Natural Language Processing, Text Mining and Machine Learning. Quality criteria for each method are discussed and a methodology for an integrated research design is proposed. An example application, the *ePol project*, is shown with respect to quality assurance in computational qualitative and quantitative content analysis. Finally, a three step approach is proposed to develop computational qualitative and quantitative content analysis research designs.

Inhaltsverzeichnis

1. Einleitung.....	5
2. Diskurs- und Inhaltsanalyse als Anwendungsfall der digital humanities	7
2.1 Forschungsstrategien: Deduktion oder Induktion	8
2.2 Diskurs- und Inhaltsanalyse in den Sozialwissenschaften	10
2.3 (Halb-) automatische Inhaltsanalysen mit Text Mining.....	12
2.4 Forschungsdesign zur (halb-) automatischen Argumentidentifikation.....	14
3. Sozialforschung und Gütekriterien	17
3.1 Gütekriterien in der quantitativen und qualitativen Forschung	18
3.2 Gütekriterien der qualitativen und quantitativen Inhaltsanalyse	19
3.3 Gütekriterien des Text Mining.....	21
3.4 Anwendung der Gütekriterien auf das Forschungsdesign.....	26
4. Fazit.....	28
Literatur	30

Abbildungs- und Tabellenverzeichnis

Abbildung 1: Methodischer Forschungsablauf in drei Schritten	15
Abbildung: 2: Aufteilung der Trainingsdaten bei der Modellselektion und der n-fold Cross Validation.....	23
Tabelle 1: Tabelle übernommen aus Alpaydin (2008, S. 354)	25

Methoden und Gütekriterien

Computergestützte Diskurs- und Inhaltsanalysen zwischen Sozialwissenschaft und Automatischer Sprachverarbeitung

1. Einleitung

Durch die Digitalisierung von Texten und die zunehmende Verbreitung digitaler Medien besteht die Chance große Datenmengen empirisch auszuwerten und zur Beantwortung sozialwissenschaftlicher Fragestellungen einzusetzen. Die sozialwissenschaftlichen Methoden der Diskurs- und Inhaltsanalyse sind aber in der Verarbeitungskapazität von textuellen Daten beschränkt, wenn manuelle Verfahren eingesetzt werden. Um die Möglichkeiten umfangreicher Analysen, die sich aus der Digitalisierung von Textdokumenten ergeben, zu nutzen ist es sinnvoll diese Methoden mit Verfahren der Automatischen Sprachverarbeitung¹ zu kombinieren. Die Anwendung und Weiterentwicklung von Verfahren der Automatischen Sprachverarbeitung ermöglichen es dann umfangreiche digitale Textkorpora computergestützt in der sozialwissenschaftlichen Forschung zu nutzen². Spezielle Verfahren, die statistische Eigenschaften von Sprache sichtbar machen, werden im Folgenden mit dem Begriff Text Mining³ umschrieben. Mit Text Mining werden „computergestützte Verfahren für die semantische Analyse von Texten bezeichnet, die automatische bzw. semi-automatische Strukturierung von Texten, insbesondere sehr großen Mengen von Texten, unterstützen“⁴ (Heyer et al. 2006, S. 3).

¹ engl. Natural Language Processing

² Diese gemeinsamen und interdisziplinären Projekte können unter dem Begriff digital humanities zusammengefasst werden. Unter digital humanities wird der Ansatz der computergestützten Forschung im Bereich der Geistes- und Sozialwissenschaften verstanden. Dabei sollen Verfahren der Informatik unterstützend bei der Forschung eingesetzt werden. Dieser Bereich ist in den deutschsprachigen Raum erst in den letzten Jahren relevanter geworden. Zur Einordnung der Methode der digital humanities in der Politikwissenschaft siehe Wiedemann/Lemke/Niekler 2012.

³ Text Mining ist verwandt mit dem Begriff Data Mining, welcher eine Disziplin beschreibt, die statistische Eigenschaften von Daten verschiedenster Art sichtbar und nutzbar machen kann.

⁴ Erweiternd definieren Feldman und Sanger (2008, S. 1) Text Mining als „a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. In a manner analogous to data mining, text mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns.“ Zur weiteren Einführung in die Konzepte und Techniken des Text Mining siehe u.a. Heyer et al. 2006; Feldman/Sanger 2008; Buch 2008.

Um sozialwissenschaftliche Fragestellungen mit Verfahren des Text Mining bearbeiten zu können, ist eine interdisziplinäre Zusammenarbeit notwendig. Ein Beispiel für eine derartige Zusammenarbeit von Sozialwissenschaft und Automatischer Sprachverarbeitung ist das Forschungsprojekt „ePol: Postdemokratie und Neoliberalismus“. Das Projekt nutzt die Potenziale computergestützter Methoden zur Beantwortung einer aktuellen Fragestellung aus der Politischen Theorie. Konkret geht es um die Prüfung einer These des Postdemokratiediskurses⁵. In diesem Diskurs postulieren verschiedene Autoren einen qualitativen Wandel westlicher Demokratien (Crouch 2008, 2011; Brown 2006). Einen zentralen Stellenwert des Diskurses nimmt die These ein, dass der Prozess der Postdemokratisierung durch die zunehmende Bedeutung der Leitideen des Neoliberalismus in der politischen Sphäre, das heißt durch eine Ökonomisierung des Politischen, geprägt sei. Der Postdemokratiediskurs wird bisher vor allem auf theoretischer Ebene geführt, eine empirische Überprüfung dieser Thesen dieses Diskurses steht derzeit noch aus. Das Erkenntnisinteresse von ePol besteht darin, die Potenziale der computergestützten Analyse digitalisierter Texte zu nutzen, um einen möglichen Wandel des politischen Sprachgebrauchs unter dem Einfluss neoliberaler Leitideen empirisch nachzuweisen.

Welche Daten können aber für dieses Forschungsvorhaben genutzt werden? Eine geeigneter Untersuchungsgegenstand sind Zeitungsdatenbanken. Begründet werden kann dies mit zwei Vorannahmen aus der Politikwissenschaft, welche Einfluss auf die Forschungsmethode und den Forschungsgegenstand haben. Erstens sind politische Entscheidungen gegenüber der Öffentlichkeit begründungsbedürftig. Die öffentliche Rechtfertigung getroffener oder zu treffender Entscheidungen stellt die Legitimation im politischen Prozess her. Zweitens gelten Medien, speziell Zeitungen, als Archiv öffentlich ausgetragener Diskurse und in ihnen kann ein Wandel im öffentlichen Sprachgebrauch untersucht werden (vgl. Niehr 1999, S. 205). Wenn sich in den Medien als *Spiegel der Wirklichkeit* (Schulz 1989) also ein Wandel des politischen Prozesses manifestiert, dann ist in Zeitungen auch einen Wandel der Begründung politischer Entscheidungen, also der empirische Nachweis eines Wandels im Sprachgebrauch in der politischen Öffentlichkeit, nachweisbar.

Das vorliegende Paper beleuchtet die methodischen Anforderungen an ein adäquates Forschungsdesign für große Textkorpora. Dabei stellt sich erstens die Frage, wie sich die Verfahren des Text Mining mit den Methoden der Diskurs- und Inhaltsanalyse kombinieren lassen. Zur Beantwortung wird in Kapitel 2. die Notwendigkeit computergestützter Verfahren am Beispiel des Projektes ePol verdeutlicht, um dann die Forschungsstrategien der beteiligten Fachrichtungen, die Methoden der Diskurs- und Inhaltsanalyse sowie ausgewählte Verfahren des Text Mining vorzustellen. Abschließend wird mit dem Forschungsdesign von ePol exemplarisch eine Kombination dieser Verfahren demonstriert. Bei der Kombination dieser Verfahren stellt sich zweitens die Frage nach der

⁵ Die Postdemokratiedebatte diskutiert damit den Verfall zentraler demokratischer Werte und Prozesse für die westlichen Demokratien (zur Orientierung Schaal/Ritzi 2012; Ritzi 2013). In der politiktheoretischen Debatte wird der Postdemokratiediskurs u.a. von Sheldon S. Wolin (2008), Jacques Rancière (1995), Wendy Brown (2006) und Colin Crouch (2008, 2011) vertreten. Postdemokratische Theoretiker sehen einen Wandel in öffentlicher Politik durch den Bedeutungsgewinn neoliberaler Theorien, welcher sich im gesellschaftlichen Bereich durch eine zunehmende Ökonomisierung auch nicht wirtschaftsaffiner Politikfelder abzeichne. Dabei wirkt die „Dominanz des Marktes“ in der Art, dass die wirtschaftliche Rationalität als Handlungsmaßstab auch im politischen Bereich aufgenommen wird (vgl. Crouch 2011, S. 51).

Qualität des Forschungsablaufs und der Ergebnisse eines solchen Forschungsdesigns. Zur Beantwortung werden in Kapitel 3. die Güterkriterien beider Fachrichtungen diskutiert. Darauf aufbauend wird beispielhaft die Anwendung dieser Kriterien der Qualitätssicherung für das Forschungsdesign von ePol dargestellt.

2. Diskurs- und Inhaltsanalyse als Anwendungsfall der digital humanities

Aus Sicht der Sozialwissenschaften besteht ein Kapazitätsproblem bei der angemessenen Methodenwahl, wenn der Nachweis veränderter Begründungsmuster der Politik in sehr großen Textkorpora zu führen ist. Dies lässt sich am Beispiel des Forschungsprojekts ePol folgendermaßen begründen. Die im Postdemokratiediskurs behaupteten Veränderungen in der politischen Hegemonie der Gesellschaft, die einen Wandel der Begründungs- und Argumentationsstruktur zur Folge haben, sind langfristige Prozesse und müssen daher in einer umfangreiche Längsschnittstudie untersucht werden. Wenn der Wandel des Sprachgebrauchs, also die Hegemonialwerdung von neoliberalen Begründungen, nicht nur an einzelnen Texten, sondern repräsentativ für den gesamten politischen Sprachgebrauch, untersucht werden soll, dann ist neben einem langen Untersuchungszeitraum eine große Tiefenschärfe bei der Analyse des gesamten politischen Diskurses notwendig, um solch subtile Veränderungen nachzuweisen. Klassische inhalts- und diskursanalytische Methoden können diesen beiden Anforderung aufgrund ihrer beschränkten manuellen Verarbeitungskapazität aber nicht gänzlich gerecht werden. Das Projekt ePol arbeitet deshalb mit einer computergestützten qualitativen und quantitativen Inhalts- und Diskursanalyse.

Die für das Forschungsprojekt benötigten Zeitungsartikel stammen aus einer digitalisierten Sammlung von Zeitungstexten. Dieser Zeitungskorpus umfasst 3,5 Millionen Dokumente aus vier überregionalen Zeitungen (Die Zeit, FAZ, taz, Süddeutsche)⁶ in einen Zeitraum von über 60 Jahren (1949–2012)⁷ und ist damit für eine manuelle Auswertung unzugänglich. Zur Analyse solcher Textkorpora sind somit Verfahren des Text Minings unabdingbar, wenn nicht auf Stichprobenverfahren zurückgegriffen und durch eine Vollerhebung eine höhere Generalisierbarkeit der Ergebnisse hergestellt werden soll.

Die Verfahren des Text Mining⁸ ermöglichen eine quantitative Grobuntersuchung des gesamten Diskurses, ebenso wie die qualitative Durchdringung und Interpretation einzelner

⁶ Diese Zeitungen wurden aufgrund ihrer Funktion als Leitmedien der deutschen Zeitungslandschaft ausgewählt. Den Leitmedien wird nach Wilke ein nicht unerheblicher Einfluss auf die Gesellschaft und die gesellschaftlichen Eliten zugeschrieben (vgl. Wilke 1999, S. 302f.). Damit können die Entwicklungen in diesen Zeitschriften Aufschluss über die Veränderungen in der gesamten Medienlandschaft liefern.

⁷ Dabei sind die Zeitungsdaten im Zeitungskorpus nicht alle über den gesamten Zeitraum vorhanden. Nur der Zeitungskorpus der *Zeit* steht über den gesamten Analysezeitraum als Volltextkorpus zur Verfügung. *Die tageszeitung* ist als Volltextkorpus ab 1986 und die *Süddeutsche Zeitung* ab 1992 verfügbar. Bei der *Frankfurter Allgemeinen Zeitung* habe wurde eine Stichprobe gezogen, die 15 % aller redaktionellen Beiträge der Ressorts Politik, Wirtschaft und Feuilleton sowie alle Kommentare des Politikressorts von 1959–2011, enthält.

⁸ Die im Rahmen des Projekts eingesetzten Verfahren des Text Mining sind z.B. Frequenzanalysen mit den Worthäufigkeiten im Zeitungskorpus untersucht werden können. Diktionärbasierte Ansätze, die induktiv oder deduktiv erzeugte Wortlisten zur schlüsselwortbasierten Suchen nutzen. Kookkurrenzanalysen die das überzufällig häufige gemeinsame Auftreten von Wörtern aufzeigen und damit den Gebrauchskontext der Wörter erschließen. Topic Modelle durch die sich globale Begriffszusammenhänge erschließen lassen. Verfahren des

gefilterter Diskursaspekte (vgl. Wiedemann/Lemke/Niekler 2013, S. 105ff.). Diese Herangehensweise an große Textmengen kann mit Moretti (2000, 2007) auch als eine Art Vogelperspektive auf Textkorpora beschrieben werden, welche er als *distant reading* bezeichnet. Dabei ermöglichte dieser Zugriff auf Texte einen besseren Überblick und erhöht die Chance durch einen quantitativ größeren Textkorpus der Wirklichkeit gerechter zu werden. An eine solche Beobachtung aus der Vogelperspektive kann die detaillierte qualitative Analyse relevanter Aspekte angeschlossen werden, das *close reading*. Diese Prozesse des *distant reading* und *close reading* können dann beliebig kombiniert werden und ermöglichen so einen mehrstufiges Forschungsdesign, das Lemke und Stulpe als *blended reading* beschreiben (Lemke/Stulpe 2014).

Text Mining Verfahren können methodische Verfahren aus der klassischen Diskurs- und Inhaltsanalyse übernehmen, da die meist qualitativen Verfahren der Diskursanalyse auch an quantitative Methoden der computergestützten Textanalyse anschlussfähig sind (Keller/Viehöver 2006, S. 104). Dies gilt im Besonderen für die Unterstützung bei der Codierung und Klassifizierung von Text (Diaze-Bone/Schneider 2004, S. 459ff.). Ebenso wie für die qualitative Inhaltsanalyse gilt es auch für computergestützte Ansätze, „die Stärken der quantitativen Inhaltsanalyse beizubehalten und auf ihrem Hintergrund Verfahren systematischer qualitativ orientierter Textanalyse zu entwickeln“ (Mayring 2010, S. 48). Nach Gerhards besteht der Nachteil von qualitativen Inhaltsanalysen darin, nur eine geringe Aussagereichweite und damit eine beschränkte Repräsentativität zu besitzen. Der Vorteil von systematischen quantitativen Inhaltsanalysen großer Textkorpora ist dagegen, dass repräsentative formulierte Aussagen statistisch bearbeitet und interpretiert werden können. Die hermeneutische Tiefenschärfe der quantitativen Inhaltsanalyse bleibt aber hinter der qualitativen Inhaltsanalyse zurück (vgl. Gerhards 2004, S. 306).

Das ePol Forschungsdesign kombiniert die Vorteile der Methoden, um umfangreiche Textkorpora mit großer Tiefenschärfe zu analysieren. Um das Forschungsdesign einordnen zu können, müssen aber zuerst die Forschungsstrategien und die Methoden der Diskurs- und Inhaltsanalyse in der Sozialwissenschaft und die Verfahren des Text Mining dargestellt werden.

2.1 Forschungsstrategien: Deduktion oder Induktion

Bei der Realisierung eines sozialwissenschaftlichen Forschungsvorhabens ist die methodische Umsetzung des Forschungsablaufes grundlegend an die Entscheidung für ein qualitatives oder quantitatives Vorgehen gebunden.⁹ Nach Flick beruht die klassische quantitative Forschung auf Modellbildung und möglicher Falsifikation von Hypothesen mit statistischen Verfahren und der Repräsentativität der Ergebnisse über den einzelnen Fall

maschinellen Lernens, die auf regelmäßige Sprachmuster trainiert werden können (vgl. Wiedemann/Lemke/Niekler 2013, S. 109).

⁹ Beide Forschungstraditionen stehen in der Geschichte der Wissenschaft nicht erst in den letzten Jahrzehnten nebeneinander, sondern sind an zwei unterschiedliche Denktraditionen der Wissenschaft gekoppelt. Die qualitative Tradition beruht auf einem aristotelischen Wissenschaftsverständnis, welches die historische Entwicklung des Gegenstandsbereiches anhand seiner Ziele und Zwecke verstehen möchte und neben deduktiven Ableitungen auch induktive Schlüsse für Einzelfälle erlaubt. Die quantitative Methode beruht auf einer galileischen Tradition, welche deduktiven Kausalerklärungen und der Suche nach Naturgesetzen, welche sich empirisch statistisch belegen lassen, den Vorzug gibt (vgl. Mayring 2002, S. 12f).

hinaus. „Theorien und Methoden erhalten hier Priorität gegenüber dem Gegenstand der Forschung.“ (Flick 2000, S. 123) Der Fokus der qualitativen Forschung liegt dagegen auf gegenstandsbezogener Theoriebildung, anhand eines konkreten Untersuchungsfeldes unter Berücksichtigung seines Kontextes (vgl. Flick 2000, S. 125). Die Anwendung von qualitativen und quantitativen Methoden und daraus folgender Untersuchungsdesigns ist demnach einerseits dem theoretischen Forschungsparadigma und andererseits der konkreten Fragestellung, also dem Erkenntnisinteresse, zugeordnet (vgl. Flick 2000, S. 40f.). Nach Flick läuft die Integration verschiedener Forschungsmethoden, die auf eine Kombination von qualitativer und quantitativer Forschung setzen, meist auf ein sequenzielles oder paralleles Design innerhalb des Forschungsprozess hinaus (vgl. Flick 2000, S. 53). Ein sequenzieller Prozess der Methodenkombination ist dann z.B. die Hintereinanderschaltung von qualitativen und quantitativen Methoden der Inhaltsanalyse unter Einbeziehung computergestützter Verfahren.

Das Projekt ePol verfolgt eine kombinierte Forschungsstrategie. Dabei bildet der deduktiv hypothesentestende Ansatz den Ausgangspunkt der Untersuchung und die induktiven und explorativen Elemente der Untersuchung nutzen Text Mining, um über den Hypothesentest hinaus, Erkenntnisse zur Debatte über die Postdemokratie zu produzieren.

Den Kern des deduktiven Forschungsansatzes bildet die empirische Prüfung der Ökonomisierungsthese des Postdemokratiediskurses, welche die Annahme eines Wandels von Begründungen, genauer eine Zunahme wirtschaftsaffiner neoliberaler Begründungen von politischen Entscheidungen in der politischen Öffentlichkeit, beinhaltet. Dafür muss die These durch Operationalisierung in falsifizierbare Hypothesen transformiert und somit empirisch zugänglich gemacht werden. Findet sich also im Zeitungskorpus eine signifikante Steigerung neoliberaler Begründungen für Politik, im theoretisch fokussierten Zeitraum, kann die These als vorläufig bestätigt angesehen werden. Eine darauf aufbauende These des Projekts ist, dass die neoliberalen Begründungen, mit zunehmender Hegemonie neoliberaler Ideen in der Politik, an Deutungsmacht gewinnen. Es wird davon ausgegangen, dass einhergehend mit der steigenden Deutungsmacht der neoliberalen Begründungen, deren Erklärungs- und Plausibilisierungsnotwendigkeit sinkt und damit eine Veränderung in der Dichte und Qualität von Begründungen eintritt. Das bedeutet, bestätigt sich die These des Postdemokratiediskurses und vermindert sich dann der Umfang der neoliberalen Begründungen, bei gleichzeitiger Präsenz ihrer qualitativen Elemente, kann auch diese These als vorläufig bestätigt gelten.

Im Gegensatz zu einer deduktiven Hypothesenprüfung ist mit Text Mining Verfahren auch ein induktives Vorgehen im Forschungsprozess möglich. Induktiv-explorative Ansätze sind besonders geeignet um die Datenlage und die Durchführbarkeit einer Hypothesenprüfung zu beurteilen. Zu Beginn des Forschungsprozesses helfen induktive Strategien bei der Formulierung trennscharfer Kategorien und damit bei der Definition eines deduktiven Forschungsansatzes. Der Vorteil solcher induktiven, also datengetriebenen, Beobachtungen des Textkorpus besteht darin, dass unbekannte Ausprägungen der zu untersuchenden Kategorien erfasst und dadurch Thesen erweitert oder verändert werden können. Dies ist z.B. für die Arbeit von ePol ein wesentlicher Bestandteil, da die Definition der messbaren Größen einer Kategorie von einer datengetriebenen Beschreibung dieser Größen abhängt.

gig ist. Zusätzlich kann durch induktive Verfahren und durch *distant reading* ein semantischer Raum und somit eine Menge an relevanten Inhalten exploriert werden, um Untersuchungsmengen und thematische Bezüge besser für die Untersuchung zu nutzen. Des Weiteren können die Ergebnisse eines deduktiven Hypothesentests einer weiteren qualitativen Analyse unterzogen werden und somit induktiv zur neuen Theoriebildung beitragen.

Bei der Verwendung induktiver und deduktiver Verfahren ist darauf zu achten, dass die beide Forschungsansätze in ihrer Anwendung durch Verfahren des Text Mining und der Inhaltsanalyse unterscheidbar und zurechenbar bleiben. Dies ermöglicht eine sozialwissenschaftlich adäquate Methodenanwendung bei gleichzeitiger Nutzung der Potenziale des Text Mining. Um die Forschungsstrategien nicht zu vermischen, muss die Prüfung der Ökonomisierungsthese also in einem eigenständigen Forschungsdesign realisiert werden, welches nach der allgemeinen Diskussion der Diskurs- und Inhaltsanalyse (Kapitel 2.2) und des Text Mining (Kapitel 2.3) in Kapitel 2.4 dargestellt wird.

2.2 Diskurs- und Inhaltsanalyse in den Sozialwissenschaften

Im Folgenden wird ein Überblick über die Forschungstradition der Inhalts- und Diskursanalyse in den Sozialwissenschaften gegeben, um die relevanten Aspekte der Anschlussfähigkeit an die Verfahren des Text Mining hervorzuheben. Die Beobachtung gesellschaftlicher Diskurse, und damit z.B. die Aufdeckung der Veränderungen politischer Begründungen, bedürfen einer eigenen empirischen Methode. Nach Hajer (vgl. 2008, S. 212) subsumieren sich unter dem Ansatz der Diskursanalyse Methoden, die argumentative Strukturen in Texten oder Aussagen zu einem bestimmten Themengebiet untersuchen und strukturieren können. Aussagen und Sprache sind nicht nur ein Medium, durch ihre Anwendung formen sie unser Verständnis der sozialen Realität. Keller (2004, S. 64) definiert Diskurse demgemäß als, „nach unterschiedlichen Kriterien abgrenzbare Aussagepraxis bzw. Gesamtheit von Aussageereignissen, die im Hinblick auf institutionell stabilisierte gemeinsame Strukturmuster, Praktiken, Regeln und Ressourcen der Bedeutungserzeugung untersucht werden“. Damit ist die Diskursanalyse auch ein Verfahren, welches „die Produktion, die Verbreitung und den historischen Wandel von Deutungen für soziale und politische Handlungszusammenhänge“ untersucht (Schwab-Trapp 2011, S. 35). Ob bei einer Diskursanalyse die detaillierte Durchdringung eines Diskurses oder die zeitliche Nachzeichnung relevanter Diskursstränge im Fokus der Untersuchung steht, ist dann eine Frage des Forschungsdesigns und der Operationalisierung der Fragestellung und welche auch die Fallauswahl beeinflusst (vgl. Hajer 2008, S. 215). Die methodische Durchführung einer Diskursanalyse folgt somit keinem Standardmodell, sondern ist von den diskurstheoretischen Grundannahmen der Forscher und des jeweiligen Forschungsinteresses abhängig¹⁰ (Keller/Viehöver 2006, S. 107).

¹⁰ Der Status der Diskursforschung als einheitliche Forschungsdisziplin ist durch die unterschiedlichen Ausrichtungen, welche zwischen einer an Habermas angelehnten Diskursethik, einer empirisch-analytischen Ausrichtung (z.B. Policy Studies oder Governance-Forschung) und von Michel Foucault inspirierten Verfahren, in den Politikwissenschaften umstritten (vgl. Kerchner 2006, S. 33f.). Nach Keller und Viehöver untersuchen Diskursanalysen „a.) Diskurse im Hinblick auf deren Regelstrukturen, ihre symbolische, semantische, pragmatische und kognitive Strukturierung (Diskurse als strukturierte Sinnsysteme) und b.) die Praktiken und (rhetorischen) Strategien ihrer Artikulation durch Akteure und Akteursgruppen (Diskurskoalitionen) in Inter-

Der Übergang zwischen einer Diskursanalyse und einer Inhaltsanalyse kann fließend sein, wenn eine Diskursanalyse angefertigt wird und diese sich ausschließlich auf Textkorpora beschränkt. In einer Diskursanalyse werden dann die Verfahren der Inhaltsanalyse genutzt und der Untersuchungsgegenstand als thematischer Diskurs eingeschränkt. Deshalb werden im Folgenden die Verfahren der Inhaltsanalyse ausführlicher beschrieben, da so die methodischen Verfahren, die dann in eine Diskursanalyse analog angewendet werden können, im Vordergrund stehen.

Die Verfahren der Inhaltsanalyse sind vielfältiger Art und beziehen sich auf unterschiedliche Forschungstraditionen¹¹. Mit diesen Verfahren wird u.a. Text als Kommunikationsinhalt wissenschaftlich untersucht, um so Aussagen über die soziale Wirklichkeit zu erhalten (vgl. Knapp 2008, S. 20). Die Inhaltsanalyse als Methode folgt dabei der etablierten Vorgehensweise aller empirischen Wissenschaften und untersucht theoretisch formulierte Aussagen über die Wirklichkeit mit Hilfe empirischer Daten (Früh 2009, S. 19). Unterschieden werden kann die Inhaltsanalyse in Verfahren die eher die qualitativen oder die quantitativen Aspekte der Analyse betonen. Diese Unterscheidung folgt stringent der Unterscheidung zwischen qualitativer und quantitativer Forschung in anderen Gebieten der Sozialforschung.¹² So wird in der qualitativen Inhaltsanalyse der Aspekt des Verstehens und in der quantitativen Inhaltsanalyse der Aspekt der Objektivität betont (vgl. Knapp 2008, S. 20f.). Neben der Objektivität des Analyseprozesses, steht in der quantitativen Inhaltsanalyse die systematische Auswertung des sprachlich verfassten Textmaterials im Vordergrund (Berelson 1971, S. 18). Inhaltsanalysen bearbeiten ihren Untersuchungsgegenstand Text dafür mit Kategoriensystemen, welche als Operationalisierung der Fragestellung die Datenerhebung ermöglichen (Früh 2009, S. 82ff.). Die quantitative Inhaltsanalyse eignet sich besonders für hypothesentestende Verfahren. Von Forschern die einen interpretativ-hermeneutischen Ansatz verfolgen, wird meist die qualitative Methode gewählt (vgl. Christmann 2006, S. 285f.). Qualitative und quantitative Inhaltsanalysen unterscheiden sich deshalb auch in der Art der Kategorienbildung. Während quantitative Ansätze für einen Hypothesentest deduktiv Kategorien aus der Theorie ableiten, entwickeln qualitative Ansätze die Kategorien induktiv am vorliegenden empirischen Material (vgl. Früh 2001, S. 119).

Im Mittelpunkt der Methodendiskussion steht bei der Inhaltsanalyse die Frage nach dem Text und dessen Interpretation bezüglich seines Inhalts und seiner Bedeutung. Nur die gültige Interpretation erlaubt eine Inferenz auf die soziale Realität. Problematisch bei einer Inhaltsanalyse ist die Annahme, einem Text könne der eine richtige Inhalt tatsächlich entnommen werden. Diese vereinfachte Sichtweise ist durch die Einsicht in die soziale Konstruktion von Sprache und damit in die Kontextgebundenheit sprachlich verfasster

Aktion (Diskurse als System von Praktiken)" (2006, S. 103). Weiterführend zu Forschungsdesigns in der Diskursanalyse u.a. Keller/Viehöver 2006; Keller 2004, 2007; Landwehr 2010.

¹¹ Für einen kurzen Überblick über die Geschichte der Inhaltsanalyse im deutschen und internationalen Kontext siehe u.a. Früh 2009, S. 11–15, Merten 1995, S. 35–47, Atteslander 2010, S. 198–202.

¹² Die qualitative Inhaltsanalyse, welche als Propaganda Analyse in Zeitungen während des ersten und zweiten Weltkriegs ihren Anfang nahm, wurde von Harold D. Lasswell und Bernard Berelson methodisch fundiert. Die qualitative Inhaltsanalyse geht auf den Ansatz von Kracauer zurück, der gegen die Vereinheitlichung der qualitativen Inhaltsanalyse die Einzelanalyse und damit den kontextuellen Sinnzusammenhang und die latenten Inhalte der Texte in den Vordergrund stellt (vgl. Christmann 2006, S. 274f.).

Texte zurückzuweisen (vgl. Knapp 2008, S. 22). Mayring merkt an, dass manchen Inhaltsanalytikern „der Begriff ‚Inhalt‘ überhaupt suspekt“ erscheint, da je nach Ausrichtung der Analyse entweder die manifesten oder die latenten Inhalte¹³ im Vordergrund stünden (Mayring 2010, S. 11). Nach Lamnek lässt sich die qualitative und die quantitative Inhaltsanalyse auch über das Begriffspaar latent und manifest unterscheiden. Methodisch erfasst die quantitative Inhaltsanalyse die manifesten Aussagen, während die qualitative Inhaltsanalyse besser die latenten Sinnzusammenhänge erfasst (vgl. Lamnek 1995, S. 191). Jedoch erlaubt die quantitative Inhaltsanalyse ebenfalls Zugriff auf latente Bedeutungen, indem Messvariablen geschickt miteinander kombiniert und ausgewertet werden¹⁴.

Bei den angesprochenen Methoden der Sozialwissenschaft ist der Einsatz computergestützter Verfahren keine Neuheit. In der quantitativen Inhaltsanalyse sind computergestützte Methoden in den Sozialwissenschaften schon seit den 1960 Jahren bekannt und seit den frühen 80er Jahren ist auch eine Zunahme computergestützter Methoden der qualitativen Inhaltsanalyse (*Qualitative Data Analysis*) zu beobachten (Kuckartz 2010, S. 218ff.). Mit Programmen zur *Qualitative Data Analysis* (QDA-Software) wie MAX QDA¹⁵ wurden in diesem Bereich große Fortschritte erzielt (Kuckartz 2006, S. 81, 2007). Auch für die Diskursforschung besteht im Einsatz von computergestützten Verfahren ein Potenzial für die Produktion valider Ergebnisse, wenn die methodologischen Grundlagen des jeweils gewählten diskursanalytischen Ansatzes eine elektronische und automatisierte Datenerhebung zulassen (vgl. Diaz-Bone/Schneider 2004, S. 490). Somit besteht in der Tradition des Faches eine bereits vielfach bewährte Computerunterstützung der klassischen Methoden der Sozialwissenschaft. Diese computergestützten Verfahren sind z.B. Text-Retrieval, also die Zusammenstellung von codierten Textstellen (vgl. Kuckartz 2010, S. 108ff.), oder Klassifikation, also automatisch typologisierende Verfahren der Texterkennung (vgl. Kuckartz 2010, S. 227ff.). Dabei bleiben diese bekannten Anwendungen aber in ihrer Verarbeitungskapazität hinter modernen (halb-) automatischen Verfahren des Text Mining zurück.¹⁶

2.3 (Halb-) automatische Inhaltsanalysen mit Text Mining

Die Inhaltsanalyse und ihre Methoden werden - wie erwähnt - mit computergestützten Ansätzen ergänzt. Für den Einsatz von computergestützten maschinellen Lernverfahren in digitalen Textkorpora sind die Begriffe Text Mining oder Natural Language Processing geläufig. Bei Verfahren des Text Mining werden einerseits überwachte Verfahren genutzt,

¹³ „*Manifest* sollen dann alle *absoluten* Merkmale eines Textes sowie die zwischen diesen herstellbaren Relationen heißen. *Latent* sind dagegen die Relationen, die sich zwischen den absoluten oder relationalen Textmerkmalen und den Benutzern des Textes (Kommunikator, Codierer, Rezipient) ausmachen lassen“ (Merten 1995, S. 57, Herv. i.O.).

¹⁴ So z.B. der kombinierten Einsatz von Wörterbücher und Kookkurrenzanalyse bei der Dokumentenselektion, hier kurz in Kapitel 2.4 oder ausführlich in Wiedemann/Niekler 2014.

¹⁵ <http://www.maxqda.de/>

¹⁶ Die darüberhinausgehende Anwendung computergestützter Verfahren der Automatischen Sprachverarbeitung sollte deshalb nicht an einer Skepsis gegenüber Automatisierung und neuartiger Unterstützung durch Text Mining bei der Inhaltsanalyse scheitern. So warnt auch Werner Früh in der Einleitung seines Buches zur Inhaltsanalyse vor einer neuen Auseinandersetzung zwischen Vertretern der computergestützten Inhaltsanalyse und der klassischen Variante. Bei der sich abzeichnenden Debatte würden nur dieselben Argumente, welche bereits zwischen qualitativen und quantitativen Forschern ausgewechselt wurden, erneuert (vgl. Früh 2009, S. 7f.).

deren Funktionsweise auf der Einbeziehung externer Beispiele basiert. Diese Verfahren besitzen deshalb einen deduktiven Charakter, da die Beispiele externes Wissen oder theoretische Annahmen repräsentieren können. Andererseits werden unüberwachte Verfahren eingesetzt, deren Aufgabe es ist gemeinsame Eigenschaften in einem Datensatz wie einer Dokumentkollektion aufzudecken und sichtbar zu machen. Die Ansätze entsprechen also einem induktiven Forschungsdesign, bei dem anhand der Daten Gesetzmäßigkeiten und Eigenschaften aufgedeckt und für die Formulierung neuer Hypothesen und Annahmen herangezogen werden können. Beim Einsatz der Verfahren des maschinellen Lernens im Text Mining kann deshalb eine Verbindung zu induktiven und deduktiven Forschungsstrategien in den Sozialwissenschaften hergestellt werden.

Die Annahme, „dass es einen Prozess gibt, der die von uns beobachteten Daten erklärt“ (Alpaydin 2008, S. 1), steht hinter der Entwicklung und Anwendung der Verfahren des maschinellen Lernens. Die Modelle, die für das Lernen aus Daten entwickelt werden, dienen einerseits dazu Vorhersagen für neue bzw. unbekannte Daten zu generieren, die von Objekten¹⁷ gleicher Art erzeugt werden oder um die Prozesse, die die Daten erzeugt haben, selbst zu beschreiben (Hastie et al. 2001, S. 2). Als ein Prozess kann z.B. die thematische Zusammensetzung eines Korpus modelliert werden (siehe Fußnote 21). Weiterhin kann der Zusammenhang zwischen dem Auftreten bestimmter Wörter in einem Dokument und einer assoziierten kategorialen Variable als ein solcher Prozess beschrieben werden. Die Ausprägungen von Dokumentobjekten, die verwendeten Wörter und das Vokabular, lassen sich unterschiedlich repräsentieren, sodass der Text als Eingabeobjekt für ein statistisches Verfahren dienen kann. Beispielsweise kann ein Dokument als Vektor über das gesamte Vokabular eines Korpus modelliert werden. Für eine Wortform, die in einem Dokument enthalten ist, wird dann an der richtigen Stelle im Dokumentvektor eine Zählung für die Auftretenshäufigkeit der Wortform im Dokument vermerkt. Ein kompletter Korpus kann demnach als sogenannte Dokument-Term-Matrix dargestellt werden, in der alle Dokumente als Zeilenvektor vermerkt sind.

Vorwiegend werden Aufgaben der Klassifikation, des Information Retrieval (IR), der Gruppierung von Dokumenten oder der Inhaltszusammenfassung an Korpora vorgenommen. An dieser Stelle zeigt sich auch eine Parallele zwischen den Methoden des maschinellen Lernens und den Methoden der Inhaltsanalyse in der Sozialwissenschaft. Denn genau wie bei der Inhaltsanalyse kann man deduktive und induktive Ansätze unterscheiden, die im Bereich des maschinellen Lernens als supervised und non-supervised Verfahren (überwacht und nicht überwacht) bekannt sind (Hastie et al., S. 3). Der Unterschied der Ansätze kann wie folgt zusammengefasst werden. Jedes Objekt in einem Korpus weist unterschiedliche Ausprägungen seiner Merkmale auf. Diese Features stellen für einen Algorithmus, oder auch Lerner, die Input-Variable dar. Beim überwachten Lernen stellt ein Objekt zusätzlich zu den Features ebenfalls eine Output-Variable bereit, z.B. eine Doku-

¹⁷ Ein Objekt kann z.B. Teil einer Sammlung aus Daten wie ein Dokumentkorpus sein. Jedes Dokument ist aus unterschiedlichen Wörtern zusammengesetzt. Die Features eines Dokuments können demnach die Wörter in den Dokumenten sein und die manifeste Wahl der Wörter der Objektzustand. Innerhalb einer Rubrik oder einer Thematisierung innerhalb unterschiedlicher Textquellen verwenden unterschiedliche Dokumente sich überschneidende Wortmengen. Diese Eigenschaft kann dann zur Klassifikation über maschinelle Lernverfahren genutzt werden.

mentklasse oder eine Kategorie, und der Lerner kann damit angeleitet werden die Ausprägungen der Features mit der Output-Variable zu assoziieren. Für überwachte Klassifikationsverfahren werden die verwendeten Kategoriensysteme auf Dokument-Basis (z.B. Rubrik, Einstellung, Thema) oder basierend auf Textstrukturen (Relationen, Mehrworteinheiten, Abschnitte) definiert. Bei IR-Systemen¹⁸ wird eine überwachte Relevanzbewertung anhand von Beispieldaten durchgeführt. Die Beispieldaten können im Fall von Dokumentkollektion Mengen von Schlüsselwörtern oder Beispieldokumente sein, zu denen ähnliche Dokumente gefunden werden sollen. Beim unüberwachten Lernen existiert diese Output-Variable nicht und man nutzt die Features, um deren Varianzstruktur bzw. deren Verteilungen zu schätzen oder ähnliche Verteilungen zu gruppieren. Im Sinne der Forschungsansätze bei der Inhaltsanalyse wird der Einsatz der überwachten Verfahren als deduktiv verstanden, da mit einer Trainingsmenge agiert wird, die vorher manuell zusammengestellt wurde. Diese Trainingsmenge aus Objekten, mit Features und je einem oder mehreren Output-Variablen, repräsentiert in diesem Fall eine Hypothese oder externe theoretische Festlegung, die in anderen unbekanntem Daten getestet wird.¹⁹ Im Gegensatz dazu ordnen und strukturieren die unüberwachten Verfahren die Daten und deren Verteilung und helfen deshalb in einer großen Menge von Objekten induktive Schlüsse über deren Zusammensetzung zu ziehen und Hypothesen oder Annahmen zu generieren.

2.4 Forschungsdesign zur (halb-) automatischen Argumentidentifikation

Nach der Beschreibung des relevanten methodischen Repertoires der beiden Fachrichtungen, kann nun die Anwendung eines kombinierten Einsatzes der Methoden der Sozialwissenschaft und des Text Mining beschrieben werden. Dafür wird exemplarisch das Forschungsdesign von ePol zur empirischen Prüfung der eingangs beschriebenen These des Postdemokratiediskurses vorgestellt, um die Potenziale des kombinierten Einsatzes von Inhaltsanalyse und Text Mining anschaulich zu demonstrieren.

Das Forschungsdesign umfasst drei Teilaufgaben zur Untersuchung eines Begründungswandels in Zeitungskorpora. Die drei Teilaufgaben zur (halb-) automatischen Argumentidentifikation sind:

1. Identifikation relevanter Dokumente aus dem Zeitungskorpus
2. manuelle Annotation neoliberaler Argumente
3. (halb-)automatische Klassifikation im Textkorpus

Die Bearbeitung der Teilaufgaben erfolgt in einer onlinebasierten Forschungsarchitektur, dem Corpus Miner, der Text Mining Werkzeuge des Projekts ePol beinhaltet (vgl. Niekler et al. 2014)²⁰.

¹⁸ „Retrieval“ steht für Abfrage einer Datenbank. Unter Information Retrieval wird die Organisation, Repräsentation und den Zugriff auf Informationen verstanden, die z.B. als Dokumente, Webseiten oder als Bibliothekskatalog vorliegen können. Zentral ist die Repräsentation und Organisation, die einen einfachen Zugriff auf die Daten im Interesse des Nutzers erlaubt (vgl. Baeza-Yates, Ribeiro-Neto S. 3).

¹⁹ So ist z.B. das Auffinden einer bestimmten Kategorie von Dokumenten in einer unbekanntem Dokumentmenge vorstellbar, sodass die Existenz bestimmter Themen in einer Textkorpus angenommen oder abgelehnt werden kann.

²⁰ Das Werkzeug wird als „Leipzig Corpus Miner“ bezeichnet (nachfolgend LCM).

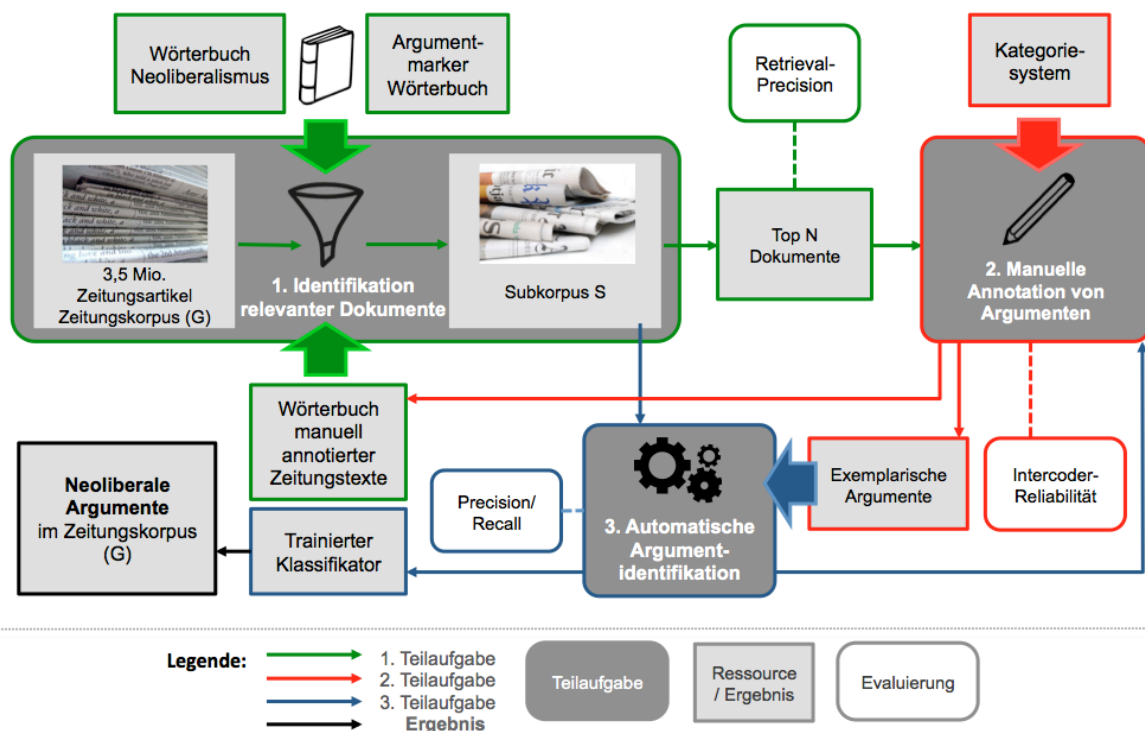


Abbildung 1: Methodischer Forschungsablauf in drei Schritten

Die erste Teilaufgabe - hier grün markiert - des Forschungsdesigns dient einer Selektion relevanter Dokumente aus dem Zeitungskorpus. Die Aufgabenstellung entspricht einer IR Aufgabe. In Standard-Retrieval Anwendungen für Dokumentendatenbanken werden meist schlüsselwortbasierte Verfahren eingesetzt. Diese selektieren Dokumente anhand der Übereinstimmung der Schlüsselwörter innerhalb von Metadaten oder des konkreten Inhalts der Dokumente. Für die Selektion relevanter Dokumente im Forschungsprozess reicht eine einfache Selektion über Schlüsselwörter nicht aus, da das Konzept „ökonomisierte Argumentation“ nicht hinreichend über eine manuell definierte Menge an Schlüsselwörtern beschrieben werden kann. Vielmehr müssen die Phrasen und Wortverwendungen für prototypische und relevante Inhalte mit Verfahren des Text Mining erkannt und daraus eine Retrieval Strategie entwickelt werden. Unser Vorgehen beinhaltet deshalb die Erstellung zweier Diktionäre, also Wörterbüchern, welche die Wortverwendung ökonomisierter Sprache (*Wörterbuch des Neoliberalismus*) und argumentativer Sprachstrukturen (*Argumentmarkerwörterbuch*) repräsentieren. Diese zwei Diktionäre werden nacheinander zur Bewertung des Zeitungskorpus angewendet, um eine geordnete Menge potenziell neoliberalen Zeitungsdokumenten mit hoher Argumentationsdichte in den vordere Rängen zu generieren. Das *Wörterbuch des Neoliberalismus* wurde dabei aus einem Referenzkorpus, bestehend aus 36 Werken neoliberaler Theoretiker, extrahiert und umfasst alle in deutscher Sprache erschienen Bücher von Autoren die Mitglieder der Mont Pelerin Society waren (z.B. Milton Friedman, Friedrich von Hayek). Damit dient die kontextualisierte Wortwahl neoliberaler Theoretiker als Auswahlgrundlage für eine nähere qualitative Auswertung in der zweiten Teilaufgabe. Mit den Verfahren der Term Extraktion

und der Topic Modelle²¹ wird so eine Liste der 500 wichtigsten neoliberalen Terme dieser Bücher extrahiert (vgl. Wiedemann/Niekler 2014). Diese Terme wurden um die gemeinsamen signifikanten Kookkurrenzen²² ergänzt, was die Berücksichtigung des Verwendungsverwendungszusammenhangs neoliberaler Sprache ermöglicht. Diese Informationen werden für die Berechnung einer Maßzahl verwendet, welche die Ähnlichkeit zwischen dem Referenzkorpus der neoliberalen Theoretiker und dem Zeitungskorpus angibt. Diese Bewertung wird genutzt, um die Zeitungsdokumente nach ihrem vermuteten Anteil an neoliberalen Inhalt in eine Reihenfolge zu bringen. Da die Suche nach neoliberalen Argumenten im Vordergrund steht, wurde diese Dokumentreihenfolge durch ein *Argumentmarkerwörterbuch* korrigiert. Das *Argumentmarkerwörterbuch* umfasst 127 Wörter, welche in Argumentationen als Schlüsselwörter gebraucht werden (vgl. Dumm/Lemke 2013). Auf diese Weise können Texte mit neoliberalen Inhalt anhand der Argumentmarker nach der vermuteten Argumentationsdichte umsortiert werden. Das Ergebnis der ersten Teilaufgabe ist eine geordnete Menge an potenziell thematisch passenden Zeitungstexten, bei denen in den vorderen Rängen eine hohe Dichte an Argumentstrukturen erwartet werden kann. Dieser wörterbuchbasierte Ansatz ermöglicht einen möglichst breiten und objektiven Einstieg für die Suche im Gesamtkorpus. (Wiedemann/Niekler 2014).

In der zweiten Teilaufgabe – in Abbildung 1 rot markiert - werden die so selektierten Dokumente analog einer klassischen Inhaltsanalyse manuell anhand eines Kategoriensystems annotiert und so neoliberale Sprachgebrauchsmuster als Argumente identifiziert.²³ Wobei neben dem eigentlichen Argument auch die Absätze, in denen das Argument erscheint, annotiert werden. Auf diese Weise kann bei der computergestützten Suche in der dritten Teilaufgabe, also bei der (halb-) automatischen Klassifikation von

²¹ Im Falle von Textdokumenten können unter der Modellannahme, dass ein Dokument aus mehreren Wortklassen (Topics) gebildet ist und dass eine bestimmte Anzahl von Topics einem Textkorpus zugrunde liegt, diese Wortklassen durch eine Inferenz berechnet werden. Die dabei entstehenden Klassen, beinhalten Wörter, die in den Dokumenten gemeinsam verwendet werden. Unter der Annahme, dass ein Dokument auch mehrere Klassen beinhaltet, können mehrere semantische latente Themen per Dokument festgestellt werden. Die Dokumente werden dann als Mischung thematischer Klassen dargestellt (Blei/Ng/Jordan, 2003). Eine Klasse wird durch die Abbildung der Wörter auf eine multinominale Wahrscheinlichkeitsverteilung gebildet. Dabei sind spezifische Wörter einer Klasse bzw. eines latenten semantischen Zusammenhangs mit einer höheren Wahrscheinlichkeit versehen als andere. Die Reihenfolge der Wichtigkeit bestimmter Terme für eine Fachsprache nach deren Verwendung in abgeschlossenen Klassen (Wahrscheinlichkeit) und der Verwendungshäufigkeit ermittelt. Innerhalb jeder Klasse wird für jedes Wort die Termwahrscheinlichkeit in einer Klasse mit der Termfrequenz im Korpus multipliziert. Für ein Wort werden diese Werte über alle Klassen summiert. Sehr themenspezifisch verwendete, aber hochfrequent auftretende Wörter erhalten einen höheren Rang (vgl. Wiedemann/Niekler 2014).

²² Die Grundlage der Kookkurrenzanalyse bildet das gemeinsame Auftreten von Wörtern in einer Analyseeinheit, welche den Umfang von Dokumenten, Sätzen oder Absätzen haben kann. Sie werden deshalb als Dokument-, Absatz-, oder Satzkoookkurrenzen bezeichnet. Gemessen wird bei der Analyse nicht die Anzahl des gemeinsamen Auftretens, sondern eine statistische Signifikanz. Dies bedeutet, dass Wortpaare, die statistisch signifikant gemeinsam auftreten auf einen Verwendungszusammenhang hinweisen, der semantisch ist und z.B. Eigenschaft eines Themas oder einer Kategorie ist (vgl. Heyer, 2006, S. 134ff.).

²³ Die manuelle Textannotation basiert auf einem klassischen Kategoriensystem. Dieses Kategoriensystem beinhaltet das Verständnis von Argumentation in Zeitungstexten, wie es für den vorgeschlagenen Forschungsprozess definiert wird. Für den Aufbau von Argumenten orientiert sich das Kategoriensystem an dem Argumentationsschema von Stephan Toulmin (1996). Dabei ist darauf hinzuweisen, dass in Zeitungen der Anspruch an Argumentationen nicht in der Weise hoch angesetzt werden darf, wie vielleicht für eine philosophische Diskussion. Der Anspruch an Argumentation umfasst, dass eine Behauptung oder einen Tatbestand mit dem Verweis auf eine Begründung desselben abgesichert wird. Für die Annotation wurde im LCM eine eigene Funktion bereitgestellt. Bei der Annotation der Argumente wird ein hierarchisches Kategoriensystem angewendet.

Argumenten, auch der Kontext der Argumenten berücksichtigt werden. Neben den neoliberalen Argumenten werden zusätzliche Information zum Text annotiert, wie z.B. ob die im Text auftauchenden Argumente eine affirmative Haltung gegenüber den neoliberalen Inhalten aufweisen oder diese eher kritisch bewerten. Die manuell annotierten Argumente stellen Trainingsdaten für die dritte Teilaufgabe dar.

Die dritte Teilaufgabe – in Abbildung 1 blau markiert - setzt die computergestützten Methoden des überwachten Lernens ein (überwachtes maschinelles Lernen bzw. überwacht Text Mining), um eine (halb-) automatische Klassifikation neoliberaler Argumente durchzuführen. Dafür ist im LCM ein Klassifikator integriert, der die Features der manuell annotierten Argumente nutzt²⁴. Die Forscher sind an diesem Prozess aktiv beteiligt, indem die klassifizierten Texte durch sie bestätigt oder falsifiziert werden. Auf diese Weise können große Mengen an Trainingsdaten ergänzt und die Klassifikationsqualität bis zur gewünschten Güte erhöht werden. Dieses Vorgehen erlaubt den Forschern eine Kontrolle der Ergebnisse der automatischen Klassifikation, deshalb ist diese Methode der Klassifikation (halb-) automatisch. Die vollständige Kontrolle auch über automatische Prozesse zu behalten ist eine der wichtigsten Anforderungen an die interdisziplinäre Arbeit. In einer abschließenden Klassifikation werden automatisch Textstellen für die antrainierten Kategorien identifiziert, die Aussagen darüber ermöglichen, ob sich die neoliberale Argumentation in der Begründung von Politik quantitativ und qualitativ im Zeitverlauf verändert hat. Damit ist dann die Beantwortung der Forschungshypothesen möglich.

In jeder dieser Teilaufgaben müssen Gütekriterien zur Anwendung kommen, um die Validität der Ergebnisse sicherzustellen. Entsprechend kommen unterschiedliche Verfahren aus der Sozialforschung, der Inhaltsanalyse und der Informatik zum Einsatz, auf die im Folgenden eingegangen wird.

3. Sozialforschung und Gütekriterien

Bei neuen und innovativen Methoden ist die Rückbindung an die Standards in der Fachdisziplin notwendig, um die Validität der Forschungsergebnisse zu sichern und die Vorbehalte gegenüber computergestützten Verfahren der Diskurs- und Inhaltsanalyse zu mindern. Die konsequente Anwendung etablierter Gütekriterien sichert die Qualität der Ergebnisse und Prozesse sowie die Anschlussfähigkeit an die gängigen Methoden der empirischen Sozialforschung. Die Anwendung von Gütekriterien auf das Forschungsdesign von ePol wird deshalb, nach der theoretischen und überblicksartigen Ausführungen der Gütekriterien der quantitativen und qualitativen Forschung und der ausführlicheren Darstellung der Gütekriterien Inhaltsanalyse und des Text Mining, im abschließenden Unterkapitel beschrieben.

²⁴ Dabei werden Algorithmen zur Text-Klassifikation eingesetzt, die aus Trainingsbeispielen lernen. Diese Trainingsbeispiele sind in dem beschriebenen Forschungsansatz Textstellen, die als Argument identifiziert wurden. Die Trainingsbeispiele ermöglichen eine Unterscheidung von Textstellen nach deren Zugehörigkeit zur Kategorie des neoliberalen Arguments. Zur weiteren Ausführung wird auf Kapitel 3.4 verwiesen.

3.1 Gütekriterien in der quantitativen und qualitativen Forschung

In der empirischen Forschungspraxis gibt es keine einheitlichen Gütekriterien, die gleichermaßen für die qualitative und quantitative Sozialforschung zur Anwendung kommen (vgl. Flick 2011, S. 109; Steinke 2010). Dieser Fakt ist u.a. dem Methodenstreit in der empirischen Forschung zwischen qualitativen und quantitativen Sozialforschern geschuldet²⁵. Nicht zuletzt durch den verstärkten Einsatz von Triangulation und Mixed-Method-Ansätzen in den Sozialwissenschaften kam es in den letzten Jahren zu einer verstärkten Diskussion über die Gütekriterien und die Erkenntnisreichweite kombinierter Forschungsansätze (vgl. Flick 2000, S. 40). In der Literatur wird argumentiert, dass jenseits der Unterscheidung von qualitativer und quantitativer Forschung allgemeine wissenschaftliche Gütekriterien existieren. Auf einer abstrakteren Ebene lassen sich also allgemeine Gütekriterien formulieren, die bei einer Kombination der beiden Forschungstraditionen Anwendung finden können (vgl. Schnapp et al. 2006, Steinke et al. 2007). Dazu zählen die Frage nach der Relevanz der Fragestellung, die Angemessenheit der Methodenwahl, die Dokumentation der Methode, die kritische Prüfung von Theorien mittels Falsifikation, die Beachtung der externen Validität und das ethische Vorgehen im Forschungsprozess (vgl. Steinke 2007, S. 180f.). Neben diesen allgemeinen Kriterien besteht in beiden Forschungstraditionen die Forderung nach der Reflexion der Erhebungs- und Analyseprozesse und Methoden als ein zentrales Element guter Forschung. (vgl. Schnapp et al. 2006, S. 17).

Die klassischen Gütekriterien quantitativer Forschung sind Reliabilität, Validität und Objektivität (vgl. Schnapp et al. 2006, S. 17). Unter Reliabilität wird die Zuverlässigkeit des Messinstrumentes verstanden (vgl. Schnell et al. 1999, S. 145f.). Dabei stehen die „Stabilität und Genauigkeit der Messung sowie der Konstanz der Messbedingungen“ im Vordergrund (Friedrichs 1973, S. 102). Dies stellt die Zuverlässigkeit eines Messinstrumentes und den Grad der Reproduzierbarkeit von Messergebnissen sicher (vgl. Merten 1995, S. 302). Bei der Validität handelt es sich um die Gültigkeit der Messung. Sie gibt an, ob ein Messinstrument tatsächlich erfasst, was untersucht werden soll (vgl. Schnell et al. 1999, S. 148; Friedrichs 1973, S. 100). Damit steht bei der Validität „die Korrespondenz von Analyseergebnis und sozialer Wirklichkeit“ im Vordergrund (Merten 1995, S. 302).²⁶ Das Kriterium der Objektivität besagt, dass die Anwendung eines Messinstrumentes durch unterschiedliche Personen zu gleichen Ergebnissen führen muss (vgl. Diekmann 2003, S. 216).

Während in der quantitativen Forschung die genannten Gütekriterien unbestritten sind, ist die Situation in der qualitativen Forschung nicht so eindeutig. Die Uneinheitlichkeit der Gütekriterien ist der Vielfalt der qualitativen Methoden geschuldet, denn von der Interviewführung bis zur teilnehmenden Beobachtung erstreckt sich ein weites Feld unterschiedlicher Forschungsansätze. Eine einheitliche Position, ähnlich der in der quantitativen Forschung, ist derzeit nicht abzusehen (vgl. Flick 2007, S. 188f., vgl. Lüders 2011).

²⁵ Zum Methodenstreit in den Sozialwissenschaften und dessen erkenntnistheoretischen Grundlagen siehe zusammenfassend u.a. Behrens (2003, S. 217ff.).

²⁶ Die Validität einer Messung ist dabei das „weitergefaßte Kriterium“, als die Reliabilität. Denn ein Messinstrument kann reliabel etwas messen, was aber in der sozialen Wirklichkeit keine Gültigkeit besitzt (vgl. Merten 1995, S. 302).

Der Stand der Diskussion um einheitliche Kriterien der qualitativen Forschung lässt nach Flick deshalb eine „konkret-pragmatische Antwort“ auf die Frage nach der Absicherung der Qualität qualitativer Studien kaum zu (vgl. Flick 2007, S. 192). Nach Flick lassen sich drei Grundpositionen in der Debatte um die Gütekriterien der qualitativen Forschung auffinden. Die erste Grundposition fordert die Verwendung der klassischen drei Gütekriterien der quantitativen Forschung auch in der qualitativen Forschung. Das Problem der Übertragung der Gütekriterien Reliabilität, Validität und Objektivität darin besteht, dass die qualitative Forschung nicht dem Grad der Standardisierung unterliegt wie die quantitative Forschung²⁷ (vgl. Flick 2007, S. 192f.). Die zweite Grundposition ist die der Reformulierung klassischer Gütekriterien für die qualitative Forschung. Für das klassische Reliabilitätskonzept ist z.B. die prozessuale Komponente der Datenerhebung und Auswertung durch Dokumentation aufzuzeigen und dadurch die Reliabilität der Untersuchung zu erhöhen (vgl. Flick 2007, S. 194). Die dritte Position ist die Forderung nach einer Neuformulierung methodenangemessener Gütekriterien, z.B. Triangulation zur Qualitätssicherung einzusetzen (vgl. Flick 2007, S. 197ff.). Etwas abweichend fasst auch Steinke die Positionen der Gütekriterien Diskussion in drei Gruppen zusammen. Es besteht die Position, welche die Kriterien der quantitativen Forschung auch in der qualitativen Forschung anwenden will. Die entgegengesetzte Position dagegen fordert eigenständige Gütekriterien der qualitativen Forschung. Wobei die dritte Position die Möglichkeit Gütekriterien für die qualitative Forschung zu formulieren generell ablehnt (vgl. Steinke 2010, S. 319f.). Steinke selbst spricht sich für die Etablierung eigenständiger Gütekriterien der qualitativen Forschung aus, die für die eingesetzte Untersuchungsmethode abgewägt werden müssen (vgl. Steinke 2010, S. 322f.; 1999). Die zentralen Gütekriterien qualitativer Forschung sind nach Steinke, die „Indikation der methodischen Vorgehensweise, die empirische Verankerung der gewonnenen Theorie in den Daten, das Aufzeigen der Verallgemeinerbarkeit der Ergebnisse und die Herstellung Intersubjektiver Nachvollziehbarkeit“²⁸ (Steinke 2007, S. 187).

3.2 Gütekriterien der qualitativen und quantitativen Inhaltsanalyse

Die Inhaltsanalyse beinhaltet sowohl qualitative als auch quantitative Elemente, deshalb müssen sich die zu definierenden Gütekriterien auf beide inhaltsanalytischen Ausrichtungen beziehen. Dazu gehören für die qualitativen Aspekte der Inhaltsanalyse die theoretische Herleitung der Forschungsfrage und der Kategorien sowie die Reliabilität der Codierung. Für die quantitativen Aspekte der Inhaltsanalyse sind die Systematik und Objektivität der Kategorien und der Analyse zu nennen. Ein regelgeleiteter Forschungsprozess ist eine Grundvoraussetzung jeder Inhaltsanalyse.

²⁷ Zum Problem der Anwendung der klassischen Gütekriterien der quantitativen Forschung auf den qualitativen Forschungsprozess siehe weiterführend u.a. Flick 2007; Steinke 1999.

²⁸ Ein oft genanntes Gütekriterium ist die intersubjektive Nachvollziehbarkeit. Nach Steinke kann bei der qualitativen Forschung der Anspruch auf intersubjektive Nachvollziehbarkeit nur durch den Fokus auf Nachvollziehbarkeit erfüllbar sein. Diese Nachvollziehbarkeit der Ergebnisse kann u.a. durch die Dokumentation des Forschungsprozesses hergestellt werden. Die Dokumentation erstreckt sich dabei auf: das Vorverständnis der Forschenden, die Erhebungsmethode, Transkriptionsregeln, Untersuchungsdaten, Auswertungsmethode, Informationsquellen, dies ermöglicht anderen Forschenden damit Ergebnisse und Interpretationen nachzuvollziehen (vgl. Steinke 2010, S. 324f.). Dieses Gütekriterium der Forschung scheint uns gerade in einem neu zu etablierenden Forschungsfeld von besonderer Bedeutung zu sein.

Ein zentrales Kriterium jeder Art von Inhaltsanalyse ist die theoretische Verankerung der Analyse. Eine gute Inhaltsanalyse: „analysiert ihr Material unter einer theoretisch ausgewiesenen Fragestellung; die Ergebnisse werden vom jeweiligen Theoriehintergrund her interpretiert und auch die einzelnen Analyseschritte sind von theoretischen Überlegungen geleitet“ (Mayring 2010, S. 13). In der ersten Phase des Forschungsprozesses muss demnach der theoretischen Vorarbeit eine herausragende Stellung eingeräumt werden. Diese muss u.a. darin bestehen, die theoretischen Implikationen der Forschungsfrage exakt zu operationalisieren²⁹. Darauf aufbauend können dann hypothesenprüfende methodische Vorgehensweisen formuliert und mit quantitativen Verfahren geprüft werden (vgl. Lamnek 1995, S. 189).

Die Gültigkeit der Ergebnisse einer Inhaltsanalyse ist maßgeblich durch die genaue und exakte Kategorienbildung zu optimieren (vgl. Knapp 2008, S. 35; Atteslander 2010, S. 203f.). In der Inhaltsanalyse werden Kategorien synonym mit Variablen, Merkmalen oder Merkmalsausprägungen beschrieben (Kuckartz 2010, S. 58). Während quantitative Ansätze deduktiv Kategorien aus der Fragestellung ableiten, entwickeln qualitative Ansätze die Kategorien induktiv am vorliegenden empirischen Material (vgl. Früh 2001, S. 119). Knapp verweist aus linguistischer Sicht darauf, dass bei der Inhaltsanalyse der Schritt der Kategorienbildung von besonderer Bedeutung ist. Deshalb gilt, dass für die Kategorien und deren Operationalisierung besondere Sorgfalt gelten muss, wenn die Validität der Messung gewährleistet sein soll (Knapp 2008, S. 28). Als Gütekriterien für Kategoriensysteme gelten die vollständige Erfassung der möglichen Ausprägungen des Untersuchungsobjektes sowie die Trennschärfe der Kategorien (vgl. Gerhards 2004, S. 317). Das gesamte Textmaterial bei der Inhaltsanalyse in einem Kategoriensystem aufzunehmen und damit Aussagen über den gesamten Text zu machen, kann aber nur in überschaubaren Korpora zur Anwendung gebracht werden (vgl. Früh 2001, S. 122).

Ein Aspekt der Güte von quantitativen Inhaltsanalysen betrifft die Objektivität und Systematik des Forschungsprozesses (vgl. Merten 1995, S. 49). Unter Objektivität wird nach Berelson³⁰ die exakte Operationalisierung der verwendeten Begrifflichkeiten und damit die Ermöglichung der intersubjektiven Nachvollziehbarkeit verstanden. Das Kriterium der Systematik bezieht sich auf die genaue Vorentscheidung, Beschreibung und Eingrenzung des zu analysierenden Datenmaterials (vgl. Lamnek 1995, S. 188f.). Für eine systematisches Vorgehen bei einer qualitativen Inhaltsanalyse ist für Mayring besonders die vorab festgelegte Orientierung an Regeln der Textanalyse und die Anpassung des Vorgehens an das Material sowie die Integration aller Schritte in ein Ablaufmodell wichtig (vgl. Mayring 2010, S. 48f). Dabei ist ein solches regelgeleitetes Vorgehen auch für die intersubjektive Nachvollziehbarkeit der Analyse wichtig (Mayring 2010, S. 12f.).

In Anlehnung an das Konzept der Reliabilität der quantitativen Forschung wird in der qualitativen Inhaltsanalyse die Güte der Ergebnisse durch Intercoderreliabilität messbar³¹ (vgl.

²⁹ Früh formuliert die Operationalisierung als „adäquate und vollständige Übersetzung der Bedeutungsstruktur“ des theoretischen Konstrukts in ein empirisches Forschungsmodell (vgl. Früh 2009, S. 24).

³⁰ Berelson definiert die Inhaltsanalyse als „research technique for the objective, systematic, and quantitative description of the manifest content of communication“ (Berelson 1971, S. 18).

³¹ Eigentlich handelt es sich bei der Intercoderreliabilität um die Sicherstellung der Objektivität der Messung. Also, ob verschiedene Personen dieselben Inhalte in einer Kategorie verorten (vgl. Mayring 2010, S. 117). Bei

Mayring 2010, S. 116f). Unter Intercoderreliabilität wird die Übereinstimmung der Codierung verschiedener Forscher verstanden (vgl. Früh 2009, S188; Mayring 2010, S. 51f). Dabei soll durch einen Reliabilitätstest die Vergleichbarkeit der Codierung zwischen den Forschern sichergestellt und Inkonsistenzen bei der Codierung vermieden werden (vgl. Kühne 2013, S. 36). Nach Früh ist die Anwendung einer Reliabilitätsprüfung eine notwendige, aber keine hinreichende Bedingung für die Zuerkennung der Güte einer Messung. Die Reliabilitätsprüfung stellt die Konsistenz des Messinstrumentes sicher, wodurch aber noch nicht gewährleistet ist, dass mit dem Messinstrument auch die zu messenden Sachverhalte abgebildet werden (vgl. Früh 2009, S. 189). Für die Messung der Intercoderreliabilität wird eine Stichprobe des zu codierenden Textmaterials erzeugt, welche als repräsentativ für das Textmaterial gelten kann. „Ist die Stichprobe für den Reliabilitätstest bestimmt, wird das Textmaterial von jedem Codierer einzeln unter normalen Codierbedingungen verschlüsselt“ (Früh 2009, S. 190). Anhand der codierten Stichproben können die Textstellen der Codierer anhand eines Koeffizienten gemessen und somit die Übereinstimmung der Codierung bewertet werden. Dabei wird entweder die paarweise Übereinstimmung der Codierer, die gesamte gemeinsame Übereinstimmung der Codierer oder ob die Mehrheit der Codierer übereinstimmt gemessen (vgl. Merten 1995, S. 304). Nach Früh ist eines der bekanntesten Verfahren zur Bestimmung der Intercoderreliabilität das Verfahren nach Holsti. Bei dieser Methode werden die Codierer paarweise verglichen. Dafür wird die Anzahl der einheitlichen Codierungen mit der Gesamtzahl der Codierungen beider Codierer in ein Verhältnis gesetzt und so die Bewertung der Güte der Codierung ermöglicht³² (vgl. Früh 2009, S. 189ff.). Nachdem nun die grundlegenden Gütekriterien der Inhaltsanalyse benannt sind, werden im Folgenden die Verfahren der Gütebestimmung des Text Mining näher beschrieben.

3.3 Gütekriterien des Text Mining

Die Methoden des maschinellen Lernens, welche im Text Mining für die Unterstützung einer Inhaltsanalyse zur Klassifikation eingesetzt werden, sind als überwachte und unüberwachte Ansätze beschrieben worden. Analog zu den Gütekriterien der Sozialforschung werden auch für Ansätze des maschinellen Lernens Methoden zur Qualitätssicherung eingesetzt. Im Folgenden wird das Verfahren der k-fachen Kreuzvalidierung (k-Fold Cross Validation) für die überwachten Verfahren des Text Mining vorgestellt. Weiterhin werden die Qualitätskennzahlen Fehlerrate, Precision und Recall zur Beurteilung der Güte von Klassifikationen dargestellt. Diese Maße sind verwandt zu Reliabilitätsmaßen aus den Methoden der Sozialwissenschaften und deswegen gut in die

der Reliabilitätsprüfung kann zwischen einer Intercoderreliabilität und einer Intracoderreliabilität unterschieden werden. Intercoder- und Intracoderreliabilität sind dabei zwei Maße zur Sicherstellung der Güte inhaltsanalytischer Messungen. Während bei der Intercoderreliabilität die Reliabilität zwischen verschiedenen Codierern überprüfbar wird, steht bei der Intracoderreliabilität die Reliabilität eines Codierers zu unterschiedlichen Zeitpunkten im Vordergrund. Dabei ist aber durch diese Reliabilitätstests nur die Genauigkeit der Messung, nicht aber die richtige Erfassung der Indikatoren sichergestellt (vgl. Früh 2009, S. 188f.).

³² Die Bestimmung der Intercoderreliabilität erfolgt dabei nach folgender Formel: $CR = \frac{2\ddot{U}}{C1 + C2}$. Dabei wird die doppelte Anzahl der übereinstimmenden Codierungen einer Kategorie von zwei Codierern (\ddot{U}) durch die Anzahl der Codierungen derselben Kategorie von Codierer eins (C1) und Codierer zwei (C2) geteilt. Wenn mehr als zwei Codierer beteiligt sind wird die Gesamtreliabilität der Kategorie durch den Mittelwert der einzelnen paarweisen Intercoderreliabilitäten ermittelt (vgl. Früh 2009, S. 190).

Forschungsprozesse von ePol integrierbar. Für die Produktion von Relevanzbewertungen innerhalb der Textkollektionen in ePol wird ein gemischter Ansatz verfolgt, der aus überwachten und unüberwachten Verfahren eine Relevanzbewertung definiert. Dieser Ansatz bewertet Dokumenten anhand ihres Kontextes bezüglich eines Referenzkorpus, um so für den Forschungsprozess bedeutende Dokumente zu finden. Dieser Prozess grenzt sich damit von der reinen Schlüsselwortsuche ab. Für die vorgestellte Retrieval Aufgabe wird die Qualitätskennzahl *Precision at k* ($P@k$) und das Data Fusion Verfahren als Gütekriterium vorgestellt.

In einem überwachten Verfahren stehen die Einzelobjekte, $X = (X_1, \dots, X_n)$, und deren Eigenschaften $X_n = (X_{n1}, \dots, X_{ni})$ als einzelne Instanzen zur Verfügung. Zusätzlich besitzen die Trainingsdaten zu den Features ebenfalls eine Output-Variable $Y = (Y_1, \dots, Y_k)$.³³ Der Lerner kann damit angeleitet werden, die Ausprägungen der Features mit der Output-Variable eines Objektes über eine Funktion $f(X)$ zu assoziieren. Die Assoziation der Output-Variable, unter Einbeziehung der Input-Variablen, unterliegt einem Fehler, der in Form einer Verlustfunktion³⁴ definiert wird (Hastie 2009, S. 18). Um ein Modell zu erstellen, welches die Output-Variable unbekannter Daten anhand ihrer Features bestimmt, muss das Modell antrainiert werden. Dazu werden bereits bekannte Daten verwendet, die aus der angenommenen gleichen Population stammen, wie die zu bestimmenden Daten. Output-Variablen können im Fall von Textdaten Dokumentkategorien, Dokumentsprachen oder Texttypen sein. Das Ziel beim Trainieren ist das Modell zu wählen, welches unter Einbeziehung aller Trainingsdaten einen minimalen Gesamtfehler³⁵ verursacht. Dies bedeutet, dass die abgeleiteten Regelmäßigkeiten in einem Modell die Zusammenhänge zwischen X und Y hinreichend genau, aber nicht zwingend exakt, abbilden müssen. Da die Regelmäßigkeiten meist durch Funktionen einer bestimmten Komplexität ausgedrückt werden, können nicht alle Abweichungen und Komplexitäten in den Trainingsdaten beschrieben werden und das Modell macht Fehler. Output-Variablen, die nicht auf der Intervallskala, sondern auf der Nominalskala liegen, sind kategoriale Variablen. Diese sind im Forschungsdesign von ePol ein wesentlicher Bestandteil aller überwachten Verfahren. In diesem Fall wird die Zuordnung einer Kategorie \hat{Y} aus einer Menge Kategorien Y zu einem Dokument oder Textabschnitt als Klassifikation bezeichnet. Dabei muss der Fehler von \hat{Y} im Gegensatz zu allen anderen Kategorien in Y bezüglich der Features des Dokuments oder Textabschnitts minimal sein.³⁶

³³ Beispiele hierfür können Patientendaten sein, deren Features jeweils Blutwerte, Blutdruck o.ä. sein kann.

³⁴ Eine Verlustfunktion ist ein formulierter Zusammenhang zwischen einer Funktion $f(x)$ und deren Ergebnis Y . Bei maschinellen überwachten Lernprozessen wird dieser Zusammenhang durch Trainingsbeispiele und einem Algorithmus modelliert. Dieser Prozess muss somit eine Regelmäßigkeit innerhalb der Daten interpretieren. Damit diese Regelmäßigkeit nicht nur innerhalb der Trainingsdaten funktioniert muss ein solcher Prozess generell für alle Daten einer Population funktionieren. Dies führt dazu, dass der Zusammenhang zwischen $f(x)$ und Y immer fehlerbehaftet ist. Dieser Fehler wird ausgedrückt durch $L(Y, f(X))$, die Verlustfunktion. Das L steht in diesem Fall für Loss.

³⁵ Ein minimaler Gesamtfehler wird formal durch $\operatorname{argmin} L(Y, f(X))$ ausgedrückt.

³⁶ Die Verlustfunktion wird hier ausgedrückt durch $\hat{Y}(x) = \operatorname{argmin}_{y \in Y} L(y, f(x_n))$. Die klein geschriebenen Variablen zeigen an, dass jeweils eine Kategorie und ein Datenpunkt, wie z.B. ein Dokument in die Verlustfunktion eingesetzt werden.

Bei der Modellauswahl besteht das Problem, dass die zu klassifizierenden Daten nicht bekannt sind und entschieden werden muss, welche Trainingsdaten in welcher Menge vorhanden sein müssen, um mit ausreichender Genauigkeit zu arbeiten. Wenn ein Modell selektiert wird, kann der Modellfehler anhand der Trainingsdaten gemessen werden, da die Kategorien Y für alle Trainingsdaten vorgegeben sind. Die Minimierung des Fehlers bei der Anpassung an die Trainingsdaten kann ein Modell immer komplexer werden lassen. Eine Überanpassung³⁷ führt allerdings dazu, dass ein Modell nicht für andere Daten außer den Trainingsdaten anwendbar ist. Das Modell ist in diesem Fall zu komplex und generalisiert nicht stark genug (Hastie 2009, S. 38). Aus diesem Grund werden verschiedene Modellselektionsstrategien vorgeschlagen, um die Modellqualität zu beurteilen und eine Über- oder Unteranpassung des Modells an die Trainingsdaten zu vermeiden. Eine grundsätzliche Überlegung ist die Trainingsdaten in drei verschiedene Untermengen zu verteilen. Daraus entstehen das Trainingsset, das Testset und das Validierungsset. Mit dieser Unterteilung wird verhindert, dass die Modellqualität bzw. der Modellfehler anhand der Daten evaluiert wird, die auch für das Training verwendet wurden. Die Vorgehensweise dabei ist, dass anhand der Trainingsdaten Modelle mit unterschiedlichen Parametern erstellt werden und anhand des Validierungssets³⁸ überprüft wird, welche Qualität die Modelle besitzen. Darauf aufbauend wird das beste Modell gewählt und dessen Vorhersagefehler³⁹ anhand des Testsets bestimmt (Hastie 2009, S. 222).

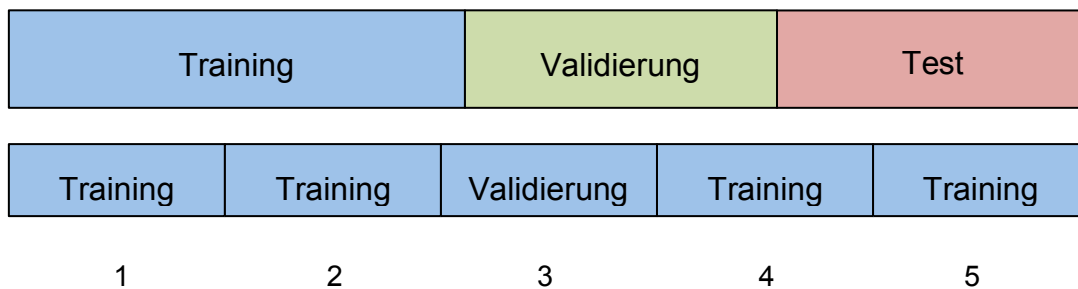


Abbildung: 2: Aufteilung der Trainingsdaten bei der Modellselektion und der n -fold Cross Validation

Modelle mit hoher Komplexität bzw. hoher Dimensionalität neigen dazu sehr viele Trainingsdaten zu benötigen, um gut generalisierbare Modelle selektieren zu können (Hastie 2009, S. 229).⁴⁰ In manchen Fällen, so auch in typischen Problemstellungen der Dokumentklassifikation in den Geistes- und Sozialwissenschaften⁴¹, stehen allerdings nicht genügend Trainingsdaten zu Verfügung, um eine generalisierbares Modell für die oft sehr abstrakten Kategorien zu wählen. Die Selektion und Qualitätsbestimmung des Modells wird in diesem Fall schwer, da die Wahl eines Validierungs- und Testsets die eigentliche Trainingsmenge verringert und ein Modell die Folge wäre, welches einen kleinen Trai-

³⁷ In der Literatur wird dieser Effekt meist mit dem englischen Wort „Overfitting“ angesprochen.

³⁸ Die Fehler unterschiedlicher Modelle bzgl. der Trainingsdaten und des Validierungssets werden als Trainingsfehler bezeichnet.

³⁹ Der Vorhersagefehler wird auch als Generalisierungsfehler bezeichnet und gibt an, welche Qualität das beste zu wählende Modell in Bezug auf unbekannte Daten erreichen kann.

⁴⁰ Diese Fälle werden als „High Variance Problem“ bezeichnet, da unterschiedliche Untermengen der Trainingsdaten jeweils andere optimale Modelle vorschlagen und dadurch eine große Auswahl an Modellen besteht. Dieser Effekt wird geringer, je größer die Trainingsdatenmenge gewählt wird.

⁴¹ So auch im vorgeschlagenen Forschungsprozess, bei dem für die Klassifikation nur begrenzt annotierte neoliberale Argumente als Trainingsdaten zur Verfügung stehen.

ningsfehler, aber einen hohen Generalisierungsfehler aufweisen würde. Um mit wenigen Trainingsdaten Modelle zu trainieren und den Erwartungswert des Generalisierungsfehlers dennoch schätzen zu können eignen sich Cross-Validation Verfahren. Bei der K -fachen Kreuzvalidierung wird der Trainingsdatensatz in K gleiche Untermengen, wie in Abbildung 2 dargestellt, unterteilt. Es werden nun K Modelle unter Einbeziehung aller Trainingsdaten, außer dem jeweiligen Teil k , trainiert.⁴²

Als Fehler wird der Mittelwert aller Trainingsfehler aus den K trainierten Modellen angegeben. Der geschätzte Fehler entspricht bei dieser Prozedur dem Erwartungswert des Generalisierungsfehlers und eben nicht dem Trainingsfehler, der nur über die Einbeziehung der Trainingsdaten errechnet wird (Hastie 2009, S. 242ff). Diese Prozedur bietet sich immer dann an, wenn durch die Unterteilung in Trainings- und Validierungsset die Trainingsdatenmenge zu gering werden würde.

Um den Fehler für nominale Output-Variablen schätzen zu können, wird eine Konfusionsmatrix, wie in Tabelle 1 dargestellt, herangezogen. Darin werden unterschiedliche Zählungen unterschieden. Diese teilen sich danach auf, ob die Vergabe oder die Nichtvergabe einer Kategorie korrekt vorgenommen wurde. Wird beispielsweise für ein Dokument in einer Dokumentkollektion eine Kategorie vergeben, obwohl das Dokument dieser Kategorie nicht zugeordnet werden kann, so wird die Zählung für die falschen Positivzuordnungen (FP) um 1 erhöht. Da bei einem sogenannten 0/1-Verlust⁴³ kein direkter Fehlerwert berechnet werden kann, werden alle durch den Klassifikator verursachten Fehler gleich schlecht behandelt (Alpaydin 2008, S. 354). Dabei werden die Fehlerrate (Error Rate), Genauigkeit (Accuracy), Precision und Recall als Gütekriterien unterschieden.⁴⁴

- Die Fehlerrate bestimmt den Anteil der falsch vergebenen (FP) und fälschlich nicht vergebenen Kategorien (FN) an unter allen (N) vorgenommenen Klassifikationen.
- Die Genauigkeit gibt dagegen den Anteil der richtig vergebenen (WP) und richtig nicht vergebenen (WN) Kategorien unter allen (N) Klassifikationen an.
- Die Precision repräsentiert den Anteil der richtig vergebenen (WP) Kategorien hinsichtlich aller vergebenen Kategorien inklusive der falsch zugeordneten Kategorien (FP). Diese Kennzahl gibt somit an, welcher Anteil aller vergebenen Kategorien (WP + FP) richtig ist.
- Mit dem Recall wird der Anteil der richtig vergebenen (WP) Kategorien an allen zu vergebenden Kategorien gemessen. Dazu zählen auch die fälschlicherweise nicht

⁴² Die Kreuzvalidierung wird angegeben als $CV(f, \alpha) = \frac{1}{N} \sum_{i=1}^N L(Y_i, f^{-k(i)}(X_i, \alpha))$. Dabei ist f^{-k} das Modell für die Trainingsmenge exklusive der Menge k . Der Generalisierungsfehler wird als Mittelwert über alle K Modelle angegeben. Der Parameter α repräsentiert einen Modellparameter, der angepasst werden kann, um den Generalisierungsfehler zu minimieren (Hastie, S. 242ff.).

⁴³ Der Begriff bezieht sich darauf, dass bei der Vergabe einer Kategorie bei der Klassifikation kein Wert oder eine Zahl geschätzt wird, sondern die Unterscheidung zwischen zwei Zuständen (Kategorie vergeben oder nicht vergeben) stattfindet. Die Vergabe ist entweder richtig oder falsch.

⁴⁴ $Error\ Rate = \frac{|FN| + |FP|}{N}$, $Accuracy = \frac{|WP| + |WN|}{N}$, $Precision = \frac{|WP|}{|TP| + |FP|}$, $Recall = \frac{|WP|}{|WP| + |FN|}$

zugeordneten (FN). Dieses Maß repräsentiert damit welcher Anteil der zu vergebenden Kategorien (WP + FN) überhaupt auch zugeordnet wurde.

	vorhergesagte Kategorie	
wahre Kategorie	ja	nein
ja	WP: wahres Positiv	FN: falsches Negativ
nein	FP: falsches Positiv	WN: wahres Negativ

Tabelle 1: Tabelle übernommen aus Alpaydin (2008, S. 354)

Anhand der vorgestellten Kennzahlen ist es möglich die Genauigkeit und Leistungsfähigkeit einer Klassifikation zu beurteilen. Zusätzlich kann durch die Optimierung der Precision dafür gesorgt werden, dass möglichst alle vergebenen Kategorien auch korrekt sind. Dies steht immer im Widerspruch mit der Optimierung des Recall, sodass möglichst alle zu vergebenden Label gefunden werden. Dadurch macht der Klassifikationsprozess aber auch mehr Fehler. Welche Kennzahl optimiert wird, hängt immer an der jeweiligen Aufgabe des Forschungsprozesses. Für die Argumentklassifikation im Projekt ePol ist eine Optimierung der Precision anzustreben, sodass die vergebenen Kategorien wenige Fehler enthalten. Das Zusammenspiel beider Kennzahlen wird durch deren harmonisches Mittel als F_1 -Maß ausgedrückt, wenn die Entscheidung nicht eindeutig auf die Optimierung eines der beiden Ziele erfolgen kann.

Für die Identifikation relevanter Dokumente wird im Projekt ePol ein Retrieval Prozess eingesetzt, um eine sortierte Liste der Dokumente in der Datenbank zu erstellen. Die Sortierung erfolgt anhand einer Relevanzbewertung innerhalb der Dokumente und prüft deren Gehalt an neoliberalen und argumentativen Zusammenhängen. Für die Evaluierung von Retrieval Prozessen müssen andere Verfahren herangezogen werden. Wie angesprochen, bestehen die Ergebnisse eines Retrieval Prozesses nicht aus abgeschlossenen Dokumentmengen und daran vorgenommenen Klassifikationen, sondern meist aus sortierten Listen, deren Länge nicht bekannt ist. Hier wird für den Forschungsprozess in ePol zwischen zwei Verfahren zur Evaluierung unterschieden. Einerseits kann das Konzept der Precision auf eine Anzahl von Dokumenten an verschiedenen Rängen einer Retrieval-Liste übertragen werden. Dafür wird die Precision innerhalb einer Dokumentmenge (z.B. 10) ab einem bestimmten Rang in der sortierten Liste gemessen. Das Verfahren wird als *Precision at k* ($P@k$) bezeichnet. Damit kann bestimmt werden, in welcher Weise die Qualität der Liste in den hinteren Rängen nachlässt. Dieses Verfahren eignet sich sehr gut für manuelle Evaluierungen, wenn für die wahre Ergebnismenge kein Testsatz vorhanden ist und die Relevanz der Dokumente in der sortierten Liste von einem Nutzer des Retrieval beurteilt werden muss. Dafür müssen nicht viele Daten betrachtet werden, dennoch sind Aussagen über die Qualität der gesamten Liste möglich.

Wenn keine ausreichenden Testdatensätze vorhanden sind, ist die Wahl einer bestimmten Parametrisierung für ein Retrieval System ebenfalls schwierig. Um die Leistungsfähigkeit verschiedener Retrieval Systeme für eine zu lösende Aufgabe zu vergleichen, kann mit einem sogenannten Data Fusion Verfahren (vgl. Nuray et al. 2006) gearbeitet werden. Diese Verfahren erzeugen aus allen Retrieval Systemen, die getestet

werden, eine verschmolzene Ergebnismenge. Dies ist vorteilhaft, da theoretisch alle Systeme unterschiedliche Ergebnisse bzw. Ränge der Dokumente in den Ergebnislisten liefern. Aus dieser Menge kann geschätzt werden, welches System die meisten Treffer in dieser Menge und somit mehr relevante Dokumente hervorbringt als andere Systeme (vgl. Wiedemann/Niekler 2014).

3.4 Anwendung der Gütekriterien auf das Forschungsdesign

Nach der Beschreibung der Gütekriterien für die relevanten Forschungsmethoden, folgt nun deren Anwendung auf das in Kapitel 2.4 beschriebene Forschungsdesign von ePol. Dafür wird erklärt, wie die Güte erhoben, im Forschungsablauf verortet und die Validität des Forschungsprozesses nachzuweisen ist. Durch die exemplarische Anwendung der Gütekriterien auf das Forschungsdesign wird aufgezeigt, wie für diesen Fall einer computergestützten qualitativen und quantitativen Inhaltsanalyse die Qualitätssicherung durchzuführen ist.

Da die Übernahme aller angesprochenen Gütekriterien der verschiedenen Forschungsrichtungen bei einem Forschungsdesign, das verschiedene Methoden kombiniert, nicht vollumfänglich möglich ist, müssen die Gütekriterien an den Forschungsprozess angepasst bzw. passend für die einzelnen Teilaufgaben zugeschnitten werden. Die für ein Forschungsdesign angewendeten Gütekriterien sollten dokumentiert werden, um die Güte der Ergebnisse zu sichern und den Standards der Forschung zu genügen. Die im Rahmen der allgemeinen Gütekriterien angesprochene Dokumentation des Forschungsprozesses ist auch für die computergestützte Inhaltsanalyse notwendig, um die intersubjektive Nachvollziehbarkeit der Ergebnisse zu gewährleisten. Dafür wird die Forschung von ePol in wissenschaftlichen Aufsätzen, Sammelbänden und Discussion Papers dokumentiert und so der Forschungsstand und die Methode der Öffentlichkeit zugänglich gemacht.

Die Gütekriterien der Systematik und der Regelgeleitetheit einer Inhaltsanalyse wurden von unserer Seite durch die Formulierung eines Forschungsdesigns erfüllt. Dabei werden die Teilaufgaben der Argumentationsidentifikation, mit ihren quantitativen und qualitativen Elementen, strukturiert und in ein Ablaufmodell integriert. Zur Qualitätssicherung in den drei identifizierten Teilaufgaben werden die Gütekriterien aus den unterschiedlichen Bereichen der qualitativen und quantitativen Sozialforschung, der Inhaltsanalyse und des Text Mining je nach ihrer Eignung im Prozess eingesetzt.⁴⁵

Die erste Teilaufgabe des Forschungsdesigns, die diktionärbasierte Vorselektion des Zeitungskorpus, ist eine angemessene Methode und ermöglicht einen validen Einstieg in die Analyse. Diese Vorselektion ist ein objektives und intersubjektiv nachvollziehbares Verfahren, da durch die ungefilterten Ansichten und Aussagen der Theoretiker des Neoliberalismus im *Wörterbuch des Neoliberalismus* keine subjektiven Vorannahmen der beteiligten Forscher in den Prozess eingebracht wurden. Die Verwendung und Erstellung der Diktionäre ist in den Discussion Papers des Projekts dokumentiert (Wiedemann/Niekler 2014; Dumm/Lemke 2013).

⁴⁵ Zur Verdeutlichung wird an dieser Stelle auf Abbildung 1 hingewiesen, in der die einzelnen Evaluierungsschritte, und damit Zeitpunkte für die Erhebung der Gütekriterien, dargestellt sind.

Vor der Diktionärerstellung wurden, in einem diskursiven Prozess, die Grundannahmen des Neoliberalismus durch die Projektteilnehmer der Politischen Theorie analysiert. Dabei wurde die Mont Pelerin Society als theoretische Keimzelle des Neoliberalismus im Sinne der Chicago School identifiziert und so als Referenzkorpus für das *Wörterbuch des Neoliberalismus* bestimmt (vgl. Schaal/Ritzi 2012; Lemke 2012). Die vorangegangenen Analysen der Grundlagen des Neoliberalismus bilden damit die theoretische Fundierung der Forschungsfrage und somit die Grundlage der Operationalisierung der Ökonomisierungsthese.

Die Anwendung des Diktionärverfahrens wurde mit dem aus dem Information Retrieval stammenden *Precision at k* Verfahren evaluiert. Für die Erstellung des Diktionärs sind verschiedene Verfahren der Textextraktion erprobt worden. Wobei die Anwendung unterschiedlich komplexer sprachlicher Elemente z.B. Unigramme (Einworteinheiten), Bigramme (Zweiworteinheiten) oder Wortkookkurrenzen als Auswahlkriterium geprüft wurde. Die Ergebnisse dieser Auswahlverfahren sind manuell anhand der Überschriften und des Inhalts in ihrer Relevanz bewertet worden. Das Diktionärverfahren bei dem die Existenz von Unigrammen und deren Kookkurrenz untereinander als Bewertungsmaßstab der Relevanz eines Dokuments dienen, ist nach der Evaluierung seiner Validität durch das *Precision at k* Verfahren eingesetzt worden (Wiedemann/Niekler 2014).

Für die zweite Teilaufgabe, die manuelle Identifikation von neoliberalen Argumenten, wurde ein Kategoriensystem ausgearbeitet und zur Erhöhung der Reliabilität in einem Codebuch dokumentiert. Das Codebuch enthält damit die inhaltliche und formale Dimension von Argumenten im Verständnis der Forschungsfrage. Die Validität der annotierten Argumente wird außerdem erreicht, indem die formale Definition der Argumentkategorien eng an linguistische Theorien und die inhaltliche Definition an die Sprachverwendung neoliberaler Protagonisten angebunden ist. Das Kategoriensystem wird zur Bewertung der Intercoderreliabilität und damit zur Bewertung der Codebuchqualität einem Test unterzogen, wobei die Reliabilität nach Holsti bestimmt wurde. Das strukturierte Vorgehen bei der Kategorienerstellung und deren Anwendung ist damit durch eine formale Definition der Argumente und einen standardisierten Codierprozess gewährleistet.

Im dritten Teil des Forschungsdesigns, der (halb-) automatischen Argumentidentifikation zur Erzeugung des Testdatensatzes, wird die Fehlermessung mit Precision und Recall zur Evaluierung eingesetzt. Da im Forschungsprozess die automatisch erkannten und manuell validierten Argumente automatisch in die Trainingsmenge für einen neuen Klassifikator einfließen, verbessert sich die Erkennungsrate und die Güte der so iterativ erzeugten Argument-Datensätze. Durch die manuelle Prüfung, der vom Klassifikator vorgeschlagenen Textabschnitte, wird die Validität der (halb-) automatischen Messinstrumente sichergestellt, denn die Fehlermessung mit Precision und Recall sichert zwar die Qualität des Klassifikators bezüglich der Trainingsbeispiele ab, bewertet jedoch nicht die daraus resultierende Relevanz für die Forschungsfrage. Durch den manuellen Eingriff in die Klassifikationsergebnisse durch die Forscher bleibt der theoriegeleitete Blick auf die Ergebnisse erhalten und der Algorithmus dominiert nicht die Ergebnisse.

4. Fazit

Dieses Paper hat aufgezeigt, wie die Verfahren des Text Mining mit den klassischen Methoden der Inhaltsanalyse kombiniert werden können. Durch die sequenzielle Anwendung von Verfahren beider Fachrichtungen lassen sich kombinierte Forschungsdesigns mit emergenten Effekten entwickeln. Der große Vorteil einer solchen Verbindung von Text Mining mit klassischen Methoden der Sozialwissenschaften besteht dann in der Beantwortung von Fragestellungen, die aufgrund der Menge an Textdaten nicht mehr manuell zu bearbeiten sind. Auf diese Weise sind zukünftige Forschungsvorhaben in der Lage, die durch die Digitalisierung großer Textmengen für die wissenschaftliche Forschung entstandenen Möglichkeiten für weitreichende Untersuchungen zu nutzen. Denn die Verfahren des Text Mining sind in ihrer Verarbeitungskapazität den bereits etablierten Methoden der Computerunterstützung (QDA-Software) in der Sozialwissenschaft überlegen. In Kapitel 2.4 wurde ein Forschungsdesign vorgestellt, das exemplarisch eine Hypothesenprüfung durch den Einsatz von Text Mining Verfahren und inhaltsanalytischer Methoden ermöglicht. Dieses Design basiert auf der Aufteilung des Forschungsprozesses in Teilaufgaben. Der kombinierte Einsatz der vorgestellten Methoden und Verfahren ist aber nicht auf das beschriebene Forschungsdesign beschränkt. Andere Designs, welche die vorgestellten Methoden und Verfahren beinhalten, sind je nach Forschungsfrage anders umsetzbar.

In der Diskussion der Gütekriterien der beiden Fachrichtungen wurde gezeigt, dass die verschiedenen Forschungsansätze unterschiedliche Gütekriterien bei der Qualitätssicherung der Forschung nutzen. Daraus folgt, dass es nicht zur Anwendung eines klassischen Kanons von Gütekriterien kommen kann, wenn unterschiedliche Forschungsansätze kombiniert werden. Dies bedeutet zwar einerseits, dass nicht alle mit unserem Verfahrensrepertoire durchgeführten Analysen mit demselben Set an Gütekriterien zu evaluieren sind. Aber andererseits liegt mit den vorgestellten Verfahren und Kriterien zur Bewertung von einzelnen Aspekten ein Bewertungskatalog für Teilaufgaben einer computergestützten qualitativen und quantitativen Inhaltsanalyse vor.

Für die Erstellung eines adäquaten Forschungsdesigns einer computergestützten qualitativen und quantitativen Inhaltsanalyse ist die Trennung der einzelnen Teilaufgaben zu Beginn des Forschungsprozesses und deren Integration in ein Ablaufmodell von essentieller Bedeutung. Zusammenfassend kann die Identifikation angemessener Methoden und Gütekriterien für andere Forschungsfragen in drei Schritte untergliedert werden:

1. Die gesamte Forschungsfrage muss in geeignete Teilaufgaben zerlegt werden. Dies beinhaltet die Analyse des Forschungsprozesses und Identifikation von Teilergebnissen, die zur Beantwortung der Forschungsfrage beitragen.
2. In jeder der Teilaufgaben muss die Forschungsmethode identifiziert und festgelegt werden. Es muss definiert werden, ob eine Teilaufgabe quantitativer oder qualitativer Natur ist und ob die Erkenntnisse deduktiv oder induktiv erzeugt werden.
3. Durch die genaue Einteilung der Teilaufgaben ist es möglich, die aus den Methodenkatalogen der Disziplinen gültigen Mess-, Bewertungs-, Güte- oder Evaluierungsverfahren zu identifizieren und einzusetzen.

Durch ein so erstelltes Forschungsdesign lassen sich die Vorteile der Verfahren und Methoden beider Fachrichtungen kombinieren. So ist es einerseits möglich die Tiefenschärfe der qualitativen Methoden zu nutzen und durch eine *close reading* Perspektive nahe am Text zu arbeiten. Andererseits ist durch die (halb-) automatische computergestützte Vorgehensweise mit Verfahren des Text Mining die Bearbeitung großer Textkorpora möglich. Durch die manuelle und computergestützte Evaluation der einzelnen Schritte im Forschungsablauf bleibt der Einfluss der Forscher auf die Qualität und Validität der Ergebnisse sichergestellt.

Literatur

- Alpaydin, Ethem (2008): Maschinelles Lernen. München: Oldenbourg.
- Atteslander, Peter; Cromm, Jürgen (2010): Methoden der empirischen Sozialforschung. 13., neu bearb. und erw. Aufl. Berlin: Schmidt (ESV basics).
- Baeza-Yates, Ricardo; Ribero-Neto, Berthie (2011): Modern information retrieval: the concepts and technology behind search. 2. Aufl. New York: Addison Wesley.
- Behrens, Maria (2003): Quantitative und qualitative Methoden in der Politikfeldanalyse. In: Klaus Schubert und Nils C. Bandelow (Hg.): Lehrbuch der Politikfeldanalyse. München [u.a.]: Oldenbourg, S. 203–235.
- Berelson, Bernard (1971): Content Analysis. In communication Research. New York, Hafner Publishing Company.
- Blei, David. M.; Ng, Andrew. Y.; Jordan, Michael. I. (2003): Latent dirichlet allocation. In: The Journal of Machine Learning Research, Jg. 3, S. 993–1022 .
- Brown, Wendy (2006): American Nightmare: Neoliberalism, Neoconservatism, and De-Democratization. In: Political Theory, 34 (6), S. 690-714.
- Buch, Bastian (2008): Text Mining. Zur automatischen Wissensextraktion aus unstrukturierten Dokumenten. Saarbrücken: VDM Verl. Müller.
- Christmann, Gabriela B: „Inhaltsanalyse“. In: Ayaß, Ruth (2006): Qualitative Methoden der Medienforschung. Orig.-Ausg. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl (Rororo Rowohlt's Enzyklopädie, 55665), S. 274-292.
- Crouch, Colin (2008): Postdemokratie, Frankfurt (Main).
- Crouch, Colin (2011): Das befremdliche Überleben des Neoliberalismus. Postdemokratie II, Berlin
- Diaz-Bone, Rainer; Schneider, Werner (2004): Qualitative Datenanalysesoftware in der sozialwissenschaftlichen Diskursanalyse - Zwei Praxisbeispiele. In: Reiner Keller, Andreas Hierseland, Werner Schneider und Willy Viehöver (Hg.): Handbuch sozialwissenschaftliche Diskursanalyse. Band 2: Forschungspraxis. 2. Aufl. Wiesbaden: VS Verl. für Sozialwiss, 457-494.
- Diekmann, Andreas (2003): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek: Rowolt.
- Dumm, Sebastian; Lemke, Matthias (2013): Argumentmarker. Definition, Generierung und Anwendung im Rahmen eines semi-automatischen Dokument-Retrieval-Verfahrens. Hamburg / Leipzig. (Schriftenreihe des Verbundprojekts Postdemokratie und Neoliberalismus, Discussion Paper Nr. 3.)
- Feldman, Ronen; Sanger, James (2008): The text mining handbook. Advanced approaches in analyzing unstructured data. Reprinted. Cambridge: Cambridge Univ. Press. Online.
- Flick, Uwe (2000): Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften. 5. Aufl. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl.
- Flick, Uwe (2007): Zur Qualität qualitativer Forschung – Diskurse und Ansätze. In: Udo Kuckartz, Heiko Grunenberg und Thorsten Dresning (Hg.): Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis. Wiesbaden: VS, Verl. für Sozialwissenschaften, S. 188–209.

- Flick, Uwe (2011): Triangulation Eine Einführung. Wiesbaden: VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden, Wiesbaden.
- Friedrichs, Jürgen (1973): Methoden empirischer Sozialforschung. Reinbek: Rowohlt.
- Früh, Werner (2001): Kategorienexploration bei der Inhaltsanalyse. Basiswissengeleitete offene Kategorienfindung (BoK). In: Werner Wirth und Edmund Lauf (Hg.): Inhaltsanalyse. Perspektiven, Probleme, Potentiale. Köln: Halem, S. 117–139.
- Früh, Werner (2009): Inhaltsanalyse. Theorie und Praxis. 1. Aufl. Konstanz: UVK Verl.-Ges (UTB Medien- und Kommunikationswissenschaft, Psychologie, Soziologie, 2501).
- Gerhards, Jürgen (2004): Diskursanalyse als systematische Inhaltsanalyse. Die öffentliche Debatte über Abtreibungen in den USA und in der Bundesrepublik Deutschland im Vergleich. In: Reiner Keller, Andreas Hiesland, Werner Schneider und Willy Viehöver (Hg.): Handbuch sozialwissenschaftliche Diskursanalyse. Band 2: Forschungspraxis. 2. Aufl. Wiesbaden: VS Verl. für Sozialwiss, S. 299–324.
- Hajer, Maarten (2008): Diskursanalyse in der Praxis: Koalitionen, Praktiken und Bedeutung; in Janning, Frank; Toens, Katrin (2008): Die Zukunft der Policy-Forschung. 1. Aufl. s.l: VS Verlag für Sozialwissenschaften, S. 211-223
- Hastie, Trevor; Tibshirani, Robert; Friedman, J. H. (2009): The elements of statistical learning: data mining, inference, and prediction. 2. Aufl. New York: Springer.
- Keller, Reiner (2004): Diskursforschung. Eine Einführung für SozialwissenschaftlerInnen. Opladen: Leske+ Budrich
- Keller, Reiner; Hiesland, Andreas; Schneider, Werner; Viehöver, Willy (Hg.) (2004): Handbuch sozialwissenschaftliche Diskursanalyse. Band 2: Forschungspraxis. 2. Aufl. Wiesbaden: VS Verl. für Sozialwiss.
- Keller, Reiner; Viehöver, Willy (2006): Diskursanalyse; in Behnke, Joachim, Gschwend, Thomas; Schindler, Delia; Schnapp, Kai-Uwe : Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren, Baden-Baden, Nomos, S. 103-112.
- Knapp, Werner (2008): Die Inhaltsanalyse aus linguistischer Sicht. In: Philipp Mayring (Hg.): Die Praxis der qualitativen Inhaltsanalyse. 2. Aufl. Weinheim, Basel: Beltz, S. 20–36.
- Krippendorff, K. (2004). Content analysis : an introduction to its methodology. 2. Aufl. Thousand Oaks, Calif.: Sage.
- Kuckartz, Udo (2007): QDA Software im Methodendiskurs: Geschichte, Potentiale, Effekte; in Kuchartz, Udo; Grunenberg, Heiko; Dresing, Thorsten: Qualitative Datenanalyse: computergestützt, 2. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 15-31.
- Kuckartz, Udo (2010): Einführung in die computergestützte Analyse qualitativer Daten. 3., aktualisierte Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kuckartz, Udo(2006): Computergestützte qualitative Inhaltsanalyse; in Behnke, Joachim, Gschwend, Thomas; Schindler, Delia; Schnapp, Kai-Uwe : Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren, Baden-Baden, Nomos, S. 81-91.
- Kühne, Rinaldo (2013): Konzeptspezifikation und Messung. In: Wiebke Möhring und Daniela Schlütz (Hg.): Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft. Wiesbaden: Springer VS, S. 23–40.

- Lamnek, Siegfried (1995): *Qualitative Sozialforschung. Methoden und Techniken*. 3. Aufl. (Qualitative Sozialforschung, 2).
- Landwehr, Achim (Hg.) (2010): *Diskursiver Wandel*. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage, Wiesbaden.
- Lemke, Matthias (2012): *Die Ökonomisierung des Politischen. Entdifferenzierungen in kollektiven Entscheidungsprozessen*. Schriftenreihe des Verbundprojekts Postdemokratie und Neoliberalismus, Discussion Paper Nr. 2 Helmut-Schmidt-Universität Hamburg (UniBw) und Universität Leipzig.
- Lemke, Matthias; Stulpe, Alexander (2014): *Text und soziale Wirklichkeit. Theoretische Grundlagen und empirische Anwendung durch Text Mining Verfahren am Beispiel des Bigrams ‚soziale Marktwirtschaft‘*, erscheint in *Zeitschrift für Germanistische Linguistik*, Themenheft „Automatisierte Textanalyse“.
- Lüders, Christian (2011): *Gütekriterien*. In: Bohnsack, Ralf; Marotzki, Winfried; Meuser, Michael (2011): *Hauptbegriffe qualitativer Sozialforschung*. 3. Aufl. Opladen: Budrich (UTB Erziehungswissenschaft, Sozialwissenschaft, 8226), S.80-82.
- Mayring, Philipp (2002): *Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken*. 5. Aufl. Weinheim: Beltz-Verl.
- Mayring, Philipp (2010): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 11. Aufl. Weinheim: Beltz (Studium Paedagogik).
- Merten, Klaus (1995): *Inhaltsanalyse*. Opladen: Westdeutscher Verl.
- Moretti, Franco (2000): *Conjectures on World Literature*. In: *New Left Review* (1), S. 54–68.
- Moretti, Franco (2007): *Graphs, maps, trees. Abstract models for literary history*. London; New York: Verso.
- Niehr, Thomas (1999): *Halbautomatische Erforschung des öffentlichen Sprachgebrauchs oder Vom Nutzen computerlesbarer Textkorpora*; in *Zeitschrift für germanistische Linguistik (ZGL)* 27.2, S.205-214.
- Niekler, Andreas; Wiedemann, Gregor / Heyer, Gerhard (2014): *Leipzig Corpus Miner - A Text Mining Infrastructure for Qualitative Data Analysis*. In: *Terminology and Knowledge Engineering 2014*, Berlin.
- Nuray, R., Can, F. (2006): *Automatic ranking of information retrieval systems using data-fusion*. *Information Processing & Management* 42(3), 595–614.
- Rancière, Jacques (1995): *La Méésentente: Politique et philosophie*, Paris.
- Ritzi, Claudia (2013): *Die Postdemokratisierung politischer Öffentlichkeit*, Wiesbaden.
- Schaal, Gary S.; Ritzi, Claudia (2012): *Neoliberalismus und Postdemokratie. Bausteine einer kritischen Gesellschaftstheorie*. In: *Ethik und Gesellschaft*, 2/2012.
- Schnapp, Kai-Uwe; Schindler, Delia; Gschwend, Thomas; Behnke, Joachim (2006): *Qualitative und quantitative Zugänge: Eine integrative Perspektive*. In: Joachim Behnke (Hg.): *Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren*. 1. Aufl. Baden-Baden: Nomos (Forschungsstand Politikwissenschaft), S. 11–26.
- Schnell, Rainer; Hill, Paul Bernhard; Esser, Elke (1999): *Methoden der empirischen Sozialforschung*. 6. völlig überarbeitete and erw. Aufl. München: R. Oldenbourg.
- Schulz, Winfried (1989): *Massenmedien und Realität. Die "ptolemäische" und die "kopernikanische" Auffassung*. In: Max Kaase und Winfried Schulz (Hg.):

- Massenkommunikation. Theorien, Methoden, Befunde. Wiesbaden: Westdeutscher Verlag (Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft, 30), S. 135–149.
- Schwab-Trapp, Michael (2011): Diskursanalyse, in :Bohnsack, Ralf; Marotzki, Winfried; Meuser, Michael (2011): Hauptbegriffe qualitativer Sozialforschung. 3. Aufl. Opladen: Budrich (UTB Erziehungswissenschaft, Sozialwissenschaft, 8226), S.35-39.
- Steinke, Ines (1999): Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung. Weinheim: Juventa.
- Steinke, Ines (2007): Qualitätssicherung in der qualitativen Forschung. In: Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis. Wiesbaden: VS, Verl. für Sozialwissenschaften, S. 176–187.
- Steinke, Ines (2010): Gütekriterien qualitativer Forschung. In: Uwe Flick, Ernst von Kardorff und Ines Steinke (Hg.): Qualitative Forschung. Ein Handbuch. 8. Aufl. Reinbek bei Hamburg: Rowohlt Taschenbuch-Verl (Rororo Rowohlts Enzyklopädie, 55628), S. 319–331.
- Toulmin, Stephen Edelston (1996): Der Gebrauch von Argumenten. 2. Aufl. Weinheim: Beltz, Athenäum (Neue wissenschaftliche Bibliothek).
- Wiedemann, Gregor / Lemke, Matthias / Niekler, Andreas (2013): Postdemokratie und Neoliberalismus – Zur Nutzung neoliberaler Argumentation in der Bundesrepublik Deutschland 1949–2011. In: Zeitschrift für Politische Theorie, 4(1), 99–115.
- Wiedemann, Gregor / Niekler, Andreas (2014): Document Retrieval for Large Scale Content Analysis using Contextualized Dictionaries .In: Terminology and Knowledge Engineering 2014, Berlin.
- Wilke, Jürgen (1999): Mediengeschichte der Bundesrepublik Deutschland. Bonn: Bundeszentrale für Politische Bildung (Schriftenreihe, Bd. 361).

Schriftenreihe des Verbundprojekts
„Postdemokratie und Neoliberalismus“

Auswahl der Diskussionspapiere

(Stand Juli 2014)

- Nr. 1 Neoliberalismus und Postdemokratie: Bausteine einer kritischen Gesellschaftstheorie.
Gary S. Schaal & Claudia Ritzl
- Nr. 2 Die Ökonomisierung des Politischen.
Entdifferenzierungen in kollektiven Entscheidungsprozessen
Matthias Lemke
- Nr. 3 Argumentmarker. Definition, Generierung und Anwendung im Rahmen eines semi-automatischen Dokument-Retrieval-Verfahrens
Sebastian Dumm & Matthias Lemke
- Nr. 4 Methoden und Gütekriterien. Computergestützte Diskurs- und Inhaltsanalysen zwischen Sozialwissenschaft und Automatischer Sprachverarbeitung
Sebastian Dumm & Andreas Niekler

Die Arbeitspapiere können bestellt werden/The discussion papers can be ordered:

Helmut-Schmidt-Universität Hamburg
Forschungsprojekt Postdemokratie und Neoliberalismus
Professur für Politische Theorie und Ideengeschichte
z. H. Susanne Kirst
Holstenhofweg 85
22043 Hamburg